THIS WEEK

Facing up to flu

The potential for mutant–flu research to improve public health any time soon has been exaggerated. Timely production of sufficient vaccine remains the biggest challenge.

Amid the scientific controversy over lab-created strains of the H5N1 avian influenza virus that can skip between mammals, it is easy to lose sight of an important public-health question: what will help the wider world to prepare for a flu pandemic? The question is crucial, because when it comes to setting priorities, the fuss over how to regulate the controversial research must not be allowed to distract from a much bigger concern. The world is ill-prepared for a severe flu pandemic of any type. In particular, it cannot yet produce enough vaccine to protect more than just a small proportion of people.

The problem was demonstrated by the 2009 pandemic of H1N1 flu. Vaccines only became available months after the outbreak began, and after the first wave had peaked in many countries. Health systems were stretched despite the relative mildness of the pandemic. The mutant-flu research does nothing to prevent a repeat of this situation.

Research to create mammalian-transmissible strains is vital basic science that could deepen our understanding of flu viruses, and of what allows a virus to jump from other species and spread easily in humans. These insights may one day produce better ways to tackle a pandemic, including ones we cannot picture today. But scientists need to be more modest and realistic with their claims about the short-term public-health benefits of such research, and provide better explanations that include the caveats.

For example, many commentators say that the biggest public-health benefit promised by the research is in the field of disease surveillance. The experiments reveal one combination of mutations that allowed the H5N1 virus to jump between species and then spread; in theory, animal-health experts can now watch out for these mutations in affected animals such as pigs and birds.

In practice, the immediate benefits are minimal. Surveillance of influenza in animals is slow and patchy at best, and follow-up sequencing of samples more so. And the mutations that we know about are likely to be outnumbered by those about which we are still ignorant.

Consider H5N1 in pigs. There is almost no systematic flu surveillance in the animals (see *Nature* **459,** 894–895; 2009). Infections are infrequent, symptoms are mild and the pig industry is concerned that talk of swine flu could unfairly taint the image of pork. As a result, the world's one billion or so pigs have yielded partial DNA sequences of just 24 H5N1 isolates, meaning that were a pandemic H5N1 virus to emerge from pigs, just as H1N1 did in 2009, there would be little or no possibility of detecting it in advance.

That does not mean that the idea of using the mutant-flu research to improve surveillance is without merit; far from it. Further work could yield a more comprehensive bank of mutations, and greater investment could create specialized centres to screen more samples in affected countries, in real time. Improving flu-virus surveillance should be a public-health priority, but international groups and governments have, in the past, been reluctant to fund it adequately. If the world is serious about preparing for a pandemic, this must change.

Done properly, surveillance could one day give early warning of an approaching pandemic. What then?

At present, such advance knowledge would make little difference to the world's limited abilities to manufacture and distribute vaccines. Current techniques can produce vaccine only six months after a pandemic emerges. Doing so faster and in much larger quantities is the most urgent public-health priority when it comes to planning for the next pandemic.

> *"The mutant–flu studies offer no serious immediate application in vaccine research."*

The mutant-flu studies contribute little to this goal. They offer no serious immediate application in vaccine research (see page 142). Any benefits to drug development — which are important, but less so than churning out vaccine for a pandemic — are more likely to flow from longer-term basic research. The mutant-flu work could certainly help this research. Yet the work itself carries a risk. An accidental, or intentional, release of the mutant viruses from a lab could spark an H5N1 pandemic that we are currently in no position to mitigate.

The fact that the risks seem to far outweigh the public-health benefits of the research, at least in the short term, means that there is no need to rush headlong into an expansion of the work. Rather, regulators and flu researchers must take whatever time they need to decide the best way for such work to proceed safely. ■

Gas and air

Natural–gas operations could leak enough methane to tarnish their clean image.

How clean is natural gas? Although it is often lumped in with coal and oil, many in the energy industry are at pains to point out that burning gas to generate electricity produces fewer greenhouse-gas emissions than does burning other fossil fuels. Certainly, countries claim reductions in carbon emissions when they switch from coal to gas, as Britain did on a large scale in the 1990s. The growing popularity of shale formations as a source of gas has re-energized the debate over its environmental impact. To release the gas, engineers must split the rock by injecting fluid under high pressure, a process called fracking. Last year, researchers from Cornell University in Ithaca, New York, said that with this taken into account, carbon emissions associated with shale gas were no better — or were worse — than those from coal.

Industry maintains that the problem has been exaggerated, and many scientists agree. Sorting fact from fiction has been difficult, however, because nobody had any independent data — until now.

As discussed on page 139, a study led by scientists from the US National Oceanic and Atmospheric Administration (NOAA), headquartered in Washington DC, and the University of Colorado in Boulder looked at methane and other emissions from a natural-gas field north of Denver, where fracking methods are used to open up sand formations.

They estimated cumulative emissions from the field using not industry reports or conceptual models, but concentrations of pollutants in air samples. This is important because the atmosphere does not misrepresent data or make mistakes; nor does it bend to ideology or political will.

The data suggest that methane emissions from natural-gas operations could be substantially higher — and so be worse for global warming — than was thought. At works in the Denver-Julesburg Basin, methane emissions were roughly double the official estimate.

This will by no means settle the debate. The NOAA scientists had to make assumptions to convert atmospheric data to cumulative emissions from a vast energy complex. They readily acknowledge substantial uncertainty in their calculations, and estimate that between 2% and 8% of the methane produced from wells in the Denver-Julesburg Basin is lost to the atmosphere, with a best guess of 4%.

These numbers, which are higher than estimates from Cornell and the US Environmental Protection Agency (EPA), should serve as a red flag to the gas industry, policy-makers and the academic community. Researchers will need to confirm the findings, reduce the uncertainties and determine emissions from other locations. But the issue clearly warrants attention. The study should also be a reminder that although it is necessary for the industry to collect data on its practices and run calculations, independent monitoring and verification are needed.

More generally, the study further complicates understanding of what is considered the world's cleanest fossil fuel. Many in industry and science have talked about using gas as a bridge fuel for the transition from coal to cleaner sources of electricity, but the picture is unclear.

In many places, including the United States, gas-fired electricity is

> *"Emissions from natural-gas operations could be substantially higher than was thought."*

likely to be significantly cleaner than coal in terms of carbon emissions even with the extra methane leakage — if only because newer gas-fired plants are much more efficient than the behemoths that provide most coal-fired electric generation. By contrast, a modelling study by Tom Wigley, a climate scientist at the US National Center for Atmospheric Research in Boulder, last year found that switching from coal to natural gas would actually increase global temperatures for decades, by reducing emissions of pollutants that reflect solar radiation back into space (T. M. L. Wigley *Climatic Change* **108**, 601–608; 2011). In the end, natural gas might be preferable to coal just because it reduces harmful air pollution. But the climatic benefits are murky at best.

The good news is that the natural-gas industry has the capacity to reduce methane leakage by cleaning up its operations. Technologies are already available to capture methane during fracking rather than venting it into the atmosphere when bringing a gas well online. As it happens, the EPA is currently considering mandatory regulations that encourage such activities by limiting various pollutants from natural-gas operations. These regulations would indirectly reduce methane emissions, and the EPA must press forward. ■

Hypocritical oaths

History judges some research as unethical, despite approval at the time.

Ethical boundaries for experiments on humans can be stated very simply. "The limits of justifiable experimentation upon our fellow creatures are well and clearly defined," Canadian physician William Osler, one of the grand old men of US medicine, wrote more than a century ago. "For man absolute safety and full consent are the conditions which make such tests allowable."

Although US standards have evolved, the concepts of informed consent and safety still underpin research on humans. How, then, could leading health officials in the United States approve a set of barbarous experiments in the 1940s, in which government physicians intentionally infected hundreds of people in Guatemala with venereal diseases?

The people were labelled volunteers, but evidence suggests that they did not provide consent. And as the News Feature on page 148 shows, records indicate that some of the people exposed to syphilis, gonorrhea and chancroid subsequently went untreated.

Such recklessness seems abhorrent now, but this is far from an isolated case. In 1941, US physician William Black infected children, including a 12-month-old baby, with the herpes virus. When Black submitted his paper to the *Journal of Experimental Medicine*, it was rejected. Francis Peyton Rous, the journal's editor, told Black that his work was "an abuse of power". Nonetheless, the paper was published soon after by the *Journal of Pediatrics*.

And Rous was less concerned about a study in which residents of a psychiatric hospital in Michigan were infected with influenza, even though it seems that at least some of the patients could not give their consent. It might be tempting to explain away such research abuses as the work of rogue scientists, but the Michigan study was conducted by a leading researcher of the time, Thomas Francis Jr, and his young colleague, Jonas Salk, who went on to develop the polio vaccine.

And two decades later, in 1963, a team run by Chester Southam injected tumour cells into extremely infirm patients at the Jewish Hospital for Chronic Disease in New York without informing them that the shots contained cancer. Southam was later put on probation by the New York State medical licensing board, but many researchers defended the work and he was later elected president of the American Association for Cancer Research.

What kind of work deemed as accepted today will be denounced by future generations? The question is one that all researchers should bear in mind, because history may judge them more harshly than their peers do. One example could be denial of treatment to sick people through the use of placebos in clinical trials and the ways in which some of these trials are carried out in developing nations, amid accusations of abuse of poor, uneducated participants. Broadening to other types of research, attitudes to work on embryonic stem cells may harden. And future generations may extend the protection currently in place for humans to cover other species, such as chimpanzees.

In the case of chimpanzees, Gabon and the United States are the only nations known to still use them for research, and a committee of the US National Research Council last year recommended that the United States should sharply limit their use, but stopped short of calling for a complete ban. Meanwhile, some researchers have been able to avoid bans in their own countries by travelling to the United States. Since 2005, foreign scientists have conducted at least 27 experiments at US chimpanzee centres (see *Nature* **474**, 268–271; 2011).

There is, of course, clear water between the Guatemalan experiments and chimpanzee research. The Guatemala research was illegal, even in the 1940s, and most of the data did not prove useful and went unpublished. Still, as with research on embryonic stem cells, there is considerable debate about the ethics of using chimpanzees as experimental subjects. In these and other cases, nations would do well to heed some of the lessons that emerged from the investigation of the experiments in Guatemala. Governments and other funders of research must exert full oversight, provide as much transparency as possible and ensure that regulations are clear, strong and evolve with the times. ■

ELIZA GREGORY

Finding the true value of US climate science

A new strategy for addressing climate change takes a realistic approach to the challenge of making science useful, says **Ryan Meyer**.

National science efforts are rarely short of ambition and grandiose promises. Focusing on energy, health, global warming or whatever, they all argue that their research will make the world a better place.

Take the US$3-billion Human Genome Project and the breathless promises of cures and treatments that it would bring. In fact, the benefits have been modest because solving societal problems is a lot more complicated and difficult than generating new knowledge.

Is there an alternative? Is it possible to be realistic and nuanced about the limited role that science often has, but still to offer a compelling case for public support? The US Global Change Research Program (USGCRP) will shortly release a strategic plan that does just that.

Over the past two decades, the USGCRP, which coordinates 13 federal agencies and departments, has spent more than $30 billion on climate-change research. In doing so, it has improved our understanding of climate systems. But, as the National Research Council pointed out in 2009, when it comes to fulfilling its legal mandate of supporting decision makers with useful information, the USGCRP has been a disappointment.

At the core of the programme's difficulties is the (faulty) assumption that better information leads to better decisions. Better information is rarely sufficient. Repeated studies have shown that making information useful demands engagement with those who will use it. This is about more than just communicating science effectively. It is about responsive scientists and science institutions. Although the USGCRP aims to serve a broad range of users, from policy-makers and natural-resource managers to fishermen and urban planners, historically it has not canvassed or accounted for their needs.

This is a long-standing problem, and in 2003 the USGCRP did produce a strategic plan that tried to address it. Littered with the word 'stakeholder', the plan invoked ideas such as participatory research, integration of natural and social sciences, and better communication and education efforts. These are important ideas, strongly advocated by those who study the challenge of how to connect knowledge with action. But the new discourse rang hollow. There was no coherent plan (let alone resources) to implement the concepts, and the central goals of the programme remained entirely focused on advancing knowledge. The USGCRP did not provide any coherent account of how doing science in this way would be different from what had gone before, or how science institutions would need to change in order to deliver better value to society.

What, then, is different this time? In its 2012 report, the USGCRP has expressed a more nuanced and humble account of the role of science in society's responses to climate change.

For example, the draft plan provocatively states: "scientific knowledge is only one part of a much broader process. Information may be scientifically relevant without being decision relevant." This idea is echoed throughout its pages and is an important logical policy step. Research may offer, for example, marginal improvements in climate prediction, new data sets, or information on the distribution of a particular animal species. But these results will be irrelevant if framed poorly, or delivered at the wrong time, to the wrong people. Decision makers do not read journal articles, nor are they likely to adjust their practices to accommodate the scale or inherent uncertainty of a new model or indicator. For example, researchers examining the use of climate forecasts by water-resource managers found various barriers and constraints. These obstacles are mainly cultural and institutional, and so increases in the quality of the forecasts themselves are unlikely to stimulate increased use.

Although the USGCRP was previously organized around five goals, all concerned with increasing scientific knowledge, this time, advancing science is just one of four stated objectives. The other three — to inform decisions, to sustain assessments and to communicate and educate — are woven in with the scientific activities. This should help to make the programme's substantial science investment more relevant to local, regional and national societal needs.

The latest plan also acknowledges difficult but crucial science-policy trade-offs. For example, it discusses the "dynamic tension" between increasing model complexity and policy-makers' needs for simplicity and tractability. For a government science programme to explicitly recognize these choices as a proper concern of science management is a new and welcome step.

Will this bold vision be realized? The USGCRP does not yet have a strong mechanism for allocating funds among its new priorities. Some in the research community will surely lobby against trade-offs that seem to threaten the status quo. And, as it has in the past, the National Research Council reviewed this plan with a critical eye, pointing out that the USGCRP will need more resources and greater leverage over agency budgets and priorities to make it happen. Without these ingredients, the idea will probably run into the sand.

Despite these doubts, the USGCRP deserves applause for taking such an important conceptual step in the right direction. It has produced a plan for science that feels compelling, plausible and ambitious. It is a useful example for other science-policy organizations to follow. ∎

> THE REPORT **EXPRESSES** A MORE NUANCED AND HUMBLE ACCOUNT OF THE ROLE OF SCIENCE IN SOCIETY'S **RESPONSES** TO CLIMATE CHANGE.

Ryan Meyer *is the science integration fellow at the California Ocean Science Trust in Oakland.*
e-mail: ryan.meyer@calost.org

MARINE METAGENOMICS

Sequencing from scratch

Although most microorganisms cannot currently be cultured, their genomes may soon be accessible.

Until now, metagenomic analyses have been able to identify only dominant members of a microbial community or those sequenced previously. Virginia Armbrust and her group at the University of Washington in Seattle developed computational tools to tame the massive amount of data produced by next-generation sequencers. The method successfully sequenced two of 14 candidate genomes identified in samples from Puget Sound, most notably a microbe of low abundance but great interest — a representative of the mysterious, as yet uncultured organisms known as marine group II *Euryarchaeota*.

Researchers now have a way to peer into the secret lives of the uncultured majority.
Science 335, **587–590 (2012)**

MATERIALS

Printing tiny coiled antennas

Typically, the largest circuit component in wireless electronic devices such as mobile phones is the antenna, which sends and receives electromagnetic waves. The tiniest antennas available are made up of wires twisted into three-dimensional coils to save

P. MORRIS/ARDEA.COM

EVOLUTION

Glad rags for a blind mole

Golden moles have a blue-green sheen to their coats that is a rare example of iridescence in mammals, report Matthew Shawkey at the University of Akron in Ohio and his colleagues.

The group conducted the first detailed study of iridescent outer hairs and non-iridescent downy hairs from four species of golden mole. Iridescent hairs were highly flattened with much smaller scales than their less eye-catching counterparts. The scales form multiple layers, which alternate in colour between light and dark, and probably produce colour as light passes between layers in a phenomenon called thin-film interference.

All four mole species are blind, so it is unlikely that the hairs evolved as sexual ornamentation. The authors suggest that the iridescence of these burrowing animals is a by-product of adaptations for durable, low-friction pelts.
Biol. Lett. http://dx.doi.org/10.1098/rsbl.2011.1168 (2012)

on space while maintaining high radiation efficiency and wide bandwidth. But bending wires is cumbersome and expensive.

Stephen Forrest and Anthony Grbic at the University of Michigan in Ann Arbor and their colleagues report a way to rapidly transfer metallic patterns directly onto a curved polymer, which can be pre-moulded to a desired shape. Stamping the pattern onto a hemispherical polymer, for instance, produces miniature high-performance antennas curled in spherical helices (**pictured**).
Adv. Mater. http://dx.doi.org/10. 1002/adma/201104290 (2012)

NETWORKS

Patchy communication

People tend to communicate with each other in bursts, exchanging clusters of messages over short time periods, and following these up with longer gaps in communication. But are these patterns simply the result of a tendency to talk more during the day and the working week?

Hang-Hyun Jo of Aalto University in Finland and his colleagues found that these temporal cycles are not sufficient to explain the bursts. They analysed 322 million mobile-phone calls between more than 5 million users over 119 days in 2007. After removing the effects of the day–night and working-week cycles, the bursts remained.

The authors suggest that the patterns reflect something fundamental in the way that people communicate.
N. J. Phys. 14, **013055 (2012)**

CANCER DRUGS

Chemo spans generations

Some commonly used cancer drugs not only generate mutations in treated mice, but scar the genomes of their

C. PFEIFFER & P. HALEY

offspring as well.

Radiation is known to cause genomic instability, leading to mutations that are passed down to the first- and even second-generation progeny of exposed mice. Colin Glen and Yuri Dubrova at the University of Leicester, UK, reasoned that the same could be true of DNA-damaging chemotherapies.

The duo tested three such drugs in male mice at concentrations similar to those used in humans, and found that the offspring of exposed mice harboured up to twice as many mutations as their exposed parent at the genome location studied. Moreover, mutations were present in both the copy of the genome inherited from the exposed parent and that from the unexposed parent.
Proc. Natl Acad. Sci. USA
http://dx.doi.org/10.1073/pnas.1119396109 (2012)

HUMAN EVOLUTION

Hobbit small, but not stunted

Evidence is mounting for the argument that the 'hobbit' of Flores Island was not the same species as modern humans.

The first of the 17,000-year-old *Homo floresiensis* fossils were discovered in 2003; since then there has been fierce debate over whether they represent a new diminutive *Homo* species, or *Homo sapiens* with the medical condition cretinism. Peter Brown at the University of New England in Armidale, Australia, analysed *H. floresiensis* traits such as brain mass, skeletal proportions and tooth development, and compared them with those of people with cretinism.

Brown found no signs in the small-bodied, small-brained *H. floresiensis* of the delayed growth associated with cretinism. He says that earlier studies may have confused damage caused by the fossilization process with features of the disorder.
J. Hum. Evol. http://dx.doi.org/10.1016/j.jhevol.2011.10.011 (2012)

ASTRONOMY

Core-collapse and star formation

When massive stars accumulate more iron than their centres can hold, they explode in what is known as a core-collapse supernova. Such supernovae enrich the surrounding environment with elements, seeding the formation of other stars. Astronomers have linked the number of core-collapse supernovae in a galaxy to the rate of star formation.

Maria-Teresa Botticella at the Padua Astronomical Observatory in Italy and her colleagues compared star-formation estimates based on core-collapse explosions to those based on more conventional measurements of galactic brightness. They found good agreement between their method and one of the two others studied.

The authors also used their measurements to estimate the mass range over which iron-rich stars explode. The study should improve our understanding of these supernovae and may lead to a new way of studying star formation in distant galaxies.
Astron. Astrophys. 537, A132 (2012)

PALAEONTOLOGY

Early bird was black

The plumage of the world's first known bird contained at least some black, researchers report.

A team led by Ryan Carney at Brown University in Providence, Rhode Island, examined a fossilized feather (**pictured**) from the bird *Archaeopteryx*, which lived 150 million years ago. Using electron microscopy, they spotted rod-shaped, pigmented organelles called melanosomes inside preserved cells. Statistical comparison of the shape of these organelles with those of 87 extant birds identified similarities to melanosomes from birds with black plumage.

The most viewed papers in science

NUCLEAR ENERGY

Testing the waters for radionuclides

HIGHLY READ on pubs.acs.org in December

A relatively reassuring study about radioactive particles released into the ocean as a result of the accident at Japan's Fukushima Daiichi nuclear power plant last March has proved popular reading.

Ken Buesseler at the Woods Hole Oceanographic Institution in Massachusetts and his colleagues gathered data on caesium and iodine isotopes collected after the accident by the Tokyo Electric Power Company and the Ministry of Education, Culture, Sports, Science and Technology, and compared these with pre-accident measurements for the same isotopes. Radionuclide levels peaked one month after the accident, owing partly to releases of cooling sea water used to manage the accident.

Ultimately, the team predicts "minimal impact on marine biota or humans", but suggests that more study is warranted, especially on potential radionuclide accumulation in seafood.
Environ. Sci. Technol. 45, 9931–9935 (2011)

The melanin responsible for black pigmentation provides structural support as well as colour. The authors suggest that this would have improved the feathers' strength and durability — an advantage during this early evolutionary stage of dinosaur flight.
Nature Commun. http://dx.doi.org/10.1038/ncomms1642 (2012)

NANOTECHNOLOGY

Electrons explain zeolite complexity

A potentially useful catalyst with a porous structure akin to that of nanoscale Swiss cheese has had its structure revealed by electron crystallography.

Zeolites are microporous aluminosilicates with many applications, but their small size and the intergrowth of their crystals can make it difficult to determine the details of their structures.

Xiaodong Zou of Stockholm University, Avelino Corma at the Polytechnic University of Valencia in Spain and their team collected high-resolution transmission electron microscopy images and data on electron diffraction for a kind of zeolite called ITQ-39. From this, they determined the three-dimensional structure of the material — the most complex zeolite structure ever elucidated — and found that it is made up of three different arrangements of the same basic layer. They say that its unusual intersecting channel system makes it a promising catalyst for converting naphtha to diesel.
Nature Chem. http://dx.doi.org/10.1038/nchem.1253 (2012)

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

SEVEN DAYS
The news in brief

Vostok drilling

Russian researchers drilling down to the subglacial Lake Vostok, 3,750 metres under Antarctica's ice sheet, were confident that they had reached the lake's surface this week. As *Nature* went to press, Valery Lukin, director of the Russian Antarctic programme, said that researchers at the site were processing data to check whether they had made the breakthrough. As summer operations in the Antarctic are now ending, the team will soon leave the borehole and will return to do further analysis in December. See go.nature.com/idwdy8 for more.

Fire damage

A fire has caused extensive damage to one of Russia's major facilities for nuclear physics. The blaze broke out on 5 February among power cables for the heavy-ion accelerator at the Institute for Theoretical and Experimental Physics in Moscow. No one was injured and damage assessments are ongoing, according to a source at the institute.

FUNDING

French Ivy League

Five higher-education clusters were selected by the French government on 3 February to get a slice of a €7.7-billion (US$10-billion) funding pot intended to create an 'Ivy League' of world-class universities. The institutional alliance Sorbonne Paris City, Sorbonne University and campuses in Toulouse, Aix-Marseille and Saclay will share the pot with the three existing clusters in Bordeaux, Strasbourg and Paris. The windfall marks a break with the country's egalitarian higher-education system and, like other recent cash boosts

African land-grabs threaten ecosystems

A scramble to buy land in Africa is threatening sustainable development on the continent, say reports released on 1 February by the Rights and Resources Initiative (RRI), an international coalition of groups working to increase community ownership of forests. The reports say that millions of Africans are being dispossessed, and that resources such as wetlands, forests and rangelands are being jeopardized by the sale of land to private investors for mineral mining (pictured, iron-ore mining in Sierra Leone), logging and biofuel production. "Three-quarters of Africa's population and two-thirds of the landscape are at risk," says Andy White, who coordinates the RRI, based in Washington DC. See go.nature.com/kajiqy for more.

for research, comes from a €35-billion economic stimulus package announced in 2009 (see *Nature* **462,** 838; 2009).

POLICY

FDA whistle-blowing

The way the US Food and Drug Administration (FDA) treats whistle-blowers is attracting congressional scrutiny. Three years ago, agency employees flagged up concerns to the press and to Congress about how medical devices had been approved by officials despite having received poor scientific review. On 25 January, the whistle-blowers filed a lawsuit alleging that their e-mails had been secretly monitored, and on 31 January, Senator Charles Grassley (Republican, Iowa) told FDA commissioner Margaret Hamburg to disclose details of any spying. The FDA told *Nature* that it does not comment on pending or ongoing litigation.

EU-wide research

Research agencies, government departments and centres of excellence in 25 European countries are coordinating their study of neurodegenerative diseases such as Alzheimer's and Parkinson's. On 7 February, they launched the European Union's Joint Programme in Neurodegenerative Disease Research (JPND), the first in a planned series of Joint Programming initiatives to address societal challenges by coordinating and prioritizing research across borders. The JPND will develop treatment, prevention and care strategies. See go.nature.com/aaznbp for more.

Failed Mars probe

Russia's Phobos-Grunt spacecraft, which failed to escape Earth orbit in its attempt to reach Mars's moon Phobos last year, was doomed by electronics components not certified for use in space, which in turn led to a computer glitch, according to an official analysis commissioned by the country's space agency, Roscosmos. Its main conclusions were released on 3 February. Once the craft reached orbit, two electronics

chips suffered radiation damage (which they had not been designed to withstand), causing two processors to reboot and crashing the on-board computer program.

Nuclear restarts

The International Atomic Energy Agency (IAEA) has endorsed Japan's nuclear-reactor 'stress tests' — raising hopes that some of the country's shuttered plants might restart despite public protest. Only 3 of Japan's 54 reactors are currently operating, and reactors that were closed after an earthquake and tsunami struck the Fukushima Daiichi plant last March must pass checks showing that they can withstand similar disasters before they reopen. Reactors at the Ohi plant in Fukui prefecture passed a first round of safety checks in January — but protesters and some nuclear analysts said the tests weren't good enough. IAEA inspectors approved the procedures on 31 January.

Newborn screening

Minnesota's state health department last week began to destroy its store of blood-spot samples collected from newborn babies. The move follows a November 2011 court ruling that the state must receive informed consent from parents to store

such samples. Blood spots (**pictured**, being taken from a 1-day-old-baby's heel) from babies are often kept and later used in epidemiological studies or to develop tests for disorders. Biomedical scientists in Minnesota worry that the loss of the samples could harm such research — but campaigners for informed consent say it is more important to preserve parents' trust in the scientific enterprise. See go.nature.com/hzfxer for more.

Malaria deaths

Malaria researchers are disputing a high-profile paper that suggests the disease may kill twice as many people worldwide as previously estimated. The research (C. J. L. Murray *et al. Lancet* **379,** 413–431; 2012), published on 2 February, puts malaria deaths in 2010 at 1.24 million,

whereas the World Health Organization estimates 655,000. The paper relies on contentious 'verbal autopsies' — interviews with friends and family of the deceased — to compile the death-toll estimate. See go.nature.com/mpq4gu for more.

Pharma cuts

Drug giant AstraZeneca announced on 2 February that it would cut 2,200 jobs in research and development, part of a wider cull of 7,300 (about 12% of its workforce). Following other firms that have abandoned drug-discovery programmes for brain disorders (see *Nature* **480,** 161–162; 2011), the company also said it would end research activity at two sites focused on neuroscience: Södertälje in Sweden and Montreal in Canada. It will keep just 40 or 50 staff in neuroscience, working largely in external collaborations with academia and industry partners.

Cystic fibrosis

Twenty-three years after scientists announced that they had identified the gene underlying cystic fibrosis, the US Food and Drug Administration has approved a landmark drug that targets a specific mutation in the gene. But ivacaftor (Kalydeco), made

by Vertex Pharmaceuticals of Cambridge, Massachusetts, is effective in only the 4% of US patients with cystic fibrosis who have that mutation, and costs US$294,000 a year per person. It was approved for sale on 31 January. See page 145 for more.

Satellite contract

The European Space Agency has signed a €250-million (US$327-million) contract for the construction of eight more satellites for Galileo, Europe's global navigation system. The European Commission, which co-funds the network, says that first services should start by 2014. The winning consortium — OHB System in Bremen, Germany, and partner Surrey Satellite Technology in Guildford, UK — was also contracted to build the first 14 Galileo satellites. The network remains on course for completion (up to 30 satellites) by the end of the decade.

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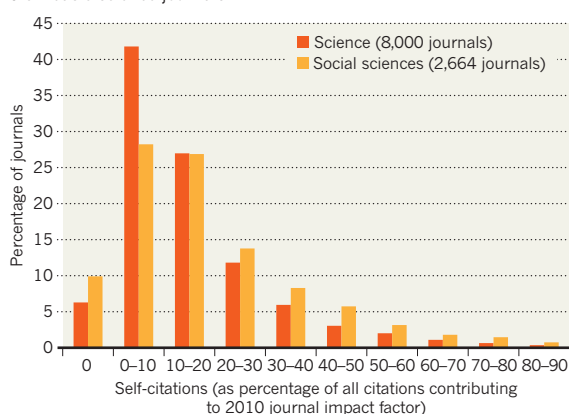For daily news updates see:
www.nature.com/news

TREND WATCH

One in five academics in a variety of social-science and business fields say journal editors have asked them to pad their papers with superfluous references to the journal in question, according to a 2 February survey (A. W. Wilhite and E. A. Fong *Science* **335,** 542–543; 2012). Such self-citations can inflate a journal's impact factor, although Thomson Reuters — which publishes the impact factor — says four-fifths of journals keep self-citations below 30% (see chart). See go.nature.com/admydy for more.

SELF-CITATIONS IN RESEARCH JOURNALS
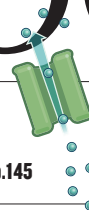Social-science journals tend to have more self-citations than basic-science journals.

Bar chart. X-axis: Self-citations (as percentage of all citations contributing to 2010 journal impact factor), bins 0, 0–10, 10–20, 20–30, 30–40, 40–50, 50–60, 60–70, 70–80, 80–90. Y-axis: Percentage of journals, 0–45. Series: Science (8,000 journals) in red; Social sciences (2,664 journals) in orange.

NEWS IN FOCUS

Natural-gas operations in areas such as Wyoming's Jonah Field could release far more methane into the atmosphere than previously thought.

CLIMATE CHANGE

Air sampling reveals high emissions from gas field

Methane leaks during production may offset climate benefits of natural gas.

BY JEFF TOLLEFSON

When US government scientists began sampling the air from a tower north of Denver, Colorado, they expected urban smog — but not strong whiffs of what looked like natural gas. They eventually linked the mysterious pollution to a 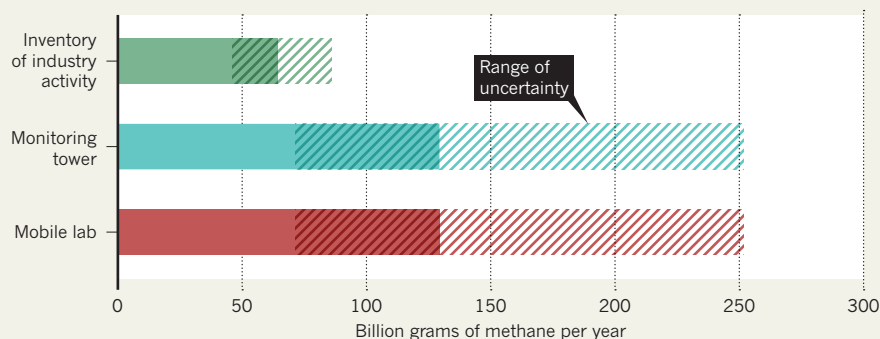nearby natural-gas field, and their investigation has now produced the first hard evidence that the cleanest-burning fossil fuel might not be much better than coal when it comes to climate change.

Led by researchers at the National Oceanic and Atmospheric Administration (NOAA) and the University of Colorado, Boulder, the study estimates that natural-gas producers in an area known as the Denver-Julesburg Basin are losing about 4% of their gas to the atmosphere — not including additional losses in the pipeline and distribution system. This is more than double the official inventory, but roughly in line with estimates made in 2011 that have been challenged by industry. And because methane is some 25 times more efficient than carbon dioxide at trapping heat in the atmosphere, releases of that magnitude ▶

A LOSING BATTLE

Estimates of methane losses from gas fields near Denver, Colorado, based on air sampling differ considerably from calculations based on industry activity.

Inventory of industry activity

Monitoring tower

Mobile lab

Range of uncertainty

0   50   100   150   200   250   300
Billion grams of methane per year

▶ could effectively offset the environmental edge that natural gas is said to enjoy over other fossil fuels.

"If we want natural gas to be the cleanest fossil fuel source, methane emissions have to be reduced," says Gabrielle Pétron, an atmospheric scientist at NOAA and at the University of Colorado in Boulder, and first author on the study, currently in press at the *Journal of Geophysical Research*. Emissions will vary depending on the site, but Pétron sees no reason to think that this particular basin is unique. "I think we seriously need to look at natural-gas operations on the national scale."

The results come as a natural-gas boom hits the United States, driven by a technology known as hydraulic fracturing, or 'fracking', that can crack open hard shale formations and release the natural gas trapped inside. Environmentalists are worried about effects such as water pollution, but the US government is enthusiastic about fracking. In his State of the Union address last week, US President Barack Obama touted natural gas as the key to boosting domestic energy production.

LACK OF DATA

Natural gas emits about half as much carbon dioxide as coal per unit of energy when burned, but separate teams at Cornell University in Ithaca, New York, and at the US Environmental Protection Agency (EPA) concluded last year that methane emissions from shale gas are much larger than previously thought. The industry and some academics branded those findings as exaggerated, but the debate has been marked by a scarcity of hard data.

"It's great to get some actual numbers from the field," says Robert Howarth, a Cornell researcher whose team raised concerns about methane emissions from shale-gas drilling in a pair of papers, one published in April last year and another last month (R. W. Howarth *et al. Clim. Change Lett.* **106,** 679–690; 2011; R. W. Howarth *et al. Clim. Change* in the press). "I'm not looking for vindication here, but [the NOAA] numbers are coming in very

close to ours, maybe a little higher," he says.

Natural gas might still have an advantage over coal when burned to create electricity, because gas-fired power plants tend to be newer and far more efficient than older facilities that provide the bulk of the country's coal-fired generation. But only 30% of US gas is used to produce electricity, Howarth says, with much of the rest being used for heating, for which there is no such advantage.

ON THE SCENT

The first clues appeared in 2007, when NOAA researchers noticed occasional plumes of pollutants including methane, butane and propane in air samples taken from a 300-metre-high atmospheric monitoring tower north of Denver. The NOAA researchers worked out the general direction that the pollution was coming from by monitoring winds, and in 2008, the team took advantage of new equipment and drove around the region, sampling the air in real time. Their readings led them to the Denver-Julesburg Basin, where more than 20,000 oil and gas wells have been drilled during the past four decades.

Most of the wells in the basin are drilled into 'tight sand' formations that require the same fracking technology being used in shale formations. This process involves injecting a slurry of water, chemicals and sand into wells at high pressure to fracture the rock and create veins that can carry trapped gas to the well. Afterwards, companies need to pump out the fracking fluids, releasing bubbles of dissolved gas as well as burps of early gas production. Companies typically vent these early gases into the atmosphere for up to a month or more until the well hits its full stride, at which point it is hooked up to a pipeline.

"A big part of it is just raw gas that is leaking from the infrastructure."

The team analysed the ratios of various

pollutants in the air samples and then tied that chemical fingerprint back to emissions from gas-storage tanks built to hold liquid petroleum gases before shipment. In doing so, they were able to work out the local emissions that would be necessary to explain the concentrations that they were seeing in the atmosphere (see 'A losing battle'). Some of the emissions come from the storage tanks, says Pétron, "but a big part of it is just raw gas that is leaking from the infrastructure". Their range of 2.3–7.7% loss, with a best guess of 4%, is slightly higher than Cornell's estimate of 2.2–3.8% for shale-gas drilling and production. It is also higher than calculations by the EPA, which revised its methodology last year and roughly doubled the official US inventory of emissions from the natural-gas industry over the past decade. Howarth says the EPA methodology translates to a 2.8% loss.

The Cornell group had estimated that 1.9% of the gas produced over the lifetime of a typical shale-gas well escapes through fracking and well completion alone. NOAA's study doesn't differentiate between gas from fracking and leaks from any other point in the production process, but Pétron says that fracking clearly contributes to some of the gas her team measured.
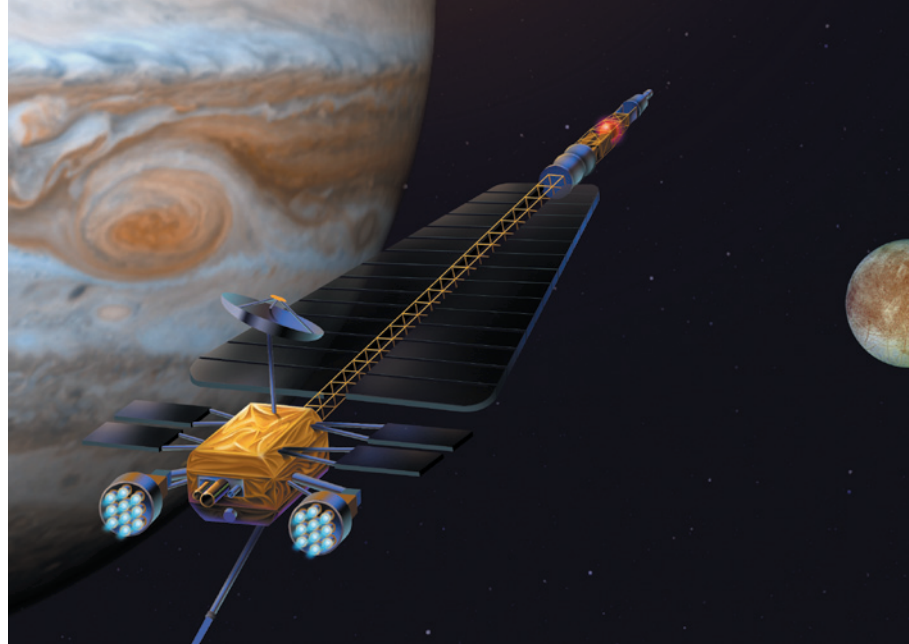
Capturing and storing gases that are being vented during the fracking process is feasible, but industry says that these measures are too costly to adopt. An EPA rule that is due out as early as April would promote such changes by regulating emissions from the gas fields.

Officials with America's Natural Gas Alliance, based in Washington DC, say that the study is difficult to evaluate based on a preliminary review, but in a statement to *Nature* they add that "the findings raise questions and warrant a closer examination by the scientific community". Environmental groups are pushing the EPA to strengthen pollution controls in the pending rule, but industry is pushing to relax many of the requirements. Many companies are already improving their practices and reducing emissions throughout the country, either voluntarily or by regulation, the alliance says.

Not all studies support the higher methane numbers. Sergey Paltsev, assistant director for economic research at the Massachusetts Institute of Technology Energy Initiative in Cambridge, and his colleagues are gathering information about industry practices for a study on shale-gas emissions. He says that their figures are likely to come in well below even the lower EPA estimate. He calls the NOAA results "surprising" and questions how representative the site is.

Pétron says that more studies are needed using industry inventories and measurements of atmospheric concentrations. "We will never get the same numbers," she says, "but if we can get close enough that our ranges overlap in a meaningful way, then we can say we understand the process." ∎

G. PÉTRON ET AL. J. GEOPHYS. RES.

A mission to Jupiter's large icy moons, cancelled in 2006, would have been powered by a nuclear reactor.

SPACE SCIENCE

Fission power back on NASA's agenda

Space–technology report prioritizes nuclear propulsion.

BY ERIC HAND

Michael Houts wants astronauts to ride a nuclear reactor to Mars. He is convinced that small amounts of uranium-235 — which has an energy density one million times greater than that of liquid fuels — could power rockets efficiently, using the heat of fission to accelerate small stores of lightweight hydrogen propellant. But although Houts, the nuclear-research manager at NASA's Marshall Space Flight Center in Huntsville, Alabama, has an unwavering belief in the potential of space-based nuclear power and propulsion, the funding to develop that technology has been inconsistent. This year, he is leading a nuclear-propulsion project with a budget of US$3 million — minuscule in comparison with the $1.3 billion that NASA will spend on space-technology research and development in the 2012 fiscal year. "The funding at times has gone to zero," says Houts. "You lose the teams and the momentum."

Yet a report released on 1 February by the US National Research Council could change Houts's fortunes. *Space Technology Roadmaps and Priorities* is the first ever community-based document to set priorities for NASA's space-technology division. The report's steering committee spent a year canvassing opinion in both industry and academia to create a ranked list of the 16 most important areas of technology development, out of a potential 320 topics. Nuclear power and propulsion came high on the list. "It would change exploration in a fundamental way forever," says Raymond Colladay, chairman of the committee and former president of Lockheed Martin Astronautics in Denver, Colorado.

Other technologies were ranked higher. For instance, the committee put an emphasis on developing 'star shades' and coronagraphs to block the light of distant stars and allow space telescopes to discern the faint light of planets orbiting them. And the report prioritized the development of ways to protect astronauts from radiation on long missions.

But the committee also said that small

POWER DRIVE

Of the available sources of energy for space flight, only nuclear fission offers both high power and long duration.

fission reactors could revolutionize the exploration of the Solar System by both humans and robots. Reactors could support long-lasting experiments on the surface of planets and power missions to the outer Solar System, where the Sun is too distant to provide much power for even the most efficient solar panels. And once human space exploration gets going, nuclear propulsion systems may be essential for multi-year trips to the asteroids or Mars. With twice the efficiency of chemical rockets, reactors could push astronauts not just farther, but also faster than ever before (see 'Power drive') — which could help to reduce explorers' exposure to space radiation.
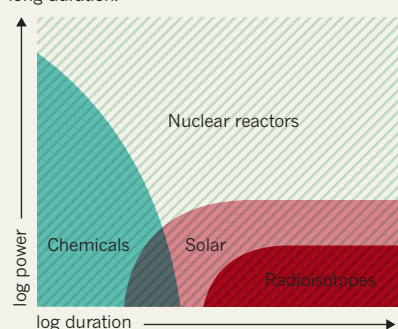
Mason Peck, NASA's chief technologist, says that he will use the priority list as a guide when setting funding in future. However, developing fission power for space will require not only money, but also political will: the image of a nuclear-powered spacecraft blowing up on the launch pad or on its way to orbit is a powerful deterrent. Houts says that the risk of nuclear material contaminating Earth after an accident is negligible because the reactor would not be started until the system were in orbit. Nevertheless, past attempts to demonstrate the technology have faltered. In 2003, NASA began Project Prometheus, which supported the development of a nuclear reactor that would drive an electric ion thruster to power a probe to Jupiter. The programme received as much as $430 million in 2005, but was cancelled a year later as NASA shifted its resources towards returning to the Moon — a destination for which nuclear propulsion was not needed.

Although the project has disappeared, it did support work that is now bearing fruit in the form of a new radioisotope power generator — a power source that does not use fission, but instead relies on the natural heat from the decay of plutonium. The Advanced Stirling Radioisotope Generator (ASRG) is lighter and more efficient than previous examples, and the space-technology report identified it as a "tipping point" technology that is almost ready for in-flight demonstration. Two mission proposals that include the ASRG — one to explore the hydrocarbon seas of Saturn's moon Titan in a boat, the other to hop from comet to comet — are under consideration at NASA.

Houts thinks that the radioactive power source for these missions would not generate much political controversy — certainly nothing like the protests when the Cassini–Huygens mission was sent to Saturn in 1997 with an earlier version of a radioistotope generator. Nowadays, Houts often opens his academic talks by asking whether the audience is aware that there is plutonium on board the Mars Science Laboratory, a mission that was launched in November 2011 to take a massive rover to Mars. About half are not, he says. "In a strange way, I feel that's good news," says Houts. "It seems like it's becoming a very accepted technology." ■

To make flu vaccines, the virus is grown in hens' eggs — a slow, complex process adequate for seasonal vaccines but ill-suited to responding to a pandemic.

PUBLIC HEALTH

Lab flu may not aid vaccines

Game-changing vaccine technologies are needed to strengthen global pandemic defences.

BY DECLAN BUTLER

Now that laboratory studies have yielded a glimpse of H5N1 flu viruses that might spread rapidly in humans and cause a devastating pandemic, vaccine makers will be better prepared if one develops. Or will they?

It is an appealing argument, and one that some scientists have made in recent weeks as controversy has swirled over two experiments that created H5N1 strains able to spread in mammals. But most experts contacted by *Nature* say that the work is unlikely to speed up the vaccine response in a pandemic. Jeremy Farrar, director of the Oxford University Clinical Research Unit in Ho Chi Minh City, Vietnam, calls such expectations "a red herring".

Farrar and others say that the studies, from teams led by Ron Fouchier at the Erasmus Medical Center in Rotterdam, the Netherlands, and Yoshihiro Kawaoka at the University of Wisconsin-Madison, could benefit flu surveillance in the long term. The detection in naturally circulating viruses of genetic changes similar to those of the lab strains might provide an early warning of a pandemic — although researchers stress that current surveillance systems are nowhere near up to that job, and that a pandemic might be heralded by a completely

different set of mutations (see *Nature* **481**, 417–418; 2012). But many agree with Richard Webby, a flu virologist at the St. Jude Children's Research Hospital in Memphis, Tennessee, who says, "I think the research is important, but not for vaccine purposes."
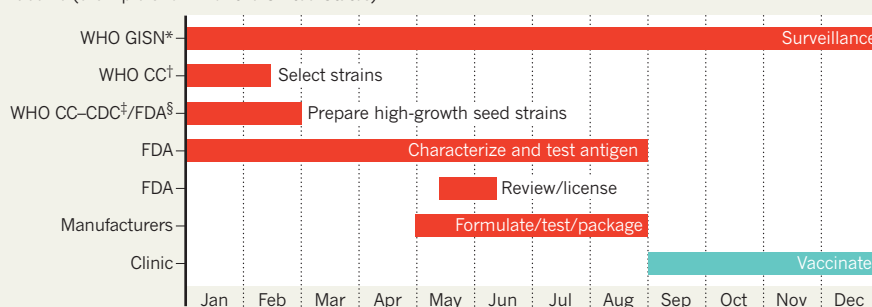
Producing vaccine faster, and in larger amounts, when a pandemic breaks out is a key public-health goal (see 'Jab lag'). In 2009, a vaccine only became available months after the H1N1 pandemic began — and even then there was enough for only about 20% of the world's population. Fouchier and Kawaoka have both argued that their research could improve pandemic preparedness. Should a wild H5N1 virus

be detected with combinations of mutations similar to those of the lab viruses, they say, manufacturers could ramp up production of the 'pre-pandemic' vaccines that have been produced in small quantities against current H5N1 viruses.

In a Comment piece in *Nature* this week (see page 155), Kawaoka notes that existing H5N1 vaccines are effective against his lab-created strain, suggesting that these might provide some cross-protection against any natural human pandemic strain resembling it. Anthony Fauci, director of the US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, adds that seeing the lab

JAB LAG

It can take at least six months to produce a new influenza vaccine (example shown for the United States).

*World Health Organization Global Influenza Surveillance Network; †WHO Collaborating Centres; ‡US Centers for Disease Control and Prevention; §US Food and Drug Administration.

mutations accumulating in field isolates of H5N1 could "get us to be a little more attentive to revving up the capability of expanding the doses of vaccine we have".

But others say that such mutants are unlikely to be spotted in time. Moreover, they say, governments are unlikely to take the political or economic risks of plunging ahead with a vaccine on the basis of a putative pandemic threat. "Nobody is going to ramp up production of a pre-pandemic vaccine based on these two experimental viruses," says Webby. "That's 100% sure." Ab Osterhaus, a co-author on Fouchier's paper also at Erasmus, agrees that industry will wait on the actual pandemic strain for any major roll-out, but says that screening for mutations could detect variants that could be used to make new seed strains, which might help with the initial response.

Daniel Perez, a virologist at the University of Maryland in College Park, has argued that the lab strains themselves could be added to the current panel of pre-pandemic vaccine strains. But the existing pre-pandemic vaccines are probably just as good. For vaccine purposes, what counts most is the overall antigenic properties of the virus's two surface proteins, haemagglutinin (HA) and neuraminidase (NA).

"The antigenicity of the virus depends less on any mutations than on where the HA and NA come from," says Ilaria Capua, an avian-flu researcher at the Veterinary Public Health Institute in Legnaro, Italy. Seed strains representing the major variants of circulating HA and NA are regularly revised for use in pre-pandemic vaccines.

In any case, vaccine makers seem unlikely to take substantial pre-emptive action, regardless of any ominous changes in wild H5N1. Bram Palache, the global government affairs director for vaccines at Abbott Biologicals in Weesp, the Netherlands, says that industry will not switch its limited plant capacity from making the seasonal flu vaccines to making a pandemic vaccine until a human pandemic has actually emerged and government orders are in hand. And once a pandemic is under way, neither industry nor governments will be content to use existing pre-pandemic vaccines — they will insist on one matched to the pandemic strain itself, says Palache.

Given the current technology and infrastructure, developing and manufacturing such a vaccine will take many months. Now that the mutant-flu studies have suggested that an H5N1 pandemic is a real possibility, health authorities should focus on shortening that timescale, says Farrar. He urges much greater investment in better and faster vaccine technologies, including universal flu vaccines — because H5N1 is far from the only possible pandemic strain. ■ **SEE EDITORIAL P.131**

Japan finds a key to unlock philanthropy

Latest Kavli centre beats legal hurdle to using endowments.

BY DAVID CYRANOSKI

Japan's universities and research institutes have long had to make do with few philanthropic donations. Strict laws governing university finances, and the lack of a philanthropic tradition, have discouraged the gifts that serve Western institutions so well.

But change is coming. This week, the University of Tokyo unveils the country's first institute named after a foreign donor: the Kavli Institute for the Physics and Mathematics of the Universe.

The announcement adds Norwegian philanthropist Fred Kavli's name, along with a US$7.5-million endowment, to one of Japan's most successful institutes. Launched at the university in 2007 as one of the country's five World Premier International Research Centers (WPIs) (see *Nature* **447,** 362–363; 2007), the Institute for the Physics and Mathematics of the Universe (IPMU) has become an international force in black-hole and dark-matter research. Run by Hitoshi Murayama of the University of California, Berkeley, the IPMU has an international mix that is uncharacteristic in Japan, something that helped it to gain the only 'S' (superior) grade among the WPIs from Japan's increasingly strict science funders in last year's interim evaluations. Last year, the university created the Todai Institutes for Advanced Study — which gives the IPMU a permanent organizational home. But converting the IPMU into a Kavli institute was a more arduous feat.

"It actually caused stirs and ruffled some feathers," says Murayama. Some argued that the endowment was insufficient to merit naming rights, and accused the researchers of "selling our souls for money", he says.

But the government has pushed Japan's cash-starved universities to seek external funds, and the IPMU, despite gaining a place in the University of Tokyo's organizational unit, did not have secure funding. So when the Kavli Foundation in Oxnard, California, approached Murayama to offer its support, he jumped at the chance.

Japan's university law, however, does not allow public universities to put money into high-yield but risky investment schemes. That makes it nearly impossible for institutions

"Japan's system is catching up with the West."

to use the returns on an endowment to continuously support themselves, as the other 15 Kavli institutes do. "You're better off just spending the endowment," says Murayama.

To get around this, instead of handing the endowment over to the IPMU, the Kavli Foundation will continue to manage the sum, giving the institute the return on the funds. "It has told us that if the system in Japan ever changes, the money is ours," says Murayama.

Murayama says that the money will allow the IPMU to continue wooing foreign researchers by, for example, finding jobs for spouses and helping to place researchers' children in international schools. The ministry considers such expenditures to be "personal matters" and not reimbursable with public funds.

Having found a way to solve the philanthropy problem, the university is keen to try it again. "I hope to build on this momentum and redouble our efforts to pursue reform" of the systems for managing donated funds, says university president Junichi Hamada.

Hiroshi Komiyama, chairman of the Mitsubishi Research Institute in Tokyo, and past-president of the University of Tokyo, hopes that the Kavli deal "will be a great trigger for other universities" to get more external funding through endowments.

That will require a much broader change, however. In 2007, charitable donations in Japan amounted to only 0.11% of the country's gross domestic product, compared with 2.2% in the United States and 0.80% in the United Kingdom, which Komiyama blames on the relatively even distribution of wealth in Japan. Tough rules on the creation of non-profit organizations that aim to organize charitable giving have also held back giving (see *Nature* **450,** 24–25; 2007).

A law enacted last year — and another that will come into effect in April — should ease the rules on setting up non-profit organizations. The laws also increase the tax exemptions for individuals donating to non-profit organizations or private universities. These will not bring direct benefit to public universities, but they could open the way fo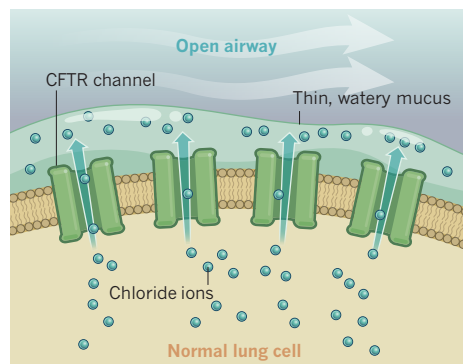r new non-profit organizations that could collect and manage funds for the universities. "Japan's system is catching up with the West," says Komiyama. ■
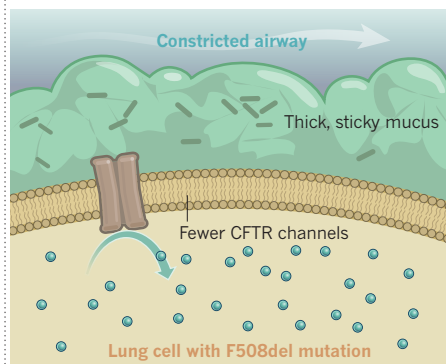
REOPENING A CHANNEL

The gene mutated in cystic fibrosis normally produces a channel that allows chloride ions to pass through the cell membrane, keeping the protective layer of mucus fluid.

A new drug corrects the effects of the rare G551D mutation, which stops the channel opening correctly, causing thick mucus that promotes bacterial infection.

Most people with cystic fibrosis have the F508del mutation, which stops the CFTR protein from folding properly and limits the number of channels.

Open airway — CFTR channel — Thin, watery mucus — Chloride ions — Normal lung cell

Constricted airway — Bacteria — Mutated CFTR — Thick, sticky mucus — Lung cell with G551D mutation

Constricted airway — Thick, sticky mucus — Fewer CFTR channels — Lung cell with F508del mutation

HEALTH

Drug bests cystic–fibrosis mutation

First treatment to tackle protein behind the disease wins approval — but only a small fraction of patients will benefit.

BY HEIDI LEDFORD

In a blinded clinical trial, neither the patient nor the clinician should know who is receiving placebo and who the active drug. But during a trial of Kalydeco (ivacaftor), a cystic-fibrosis treatment approved by the US Food and Drug Administration on 31 January, Drucy Borowitz says it was sometimes easy to tell the difference. "We had two brothers in the trial," says Borowitz, a paediatric pulmonologist at the State University of New York in Buffalo. After two weeks, she says, the pair stepped out of the lift together and it was clear who was taking the drug. "The younger brother looked sturdier," she says. "It reminded me of the change in appearance that we see in patients with cystic fibrosis after they have lung transplants."

Kalydeco, made by Vertex Pharmaceuticals of Cambridge, Massachusetts, is the first drug to target a cause of cystic fibrosis rather than the condition's symptoms. In doing so, it fulfils a promise made more than 20 years ago when a mutated gene, called cystic fibrosis transmembrane conductance regulator (*CFTR*), was first discovered and researchers spoke optimistically about developing drugs to correct it.

"This is a seminal turning point in the treatment of cystic fibrosis," says Matthew Reed, chief executive of the Cystic Fibrosis Trust in London. "But there is much further to go until we've cracked the cystic-fibrosis problem." As many as 1,500 different mutations have been found to affect *CFTR*, and Kalydeco targets just one — G551D, found in 4% of patients.

The CFTR protein forms a channel that allows chloride ions to cross the cell membrane — a key step in the production of mucus, digestive enzymes and sweat (see 'Reopening a channel'). In patients with the G551D mutation, the channel fails to open properly, so ions are unable to pass through. About 90% of people with cystic fibrosis have a different mutation, called F508del, which results in proteins that do not fold into their proper shape and so get targeted for degradation, reducing the number of channels. Either way, the resulting imbalance of ions causes mucus to become thick and sticky, blocking airways and opening the door to infection.

In the beginning, the idea that a chemical could correct the cystic-fibrosis protein was a tough sell. Most drugs work by blocking a protein's function, often by binding to an important site on the protein to gum up its activity. "It's easy to break things," says Eric Olson, head of cystic-fibrosis research at Vertex. "It's a lot harder to think about how to fix a protein."

Drugs are on the horizon for countering other cystic-fibrosis mutations. VX-809, another Vertex compound, seems to protect proteins affected by the F508del mutation from degradation. Trials of this drug in combination with Kalydeco are under way to see

whether VX-809 will get the protein to the cell membrane so that Kalydeco could then get it working. Last year, Vertex announced that early tests of this combination reduced the amount of chloride in sweat — a marker used to judge how well CFTR is functioning. But whereas Kalydeco nearly halved sweat chloride in people with G551D, the combination therapy cut it by just 13% in those with F508del. That, says Borowitz, may be because the dose used was not high enough. A larger dose is now being tested in a clinical trial, with results expected later this year.

Meanwhile, PTC Therapeutics, based in South Plainfield, New Jersey, is developing a therapy to target a set of mutations that insert a 'stop' signal in the middle of *CFTR*, preventing the cell from producing a full-length protein. That therapy, called ataluren, is in late-stage clinical trials, with results due later this year. And several institutions in the United Kingdom, working with Genzyme, a biotech company also based in Cambridge, are trying to use gene therapy to help the cell to express normal CFTR. They expect to get backing from the British government for their next clinical trial, says Reed.

The Cystic Fibrosis Foundation in Bethesda, Maryland, has helped to drive the development of new drugs. It supports another project at Genzyme and is co-sponsoring a project with Pfizer in New York, which recently bought FoldRx, a firm based in Cambridge that develops therapies for misfolded proteins. The foundation hopes that more sophisticated chemical screens — for instance, testing drugs in cells cultured from patients with cystic fibrosis rather than in rat cells, as was done originally — may yield new hits.

Despite the latest success, researchers and patient advocates say that combinations of all these approaches may be needed to tackle the disease fully. When *CFTR* was first cloned, "we were naive as to the complexity of curing a genetic disease", says Jack Riordan, a biochemist at the University of North Carolina at Chapel Hill who worked on the team that discovered *CFTR*. "We thought we could find a silver bullet, but we don't use that terminology any more." ∎

MISCONDUCT

Duplicate–grant case puts funders under pressure

Critics call for tighter checks to stop researchers being funded twice for the same work.

BY EUGENIE SAMUEL REICH

It sounds like every researcher's dream: two or more agencies are falling over each other to fund your grant proposal.

But for those tempted to accept funding for the same piece of research from more than one agency, grant fraud charges brought by the US authorities on 31 January are a sober warning. The incident has also sparked renewed calls for funding agencies to work harder to avoid grant duplication.

The recent charges were brought against Craig Grimes, who until 2010 was a professor of electrical engineering at Pennsylvania State University. Last month, he pleaded guilty to charges that included accepting grants from the Department of Energy (DOE) and the US National Science Foundation (NSF) to fund the same research on solar conversion of carbon dioxide into hydrocarbons. "It is not a problem to apply for funds for the same research at different funding agencies, but it is illegal to accept and use the funding," says Christine Boesz, a former inspector-general for the NSF. Such duplicate funding is banned in many leading scientific nations. Boesz says that there is no way of knowing how prevalent the problem is, but that cases tend to come to light only if peer reviewers spot similarities in grant applications.

Grimes raised money for his research on carbon dioxide conversion through a 2009 NSF grant, and went on to accept a second grant later in the year from the DOE's Advanced Research Projects Agency-Energy (ARPA-E), while claiming that he had no other source of funding. Lisa Powers, a spokeswoman for his former university, says that university administrators did question Grimes about having two grants that sounded very similar, but that he assured them there was no overlap. Yet in a 2010 paper (S. C. Roy *et al. ACS Nano* **4,** 1259–1278; 2010) he openly acknowledged both the NSF and ARPA-E for supporting the same work. That year, the DOE inspector-general spotted the similarity between the grants, the NSF began its investigation, and Grimes resigned his university position.

The charges against Grimes also include misappropriating National Institutes of Health funds intended to test a blood sensor in newborn babies. "Due to the pending criminal case, Dr Grimes has no comment other than to say that he is dedicated to his scientific research and is hopeful that he will be able to continue to make progress in his work," says Grimes's attorney, Tina Miller of Farrell & Reisinger in Pittsburgh, Pennsylvania.

Republicans in the House of Representatives, who criticized ARPA-E in a 24 January hearing for not having sufficient checks in place to prevent duplication with other sources of funding, say the case underscores their concerns. But Paul Tonko (Democrat, New York) commends the agency for its prompt action over the fraud, adding that it "should discourage others from trying to slip something by ARPA-E".

Congressman Paul Broun (Republican, Georgia), however, wonders why ARPA-E didn't notice the duplication before awarding its grant. Both the NSF and the DOE told *Nature* that their agencies take multiple precautions to prevent duplicate funding. The primary protection is to require grant applicants to declare other sources of private or public funding when they apply, which Grimes failed to do. Agencies also expect an applicant to turn down a grant if it overlaps with another one.

> *It is not a problem to apply for funding for the same research at different agencies, but it is illegal to accept it.*

But the additional funding may pose too great a temptation for some researchers, and agencies could take further steps to nip duplicate awards in the bud, says Harold Garner, a bioinformatician at Virginia Tech in Blacksburg. In general, agencies do not cross-check federal grants against their own new awards. Garner has developed software that could indicate whether projects being described are the same, and he says that agencies could benefit from using it.

For example, the discovery of duplicated text triggered a 2010 inquiry into electrical engineer Guifang Li of the University of Central Florida in Orlando, who was accused of plagiarizing material from another research group's paper in his grant proposal to the US Air Force. Air Force and NSF investigations subsequently revealed that duplicate text had appeared in successful applications submitted to the Air Force, the Pentagon's research agency DARPA and the NSF. Concluding that this was a case of duplicate funding for the same work, the NSF barred Li from applying for federal funding for two years. It referred his case to the US Department of Justice, which did not prosecute because of the low amounts of money involved, and because there was no proof that Li had criminal intent.

Li says that he disagrees with the NSF's assessment of the case and its conclusion that he broke the rules. Although he submitted partly identical grant proposals to save time, he says, other parts of the proposals differed and would have led to different research projects. But he adds that he is trying to put the episode behind him, and is continuing with his research career.

"It was a unique situation in which everyone wanted to fund" the work, he says. "Basically, I had an idea that everybody loved, and that's the sad part of it." ∎

MORE ONLINE

A. FRANCA/AP

IBSC/SPL

FIRST, DO HARM

In the 1940s, US doctors deliberately infected thousands of Guatemalans with venereal diseases. The wound is still raw.

BY MATTHEW WALTER

The injections came without warning or explanation. As a low-ranking soldier in the Guatemalan army in 1948, Federico Ramos was preparing for weekend leave one Friday when he was ordered to report to a clinic run by US doctors.

Ramos walked to the medical station, where he was given an injection in his right arm and told to return for another after his leave. As compensation, Ramos's commanding officer gave him a few coins to spend on prostitutes. The same thing happened several times during the early months of Ramos's two years of military service. He believes that the doctors were deliberately infecting him with venereal disease.

Now 87 years old, Ramos says that he has suffered for most of his life from the effects of those injections. After leaving the army, he returned to his family's remote village, on a steep mountain slope northeast of Guatemala City. Even today, Las Escaleras has no electricity or easy access to medical attention. It wasn't until he was 40, nearly two decades after the injections, that Ramos saw a doctor and was diagnosed with syphilis and gonorrhoea. He couldn't pay for medication.

"For a lack of resources, I was here, trying to cure myself," says Ramos. "Thanks to God, I would feel some relief one year, but it would come back." Over the decades, he has endured bouts of pain and bleeding while urinating, and he passed the infection onto his wife and his children, he told *Nature* last month in an interview at his home.

Ramos's son, Benjamin, says that he has endured lifelong symptoms, such as irritation in his genitals, and that his sister was born with cankers on her head, which led to hair loss. Ramos and his children blame the United States for their decades of suffering from venereal disease. "This was an American experiment to see if it caused harm to human beings," says Benjamin.

Ramos is one of a handful of survivors from US experiments on ways to control sexually

US doctors experimented on patients with psychiatric disorders without consent.

transmitted diseases (STDs) that ran in semi-secrecy in Guatemala from July 1946 to December 1948. US government researchers and their Guatemalan colleagues experimented without consent on more than 5,000 Guatemalan soldiers, prisoners, people with psychiatric disorders, orphans and prostitutes. The investigators exposed 1,308 adults to syphilis, gonorrhoea or chancroid, in some cases using prostitutes to infect prisoners and soldiers. After the experiments were uncovered in 2010, Ramos and others sued the US government, and US President Barack Obama issued a formal apology. Obama also asked a panel of bioethics advisers to investigate, and to determine whether current standards adequately protect participants in clinical research supported by the US government.

When details of the Guatemalan experiments came to light, US health officials condemned them as 'repugnant' and 'abhorrent'. Last September, the Presidential Commission for the Study of Bioethical Issues went further, concluding in its report[1], that "the Guatemala experiments involved unconscionable violations of ethics, even as judged against the researchers' own understanding of the practices and requirements of medical ethics of the day" (see 'Evolving ethics').

Yet that report and documents written by the researchers involved in the Guatemalan work paint a more complex picture. John Cutler, the young investigator who led the Guatemalan experiments, had the full backing of US health officials, including the surgeon general.

"Cutler thought that what he was doing was really important, and he wasn't some lone gunman," says Susan Reverby, a historian at Wellesley College in Massachusetts, whose discovery of Cutler's unpublished reports on the experiments led to the public disclosure of the research[2].

Cutler and his superiors knew that some parts of society would not approve. But they viewed the studies as ethically defensible because they believed that the results would have widespread benefits and help Guatemala to improve its public-health system. Those rationalizations serve as a warning about the potential for medical abuses today, as Western clinical trials increasingly move to developing countries to take advantage of lower costs and large populations of people with untreated disease. Bioethicists worry that laxer regulations and looser ethical standards in some countries allow researchers to conduct trials that would not be allowed at home. "The strongest lesson should be that the same rules, same principles, same ethics should apply no matter where you are," says Christine Grady, acting chief of the Department of Bioethics at the National Institutes of Health (NIH) Clinical Center in Bethesda, Maryland, and a member of the bioethics commission.

THE WAR AGAINST SYPHILIS

In the early decades of the twentieth century, US health officials were consumed by the battle against STDs, much as subsequent generations of researchers have fought cancer and HIV. In 1943, Joseph Moore, then chairman of the US National Research Council's Subcommittee on Venereal Diseases, estimated that the military would face 350,000 new infections of gonorrhoea annually, "the equivalent of putting out of action for a full year the entire strength of two full armored divisions or of ten aircraft carriers". The government launched vigorous campaigns of research, treatment and advertising to combat the problem. "She may look clean — but pick-ups, 'good time' girls, prostitutes spread syphilis and gonorrhea", read one poster issued by the US Public Health Service, which promotes health initiatives and medical research.

Many of the country's leading health officials were veterans of that fight. The surgeon general who would approve the proposal for the Guatemalan experiments, Thomas Parran, had previously run the Public Health Service's Venereal Disease Research Lab (VDRL) in New York, and had written two books on the topic. And the associate director of that lab went on to serve as the chief of the research grants office at the NIH, which would fund the Guatemalan work in early 1946.

"You had a very active venereal-disease division," says John Parascandola, a former historian of the Public Health Service and author of *Sex, Sin and Science: A History of Syphilis in America* (Praeger, 2008). Even after researchers demonstrated in 1943 that penicillin was an effective treatment for syphilis and gonorrhoea, they had many questions about

EVOLVING ETHICS

Sexually transmitted diseases (STDs) were a prime concern for health officials in the 1940s, and many medical studies — including the US experiments in Guatemala — used methods that would be considered unethical today. Although standards improved over the decades, clinical researchers continued to push the boundaries of acceptable science.

1941 A 12-month old baby in California is deliberately infected with herpes as part of an experiment. One editor objects to publishing the work, but it appears in 1942 in the *Journal of Pediatrics*.

1943 Penicillin demonstrated to be an effective treatment for syphilis and gonorrhoea.

The US military and Public Health Service combat STDs through an all-out campaign of research, treatment programmes and advertising.

1944 John Cutler works on a prison experiment in Terre Haute, Indiana, in which inmate volunteers are infected with STDs to test a prophylactic method.
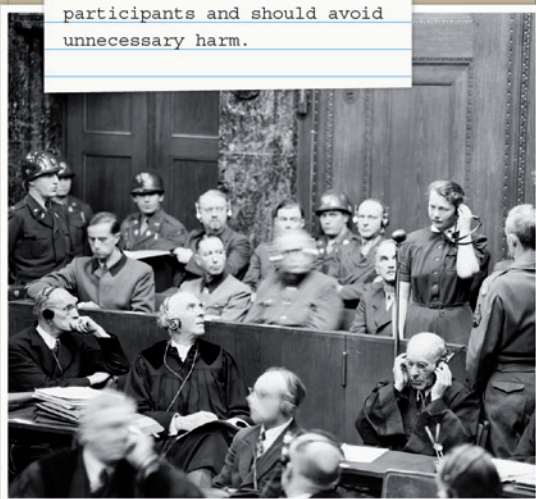
1946

Cutler starts experiments in **Guatemala** that eventually expose 1,308 prisoners, soldiers and patients at a psychiatric hospital to STDs. The US team also takes blood from 1,384 children to assess STD diagnostic tests. Evidence suggests that the participants in the study did not give their consent.

1946–7

The trial of Nazi doctors leads to the establishment of the **Nuremberg Code** of medical ethics, which holds that experimenters must obtain voluntary consent from participants and should avoid unnecessary harm.

1953

Cutler starts an experiment at Sing Sing prison in New York, in which he **injects syphilis into inmate volunteers**, some of whom had received an experimental vaccine.

preventing and treating those diseases and others. "You still had all these people who cut their teeth with venereal diseases and were interested in that topic. Certainly, the venereal-disease division in the 1940s didn't think the problem was licked."

The military, in particular, wanted to develop prophylactic techniques better than the 'pro kit' that had been in use for decades. After sex, servicemen were supposed to inject a solution containing silver into their penises to prevent gonorrhoea, and rub a calomel ointment over their genitals to prevent syphilis. The methods were painful, messy and not very effective.

To test treatments and prophylaxis, the Public Health Service had argued in late 1942 that it was crucial to give the disease to people under controlled conditions. Officials debated the legality and ethics of this, and even solicited the input of the US attorney general. They decided to do the work at a federal prison in Terre Haute, Indiana, using volunteer inmates.

Cutler was one of the doctors charged with carrying out the work. When the prison study began in September 1943, Cutler was 28, and had finished medical school only two years before. The researchers tried to infect prisoners by depositing bacteria — sometimes gathered from prostitutes arrested by the Terre Haute police — directly on the end of the penis. The experiment established several practices that Cutler would go on to use in Guatemala, including working with local law-enforcement agencies and prostitutes. But the researchers could not develop a means to effectively infect people — a necessary step towards testing prophylactic techniques. Within ten months, the experiments were abandoned.

CAPTIVE POPULATION

After Terre Haute, researchers began to plan a more ambitious study. They wanted to try causing infections through what they called normal exposure, in which people would have sex with infected prostitutes.

In 1945, a Guatemalan health official who was working for a year at the VDRL offered to host studies in his country. As director of the Guatemalan Venereal Disease Control Department, Juan Funes was uniquely positioned to help. Prostitution was legal in his country at the time, and sex workers were required to visit a clinic twice a week for examinations and treatment. Funes oversaw one of the main clinics, so he could recommend infected prostitutes for experiments. Cutler and other scientists at the VDRL were quickly sold on the idea: they proposed a programme, which was approved with a budget of US$110,450.

According to a Guatemalan report[3], the US plan was a clear violation of contemporary Guatemalan law, which made it illegal to knowingly spread venereal diseases. But the country was experiencing political upheaval in the mid-1940s and the bureaucracy did not object to the US plan. Government officials as high up as Luis Galich, head of the Guatemalan ministry of public health, were involved in the US study, and even President Juan José Arévalo, who had been elected in 1945, was at least aware of a syphilis experiment being done by US scientists. The study presented a chance to tap into US funding to upgrade Guatemala's inadequate public-health infrastructure, and to import scientific expertise.

Cutler arrived in the country in August 1946 and began setting up experiments. He planned to assess diagnostic blood tests, and to determine the effectiveness of penicillin and an agent called orvus-mapharsen in preventing STDs. At first, Cutler tried using infected prostitutes to spread gonorrhoea to soldiers: he and his team used various bacterial strains to inoculate sex workers, who then had intercourse with many men. Records show that one prostitute had sex with 8 soldiers in a period of 71 minutes. The team also carried out similar experiments using sex workers at a prison.

But it was hard to induce infections by the 'natural' method. So researchers turned to inoculation, swabbing the urethra with an infected solution, or using a toothpick to insert the swab deep into the urethra. At the National Psychiatric Hospital of Guatemala, scientists scratched male patients' penises before artificial exposure to improve infection rates, and injected syphilis into the spinal fluid of seven female patients.

According to the US bioethics commission's report, Cutler's team exposed 558 soldiers, 486 patients at the psychiatric hospital, 219 prisoners, 6 prostitutes and 39 other people to gonorrhoea, syphilis or

chancroid. But the commission was unable to determine how many people actually developed infections or how many of the participants were treated. Researchers also measured the accuracy of diagnostic tests in experiments that involved orphans and people with leprosy, as well as people from the psychiatric hospital, prison and the army.

The commission says there is no evidence that Cutler sought or obtained consent from participants, although in some cases he did get permission from commanding officers, prison officials and doctors who oversaw the patients at the psychiatric hospital. In a letter to his supervisor, John Mahoney, director of the VDRL, Cutler openly admits to deceiving patients at the psychiatric hospital, whom he was injecting with syphilis and later treating. "This double talk keeps me hopping," Cutler wrote.

Cutler and his colleagues treated some people brutally. In one case, detailed by the bioethics commission, the US doctors infected a woman named Berta, a patient at the psychiatric hospital, with syphilis, but did not treat her for three months. Her health worsened, and within another three months Cutler reported that she seemed close to death. He re-infected Berta with syphilis, and inserted pus from someone with gonorrhoea into her eyes, urethra and rectum. Over several days, pus developed in Berta's eyes, she started bleeding from her urethra and then she died.

Yet Cutler did do some good in Guatemala. He took steps to improve public health, initiating a venereal-disease treatment programme at the military hospital and developing a prophylactic plan for the army. He treated orphans for malaria, lobbied his supervisors to supply the army with penicillin — he was turned down — and trained local doctors and technicians. And he provided treatment for 142 people who may have had venereal disease but had not been exposed to it as part of the research.

At the prison, he reported that "we have found a very ready acceptance of our group, both on the part of the prison officials and the part of the inmates, which we think stems from the fact that we now have given them a program of care for venereal disease, which they have lacked in the past. Thus we feel that our treatment program is worthwhile and fully justified."

In the end, Cutler could claim no real success in his experiments, in part because he was never able to infect people reliably without resorting to extreme methods. He secured an extension to continue the experiments from June to December 1948, and he left Guatemala at the end of that year. Other researchers published some of the blood-test results, but Cutler did not publish his work on prophylactic methods. The experiments were not only unconscionable violations of ethics, the bioethics commission charges, they were also poorly conceived and executed.

A DISTINGUISHED CAREER

Despite the failures, the work burnished Cutler's credentials. A few months after he arrived home, the World Health Organization sent Cutler to India to lead a team demonstrating how to diagnose and treat venereal diseases. In the 1960s, he became a lead researcher in the infamous Tuskegee experiment in Alabama, in which hundreds of black men with syphilis were studied for decades without receiving treatment. He flourished in the Public Health Service and later became a professor of international health at the University of Pittsburgh in Pennsylvania. He died in 2003, well before details of the Guatemala experiments were exposed.

Michael Utidjian was an epidemiologist at Pittsburgh in the late 1960s and co-authored two papers with Cutler. He describes his former colleague as devoted to venereal-disease studies and enthusiastic about international research. "He did some pioneer work out in India using penicillin to treat the commoner STDs." But Utidjian says that Cutler was a flawed researcher. "I wouldn't rank him as a top-flight scientist or designer of studies." The two scientists collaborated on a study to test the effectiveness of a topical prophylaxis in prostitutes at a brothel in Nevada. However, the poor implementation of the experiment led to "pretty worthless" results, says Utidjian.

The participants in Cutler's Guatemalan study fared far worse than the doctor himself. Shuffling among the tin-roofed homes in Las Escaleras, Ramos is bone thin and speaks in a mumble, made worse by his lack of teeth. He says that he put off treatment until about ten years ago, when it became too painful to urinate. His son rushed him to a hospital,

1956

Researchers launch a 15-year study at the **Willowbrook State School** in New York, where they infect children with mental disabilities with hepatitis.

1963

Doctors **inject live cancer cells into elderly residents** at the Jewish Chronic Disease Hospital in New York without their consent. Outrage over the experiment helps to spark a debate about research ethics and prompts federal authorities to develop tighter controls on research.

1972

The United States halts a 40-year experiment in **Tuskegee**, Alabama, where hundreds of African American men with syphilis were observed for decades without being treated.

2010

Historian **Susan Reverby** exposes the abuses of Cutler's Guatemalan experiments, prompting **US President Barack Obama to apologize to Guatemala** and launch an investigation, which finds that the "experiments involved unconscionable violations of ethics".

Federico Ramos has suffered excruciating symptoms after US experiments.

where doctors inserted a catheter and later performed an operation.

Gonzalo Ramirez Tista lives in the same village as Ramos and says that his father, Celso Ramirez Reyes, also participated in the experiments during his three years in the army. He was required by the scientists to have sex with infected prostitutes. "They gave him an order, and it came from a superior," says Tista. They also gave him injections, and within days he noticed pus coming out of his penis. "He still had these symptoms when he left [the military], and he infected my mother." After his service, gonorrhoea left Reyes with sores, poor eyesight and lethargy.

Like Ramos's family, Tista is a party to the lawsuit seeking compensation from the US government. Neither man could provide documents to support their claims. But Pablo Werner, a doctor with Guatemala's Human Rights Ombudsman's office, has reviewed the cases and found that Ramos's and Reyes's stories are supported by the timing of their military service and details in the medical histories that they gave. Reyes is also named in a database of research participants that was compiled by Guatemala's National Police Historical Archive from Cutler's papers.

NEVER AGAIN

The US Department of Justice requested last month that the compensation case be dismissed, arguing that the courts are not the "proper forum" for it. But last September, a panel of the presidential bioethics commission recommended[4] that the government set up a general compensation system for test participants harmed by federally funded research.

This January, the US Department of Health and Human Services committed nearly $1.8 million to improving the treatment of STDs in Guatemala and strengthening ethics training there regarding research on humans. The plaintiffs are not satisfied and intend to press their case, says Piper Hendricks, an international human-rights lawyer with Conrad & Scherer in Fort Lauderdale, Florida, who is representing them.

As the case moves forward, researchers are wrestling with how to judge the actions of Cutler's team, and how to prevent such abuses from happening again. The bioethics commission argues that Cutler and his superiors knew that they were violating the medical ethics of their day, because they had sought the consent of participants in Terre Haute. And in Guatemala, the researchers took steps to suppress knowledge of their work. One colleague told Cutler that the US surgeon general "is very much interested in the project and a merry twinkle came into his eye when he said, 'You know, we couldn't do such an experiment in this country'."

But the ethical landscape was evolving rapidly at the time. The standards of the 1940s were "a lot murkier" than those of today, says Susan Lederer, a bioethicist at the University of Wisconsin–Madison. "The idea that it was so clear in 1946 to me doesn't ring true."

In late 1946, after Cutler had started his work in Guatemala, 23 Nazi doctors and officials went on trial in Nuremberg, Germany, for the inhuman experiments that they had carried out in concentration camps during the Second World War. From that trial emerged the Nuremberg Code, a set of principles that mandated that experimenters obtain voluntary consent from participants, that participants be capable of giving such consent and that experiments avoid unnecessary physical and mental harm.

Although such tight standards were not entirely foreign to researchers before the Nuremberg trials, few followed them. In 1935, for example, the Supreme Court of Michigan stated that researchers could get consent from caregivers of participants, which Cutler did in a sense when he consulted commanding officers and other officials. Many of Cutler's participants were poor, uneducated people from indigenous populations, whom the scientists viewed as incapable of understanding the experiments.

At the time, some of the United States's top researchers worked without obtaining consent from individuals. Jonas Salk, who later earned fame for developing the polio vaccine, and Thomas Francis Jr, a leading influenza researcher, intentionally infected patients at a psychiatric hospital in Ypsilanti, Michigan, with influenza in 1943 (ref. 5). There is evidence that the patients did not all consent to the experiments.

Cutler and his superiors apparently thought it was acceptable in Guatemala to cross ethical lines that they would not have breached at home — an issue that raises concern today, with Western companies increasingly running clinical trials in foreign countries, particularly in developing nations. In 2010, the US Department of Health and Human Services investigated all requests by companies to market their drugs in the United States, and found that in 2008, nearly 80% of approved applications used data from clinical trials in other countries.

Developing nations often have lower medical standards than developed countries, and can't enforce rules as effectively. In India, for example, human-rights activists and members of parliament say that foreign drug companies often test experimental drugs on poor, illiterate people without obtaining their consent or properly explaining the risks.

And in 2009, the pharmaceutical giant Pfizer agreed to pay up to $75 million to settle lawsuits over the deaths of Nigerian children who had participated in tests of an experimental antibiotic. Nigerian officials and activists had claimed that the company had acted improperly by, for example, not obtaining proper authorization or consent. But Pfizer denies the allegations and did not admit any wrongdoing in the settlement.

Ethicists also warn about practices viewed as acceptable today, such as testing medications on patients who are extremely ill, and who see new treatments as their only hope, no matter how dangerous they are. Lederer notes that some trials of cancer drugs involve particularly toxic compounds. In the future, she says, "people might say, 'how can people who are so sick make informed decisions?'"

For Grady, the lessons from Guatemala are fundamental tenets of bioethics: not every method is acceptable, transparency is key and scientists should remember that they are working with human beings.

But in clinical research, she says, the ethical lines aren't always well defined. "When you get to the details of what that means in a particular case, people disagree." And that may be the most troubling lesson of the Guatemalan experiments. In any era, many if not most researchers might agree that a certain practice or rule is justified and necessary. But for later generations, the barbarism of the past seems only too obvious. ■ **SEE EDITORIAL P.132**

Matthew Walter *is a freelance writer in New York. Additional reporting provided by* **Richard Monastersky**.

1. Presidential Commission for the Study of Bioethical Issues *Ethically impossible: STD Research in Guatemala from 1946 to 1948* (2011).
2. Reverby, S. M. *J. Policy Hist.* **23**, 6–28 (2011).
3. Comisión Presidencial para el Esclarecimiento de los Experimentos Practicados con Humanos en Guatemala, durante el periodo 1946 a 1948 *Consentir el Daño* (2011).
4. Presidential Commission for the Study of Bioethical Issues *Research Across Borders* (2011).
5. Francis, T. Jr, Salk, J. E., Pearson, H. E. & Brown, P. N. *J. Clin. Invest.* **24**, 536–546 (1945).

COMMENT

Pathogenic H5N1 avian influenza has led to the culling of hundreds of millions of birds. A human-transmissible form could have much worse consequences.

Adaptations of avian flu virus are a cause for concern

Members of the US **National Science Advisory Board for Biosecurity** explain its recommendations on the communication of experimental work on H5N1 influenza.

We are in the midst of a revolutionary period in the life sciences. Technological capabilities have dramatically expanded, we have a much improved understanding of the complex biology of selected microorganisms, and we have a much improved ability to manipulate microbial genomes. With this has come unprecedented potential for better control of infectious diseases and significant societal benefit. However, there is also a growing risk that the same science will be deliberately misused and that the consequences could be catastrophic. Efforts to describe or define life-sciences research of particular concern have focused on the possibility that knowledge or products derived from such research, or new technologies, could be directly misapplied with a sufficiently broad scope to affect national or global security. Research that might greatly enhance the harm caused by microbial pathogens has been of special

concern[1–3]. Until now, these efforts have suffered from a lack of specificity and a paucity of concrete examples of 'dual use research of concern'[3]. Dual use is defined as research that could be used for good or bad purposes. We are now confronted by a potent, real-world example.

Highly pathogenic avian influenza A/H5N1 infection of humans has been a serious public-health concern since its identification in 1997 in Asia. This virus rarely infects humans, but when it does, it causes severe disease with case fatality rates of 59% (ref. 4). To date, the transmission of influenza A/H5N1 virus from human to human has been rare, and no human pandemic has occurred. If influenza A/H5N1 virus acquired the capacity for human-to-human spread and retained its current virulence, we could face an epidemic of significant proportions. Historically, epidemics or pandemics with high mortalities have been documented when

humans interact with new agents for which they have no immunity, such as with *Yersinia pestis* (plague) in the Middle Ages and the introduction of smallpox and measles into the Americas after the arrival of Europeans.

Recently, several scientific research teams have achieved some success in modifying influenza A/H5N1 viruses such that they are now transmitted efficiently between mammals, in one instance with maintenance of high pathogenicity. This information is very important because, before these experiments were done, it was uncertain whether avian influenza A/H5N1 could ever acquire the capacity for mammal-to-mammal transmission. Now that this information is known, society can take steps globally to prepare for when nature might generate such a virus spontaneously. At the same time, these scientific results also represent a grave concern for global biosecurity, biosafety and public health. Could ▶

▶ this knowledge, in the hands of malevolent individuals, organizations or governments, allow construction of a genetically altered influenza virus capable of causing a pandemic with mortality exceeding that of the 'Spanish flu' epidemic of 1918? The research teams that performed this work did so in a well-intended effort to discover evolutionary routes by which avian influenza A/H5N1 viruses might adapt to humans. Such knowledge may be valuable for improving the public-health response to a looming natural threat. And, to their credit and that of the peer reviewers selected by the journals *Science* and *Nature*, the journals themselves, as well as the US government, it was recognized before their publication that these experiments had dual use of concern potential.

The US government asked the National Science Advisory Board for Biosecurity (NSABB; go.nature.com/oeryit) to assess the dual-use research implications of two as-yet-unpublished manuscripts on the avian influenza A/H5N1 virus, to consider the risks and benefits of communicating the research results and to provide findings and recommendations regarding the responsible communication of this research. In our deliberations, we first assessed the potential risks and consequences of the misuse of the information to cause harm to the public.

Risk assessment of public harm is challenging because it necessitates consideration of the intent and capability of those who wish to do harm, as well as the vulnerability of the public and the status of public-health preparedness for both deliberate and accidental events. We found the potential risk of public harm to be of unusually high magnitude. In formulating our recommendations to the government, scientific journals and to the broader scientific community, we tried to balance the great risks against the benefits that could come from making the details of this research known. Because the NSABB found that there was significant potential for harm in fully publishing these results and that the harm exceeded the benefits of publication, we therefore recommended that the work not be fully communicated in an open forum. The NSABB was unanimous that communication of the results in the two manuscripts it reviewed should be greatly limited in terms of the experimental details and results.

This is an unprecedented recommendation for work in the life sciences and our analysis was conducted with careful consideration both of the potential benefits of publication and of the potential harm that could occur from such a precedent. Our concern is that publishing these experiments in detail would provide information to some person, organization or government that would help them to develop similar mammal-adapted influenza A/H5N1 viruses for harmful purposes. We believe that as scientists and as members of the general public, we have a primary responsibility 'to do no harm' as well as to act prudently and with some humility as we consider the immense power of the life sciences to create microbes with novel and unusually consequential properties. At the same time, we acknowledge that there are clear benefits to be realized for the public good in alerting humanity of this potential threat and in pursuing those aspects of this work that will allow greater preparedness and the potential development of novel strategies leading to future disease control. By recommending that the basic result be communicated without methods or details, we believe that the benefits to society are maximized and the risks minimized. Although scientists pride themselves on the creation of scientific literature that defines careful methodology that would allow other scientists to replicate experiments, we do not believe that widespread dissemination of the methodology in this case is a responsible action.

The life sciences have reached a crossroads. The direction we choose and the process by which we arrive at this decision must be undertaken as a community and not relegated to small segments of government, the scientific community or society. Physicists faced a similar situation in the 1940s with nuclear weapons research, and it is inevitable that other scientific disciplines will also do so.

> **"We found the potential risk of public harm to be of unusually high magnitude."**

Along with our recommendation to restrict communication of these particular scientific results, we discussed the need for a rapid and broad international discussion of dual-use research policy concerning influenza A/H5N1 virus with the goal of developing a consensus on the path forward. There is no doubt that this is a complex endeavour that will require diligent and nuanced consideration. There are many important stakeholders whose opinions need to be heard at this juncture. This must be done quickly and with the full participation of multiple societal components.

We are aware that the continuing circulation of the highly pathogenic avian influenza A/H5N1 virus in Eurasia — where it is constantly found to cause disease in animals of particular regions — constitutes a continuing threat to humankind. A pandemic, or the deliberate release of a transmissible highly pathogenic influenza A/H5N1 virus, would be an unimaginable catastrophe for which the world is currently inadequately prepared. It is urgent to establish how best to facilitate the much-needed research as well as minimize potential dual use.

To facilitate and motivate this process, we also discussed the possibility of the scientific community participating in a self-imposed moratorium on the broad communication of the results of experiments that show greatly enhanced virulence or transmissibility of such potentially dangerous microbes as the influenza A/H5N1 virus. This moratorium would run until consensus is reached on the balance that must be struck between academic freedom and protecting the greater good of humankind from potential danger. With proper diligence and rapid achievement of a consensus on a proper path forward, this could have little detrimental effect on scientific progress but significant effect on diminishing risk.

There are many parallels with the situation in the 1970s and recombinant DNA technologies[5–7]. The Asilomar Conference in California in 1975 was a landmark meeting important to the identification, evaluation and mitigation of risks posed by recombinant DNA technologies. In that case, the research community voluntarily imposed a temporary moratorium on the conduct of recombinant DNA research until they could develop guidance for the safe and responsible conduct of such research. We believe that this is another Asilomar-type moment for public-health and infectious-disease research that urgently needs our attention. ∎

Kenneth I. Berns, **Arturo Casadevall**, **Murray L. Cohen**, **Susan A. Ehrlich**, **Lynn W. Enquist**, **J. Patrick Fitch**, **David R. Franz**, **Claire M. Fraser-Liggett**, **Christine M. Grant**, **Michael J. Imperiale**, **Joseph Kanabrocki**, **Paul S. Keim**, **Stanley M. Lemon**, **Stuart B. Levy**, **John R. Lumpkin**, **Jeffery F. Miller**, **Randall Murch**, **Mark E. Nance**, **Michael T. Osterholm**, **David A. Relman**, **James A. Roth** and **Anne K. Vidaver** *are members of the US National Science Advisory Board for Biosecurity.*

1. National Research Council Committee on Research Standards and Practices to Prevent the Destructive Application of Biotechnology. *Biotechnology Research in an Age of Terrorism* (National Academies Press, 2004); available at http://go.nature.com/4vl3ye
2. National Research Council Committee on Advances in Technology and the Prevention of Their Application to Next Generation Biowarfare Threats. *Globalization, Biosecurity, and the Future of the Life Sciences* (National Academies Press, 2006); available at http://go.nature.com/hktvtc
3. National Science Advisory Board for Biosecurity. *Strategic Plan for Outreach and Education on Dual Use Research Issues* (NSABB, 2008); available at http://go.nature.com/nuriw4
4. World Health Organization. *Cumulative Number of Confirmed Human Cases for Avian Influenza A(H5N1) reported to WHO, 2003–2012*; available at http://go.nature.com/epb7ts
5. Singer, M. & Soll, D. *Science* **181**, 1114 (1973).
6. Berg, P., Baltimore, D., Brenner, S., Roblin, R. O. & Singer, M. F. *Proc. Natl Acad. Sci. USA* **72**, 1981–1984 (1975).
7. Singer, M. & Berg, P. *Science* **193**, 186–188 (1976).

Flu transmission work is urgent

Yoshihiro Kawaoka explains that research on transmissible avian flu viruses needs to continue if pandemics are to be prevented.

Highly pathogenic avian H5N1 influenza viruses first proved lethal in humans in 1997 in Hong Kong. Since 2003, 578 confirmed infections have resulted in 340 deaths (go.nature.com/epb7ts). Now widespread in parts of southeast Asia and the Middle East, H5N1 viruses have killed or led to the culling of hundreds of millions of birds.

To date, H5N1 viruses have not been transmitted between humans. Some experts have argued that it is impossible. But given the potential consequences of a global outbreak, it is crucial to know whether these viruses can ever become transmissible. Work by my group (accepted by *Nature*) and an independent study (accepted by *Science*) led by Ron Fouchier of the Erasmus Medical Center in Rotterdam, the Netherlands, suggest that H5N1 viruses have the potential to spread between mammals. As the risks of such research and its publication are debated by the community, I argue that we should pursue transmission studies of highly pathogenic avian influenza viruses with urgency.

To determine whether H5N1 viruses could be transmitted between humans, my team generated viruses that combined the H5 haemagglutinin (HA) gene with the remaining genes from a pandemic 2009 H1N1 influenza virus. Avian H5N1 and human pandemic 2009 viruses readily exchange genes in experimental settings, and those from a human virus may facilitate replication in mammals. Indeed, we identified a mutant H5 HA/2009 virus that spread between infected and uninfected ferrets (used as models to study the transmission of influenza in mammals) in separate cages via respiratory droplets in the air. Thus viruses possessing an H5 HA protein can transmit between mammals.
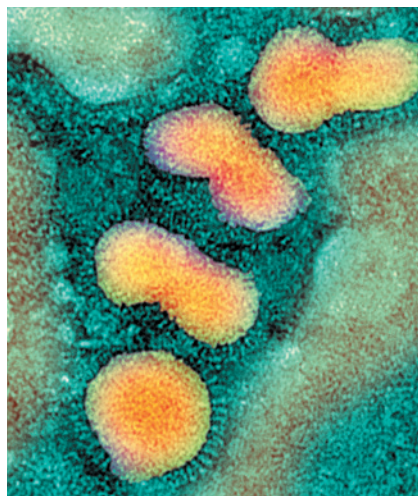
Our results also show that not all transmissible H5 HA-possessing viruses are lethal. In ferrets, our mutant H5 HA/2009 virus was no more pathogenic than the pandemic 2009 virus — it did not kill any of the infected animals. And, importantly, current vaccines and antiviral compounds are effective against it.

Fouchier and his team also generated a transmissible H5 HA-possessing virus — meaning that two independent studies have demonstrated the potential for transmissibility of H5 HA-possessing viruses between ferrets. Their mutant H5 HA virus, generated in the genetic background of an H5N1 virus, did kill infected ferrets.

Some people have argued that the risks of such studies — misuse and accidental release, for example — outweigh the benefits. I counter that H5N1 viruses circulating in nature already pose a threat, because influenza viruses mutate constantly and can cause pandemics with great losses of life. Within the past century, 'Spanish' influenza, which stemmed from a virus of avian origin, killed between 20 million and 50 million people. Because H5N1 mutations that confer transmissibility in mammals may emerge in nature, I believe that it would be irresponsible not to study the underlying mechanisms.

The new work has implications for pan-

H5N1 avian influenza virus particles.

demic preparedness. There is an urgent need to expand development, production and distribution of vaccines against H5 viruses, and to stockpile antiviral compounds. Both studies identify specific mutations in HA that confer transmissibility in ferrets to H5 HA-possessing viruses. A subset of these mutations has been detected in H5N1 viruses circulating in certain countries. It is therefore imperative that these viruses are monitored closely so that eradication efforts and countermeasures (such as vaccine-strain selection) can be focused on them, should they acquire transmissibility.

Consequently, I believe that the benefits of these studies — the knowledge that H5 HA-possessing viruses pose a risk and the ability to monitor them and develop countermeasures — outweigh the risks. High biosafety and security standards can be met. Our experiments were carried out in a high-containment facility by a small group of highly trained individuals who operate under strict procedures to prevent the accidental release of viruses.

However, the US National Science Advisory Board for Biosecurity (NSABB) has recommended that details of both studies (including the mutations that confer transmissibility) should be restricted, and released only to select individuals on a 'need-to-know' basis. I acknowledge the advisory role of the NSABB, but I do not concur with its decision.
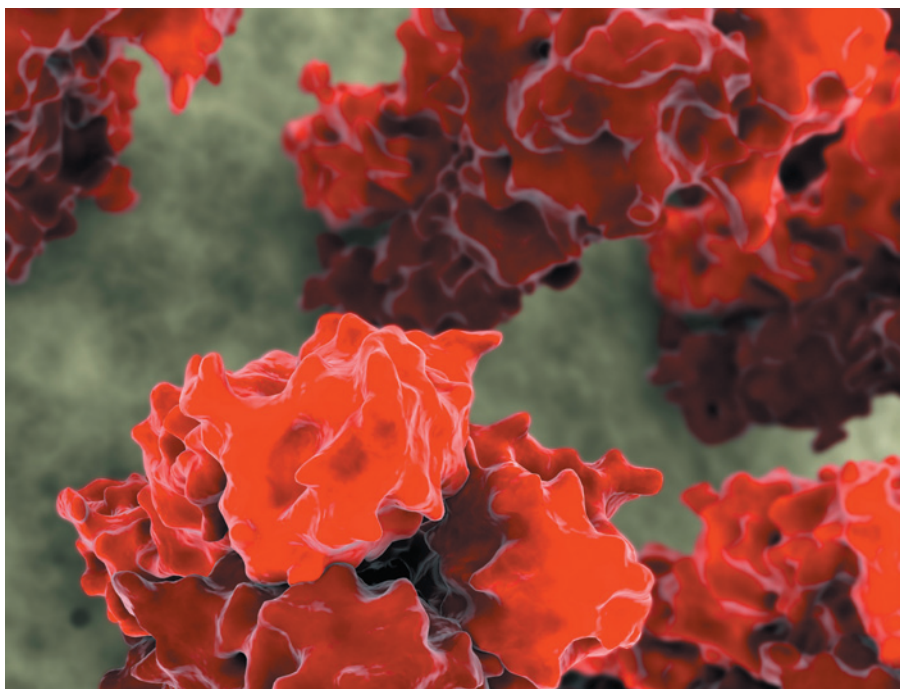
The primary justification for the NSABB's recommendation is that publication of our data "could enable replication of the experiments by those who would seek to do harm" (go.nature.com/nywkdy). But redacting our papers will not eliminate that possibility — there is already enough information publicly available to allow someone to make a transmissible H5 HA-possessing virus.

The mechanism that the US government proposes for releasing data would also be unwieldy. Thousands of applications to access the research are likely to be filed, and potential background checks would create a huge administrative burden. We cannot afford to lose time if we are to combat emerging pandemic threats. Even if an efficient process can be established, it would be difficult to enforce continued confidentiality in the scientific community.

By contrast, wide data dissemination will attract researchers from other areas to contribute to the field. This is crucial, because new ideas are needed to answer some of the most urgent questions. For example, the specific mutations that we identified suggest that influenza transmission is more complex than anticipated and involves not only the receptor-binding properties of HA, but other biological and physical properties.

The redaction of our manuscript, intended to contain risk, will make it harder for legitimate scientists to get this information while failing to provide a barrier to those who would do harm. To find better solutions to dual-use concerns, the international community should convene to discuss how to minimize risk while supporting scientific discovery. Flu investigators (including me) have agreed to a 60-day moratorium on avian flu transmission research (go.nature.com/ttivj5) because of the current controversy. But our work remains urgent — we cannot give up. ∎

Yoshihiro Kawaoka *is at the University of Tokyo and the University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.*
e-mail: kawaokay@svm.vetmed.wisc.edu

NIBSC/SPL

Influenza viruses use the haemagglutinin (red) on their surfaces to bind to host cells.

R. ANDRADE/3DCIENCIA/SPL

Q&A Paul S. Keim for the NSABB

Reasons for proposed redaction of flu paper

The US National Science Advisory Board for Biosecurity (NSABB) recommends[1] (see page 153) that two papers reporting experimental adaptations of influenza viruses should be published in a form that withholds essential information. Yoshihiro Kawaoka of the University of Wisconsin-Madison and his colleagues show[2] that a mutated H5 haemagglutinin combined with genes from a pandemic human H1N1 virus is transmissible in respiratory droplets between ferrets. Ron Fouchier of the Erasmus Medical Center in Rotterdam, the Netherlands, and his colleagues report[3] adaptations that make highly pathogenic avian H5N1 virus transmissible. To assist decision-making in response to the NSABB's challenging recommendation, Nature *asked the board to explain the reasoning behind its conclusion for the Kawaoka paper. Acting NSABB chair Paul S. Keim coordinated the board's answers.*

You have recommended that both papers be published in a redacted form even though the Kawaoka paper does not report transmissibility of fully avian H5N1. Why?
The Kawaoka and Fouchier manuscripts are indeed different, and the committee spent a considerable amount of time on each. The concern for the Kawaoka paper is that the authors provide a method for producing a transmissible H5N1 reassortant virus. They demonstrate the compatibility of segments of the 2009 pandemic influenza (A(H1N1) pdm09) backbone with H5 haemagglutinin (HA) to produce a virus that can be transmitted between ferrets.

Another concern is the high transmissibility of this strain of H1N1 and the ability of influenza viruses to continually reassort in pigs. The detection of novel H3N2 reassortants in humans in 2011 raises additional worries about the use of reassortant H5 haemagglutinin on an H1N1 backbone.

Publication of the experimental details from either paper would, in our view, allow others to replicate the experiments and move closer to production capability of an avian flu virus for humans that is highly pathogenic and transmissible by respiratory aerosols.

Perhaps in time, the global community will decide that the availability of these data is not of concern. However, a decision to release these data now is not reversible, whereas delayed publication of these details can easily be reversed on some future date.

The focus of your statement[1] is on the risks of high pathogenicity. As Kawaoka's H5 HA/H1N1 virus is not highly pathogenic, why consider it a public-health risk?
Although the reported pathogenicity[2] of the generated viruses is no greater than that of A(H1N1)pdm09, we believe that the techniques described could be used to generate other viruses with H5 HA that have potentially much greater pathogenicity.

The fact that humans have no previous immunological experience with H5 infections could lead to a more widespread pandemic than that of 2009, as part of the population had pre-existing immunity to A(H1N1)pdm09 because of exposure to other H1N1 strains before 1957.

Although the work performed by Kawaoka and his colleagues[2] is an important contribution, it crosses a line and raises the question of when the risks of widely disseminated experimental detail outweigh the benefits.

Once mammal-to-mammal respiratory transmission is accomplished, the virus is likely to adapt on its own to enhance the efficiency of respiratory transmission. If such viruses were misused or escaped from the lab, they would evolve in ways that cannot be predicted. H5N1 has been in birds since at least 1996, and despite the almost 600 human infections of which we are aware, this virus has not yet become efficiently transmissible between mammals. There might be good reasons why it hasn't achieved this capacity, including inherent biological limitations.

The artificial evolution of a new mammal-adapted H5N1 virus, as reported in these two papers, has removed the natural barriers that might have existed. Accomplishing this in the lab, however, doesn't mean that it can occur naturally.

We also need to consider the potential role of, and impact on, other species. Pigs are a well-known 'mixing vessel' for influenza viruses. This mixing could lead to the emergence of new antigenically shifted viruses.

Sialic acids in the respiratory tracts of pigs attach to sugars by both α2,6-linkage (favoured as receptors by H1N1) and α2,3-linkage (favoured by H5N1). This makes pigs susceptible to infection by the H5N1 virus and by the pandemic A(H1N1)pdm09 virus. However, like humans, pigs do not easily transmit the H5N1 virus and it is not very pathogenic in pigs. Adapting the H5 HA to bind to α2,6-linked saccharides is likely to enhance transmission of H5 viruses between pigs.

Dogs and cats are also susceptible to H5N1 and A(H1N1)pdm09. Because humans have close contact with all three of these species, we need to be concerned about the possibility that the mutated H5 viruses may be

transmitted to and from humans and any of these species. This could greatly complicate disease control. There are an estimated 2 billion domestic pigs globally. An influenza virus with high morbidity and mortality in pigs that could be transmitted between them could have a devastating effect on the world's food supply; pigs would then serve as an intermediate host of virus adaptation to humans.

It is likely that the H5 HA/H1N1 mutations described in the Kawaoka paper are insufficient to provide a blueprint to construct a transmissible and highly pathogenic wholly avian H5N1 virus — additional mutations may be required. Why do you think publication is still risky?
The fact that Kawaoka's specific virus and mutations might not be the feared H5N1 pandemic strain is not the point. It is that this laboratory created a virus that has now bypassed apparent barriers to evolution in the wild. If this virus were to escape by error or by terror, we must ask whether it would cause a pandemic. The probability is unknown, but it is not zero.

Kawaoka's work establishes the feasibility and pinpoints those particular viruses as candidates for producing a transmissible virus. Advances in basic science are incremental and cumulative. This work significantly advances the ability to construct an H5 virus with catastrophic potential. This altered H5 HA gene could be combined with other influenza virus genes possibly leading to a pandemic. A major concern is that the human population does not have immunity to H5.

Do we want to take this gamble and thereby potentially jeopardize public health and safety as well as risk the resulting economic consequences, based simply on a belief that this probably won't occur?

Several of our independent advisers felt that the likelihood of H5 HA/H1N1-based influenza being used as an agent of bioterrorism is low, given that it cannot be targeted to a specific population, and vaccines and drugs exist to combat it, as Kawaoka reports. Why are you concerned?
No one should presume to know all the ways in which influenza virus could be misused, and the motivations for doing so, but the consequences could be catastrophic. There are many scenarios to consider, ranging from mad lone scientists, desperate despots and members of millennial doomsday cults to nation states wanting mutually assured destruction options, bioterrorists or a single person's random acts of craziness. These are low-probability events, but they could introduce a new evolutionary H5N1 seed into the environment that seems not to exist in nature. This might not cause

◯ NATURE.COM
See the mutant flu
special at:
go.nature.com/mhmibi

a pandemic instantly, but it could start the virus on a new path for pandemic evolution.

H5N1 vaccines and a class of antivirals do exist, but not in sufficient quantities anywhere in the world — and in the event of an H5N1 influenza pandemic, they may have limited value. Neither the vaccine-generating capacity nor our pharmaceutical industry could cope with the rapidity of a pandemic that could potentially affect 7 billion people. The world population is immunologically naive to the H5 family of viruses. Work published in October 2011 indicates that current influenza virus vaccines are less effective than previously thought[4]. There is only one family of antiviral drugs available (the neuraminidase inhibitors) and resistance in A/H5N1 has already been documented[5].

Several of our independent advisers also felt that the Kawaoka paper has important messages for surveillance and prevention preparedness. Do you disagree?
The major benefit of the work is to alert humanity to the potential threat posed by H5N1. It is important to convey how unprepared, on every level, the world is for an H5N1 pandemic. Initially, some NSABB members also advocated for communication to enhance surveillance. But further examination of the issue lessened our enthusiasm.

The practical benefits of this work may be limited, because there are many paths to the evolution of human-transmissible H5N1. Consequently, the utility of the specific mutations presented in this manuscript for surveillance and countermeasure development is unknown — and may even be misleading for surveillance. Furthermore, it is unlikely that the detection of these mutations in a single virus either in humans or in other animals will provide sufficient lead time so that effective public-health and safety action could be taken to pre-empt a pandemic.

In 2005, the highly pathogenic 1918 influenza virus was reconstructed and the information was published in full[6]. Why do you feel that the work on transmissible H5N1 is riskier?
Our unanimous decision in 2005 to recommend publication of the 1918 papers was a difficult one, reached after much debate. This judgement of the 1918 papers was made in the context of the time and with an awareness that this dual-use research — research that could be used for good or bad purposes — was very close to a line beyond which information would need to be restricted. There were two primary reasons that the NSABB did not consider the 1918 research paper to be of sufficient concern to warrant limited dissemination.

First, the 1918 H1N1 virus had already existed in nature as result of the 1918 pandemic. Given that this virus (and derivative strains) continued to circulate in the

population until 1957, and then returned in 1977, it was suggested that there may be a critical level of pre-existing population-level immunity from these exposures. In 2009, we learned that was not the case. However, exposure to A(H1N1)pdm09 does provide some cross protection against the 1918 virus, making transmission of the 1918 virus today much less likely.

Second, the eight mutations that resulted in the construction of the 1918 virus were obtained only from forensic investigation of pathology slides of lungs from fatal cases of 1918 influenza or from tissue from exhumed bodies of deceased patients who were buried in Alaskan permafrost. In 2005, it was considered highly unlikely that anyone with an intent for harm could reasonably assemble the required genes to make a 1918 virus. Now this argument holds little validity.

Are you only worried about research that specifically investigates H5N1 transmission?
We remain concerned about any type of research that enhances the virulence of influenza virus, facilitates trans-species or host-to-host infection or renders the virus resistant to all available drugs or vaccine-induced immunity. Consequently, we are concerned about the many variants of influenza virus that currently infect other animal species, because these studies[2,3] show that some of these variants could also potentially be adapted for mammalian transmission or could provide raw materials for enhancing the virulence of mammalian-adapted viruses.

For avian and other highly pathogenic flu strains, experiments should be vetted carefully before being conducted. At this time, there is no formal, standardized mechanism for screening proposals and papers that contain dual-use research of concern, apart from assessments by authors, editors and reviewers. The NSABB has recommended a broad oversight mechanism for such research in the United States[7].

We believe that a discussion ought to take place across the scientific, public-health and policy communities about those experiments that fall within the criteria of dual-use research of concern. ∎

Paul S. Keim *is acting chair of the NSABB.*
e-mail: paul.keim@nau.edu

1. Berns, K. I. *et al. Nature* **482,** 153–154 (2012).
2. Kawaoka, Y. *et al. Nature* (in the press).
3. Fouchier, R. *et al. Science* (in the press).
4. Osterholm, M. T., Kelley, N. S., Sommer, A. & Belongia, E. A. *Lancet Infect. Dis.* **12,** 36–44 (2012).
5. de Jong, M. D. *et al. N. Engl. J. Med.* **353,** 2667–2672 (2005).
6. Tumpey, T. M. *et al. Science* **310,** 77–80 (2005).
7. National Science Advisory Board for Biosecurity. *Proposed Framework for the Oversight of Dual Use Life Sciences Research: Strategies for Minimizing the Potential Misuse of Research Information* (NSABB, 2007); available at http://go.nature.com/s7cpo7

Young chimpanzees watch and mimic an adult as she digs for termites, showing that the ability to learn by observing others is not unique to humans.
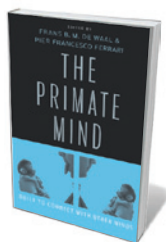
PRIMATE COGNITION

Copy that

The past decade has seen a revolution in our perception of primates' social brains, says **Christian Keysers**.

Every time my 18-month-old daughter sees me using a tool, she tries to copy me. She steals my pen to write, and excitedly brushes the few teeth she has when I brush mine. Such a capacity for connecting with and learning from other minds also manifests itself in the empathy we feel with other people's emotions, and in our ability to understand others' goals and help them. Through that ability, we can create and manage the complex social world that is arguably the key to our species' dominance.

Ten years ago, human minds were thought to be unique in their ability to connect. But as *The Primate Mind* shows, there has been a revolution in our understanding. This collection of essays, the result of a 2009 conference organized by primatologist Frans de Waal and ethologist Pier Francesco Ferrari, presents an authoritative, surprising and enriching picture of our monkey and ape cousins. We now know that they have remarkably sophisticated social minds, and that their poor performance in social tasks set by humans was more a result of

The Primate Mind: Built to Connect With Other Minds
EDITED BY FRANS B. M. DE WAAL AND PIER FRANCESCO FERRARI
Harvard University Press: 2012. 416 pp. $49.95, £36.95

researchers asking the wrong questions than deficiencies in their experimental subjects.

For example, a chapter by psychologists April Ruiz and Laurie Santos explores whether non-human primates can monitor where others are looking and use that information in their own decision-making — a test of whether the animal understands what another perceives. Primatologists first tested this by seeing whether monkeys followed an experimenter's gaze to find a box containing food. The animals performed unexpectedly poorly. But changing the task from cooperation to competition unleashed the primates' true potential: macaques readily stole food from humans who looked

away, but refrained from doing so when watched. Placing the task in a setting more relevant to macaque social life, which is less cooperative than our own, emphasized the continuity between our social mind and that of our primate ancestors.

The study of imitation has followed a similar path. In the early 2000s, when I began comparing how the macaque and human brains respond to the sight and sound of others' actions, I was struck by the similarities between the two species' frontal and parietal mirror systems. This system maps the actions of others onto the observer's own actions. In humans, it is thought to allow imitation, and we suspected that the same should go for monkeys. But to my surprise, primatologists at the time believed that monkeys and apes could not imitate.

The Primate Mind shows how this discrepancy between neural similarities and behavioural dissimilarities has been resolved. There is more than one way to copy others: one can either mimic every detail, or achieve the same goal by different means. Recent studies, reviewed in chapters by cognitive biologist Ludwig Huber and by primatologist Andrew Whiten and his colleagues, reveal that apes will rationally shift between these alternatives.

If apes see a man pressing a button with his head because his hands are occupied holding a blanket, they will press the button with their hands. Apes thus demonstrate something smarter than simple imitation — the ability to infer why a person is doing something in a particular way. But

if the man's hands are not occupied, giving the ape no clue as to why the person would push a button with his head, chimpanzees tend also to use their heads. It is one of many illustrations of how easy it is to misinterpret experimental results: the apes' ability to copy the details of an action only when it makes sense was misinterpreted as an inability to imitate fine details.

The essay by Whiten and his colleagues shows us that primate imitation is sometimes best studied with ape demonstrators. Chimps in a sanctuary in Uganda have been seen moving their hands up and down in synchrony with a chimp cracking a nut with a stone, and so acquiring the same skill. Many similar examples give uncontestable evidence that apes and monkeys, and even dogs and birds, can learn how to perform an action by observing others in natural environments. The disagreement between the neurological and behavioural evidence has dissolved.

Instead, we now understand that mirror neurons map the sight of observed actions onto motor programmes that allow an animal to achieve the observed goal through a variety of different actions. By bringing neuroscience and behaviour together, these findings pave the way to a deep biological understanding of how we learn by observing others. One day, this knowledge might inspire the design of robots that watch your skilled actions, and then do as you did.

> *"Our primate cousins share empathy and the inclination to cooperate."*

One by one, claims to human uniqueness have fallen. Other essays by de Waal and anthropologists Brian Hare and Jingzhi Tan show that our primate cousins share empathy and the inclination to cooperate. Apes console other apes after conflict. Chimps overcome their fear of water to save a drowning chimp. Monkeys can favour actions that benefit other monkeys. Apes even recruit other apes to collaborate with them, and will negotiate a fair distribution of pay-offs.

Clearly, we are different from other primates. I have never seen macaques display anything like a toddler's eagerness to imitate. *The Primate Mind* suggests that it may not be the capacity to imitate, but the motivation to do so that sets us apart from other animals. Like all good suggestions, this opens the door to more questions about the mechanisms and evolution of such motivation — and, ultimately, about how our own social minds evolved from the deeply interconnected minds of our primate cousins. ∎

Christian Keysers *is a professor for the social brain at the Netherlands Institute for Neuroscience, Amsterdam, the Netherlands, and author of* The Empathic Brain.
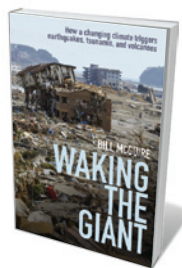*e-mail: c.keysers@nin.knaw.nl*

Books in brief

The Quantum Exodus: Jewish Fugitives, the Atomic Bomb, and the Holocaust
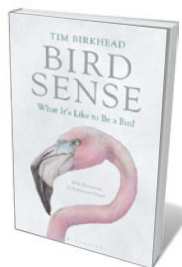*Gordon Fraser* OXFORD UNIVERSITY PRESS *267 pp. £25 (2012)*
It is no accident that the Holocaust and the Manhattan Project occurred at the same time, says science writer Gordon Fraser. Adolf Hitler's policies created a diaspora of exceptional Jewish physicists, who realized both the potential of atomic weaponry and the ambitions of the Nazi regime. Fear of the regime drove them to develop the weapons, convinced that they were locked in a race, Fraser says. However, as he notes, the Nazis' focus on the Final Solution actually distracted them from pursuing the bomb.

Waking the Giant: How a Changing Climate Triggers Earthquakes, Tsunamis, and Volcanoes
Bill McGuire OXFORD UNIVERSITY PRESS *320 pp. £18.99 (2012)*
Volcanologist Bill McGuire uses the relationship between atmosphere and geosphere as his springboard for a wide discussion of how climate change could affect what happens on and below Earth's surface. Arguing that sea-level rise, melting ice and other factors could trip already unstable geological systems such as active fault lines, he trawls deep history and new research to examine the evidence. He makes the case for a subterranean dimension to the unfolding drama of climate change.

Bird Sense: What it's Like to Be a Bird
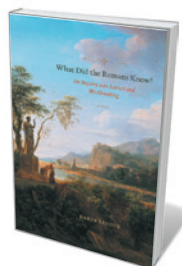*Tim Birkhead* BLOOMSBURY *266 pp. £16.99 (2012)*
Anyone who has watched a soaring gull must have wondered how it feels to be up there, alone and aloft. Animal-behaviour expert Tim Birkhead seeks to tell us, one sense at a time. Even familiar capabilities have alien elements in birds — many species can see ultraviolet light, for example. Sight also has a crucial role in birds' ability to navigate using Earth's magnetic field: a robin with a blurry contact lens on its right eye, for example, loses its sense of direction. Finally, Birkhead speculates about birds' emotions. Is a goose that seems to stand vigil over its dead partner truly grieving?

How Economics Shapes Science
Paula Stephan HARVARD UNIVERSITY PRESS *367 pp. $45 (2012)*
A big biomedical lab spends 18 cents a day to keep one lab mouse, amounting to hundreds of thousands of dollars for animals each year. Economist Paula Stephan takes an exhaustive look at how publicly funded science pays such bills, and how this affects research, researchers and the economy. She argues that expanding universities and stagnant budgets have made funders and scientists more risk-averse, and stunted the development of young investigators. She recommends decoupling research and training to reduce the over-production of PhDs, and forcing universities to bear more salary costs.

What did the Romans Know? An Inquiry into Science and Worldmaking
Daryn Lehoux UNIVERSITY OF CHICAGO PRESS *288 pp. $45 (2012)*
If you rub a magnet with garlic, wrote the Roman philosopher Plutarch, it loses its power to attract. The tale inspired classicist Daryn Lehoux to investigate how these educated people came to believe silly things, and why we now realize they're risible. He defends Roman knowledge, arguing that figures such as Galen, Ptolemy and Cicero forged a distinctive investigative approach shaped by their religious, cultural and political environment.

HISTORY

Impractical magic

A biography of alchemist John Dee sidesteps his impact on science, suggests **Philip Ball**.

The late-sixteenth-century mathematician and alchemist John Dee still exerts a powerful grip on the public imagination. Several novels have centred on him, including Peter Ackroyd's 1993 book *The House of Doctor Dee*. Damon Albarn of British band Blur debuted the pop opera *Dr Dee* in 2011. Now, in *The Arch Conjuror of England*, historian Glyn Parry gives us probably the most meticulous account so far of Dee and his career.

In some ways, all this attention seems disproportionate. Dee was less important in the philosophy of natural magic than the now relatively obscure Giambattista Della Porta and Cornelius Agrippa, and less significant as a transitional figure between magic and science than Della Porta and his contemporaries Bernardino Telesio and Tommaso Campanella, both anti-Aristotelian empiricists from Calabria in Italy. Dee's works, such as *Monas Hieroglyphica*, in which the unity of the cosmos was represented in a mystical symbol, were widely deemed impenetrable even in his own day.

Yet Dee was prominent during the Elizabethan age. He was probably the model for both William Shakespeare's Prospero in *The Tempest* and Ben Jonson's charlatan Subtle in the satire *The Alchemist*. Dee's glamour stems mostly, however, from the same source as that of Walter Raleigh and Francis Drake: they all fell within the orbit of Queen Elizabeth I herself. Benjamin Woolley's 2001 biography of Dee draws explicitly on this connection, calling him 'the queen's conjuror'. And he was precisely that, on and off.

There is no way to make sense of Dee without embedding him within the magical cult of Elizabeth, which also holds the keys to Edmund Spenser's epic poem *The Faerie Queen* and to the flights of fancy in *A Midsummer Night's Dream*. To the English, Elizabeth's reign heralded a mystical Protestant awakening. In Germany, that dream would die in the brutal Thirty Years' War; in England, it would spawn an empire. Dee coined the phrase 'the British Empire', but he looked less towards a colonial future than back to an imagined, magical realm of King Arthur.

As well as being versed in the 'occult arts' of alchemy and astrology, Dee was an able mathematician and an authority on navigation, cartography, cryptography and calendar reform. As Parry illustrates, there were no boundaries between these practical, intellectual and mystical disciplines in Elizabethan culture. One of the book's strengths is its portrayal of how magic and the occult sciences were deeply woven into the fabric of that age.

Dee's relationship with the slippery Edward Kelley also feeds the popular fascination. Kelley claimed to be able to converse with angels through a crystal ball, and Dee's faith in his prophecies and angelic commands never wavered, even when the increasingly deranged Kelley told him that the angels had commanded them to swap wives. During their ill-fated excursion to Poland and Prague in 1583, when they sought the patronage of Holy Roman Emperor Rudolf II, the servant–master relationship became inverted. Dee was reduced to a pathetic figure by the end of the trip.

He had left England after damaging his standing in Elizabeth's court, partly by throwing in his lot with a dubious visiting noble from Poland. He ruined his chances of receiving Rudolf's favour too, by passing on Kelley's angelic reprimand to the emperor for his errant ways.

The Arch Conjuror of England: John Dee
GLYN PARRY
Yale University Press: 2012. 384 pp. £25, $55

John Dee and Edward Kelley 'raising the dead'.

Dee was always making such misjudgements: he was hopeless at court politics. But he can't be held entirely to blame. As Parry highlights, negotiating the convoluted currents of the courts was fiendishly difficult, especially in England, where the memory of the bloody reign of Queen Mary I still hung in the air, along with a fear of papist plots. Unfortunately, Parry's presentation of these political intrigues often become as baffling as they must have been for Dee.

What I really missed was context: an indication of why Dee's magical enthusiasms were emblematic of the times and still felt in the 'scientific revolution' that followed. It is hard to locate Dee in history without hearing about contemporary figures who also sought to expand natural philosophy, such as Della Porta and Francis Bacon. Bacon, in particular, was another intellectual whose grand schemes and attempts to gain the queen's ear were hampered by court rivalries.

We need more than a cradle-to-grave story to understand Dee's significance. For example, although Parry explains the numerological and symbolic mysticism of his *Monas Hieroglyphica*, its preoccupation with divine and Adamic languages seems merely quirky if we are not told that this was a persistent concern, pursued later by the likes of the German Jesuit Athanasius Kircher (the most Dee-like figure of the early Enlightenment) and John Wilkins, a founder of the Royal Society.

Likewise, it would have been easier to evaluate Dee's mathematics if we knew that, until the mid-seventeenth century, maths was closely associated with both witchcraft and mechanical ingenuity, at which Dee excelled. Wilkins can provide orientation here too: he delighted in automata and devices, and describes them in his 1648 account *Mathematical Magick*, a direct descendant of Dee's famed 'Mathematical Preface' to a new translation of Euclid's *Elements*.

We would never know from *The Arch Conjuror of England* that Dee influenced the early modern scientific world through such transitional scholars as Robert Fludd, Elias Ashmole and Margaret Cavendish — nor that his works were studied by Robert Boyle, and probably by Isaac Newton. Parry has assembled an important contribution to our understanding of how magic became science. It is a shame that he didn't see it as part of his task to make that connection. ∎

Philip Ball *is a writer based in London.*

BRIDGEMAN ART LIBRARY

Fossils from the past 150 million years are on show at *Past Worlds*, including this *Tyrannosaurus* skull.

PALAEONTOLOGY

Beyond the Jurassic

Brian Switek winds his way through prehistory at Utah's rehoused museum of natural history.

Utah is spectacularly gifted with dinosaur remains. Most of its museums, however, focus on finds from the Upper Jurassic period 150 million years ago. Now, an exhibition at the Natural History Museum of Utah that includes fossil treasures from subsequent periods is set to broaden our understanding of prehistory.

Past Worlds is housed alongside several other installations in a brand-new, copper-coated sustainable building set against the hills of Salt Lake City. A cohort of Pleistocene mammals — a mammoth, giant ground sloth and sabre-toothed cat among them — stand guard at the upper entrance, representing what life was like around 13,000 years ago, when Lake Bonneville filled the valley now cradling the city.

The exhibition features a circuitous route through some wonderful illustrations — skeletal diagrams and immense, intricately rendered prehistoric scenes — by artists Douglas Henderson, Scott Hartman and Victor Leshyk. Visitors are taken through Utah's prehistory, from youngest (the last ice age, which ended almost 12,000 years ago) to oldest. From the winding walkway, you can view Pleistocene carnivores,

Past Worlds
Natural History Museum of Utah, Salt Lake City. Permanent exhibition.

an Eocene lakeshore and Cretaceous dinosaurs from more than 65 million years ago, before finding yourself among *Marshosaurus*, *Stegosaurus* and other Jurassic giants. The walkway then doubles back — a feature I found odd because it disrupts the chronology, particularly given that the eras are represented on different levels of the building.

The only linear story in the exhibit hangs overhead. Starting with modern pelicans near the upper entrance, a simplified rendition of avian evolution is revealed by birds that are representative of each time period, culminating with a suspended *Archaeopteryx* reconstruction (a creature discovered in Germany, not Utah, but which serves here as a Jurassic stand-in).

Despite the needless convolutions through time, it is a fabulous spectacle. An Eocene lake boasts sleek crocodilians and the rotting body of an archaic horse, while a gaggle of skeletal waterbirds assembles along the upper shore.

⟳ NATURE.COM
For more sauropods on show:
go.nature.com/xjawmi

The scene evokes the habitat in which each of the fish-bearing fossil slabs displayed in the lower gallery was formed about 50 million years ago.

But the true draws of the fossil hall are the dinosaurs. Real fossils are displayed alongside many of the skeletal casts, tying them to authentic finds in the field.

The Cretaceous exhibit is well stocked with dinosaurs exhumed from the state's Grand Staircase-Escalante National Monument. Other fauna from the era include the enormous alligatoroid *Deinosuchus*, and a juvenile–adult pairing of the recently described tyrannosaur *Teratophoneus* that prowls the upper deck. In the lower gallery, a skeleton of the hadrosaur *Gryposaurus* — a herbivore with spoon-like jaws and a high ridge running along the snout — strikes an alert pose.

Horned dinosaur skulls are ranged against the back wall in evolutionary ranks, some of which represent new and as-yet-undescribed genera. The most intriguing specimen is the 'hadrosaur under the floor' — an actual *Gryposaurus* fossil preserved with skin impressions and laid out beneath transparent floor tiles.

The same trick is used in the Jurassic display that follows. Here, bone casts are scattered beneath transparent panes in a reconstruction of the Cleveland-Lloyd dinosaur quarry. At least 46 *Allosaurus* individuals died at this site near Price, Utah, together with rarer carnivorous and herbivorous dinosaurs. Exactly what killed them is disputed. The museum asks visitors to vote by dropping pennies into cylinders marked with the various hypotheses. ('Poisoning' was the uncontested leader when I last checked.)

The Jurassic section also contains the best thrill of the entire hall: a skeletal mount of a hapless *Barosaurus* being harried by a comparatively tiny *Allosaurus*. The sauropod's ludicrous neck arcs high into the air as a young *Allosaurus* pounces on to its back and a family of the carnivores surrounds it. Whether such a scene ever took place is a matter of speculation; the vignette is an extrapolation from the many bones found together in the quarry.

The visual splendour is not matched by the signposting, however. Explanations on the plaques accompanying the displays give details of the daily lives of these prehistoric animals without explaining the pathways to that knowledge. That said, *Past Worlds* brings Utah's awe-inspiring prehistory to vibrant life, all the way from the animals' skeletons through to their habitats. ∎

Brian Switek *is a freelance writer based in Salt Lake City, Utah. He is author of* Written in Stone.
e-mail: evogeek@gmail.com

Correspondence

Whaling: quota trading won't work

Anti-whaling organizations are often presented as conservationists (*Nature* **481,** 114; 2012). But for conservation efforts to advance, we need to resolve the differences between animal welfare, which is concerned with individuals, and environmental conservation, which focuses on maintaining populations, species and ecosystems.

Anti-whaling organizations spend millions of dollars every year trying to stop the Japanese whaling fleet from hunting the common minke whale (*Balaenoptera acutorostrata*), which is not endangered (*Nature* **481,** 139–140; 2012). Their use of financial resources is justifiable only from an animal-welfare perspective.

If the anti-whaling lobby were interested in whale conservation, it would use its financial power to help to assess the population ecology and dynamics of the many whale species listed as 'data deficient' by the International Union for Conservation of Nature. This would enable evidence-based quotas to be set for countries that choose to exploit this resource.

The quota-trading scheme proposed by Christopher Costello and his colleagues is a promising market-based solution for whale conservation, but is unlikely to succeed. For some countries, such as Japan, whaling is a symbol of national and cultural identity, so the economic returns may not provide sufficient incentive. Also, this is strictly a moral issue for the anti-whaling lobby, driven not by environmental conservation but by the suffering imposed on individual whales.

Over the past decade, the two sides have grown further apart. If a compromise is to be reached, environmental conservationists must inform decision-makers and public opinion in the same way that the anti-whaling lobby has used its financial muscle to push its agenda over the years.
Diogo Veríssimo, Kristian Metcalfe *Durrell Institute of Conservation and Ecology, University of Kent, Canterbury, UK. dv38@kent.ac.uk*

Scientists cannot compete as lobbyists

Suggestions that scientists should run for political office or campaign to promote their work are counterproductive and ultimately self-defeating (*Nature* **480,** 153; 2011). Science needs a permanent pipeline into policy, not temporary windows cracked open by individual researchers.

Lobbying takes time and money: more than US$3.5 billion was spent in 2010 on lobbying US Congress members. Academic scientists simply cannot compete on that scale.

Scientists must be impartial arbiters of data, not political agents. They need to be able to negotiate with governments, irrespective of their political hue, and to advise politicians in a useful and timely way.

Scientific-liaison offices would give scientists an apolitical route to policy formation. These would have a cross-ministerial mandate to make research results accessible and enable politicians and policy-makers to reach informed decisions.

When politicians ignore science, it is a failure of our system of governance rather than of individual scientists to act as lobbyists for their research.
Brett Favaro *Simon Fraser University, Burnaby, British Columbia, Canada. bfavaro@sfu.ca*

Expand Australia's sustainable fisheries

We do not believe that marine protected areas (MPAs) currently offer effective conservation in Australia. They do not address pollution or climate change (*Nature* **480,** 151–152; 2011), and overfishing there has largely been rectified. MPAs are also inadequate for managing the major threat of introduced organisms, of which more than 400 have already been identified in Australian waters.

Terry Hughes' call to protect coral reefs from catch-and-release fishing (*Nature* **480,** 14–15; 2011), by closing a further 480,000 square kilometres of ocean in Australia's Coral Sea in addition to the adjacent 507,000 km^2 already proposed, is an example of exaggerated restriction of fishing. We contend that sustainable fisheries need to be expanded, not restricted.

Australia has well-managed fisheries but imports more than 70% of its seafood. By continuing to import while closing more of its exclusive economic zones to fishing, Australia is diverting pressure on seafood resources and the responsibility for their sustainable exploitation to other countries, most of which do not have Australia's effective governance of fishing.
Robert Kearney *University of Canberra, Australia. bob.kearney@canberra.edu.au*
Graham Farebrother *Institute for Marine and Antarctic Studies, University of Tasmania, Australia.*

Use snail ecology to assess dam impact

It is not yet clear whether dam construction in the Mekong Basin will increase the impact of schistosomiasis in the region (A. R. Blaazer *Nature* **479,** 478; 2011). We need a better understanding of the parasite's transmission ecology to improve disease prediction and to determine the best dam locations.

Comparisons with dams in other countries can be misleading. In Africa, schistosome parasites are transmitted by snails with different habitat requirements from *Neotricula aperta*, a snail that is found only in calcium-rich waters in the Mekong Basin and the sole intermediate host of *Schistosoma mekongi*.

In fact, densities of *N. aperta* have declined to undetectable levels downstream of the Nam Theun 2 dam in Laos (S. W. Attwood *et al. Ann. Trop. Med. Parasitol.* **98,** 221–230; 2004) — possibly as a result of flooding, decreased calcium levels and silting. Densities are also falling farther downstream in Thailand, even though habitats there are apparently unaffected (my unpublished observations).
Stephen W. Attwood *Sichuan University, Chengdu, China. swahuaxi@yahoo.com*

Asian medicine: a way to compare data

To help to integrate traditional Asian medicine with Western medicine (S. Cameron *et al. Nature* **482,** 35; 2012), the World Health Organization (WHO) is developing common systems for collecting statistics from both. This information — known as the International Classification of Traditional Medicine (see go.nature.com/mv3iux) — is being incorporated into a revision of the WHO International Classification of Diseases, to be released in 2015.

Clean, standardized data from several countries will allow proper comparison of the effectiveness, cost and safety of the different approaches.
Kenji Watanabe, Xiorui Zhang, Seung-Hoon Choi *WHO ICTM Project Team, Center for Kampo Medicine, Tokyo, Japan. watanabekenji@a6.keio.jp*

NEWS & VIEWS

The fridge gate

Logic gates are the elementary building blocks of computers. The finding that a single logic gate may drive a refrigerator is a beautiful demonstration that information–processing devices can have useful thermodynamic properties.

RENATO RENNER

Minimalism is a popular trend in design, striving to expose the essence of an object through the elimination of all non-essential parts. Writing in the *Journal of Physics A*, Skrzypczyk et al.[1] have now applied this approach to the study of thermal machines such as heat engines and refrigerators. Reducing the complexity of a refrigerator to its extreme, they arrived at a device as simple as a single logic gate. What's more, this minimalist fridge works at optimal efficiency.

To understand the conceptual significance of this result, it is worth taking a sideways glance at the theory of computing. If you were asked to multiply two large numbers on paper, you would probably apply a method (commonly taught in secondary school) to split this task into a number of small steps, each involving the addition or multiplication of only single digits. This is also the way computers work: a computation is divided into a sequence of elementary operations, each of which involves only binary digits — the bits. Logic gates are the devices (usually implemented by electronic components such as transistors) that carry out these elementary operations. What's nice about them is their simplicity: they operate on the bits, which can take only two values (0 or 1). For example, the Toffoli gate[2] acts on three bits: it checks whether the first two are both equal to 1 and, if this is the case, alters the value of the third.

At first sight, refrigerators have little in common with such information-processing gates. But as early as the 1960s, information processing was studied from a thermodynamic perspective. It was found[3] that, for instance, initializing the memory of a computer (that is, setting it into a known state) is basically a process that cools the memory, while consuming energy and dissipating heat into the environment. More generally, it turns out that knowledge can always be traded for coldness[4–6]. It is therefore no surprise that information-processing devices, such as logic gates, can have useful thermodynamic properties[5,7]. The minimalist fridge proposed by Skrzypczyk and colleagues[1] is a beautiful manifestation of this. Although it operates like a simple
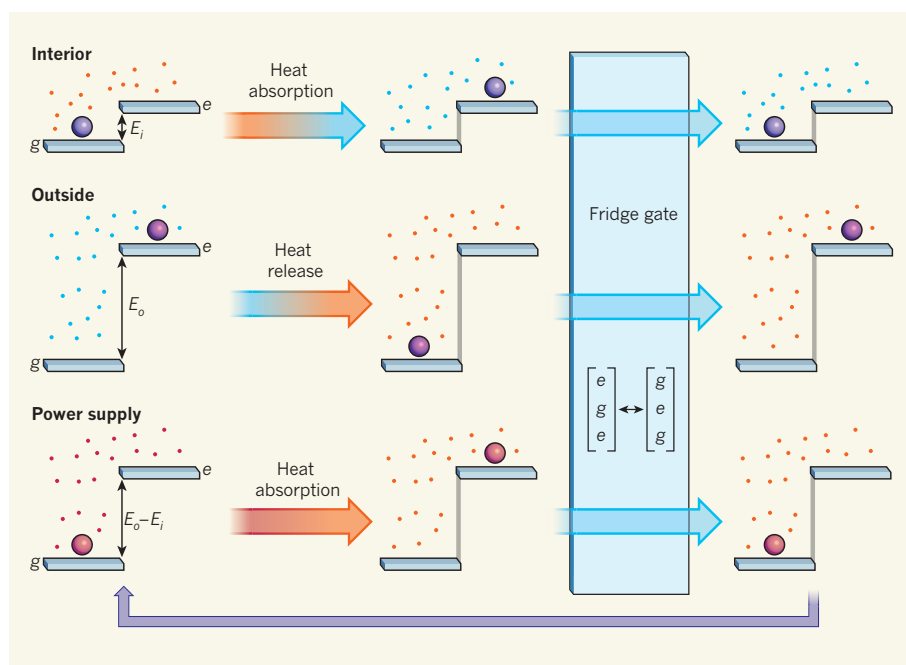

Figure 1 | The three-bit fridge. Refrigerators cool by moving heat from their interior to the outside, while drawing energy from a power supply. The minimalist fridge proposed by Skrzypczyk et al.[1], a discretized version of which is shown, involves only three particles (spheres). Each of them can be in two possible states, the ground state (*g*) and the excited state (*e*). The first particle is connected to the interior of the refrigerator. In state *e*, it carries a little more energy than in state *g*, but the energy difference (E_i) is small enough that the transition from *g* to *e* occurs with high probability. To conserve the total energy, this transition must absorb heat from the ambient molecules (dots), thereby cooling them. Conversely, the energy difference E_o between the levels of the second particle, placed at the outside of the refrigerator, is large in order to make transitions from *e* to *g* very likely. The third particle is connected to a power supply, which heats it so that it often reaches state *e*. The fridge gate induces an energy-neutral swap between the particles' states as indicated. Its net effect on the first particle is to reset it to state *g*, so that the cooling cycle can be restarted (purple arrow).

three-bit logic gate, it has the functionality of a fully fledged refrigerator.

The authors' construction exploits the thermodynamic properties of two-level systems — the physicists' counterparts of the computer scientists' bits. Two-level systems have two possible states with different energies, called the ground state and the excited state. As an example, imagine a tiny particle that lies either on the floor (corresponding to its ground state) or on a slightly elevated stage (corresponding to its excited state). The particle may be hit by ambient molecules, which sometimes catapult it from the floor to the elevated stage. When this happens, the

particle gains potential energy, which — as required by conservation of energy — must be taken away from the ambient molecules, thereby cooling them. Conversely, if the particle falls from the elevated stage onto the floor, its potential energy is released, heating the ambient molecules (Fig. 1).

In thermal equilibrium, the transition from the ground state to the excited state and the reverse process occur equally often, maintaining a balance between the heating and the cooling effect. However, imagine now a mechanism that, whenever the particle reaches its excited state, replaces the particle with a fresh one in the ground state. This would lead to transitions

only from the ground state to the excited state, thus resulting in a net cooling effect. Yet, this cooling mechanism does not run for free. It requires a constant supply of fresh particles in the ground state.

This is where the fridge gate comes into play. It acts on three different two-level particles, one connected to the interior of the refrigerator (which we want to cool), one to the outside, and a third to a power supply (Fig. 1). The gate merely performs an (energy-neutral) permutation of the particles' states. But the net effect on the particle in the interior of the refrigerator is the same as if it were replaced by a fresh one in the ground state — as required for the above-described cooling mechanism to work. Closer inspection reveals that the fridge gate induces a heat transport from this interior particle to the one outside, thereby drawing power from the third particle. Such a heat transfer towards the outside is pretty much what a conventional refrigerator does. But the fridge gate achieves this not only with minimal complexity and size, but also

with minimal energy usage: any cooling mechanism with better energy efficiency than the fridge gate would necessarily violate the laws of thermodynamics.

Can we now build refrigerators driven by a single logic gate? The answer is yes, provided that we are able to realize a fridge gate that works without dissipating any excess heat, which may undermine the cooling effect. Resorting to quantum mechanics, Skrzypczyk and colleagues[1] show that it suffices to let the three two-level particles interact with each other in a constant but specific manner. Although designing this specific interaction is certainly a non-trivial task, the latest developments in quantum engineering give rise to optimism: several research groups have already presented reversible implementations[8–11] of the Toffoli gate described earlier, which is similar in complexity to the fridge gate. These experimental efforts have so far been motivated largely by the prospect of quantum computing. But insights into thermodynamics[1,4–7] obtained from information theory uncover

novel applications and challenges for quantum engineering — implementing the fridge gate being just one. ∎

Renato Renner *is at the Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland.*
e-mail: renner@phys.ethz.ch

1. Skrzypczyk, P., Brunner, N., Linden, N. & Popescu, S. *J. Phys. A* **44**, 492002 (2011).
2. Toffoli, T. in *Automata, Languages and Programming, 7th Colloq.* (eds de Bakker, J. W. & van Leeuwen, J.) 632–644 (Springer, 1980).
3. Landauer, R. *IBM J. Res. Dev.* **5**, 183–191 (1961).
4. Bennett, C. *Int. J. Theor. Phys.* **21**, 905–940 (1982).
5. Schulman, L. & Vazirani, U. in *Proc. 31st Annu. ACM Symp. Theory of Computing* (eds Vitter, J. S., Larmore, L. & Leighton, T.) 322–329 (ACM Press, 1999).
6. del Rio, L., Aberg, J., Renner, R., Dahlsten, O. & Vedral, V. *Nature* **474**, 61–63 (2011).
7. Boykin, P. O., Mor, T., Roychowdhury, V., Vatan, F. & Vrijen, R. *Proc. Natl Acad. Sci. USA* **99**, 3388–3393 (2002).
8. Cory, D. G. *et al. Phys. Rev. Lett.* **81**, 2152–2155 (1998).
9. Lanyon, B. P. *et al. Nature Phys.* **5**, 134–140 (2009).
10. Monz, T. *et al. Phys. Rev. Lett.* **102**, 040501 (2009).
11. Fedorov, A., Steffen, L., Baur, M., da Silva M. P. & Wallraff, A. *Nature* **481**, 170–172 (2012).

GENETICS

How intelligence changes with age

An analysis of common genetic variants shows that hereditary factors that influence intelligence in childhood also affect it in old age. Such work could signal the end of the nature–nurture controversy. SEE LETTER P.212

ROBERT PLOMIN

Francis Galton — Charles Darwin's half-cousin — argued 150 years ago that "there is no escape from the conclusion that nature prevails enormously over nurture". Since then, the nature–nurture (or genetics–environment) controversy has never been more contentious than when it concerns human intelligence. A report by Deary and colleagues[1] on page 212 of this issue, however, may mark the beginning of the end of this debate. Instead of estimating genetic influence on intelligence indirectly by using special groups such as twins and adoptees, the authors use DNA data from unrelated people.

A traditional approach to estimate the heritability of a trait, or phenotype, has been to compare groups of known genetic relatedness, such as identical twins (100% relatedness) and fraternal twins (roughly 50%). The strength of this approach is that it estimates the net effect of genetic influence without the need to know which genes are responsible. However, this absence of DNA sequence information is also its weakness.

So, following the sequencing of the human genome, researchers had great expectations from genome-wide association studies (GWAS). It was hoped that GWAS would identify enough associations between DNA sequence variants (typically, single nucleotide polymorphisms, or SNPs) and complex traits such as intelligence to account for most of the heritability of the traits. But such analyses, sometimes involving hundreds of thousands of individuals, have detected only a small portion of genetic influence, even for highly heritable traits such as height and weight. For instance, initial GWAS of intelligence[2,3] have indicated contributions of many small genetic effects. This is because the genomic differences identified so far between individuals make only a small total contribution to the heritability of this trait — an issue that has been dubbed the missing heritability problem[4].

Deary *et al.*[1] use a variation of genome-wide complex-trait analysis (GCTA), a method that complements GWAS. In GCTA, researchers use DNA data for hundreds of thousands of SNPs from unrelated individuals to estimate genetic influence on a particular trait.

Unlike the hypothesis-testing approach of GWAS, GCTA does not specify which DNA variants are associated with a measured trait. Instead, it is a parameter-estimation approach that relates similarity in SNPs to phenotypic similarity between pairs of individuals. The use of a large sample, together with pair-by-pair comparisons, allows amplification of the weak signal derived from the low genetic similarity between unrelated subjects. Heritability is estimated as the extent to which genetic similarity can account for phenotypic similarity.

GCTA has been applied to estimate heritability for traits such as height[5] and weight[6], psychiatric and other medical disorders[7], and intelligence[3]. To estimate how genetic factors influence the stability of intelligence and how it changes with age, Deary *et al.*[1] applied this approach to SNP data and intelligence-test scores from almost 2,000 unrelated people from Scotland. What is especially exciting about this report is that, in contrast to previous GCTA studies[3–7], the authors extend their analysis to the multivariate case and obtain a noteworthy result. Essentially, the multivariate extension of GCTA evaluates relatedness between each pair of individuals for different traits. In Deary and colleagues' report, the different traits are intelligence assessed at two stages of life in the same people: in childhood (at age 11) and, half a century later, in old age. Specifically, they estimate genetic change and continuity as the extent to which similarity in SNPs between two individuals can account for similarity in change and continuity in their intelligence.

The authors find a

substantial genetic correlation (0.62) between intelligence in childhood and in old age, which means that many of the same genetic elements are associated with this trait throughout life. The analysis also estimates the genetic influence on cognitive change across life: nearly a quarter of the variation in the change in cognitive scores that occurs throughout life could be explained by different genes being associated with this trait in childhood and later life. These findings are consistent with previous results[8] from family-based genetic research, although no family-based studies have extended from childhood to old age.

Deary *et al.* show appropriate caution about their estimates. Their results are valuable because such data are rare, but the results come with large standard errors (a measure of how the data values spread around the mean) and are not statistically significant by conventional measures. This is because, in GCTA, a tiny signal is extracted from a lot of noise, and so samples in the tens of thousands — much larger than the one used in this report — are needed for accurate estimates.

Nonetheless, GCTA will stimulate research on the genetics of intelligence, because this method does not require special samples such as groups of twins or adoptees. Indeed, multivariate GCTA could be used to test findings from family-based research on intelligence. These findings include that the same genetic factors affect different cognitive abilities and disabilities, and that genetic propensities for intelligence correlate and interact with cognitively relevant experiences[9].

The prerequisites for GCTA — very large samples in which huge numbers of SNPs have been analysed — seem daunting. But these are the same resources required for GWAS, and many such samples are already available for several traits, including intelligence. Another caveat of GCTA, however, is that it underestimates heritability because it is limited to SNPs that have been mapped on the genome and to DNA variants correlated with those SNPs (that is, variants that are in linkage disequilibrium with them). By contrast, traditional family-based genetic designs capture variation due to all causal variants in the genome.

Regardless of such caveats, GCTA may provide crucial clues for solving the missing heritability problem. It has been suggested[10] that, to find genes associated with complex traits such as intelligence, researchers need to analyse rare genetic variants in addition to the common ones that are detected by available microarray tools. However, to the extent that GCTA estimates of heritability account for heritability estimates derived from family-based studies, this suggests that common SNPs can powerfully predict intelligence if sample sizes are sufficiently large. If true, this means that intelligence is similar to height in terms of genetic architecture and that — with

similar sample sizes to those used for research on heritability of height — many associations between DNA and intelligence will be found.

So, although GCTA cannot quite mark the end of the nature–nurture controversy, it might be the beginning of the end. Similar to family-based genetic methods, this approach is limited to estimating genetic influence indirectly from genetic similarity between pairs of individuals, rather than directly from specific genes, which is the ultimate goal. But it is much more difficult to dispute GCTA results based on DNA data than it is to quibble about twin and adoptee studies. ∎

Robert Plomin *is at King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London SE5 8AF, UK.*
e-mail: robert.plomin@kcl.ac.uk

1. Deary, I. J. *et al. Nature* **482,** 212–215 (2012).
2. Davis, O. S. P. *et al. Behav. Genet.* **40,** 759–767 (2010).
3. Davies, G. *et al. Mol. Psychiatry* **16,** 996–1005 (2011).
4. Maher, B. *Nature* **456,** 18–21 (2008).
5. Yang, J. *et al. Nature Genet.* **42,** 565–569 (2010).
6. Yang, J. *et al. Nature Genet.* **43,** 519–525 (2011).
7. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. *Am. J. Hum. Genet.* **88,** 294–305 (2011).
8. Lyons, M. J. *et al. Psychol. Sci.* **20,** 1146–1152 (2009).
9. Plomin, R., DeFries, J. C., McClearn, G. E. & McGuffin, P. *Behavioral Genetics* 5th edn (Worth, 2008).
10. Cirulli, E. T. & Goldstein, D. B. *Nature Rev. Genet.* **11,** 415–425 (2010).

An Earth–sized duo

The first Earth–sized planets orbiting a Sun–like star outside the Solar System have at last been detected. The discovery paves the way to finding Earth–like worlds. SEE LETTER P.195

DIDIER QUELOZ

Less than two decades after the first discoveries[1,2] of extrasolar planets, we know that a significant number of stars in our Galaxy host planetary companions. Most of the planets detected have been of the giant kind: they are similar in mass and size to Jupiter. But with advances in the radial-velocity planet-hunting technique[3] and the successful launch of the Kepler space telescope, which uses the transit detection method, planetary systems composed of small and low-mass planets are being detected. These discoveries suggest that such planetary systems are common in our Galaxy[4]. On page 195 of this

issue, Fressin *et al.*[5] describe the detection of two Earth-sized planets in the planetary system Kepler-20. The discovery of planets of such a small size shifts the *terra incognita* of the planetary landscape to objects smaller than Earth.

Kepler-20 is a Sun-like star located about 300 parsecs (roughly 1,000 light years) from Earth. Observations by the Kepler telescope have revealed[6] three sub-Neptune-sized planets (Kepler-20 b, Kepler-20 c and Kepler-20 d) and two Earth-sized candidate planets (Kepler-20 e and Kepler-20 f) that transit (pass in front of) this star, their orbital periods being respectively 3.7, 10.9, 77.6, 6.1 and 19.6 days. The compactness of this system — the planets
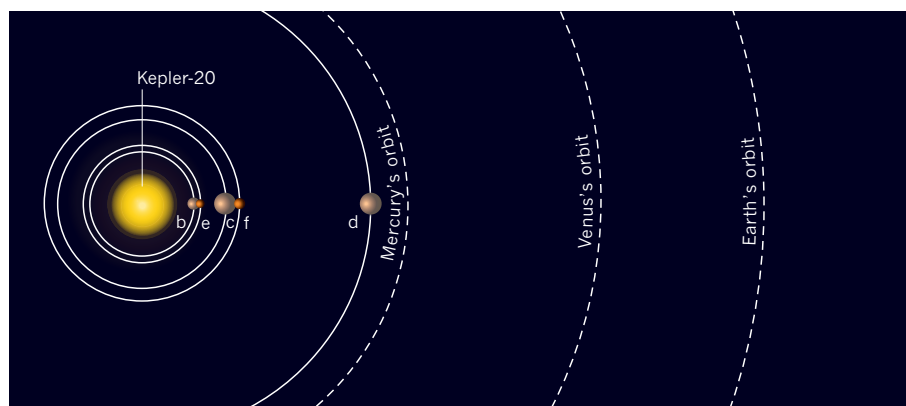

Figure 1 | Orbital configuration of Kepler-20. The Earth-sized planets detected by Fressin *et al.*[5], Kepler-20 e and Kepler-20 f, are part of a compact system of five planets (b–f). The orbits of these planets around their host star Kepler-20 are closer than those of the inner planets around the Sun. Planet sizes are not shown to scale with distance. (Modified from a graphic by K. Tate, Space.com.)

orbit their host star at distances closer than Mercury's orbit around the Sun (Fig. 1) — is not unusual for an extrasolar planetary system.

Many similar dense configurations have been detected by the radial-velocity (Doppler) technique since the discovery[7] in 2006 of three Neptune-mass planets on compact orbits around star HD 69830. The Doppler technique measures tiny Doppler shifts in a parent star's light that are caused by the gravitational tug of an orbiting planet. The most dense multi-planet configurations have been observed in a system of seven planets[8] orbiting star HD 10180 and a system of six planets[9] transiting star Kepler-11.

In some dense planetary systems, planets can affect each other through dynamical effects related to their orbits. This phenomenon perturbs the regularity of the orbits, generating small time delays. The delays can be detected with planetary transits and are known as transit-timing variations.

From a combination of the dynamical effect of the planets on their host stars (obtained through Doppler and transit-timing measurements) and transit observations — which measure the dimming of the star as a planet transits it — a planet's mean density can be calculated. This calculation provides insight into the object's overall structure, for example whether it is a gas giant or a small, rocky planet. However, there are few planets below the sub-Saturn mass range for which the density is known. The stars explored by space-based planetary-transit missions such as Kepler are too faint for accurate ground-based Doppler follow-up observations of the smallest planetary candidates, and detection of the timing of the transits is restricted to specific planetary-system configurations[9].

Ideally, the detection of a planet's gravitational tug on its star is required to confirm that a transiting candidate is a planet. Practically, however, for most candidates detected by Kepler, measuring this dynamical effect is challenging. Other possible candidates, including eclipsing two-star systems, can lead to a transiting signal similar to that of a planet, and can be eliminated only through a combination of complementary measurements and statistical analysis. This is the approach that Fressin and colleagues[5] take in their study. Their identification of two Earth-sized planets — Kepler-20 e and Kepler-20 f — relied on a statistical analysis of previous Kepler measurements[6] to establish that the transiting signals are indeed of planetary origin.

To consider the two new planets in the wider planetary landscape, Fressin et al. produced a mass–radius diagram of all known 'super-Earth' planets (see Fig. 3 of the paper[5]). A super-Earth is a planet that has a mass between those of Earth and Neptune, irrespective of its

internal structure. The diagram is a striking illustration of the potential diversity of planets in this mass domain: objects of the same mass can be a gas giant or a dense, iron-core planet. This result will prompt researchers to explore the origin of such diversity in the context of planet-formation models.

Although the masses of Kepler-20 e and Kepler-20 f are unknown, the authors show[5] that the two planets are without doubt located in the low-end corner of the mass–radius diagram, where Earth-like planets lie. But because the planets' mass is unknown, their composition cannot be determined unambiguously. Interestingly, however, some compositional knowledge exists for Kepler-20 b and Kepler-20 c, from the detection and upper limit of the Doppler signal[6] originating from these two more massive planets. This information already suggests a possible broad range of composition for the two planets: from magnesium silicate to water ice; or they may even be gas giants[6].

The existence of a series of small planets such as Kepler-20 e and Kepler-20 f identify them as key objects in the steadily expanding list of planetary systems. This is because, in contrast to the Solar System, where small, rocky planets lie close to the Sun but gas giants are found far from it, these planets have no obvious hierarchical orbital location. The next, pivotal, step in extrasolar planetary research will be to detect the dynamical effect of each of these small planets on their host star and to determine their mass. ∎

Didier Queloz *is at the Observatoire de Genève, 1290 Sauverny, Switzerland.*
e-mail: didier.queloz@unige.ch

1. Wolszczan, A. & Frail, D. A. *Nature* **355,** 145–147 (1992).
2. Mayor, M. & Queloz, D. *Nature* **378,** 355–359 (1995).
3. Udry, S., Fischer, D. & Queloz, D. in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 685–699 (Univ. Arizona Press, 2007).
4. Pepe, F. *et al. Astron. Astrophys.* **534,** A58 (2011).
5. Fressin, F. *et al. Nature* **482,** 195–198 (2012).
6. Gautier, T. N. III *et al. Astrophys. J.* (in the press); preprint at http://arxiv.org/abs/1112.4514v2 (2011).
7. Lovis, C. *et al. Nature* **441,** 305–309 (2006).
8. Lovis, C. *et al. Astron. Astrophys.* **528,** A112 (2011).
9. Lissauer, J. J. *et al. Nature* **470,** 53–58 (2011).

Genomics decodes drug action

Drugs used to treat African sleeping sickness are outdated, and how they enter cells and exert biological effects is poorly understood. A genome–wide study using RNA interference provides valuable insight. SEE LETTER P.232

ALAN H. FAIRLAMB

African trypanosomiasis, or sleeping sickness, is a deadly yet neglected human disease caused by the single-celled parasites *Trypanosoma brucei gambiense* and *T. b. rhodesiense*. The origins of some antitrypanosome drugs, including suramin and melarsoprol, date back to pioneering studies with coloured dyes and organic arsenical compounds at the beginning of the twentieth century. Nonetheless, the modes of action of these and the three other drugs currently used to treat sleeping sickness (pentamidine, nifurtimox and eflornithine) are incompletely understood. On page 232 of this issue, Alsford et al.[1] identify some of the biological pathways used by these drugs, offering insight into how they reach their cellular targets and how drug resistance can arise. The results pave the way for the development of new therapeutic strategies.

The existing antitrypanosome drugs are typically given by injection. Moreover, some of them cannot cross the blood–brain barrier, making them ineffective against late-stage disease, when parasites invade the brain. To facilitate the discovery of drugs that lack these unsatisfactory features, there is a need to identify additional drug targets. With this aim in mind, Alsford et al. conducted a genome-wide RNA interference (RNAi) screen on trypanosomes. When induced artificially, RNAi — which works by silencing messenger RNA transcripts[2] — is a powerful tool for probing the biological function of specific genes. It allows researchers to study the cellular effects of the loss of a specific protein, and aids in determining whether a protein has a structural, regulatory, transport or metabolic function.

An inducible RNAi system has previously been set up[3] in *T. brucei* and has already been used for target-based drug discovery, to assess whether specifically selected genes are essential to trypanosome survival[4]. However, this is a relatively slow approach and can suffer from investigator bias. In addition, the sequencing
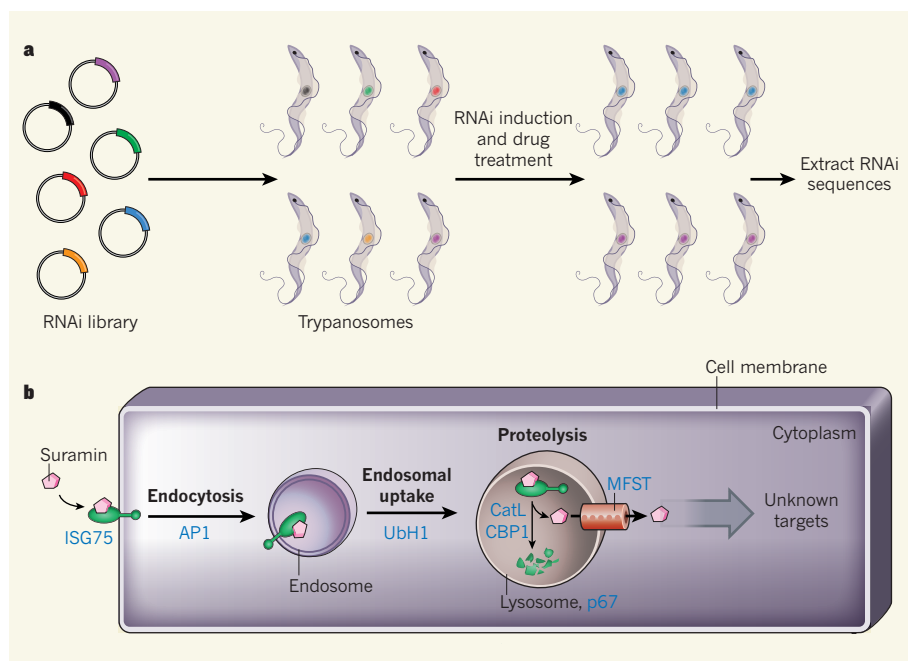
Figure 1 | A genome-wide RNAi screen to identify antitrypanosome drug activity. a, By introducing a genome-spanning panel of RNAi molecules into trypanosomes, Alsford *et al.*[1] identified those RNA sequences that, when silenced, allowed the cells to survive treatment with each of five drugs used to treat trypanosomiasis. The authors mapped these sequences onto the trypanosome genome to reveal proteins involved in drug action. Here, trypanosomes containing RNAi sequences that confer resistance to the drug suramin are shown as blue and purple. **b,** Proteins (labelled in blue) pinpointed by the suramin screen are shown as an example of how this technique can reveal stages of the pathway a drug takes, including the drug's entry to the cell by binding to membrane protein ISG75, its uptake into the endosome by endocytosis (which relies on four protein subunits that form adaptor protein complex 1, AP1) and transfer of the suramin–ISG75 complex to the lysosome and its breakdown by proteolysis there (involving the proteins UbH1, CatL, CBP1 and p67). The free drug is then released into the cytoplasm (with the aid of transport protein MFST), where it affects as-yet-unknown targets.

of the *T. brucei* genome[5] allowed application of a genome-wide RNAi screen[6] to examine key features of trypanosome biology. In this method, random fragments of genomic DNA were expressed as inducible RNAi molecules, and the sequence of fragments that, when expressed, caused trypanosome death was determined. This system linked hundreds of previously uncharacterized proteins, some of which may represent new drug targets, to essential functions at various stages of the parasite's life cycle.

Alsford *et al.* have now used the RNAi approach to ask a different question: which non-essential gene products, when down-regulated, confer a selective advantage on drug-treated trypanosomes? The authors conducted the screens with each of the five existing drugs. In response to drug exposure, trypanosome growth was initially curtailed, but a drug-resistant population subsequently emerged. Using high-throughput sequencing, Alsford and colleagues mapped the sequence of each RNAi molecule extracted from these surviving trypanosomes onto the reference genome (Fig. 1a). In such screens, whenever loss of function of a protein increases drug tolerance, its corresponding RNAi target sequence shows up more frequently on the maps than do

target sequences for non-essential proteins not conferring a selective advantage.

The current study reveals a fascinating pattern of genes involved in diverse areas of metabolism and cell biology. The authors' screens not only support previous findings from decades of painstaking biochemical and genetic approaches[7], but also reveal previously unknown pathways involved in drug uptake, activation and action.

Knockdown of known or potential drug-uptake mechanisms — through decreased expression of proteins involved in drug transport (for eflornithine and melarsoprol) and cellular uptake[8] (for suramin) — is evident in three of the screens. For suramin, the authors identified multiple proteins that increase resistance to the drug, revealing details of the pathway it follows in trypanosomes (Fig. 1b). They postulate that suramin is initially bound to proteins in the blood plasma and subsequently binds to ISG75, a transmembrane glycoprotein of unknown function on the trypanosome surface. ISG75 is then endocytosed (absorbed by engulfing) by the cell and tagged with the small protein ubiquitin. The ubiquitin tag directs the suramin–ISG75 complex to cellular organelles called lysosomes, where it is broken down by protein-degrading enzymes

(CatL and CBP1). This releases suramin, which is then proposed to enter the cell cytoplasm via a transporter protein, MFST, to affect as-yet-unknown intracellular targets.

Alsford *et al.* also discovered an unexpected drug–drug interaction. Knockdown of three enzymes involved in the *de novo* biosynthesis of spermidine, a polyamine compound essential for trypanosome growth, confers resistance to suramin. It also emerged that one of the other drugs, eflornithine, can antagonize suramin's trypanosome-killing activity by inhibiting the synthesis of polyamines.

Of the five drugs studied, nifurtimox differs in that it requires activation by the enzyme nitroreductase to form reactive products, the downstream targets of which remain unknown. Knockdown of nitroreductase (or of its cofactor FMN) leads to resistance to nifurtimox, consistent with previous studies[9]. The nifurtimox RNAi screen also identified reduced synthesis of the electron-transport molecule ubiquinone, suggesting that this is the substrate for nitroreductase in trypanosomes.

Another of the drugs tested, melarsoprol, is an arsenic-based compound that binds to trypanothione, a trypanosome-specific antioxidant metabolite essential to the parasite's survival. The melarsoprol screen shows that resistance to this drug is associated with reduced expression of the transporter protein P2 and with trypanothione biosynthesis, suggesting that the complex formed between trypanothione and melarsoprol is itself toxic.

Cross-resistance — whereby resistance to one drug also confers resistance to another class of drug — occurs between pentamidine and melarsoprol, and other drugs of these respective classes, but, again, the resistance mechanism is unclear. One RNAi effect that Alsford and colleagues hit upon in screens using these drugs identified two closely related aquaglyceroporin proteins. The authors generated trypanosomes lacking these proteins, and found them to be less susceptible to both drugs. This suggests that aquaglyceroporins may be partly responsible for cross-resistance.

Alsford and colleagues' findings[1] should stimulate further research, particularly to determine the functions of other down-regulated genes that encode as-yet-uncharacterized proteins, and other genes and pathways reported in their study. However, not all resistance mechanisms involve loss of protein function — drug-efflux pathways being one example — so these would be missed by RNAi screens. This approach is also unable to identify essential proteins that are drug targets, because targeting these with RNAi leads to cell death.

Therefore, to further unravel the complex mode of action of these drugs, analysis using a genome-wide overexpression system will be required. This method has already been used successfully to demonstrate the effects of

enzyme inhibitors against genetically validated targets such as *N*-myristoyltransferase[10] and trypanothione synthetase[11]. Given the resurgence of interest in screening libraries of chemical compounds for those that inhibit trypanosome growth, to identify novel starting points for drug discovery, a genome-wide strategy would greatly accelerate this expensive and laborious process. ∎

Alan H. Fairlamb *is in the Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK.*
e-mail: a.h.fairlamb@dundee.ac.uk

1. Alsford, S. *et al. Nature* **482,** 232–236 (2012).
2. Meister, G. & Tuschl, T. *Nature* **431,** 343–349 (2004).
3. Wirtz, E., Leal, S., Ochatt, C. & Cross, G. A. M. *Mol. Biochem. Parasitol.* **99,** 89–101 (1999).
4. Kolev, N. G., Tschudi, C. & Ullu, E. *Eukaryot. Cell* **10,** 1156–1163 (2011).
5. Berriman, M. *et al. Science* **309,** 416–422 (2005).
6. Alsford, S. *et al. Genome Res.* **21,** 915–924 (2011).
7. Jacobs, R. T., Nare, B. & Phillips, M. A. *Curr. Top. Med. Chem.* **11,** 1255–1274 (2011).
8. Field, M. C., Lumb, J. H., Adung'a, V. O., Jones, N. G. & Engstler, M. *Int. Rev. Cell Mol. Biol.* **278,** 1–67 (2009).
9. Wilkinson, S. R., Taylor, M. C., Horn, D., Kelly, J. M. & Cheeseman, I. *Proc. Natl Acad. Sci. USA* **105,** 5022–5027 (2008).
10. Frearson, J. A. *et al. Nature* **464,** 728–732 (2010).
11. Wyllie, S. *et al. Mol. Microbiol.* **74,** 529–540 (2009).

QUANTUM OPTICS

Controlling the light

Means to access and manipulate X-rays have been developing at a slow pace. But quantum–optical effects in ensembles of nuclei offer a way to tackle the control of this energetic radiation. SEE LETTER P.199

BERNHARD W. ADAMS

Our world is one of electrons in chemical bonds, and our sensory perception of it is based on quantum energies of a few electronvolts at most. But during the past century we have uncovered the existence of another realm — that of nuclear physics and electrons in inner atomic shells, where quantum energies fall in the X-ray regime and are of the order of kilo- and mega-electronvolts. Although many applications have been found for X-ray and nuclear science, our ability to control this world has been limited. On page 199 of this issue, Röhlsberger *et al.*[1] demonstrate detailed quantum-physical control of the emission of light occurring at nuclear energy scales. Although this result is unlikely to have an immediate application, new capabilities are expected to emerge from a detailed quantum-level control of X-rays. Among these are types of spectroscopy to probe chemical dynamics, or a drastic reduction in the radiation dose required for biological X-ray applications.

Among some other investigations, such as the control of nuclear γ-ray emission by magnetic fields[2], Röhlsberger and colleagues' study[1] can be seen as a step in a progression towards extending the exquisite control of X-rays and γ-rays. This progression re-traces steps taken previously at lower photon energies than those of X-rays — first with radio-frequency waves, which are of sub-microelectronvolt photon energy and can be easily controlled in amplitude and phase (where a wave's peaks and troughs lie) and, more recently, with near-visible-light lasers, which have photon energies

of a few electronvolts. The latter has grown into the field of photonics (smart photons), thanks to progress in precision optics, coherent light sources (those in which light is of well-defined amplitude and phase) and nonlinear optics, which couples light waves instead of allowing them to pass unhindered through each other.

These developments have led to the point at which laser-based high-harmonic generation (HHG) reaches the soft X-ray regime up to about 1 keV. HHG is the nonlinear optical process by which lower-energy photons 'stack up' to generate more energetic ones, and requires meticulous control of the light with respect to coherence and nonlinear optics. Although X-ray and γ-ray photons stand out from thermal or electronic background noise in detectors much more clearly than do visible photons, the development of ways to access and manipulate this energetic radiation has been slow. Now, however, X-ray quantum optics is poised to take off and tackle this radiation regime.

In their study, Röhlsberger *et al.*[1] demonstrate the application of the quantum-optical concepts of superradiance[3] and electromagnetically induced transparency[4] to the control of X-ray scattering. Superradiance is the phenomenon of collective spontaneous emission of radiation. Superradiance, as well as the related effect of subradiance, occurs when an ensemble of atoms or nuclei is prepared in an entangled state (a defining feature of quantum physics) of excitation, and then emits radiation. When the experiment is done such that there is, in principle, no way of telling which atoms or nuclei in the ensemble were excited, it doesn't make sense to consider them

individually. Rather, the whole ensemble emits radiation collectively and may show telltale signs of superradiance, namely directional and accelerated light emission. Here, the authors attained X-ray superradiance from a collection of iron-57 nuclei by exciting them with X-rays at a photon energy of 14.4 keV.

To return to the control of light, from radio waves to X-rays, the classical analogue of superradiant directionality is the directional radio signal obtained from a device known as a phased-array antenna, which is commonly used in radar technology. However, in contrast to the phased-array antenna, superradiance occurs even at the extreme quantum limit of ensemble excitation by a single photon[5], as observed in the authors' experiment. At this limit, we cannot speak of the emission of classical waves from many atoms or nuclei undergoing constructive interference. It is the interference of multiple possible excitation–emission pathways for a single photon that leads to collective emission. Given the right tools, this interference can be controlled.

One of these tools is electromagnetically induced transparency (EIT), in which light absorption due to a transition from one atomic energy level, *g*, to another, *e*, is suppressed as a result of coherence induced by an auxiliary laser. In the typical case, a laser strongly drives a transition from *e* to a metastable state, *f*, and back to *e*; this is technically known as Rabi flopping, and occurs at the Rabi frequency, which is proportional to the square root of the auxiliary-laser intensity. Rabi flopping leads to a splitting of *e*, so that there are now two levels (sidebands) at energies symmetrically above and below that of *e*. For incident photons at the original transition energy from *g* to *e*, the contributions of the two sidebands to the absorption cancel out.

In their experiment, the authors[1] observed superradiance from ^{57}Fe nuclei in an X-ray waveguide because X-ray emission from the nuclei interferes constructively with that from their mirror images in the waveguide walls. The strength of this collective emission of the nuclei, together with their images, depends on the coupling to the guided mode — that is, on the position of the nuclei relative to the standing-wave X-ray pattern that is created in the waveguide. In a configuration in which one layer, A, of nuclei is coupled strongly to a guided mode and another layer, B, is not, they also detected a sharp dip in the ensemble's absorption spectrum characteristic of EIT. No auxiliary laser was used to couple between two energy levels, as in conventional EIT. Instead, excitations of the nuclear ensembles A and B correspond, respectively, to those of states *e* and *f* in conventional EIT, and the coupling of the two is due to the waveguide. In this case, ensemble B, which lacks superradiance owing to its weak waveguide coupling, takes the role of the metastable state *f* because it has a longer excited-state

lifetime than superradiant ensemble A.

Röhlsberger and colleagues' elegant experiment takes its place among the few demonstrations of quantum-optical concepts in the X-ray or γ-ray regime. Others are X-ray parametric downconversion (the spontaneous splitting of photons into two, possibly entangled, photons), and X-ray spectroscopy of atoms in intense laser fields. With the growing significance of X-ray free-electron lasers,

quantum-optical effects will be observed much more often than in the above, synchrotron-based studies. Understanding them will be crucial for a proper interpretation of experiments based on these lasers. Quantum optics can also be used to devise novel experimental techniques for X-ray free-electron lasers — smart photons now also for X-rays. ■

Bernhard W. Adams *is at the Argonne*

National Laboratory, Lemont, Illinois 60439, USA.
e-mail: adams@aps.anl.gov

1. Röhlsberger, R., Wille, H.-C., Schlage, K. & Sahoo, B. *Nature* **482**, 199–203 (2012).
2. Shvyd'ko, Y. V. *et al. Phys Rev. Lett.* **77**, 3232–3235 (1996).
3. Dicke, R. H. *Phys. Rev.* **93**, 99–110 (1954).
4. Harris, S. E. *Phys. Today* **50**, 36–42 (1997).
5. Scully, M. O.,Fry, E. S., Ooi, C. H. R. & Wódkiewicz, K. *Phys. Rev. Lett.* **96**, 010501 (2006).

CELL BIOLOGY

Destruction deconstructed

Correctly dismantling a structure can be as challenging as assembling it. The architecture of the yeast proteasome reveals this enzyme's intricate machinery for protein degradation. SEE ARTICLE P.186

GENG TIAN & DANIEL FINLEY

The ensemble of proteins that makes up a cell is constantly changing. The malleability of a cell's protein content on a short timescale is largely due to the proteasome — a complex member of the protease family of enzymes, which break down proteins. The proteasome, found in eukaryotic organisms (such as plants, animals and yeast), is composed of two main parts: the 20S core particle, which is a 28-subunit protein-destruction complex, and the 19S regulatory particle, a 19-subunit complex that mediates substrate selection. On page 186 of this issue, Lander *et al.*[1] report the complete subunit structure of the yeast 19S regulatory particle. The results include many surprises and some puzzles that will challenge the proteasome field.

Unlike extracellular proteases (Fig. 1a), many of which break down proteins almost indiscriminately, the major intracellular proteases are large and highly selective protein complexes. The eukaryotic proteasome seems to have evolved from a protease known as PAN (or something comparable to this enzyme), which is found in microorganisms called archaea (Fig. 1b). The proteolytic sites of both PAN and the proteasome are hidden in an internal compartment, which a substrate protein can reach by passing through a narrow pore; the pore prevents the entry of properly folded proteins. Movement and unfolding of the substrate require energy, which the enzymes derive from the hydrolysis of ATP molecules[2]. The ATP-hydrolysing components, called ATPases, form a ring directly atop the pore. Any protein that passes through the pore is unlikely to survive the journey.

In simple ATP-dependent proteases such as

PAN, the ATPases select substrates for degradation from the pool of cellular proteins. In eukaryotes, however, chains of a protein called ubiquitin 'tag' those proteins that are destined for degradation[3]. Ubiquitin is then recognized by the regulatory particle of the proteasome, and the substrate is channelled into the core particle. A large family of ubiquitin ligase
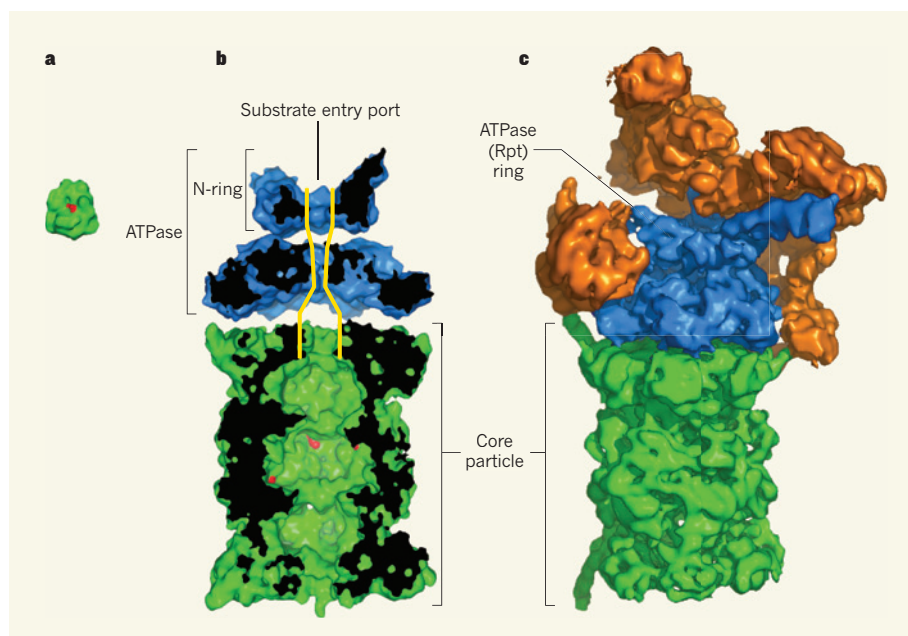
enzymes, which probably numbers more than 500 in humans, can thus condemn specific proteins by attaching ubiquitin groups to them. This dramatic evolutionary elaboration of the protein-degrading machinery is reflected in the fact that the proteasome has assumed regulatory functions in virtually all aspects of eukaryotic cell biology.

The evolution of ubiquitin tagging also coincided with a transformation of the proteasome's structure. In the PAN complex, the regulatory particle is composed solely of the ATPase ring[4,5] (Fig. 1b). By contrast, the eukaryotic proteasome contains 13 additional subunits (Fig. 1c), nine of which make up the proteasome lid, with the other four, in combination with the ATPase (Rpt) ring, forming the base[3].

Lander and colleagues' study[1], together with three other recent papers[6–8], provides a comprehensive picture of the proteasome components that are specific to eukaryotes

Figure 1 | The evolution of proteases. a, Trypsin, an example of a typical extracellular protease, with its proteolytic active site shown in red (model taken from ref. 13). **b,** Central cross-section of a model[4] of the archaeal PAN complex showing the six-subunit ATPase with its N-ring and the 28-subunit core particle, together with its proteolytic active sites (red). To model the structure, the ATPase, N-ring and core particle were manually placed in proximity. The substrate-translocation channel of PAN (yellow) has the entry port at its top. **c,** Surface of the eukaryotic proteasome, as obtained by Lander *et al.*[1], showing the core particle and the ATPase (Rpt) ring. Subunits comprising those parts of the regulatory particle that are specific to the eukaryotic proteasome are shown in orange.

(Figs 1c and 2). Of particular interest are the ubiquitin receptors and deubiquitinating enzymes, as well as the placement of these elements with respect to the substrate-translocation channel of the ATPase ring. At the top of the channel is the N-ring, which is thought to be the entry port (Fig. 2a) for the substrate[4,5]. Deep within the channel are the pore loops of the Rpt proteins, which are proposed to make contact with substrate proteins and use ATP-driven movements to inject them into the core particle, as has been shown for simple ATP-dependent proteases[2].

One of two deubiquitinating enzymes in the yeast proteasome, Rpn11, forms part of the lid assembly and hovers above the channel[1,6] (Fig. 2a). By removing the poly-ubiquitin chain, which does not readily fit into the narrow translocation channel, Rpn11 promotes substrate degradation. This process is ATP dependent[9,10], although Rpn11 lacks an ATP binding site. Lander and colleagues' structure[1] suggests that, when ATP hydrolysis by the Rpt ring threads the substrate through the channel, it brings the attached ubiquitin chain close to the active site of Rpn11 (refs 1, 9, 10). This arrangement could explain how Rpn11 couples deubiquitination to substrate degradation. In striking contrast to this, the other deubiquitinating enzyme, Ubp6, is located far from the entry port (Fig. 2a), and its activity is accordingly not linked to substrate degradation.

Rpn11 also restricts the accessibility of the entry port (Fig. 2a), a feature not seen in PAN or in any other ATP-dependent protease. This design idiosyncrasy may explain why the proteasome is poor at attacking some protein aggregates and complexes.

Lander *et al.* also show that the bulk of the lid unexpectedly straddles the side of the regulatory particle, with individual subunits extending like fingers to grip and presumably stabilize various elements of the base (Fig. 2b). Moreover, two of the lid subunits, Rpn5 and Rpn6, project as far down as the core particle, thereby stabilizing the interface between it and the regulatory particle. The authors of a recent paper[7] reporting the crystal structure of Rpn6 drew this same conclusion.

In both PAN and the eukaryotic proteasome, structures called coiled-coil elements project like spokes from the N-ring[1,4,5]. In the proteasome, numerous subunits form contacts with these elements[1], suggesting that the coiled coils link the Rpt ring with the rest of the regulatory particle. This is markedly different from the case for PAN, which, apart from its ATPase ring, has no additional subunits for these coiled coils to contact. It is possible that these elements have a completely different function in PAN, perhaps in substrate recognition.

The largest proteasome subunits, Rpn1 and Rpn2, are thought to be scaffolds for ubiquitin receptors, ubiquitin ligases and

Figure 2 | Subunit architecture of the yeast proteasome regulatory particle[1]. a, Tilted view over the Rpt ring (blue) between the Rpt4/5 (right) and Rpt1/2 (not visible) coiled coils. The core particle is shown in green. The centre of the N-ring constitutes the presumed substrate entry port. Highlighted in various colours are the ubiquitin receptors Rpn13 and Rpn10, with its ubiquitin-binding element (Rpn10-UIM), and the deubiquitinating enzymes Rpn11 and Ubp6 (Ubp6 model taken from ref. 14). The position of Ubp6 is approximate and was set manually on the basis of Lander and colleagues' structure[1]. All other subunits are in grey. Free ubiquitin (upper right) is shown for comparison. **b**, Lateral view of the proteasome with the lid (grey) turned to the right. Also visible are the base subunits Rpn10, Rpn13, Rpn1, Rpn2 and, in blue, the Rpt ring (note that Rpn10 straddles the base and lid).

deubiquitinating enzymes[3,11]. The new structures[1,6] show Rpn1 and Rpn2 situated to the side of and above the Rpt ring, respectively (Fig. 2b). Because the exact position of Rpn1 is variable[1,6], the factors that reversibly attach to it may not have fixed positions in the proteasome.

Intriguingly, the proteasome's two ubiquitin receptors, Rpn10 and Rpn13, sit on opposite sides of the port[1,8] (Fig. 2a). This configuration could mean that substrates can reach the port by more than one pathway, which may be an advantage given the tremendous variety of protein structures on which proteasomes must act. It might also help to explain why proteasomes prefer to bind ubiquitin chains of four or more units — longer chains would be able to engage both receptors. Lander and colleagues' structural model should allow this possibility to be tested.

The structure[1] also provides information on the coordination between the six ATPases of the Rpt ring, which work together to drive substrates into the core particle. Some studies have suggested that ATPases act randomly, whereas others indicated a more defined pattern of activity[12]. Surprisingly, Lander *et al.* show that the pore loops of the Rpt proteins form a staircase structure, which is suggestive of a rotary mechanism for ATP hydrolysis and substrate engagement. However, whether this structure represents the ATPases in an idle or active state is unclear.

Given the dynamic nature of the Rpt ring, resolving the mechanism of coordination

among the ATPases may require 'trapping' the proteasome in multiple ATP-bound states for analysis by electron microscopy. Because ATP is small and hard to resolve, the visualization of ubiquitin chains bound to the proteasome might be a more promising option. This and other objectives for future work will build on the structural model presented by Lander and colleagues, which provides a platform for answering questions about the proteasome that were previously beyond reach. ■

Geng Tian *and* **Daniel Finley** *are in the Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. e-mail: daniel_finley@hms.harvard.edu*

1. Lander, G. C. *et al. Nature* **482,** 186–191 (2012).
2. Sauer, R. T. & Baker, T. A. *Annu. Rev. Biochem.* **80,** 587–612 (2011).
3. Finley, D. *Annu. Rev. Biochem.* **78,** 477–513 (2009).
4. Zhang, F. *et al. Mol. Cell* **34,** 473–484 (2009).
5. Djuranovic, S. *et al. Mol. Cell* **34,** 580–590 (2009).
6. Lasker, K. *et al. Proc. Natl Acad. Sci. USA* http://dx.doi.org/10.1073/pnas.1120559109 (2012).
7. Pathare, G. R. *et al. Proc. Natl Acad. Sci. USA* **109,** 149–154 (2012).
8. Sakata, E. *et al. Proc. Natl Acad. Sci. USA* http://dx.doi.org/10.1073/pnas.1119394109 (2012).
9. Verma, R. *et al. Science* **298,** 611–615 (2002).
10. Yao, T. & Cohen, R. E. *Nature* **419,** 403–407 (2002).
11. Elsasser, S. *et al. Nature Cell Biol.* **4,** 725–730 (2002).
12. Smith, D. M., Fraga, H., Reis, C., Kafri, G. & Goldberg, A. L. *Cell* **144,** 526–538 (2011).
13. Walter, J. *et al. Acta Crystallogr. B* **38,** 1462–1472 (1982).
14. www.rcsb.org/pdb/explore.do?structureId=1VJV

ARTICLE

The *Drosophila melanogaster* Genetic Reference Panel

Trudy F. C. Mackay[1]*, Stephen Richards[2]*, Eric A. Stone[1]*, Antonio Barbadilla[3]*, Julien F. Ayroles[1]†, Dianhui Zhu[2], Sònia Casillas[3]†, Yi Han[2], Michael M. Magwire[1], Julie M. Cridland[4], Mark F. Richardson[5], Robert R. H. Anholt[6], Maite Barrón[3], Crystal Bess[2], Kerstin Petra Blankenburg[2], Mary Anna Carbone[1], David Castellano[3], Lesley Chaboub[2], Laura Duncan[1], Zeke Harris[1], Mehwish Javaid[2], Joy Christina Jayaseelan[2], Shalini N. Jhangiani[2], Katherine W. Jordan[1], Fremiet Lara[2], Faye Lawrence[1], Sandra L. Lee[2], Pablo Librado[7], Raquel S. Linheiro[5], Richard F. Lyman[1], Aaron J. Mackey[8], Mala Munidasa[2], Donna Marie Muzny[2], Lynne Nazareth[2], Irene Newsham[2], Lora Perales[2], Ling-Ling Pu[2], Carson Qu[2], Miquel Ràmia[3], Jeffrey G. Reid[2], Stephanie M. Rollmann[1]†, Julio Rozas[7], Nehad Saada[2], Lavanya Turlapati[1], Kim C. Worley[2], Yuan-Qing Wu[2], Akihiko Yamamoto[1], Yiming Zhu[2], Casey M. Bergman[5], Kevin R. Thornton[4], David Mittelman[9] & Richard A. Gibbs[2]

A major challenge of biology is understanding the relationship between molecular genetic variation and variation in quantitative traits, including fitness. This relationship determines our ability to predict phenotypes from genotypes and to understand how evolutionary forces shape variation within and between species. Previous efforts to dissect the genotype–phenotype map were based on incomplete genotypic information. Here, we describe the *Drosophila melanogaster* Genetic Reference Panel (DGRP), a community resource for analysis of population genomics and quantitative traits. The DGRP consists of fully sequenced inbred lines derived from a natural population. Population genomic analyses reveal reduced polymorphism in centromeric autosomal regions and the *X* chromosome, evidence for positive and negative selection, and rapid evolution of the *X* chromosome. Many variants in novel genes, most at low frequency, are associated with quantitative traits and explain a large fraction of the phenotypic variance. The DGRP facilitates genotype–phenotype mapping using the power of *Drosophila* genetics.

Understanding how molecular variation maps to phenotypic variation for quantitative traits is central for understanding evolution, animal and plant breeding, and personalized medicine[1,2]. The principles of mapping quantitative trait loci (QTLs) by linkage to, or association with, marker loci are conceptually simple[1,2]. However, we have not yet achieved our goal of explaining genetic variation for quantitative traits in terms of the underlying genes; additive, epistatic and pleiotropic effects as well as phenotypic plasticity of segregating alleles; and the molecular nature, population frequency and evolutionary dynamics of causal variants. Efforts to dissect the genotype–phenotype map in model organisms[3,4] and humans[5–7] have revealed unexpected complexities, implicating many, novel loci, pervasive pleiotropy, and context-dependent effects.

Model organism reference populations of inbred strains that can be shared among laboratories studying diverse phenotypes, and for which environmental conditions can be controlled and manipulated, greatly facilitate efforts to dissect the genetic architecture of quantitative traits[3,4]. Measuring many individuals of the same homozygous genotype increases the accuracy of the estimates of genotypic value[1] and the power to detect variants, and genotypes of molecular markers need only be obtained once. We constructed the *Drosophila melanogaster* Genetic Reference Panel (DGRP) as such a community resource. Unlike previous populations of recombinant inbred lines derived from limited samples of genetic variation, the DGRP consists

of 192 inbred strains derived from a single outbred population. The DGRP contains a representative sample of naturally segregating genetic variation, has an ultra-fine-grained recombination map suitable for precise localization of causal variants, and has almost complete euchromatic sequence information.

Here, we describe molecular and phenotypic variation in 168 re-sequenced lines comprising Freeze 1.0 of the DGRP, population genomic inferences of patterns of polymorphism and divergence and their correlation with genomic features, local recombination rate and selection acting on this population, genome-wide association mapping analyses for three quantitative traits, and tools facilitating the use of this resource.

Molecular variation in the DGRP

We constructed the DGRP by collecting mated females from the Raleigh, North Carolina, USA, population, followed by 20 generations of full-sibling inbreeding of their progeny. We sequenced 168 DGRP lines using a combination of Illumina and 454 sequencing technology: 29 of the lines were sequenced using both platforms, 129 lines have only Illumina sequence, and 10 lines have only 454 sequence. We mapped sequence reads to the *D. melanogaster* reference genome, re-calibrated base quality scores, and locally re-aligned Illumina reads. Mean sequence coverage was 21.4× per line for Illumina sequences and 12.1× per line for 454 sequences (Supplementary

[1]Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695, USA. [2]Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030 USA. [3]Genomics, Bioinformatics and Evolution Group, Institut de Biotecnologia i de Biomedicina - IBB/Department of Genetics and Microbiology, Campus Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain. [4]Department of Ecology and Evolutionary Biology, University of California - Irvine, Irvine, California 92697, USA. [5]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK. [6]Department of Biology, North Carolina State University, Raleigh, North Carolina 27695, USA. [7]Molecular Evolutionary Genetics Group, Department of Genetics, Faculty of Biology, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain. [8]Center for Public Health Genomics, University of Virginia, PO Box 800717, Charlottesville, Virginia 22908, USA. [9]Virginia Bioinformatics Institute and Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia 24061, USA. †Present addresses: FAS Society of Fellows, Harvard University, 78 Mt Auburn Street, Cambridge, Massachusetts 02138, USA (J.F.A.); Functional Comparative Genomics Group, Institut de Biotecnologia i de Biomedicina - IBB, Campus Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain (S.C.); Department of Biological Sciences, University of Cincinnati, Cincinnati, Ohio 45221, USA (S.M.R.).
*These authors contributed equally to this work.

Table 1). On average, we assayed 113.5 megabases (94.25%) of the euchromatic reference sequence with ~22,000 read mapping gaps per line (Supplementary Table 2). We called 4,672,297 single nucleotide polymorphisms (SNPs) using the Joint Genotyper for Inbred Lines (JGIL; E.A.S., personal communication), which takes into account coverage and quality sequencing statistics, and expected allele frequencies after 20 generations of inbreeding from an outbred population initially in Hardy–Weinberg equilibrium. In cases where base calls were made by both technologies, concordance was 99.36% (Supplementary Table 3).

The SNP site frequency distribution (Fig. 1a) is characterized by a majority of low frequency variants. The numbers of SNPs vary by chromosome and site class (Fig. 1b). Linkage disequilibrium[8] decays to $r^2 = 0.2$ on average within 10 base pairs on autosomes and 30 base pairs on the X chromosome (Fig. 1c and Supplementary Fig. 1). This difference is expected because the population size of the X chromosome is three quarters that of autosomes, and the X chromosome can experience greater purifying selection because of exposure of deleterious recessive alleles in hemizygous males. There is little evidence of global population structure in the DGRP (Fig. 1d and Supplementary Fig. 2). The rapid decline in linkage disequilibrium locally and lack of global population structure are favourable for genome-wide association mapping.

Not all SNPs are fixed within individual DGRP lines (Supplementary Table 4). The expected inbreeding coefficient (F) after 20 generations of full-sibling inbreeding[1] is $F = 0.986$; therefore, we expect some SNPs to remain segregating by chance. Segregating SNPs can also arise from new mutations, or if natural selection opposes inbreeding, due to true overdominance for fitness at individual loci or associative overdominance due to complementary deleterious alleles that are closely linked or in segregating inversions.

We identified 390,873 microsatellite loci, 105,799 of which were polymorphic (Supplementary Table 5); 36,810 transposable element insertion sites and 197,402 total insertions (Supplementary Table 6). On average, each line contained 1,175 transposable element insertions (Supplementary Table 6), although most transposable element insertion sites (25,562) were present in only one line (Supplementary

Table 7). We identified 149 transposable element families. The number of copies per family varied greatly from an average of 315.7 INE-1 elements per line to an average of 0.003 Gandalf-Dkoe-like elements per line (Supplementary Table 8).

Wolbachia pipientis is a maternally inherited bacterium found in insects, including Drosophila, and can affect reproduction[9]. We assessed Wolbachia infection status in the DGRP lines to account for it in analyses of genotype–phenotype associations, and found 51.2% of lines harbouring sufficient Wolbachia DNA to imply infection (Supplementary Table 9).

Polymorphism and divergence

We used the DGRP Illumina sequence data and genome sequences from Drosophila simulans and Drosophila yakuba[10] to perform genome-wide analyses of polymorphism and divergence, assess the association of these parameters with genomic features and the recombination landscape, and infer the historical action of selection on a much larger scale than had been possible previously[11–16]. We computed polymorphism (π and θ, refs 17 and 18) and divergence (k, ref. 19) for the whole genome, by chromosome arm (X, 2L, 2R, 3L, 3R), by chromosome region (three regions of equal size in Mb — telomeric, middle and centromeric), in 50-kbp non-overlapping windows, and by site class (synonymous and non-synonymous sites within coding sequences, and intronic, untranslated region (UTR) and intergenic sites) (Supplementary Tables 10 and 11).

Averaged over the entire genome, $\pi = 0.0056$ and $\theta = 0.0067$, similar to previous estimates from North American populations[16,20]. Average polymorphism on the X chromosome ($\pi_X = 0.0040$) is reduced relative to the autosomes ($\pi_A = 0.0060$) (X/A ratio = 0.67, Wilcoxon test $P = 0$), even after correcting for the X/A effective population size ($X_{4/3} = 0.0054$, Wilcoxon test $P < 0.00002$; Supplementary Table 10). Autosomal nucleotide diversity is reduced on average 2.4-fold in centromeric regions relative to non-centromeric regions, and at the telomeres (Fig. 2a and Supplementary Table 10), whereas diversity is relatively constant along the X chromosome. Thus, $\pi_X > \pi_A$ in centromeric regions, but $\pi_A > \pi_X$ in other chromosomal regions (Fig. 2a and Supplementary Table 10).
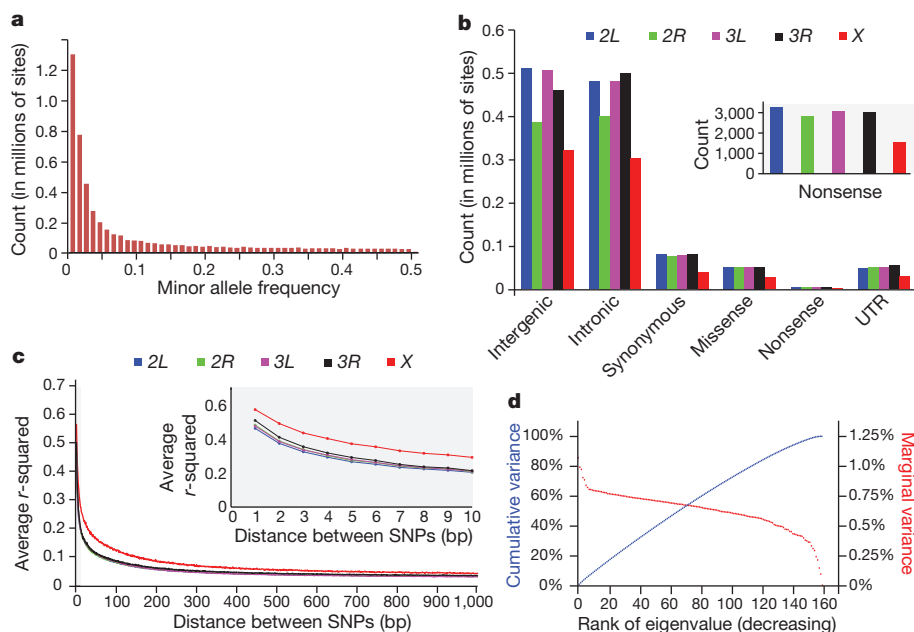

Figure 1 | SNP variation in the DGRP lines. a, Site frequency spectrum. **b**, Numbers of SNPs per site class. **c**, Decay of linkage disequilibrium (r^2) with physical distance for the five major chromosome arms. **d**, Lack of population structure. The red curve depicts the ranked eigenvalues of the genetic covariance matrix in decreasing order with respect to the marginal variance explained; the blue curve shows their cumulative sum as a fraction of the total with respect to cumulative variance explained. The partitioning of total genetic variance is balanced among the eigenvectors. The principal eigenvector explains < 1.1% of the total genetic variance.
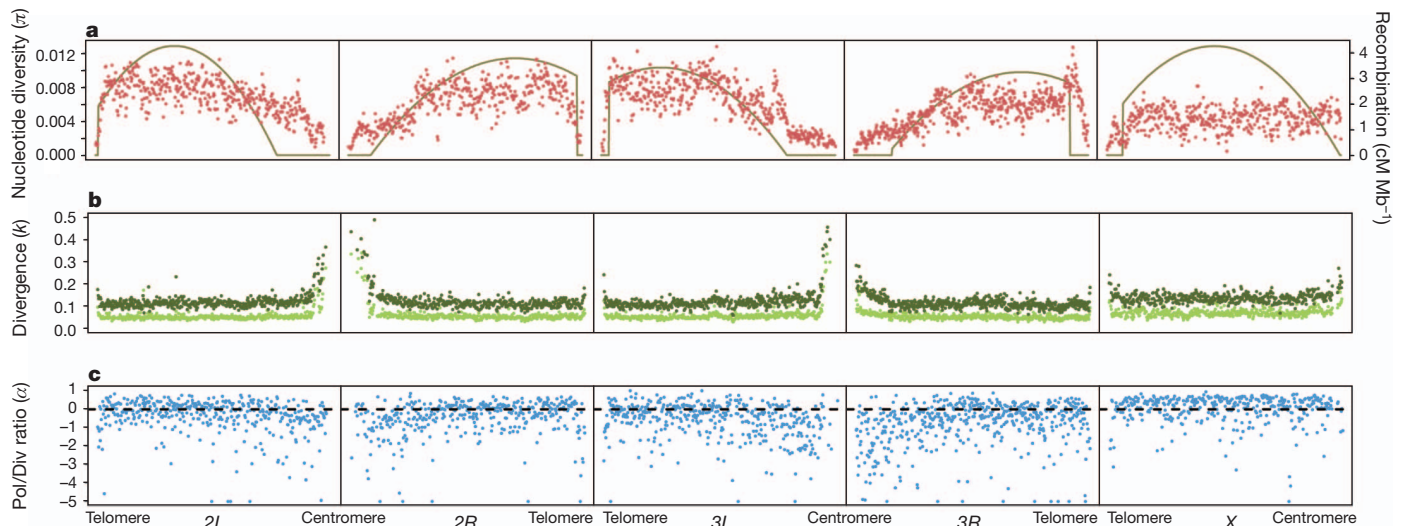
Figure 2 | Pattern of polymorphism, divergence, α and recombination rate along chromosome arms in non-overlapping 50-kbp windows. a, Nucleotide polymorphism (π). The solid curves give the recombination rate (cM Mb^{-1}). **b**, Divergence (k) for *D. simulans* (light green) and *D. yakuba* (dark green). **c**, Polymorphism to divergence ratio (Pol/Div), estimated as $1 - [(\pi_{0\text{-fold}}/\pi_{4\text{-fold}})/(k_{0\text{-fold}}/k_{4\text{-fold}})]$. An excess of 0-fold divergence relative to polymorphism ($k_{0\text{-fold}}/k_{4\text{-fold}}) > (\pi_{0\text{-fold}}/\pi_{4\text{-fold}})$ is interpreted as adaptive fixation whereas an excess of 0-fold polymorphism relative to divergence ($\pi_{0\text{-fold}}/\pi_{4\text{-fold}}) > (k_{0\text{-fold}}/k_{4\text{-fold}})$ indicates that weakly deleterious or nearly neutral mutations are segregating in the population.

Genes on the *X* chromosome evolve faster ($k_X = 0.140$) than autosomal genes ($k_A = 0.126$) (*X/A* ratio = 1.131, Wilcoxon test $P = 0$) (Fig. 2b and Supplementary Table 10). Divergence is more uniform (coefficient of variation $(CV)_k = 0.2841$) across chromosome arms than is polymorphism ($CV_\pi = 0.4265$). The peaks of divergence near the centromeres could be attributable to the reduced quality of alignments in these regions. Patterns of divergence are similar regardless of the outgroup species used (Fig. 2b and Supplementary Table 11).

The pattern of polymorphism and divergence by site class is consistent within and among chromosomes ($\pi_{k_{\text{Synonymous}}} > \pi_{k_{\text{Intron}}} > \pi_{k_{\text{Intergenic}}} > \pi_{k_{\text{UTR}}} > \pi_{k_{\text{Non-synonymous}}}$), in agreement with previous studies on smaller data sets[12,15] (Supplementary Figs 3 and 4 and Supplementary Table 11). Polymorphism levels between synonymous and non-synonymous sites differ by an order of magnitude. Variation and divergence patterns within the site classes generally follow the same patterns observed overall, with reduced polymorphism for all site classes on the *X* chromosome relative to autosomes, increased *X* chromosome divergence relative to autosomes for all but synonymous sites, decreased polymorphism in centromeric regions, and greater variation among regions and arms for polymorphism than for divergence. Other diversity measures and more detailed patterns at different window-sizes for each chromosome arm can be accessed from the Population *Drosophila* Browser (popDrowser) (Table 1 and Methods).

Recombination landscape

Evolutionary models of hitchhiking and background selection[21,22] predict a positive correlation between polymorphism and recombination rate. This expectation is realized in regions where recombination is less than 2 cM Mb^{-1} (Spearman's $\rho = 0.471$, $P = 0$), but recombination and polymorphism are independent in regions where recombination exceeds 2 cM Mb^{-1} (Spearman's $\rho = -0.0044$, $P = 0.987$) (Fig. 2a and Supplementary Table 12). The average rate of recombination of the *X* chromosome (2.9 cM Mb^{-1}) is greater than that of autosomes (2.1 cM Mb^{-1}), which may account for the low overall *X*-linked correlation between recombination rate and π. The lack of correlation between recombination and divergence (Supplementary Table 12) excludes mutation associated with recombination as the cause of the correlation. We assessed the independent effects of recombination rate, divergence, chromosome region and gene density on nucleotide variation of autosomes and the *X* chromosome (Supplementary Table 13). Recombination is the major predictor of

polymorphism on the *X* chromosome and autosomes; however, the significant effect of autosomal chromosome region remains after accounting for variation in recombination rates between centromeric and non-centromeric regions.

Selection regimes

We used the standard[23] and generalized[12,24,25] McDonald Kreitman tests (MKT) to scan the genome for evidence of selection. These tests

Table 1 | Community resources

Resource	Location
DGRP lines	Bloomington *Drosophila* Stock Center http://flystocks.bio.indiana.edu/Browse/RAL.php
Sequences	Baylor College of Medicine Human Genome Sequencing Center http://www.hgsc.bcm.tmc.edu/project-species-i-DGRP_lines.hgsc National Center for Biotechnology Information Short Read Archive http://www.ncbi.nlm.nih.gov/sra?term=DGRP Mackay Laboratory http://dgrp.gnets.ncsu.edu/
Read alignments	Baylor College of Medicine Human Genome Sequencing Center http://www.hgsc.bcm.tmc.edu/projects/dgrp/
SNPs	Baylor College of Medicine Human Genome Sequencing Center http://www.hgsc.bcm.tmc.edu/projects/dgrp/freeze1_July_2010/snp_calls/ National Center for Biotechnology Information dbSNP http://www.ncbi.nlm.nih.gov/SNP/snp_viewBatch.cgi?sbid=1052186 Mackay Laboratory http://dgrp.gnets.ncsu.edu/
Microsatellites	Baylor College of Medicine Human Genome Sequencing Center http://www.hgsc.bcm.tmc.edu/projects/dgrp/freeze1_July_2010/microsat_calls/ Mittelman Laboratory http://genome.vbi.vt.edu/public/DGRP/
Transposable elements	Mackay Laboratory http://dgrp.gnets.ncsu.edu/
Molecular population genomics	PopDrowser http://popdrowser.uab.cat
Phenotypes	Mackay Laboratory http://dgrp.gnets.ncsu.edu/
Genome-wide association analysis	Mackay Laboratory http://dgrp.gnets.ncsu.edu/

compare the ratio of polymorphism at a selected site with that of a neutral site to the ratio of divergence at a selected site to divergence at a neutral site. The standard MKT is applied to coding sequences, and synonymous and non-synonymous sites are used as putative neutral and selected sites, respectively. The generalized MKT is applied to non-coding sequences and uses fourfold degenerate sites as neutral sites. Using polymorphism and divergence data avoids confounding inference of selection with mutation rate differences, and restricting the tests to closely linked sites controls for shared evolutionary history[26–28]. We infer adaptive divergence when there is an excess of divergence relative to polymorphism, and segregation of slightly deleterious mutations when there is an excess of polymorphism over divergence. Estimates of α, the proportion of adaptive divergence, are biased downwards by low frequency, slightly deleterious mutations[29,30]. Rather than eliminate low frequency variants[31], we incorporated information on the site frequency distribution to the MKT test framework to obtain estimates of the proportion of sites that are strongly deleterious (d), weakly deleterious (b), neutral (f) and recently neutral (γ) at segregating sites, as well as unbiased estimates of α (Supplementary Methods).

Deleterious and neutral sites

Averaged over the entire genome, we infer that 58.5% of the segregating sites are neutral or nearly neutral, 1.9% are weakly deleterious and 39.6% are strongly deleterious. However, these proportions vary between the X chromosome and autosomes, site classes and chromosome regions (Supplementary Tables 14–16 and Fig. 3). Non-synonymous sites are the most constrained (d = 77.6%), whereas in non-coding sites d ranges from 29.1% in 5′ UTRs to 41.3% in 3′ intergenic regions. The inferred pattern of selection differs between autosomal centromeric and non-centromeric regions: d is reduced and f is increased in centromeric regions for all site categories (Fig. 3). We observe an excess of polymorphism relative to divergence in autosomal centromeric regions, even after correcting for weakly deleterious mutations, implying a relaxation of selection from the time of separation of D. melanogaster and D. yakuba. Because selection coefficients depend on the effective population size[32] (N_e), this could occur if the recombination rate has specifically diminished in centromeric regions during the divergence between D. melanogaster and D. yakuba; or with an overall reduction of N_e associated with the colonization of North American habitats[33,34]. In the latter case, we expect a genome-wide signature of an excess of low-frequency polymorphisms and of polymorphism relative to divergence, exacerbated in regions of low recombination. We indeed find an excess of low-frequency polymorphism relative to neutral expectation as indicated by the negative estimates of Tajima's D statistic[35]

(D = −0.686 averaged over the whole genome and D = −0.997 in autosomal centromeric regions). In contrast, the X chromosome does not show a differential pattern of selection in the centromeric region, has a lower fraction of relaxation of selection, fewer neutral alleles, and a higher percentage of strongly deleterious alleles for all site classes and regions (Fig. 3 and Supplementary Tables 14–16).

Transposable element insertions are thought to be largely deleterious. There are more singleton insertions in regions of high recombination ($\geq 2\,\mathrm{cM\,Mb^{-1}}$) and more insertions shared in multiple lines in regions of low recombination ($< 2\,\mathrm{cM\,Mb^{-1}}$) (Fisher's exact test P = 0), and comparison of observed and expected site occupancy spectra reveals an excess of singleton insertions (P = 0, Supplementary Fig. 5).

Adaptive fixation

We find substantial evidence for positive selection in autosomal non-centromeric regions and the X chromosome (Fig. 2c and Supplementary Tables 15 and 17). We estimated α by aggregating all sites in each region analysed to avoid underestimation by averaging across genes[36] in comparisons of chromosomes, regions and site classes. We also computed the direction of selection, DoS[37], which is positive with adaptive selection, zero under neutrality and negative when weakly deleterious or new nearly neutral mutations are segregating. Estimates of α from the standard and generalized MKT indicate that on average 25.2% of the fixed sites between D. melanogaster and D. yakuba are adaptive, ranging from 30% in introns to 7% in UTR sites (Supplementary Fig. 6). Estimates of DoS and α are negative for non-synonymous and UTR sites in the autosomal centromeres, consistent with underestimating the fraction of adaptive substitutions in regions of low recombination because weakly deleterious or nearly neutral mutations are more common than adaptive fixations. The majority of adaptive fixation on autosomes occurs in non-centromeric regions (Fig. 2c). We find over four times as many adaptive fixations on the X chromosome relative to autosomes. The pattern holds for all site classes, in particular non-synonymous sites and UTRs, as well as individual genes, and is not solely due to the autosomal centromeric effect (Supplementary Table 15 and Supplementary Figs 6 and 7). Finally, when we consider DoS in recombination environments above and below $2\,\mathrm{cM\,Mb^{-1}}$, we find greater adaptive propensity in genes whose recombination context is $\geq 2\,\mathrm{cM\,Mb^{-1}}$ (Wilcoxon test, P = 0; Supplementary Fig. 8).

To understand the global patterns of divergence and constraint across functional classes of genes, we examined the distributions of ω (d_N/d_S, the ratio of non-synonymous to synonymous divergence) and DoS across gene ontology (GO) categories. The 4.9% GO categories with significantly elevated DoS include the biological process categories of behaviour, developmental process involved in reproduction, reproduction and ion transport (Supplementary Table 18). Recombination context is the major determinant of variation in DoS (Supplementary Table 19) whereas GO category is as important as recombinational context for predicting variation in ω (Supplementary Table 19).

GO categories enriched for positive DoS values differ from those associated with high values of ω (Supplementary Table 18), indicating that positive selection does not occur necessarily on genes with high ω values. If adaptive substitutions are common, high values of ω reflect the joint contributions of neutral and adaptive substitutions. Further, equating high constraint (low ω) with functional importance overlooks the functional role of adaptive changes[15]. Unlike ω, DoS takes into account the constraints inferred from the current polymorphism, distinguishing negative, neutral and adaptive selection.

Genome-wide genotype-phenotype associations

We measured resistance to starvation stress, chill coma recovery time and startle response[38] in the DGRP. We found considerable genetic variation for all traits, with high broad sense heritabilities. We also found variation in sex dimorphism for starvation resistance and chill

Figure 3 | The fraction of alleles segregating under different selection regimes by site class and chromosome region, for the autosomes (A) and the X chromosome (X). The selection regimes are strongly deleterious (d, dark blue), weakly deleterious (b, blue), recently neutral (γ, white) and old neutral (f − γ, light blue). Each chromosome arm has been divided in three regions of equal size (in Mb): centromere, middle and telomere.

coma recovery with cross-sex genetic correlations significantly different from unity (Supplementary Tables 20–22).

We performed genome-wide association analyses for these traits, using the 2,490,165 SNPs and 77,756 microsatellites for which the minor allele was represented in four or more lines, using single-locus analyses pooled across sexes and separately for males and females. At $P < 10^{-5}$ ($P < 10^{-6}$), we find 203 (32) SNPs and 2 (0) microsatellites associated with starvation resistance; 90 (7) SNPs and 4 (2) microsatellites associated with startle response; and 235 (45) SNPs and 5 (3) microsatellites associated with chill coma recovery time (Fig. 4a, Supplementary Fig. 9 and Supplementary Tables 23 and 24). The minor allele frequencies for most of the associated SNPs are low, and there is an inverse relationship between effect sizes and minor allele frequency (Supplementary Fig. 10).

The DGRP is a powerful tool for rapidly reducing the search space for molecular variants affecting quantitative traits from the entire genome to candidate polymorphisms and genes. Although we cannot infer which of these polymorphisms are causal due to linkage disequilibrium between SNPs in close physical proximity as well as occasional spurious long range linkage disequilibrium (Fig. 4a and Supplementary Fig. 9), the candidate gene lists are likely to be enriched for causal variants. The majority of associations are in computationally predicted genes or genes with annotated functions not obviously associated with the three traits. However, genes previously associated with startle response[39] (*Sema-1a* and *Eip75B*) and starvation resistance[40] (*pnt*) were identified in this study; and a SNP in *CG3213*, previously identified in a *Drosophila* obesity screen[41], is associated with variation in starvation resistance. Several genes associated with quantitative traits are rapidly evolving (*psq*, *Egfr*; Supplementary Tables 17 and 23) or are plausible candidates based on SNP or gene ontology annotations (Supplementary Table 23).

Predicting phenotypes from genotypes

We used regression models to predict trait phenotypes from SNP genotypes and estimate the total variance explained by SNPs. The latter cannot be done by summing the individual contributions of the single marker effects because markers are not completely independent, and estimates of effects of single markers are biased when more than one locus affecting the trait segregates in the population. We derived gene-centred multiple regression models to estimate the effects of multiple SNPs simultaneously. In all cases 6–10 SNPs explain from 51–72% of the phenotypic variance and 65–90% of the genetic variance (Supplementary Tables 25 and 26 and Supplementary Figs 11–13). We also derived partial least square regression models using all SNPs for which the single marker effect was significant

at $P < 10^{-5}$. These models explain 72–85% of the phenotypic variance (Fig. 4b, c and Supplementary Fig. 14).

Discussion

The DGRP lines, sequences, variant calls, phenotypes and web tools for molecular population genomics and genome-wide association analysis are publicly available (Table 1). The DGRP lines contain at least 4,672,297 SNPs, 105,799 polymorphic microsatellites and 36,810 transposable elements, as well as insertion/deletion events and copy number variants and are a valuable resource for understanding the genetic architecture of quantitative traits of ecological and evolutionary relevance as well as *Drosophila* models of human quantitative traits. These novel mutations have survived the sieve of natural selection and will enhance the functional annotation of the *Drosophila* genome, complementing the *Drosophila* Gene Disruption Project[42] and the *Drosophila* modENCODE project[43].

Genome-wide molecular population genetic analyses show that patterns of polymorphism, but not divergence, differ by autosomal chromosome region, and between the *X* chromosome and autosomes. Polymorphism is lower in autosomal centromeric than non-centromeric regions, but not for the *X* chromosome. We propose that the correlation of polymorphism with recombination in regions where recombination is $< 2\,\mathrm{cM\,Mb}^{-1}$ is due to the reduced effective population size in regions of low recombination[8]. Selection is less efficient in regions of low recombination[32], consistent with our observation that the fraction of strongly deleterious mutations and positively selected sites are reduced in these regions.

All molecular population genomic analyses support the 'faster *X*' hypothesis[44]. Relative to the autosomes, the *X* chromosome shows lower polymorphism, faster rates of molecular evolution, a higher percentage of gene regions undergoing adaptive evolution, a higher fraction of strongly deleterious sites, and a lower level of weak negative selection and relaxation of selection. New *X*-linked mutations are directly exposed to selection each generation in hemizygous males, and the *X* chromosome has greater recombination than autosomes[44]; both of these factors could contribute to this observation.

Genome-wide association analyses of three fitness-related quantitative traits reveal hundreds of novel candidate genes, highlighting our ignorance of the genetic basis of complex traits. Most variants associated with the traits are at low frequency, and there is an inverse relationship between frequency and effect. Given that low-frequency alleles are likely to be deleterious for traits under directional or stabilizing selection, these results are consistent with the mutation–selection balance hypothesis[1] for the maintenance of quantitative genetic variation. Regression models incorporating significant SNPs
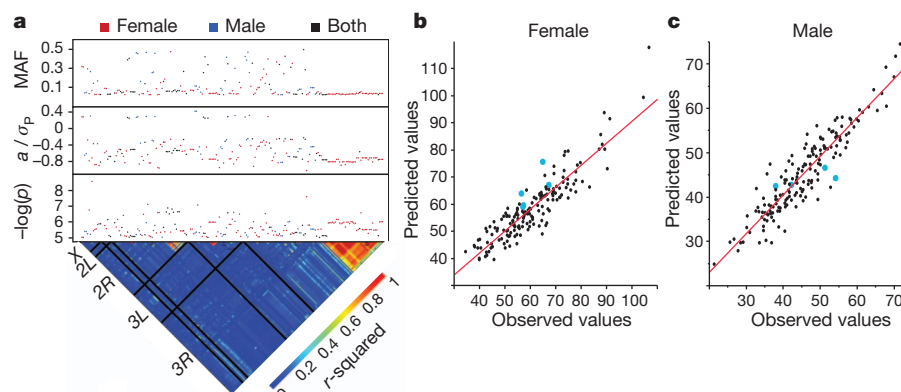

Figure 4 | Genotype–phenotype associations for starvation resistance. a, Genome-wide association results for significant SNPs. The lower triangle depicts linkage disequilibrium (r^2) among SNPs, with the five major chromosome arms demarcated by black lines. The upper panels give the significance threshold ($-\log(p)$, uncorrected for multiple tests), the effect in phenotypic standard deviation units, and the minor allele frequency (MAF). **b, c**, Partial least squares regressions of phenotypes predicted using SNP data on observed phenotypes. The blue dots represent the predicted and observed phenotypes of lines that were not included in the initial study. **b**, Females ($r^2 = 0.81$); **c**, males ($r^2 = 0.85$).

explain most of the phenotypic variance of the traits, in contrast with human association studies, where significant SNPs have tiny effects and together explain a small fraction of the total phenotypic variance[7]. If the genetic architecture of human complex traits is also dominated by low-frequency causal alleles, we expect estimates of effect size based on linkage disequilibrium with common variants to be strongly biased downwards.

In the future, the full power of *Drosophila* genetics can be applied to validating marker-trait associations: mutations, RNA interference constructs and quantitative trait loci mapping populations. The DGRP is an ideal resource for systems genetics analyses of the relationship between molecular variation, causal molecular networks and genetic variation for complex traits[4,38,45], and will anchor evolutionary studies in comparison with sequenced *Drosophila* species to assess to what extent variation within a species corresponds to variation among species.

METHODS SUMMARY

The full Methods are in the Supplementary Information. Information on sequencing and bioinformatics includes methods for DNA isolation; library construction and genomic sequencing; sequence read alignment; SNP, microsatellite and transposable element identification; genotypes for assurance of sample identity; and *Wolbachia* detection. Methods for molecular population genomics analysis include details of recombination estimates; diversity measures; linkage disequilibrium and neutrality tests; software used for population genomic analysis; data visualization (popDrowser); standard and generalized McDonald–Kreitman tests, statistical analysis methods; quality assessment and data filtering; and gene ontology analyses. Methods for quantitative genetic analyses include phenotype measures, quantitative genetic analyses of phenotypes, statistical analyses of genotype–phenotype associations and predictive models, and a web-based association analysis pipeline.

1. Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* 4th edn (Longman, 1996).
2. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, 1998).
3. Flint, J. & Mackay, T. F. C. Genetic architecture of quantitative traits in flies, mice and humans. *Genome Res.* **19,** 723–733 (2009).
4. Mackay, T. F. C., Stone, E. A. & Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. *Nature Rev. Genet.* **10,** 565–577 (2009).
5. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322,** 881–888 (2008).
6. Donnelly, P. Progress and challenges in genome-wide association studies in humans. *Nature* **456,** 728–731 (2008).
7. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461,** 747–753 (2009).
8. Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **8,** 269–294 (1966).
9. Werren, J. H. Biology of *Wolbachia*. *Annu. Rev. Entomol.* **42,** 587–609 (1997).
10. Clark, A. G. *et al.* Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450,** 203–218 (2007).
11. Smith, N. G. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415,** 1022–1024 (2002).
12. Andolfatto, P. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437,** 1149–1152 (2005).
13. Presgraves, D. C. Recombination enhances protein adaptation in *Drosophila melanogaster*. *Curr. Biol.* **15,** 1651–1656 (2005).
14. Casillas, S., Barbadilla, A. & Bergman, C. Purifying selection maintains highly conserved noncoding sequences in *Drosophila*. *Mol. Biol. Evol.* **24,** 2222–2234 (2007).
15. Sella, G. *et al.* Pervasive natural selection in the *Drosophila* genome? *PLoS Genet.* **5,** e1000495 (2009).
16. Sackton, T. B. *et al.* Population genomic inferences from sparse high-throughput sequencing of two populations of *Drosophila melanogaster*. *Genome Biol. Evol.* **1,** 449–465 (2009).
17. Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, 1987).
18. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7,** 256–276 (1975).
19. Jukes, T. H. & Cantor, C. R. in *Mammalian Protein Metabolism* vol. 3 (eds Munro, H. N. & Allison, J. B.) 21–132 (Academic Press, 1969).
20. Andolfatto, P. & Przeworski, M. Regions of lower crossing over harbor more rare variants in African *Drosophila melanogaster*. *Genetics* **158,** 657–665 (2001).
21. Begun, D. J. & Aquadro, C. F. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356,** 519–520 (1992).
22. Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134,** 1289–1303 (1993).
23. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351,** 652–654 (1991).
24. Jenkins, D. L., Ortori, C. A. & Brookfield, J. F. A test for adaptive change in DNA sequences controlling transcription. *Proc. R. Soc. Lond. B* **261,** 203–207 (1995).
25. Egea, R., Casillas, S. & Barbadilla, A. Standard and generalized McDonald–Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res.* **36,** W157–W162 (2008).
26. Sawyer, S. A. & Hartl, D. L. Population genetics of polymorphism and divergence. *Genetics* **132,** 1161–1176 (1992).
27. Nielsen, R. Statistical tests of selective neutrality in the age of genomics. *Heredity* **86,** 641–647 (2001).
28. Eyre-Walker, A. Changing effective population size and the McDonald-Kreitman test. *Genetics* **162,** 2017–2024 (2002).
29. Charlesworth, J. & Eyre-Walker, A. The McDonald-Kreitman test and slightly deleterious mutations. *Mol. Biol. Evol.* **25,** 1007–1015 (2008).
30. Eyre-Walker, A. & Keightley, P. D. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* **26,** 2097–2108 (2009).
31. Fay, J. C., Wyckoff, G. J. & Wu, C. I. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415,** 1024–1026 (2002).
32. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246,** 96–98 (1973).
33. David, J. R. & Capy, P. Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet.* **4,** 106–111 (1988).
34. Begun, D. J. & Aquadro, C. F. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365,** 548–550 (1993).
35. Tajima, F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–595 (1989).
36. Smith, N. G. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415,** 1022–1024 (2002).
37. Stoletzki, N. & Eyre-Walker, A. Estimation of the neutrality index. *Mol. Biol. Evol.* **28,** 63–70 (2011).
38. Ayroles, J. F. *et al.* Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genet.* **41,** 299–307 (2009).
39. Yamamoto, A. *et al.* Neurogenetic networks for startle-induced locomotion in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **105,** 12393–12398 (2008).
40. Harbison, S. T., Yamamoto, A. H., Fanara, J. J., Norga, K. K. & Mackay, T. F. C. Quantitative trait loci affecting starvation resistance in *Drosophila melanogaster*. *Genetics* **166,** 1807–1823 (2004).
41. Pospisilik, J. A. *et al. Drosophila* genome-wide obesity screen reveals hedgehog as a determinant of brown versus white adipose cell fate. *Cell* **140,** 148–160 (2010).
42. Bellen, H. J. *et al.* The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* **167,** 761–781 (2004).
43. The ModENCODE Consortium. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330,** 1787–1797 (2010).
44. Charlesworth, B., Coyne, J. A. & Barton, N. H. The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* **130,** 113–146 (1987).
45. Sieberts, S. K. & Schadt, E. E. Moving toward a system genetics view of disease. *Mamm. Genome* **18,** 389–401 (2007).

Author Contributions T.F.C.M., S.R. and R.A.G. conceived the project. T.F.C.M., S.R., A.B. and E.A.S. wrote the main manuscript. T.F.C.M., S.R., A.B., E.A.S., J.F.A., K.R.T., J.M.C., C.M.B. and D.M. wrote the Supplementary methods. M.M.M., C.B., K.P.B., M.A.C., L.C., L.D., Y.H., M.J., J.C.J., S.N.J., K.W.J., F. Lara, F. Lawrence, S.L.L., R.F.L., M.M., D.M.M., L.N., I.M., L.P., L.L.P., C.Q., J.G.R., S.M.R., L.T., K.C.W., Y.-Q.W., A.Y. and Y.Z. performed experiments. T.F.C.M., A.B., J.F.A., D.Z., S.C., M.M.M., J.M.C., M.F.R., M.B., D.C., R.S.L., A.M., C.M.B., K.R.T., D.M. and E.A.S. did the bioinformatics and data analysis. J.F.A., S.C., M.M.M., Z.H., P.L., M.R., J.R. and E.A.S. wrote the Methods and did the web site development. R.R.H.A. contributed resources.

ARTICLE

doi:10.1038/nature10809

Inflammasome–mediated dysbiosis regulates progression of NAFLD and obesity

Jorge Henao–Mejia[1]*, Eran Elinav[1]*, Chengcheng Jin[1,2]*, Liming Hao[3], Wajahat Z. Mehal[4], Till Strowig[1], Christoph A. Thaiss[1], Andrew L. Kau[5,6], Stephanie C. Eisenbarth[7], Michael J. Jurczak[4], Joao–Paulo Camporez[4], Gerald I. Shulman[4,8], Jeffrey I. Gordon[5], Hal M. Hoffman[9] & Richard A. Flavell[1,8]

Non–alcoholic fatty liver disease (NAFLD) is the hepatic manifestation of metabolic syndrome and the leading cause of chronic liver disease in the Western world. Twenty per cent of NAFLD individuals develop chronic hepatic inflammation (non–alcoholic steatohepatitis, NASH) associated with cirrhosis, portal hypertension and hepatocellular carcinoma, yet the causes of progression from NAFLD to NASH remain obscure. Here, we show that the NLRP6 and NLRP3 inflammasomes and the effector protein IL–18 negatively regulate NAFLD/NASH progression, as well as multiple aspects of metabolic syndrome via modulation of the gut microbiota. Different mouse models reveal that inflammasome–deficiency–associated changes in the configuration of the gut microbiota are associated with exacerbated hepatic steatosis and inflammation through influx of TLR4 and TLR9 agonists into the portal circulation, leading to enhanced hepatic tumour–necrosis factor (TNF)–α expression that drives NASH progression. Furthermore, co–housing of inflammasome–deficient mice with wild–type mice results in exacerbation of hepatic steatosis and obesity. Thus, altered interactions between the gut microbiota and the host, produced by defective NLRP3 and NLRP6 inflammasome sensing, may govern the rate of progression of multiple metabolic syndrome–associated abnormalities, highlighting the central role of the microbiota in the pathogenesis of heretofore seemingly unrelated systemic auto–inflammatory and metabolic disorders.

The prevalence of non–alcoholic fatty liver disease (NAFLD) ranges from 20–30% in the general population and up to 75–100% in obese individuals[1,2]. NAFLD is considered one of the manifestations of metabolic syndrome[3]. Whereas most patients with NAFLD remain asymptomatic, 20% progress to develop chronic hepatic inflammation (non–alcoholic steatohepatitis, NASH), which in turn can lead to cirrhosis, portal hypertension, hepatocellular carcinoma and increased mortality[4–6]. Despite its high prevalence, factors leading to progression from NAFLD to NASH remain poorly understood and no treatment has proven effective[7,8].

A "two hit" mechanism is proposed to drive NAFLD/NASH pathogenesis[9]. The first hit, hepatic steatosis, is closely associated with lipotoxicity–induced mitochondrial abnormalities that sensitize the liver to additional pro–inflammatory insults. These second hits include enhanced lipid peroxidation and increased generation of reactive oxygen species (ROS)[10]. Inflammasomes are cytoplasmic multi–protein complexes composed of one of several NLR and PYHIN proteins, including NLRP1, NLRP3, NLRC4 and AIM2. Inflammasomes are sensors of endogenous or exogenous pathogen–associated molecular patterns (PAMPs) or damage–associated molecular patterns (DAMPs)[11] that govern cleavage of effector pro–inflammatory cytokines such as pro–IL–1β and pro–IL–18 (refs 12, 13). Most DAMPs trigger the generation of ROS, which are known to activate the NLRP3 inflammasome[14]. Therefore, we propose that inflammasome–dependent processing of IL–1β and IL–18 may have an important role in the progression of NAFLD.

Results

Feeding adult mice a methionine–choline–deficient diet (MCDD) for 4 weeks beginning at 8 weeks of age induces several features of human NASH, including hepatic steatosis, inflammatory cell infiltration and ultimately fibrosis[15]. To investigate the role of inflammasomes in NASH progression, we fed MCDD to C57Bl/6 wild type (NCI), apoptosis–associated speck–like protein containing a CARD ($Asc^{-/-}$, also known as $Pycard$) and caspase 1 ($Casp1^{-/-}$) mutant mice to induce early liver damage in the absence of fibrosis (Fig. 1a–d and Supplementary Fig. 1c). Compared to wild–type animals, age– and gender–matched $Asc^{-/-}$ and $Casp1^{-/-}$ mice that were fed MCDD were characterized by significantly higher serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST) activity, by enhanced microvesicular and macrovesicular hepatic steatosis, and by accumulation of multiple immune subsets in the liver from the innate and adaptive arms of the immune system (as defined by pathological examination and flow cytometry; $n = 7–11$ mice per group; Fig. 1a–d and Supplementary Figs 1c, 2a). Remarkably, the hepatic accumulation of T and B cells seems to be dispensable for this phenotype because $Asc^{-/-}$ mice lacking adaptive immune cells ($Asc^{-/-}$; $Rag^{-/-}$) also showed more severe NASH compared to wild–type animals, and comparable degrees of pathology to $Asc^{-/-}$ animals (Supplementary Fig. 2b–d).

To test whether the increased NASH observed in Asc– and $Casp1$–deficient mice was mediated by IL–1β or IL–18, we performed similar experiments using mice deficient in either the IL–1 receptor ($Il1r^{-/-}$)
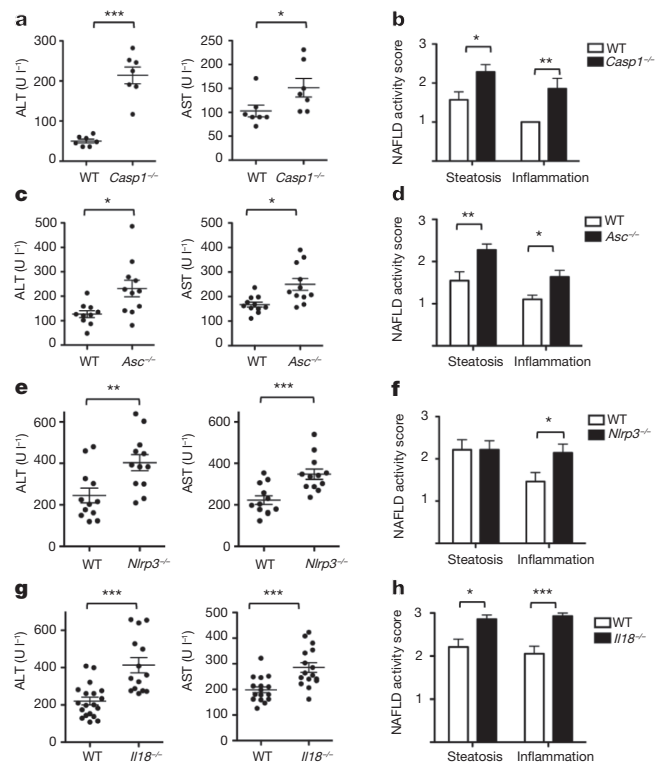
Figure 1 | Increased severity of NASH in inflammasome-deficient mice. To induce NASH, mice were fed with MCDD for 24 days. Their serum ALT and AST activities were measured and NAFLD histological activity scores were determined. **a–h**, Comparison of ALT, AST and NAFLD activity, plus histological scores for steatosis and inflammation between singly housed wild-type (WT) mice and $Casp1^{-/-}$ (**a, b**), $Asc^{-/-}$ (**c, d**), $Nlrp3^{-/-}$ (**e, f**), or $Il18^{-/-}$ (**g, h**). Data represent two independent experiments ($n = 7$–19 mice per treatment group). Error bars represent the s.e.m. of samples within a group. $*P \le 0.05$, $**P \le 0.01$, $***P \le 0.001$ (Student's t-test).

or IL-18 ($Il18^{-/-}$). $Il1r^{-/-}$ mice did not show any changes in the severity of NASH when compared to wild-type mice when fed MCDD (Supplementary Fig. 1a, b). In contrast to, but similar to $Asc^{-/-}$ and $Casp1^{-/-}$ mice, MCDD-fed $Il18^{-/-}$ animals featured a significant exacerbation of NASH severity (Fig. 1g, h and Supplementary Fig. 1c).

To assess the role of the NLRP3 inflammasome in NASH progression, we fed singly housed $Nlrp3^{-/-}$ and wild-type animals MCDD for 24 days and evaluated disease progression. $Nlrp3^{-/-}$ mice developed exacerbated NASH compared to wild-type mice as judged by increased levels of serum ALT and AST, plus NAFLD activity inflammation scores (Fig. 1e, f and Supplementary Fig. 1c). Remarkably, bone marrow chimaeric mice in which NLRP3 and ASC deficiency was limited to the haematopoietic compartment did not show any increase in the severity of NASH when compared to wild-type mice reconstituted with wild-type bone marrow (Supplementary Fig. 3a–f). Likewise, knock-in mice that specifically express a constitutively active NLRP3 inflammasome in CD11c+ myeloid cells ($Nlrp3KI$; CD11c+-Cre) or hepatocytes ($Nlrp3KI$; albumin-Cre)[16] did not feature any significant differences in MCDD-induced NASH severity as compared to wild-type mice (Supplementary Fig. 3g–l). These results indicate that aberrations in inflammasome function in cells other than hepatocytes or myeloid cells are key determinants of the enhanced disease progression in inflammasome-deficient mice.

We recently discovered that inflammasomes act as steady-state sensors and regulators of the colonic microbiota, and that a deficiency in components of two inflammasomes, NLRP6 (ref. 17) and NLRP3 (unpublished), both of which include ASC and caspase 1, and involve

IL-18 but not IL-1R, results in the development of an altered transmissible, colitogenic intestinal microbial community[17]. This microbiota is associated with increased representation of members of Bacteroidetes (Prevotellaceae) and the bacterial phylum TM7, and reductions in representation of members of the genus *Lactobacillus* in the Firmicutes phylum[17]. Moreover, electron microscopy studies disclosed aberrant colonization of crypts of Lieberkühn with bacteria with morphologic features of Prevotellaceae[17]. Therefore, we sought to investigate whether enhanced NASH severity in inflammasome-deficient mice is driven by their altered microbiota. Strikingly, co-housing of $Asc^{-/-}$ and $Il18^{-/-}$ mice with wild-type animals for 4 weeks (beginning at 4–6 weeks of age), before induction of NASH with MCDD resulted in significant exacerbation of NASH in the wild-type cage-mates (which we will refer to as WT($Asc^{-/-}$) and WT($Il18^{-/-}$), respectively, in the following text), as compared to singly housed, age- and gender-matched wild-type controls ($n = 5$–7 mice per genotype per housing condition). In co-housed wild-type mice, disease severity reached comparable levels to that of co-housed $Asc^{-/-}$ and $Il18^{-/-}$ mice (Fig. 2a–h). Moreover, significantly increased numbers of multiple inflammatory cell types were present in the liver of WT($Asc^{-/-}$) compared to wild-type mice (Supplementary Fig. 2a). Similar findings were observed in wild-type mice co-housed with $Casp1^{-/-}$, $Nlrp3^{-/-}$ and $Nlrp6^{-/-}$ mice (Supplementary Fig. 4a–f). To exclude the possibility that aberrant microbiota presented in all mice maintained in our vivarium, we co-housed wild-type mice with other strains of NLR-deficient mice that were either obtained from the same source as $Asc^{-/-}$ and $Nlrp3^{-/-}$ mice ($Nlrc4^{-/-}$, $Nlrp12^{-/-}$), or generated in our laboratory ($Nlrp4c^{-/-}$). None of these strains featured a similar phenotype (Supplementary Fig. 4g–l). These results indicate that the transmissible colitogenic microbiota present in inflammasome-deficient mice is a major contributor to their enhanced NASH. In agreement with this, combined antibiotic treatment with ciprofloxacin and metronidazole, previously shown to abrogate the colitogenic activity
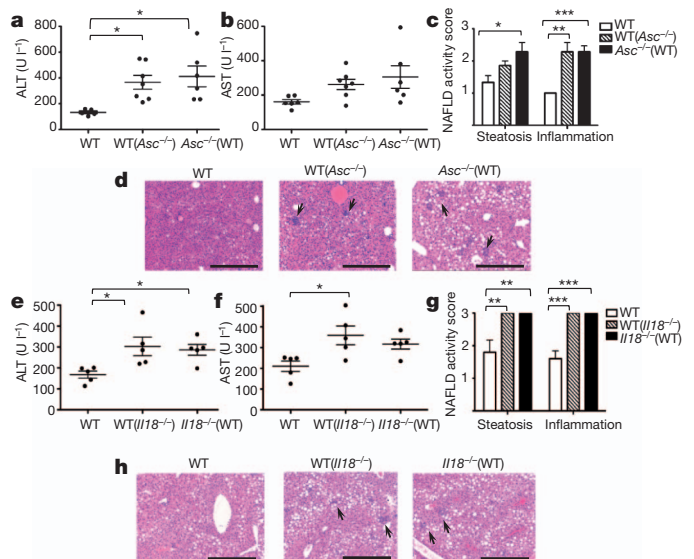

Figure 2 | Increased severity of NASH in Asc- and $Il18$-deficient mice is transmissible to co-housed wild-type animals. $Asc^{-/-}$ or $Il18^{-/-}$ mice and wild-type mice were co-housed for 4 weeks and then fed MCDD. **a–d**, ALT (**a**), AST (**b**), NAFLD activity scores (**c**), and haematoxylin and eosin-stained sections of livers (**d**) of singly housed wild-type mice (WT), wild-type mice co-housed with $Asc^{-/-}$ mice (WT($Asc^{-/-}$)), and $Asc^{-/-}$ mice co-housed with wild-type mice ($Asc^{-/-}$(WT)). **e–h**, ALT (**e**), AST (**f**), NAFLD activity histological scores (**g**), and haematoxylin and eosin-stained sections of livers (**h**) of wild-type, WT($Il18^{-/-}$) and $Il18^{-/-}$(WT). Data are representative of two independent experiments. Error bars represent s.e.m. Scale bars, 200 μm (**d, h**). $*P \le 0.05$, $**P \le 0.01$, $***P \le 0.001$.

of the microbiota associated with inflammasome-deficient mice associated microbiota[17], significantly reduced the severity of NASH in $Asc^{-/-}$ mice, and abolished transmission of the phenotype to WT($Asc^{-/-}$) animals (Supplementary Fig. 5).

To ascertain the effects of MCDD on the gut microbiota, we performed a culture-independent analysis of amplicons generated by primers directed against variable region 2 of bacterial 16S ribosomal RNA genes of faecal samples collected from wild-type mice co-housed with $Asc^{-/-}$ animals (WT($Asc^{-/-}$)), their $Asc^{-/-}$ cage-mates ($Asc^{-/-}$(WT)) as well as singly housed wild-type controls 1 day and 12 days before, and 7, 14 and 19 days after initiation of this diet ($n = 20$ animals; 8 singly housed wild-type, 6 co-housed wild-type and 6 $Asc^{-/-}$ mice). The structures of bacterial communities were compared based on their phylogenetic content using unweighted UniFrac. The results are illustrated in Fig. 3. Supplementary Table 1 provides a list of all phylotypes that, based on criteria outlined in Methods, discriminate co-housed WT($Asc^{-/-}$) from their singly housed wild-type counterparts. Prior to MCDD, and consistent with our previous findings[17], the faecal microbiota of WT($Asc^{-/-}$) mice adopted a configuration similar to $Asc^{-/-}$ cage-mates, including the appearance of Prevotellaceae (Supplementary Table 1 and Fig. 3 a–c). There was also a significant increase in proportional representation of members of the family Porphyromonadaceae (primarily in the genus *Parabacteroides*) in WT($Asc^{-/-}$) mice compared to their singly

housed wild-type counterparts (Fig. 3d,e). The representation of Porphyromonadaceae was greatly increased in both the co-housed wild-type and $Asc^{-/-}$ mice (but not in singly housed wild-type) when they were switched to a MCDD diet ($P < 0.01$; t-test; Fig. 3d). A dramatic increase in the family Erysipelotrichaceae (phylum Firmicutes) also occurred with MCDD in both singly and co-housed WT animals, to a level that was >10% of the community (Fig. 3f). Although the Prevotellaceae decreased when co-housed WT($Asc^{-/-}$) mice were placed on MCDD, their relative abundance remained significantly higher than in singly housed wild-type animals (Fig. 3c).

Together, these results pointed to the possibility that members of the altered intestinal microbiota in inflammasome-deficient MCDD-treated mice may promote a signalling cascade in the liver upon translocation, resulting in progression to NASH in susceptible animals. Toll-like receptors (TLR) have a major role in NAFLD pathophysiology due to the liver's exposure to relatively large amounts of PAMPs derived from the intestine and delivered via the portal circulation[18–20]. Therefore, we propose that TLR signalling mediates the increased susceptibility to progression to NASH in mice exposed to the gut microbiota of $Asc^{-/-}$ animals. $Myd88^{-/-}$;$Trif^{-/-}$ mice are devoid of all TLR signalling pathways. When co-housed with $Asc^{-/-}$ ($Myd88^{-/-}$;$Trif^{-/-}$($Asc^{-/-}$))mice between 5 and 9 weeks of age, they showed decreased severity of NASH after exposure to MCDD for 24 days, compared to WT($Asc^{-/-}$) mice (Supplementary Fig. 6a, b).
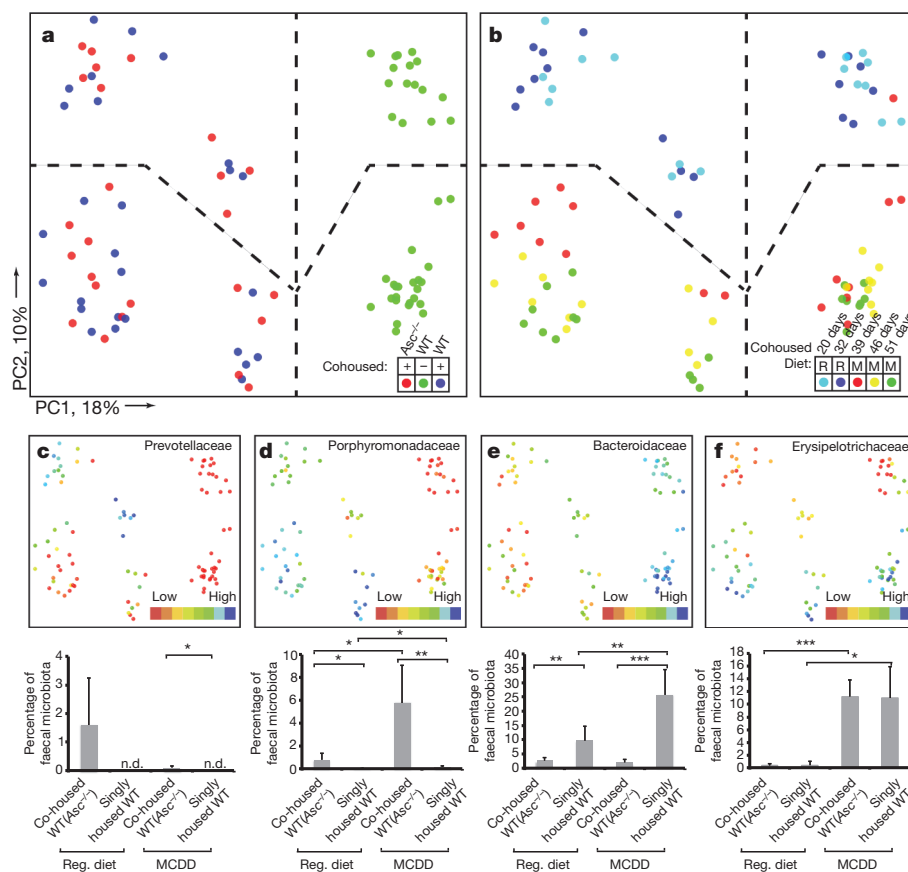

Figure 3 | 16S rRNA sequencing demonstrates diet and co-housing associated changes in gut microbial ecology. a, Principal coordinates analysis (PCoA) of unweighted UniFrac distances of 16S rRNA sequences demonstrating clustering according to co-housing status on principal coordinate 1 (PC1). **b**, PCoA of same plot as in **a** coloured for experimental day. Mice were co-housed and fed a regular diet (R) for the first 32 days of the experiment (two time points taken at day 20 and 32) before being switched to MCDD (M, sampled at days 39, 46 and 51 of the experiment). **c–f**, PCoA and bar graphs of family level taxa Prevotellaceae, Porphyromonadaceae, Bacteroidaceae and Erysipelotrichaceae demonstrating diet- and microbiota-

dependent differences in taxonomic representation. PCoA plots contain spheres representing a single faecal community coloured according to relative representation of the taxon (blue represents relatively higher levels; red indicates lower levels). Bar graphs represent averaged taxonomic representation for singly or co-housed mouse while on either regular or MCD diet ($n = 8$ for singly housed wild-type, $n = 12$ co-housed $Asc^{-/-}$(WT) and WT($Asc^{-/-}$) animals). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ by t-test after Bonferroni correction for multiple hypotheses. n.d., not detected; Reg. diet, regular diet.

To define which specific TLRs were responsible for the inflammatory response, we co-housed $Tlr4$-, $Tlr9$- or $Tlr5$-deficient mice with $Asc^{-/-}$ animals and induced NASH with MCDD as previously described. Similar to wild-type mice, $Tlr5^{-/-}$ mice co-housed with $Asc^{-/-}$ mice ($Tlr5^{-/-}(Asc^{-/-})$) featured a statistically significant exacerbation of hepatic injury, steatosis and inflammation, when compared to singly housed $Tlr5^{-/-}$ controls (Fig. 4c and Supplementary Fig. 6g, h), indicating that TLR5 does not mediate the microbiota-mediated exacerbation in disease severity. In contrast, $Tlr4^{-/-}(Asc^{-/-})$ and $Tlr9^{-/-}(Asc^{-/-})$ mice did not show the customary increase in disease severity when compared to their singly housed $Tlr4^{-/-}$ and $Tlr9^{-/-}$ counterparts (Fig. 4a, b and Supplementary Fig. 6c–f).

These observations indicate that intact bacteria or bacterial products derived from the intestine trigger TLR4 and TLR9 activation, which results in an increased rate of disease progression in mice that house a colitogenic gut microbiota associated with inflammasome deficiency (that is, $Asc^{-/-}$ and WT($Asc^{-/-}$) mice). Efforts to sequence 16S rRNA genes that might be present in total liver DNA, microbial quantitative PCR assays of portal vein blood DNA, histologic analysis of intact

liver, and aerobic and anaerobic cultures of liver homogenates did not reveal any evidence of intact bacteria in wild-type or $Asc^{-/-}$ mice fed MCDD (data not shown). Notably, transmission electron microscopy studies of colon collected from wild-type and $Asc^{-/-}$ mice revealed an abundance of electron-dense material, suggestive of some black-pigmented bacterial species, in colonic epithelial cells and macrophages located in the lamina propria of $Asc^{-/-}$ mice but not in wild-type animals (Fig. 4e and Supplementary Fig. 7c). In agreement with previous results, we did not detect any translocation of intact bacteria (Fig. 4e and Supplementary Fig. 7c).

These observations provide evidence for the uptake of bacterial products from locally invasive gut microbes in $Asc^{-/-}$ mice (Fig. 4e and Supplementary Fig. 7c). If microbial components, rather than whole organisms, were transmitted to the liver then they should be detectable in the portal circulation. Indeed, levels of TLR4 and TLR9 agonists, but not TLR2 agonists (assayed by their ability to activate TLR reporter cell lines), were markedly increased in the portal circulation of MCDD-fed WT($Asc^{-/-}$), and $Asc^{-/-}$ mice compared to wild-type controls ($n = 13–28$ mice per group; Fig. 4d and Supplementary Fig. 7a, b). Altogether, these results indicate a mechanism whereby TLR4 and TLR9 agonist efflux from the intestines of inflammasome-deficient mice or their co-housed partners, through the portal circulation, to the liver where they trigger TLR4 and TLR9 activation that in turn results in enhanced progression of NASH.

We next explored the downstream mechanism whereby microbiota-induced TLR signalling enhances NASH progression. Pro-inflammatory cytokines, and in particular TNF-α, a downstream cytokine of TLR signalling, are known to contribute to progression of hepatic steatosis to steatohepatitis and eventually hepatic fibrosis in a number of animal models and in human patients[21,22]. Following induction of NASH by MCDD, hepatic Tnf mRNA expression was significantly upregulated in $Asc^{-/-}$ and $Il18^{-/-}$ mice, which show exacerbated disease, but not in $Il1r^{-/-}$ mice, which do not (Supplementary Fig. 8a–c). Moreover, Tnf mRNA levels were significantly increased in wild-type mice that had been previously co-housed with $Asc^{-/-}$ or $Il18^{-/-}$ mice and then fed MCDD (Supplementary Fig. 8d, e), indicating that its enhanced expression was mediated by elements of the microbiota responsible for NASH exacerbation. In contrast, we did not observe any changes in $Il6$ or $Il1b$ mRNA levels in the livers of $Asc^{-/-}$, $Il18^{-/-}$ or $Il1r^{-/-}$ mice compared to wild-type controls (Supplementary Fig. 8a–c). Furthermore, whereas MCDD-administered singly housed $Tnf^{-/-}$ mice had comparable NASH severity to singly housed wild-type animals (Fig. 4f–h and Supplementary Fig. 8f), co-housing with Asc-deficient mice before MCDD induction of NASH resulted in increased liver injury, hepatic steatosis and inflammation in wild-type mice but not in $Tnf^{-/-}$ mice (Fig. 4f–h and Supplementary Fig. 8f). These results indicate that TNF-α mediates the hepatotoxic effects downstream of the transmissible gut microbiota present in $Asc^{-/-}$ mice.

The aberrant gut microbiota in NLRP3 and NLRP6 inflammasome-deficient mice induces colonic inflammation through epithelial induction of CCL5 secretion[17]. To test whether this colon inflammation influences TLR agonist influx into the portal circulation and NASH progression, we induced NASH in wild-type and $Ccl5^{-/-}$ mice that had been either singly housed or co-housed. MCDD-fed, singly housed wild-type and $Ccl5^{-/-}$ mice showed equivalent levels of NASH severity (Supplementary Fig. 9a–c), indicating that CCL5 does not have a role in the early stages of NAFLD/NASH in the absence of the inflammasome-associated colitogenic microbiota. However, we documented significantly increased levels of liver injury, inflammation and steatosis in WT($Asc^{-/-}$) but not $Ccl5^{-/-}(Asc^{-/-})$ mice (Fig. 5a–c), which led us to conclude that CCL5 is required for the exacerbation of disease through cohousing with inflammasome-deficient mice. Moreover, $Ccl5^{-/-}(Asc^{-/-})$ animals showed significantly reduced levels of TLR4 and TLR9 agonists in their portal vein blood than WT($Asc^{-/-}$) mice (Supplementary Fig. 9d–f). Together, these results
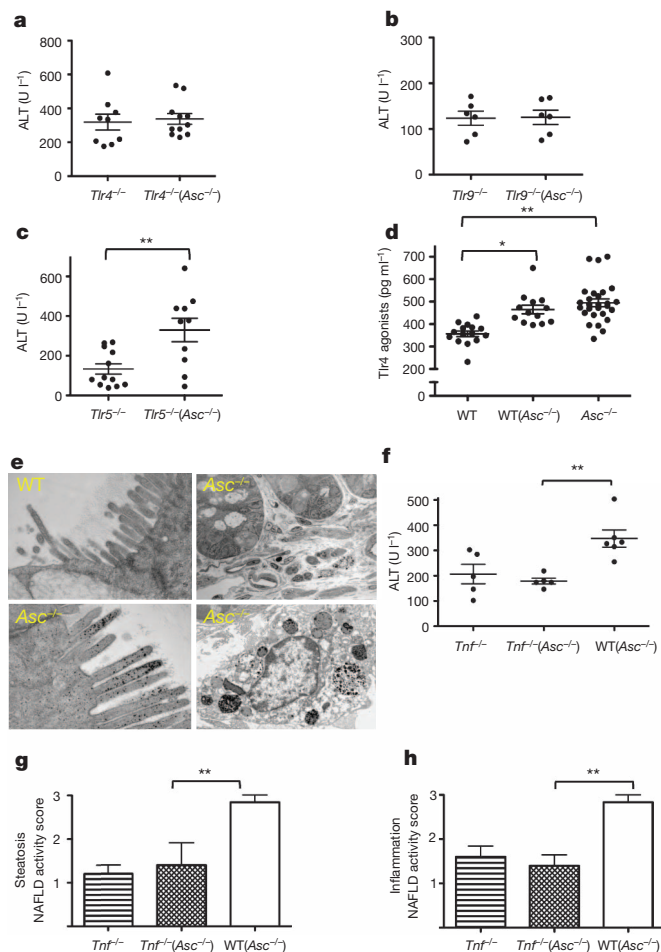

Figure 4 | Increased severity of NASH in Asc-deficient and co-housed wild-type animals is mediated by TLR4, TLR9 and TNF-α. $Asc^{-/-}$ mice were co-housed with wild-type, $Tnf^{-/-}$, $Tlr4^{-/-}$, $Tlr9^{-/-}$ or $Tlr5^{-/-}$ mice for 4 weeks and then fed MCDD. **a–c**, ALT levels of $Tlr4^{-/-}(Asc^{-/-})$ (**a**), $Tlr9^{-/-}(Asc^{-/-})$ (**b**), and $Tlr5^{-/-}(Asc^{-/-})$ mice (**c**) and their singly housed counterparts. **d**, TLR4 agonists in portal vein sera from MCDD-fed wild-type, WT($Asc^{-/-}$) and $Asc^{-/-}$ animals. **e**, Transmission electron microscopy images of colon from wild-type and $Asc^{-/-}$ mice. **f–h**, ALT (**f**) and NAFLD (**g–h**) activity histological scores of $Tnf^{-/-}$, WT($Asc^{-/-}$) and $Tnf^{-/-}(Asc^{-/-})$ mice. Data are representative of two independent experiments. Error bars represent s.e.m. *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$.

indicate that microbiota-induced subclinical colon inflammation is a determining factor in the rate of TLR agonist influx from the gut, and in NAFLD/NASH progression.

The MCDD system is a common model for studying inflammatory processes associated with progression from NAFLD to NASH, yet it lacks many of the associated metabolic phenotypes of NAFLD, such as obesity and insulin resistance[23]. As such, our results in this model might conceivably be limited to the way dysbiosis can influence NASH progression in patients with enhanced intestinal permeability, such as those with inflammatory bowel disease[24], but not for the majority of patients who suffer from NASH in the context of metabolic syndrome. To test whether alterations in the gut microbiota of inflammasome-deficient mice may affect the rate of progression of NAFLD and other features associated with metabolic syndrome, we extended our studies to genetically obese mice and mice fed with high-fat diet (HFD).

Leptin-receptor deficient (db/db; db is also known as Lepr) animals develop multiple metabolic abnormalities, including NAFLD and impaired intestinal barrier function[25], that closely resemble the human disease[26]. However, significant hepatocyte injury, inflammation, and fibrosis are not observed in the absence of a "second hit"[27]. Upon co-housing of db/db mice with $Asc^{-/-}$ ($db/db(Asc^{-/-})$) or WT mice (db/db(WT)) for a period of 12 weeks, and as previously shown for $Asc^{-/-}$ mice[17], the colon and ileum of all $db/db(Asc^{-/-})$ mice showed mild to moderate mucosal and crypt hyperplasia (Fig. 5d–f) that was not seen in db/db(WT) mice.

Strikingly, co-housed $db/db(Asc^{-/-})$ mice also showed increased levels of hepatocyte injury as evidenced by higher levels of ALT and AST in their sera, and significantly exacerbated steatosis and hepatic inflammation scores when compared with db/db(WT) mice (Fig. 5g–i). In addition to a parenchymal inflammatory exudate, patchy areas of markedly degenerated hepatocytes and hepatocytes undergoing necrosis were observed, but only in $db/db(Asc^{-/-})$ animals (Fig. 5f). Furthermore, some areas of congestion were seen in the centro-lobular zone as well as in the hepatic parenchyma — features that resemble peliosis hepatis, a condition observed in a variety of pathological settings including infection (data not shown). In accord with our MCDD results, hepatic Tnf mRNA levels were significantly higher in co-housed $db/db(Asc^{-/-})$ mice than in db/db(WT) animals (Fig. 5j). Again, no significant differences were observed in hepatic Il6 or Il1b mRNA levels (Fig. 5j).

Interestingly, $db/db(Asc^{-/-})$ mice developed significantly more weight gain compared to db/db(WT) mice after 12 weeks of co-housing (Fig. 6a), indicating that the inflammasome-associated gut microbiota could exacerbate additional processes associated with the metabolic syndrome, such as obesity. To address this possibility, we monitored multiple metabolic parameters in wild-type, WT($Asc^{-/-}$) and $Asc^{-/-}$ mice fed a high-fat diet (HFD) for 12 weeks. Strikingly, $Asc^{-/-}$ mice gained body mass more rapidly and featured enhanced hepatic steatosis (Fig. 6b, c and Supplementary Fig. 11f). $Asc^{-/-}$ mice also showed elevated fasting plasma glucose and insulin levels, and decreased glucose tolerance compared to singly housed weight-matched wild-type mice (Fig. 6d–f). Interestingly, WT($Asc^{-/-}$) mice recapitulated
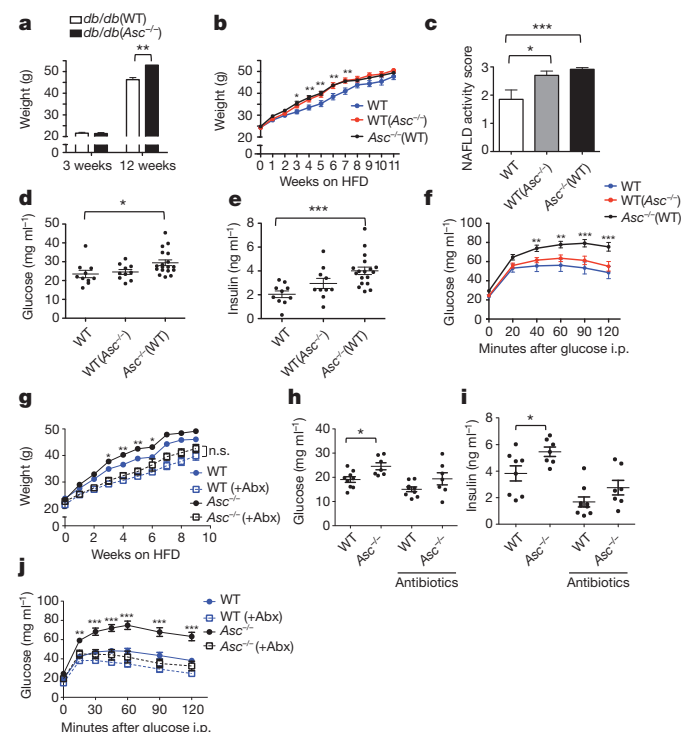

Figure 5 | Increased severity of NASH in *Asc*-deficient mice is transmissible to *db/db* by co-housing and is mediated by CCL5-induced intestinal inflammation. a–c, ALT (a), AST (b) and NAFLD (c) activity histological scores of WT($Asc^{-/-}$) and $Ccl5^{-/-}(Asc^{-/-})$ mice. Data represents two independent experiments. d–j, db/db mice were co-housed with wild-type or $Asc^{-/-}$ mice for 12 weeks. d–f, Representative haematoxylin and eosin-stained sections of colon (d), terminal ileum (e) and liver (f) from db/db(WT) and db/db($Asc^{-/-}$) mice fed a standard chow diet. Mucosal and crypt hyperplasia (arrow). Hepatocyte degeneration (arrowhead). Scale bars, 500 μm (d–e), 200 μm (f). g–i, ALT (g), AST (h) and NAFLD (i) activity scores of db/db(WT) and db/db($Asc^{-/-}$) mice. j, Hepatic Tnf, Il6 and Il1b mRNA levels. Error bars represent s.e.m. *$P \le 0.05$, **$P \le 0.01$, ***$P \le 0.001$.

Figure 6 | *Asc*-deficient mice develop increased obesity and loss of glycaemic control on HFD. a, Weight of db/db(WT) or db/db($Asc^{-/-}$) mice at 3 weeks of age and at 12 weeks of co-housing. b–f, $Asc^{-/-}$ and wild-type mice were co-housed for 4 weeks and then fed HFD. b, Body weights. c, NAFLD histological activity score. d, e, Fasting plasma glucose and insulin after 11 weeks of HFD. f, Intraperitoneal (i.p.) glucose tolerance test after 12 weeks of HFD. g–j, Mice were untreated, or treated orally with antibiotics (Abx), for 3 weeks before HFD feeding for 12 weeks. g, Body weights. h, i, Fasting plasma glucose and insulin levels after 8 weeks on a HFD. j, Intraperitoneal glucose tolerance test after 10 weeks of HFD. Error bars represent s.e.m. *$P \le 0.05$, **$P \le 0.01$, ***$P \le 0.001$.

the same increased rate of body mass gain and steatosis when compared to singly housed wild-type controls, although they did not show significant alterations in glucose homeostasis (Fig. 6d–f). Nevertheless, antibiotic treatment (ciprofloxacin and metronidazole) abrogated all these abnormalities, including altered rate of gain in body mass, glucose intolerance and fasting plasma insulin levels in $Asc^{-/-}$ mice compared to wild-type mice (Fig. 6g–j). Alterations of these metabolic parameters were not caused by changes in feeding behaviour between the antibiotic-treated and untreated groups (data not shown). These results indicate different levels of microbiota-mediated regulation of the various manifestations of the metabolic syndrome: that is, some features (obesity, steatosis) are pronounced and transmissible by co-housing, whereas others (glycaemic control) are affected by alterations in the microbiota but not readily transferable by co-housing. Additionally, we performed a 16S rRNA-based analysis of the faecal microbiota of $Asc^{-/-}$ and wild-type animals that were treated with or without ciprofloxacin and metronidazole (4 weeks) before switching to HFD for 4 additional weeks. Importantly, the analysis demonstrated that Prevotellaceae and Porphyromonadaceae, two family-level taxa, were undetectable in $Asc^{-/-}$ mice 8 weeks after antibiotic treatment (Supplementary Fig. 12a–c; Supplementary Table 2).

To assess whether these metabolic abnormalities are specific to $Asc^{-/-}$ mice, we performed similar experiments with $Nlrc4^{-/-}$ mice. These mice showed an equal rate of body mass gain, and similar glucose tolerance phenotypes as singly housed wild-type mice, confirming the specificity of the phenotype (Supplementary Fig. 10a–d). 16S rRNA analysis revealed that there was an increased representation of Porphyromonadaceae in $Nlrc4^{-/-}$ mice when compared to wild-type mice (Supplementary Table 3). These results indicate that (1) some metabolic aberrations associated with the dysbiosis of inflammasome-deficient mice can be horizontally transferred from one mouse to another, (2) the gut microbiota of inflammasome-deficient mice has a negative effect on NAFLD progression and glucose homeostasis, and (3) configurational changes in the microbiota, which involve overrepresentation Porphyromonadaceae in combination with alterations in additional taxa, are likely required to produce these host phenotypes.

Discussion

The results presented here provide evidence that modulation of the intestinal microbiota through multiple inflammasome components is a critical determinant of NAFLD/NASH progression as well as multiple other aspects of metabolic syndrome such as weight gain and glucose homeostasis. Our results demonstrate a complex and cooperative effect of two sensing protein families, namely NLRs and TLRs, in shaping metabolic events. In the gut, the combination of host-related factors such as genetic inflammasome deficiency-associated dysbiosis result in abnormal accumulation of bacterial products in the portal circulation. The liver, being a 'first pass' organ and thus exposed to the highest concentration of portal system products such as PAMPs, is expected to be most vulnerable to their effects, particularly when pre-conditioned by sub-clinical pathology such as lipid accumulation in hepatocytes. Indeed in our models, accumulation of TLR agonists was sufficient to drive progression of NAFLD/NASH even in genetically intact animals.

This 'gut–liver axis', driven by alterations in gut microbial ecology, may offer an explanation for a number of long-standing, albeit poorly understood, clinical associations. One example is the occurrence of primary sclerosing cholangitis (PSC) in patients with inflammatory bowel disease, particularly those with inflammation along the length of the colon. Coeliac disease, another inflammatory disorder with increased intestinal permeability, is associated with a variety of liver disorders, ranging from asymptomatic transaminasaemia, NAFLD, to primary biliary cirrhosis (PBC). In fully developed cirrhosis, complications associated with high mortality such as portal hypertension, variceal bleeding, spontaneous bacterial peritonitis and encephalopathy

are triggered by translocation of bacteria or bacterial components, providing another important example of the importance of the interplay between the microbiome, the immune response and liver pathology[28].

Recent reports suggest a complex role of inflammasome function in multiple manifestations of the metabolic syndrome. Activation of IL-1β, mainly through cleavage by the NLRP3 inflammasome, promotes insulin resistance[29,30], atherosclerotic plaque formation[31], and β cell death[32,33]. Moreover, caspase-1 activation seems to direct adipocytes towards a more insulin-resistant phenotype[34]. Conversely, $Il18$-deficient mice are prone to develop obesity, hyperphagia and insulin resistance[35]. These discrepancies most probably reflect a hierarchical contribution of multiple inflammasome components in different metabolic processes, tissues and mouse models. In agreement with previous studies, we found increased obesity and insulin resistance in $Il18$-deficient mice fed with a HFD (data not shown). However, and in contrast to two previous reports[30,34], we showed that $Asc^{-/-}$ mice are prone to obesity induction and hepatosteatosis, as well as impaired glucose homeostasis when fed a HFD. We propose that alterations in intestinal microbiota communities associated with multiple inflammasome deficiencies could account for these discrepancies and it should be added to the list of major environmental/host factors affecting manifestations and progression of metabolic syndrome in susceptible populations.

In the inflammasome-deficient setting, a significant expansion of Porphyromonadaceae was found following administration of MCDD and HFD, which was abolished by antibiotic treatment. Interestingly, one member of the family, $Porphyromonas$, has been associated with several components of the metabolic syndrome in both mice and humans, including atherosclerosis and diabetes mellitus[36,37]. Moreover, expansion of this taxa is strongly associated with complications of chronic liver disease[36]. More work is needed to further delineate the relevance of the suggested taxa discovered in our work to the pathogenesis and progression of human NAFLD/NASH and other features of the metabolic syndrome. Elucidation of similar or distinct mechanisms to the ones presented here, possibly linking Porphyromonadaceae expansion to a propensity for development of the metabolic syndrome, would be of importance to the field.

METHODS SUMMARY

Six- to eight-week-old male mice were fed a methionine-choline-deficient diet for 24 days. Eight- to ten-week-old male mice were fed a HFD ad libitum. This diet consists of 60% calories from fat and was administered for 10–12 weeks. Standard histology of liver, terminal ileum and colon were described previously[17]. The presence of immune cells in liver tissue was analysed by flow cytometry on livers digested with $0.5\,mg\,ml^{-1}$ collagenase. Glucose tolerance test were performed after 10–12 weeks of consuming the HFD and mice were fasted overnight ($\sim 14\,h$), and injected intraperitoneally with D-glucose. Transmission electron microscopy was performed as previously described[17]. Data are expressed as mean ± s.e.m. Differences were analysed by Student's t-test or ANOVA and post hoc analysis for multiple group comparison. P values ≤ 0.05 were considered significant.

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

1. Sheth, S. G., Gordon, F. D. & Chopra, S. Nonalcoholic steatohepatitis. *Ann. Intern. Med.* **126**, 137–145 (1997).
2. Ludwig, J., Viggiano, T. R., McGill, D. B. & Oh, B. J. Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. *Mayo Clin. Proc.* **55**, 434–438 (1980).
3. Marchesini, G. et al. Nonalcoholic fatty liver, steatohepatitis, and the metabolic syndrome. *Hepatology* **37**, 917–923 (2003).
4. Caldwell, S. H. et al. Cryptogenic cirrhosis: clinical characterization and risk factors for underlying disease. *Hepatology* **29**, 664–669 (1999).
5. Shimada, M. et al. Hepatocellular carcinoma in patients with non-alcoholic steatohepatitis. *J. Hepatol.* **37**, 154–160 (2002).
6. Propst, A., Propst, T., Judmaier, G. & Vogel, W. Prognosis in nonalcoholic steatohepatitis. *Gastroenterology* **108**, 1607 (1995).

7. Charlton, M. Cirrhosis and liver failure in nonalcoholic fatty liver disease: molehill or mountain? *Hepatology* **47,** 1431–1433 (2008).
8. Hjelkrem, M. C., Torres, D. M. & Harrison, S. A. Nonalcoholic fatty liver disease. *Minerva Med.* **99,** 583–593 (2008).
9. Day, C. P. & James, O. F. Steatohepatitis: a tale of two "hits"? *Gastroenterology* **114,** 842–845 (1998).
10. Sanyal, A. J. *et al.* Nonalcoholic steatohepatitis: association of insulin resistance and mitochondrial abnormalities. *Gastroenterology* **120,** 1183–1192 (2001).
11. Sutterwala, F. S., Ogura, Y. & Flavell, R. A. The inflammasome in pathogen recognition and inflammation. *J. Leukoc. Biol.* **82,** 259–264 (2007).
12. Martinon, F., Burns, K. & Tschopp, J. The inflammasome. *Mol. Cell* **10,** 417–426 (2002).
13. Agostini, L. *et al.* NALP3 forms an IL-1β-processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. *Immunity* **20,** 319–325 (2004).
14. Zhou, R., Yazdi, A. S., Menu, P. & Tschopp, J. A role for mitochondria in NLRP3 inflammasome activation. *Nature* **469,** 221–225 (2011).
15. Varela-Rey, M. *et al.* Non-alcoholic steatohepatitis and animal models: understanding the human disease. *Int. J. Biochem. Cell Biol.* **41,** 969–976 (2009).
16. Brydges, S. D. *et al.* Inflammasome-mediated disease animal models reveal roles for innate but not adaptive immunity. *Immunity* **30,** 875–887 (2009).
17. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145,** 745–757 (2011).
18. Rivera, C. A. *et al.* Toll-like receptor-4 signaling and Kupffer cells play pivotal roles in the pathogenesis of non-alcoholic steatohepatitis. *J. Hepatol.* **47,** 571–579 (2007).
19. Miura, K. *et al.* Toll-like receptor 9 promotes steatohepatitis by induction of interleukin-1β in mice. *Gastroenterology* **139,** 323–334 e7 (2010).
20. Seki, E. *et al.* TLR4 enhances TGF-β signaling and hepatic fibrosis. *Nature Med.* **13,** 1324–1332 (2007).
21. Crespo, J. *et al.* Gene expression of tumor necrosis factor α and TNF-receptors, p55 and p75, in nonalcoholic steatohepatitis patients. *Hepatology* **34,** 1158–1163 (2001).
22. Li, Z. *et al.* Probiotics and antibodies to TNF inhibit inflammatory activity and improve nonalcoholic fatty liver disease. *Hepatology* **37,** 343–350 (2003).
23. Diehl, A. M. Lessons from animal models of NASH. *Hepatol. Res.* **33,** 138–144 (2005).
24. Broomé, U., Glaumann, H. & Hultcrantz, R. Liver histology and follow up of 68 patients with ulcerative colitis and normal liver function tests. *Gut* **31,** 468–472 (1990).
25. Guo, X. *et al.* Leptin signaling in intestinal epithelium mediates resistance to enteric infection by *Entamoeba histolytica. Mucosal Immunol.* **4,** 294–303 (2011).
26. Ikejima, K. *et al.* The role of leptin in progression of non-alcoholic fatty liver disease. *Hepatol. Res.* **33,** 151–154 (2005).
27. Guebre-Xabier, M. *et al.* Altered hepatic lymphocyte subpopulations in obesity-related murine fatty livers: potential mechanism for sensitization to liver damage. *Hepatology* **31,** 633–640 (2000).
28. Almeida, J., Galhenage, S., Yu, J., Kurtovic, J. & Riordan, S. M. Gut flora and bacterial translocation in chronic liver disease. *World J. Gastroenterol.* **12,** 1493–1502 (2006).
29. Vandanmagsar, B. *et al.* The NLRP3 inflammasome instigates obesity-induced inflammation and insulin resistance. *Nature Med.* **17,** 179–188 (2011).
30. Wen, H. *et al.* Fatty acid-induced NLRP3-ASC inflammasome activation interferes with insulin signaling. *Nature Immunol.* **12,** 408–415 (2011).
31. Duewell, P. *et al.* NLRP3 inflammasomes are required for atherogenesis and activated by cholesterol crystals. *Nature* **464,** 1357–1361 (2010).
32. Zhou, R., Tardivel, A., Thorens, B., Choi, I. & Tschopp, J. Thioredoxin-interacting protein links oxidative stress to inflammasome activation. *Nature Immunol.* **11,** 136–140 (2010).
33. Masters, S. L. *et al.* Activation of the NLRP3 inflammasome by islet amyloid polypeptide provides a mechanism for enhanced IL-1β in type 2 diabetes. *Nature Immunol.* **11,** 897–904 (2010).
34. Stienstra, R. *et al.* Inflammasome is a central player in the induction of obesity and insulin resistance. *Proc. Natl Acad. Sci. USA* **108,** 15324–15329 (2011).
35. Netea, M. G. *et al.* Deficiency of interleukin-18 in mice leads to hyperphagia, obesity and insulin resistance. *Nature Med.* **12,** 650–656 (2006).
36. Bajaj, J. S. *et al.* Linkage of gut microbiome with cognition in hepatic encephalopathy. *Am. J. Physiol. Gastrointest. Liver Physiol.* **302,** 168–175 (2011).
37. Makiura, N. *et al.* Relationship of *Porphyromonas gingivalis* with glycemic level in patients with type 2 diabetes following periodontal treatment. *Oral Microbiol. Immunol.* **23,** 348–351 (2008).

METHODS

Mice. $Casp1^{-/-}$ (Casp1^{tm1Flv}) and $Nlrp4c^{-/-}$ mice were generated in our laboratory[38]. Production of $ASC^{-/-}$ (Pycard^{tm1Flv}), $Nlrp3^{-/-}$, $Nlrp6^{-/-}$, $Nlrc4^{-/-}$ and $Nlrp12^{-/-}$ mice is described elsewhere[17]. $Il18^{-/-}$ (Il18^{tm1Aki}), $Il1r^{-/-}$ (Il1r1^{tm1Imx}), $Tnf^{-/-}$ (Tnf^{tm1Gkl}), $Tlr4^{-/-}$ (Tlr4$^{lps-del}$), $Tlr5^{-/-}$ (Tlr5^{tm1Flv}), $Myd88^{-/-}$ (Myd88^{tm1Defr}), $Ccl5^{-/-}$ (Ccl5^{tm1Hso}), $Rag1^{-/-}$ (Rag1^{tm1Mom}), CD11c-Cre (Itgax-cre), albumin-Cre (Alb-cre), $Trif^{-/-}$ ($Ticam1^{Lps2}$) and db/db($Lepr^{db}$) mice were obtained from Jackson Laboratories. $Tlr9^{-/-}$ mice have been described in another report[39]. Production of $Nlrp3KI$ (A350V) mice is described elsewhere[16]. Wild-type C57Bl/6 mice were purchased from the NCI. For co-housing experiments, age-matched wild-type and KO mice at the age of 4–6 weeks were co-housed in sterilized cages for 4 or 12 weeks at a ratio of 1:1 (WT:KO), with unrestricted access to food and water. No more than 6 mice in total were housed per cage. For antibiotic treatment, mice were given a combination of ciprofloxacin ($0.2\,g\,l^{-1}$) and metronidazole ($1\,g\,l^{-1}$) for 4 weeks in the drinking water. All antibiotics were obtained from Sigma Aldrich. All experimental procedures were approved by the local IACUC.

NASH model. 6–8 week-old male mice were fed a methionine-choline-deficient diet (MP Biomedicals) for 24 days. Methionine-choline-sufficient control diet was the same but supplemented with choline chloride (2 g per kg of diet) and DL-methionine (3 g per kg of diet). Mice had unrestricted access to food and water.

High fat diet model. 8-10 week-old male mice were fed a HFD ad libitum. This diet consists of 60% calories from fat (D12492i; Research Diets) and was administered for 10–12 weeks.

Histology. The intact liver was excised immediately after mice were euthanized by asphyxiation, fixed in 10% neutral buffered formalin and embedded in paraffin. Liver sections were stained with haematoxylin and eosin, or trichrome. Histological examination was performed in a blinded fashion by an experienced gastrointestinal pathologist with the histological scoring system for NAFLD[40]. Briefly, steatosis and inflammation scores ranged from 0 to 3 with 0 being within normal limits and 3 being most severe. Individual scores were assigned for each parameter. The most severe area of hepatic inflammation of representative histology sections were photographed using an Olympus microscope.

Colons were fixed in Bouin's medium and embedded in paraffin. Blocks were serially sectioned along the cephalocaudal axis of the gut to the level of the lumen; 5-μm-thick sections were stained with haematoxylin and eosin. Digital light microscopic images were recorded with a Zeiss Axio Imager.A1 microscope, AxioCam MRc5 camera and AxioVision 4.7.1 imaging software (Carl Zeiss Microimaging). Further details in ref. 17.

Gene expression analysis. Tissues were preserved in RNAlater solution (Ambion), and subsequently homogenized in TRIzol reagent (Invitrogen). RNA (1 μg) was used to generate complementary DNA using the HighCapacity cDNA Reverse Transcription kit (Applied Biosystems). Real time PCR was performed using gene-specific primer/probe sets (Applied Biosystems) and Kapa Probe Fast qPCR kit (Kapa Biosystems) on a 7500 Fast Real Time PCR instrument (Applied Biosystems). The reaction conditions were 95 °C for 20 s, followed by 40 cycles of 95 °C for 3 s and 60 °C for 30 s. Data was analysed using the Sequence Detection Software according to the ΔC_t method with $Hprt$ serving as the reference housekeeping gene.

Glucose tolerance test (GTT). GTTs were performed after 10–12 weeks of consuming the HFD. Mice were fasted overnight ($\sim$14 h), and injected intraperitoneally with 10% dextrose at a dose of 1 g per kg body weight. Blood was collected from tail vein and plasma glucose levels measured at indicated times using a YSI 2700 Select Glucose Analyzer (YSI Life Sciences). Plasma insulin levels were determined by radioimmunoassay (Linco).

Flow cytometry analysis. Livers were collected, digested with $0.5\,mg\,ml^{-1}$ collagenase IV (Sigma) for 45 min at 37 °C, homogenized and repeatedly centrifuged at $400g$ for 5 min to enrich for haematopoietic cells. Cells were stained for flow cytometry using antibodies against CD45.2, CD11b, CD11c, NK1.1, B220, CD4, CD8, TCRβ, F4/80, Gr-1, MHC class II (Biolegend) and analysed on a BD LDR II.

Portal vein blood collection. Mice were anaesthetized with ketamine 100 mg per kg and xylazine 10 mg per kg. Mice were placed on a clean surgical field, and the abdominal fur was clipped and cleaned with a two stage surgical scrub consisting of Betadine and 70% ethanol. A 1 to 1.5 cm midline incision was made in the skin and abdominal wall. The peritoneum was moved to the left and the portal vein was punctured with a 30G needle. Between 0.2 and 0.3 ml of blood were collected per mouse. Serum was recovered by centrifugation at $1,500g$ for 15 min at room temperature and then stored at -80 °C in endotoxin-free tubes until assayed.

Measurement of PAMPs. TLR2, TLR4 and TLR9 agonists were assayed in portal vein serum using HEK-blue mTLR2, HEK-blue mTLR4 and HEK-blue mTLR9 reporter cell lines (InvivoGen) and the manufacturer's protocol with modifications. In brief, 2.2×10^5 HEK-blue mTLR2, 1.0×10^5 HEK-blue mTLR4 and 2.0×10^5 HEK-blue mTLR9 cells were plated in 96-well plates containing 10 μl of heat-inactivated (45 min at 56 °C) portal vein serum. Cells were then incubated for 21 h at 37 °C under an atmosphere of 5% CO_2/95% air. Twenty microlitres of the cell culture supernatants were collected and added to 180 μl of the QUANTI-Blue substrate in a 96-well plate. The mixtures were then incubated at 37 °C in 5% CO_2/95% air for 3 h and secreted embryonic alkaline phosphatase levels were determined using a spectrophotometer at 655 nm.

Transmission electron microscopy. Mice were perfused via their left ventricles using 4% paraformaldehyde in PBS. Selected tissues were fixed in 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer pH 7.4 for 1–2 h. Samples were rinsed three times in sodium cacodylate buffer and post-fixed in 1% osmium tetroxide for 1 h, en bloc stained in 2% uranyl acetate in maleate buffer pH 5.2 for a further hour then rinsed, dehydrated, infiltrated with Epon812 resin, and baked overnight at 60 °C. Hardened blocks were cut using a Leica UltraCut UCT. 60-nm-thick sections were collected and stained using 2% uranyl acetate and lead citrate. Samples were all viewed in an FEI Tencai Biotwin TEM at 80 kV. Images were taken using Morada CCD and iTEM (Olympus) software. Further details in ref. 17.

Bone marrow chimeras. Bone marrow was flushed from femurs with DMEM with 10% FBS, red cells were lysed, and the material filtered through a 70 μm filter. 10^6 cells in 100 μl PBS were delivered by retro-orbital injection into lethally irradiated (1,000 rad) mice. For 2 weeks post-engraftment, mice were maintained on antibiotics (Sulfatrim). Six weeks after transplantation animals were switched to MCDD. A wild-type non-irradiated mouse was co-housed with the engrafted mice for 4 weeks before NASH induction. Under our standardized protocol, bone marrow chimaeras routinely show a level of engraftment of $\geq$ 93%.

Bacterial 16S rRNA amplicon sequencing. Total DNA was isolated from the livers of mice fed a MCDD diet and used for attempted PCR amplification of variable region 2 of bacterial 16S rRNA genes[20] that may be present in the tissue. Thirty cycles of amplification of liver DNA prepared from seven wild-type, and seven $Asc^{-/-}$ mice yielded detectable product (>60 ng per reaction) in three samples from the wild-type group and three samples from the $Asc^{-/-}$ group. All amplicons were then subjected to multiplex pyrosequencing with a 454 instrument using FLX Titanium chemistry (137–1,510 reads per sample, average read length, 360 nucleotides). Reads were analysed using the QIIME software package. Operational taxonomic unit (OTU) picking was performed using uclust and taxonomic assignments made with RDP[41]. This analysis demonstrated inconsistent representation of taxa between animals and taxa that largely represented organisms not associated with the gut microbiota. G-test indicated that there was no significant correlation between any of these taxa and the presence of NASH.

For analysis of the faecal microbiota of MCDD-fed $Asc^{-/-}$ (WT), WT($Asc^{-/-}$) and singly housed wild-type mice, faecal pellets were collected at the time points indicated in Fig. 3. The protocols that we used to extract faecal DNA and to perform multiplex pyrosequencing of amplicons generated by PCR from the V2 regions of bacterial 16S rRNA genes, have been previously described[20]. A total of 366,283 sequences were generated from 181 faecal samples (average $2,023 \pm 685$ reads per sample; average read length, 360 nucleotides). Sequences were de-multiplexed and binned into species-level operational taxonomic units (OTUs; 97% nucleotide sequence identity; %ID) using QIIME 1.2.1 (ref. 41). Taxonomy was assigned within QIIME using RDP. Chimaeric sequences were removed using ChimeraSlayer and OTUs were filtered to a minimum of 10 sequences per OTU and 1,000 OTUs per sample. PCoA plots were generated by averaging the unweighted UniFrac distances of 100 subsampled OTU tables. Statistical analysis was performed on the proportional representation of taxa (summarized to Phyla, Class, Order, Family and Genus levels), using paired (where possible) and unpaired t-tests. Taxa that were significantly different after multiple hypothesis testing were included in Supplementary Tables 1–3.

Statistical analysis. Data are expressed as mean $\pm$ s.e.m. Differences were analysed by Student's t-test or ANOVA and post hoc analysis for multiple group comparison. P values ≤ 0.05 were considered significant.

38. Sutterwala, F. S. et al. Critical role for NALP3/CIAS1/Cryopyrin in innate and adaptive immunity through its regulation of caspase-1. Immunity 24, 317–327 (2006).
39. Hemmi, H. et al. A Toll-like receptor recognizes bacterial DNA. Nature 408, 740–745 (2000).
40. Kleiner, D. E. et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 41, 1313–1321 (2005).
41. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7, 335–336 (2010).

ARTICLE

Complete subunit architecture of the proteasome regulatory particle

Gabriel C. Lander[1]*, Eric Estrin[2]*, Mary E. Matyskiela[2]*, Charlene Bashore[2], Eva Nogales[1,3,4] & Andreas Martin[2,4]

The proteasome is the major ATP–dependent protease in eukaryotic cells, but limited structural information restricts a mechanistic understanding of its activities. The proteasome regulatory particle, consisting of the lid and base subcomplexes, recognizes and processes polyubiquitinated substrates. Here we used electron microscopy and a new heterologous expression system for the lid to delineate the complete subunit architecture of the regulatory particle from yeast. Our studies reveal the spatial arrangement of ubiquitin receptors, deubiquitinating enzymes and the protein unfolding machinery at subnanometre resolution, outlining the substrate's path to degradation. Unexpectedly, the ATPase subunits within the base unfoldase are arranged in a spiral staircase, providing insight into potential mechanisms for substrate translocation through the central pore. Large conformational rearrangements of the lid upon holoenzyme formation suggest allosteric regulation of deubiquitination. We provide a structural basis for the ability of the proteasome to degrade a diverse set of substrates and thus regulate vital cellular processes.

The ubiquitin–proteasome system is the major pathway for selective protein degradation in eukaryotic cells. Covalent modification with a polyubiquitin chain targets damaged, misfolded and short-lived regulatory proteins for ATP-dependent destruction by the 26S proteasome, a massive 1.5 MDa proteolytic machine. The proteasome thus controls a myriad of essential cellular processes, including the cell cycle, transcription and protein quality control[1]. Despite intensive study, however, the structural basis for substrate recognition and processing by the proteasome remains poorly understood.

The proteasome contains at least 32 different subunits that form a barrel-shaped 20S proteolytic core capped on either end by a 19S regulatory particle. The active sites of the peptidase are sequestered in an internal chamber, and access is controlled by the regulatory particle, which functions in substrate recognition, deubiquitination, unfolding and translocation of the unfolded chains into the core[2–5].

The regulatory particle is composed of 19 subunits and can be divided into two subcomplexes, the lid and the base. The lid consists of nine non-ATPase proteins (Rpn3, Rpn5– Rpn9, Rpn11, Rpn12 and Sem1 in yeast), including the deubiquitinating enzyme (DUB) Rpn11, whose activity is essential for efficient substrate degradation[6,7]. The base contains six distinct AAA+ ATPases, Rpt1–Rpt6, that form a hetero-hexameric ring (in the order Rpt1, Rpt2, Rpt6, Rpt3, Rpt4, Rpt5; ref. 8) and constitute the molecular motor of the proteasome. The ATPases are predicted to use the energy of ATP binding and hydrolysis to exert a pulling force on substrate proteins, unfold them, and translocate the polypeptides through a narrow central pore into the peptidase chamber. In the presence of ATP, the carboxy termini of the ATPases bind dedicated sites on the α-subunit ring (α1–α7) of the 20S core, triggering the opening of a gated access channel and facilitating substrate entry[5,9–11]. Besides Rpt1–Rpt6, the base contains four non-ATPase subunits: Rpn1, Rpn2 and the ubiquitin receptors Rpn10 and Rpn13. Additional ubiquitin shuttle receptors (Rad23, Ddi1 and Dsk2) are recruited to the base through interactions with Rpn1, which also binds a second, non-essential DUB, Ubp6 (refs 12–14).

Whereas the proteolytic core has been well studied, there is only limited structural characterization of the regulatory particle[11,15–17].

None of the 13 non-ATPase subunits, including the ubiquitin receptors and deubiquitinating enzymes, have been localized within this assembly. Although it has been shown that efficient degradation depends on the length, linkage type and placement of an ubiquitin chain, as well as the presence of an unstructured initiation site on a substrate[3,18,19], we are missing the topological information needed to explain these requirements. Thus, elucidating the architecture of the regulatory particle and the spatial arrangement of individual subunits is crucial to understanding the molecular mechanisms for substrate recognition and processing.

Here, we present the electron microscopy structure of the proteasome holoenzyme and the lid subcomplex. A new heterologous expression system for the lid facilitated the localization of all subunits within the regulatory particle, providing a complete architectural picture of the proteasome. The resulting structural understanding offers novel insight into the mechanisms of ubiquitin binding, deubiquitination, substrate unfolding and translocation by this major eukaryotic proteolytic machine.

Recombinant expression of yeast lid in *Escherichia coli*

We developed a system for the heterologous coexpression of all nine lid subunits from *Saccharomyces cerevisiae* in *Escherichia coli*. This system allowed us to generate truncations, deletions and fusion constructs that were used to localize individual subunits and delineate their boundaries within the lid. The recombinant, purified lid was analysed in its subunit composition and stoichiometry by SDS–polyacrylamide gel electrophoresis (SDS–PAGE; Supplementary Figs 1 and 2) and tandem mass spectrometry. The small, non-essential subunit Sem1 could not be detected, neither for the recombinant nor the endogenous lid that was isolated from yeast. All other subunits were present with the expected stoichiometry, and gel-filtration analyses showed indistinguishable elution profiles for the heterologously expressed lid and its endogenous counterpart (data not shown). Furthermore, atomic emission spectroscopy confirmed that the essential Zn^{2+} ion was incorporated in Rpn11, indicating proper folding in *E. coli*.

[1]Life Sciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720, USA. [2]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. [3]Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. [4]QB3 Institute, University of California, Berkeley, California 94720, USA.
*These authors contributed equally to this work.

To compare the functionalities of recombinant and endogenous lid, we established conditions for their *in vitro* reconstitution with base and 20S core subcomplexes from yeast to yield 26S holoenzyme. These reassembled particles were assayed for their activity in ubiquitin-dependent substrate degradation by using a polyubiquitinated green fluorescent protein (GFP)–cyclin fusion protein and following the decrease in GFP fluorescence. Proteasome reconstituted with *E. coli*-expressed lid supported robust substrate degradation (Supplementary Fig. 3). Importantly, the three-dimensional electron microscopy reconstructions from negative-stained samples of both lid sub-complexes are practically identical (Fig. 1a and Supplementary Fig. 4), establishing this recombinant system as an ideal tool for our structural studies of the regulatory particle.

Localization of regulatory particle subunits

As a first step in elucidating the architecture of the regulatory particle, we compared the single-particle electron microscopy reconstructions of the yeast holoenzyme and the isolated lid subcomplex obtained at 9- and 15-Å resolution, respectively (Fig. 1b, Supplementary Figs 5–7 and Supplementary Movie 1). Docking the five-lobed, hand-shaped structure of the lid into the electron density of the holoenzyme revealed the lid's position on one side of the regulatory particle, forming extensive interactions with the base subcomplex, but also contacting the 20S core. The lid subunits Rpn3, Rpn5, Rpn6, Rpn7, Rpn9 and Rpn12 contain a C-terminal PCI (Proteasome–CSN–eIF3) domain that is assumed to have scaffolding functions and allow inter-subunit contacts[1]. Our reconstruction provided sufficient resolution to unambiguously locate the winged-helix fold and the flanking helical segments of individual PCIs (Fig. 1c and Supplementary Movie 1). The C-terminal PCI domains of the six Rpn subunits thus interact laterally to form a horseshoe-shaped anchor from which the amino-terminal domains extend radially. This arrangement demonstrates

the scaffolding function of PCI domains in the lid, and we predict that similar interactions underlie the architecture of other PCI-containing complexes.

To determine the subunit topology of the lid, we used our heterologous *E. coli* expression system, fused maltose-binding protein (MBP) to the N or C terminus of individual subunits (Supplementary Fig. 1), and localized the MBP within the tagged lid particles by negative-stain electron microscopy (Supplementary Fig. 8a). None of the MBP fusions notably affected the lid structure, and we were able to identify the positions of all eight essential lid subunits and the relative orientation of their N and C termini. In combination with the PCI docking, the resolution of secondary structures in the cryoelectron density and known molecular weights, this information allowed us to delineate approximate subunit boundaries (Fig. 2a and Supplementary Movie 1).

Overall, Rpn3, Rpn7, Rpn6, Rpn5 and Rpn9 form the fingers of the hand-shaped lid structure. Rpn8 shows an extended conformation that connects Rpn3 and Rpn9, and thus closes the PCI horseshoe. In addition, it interacts with Rpn11, the only essential DUB of the proteasome, which lies in the palm of the hand and makes extensive contacts with Rpn8, Rpn9 and Rpn5.

Using the topology determined for the isolated lid subcomplex, we delineated the individual lid subunits in the context of the holoenzyme (Fig. 2b). To complete the subunit assignment for the entire regulatory particle, the positions of Rpt1–Rpt6 in the base subcomplex were assigned according to established interactions with the core particle[15,20], whose crystal structure could be docked unambiguously into the electron microscopy density (Supplementary Fig. 9). We localized the two large non-ATPases Rpn1 and Rpn2 of the base subcomplex by antibody-labelling of a C-terminal Flag tag and N-terminal fusion of glutathione-*S*-transferase (GST), respectively (Supplementary Figs 2 and 10a–c). Rpn1 and Rpn2 had been predicted to contain numerous tetratricopeptide repeat (TPR)-like motifs and adopt α-solenoid structures[21]. Indeed, we found a high structural resemblance between Rpn1 and Rpn2, both consisting of a strongly curled solenoid that transitions into an extended arm towards the C terminus (Fig. 3a). Rpn1 contacts the C-terminal helix of the 20S core subunit α4 and, based on the variability observed in our electron microscopy images, is likely to be flexible or loosely attached to the side of the base. Previous crystallography studies of the archaeal proteasome homologue PAN revealed that the N-terminal domains of the ATPases form a separate hexameric ring (N-ring) that consists of OB domains and three protruding coiled-coil segments[17,22]. Each coiled coil is formed by the far N-terminal residues of two neighbouring ATPases in the hexamer. Although Rpt1 and Rpt2 do not seem to form an extended coiled coil, we find that the N-terminal helical portion of Rpt1 interacts with the solenoid and the C-terminal arm of Rpn1. Rpn2 is located above the N-ring and mounted atop the longest of the protruding coiled coils, formed by Rpt3 and Rpt6. These interactions strongly resemble those observed between Rpt1 and Rpn1 (Fig. 3a).

Localizing the ubiquitin receptors and DUBs within the regulatory particle is of particular interest. In addition to the DUB Rpn11 in the lid, we identified the positions of both intrinsic ubiquitin receptors, Rpn10 and Rpn13, and of the base-associated DUB Ubp6 by imaging proteasome particles from yeast deletion strains (Fig. 3b and Supplementary Fig. 10d–f). The ubiquitin receptor Rpn13 binds to Rpn2 as expected[23,24]. The globular VWA domain of the second receptor Rpn10 has been shown previously to stabilize the lid–base interaction[25,26]; however, we found that it does not contact the base directly. This domain bridges Rpn11 and Rpn9, which might increase the lid–base affinity indirectly by stabilizing Rpn11 in its Rpn2-bound conformation (see below). The flexibly attached ubiquitin interacting motif (UIM) of Rpn10 probably contacts the coiled coil formed by Rpt4 and Rpt5, stabilizing its position relative to other subunits and potentially communicating with the AAA+ motor. The DUB Ubp6 seems to be flexible and does not give rise to ordered density. Nonetheless, variance
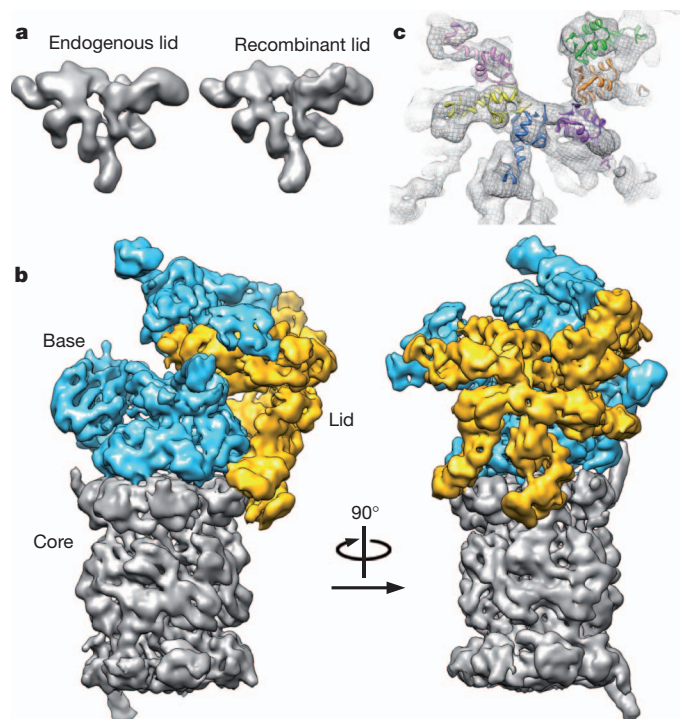

Figure 1 | The lid subcomplex within the holoenzyme assembly. a, Negative-stain three-dimensional reconstruction at approximately 15-Å resolution shows resemblance between endogenous (left) and recombinant (right) lid. **b**, Locations of lid (yellow) and base (cyan) within the subnanometre holoenzyme reconstruction. **c**, Six copies of the crystal structure of a PCI domain (PDB ID: 1RZ4) are docked into the lid electron density, showing a horseshoe-shaped arrangement of the winged-helix domains. Each domain is coloured according to its respective lid subunit (Fig. 2).
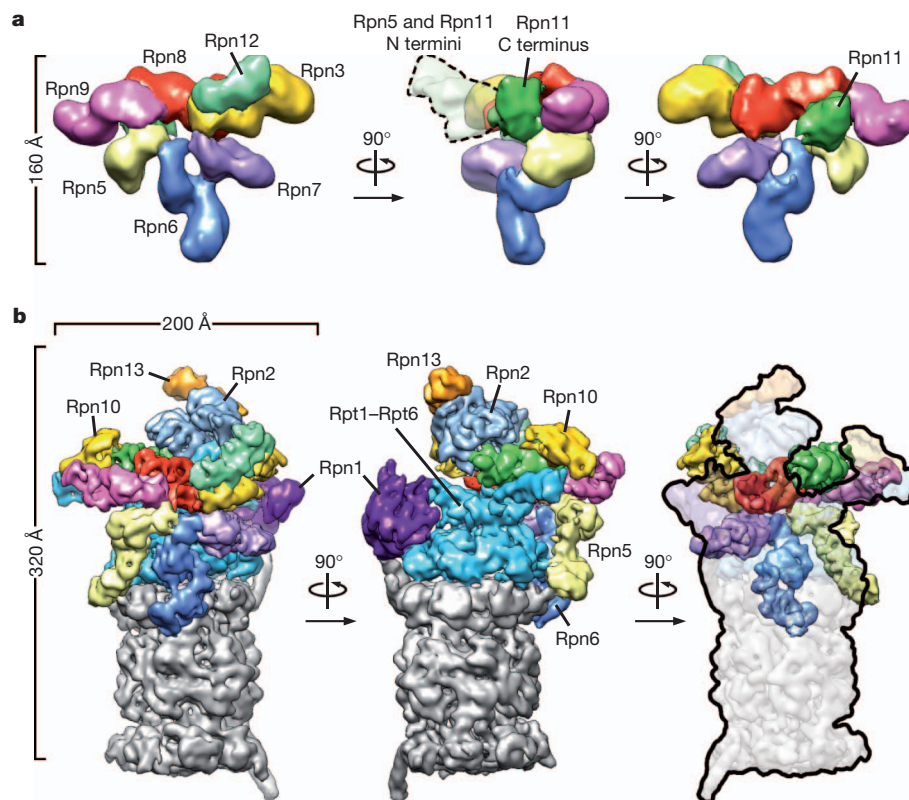
Figure 2 | Three-dimensional reconstructions of the recombinant lid subcomplex and the yeast 26S proteasome. a, Negative-stain reconstruction of the isolated lid subcomplex at 15-Å resolution, coloured by subunit and shown from the exterior (left), the side (middle) and the interior, base-facing side (right). A dotted line (middle) indicates the highly variable electron density maps indicate that it interacts with the C-terminal arm of Rpn1, as suggested by immunoprecipitations[14].

for the flexible N-terminal domains of Rpn5 and Rpn11. **b**, Subnanometre cryoelectron microscopy reconstruction of the holoenzyme, shown in three views corresponding to the isolated lid and coloured as above, with the core particle in grey.

Inter–subcomplex contacts

The complete localization of subunits within the holoenzyme revealed unexpected contacts between the lid and core subcomplexes. Rpn5 and Rpn6 form fingers that touch the C termini of the core subunits α1 and α2, respectively. We confirmed the interaction between Rpn6 and α2 by *in vitro* crosslinking, using an engineered cysteine in α2 and a 7-Å heterobifunctional crosslinker (Supplementary Fig. 11). These previously unknown direct interactions between lid and core may stabilize the entire holoenzyme assembly, and/or be part of an allosteric network that modulates the activities of either subcomplex.

Our holoenzyme structure shows that Rpn3, Rpn7, Rpn8 and Rpn11 make extensive contacts with the base. Compared to their positions in the isolated lid, Rpn8 and Rpn11 have undergone significant conformational changes in the holoenzyme (Fig. 4). The C terminus of Rpn8 is detached from Rpn3 to interact with the coiled coil of Rpt3/Rpt6, while the N-terminal MPN domain of Rpn11 extends towards the centre of the regulatory particle to bind the solenoid portion of Rpn2. Similarly, the N-terminal region of Rpn3 is more elongated than in the isolated lid and also contacts the Rpn2 solenoid, but from the opposite side. In turn, the extended C-terminal arm of Rpn2 interacts with Rpn3 and Rpn12, and thus forms a direct connection between the solenoid section of Rpn2, the coiled coil of Rpt3/Rpt6, and the lid (Fig. 3b).
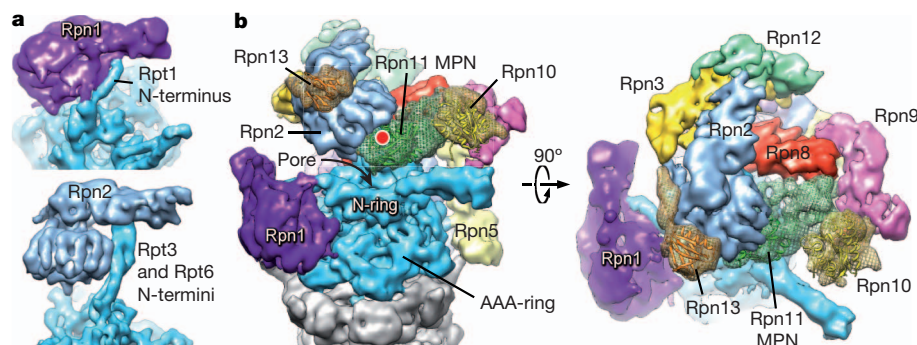

Figure 3 | Localization of Rpn1 and Rpn2, and ubiquitin-interacting subunits. a, Rpn1 (top) and Rpn2 (bottom) are oriented to emphasize similarities in their domain structure and solenoid attachment to the extended N-terminal helices of Rpt1 and Rpt3/Rpt6, respectively. **b**, Side and top views of the regulatory particle, showing the locations of the ubiquitin receptors Rpn10 and Rpn13, and the DUB Rpn11 relative to the central pore. Crystal structures for Rpn10 (PDB ID: 2X5N), Rpn13 (PDB ID: 2R2Y), and an MPN domain homologous to Rpn11 (AMSH-LP, PDB ID: 2ZNR) are shown docked into the electron microscopy density. The predicted active site of Rpn11 is indicated (red dot).
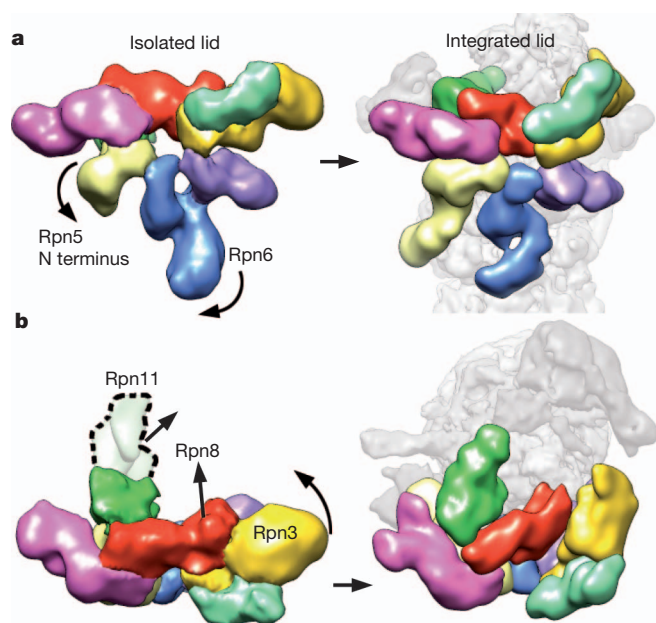
Figure 4 | Conformational rearrangements of the lid subcomplex upon integration into the holoenzyme. a, b, The lid complex in its isolated (left) and integrated (right) state is shown as viewed from the exterior (**a**) and top (**b**) of the regulatory particle. Major subunit rearrangements are depicted by arrows. The N terminus of Rpn5 (light yellow) interacts with Rpn11 in the isolated complex, and swings down to contact the core particle upon incorporation into the holoenzyme. The N-terminal domain of Rpn6 swings to the left to interact similarly with the core particle. Rpn3, Rpn8 and Rpn11 undergo notable rearrangements, in which they move towards the centre of the regulatory particle.

We speculate that Rpn2 stabilizes a lid conformation in which Rpn3, Rpn8 and the DUB Rpn11 extend towards the base (Fig. 4b). Together, the lid, Rpn2 and the coiled coils of the N-ring seem to function as a scaffold that positions the two intrinsic ubiquitin receptors Rpn10 and Rpn13, and the DUB Rpn11 for substrate binding, deubiquitination and transfer to the subjacent central pore of the AAA+ motor (Fig. 3b). Interestingly, several lid subunits interact directly with AAA+ domains of the Rpt subunits. Rpn7 contacts the AAA+ domains of Rpt2 and Rpt6, while Rpn6 and Rpn5 touch Rpt3. These interactions with contiguous motor domains are surprising, because current models for ATP-dependent unfoldases suggest significant conformational changes of individual subunits in the hexamer during ATP hydrolysis and substrate translocation[27–29]. The observed contacts between lid and the motor domains might form only transiently; alternatively, the AAA+ ring of the proteasome may be much more static than previously assumed.

Lid conformational changes may regulate DUB activity

Comparing the structures of the lid in isolation and when bound to holoenzyme revealed major conformational changes that suggest an allosteric mechanism for the regulation of Rpn11 DUB activity (Fig. 4). In the isolated lid, the N-terminal MPN domain of Rpn11 forms extensive interactions with Rpn9 and the curled up Rpn5 finger. Upon lid binding to the holoenzyme, this Rpn5 finger swings down to contact the α1 subunit of the 20S core and thereby releases Rpn11, which then extends towards the Rpn2 solenoid. Docking the MPN domain of a related DUB (PDB ID: 2ZNR) into the electron density of Rpn11 indicates the approximate location of the active site (Fig. 3b). The interactions of Rpn11 with Rpn9 and Rpn5 in the free lid probably restrict access to this active site, which would prevent futile substrate deubiquitination in the absence of base and 20S core, and explain previous observations that the lid subcomplex has DUB activity only within the holoenzyme[7] (and our unpublished data).

Functional asymmetry in the AAA+ unfoldase

Our subnanometre structure of the holoenzyme provides new insights into the architecture and potential mechanisms of the base AAA+ unfoldase. As suggested by previous electron microscopy studies[15,16], the ring of the base and the 20S core are slightly offset from a coaxial alignment, with the base shifted by approximately 10 Å towards the lid (Fig. 5a). Despite or perhaps because of this offset, the C-terminal tails of Rpt2, Rpt3 and Rpt5 are docked into their cognate 20S binding pockets at the interfaces of the subunits α3 and α4, α1 and α2, and α5 and α6, respectively. Those three Rpt tails contain the terminal HbYX motif, which is critical for triggering gate opening in the 20S core[5,10], and indeed our structure is consistent with an open-gate conformation. The tails of Rpt1, Rpt4 and Rpt6 lack this motif and were not observed to interact statically with 20S in our holoenzyme structure.

Current mechanistic models for AAA+ unfoldases predict that ATPase subunits in the hexamer are in different nucleotide states and undergo significant conformational changes driven by coordinated ATP hydrolysis[27,30,31]. Because we determined the structure of wild-type proteasome in the presence of saturating ATP, we expected that different complexes would have any given Rpt subunit in different conformations, leading to reduced electron density or low resolution when averaging thousands of these unsynchronized motors. However, our reconstruction shows highly ordered density throughout the AAA+ domains of all six Rpt subunits. Whereas the C-terminal 'small AAA+' subdomains (except for Rpt6) arrange in one plane above the 20S core, the 'large AAA+' subdomains of Rpt1–Rpt5 are oriented in a spiral staircase around the hexameric ring, with Rpt3 at the highest and Rpt2 at the lowest position (Fig. 5b and Supplementary Movie 1). The AAA+ domain of Rpt6 adopts a tilted orientation, bridging Rpt2 and Rpt3. Similar staircase arrangements have been observed previously for helicases of the AAA+ and RecA superfamilies[32,33]. It was suggested that during ATP hydrolysis, individual subunits progress through the different conformational stages of the staircase, thereby translocating substrate through the pore. The particular staircase orientation we observed identically for all proteasome particles may represent a low-energy state of the base, adopted under our experimental conditions. Alternatively, this staircase arrangement of Rpt1–Rpt6 may be static and reflect the functional state of the base, in which substrates are translocated by local motions of the pore loops while the relative positions of the motor subunits remain fixed. Future biochemical and structural studies will be required to distinguish between these two models.
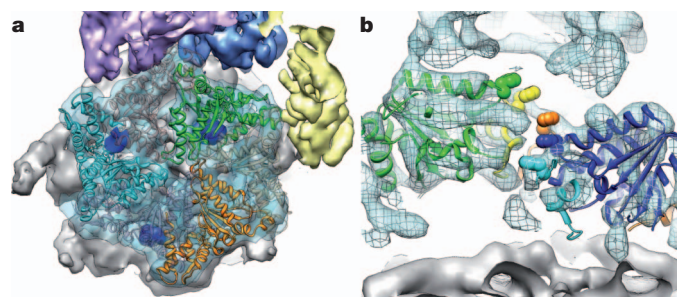

Figure 5 | Structural features of the base ATPase subunits. a, Positions of Rpt2 (cyan), Rpt3 (green) and Rpt5 (orange) within the base hexameric ring and relative to the 20S core (grey) are shown using fitted crystal structures of the homologous PAN AAA+ domain (PDB ID: 3H4M). The electron microscopy density contains the molecular envelope of the C-terminal tails (dark blue), docked into their cognate binding sites on the 20S core. Corresponding densities were not found for the tails of Rpt1, Rpt4 and Rpt6 (grey ribbon structure). **b,** Cutaway side view of the holoenzyme electron microscopy density with Rpt1–Rpt5 visible. Individually docked copies of the PAN crystal structure reveal a spiral staircase arrangement of the Rpt subunits, emphasized by space-filling representations of the PAN pore-1 loop residues (not resolved in the Rpt subunits).

Spatial arrangement of ubiquitin receptors and DUBs

Localizing all subunits of the regulatory particle enabled us to infer the requirements and potential mechanisms for the recognition and degradation of ubiquitin-tagged substrates (Fig. 6). After a substrate binds to an ubiquitin receptor, its polyubiquitin chain must be removed by Rpn11 cleavage at the proximal ubiquitin to permit subsequent fast degradation[6,7]. To allow cleavage without disengaging from the receptor, an ubiquitin chain must be long enough to span the distance between receptor and DUB. Based on our structure, both Rpn13 and the UIM of Rpn10 are located 70–80 Å from the predicted position of the Rpn11 MPN domain (Fig. 3b). The shuttle receptors Rad23, Ddi1 and Dsk2 are expected to reside ~80–120 Å away from Rpn11, depending on where they bind Rpn1 (ref. 13). For receptor interaction, at least part of the ubiquitin chain has to be in an extended conformation with the hydrophobic patches exposed[24,34,35]. Because a single ubiquitin moiety in an extended K48-linked chain contributes approximately 30 Å in length[36], it would take three ubiquitins to span the distance between Rpn10 or Rpn13 and Rpn11. Moreover, both Rpn10 and Rpn13 bind between two consecutive ubiquitin moieties[24,34], such that at least a tetra-ubiquitin chain would be required on a substrate to allow interaction with a receptor and simultaneous deubiquitination by Rpn11 (Fig. 6). This model agrees with *in-vitro* studies that indicate a minimum of four K48-linked ubiquitins is necessary for efficient substrate degradation[3], although this number may differ for other chain types[37]. Given the arrangement of Rpn10 and Rpn13, an ubiquitin chain would have to be significantly longer to interact with both receptors. However, knockout studies have shown that ubiquitin chains are not required to bind to multiple receptors simultaneously[38].

In contrast to Rpn11, Ubp6 is known to cleave within polyubiquitin chains or trim them from their distal end[39]. Of all the ubiquitin-interacting subunits in the regulatory particle, we found Ubp6 to be the furthest away from the entrance to the pore, which may allow it to clip extended or unnecessary ubiquitin chains from substrates. Because Ubp6 is located closer to Rad23, Dsk2 or Ddi1 than to Rpn10 or Rpn13, it may act preferentially on substrates delivered by these shuttle receptors.

To avoid dissociation upon deubiquitination, a substrate polypeptide must be engaged with the unfolding machinery of the base before or shortly after removal of its ubiquitin chain. Engagement by the base is known to depend on an unstructured initiation site or "tail" on the substrate[40], which needs to be long enough to reach through the narrow N-ring and into the AAA+ pore (Fig. 6). In addition, this tail would have to be sufficiently spaced from the attachment point of the poly-ubiquitin chain to allow concurrent substrate engagement by the pore and deubiquitination by Rpn11. The distance between the predicted active site of Rpn11 and the AAA+ pore below the N-ring is approximately 60 Å, which could easily be bridged by 40–45 unstructured residues or a shorter tail combined with a folded structure.

As an alternative to the above model for simultaneous receptor binding and deubiquitination, it has been proposed that commencing substrate translocation by the base might move the proximal ubiquitin from a receptor towards Rpn11 for cleavage[7]. Our structure suggests for this model that efficient substrate processing would only require a mono- or diubiquitin for receptor binding and a 50–60 Å longer spacing between the ubiquitin and the flexible tail to reach the AAA+ pore. This length dependence of engagement is consistent with recent *in vitro* degradation studies, using model substrates with different lengths and ubiquitin modifications[19]. Future experiments will be required to assess whether substrates get deubiquitinated in a translocation-dependent or -independent manner.

Concluding remarks

The work presented here defines the architecture of the entire proteasome regulatory particle and provides a much-needed structural framework for the mechanistic understanding of ubiquitin-dependent protein degradation. We localized Rpn11 directly above the entrance of the pore, surrounded by the ubiquitin receptors Rpn10 and Rpn13. This insight allows us to visualize the substrate's path towards degradation and will be critical in elucidating how the characteristics of ubiquitin modifications affect substrate recognition and processing. Moreover, our study significantly furthers the understanding of the heterohexameric AAA+ motor of the proteasome. Individual ATPase subunits were found in a spiral staircase arrangement and may operate with more limited dynamics than previously assumed for AAA+ protein unfoldases.

Unexpectedly, the lid is bound to the side of the holoenzyme and interacts with both the base and core particle. These interactions induce major conformational changes in lid subunits that may allosterically activate the DUB Rpn11, allowing critical removal of ubiquitin chains during substrate degradation in the holoenzyme, while preventing futile deubiquitination by the isolated lid. In addition, contacts between the subcomplexes could have unexplored roles in coordinating individual substrate processing steps, for instance ubiquitin binding, deubiquitination, and the onset of translocation. The intricate architecture of the proteasome highlights the complex requirements for this proteolytic machine, which must accommodate and specifically regulate a highly diverse set of substrates in the eukaryotic cell.
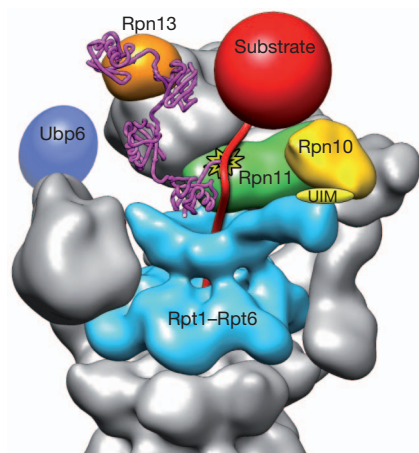

Figure 6 | Model for the recognition, deubiquitination and engagement of a polyubiquitinated substrate by the 26S proteasome. A K48-linked tetra-ubiquitin chain (magenta, PDB ID: 2KDE) is conjugated to the unstructured initiation region of a substrate (red) and bound to the ubiquitin receptor Rpn13 (orange). The substrate is poised for deubiquitination by Rpn11 (green, active site indicated by star), and its unstructured initiation region is engaged by the translocation machinery of the base (cyan). A polyubiquitin chain could alternatively bind to the UIM of Rpn10 (yellow) or interact with both receptors simultaneously. The DUB Ubp6 is localized further from the central pore, in a position to trim excess ubiquitin chains.

METHODS SUMMARY

Protein expression and purification. Endogenous holoenzyme, core particle[41] and lid subcomplex[42] were purified from *S. cerevisiae* essentially as described. The base subcomplex was purified according to protocols for the holoenzyme preparation, but with minor modifications as described in the Methods. Details of yeast strain construction are provided in Supplementary Table 1.

Yeast lid was recombinantly expressed from three plasmids in *E. coli* BL21-star (DE3), and purified on anti-Flag M2 resin and by size-exclusion chromatography (see Methods).

Electron microscopy and image analysis. All electron microscopy data were collected using the Leginon data collection software[43] and processed in the Appion electron microscopy processing environment[44]. Three-dimensional maps were calculated using libraries from the EMAN2 and SPARX software packages[45,46]. UCSF Chimera was used for volume segmentation, atomic coordinate docking and figure generation[47].

1. Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78,** 477–513 (2009).
2. Glickman, M. H., Rubin, D. M., Fried, V. A. & Finley, D. The regulatory particle of the *Saccharomyces cerevisiae* proteasome. *Mol. Cell. Biol.* **18,** 3149–3162 (1998).
3. Thrower, J. S., Hoffman, L., Rechsteiner, M. & Pickart, C. M. Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **19,** 94–102 (2000).
4. Groll, M. et al. A gated channel into the proteasome core particle. *Nature Struct. Biol.* **7,** 1062–1067 (2000).
5. Smith, D. M. et al. Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's α ring opens the gate for substrate entry. *Mol. Cell* **27,** 731–744 (2007).
6. Yao, T. & Cohen, R. E. A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419,** 403–407 (2002).
7. Verma, R. et al. Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298,** 611–615 (2002).
8. Tomko, R. J. Jr, Funakoshi, M., Schneider, K., Wang, J. & Hochstrasser, M. Heterohexameric ring arrangement of the eukaryotic proteasomal ATPases: implications for proteasome structure and assembly. *Mol. Cell* **38,** 393–403 (2010).
9. Rabl, J. et al. Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Mol. Cell* **30,** 360–368 (2008).
10. Gillette, T. G., Kumar, B., Thompson, D., Slaughter, C. A. & DeMartino, G. N. Differential roles of the COOH termini of AAA subunits of PA700 (19 S regulator) in asymmetric assembly and activation of the 26 S proteasome. *J. Biol. Chem.* **283,** 31813–31822 (2008).
11. da Fonseca, P. C. & Morris, E. P. Structure of the human 26S proteasome: subunit radial displacements open the gate into the proteolytic core. *J. Biol. Chem.* **283,** 23305–23314 (2008).
12. Elsasser, S. et al. Proteasome subunit Rpn1 binds ubiquitin-like protein domains. *Nature Cell Biol.* **4,** 725–730 (2002).
13. Gomez, T. A., Kolawa, N., Gee, M., Sweredoski, M. J. & Deshaies, R. J. Identification of a functional docking site in the Rpn1 LRR domain for the UBA-UBL domain protein Ddi1. *BMC Biol.* **9,** 33 (2011).
14. Leggett, D. S. et al. Multiple associated proteins regulate proteasome structure and function. *Mol. Cell* **10,** 495–507 (2002).
15. Bohn, S. et al. Structure of the 26S proteasome from *Schizosaccharomyces pombe* at subnanometer resolution. *Proc. Natl Acad. Sci. USA* **107,** 20992–20997 (2010).
16. Nickell, S. et al. Insights into the molecular architecture of the 26S proteasome. *Proc. Natl Acad. Sci. USA* **106,** 11943–11947 (2009).
17. Förster, F. et al. An atomic model AAA-ATPase/20S core particle sub-complex of the 26S proteasome. *Biochem. Biophys. Res. Commun.* **388,** 228–233 (2009).
18. Xu, P. et al. Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* **137,** 133–145 (2009).
19. Inobe, T., Fishbain, S., Prakash, S. & Matouschek, A. Defining the geometry of the two-component proteasome degron. *Nature Chem. Biol.* **7,** 161–167 (2011).
20. Tian, G. et al. An asymmetric interface between the regulatory and core particles of the proteasome. *Nature Struct. Mol. Biol.* **18,** 1259–1267 (2011).
21. Effantin, G., Rosenzweig, R., Glickman, M. H. & Steven, A. C. Electron microscopic evidence in support of α-solenoid models of proteasomal subunits Rpn1 and Rpn2. *J. Mol. Biol.* **386,** 1204–1211 (2009).
22. Zhang, F. et al. Structural insights into the regulatory particle of the proteasome from *Methanocaldococcus jannaschii*. *Mol. Cell* **34,** 473–484 (2009).
23. Hamazaki, J. et al. A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. *EMBO J.* **25,** 4524–4536 (2006).
24. Schreiner, P. et al. Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* **453,** 548–552 (2008).
25. Glickman, M. H. et al. A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* **94,** 615–623 (1998).
26. Verma, R., Oania, R., Graumann, J. & Deshaies, R. J. Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin-proteasome system. *Cell* **118,** 99–110 (2004).
27. Glynn, S. E., Martin, A., Nager, A. R., Baker, T. A. & Sauer, R. T. Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139,** 744–756 (2009).
28. Maillard, R. A. et al. ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145,** 459–469 (2011).
29. Aubin-Tam, M. E., Olivares, A. O., Sauer, R. T., Baker, T. A. & Lang, M. J. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* **145,** 257–267 (2011).
30. Martin, A., Baker, T. A. & Sauer, R. T. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437,** 1115–1120 (2005).
31. Hersch, G. L., Burton, R. E., Bolon, D. N., Baker, T. A. & Sauer, R. T. Asymmetric interactions of ATP with the AAA+ ClpX6 unfoldase: allosteric control of a protein machine. *Cell* **121,** 1017–1027 (2005).
32. Thomsen, N. D. & Berger, J. M. Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* **139,** 523–534 (2009).
33. Enemark, E. J. & Joshua-Tor, L. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442,** 270–275 (2006).
34. Riedinger, C. et al. Structure of Rpn10 and its interactions with polyubiquitin chains and the proteasome subunit Rpn12. *J. Biol. Chem.* **285,** 33992–34003 (2010).
35. Eddins, M. J., Varadan, R., Fushman, D., Pickart, C. M. & Wolberger, C. Crystal structure and solution NMR studies of Lys48-linked tetraubiquitin at neutral pH. *J. Mol. Biol.* **367,** 204–211 (2007).
36. Cook, W. J., Jeffrey, L. C., Carson, M., Chen, Z. & Pickart, C. M. Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). *J. Biol. Chem.* **267,** 16467–16471 (1992).
37. Bremm, A., Freund, S. M. & Komander, D. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nature Struct. Mol. Biol.* **17,** 939–947 (2010).
38. Husnjak, K. et al. Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453,** 481–488 (2008).
39. Hanna, J. et al. Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. *Cell* **127,** 99–111 (2006).
40. Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E. & Matouschek, A. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nature Struct. Mol. Biol.* **11,** 830–837 (2004).
41. Verma, R. et al. Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11,** 3425–3439 (2000).
42. Leggett, D. S., Glickman, M. H. & Finley, D. Purification of proteasomes, proteasome subcomplexes, and proteasome-associated proteins from budding yeast. *Methods Mol. Biol.* **301,** 57–70 (2005).
43. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151,** 41–60 (2005).
44. Lander, G. C. et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166,** 95–102 (2009).
45. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
46. Hohn, M. et al. SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157,** 47–55 (2007).
47. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157,** 281–287 (2007).

Author contributions E.E., M.E.M. and C.B. designed, expressed and purified proteasome constructs, and performed biochemical experiments. G.C.L. performed the electron microscopy, processing and segmentation analysis. All authors contributed to experimental design, data analysis and manuscript preparation.

METHODS

Recombinant lid construction and purification. Yeast Rpn5, Rpn6, Rpn8, Rpn9 and Rpn11–6×His were cloned into pETDuet-1 (Novagen), yeast Rpn3, Flag–Rpn7 and Rpn12 were cloned into pCOLADuet-1 (Novagen), and yeast Sem1 and Hsp90 were cloned into pACYCDuet-1 (Novagen). A T7 promoter preceded each gene and each plasmid contained a T7 terminator following the multiple cloning site. Genes for select rare transfer RNAs were included in the pACYCDuet-1 plasmid to account for codon-usage differences between yeast and *E. coli*. To ensure full-length of Rpn6 in lid particles used for biochemical experiments and the negative stain reconstruction of recombinant lid, we used a construct with the Flag tag moved from Rpn7 to Rpn6. *E. coli* BL21-star (DE3) cells were co-transformed with the three plasmids mentioned above. Lid proteins and the chaperone Hsp90 were coexpressed overnight at 18 °C after inducing cells with 0.5 mM isopropyl-β-D-thiogalactopyranoside at $D_{600} = 0.7$. Cells were collected by centrifugation (4,000*g* for 30 min), resuspended in Flag buffer (50 mM HEPES, pH 7.6, 100 mM NaCl, 100 mM KCl and 5% glycerol) supplemented with protease inhibitors and 2 mg ml^{-1} lysozyme, and sonicated on ice for 2 min in 15-s bursts. The lysate was clarified by centrifugation (27,000*g* for 30 min), and the complex was affinity-purified on anti-Flag M2 resin (Sigma-Aldrich) using an N-terminal Flag-tag on Rpn6 or Rpn7. The protein was concentrated in a 30,000 MWCO concentrator (Amicon) for further purification on a Superose 6 size-exclusion column (GE Healthcare) equilibrated in Flag buffer. Intact, assembled lid particles eluted at 13.1 ml, similar to lid purified from yeast.

His$_6$-tagged yeast Rpn10 was expressed in *E. coli* and purified by Ni-NTA affinity and size-exclusion chromatography.

Yeast strain construction. Wild-type holoenzyme was purified from the strain YYS40 (*MAT*a *ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1 RPN11::RPN11-3XFLAG (HIS3)*)[48]. To generate *RPN10*, *RPN13* and *UBP6* deletion strains, the kanMX6 sequence was integrated at the respective genomic locus, replacing the gene in YYS40. To generate the strains used to purify GST–Rpn2, GFP–Rpn5 and GFP–Rpn8 holoenzyme, sequences encoding the respective tags under the control of the P$_{GAL1}$ promoter were integrated 5′ of the respective genes in YYS40. To generate the strain used to purify Rpn1–Flag holoenzyme, a sequence encoding the Flag-tag was integrated 3′ to *RPN1* in aW303 background strain (*MAT*a *ade2-1 his3-11 leu2-3,112 trp1-1 ura3-1 can1-100 bar1*).

To generate the strains used to purify α2 mutant-containing core particle for the crosslinking experiments shown in Supplementary Fig. 11, pRS305 (*LEU2*) containing the mutant α2 and the genomic sequences found 500 nucleotides upstream and 100 nucleotides downstream of the gene was integrated at the *LEU2* locus of RJD1144 (*MAT*a, *his3Δ200 leu2-3,112 lys2-801 trpΔ63 ura3-52 PRE1-FLAG-6xHIS::Ylplac211 (URA3)*)[41], and the chromosomal copy of α2 was deleted. To generate the strain used to purify lid with Rpn6 tagged with three haemagglutinin (HA) for crosslinking, the 3×HA sequence was integrated 3′ of *RPN6* in YYS40.

Expression and purification of yeast holoenzyme and subcomplexes. Endogenous holoenzyme, core particle[41] and lid subcomplex[42] were purified from *S. cerevisiae* essentially as described. Frozen yeast cells were lysed in a Spex SamplePrep 6870 Freezer/Mill. For holoenzyme purification, lysed cells of a strain containing a Flag-tag on Rpn11 were resuspended in lysis buffer containing 60 mM HEPES pH 7.6, 50 mM NaCl, 50 mM KCl, 5 mM MgCl$_2$, 0.5 mM EDTA, 10% glycerol, 0.2% NP-40, and ATP regeneration mix (5 mM ATP, 0.03 mg ml^{-1} creatine kinase, 16 mM creatine phosphate). Holoenzyme was bound to anti-Flag M2 resin and washed with wash buffer (60 mM HEPES pH 7.6, 50 mM NaCl, 50 mM KCl, 5 mM MgCl$_2$, 0.5 mM EDTA, 10% glycerol, 0.1% NP-40 and 500 μM ATP) before elution with 3×Flag peptide and separation over Superose-6 in gel-filtration buffer (60 mM HEPES pH 7.6, 50 mM NaCl, 50 mM KCl, 5 mM MgCl$_2$, 0.5 mM EDTA, 10% glycerol and 500 μM ATP). Lid, base or core particle were purified similarly but from different yeast strains and including a salt wash to separate subcomplexes. Lid was purified from a yeast strain containing Rpn11–Flag using a 900 mM NaCl wash. Base was purified from a yeast strain containing a C-terminal Flag tag on Rpn2 and including a 500 mM NaCl wash, with 500 μM ATP present throughout the purification. Core particle was purified from a yeast strain containing a Flag–6×His tag on Pre1 and including a 500 mM NaCl wash. All subcomplexes were further purified by size-exclusion chromatography on Superose-6 in gel filtration buffer (see above).

GFP degradation assay. Proteasome holoenzyme was reconstituted from 20S core, base, Rpn10 and recombinant or endogenous yeast lid in the presence of ATP. A GFP–titin–cyclin fusion protein was modified with a K48-linked polyubiquitin chain[49] and degraded by reconstituted proteasome at 30 °C in Flag buffer with an ATP-regeneration system (5 mM ATP, 16 mM creatine phosphate, 6 μg ml^{-1} creatine phosphokinase). Degradation was monitored by the loss of fluorescence using a QuantaMaster spectrofluorimeter (PTI).

Protein crosslinking. Sulfo-MBS (Thermo Scientific) is a short (7.3 Å), heterobifunctional crosslinker, whose maleimide moiety reacts primarily with sulphydryls between pH 6.5 and 7.5, and whose NHS ester reacts with primary amines between pH 7 and 9. We purified core particle from yeast strains in which the only copy of the core α2 subunit was either wild type, a D245C mutant, or an A249C mutant. Other intrinsic cysteines of the core were found largely non-reactive towards sulphydryl-modifying agents (not shown). 10 μM reduced core particle purified from strains containing wild type, A249C and D245C α2 was incubated with 150 μM sulpho-MBS for 15 min at pH 6.5, allowing conjugation of the crosslinker to cysteines. Core particle was buffer-exchanged to remove excess crosslinker and increase the pH to 7.5, activating the amine-reactive functional group on sulpho-MBS. This core particle was added at a final concentration of 2 μM to a proteasome reconstitution mixture, containing 2 μM purified base, 10 μM purified Rpn10, 0.5 mM ATP, and 2 μM lid purified from a yeast strain in which Rpn6 was C-terminally tagged with a 3×HA tag. Crosslinking was allowed to proceed for 15 min before reactions were stopped by the addition of 0.5 mM glycine pH 7.5 and divided equally for separation by SDS–PAGE, followed by either Coomassie staining or anti-HA western blotting.

Electron microscopy. Sample preparation: negative-stain analysis of both the purified proteasome lid and holoenzyme complexes was performed using 400 mesh continuous carbon grids that had been plasma-cleaned in a 75% argon/25% oxygen atmosphere for 20 s using a Solarus plasma cleaner (Gatan). Due to the tendency for holoenzyme to adopt a preferential orientation on the carbon substrate, 5 μl of 0.1% poly L-lysine hydrobromide (Polysciences catalogue no. 09730) was placed onto the hydrophilized carbon grids and adsorbed for 90 s, washed twice with 5 μl drops of water, and allowed to dry completely. This polylysine step was skipped when preparing grids containing the lid samples, as the lid does not adopt a preferred orientation on the carbon substrate. The remaining steps were identical for both holoenzyme and lid. A 4-μl drop of sample at a concentration of 25 μM was placed onto the grid and allowed to adsorb for 1 min. The grid was blotted to near-dryness and a 4-μl drop of fresh 2% (w/v) uranyl formate was quickly placed onto the grid. To reduce the amount of glycerol remaining on the grids, they were subsequently floated on four successive 25-μl drops of the uranyl formate solution, waiting 10 s on each drop. The grids were then blotted to dryness.

Preservation of both lid and holoenzyme complexes in vitreous ice was performed in the same manner. 400-mesh C-flats containing 2 μm holes with a spacing of 2 μm (Protochips) were plasma cleaned in a 75% argon/25% oxygen atmosphere for 8 s using a Solarus plasma cleaner (Gatan). The purified sample, at a concentration of 5 μM in a buffer containing 5% glycerol, was first diluted 1:5 from 60 mM HEPES, pH 7.6, 50 mM NaCl, 50 mM KCl, 5 mM MgCl$_2$, 0.5 mM EDTA, 10% glycerol, 1 mM DTT, 0.5 mM ATP into a buffer containing 20 mM HEPES, pH 7.6, 50 mM NaCl, 50 mM KCl, 1 mM ATP, 1 mM DTT and 0.05% NP40, and 4-μl aliquots were placed onto the grids. Grids were immediately loaded into a Vitrobot (FEI company) whose climate chamber had equilibrated to 4 °C and 100% humidity. The grids were blotted for 3 s at an offset of −1 mm, and plunged into liquid ethane. The frozen grids were transferred to a grid box and stored in liquid nitrogen until retrieved for data collection.

Data collection: negative-stain analysis of the lid and holoenzyme samples was performed using a Tecnai T12 Bio-TWIN and a Tecnai F20 TWIN transmission electron microscope operating at 120 keV. Lid samples were imaged at a nominal magnification of ×68,000 (1.57 Å per pixel at the specimen level) on the T12, and ×80,000 (1.45 Å per pixel) on the F20. Holoenzyme samples were imaged at a magnification of ×49,000 (2.18 Å per pixel) on the T12, and ×50,000 (2.16 Å per pixel) on the F20. T12 data were acquired on a F416 CMOS 4Kx4K camera (TVIPS), F20 data were acquired on a Gatan 4Kx4K camera, and all micrographs were collected using an electron dose of 20 e$^-$ Å^{-2} with a randomly set focus ranging from −0.5 to −1.2 μm. The automatic rastering application of the Leginon data collection software was used for data acquisition. Between 300 and 500 micrographs were collected for each of the negatively stained data sets.

For cryoelectron microscopy, individual grids were loaded into a 626 single-tilt cryotransfer system (Gatan) and inserted into a Tecnai F20 TWIN transmission electron microscope operating at 120 keV. Data were acquired at a nominal magnification of ×100,000 (1.08 Å per pixel) using an electron dose of 20 e$^-$ Å^{-2} with a randomly set focus ranging from −1.2 to −2.5 μm. A total of 9,153 micrographs were collected of the holoenzyme using the MSI-T application of the Leginon software. While the holoenzyme was remained intact during the freezing process, the isolated lid specimen became completely disassembled during the freezing process. In an attempt to overcome this, the isolated lid was also frozen using grids onto which a thin carbon film was floated. Due to the elevated background noise from the addition of a carbon substrate, the resulting images lacked the sufficient signal-to-noise ratio necessary to solve a cryoelectron

microscopy structure of the isolated lid to a better resolution than the negative-stain structure.

Image processing of negative-stain data. All image pre-processing and two-dimensional classification was performed in the Appion image processing environment[44]. Due to the large number of data sets acquired for both the negatively stained lid and holoenzyme complexes, a generalized schema was used for image analysis. This schema also minimized user bias during comparison of tagged and deletion constructs with their wild-type counterparts. The contrast transfer function (CTF) of each micrograph was estimated concurrently with data collection using the ACE2 and CTFFind programs[50,51], providing a quantitative measurement of the imaging quality. Particle selection was also performed automatically concurrent with data collection. Negatively stained lid particles were selected from the micrographs using a difference of Gaussians (DoG) transform-based automated picker[52], and holoenzyme particles were selected using a template-based particle picker. Micrograph phases were corrected using ACE2, and both lid and holoenzyme particles were extracted using a 288 × 288-pixel box size. The data were then binned by a factor of two for processing. Each particle was normalized to remove pixels whose values were above or below 4.5 σ of the mean pixel value using the XMIPP normalization program[53].

To remove aggregation, contamination or other non-particle selections, particle stacks were decimated by a factor of 2 and subjected to five rounds of iterative multivariate statistical analysis (MSA) and multi-reference alignment (MRA) using the IMAGIC software package[54]. Two-dimensional class averages depicting properly assembled complexes were manually selected, and the non-decimated particles contributing to these class averages were extracted to create a new stack for further processing. To include a larger range of holoenzyme views, particles contributing to doubly capped proteasome averages were removed. This stack of particles went through five rounds of MSA/MRA in IMAGIC[54], and a final correspondence analysis and classification based on Eigen images using the SPIDER software package[55] was performed to generate two-dimensional class averages of the complexes.

Initial models for reconstructions of both the holoenzyme and lid were determined using the established "C1 startup" routines in IMAGIC. Two-dimensional class averages were manually inspected to select three images representing orthogonal views of the complex, which were in turn used to assign Eulers in a stepwise fashion to the entire data set of reference-free class averages. The resulting low-resolution models of the lid and holoenzyme were low-pass filtered to 60-Å resolution, and these densities were used as starting points for refinement of the three-dimensional structure.

Three-dimensional reconstructions were all performed using an iterative projection-matching and back-projection approach using libraries from the EMAN2 and SPARX software packages[45,46]. Refinement of the starting models began using an angular increment of 25°, progressing down to 2° for the lid, and 1° for the holoenzyme. The refinement only continued to the subsequent angular increment once greater than 95% of the particles showed a pixel error of less than 1 pixel. The resolution was estimated by splitting the particle stack into two equally sized data sets, calculating the Fourier shell correlation (FSC) between the resulting back-projected volumes. The estimated resolutions for the final endogenous and recombinant lid structures based on their FSC curves at 0.5 were about 15 Å.

Image processing of cryoelectron microscopy holoenzyme. Processing of the holoenzyme cryo data set proceeded in a very similar fashion to that of the negatively stained particle data sets. Only ACE2 was used to estimate CTF of the images and measure image quality, and particles were extracted using a box size of 576 pixels. Reference-free two-dimensional classification was performed to remove particles that did not contribute to averages depicting a doubly capped proteasome. Three rounds of reference-free two-dimensional classification, and particles were removed after each round. From an initial data set of 312,483 automatically selected particles, 93,679 were kept for the three-dimensional reconstruction. C2 symmetry was applied to one of the previously determined asymmetric negative-stained reconstructions to serve as a starting model for structure refinement. The reconstruction began using an angular increment of 25°, and iterated down to 0.6°. C2 symmetry was imposed during the reconstruction. Low-resolution Fourier amplitudes of the final map were dampened to match those of an experimental GroEL SAXS curve using the SPIDER software package[55].

The estimated resolution based on the FSC of the half-volumes at 0.5 was approximately 9 Å, although a local resolution calculation using the "blocres" function in the Bsoft package[56] indicated a range of resolutions within the density. The majority of the core particle subunits and the AAA+ ATPases were resolved to between 7- and 8-Å resolution, whereas the non-ATPase subunits in the regulatory particle ranged from 8- to 12-Å resolution (Supplementary Fig. 7). Notably, Rpn1 and the ubiquitin receptors Rpn10 and Rpn13 were the lowest resolution features of the holoenzyme. To filter the low-resolution portions of the map properly, without destroying the details of the better-ordered features, a resolution-driven adaptive localized low-pass filter was applied to the final volume (G. Cardone, personal communication).

The segmentation analysis was manually performed using the "Volume Tracer" tool in the UCSF Chimera visualization software[47]. This software was additionally used to perform all rigid-body fitting of crystal structures into the holoenzyme cryoelectron microscopy density, as well as to generate all renderings for figure images.

48. Sone, T., Saeki, Y., Toh-e, A. & Yokosawa, H. Sem1p is a novel subunit of the 26 S proteasome from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279,** 28807–28816 (2004).
49. Kim, H. C. & Huibregtse, J. M. Polyubiquitination by HECT E3s and the determinants of chain type specificity. *Mol. Cell. Biol.* **29,** 3307–3318 (2009).
50. Mallick, S. P., Carragher, B., Potter, C. S. & Kriegman, D. J. ACE: automated CTF estimation. *Ultramicroscopy* **104,** 8–29 (2005).
51. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
52. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166,** 205–213 (2009).
53. Sorzano, C. O. *et al.* XMIPP: a new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* **148,** 194–204 (2004).
54. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116,** 17–24 (1996).
55. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
56. Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157,** 3–18 (2007).

LETTER

A tidally distorted dwarf galaxy near NGC 4449

R. M. Rich[1,2], M. L. M. Collins[3], C. M. Black[1], F. A. Longstaff[2,4], A. Koch[5], A. Benson[6] & D. B. Reitzel[1,7]

NGC 4449 is a nearby Magellanic irregular starburst galaxy[1] with a B-band absolute magnitude of −18 and a prominent, massive, intermediate-age nucleus[2] at a distance from Earth of 3.8 megaparsecs (ref. 3). It is wreathed in an extraordinary neutral hydrogen (H I) complex, which includes rings, shells and a counter-rotating core, spanning ∼90 kiloparsecs (kpc; refs 1, 4). NGC 4449 is relatively isolated[5], although an interaction with its nearest known companion—the galaxy DDO 125, some 40 kpc to the south—has been proposed as being responsible for the complexity of its H I structure[6]. Here we report the presence of a dwarf galaxy companion to NGC 4449, namely NGC 4449B. This companion has a V-band absolute magnitude of −13.4 and a half-light radius of 2.7 kpc, with a full extent of around 8 kpc. It is in a transient stage of tidal disruption, similar to that of the Sagittarius dwarf[7] near the Milky Way. NGC 4449B exhibits a striking S-shaped morphology that has been predicted for disrupting galaxies[7,8] but has hitherto been seen only in a dissolving globular cluster[9]. We also detect an additional arc or disk ripple embedded in a two-component stellar halo, including a component extending twice as far as previously known, to about 20 kpc from the galaxy's centre.

We obtained deep imaging of NGC 4449 during the time period 29 May 2011 to 1 June 2011, in the course of commissioning a 0.7-m telescope[10] designed to study low-surface-brightness structures in the vicinity of other galaxies. We discovered the profoundly tidally distorted dwarf galaxy NGC 4449B, and recover an additional lower-luminosity arc or disk ripple, deeper in its halo (Fig. 1). Our photometry reveals that the original exponential halo terminates in a dumb-bell-shaped shelf, beyond which we measure a de Vaucouleurs $r^{1/4}$ surface brightness profile to 20 kpc (here r is the angular distance from the centre of NGC 4449). (Figs 1 and 2). Although we do not measure a change in the $g - r$ colour of the outer halo, the break in structure might imply a different origin for the $r^{1/4}$ component.

The lower-luminosity arc or ripple mentioned above is revealed by subtracting a model halo profile, but can also be clearly seen in unprocessed images (Fig. 1) and is also faintly visible and noted in earlier images[1]. However, we detect no additional components of a putative shell system as might be expected if this arc were part of a typical shell complex (even induced via an unusual collision geometry[11,12]). The arc or ripple might plausibly be a disk ripple, owing its origin to the interaction with NGC 4449B or a different event[13]. The ripple is
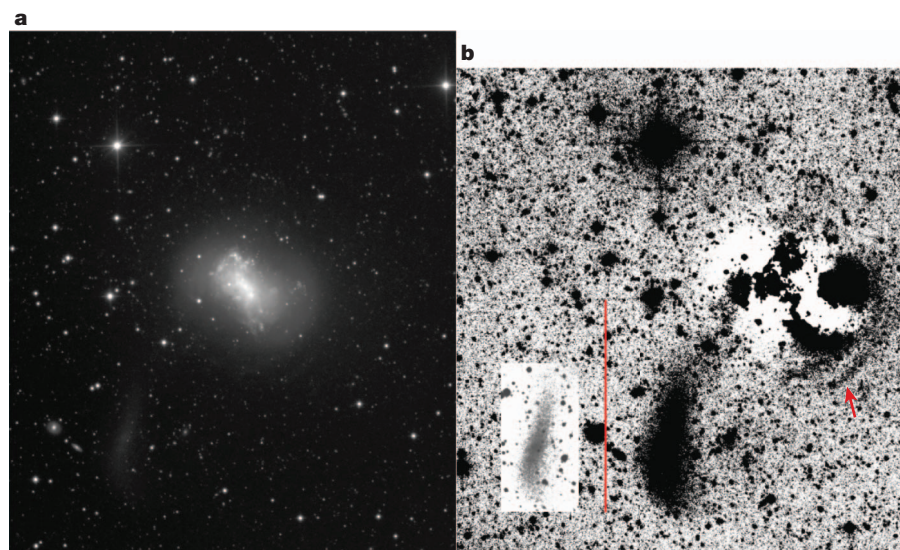

Figure 1 | Image and halo-subtracted imagery of NGC 4449. a, Positive image of NGC 4449 and NGC 4449B. (This is a 3.2-h luminance filter image from an STL 11000m camera, obtained using the Saturn Lodge 0.7-m Centurion[10] telescope.) **b,** Image (same scale as **a**) obtained by subtracting from **a** a model halo, using ELLIPSE within IRAF. Image shows detail of NGC 4449B, including a plume extended northwestwards towards the nucleus of NGC 4449. Inset, a softer stretch, revealing the S-shape distortion characteristic of a galaxy that has undergone tidal disruption. The fainter arc/disk ripple (indicated with a red arrow) can be easily seen to the southwest of the nucleus, and can be recovered as well in **a**. The arc/ripple lacks the edge or counter-arc structures characteristic in classical shells. A well defined shelf in the halo of NGC 4449 is evident in **a** and can be clearly seen in the surface brightness profile of Fig. 2. North is up, east is left. The red scale bar is 10 arcmin = 11.11 kpc, adopting a distance[1] of 3.82 Mpc for NGC 4449. Integration times were 3.2 h in a broadband Astrodon I-series Luminance (L) filter and 45 min each in the B and R filters. The wide L filter is a square pass filter spanning 400–700 nm that yields the deepest images, while the B and R filters are square pass filters covering 400–500 nm and 600–700 nm, respectively. Because NGC 4449 is within the SDSS footprint, we use catalogued SDSS stars to calibrate B and R to SDSS g and r photometry. The total r magnitude for NGC 4449B was obtained by calibrating the L filter to SDSS r with the total magnitude from ELLIPSE, after subtracting stellar sources from the footprint of the dwarf.

[1]Department of Physics and Astronomy, 430 Portola Plaza, Box 951547, University of California, Los Angeles, California 90095-1547, USA. [2]Polaris Observatory Association, Frazier Park, California 93225, USA. [3]Max-Planck-Institut für Astronomie, Königstuhl 17, Heidelberg D-69117, Germany. [4]UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles, California 90095-1481, USA. [5]Zentrum für Astronomie der Universitat Heidelberg, Landessternwarte, Königstuhl 12, 69117 D-Heidelberg, Germany. [6]Department of Astronomy, California Institute of Technology, MC 249-1, 1200 East California Boulevard, Pasadena, California 91125, USA. [7]Griffith Observatory, 2800 East Observatory Road, Los Angeles, California 90027, USA.
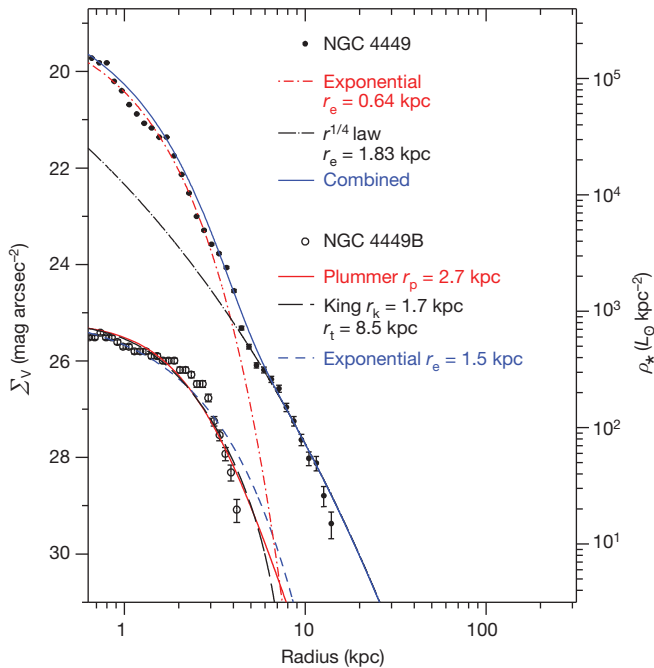
Figure 2 | Surface photometry of NGC 4449 and NGC 4449B. Surface brightness profiles (data points) and model fits (solid and broken lines) for the halo of NGC 4449 and for NGC 4449B. Σ_V is the surface brightness, ρ_* is the stellar surface density, and $L_\odot$ is solar luminosity. The halo of NGC 4449 exhibits a dumb-bell-shaped shelf (Fig. 1) at 5 kpc, coincident with the break in the surface brightness profile. The inner portion of NGC 4449 is best fitted by an exponential[5] with $r_e = 0.64$ kpc and the outer envelope by an $r^{1/4}$ law with $r_e = 1.83$ kpc that can be traced to 20 kpc radius (here r_e is the de Vaucouleurs half-light enclosed or effective radius). Beyond 3 kpc the halo colour is $g - r = 0.5$, similar to that of the dwarf, and we do not detect any change in $g - r$ colour at the shelf; position angle and ellipticity change at the boundary of the outer halo. The exponential portion may be related to the bar, while the outer halo may have an accretion origin. NGC 4449B has a Plummer half-light radius of $r_p = 2.7$ kpc, but we are unable to find any analytical profile that provides a good fit, consistent with a system undergoing tidal disruption[7]. (We include for information the attempted fit to the King profile, with scale radius r_k and tidal radius r_t.) Error bars, s.d.

2.6 kpc long with an r-band magnitude of 19.1 (corresponding to an absolute magnitude $M_r = -8.91 \pm 0.1$, faint even relative to Milky Way dwarfs); our halo subtraction uncovered no additional arcs or candidate dwarfs.

NGC 4449B (also known as NGC 4449 J1228.8+4357.8) lies at a projected distance of 9 kpc from the nucleus of NGC 4449 at right ascension (2000) 12 h 28 min 45 s, declination (2000) +43° 57′ 44″, which is 39.8″ E and 8′ 07.2″ S of the nucleus. The halo model subtraction in Fig. 1 reveals the complete extent of the dwarf, including a plume of faint emission extending northwest towards the nucleus. We derive $r = 14.47 \pm 0.1$ mag and adopting $(m - M)_o = 27.91$ (ref. 3), calculate $M_r = -13.44 \pm 0.1$ mag. The colour of $g - r = 0.48 \pm 0.22$ is that of a non-star forming dwarf galaxy, consistent with the lack of structure at this location in published H I maps[1,4] and non-detection on archival GALEX[14] satellite images of 3,283 s duration in the near-ultraviolet and 1,685 s in the far-ultraviolet. It is noteworthy that the position of NGC 4449B misses by >3 kpc any catalogued[1] H I cloud or shell, although the position falls on the southern edge of the main H I ring[4]. Although GALEX far-ultraviolet imaging usually detects stellar emission in H I tidal tails[15], the extensive H I complex near NGC 4449 is surprisingly undetected in the GALEX imagery. DDO 125 ($M_V = -15.57$) is detected easily in H I and GALEX near- and far-ultraviolet[16], but is disjoint from the main H I complex and, lying 31 kpc to the south of NGC 4449B, is uninvolved with the dwarf.

Optical emission from NGC 4449B is traceable to a full extent of $2.65′ \times 4.0′$ or 2.9×7.4 kpc in extent; we calculate a stellar mass[17] of 3.5×10^7 solar masses.

The S-shaped morphology qualitatively resembles a model[7] that places a dwarf galaxy on a highly eccentric orbit, and tracks its evolution from encounter through the close approach of the dwarf galaxy (the 'impactor') to the nucleus of the primary galaxy. Such extreme orbits are proposed for other systems: for example, And XIV has kinematics consistent with a first-encounter plunge orbit with M31 (ref. 18). NGC 4449B appears to fall somewhere between time steps 2 and 3 (as shown in figure 1 of the simulation reported in ref. 7), or ~5–10 crossing times past closest approach between the dwarf and nucleus, a point in the simulation where most of the dark matter still remains bound to the dwarf. The encounter geometry that we observe for NGC 4449B is also consistent with the modelled timescale[7] over which the dwarf evolves from a compact spheroid to the 'nucleus and tails' morphology prominent in the disrupted globular cluster Palomar 5[8,9]. The width of the dwarf galaxy's central region ($28″ = 516$ pc) constrains a length scale for the pre-encounter system even though we do not discern the nucleus. If we adopt an effective radius of ~200 pc for the pre-tidal dwarf and internal velocity dispersion $\sigma = 10$ km s^{-1}, figure 1 of the simulation[7] gives a morphology evolution timescale $t - t_p \approx 10 R_c \approx 2 \times 10^8$ yr, where $t - t_p$ is the time since pericentre, or closest encounter, and R_c is the core radius of the dwarf. Assuming that the orbit plane is roughly perpendicular to our sightline (based on the S morphology), we find a timescale of 10^8 yr to traverse 9 kpc at 100 km s^{-1}, in good agreement with the simulation timescale.

We speculate that NGC 4449B is on its first encounter with NGC 4449 and experienced a close passage near the nucleus of NGC 4449. This conclusion is supported by the morphology of NGC 4449B, the plume pointing at the nucleus, and the approximate agreement with the structure and timescales of the simulation[7]. The calculated timescales would not contradict the hypothesis that the NGC 4449B encounter played a role in igniting the present epoch of star formation in NGC 4449. The simulation[7] also predicts that a morphology resembling that of NGC 4449B survives only for a relatively brief interval of ~5 crossing times, or ~10^8 yr, which may, along with its low surface brightness, account for its uniqueness.

1. Hunter, D. A., van Woerden, H. & Gallagher, J. S. Neutral hydrogen and star formation in the irregular galaxy NGC 4449. *Astron. J.* **118,** 2184–2210 (1999).
2. Böker, T. *et al.* A young stellar cluster in the nucleus of NGC 4449. *Astron. J.* **121,** 1473–1481 (2001).
3. Annibali, R. *et al.* Starbursts in the local Universe: New Hubble Space Telescope Advanced Camera for Surveys observations of the irregular galaxy NGC 4449. *Astron. J.* **135,** 1900–1916 (2008).
4. Hunter, D. *et al.* The nature of extended HI gas around NGC 449: the Dr. Jekyll/Mr. Hyde of irregular galaxies. *Astrophys. J.* **495,** L47–L50 (1998).
5. de Vaucouleurs, G. in *Galaxies and the Universe* (eds Sandage, A., Sandage, M. & Kristian, J.) 557–597 (Univ. Chicago Press, 1975).
6. Theis, C. & Koble, S. Multi-method-modeling of interacting galaxies. I. A unique scenario for NGC 4449? *Astron. Astrophys.* **370,** 365–383 (2001).
7. Peñarrubia, J. *et al.* The signature of Galactic tides in Local Group dwarf spheroidals. *Astrophys. J.* **698,** 222–232 (2009).
8. Capuzzo Dolcetta, R., di Matteo, P. & Miocchi, P. Formation and evolution of clumpy tidal tails around globular clusters. *Astron. J.* **129,** 1906–1921 (2005).
9. Odenkirchen, M. *et al.* The extended tails of Palomar 5: a 10° arc of globular cluster tidal debris. *Astron. J.* **126,** 2385–2407 (2003).
10. Brosch, N. *et al.* The Centurion 18 telescope of the Wise Observatory. *Astrophys. Space Sci.* **314,** 163–176 (2008).
11. Schombert, J. & Wallin, J. F. Arp 227: a case for shells without mergers? *Astron. J.* **94,** 300–305 (1987).
12. Wallin, J. F. & Struck-Marcell, C. A collisional model for the formation of ripples in early-type disk galaxies. *Astron. J.* **96,** 1850–1860 (1988).
13. Purcell, C. W. *et al.* The Sagittarius impact as an architect of spirality and outer rings in the Milky Way. *Nature* **477,** 301–303 (2011).
14. Martin, D. C. *et al.* The Galaxy Evolution Explorer: a space ultraviolet survey mission. *Astrophys. J.* **619,** L1–L6 (2005).
15. Neff, S. G. *et al.* Ultraviolet emission from stellar populations within tidal tails: catching the youngest galaxies in formation? *Astrophys. J.* **619,** L91–L94 (2005).
16. Dale, D. A. *et al.* The Spitzer Local Volume Legacy: survey description and infrared photometry. *Astrophys. J.* **703,** 517–556 (2009).

17. Wang, J. L. *et al.* Luminous infrared galaxies in the local Universe. *Astrophys. J.* **649,** 722–729 (2006).
18. Majewski, S. R. *et al.* Discovery of Andromeda XIV: a dwarf spheroidal dynamical rogue in the Local Group? *Astrophys. J.* **670,** L9–L12 (2007).

Author Contributions R.M.R. conceived the project, obtained the data and coordinated the activity. M.L.M.C. fitted the surface photometry of NGC 4449 and NGC 4449B. C.M.B., F.M.L. and D.B.R. analysed and reduced various aspects of the dataset, including the surface photometry. F.A.L. and R.M.R. implemented the Saturn Lodge 0.7-m telescope and detector system. A.K. provided insight on dwarf galaxies and discussion, and A.B. provided a discussion of theoretical implications.

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

LETTER

Two Earth-sized planets orbiting Kepler-20

Francois Fressin[1], Guillermo Torres[1], Jason F. Rowe[2], David Charbonneau[1], Leslie A. Rogers[3], Sarah Ballard[1], Natalie M. Batalha[4], William J. Borucki[2], Stephen T. Bryson[2], Lars A. Buchhave[5,6], David R. Ciardi[7], Jean-Michel Désert[1], Courtney D. Dressing[1], Daniel C. Fabrycky[8], Eric B. Ford[9], Thomas N. Gautier III[10], Christopher E. Henze[2], Matthew J. Holman[1], Andrew Howard[11], Steve B. Howell[2], Jon M. Jenkins[12], David G. Koch[2], David W. Latham[1], Jack J. Lissauer[2], Geoffrey W. Marcy[11], Samuel N. Quinn[1], Darin Ragozzine[1], Dimitar D. Sasselov[1], Sara Seager[3], Thomas Barclay[2], Fergal Mullally[12], Shawn E. Seader[12], Martin Still[2], Joseph D. Twicken[12], Susan E. Thompson[12] & Kamal Uddin[13]

Since the discovery of the first extrasolar giant planets around Sun-like stars[1,2], evolving observational capabilities have brought us closer to the detection of true Earth analogues. The size of an exoplanet can be determined when it periodically passes in front of (transits) its parent star, causing a decrease in starlight proportional to its radius. The smallest exoplanet hitherto discovered[3] has a radius 1.42 times that of the Earth's radius ($R_\oplus$), and hence has 2.9 times its volume. Here we report the discovery of two planets, one Earth-sized ($1.03R_\oplus$) and the other smaller than the Earth ($0.87R_\oplus$), orbiting the star Kepler-20, which is already known to host three other, larger, transiting planets[4]. The gravitational pull of the new planets on the parent star is too small to measure with current instrumentation. We apply a statistical method to show that the likelihood of the planetary interpretation of the transit signals is more than three orders of magnitude larger than that of the alternative hypothesis that the signals result from an eclipsing binary star. Theoretical considerations imply that these planets are rocky, with a composition of iron and silicate. The outer planet could have developed a thick water vapour atmosphere.

Precise photometric time series gathered by the Kepler spacecraft[5] over eight observation quarters (670 days) have revealed five periodic transit-like signals in the G8 star Kepler-20, of which three have been previously reported as arising from planetary companions[4] (Kepler-20 b, Kepler-20 c and Kepler-20 d, with radii of $1.91R_\oplus$, $3.07R_\oplus$ and $2.75R_\oplus$, and orbital periods of 3.7 days, 10.9 days and 77.6 days, respectively). The two, much smaller, signals described here recur with periods of 6.1 days (Kepler-20 e) and 19.6 days (Kepler-20 f) and exhibit flux decrements of 82 parts per million (p.p.m.) and 101 p.p.m. (Fig. 1), corresponding to planet sizes of $0.868^{+0.074}_{-0.096}R_\oplus$ (potentially smaller than the radius of Venus, $R_{Venus} = 0.95R_\oplus$) and $1.03^{+0.10}_{-0.13}R_\oplus$. The properties of the star are listed in Table 1.

A background star falling within the same photometric aperture as the target and eclipsed by another star or by a planet produces a signal that, when diluted by the light of the target, may appear similar to the observed transits in both depth and shape. The Kepler-20 e and Kepler-20 f signals have undergone careful vetting to rule out certain false positives that might manifest themselves through different depths of odd- and even-numbered transit events, or displacements in the centre of light correlated with the flux variations[6]. High-spatial-resolution imaging shows no neighbouring stars capable of causing the signals[4]. Radial-velocity measurements based on spectroscopic observations with the Keck I telescope rule out stars or brown dwarfs orbiting the primary star, but they are not sensitive enough to detect the acceleration of the star due to these putative planetary companions[4].
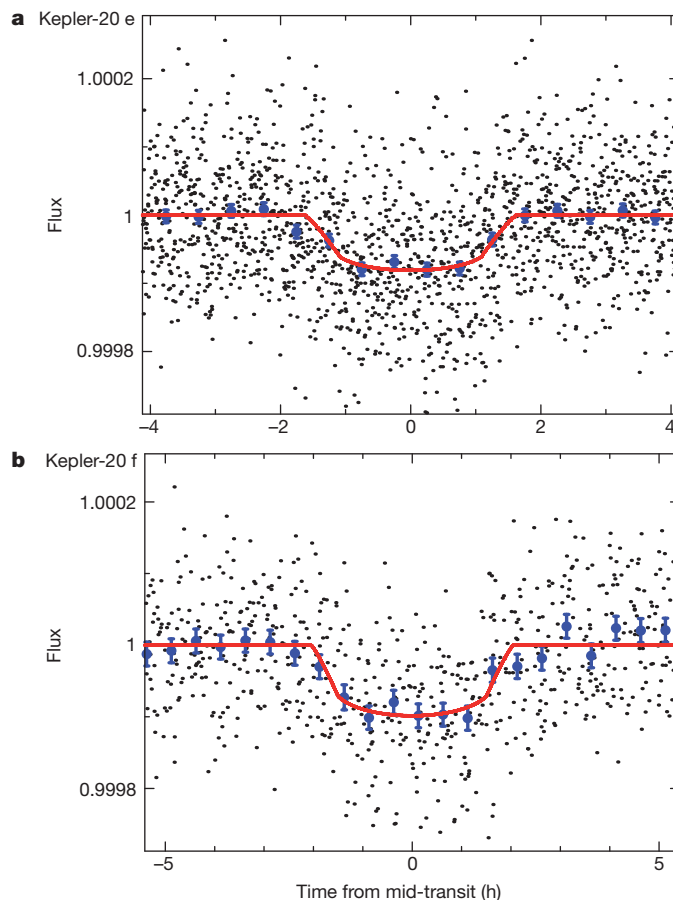

Figure 1 | Transit light curves. Kepler-20 (also designated as KOI 070, KIC 6850504 and 2MASS J19104752+4220194) is a G8V star of Kepler magnitude 12.497 and celestial coordinates right ascension α = 19 h 10 min 47.5 s and declination δ = +42° 20′ 19.38′′. The stellar properties are listed in Table 1. The photometric data used for this work were gathered between 13 May 2009 and 14 March 2011 (quarter 1 to quarter 8), and comprise 29,595 measurements at a cadence of 29.426 min (black dots). The Kepler photometry phase-binned in 30-min intervals (blue dots with 1σ standard error of the mean (s.e.m.) error bars) for Kepler-20 e (a) and Kepler-20 f (b) is displayed as a function of time, with the data detrended[4] and phase-folded at the period of the two transits. Transit models (red curves) smoothed to the 29.426-min cadence are overplotted. These two signals are unambiguously detected in each of the eight quarters of Kepler data, and have respective signal-to-noise ratios of 23.6 and 18.5, which cannot be due to stellar variability, data treatment or aliases from the other transit signals[4].

[1]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [2]NASA Ames Research Center, Moffett Field, California 94035, USA. [3]Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Department of Physics and Astronomy, San Jose State University, San Jose, California 95192, USA. [5]Niels Bohr Institute, University of Copenhagen, DK-2100, Copenhagen, Denmark. [6]Center for Star and Planet Formation, University of Copenhagen, DK-1350, Copenhagen, Denmark. [7]NASA Exoplanet Science Institute/California Institute of Technology, Pasadena, California 91125, USA. [8]Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA. [9]Astronomy Department, University of Florida, Gainesville, Florida 32111, USA. [10]Jet Propulsion Laboratory/California Institute of Technology, Pasadena, California 91109, USA. [11]Department of Astronomy, University of California, Berkeley, California 94720, USA. [12]SETI Institute/NASA Ames Research Center, Moffett Field, California 94035, USA. [13]Orbital Sciences Corporation/NASA Ames Research Center, Moffett Field, California 94035, USA.

Table 1 | Stellar and planetary parameters for Kepler-20

Stellar properties	Kepler-20
Effective temperature, T_{eff}	$5,466 \pm 93$ K
Surface gravity $\log[g \, (cm \, s^{-2})]$	4.443 ± 0.075
Metallicity [Fe/H]	0.02 ± 0.04
Projected rotational velocity, $v\sin i$	0.4 ± 0.5 km s^{-1}
Stellar mass, M_s	$(0.912 \pm 0.035) M_\odot$
Stellar radius, R_s	$0.944^{+0.060}_{-0.095} R_\odot$
Stellar density, ρ_s	1.51 ± 0.39 g cm^{-3}
Luminosity, L_s	$(0.853 \pm 0.093) L_\odot$
Distance, D	290 ± 30 pc

Planetary parameters	Kepler-20 e (KOI 070.04)	Kepler-20 f (KOI 070.05)
Orbital period, P	6.098493 ± 0.000065 days	19.57706 ± 0.00052 days
Time of centre of transit, T_c	$2,454,968.9336 \pm 0.0039$ BJD	$2,454,968.219 \pm 0.011$ BJD
Eccentricity, e	<0.28	<0.32
Planet/star radius ratio, R_p/R_s	$0.00841^{+0.00035}_{-0.00054}$	$0.01002^{+0.00063}_{-0.00077}$
Scaled semi-major axis, a/R_s	$11.56^{+0.21}_{-0.29}$	$25.15^{+0.47}_{-0.63}$
Impact parameter, b	$0.630^{+0.070}_{-0.053}$	$0.727^{+0.054}_{-0.053}$
Orbital inclination, i	$87.50^{+0.33}_{-0.34}$ degrees	$88.68^{+0.14}_{-0.17}$
Planetary radius, R_p	$0.868^{+0.074}_{-0.096} R_\oplus$	$1.03^{+0.10}_{-0.13} R_\oplus$
Planetary mass, M_p	$<3.08 M_\oplus$ (spectroscopic limit); $0.39 M_\oplus < M_p < 1.67 M_\oplus$ (theoretical considerations)	$<14.3 M_\oplus$ (spectroscopic limit); $0.66 M_\oplus < M_p < 3.04 M_\oplus$ (theoretical considerations)
Planetary equilibrium temperature, T_{eq}	$1,040 \pm 22$ K	705 ± 16 K

$M_\odot$, mass of the Sun; $R_\odot$, radius of the Sun. The effective temperature, surface gravity, metallicity and projected rotational velocity of the star were spectroscopically determined[21] from our Keck/HIRES spectrum. With these values and the use of stellar evolution models[22], we derived the stellar mass, radius, luminosity, distance and mean density. The transit and orbital parameters (period, time of centre of transit, radius ratio, scaled semi-major axis, impact parameter and orbital inclination) for the five planets in the Kepler-20 system were derived jointly based on the Kepler photometry using a Markov-chain Monte Carlo procedure with the mean stellar density as a prior[4]. The parameters above are based on an eccentricity constraint: that the orbits do not cross each other. After calculating the above parameters, we performed a suite of N-body integrations to estimate the maximum eccentricity for each planet consistent with dynamical stability[4]. The N-body simulations provide similar constraints on the maximum eccentricity and justify the assumption of non-crossing orbits. The planetary spectroscopic mass limits are the 2σ upper limits determined from the radial velocity analysis based on the Keck radial velocity measurements. Planet interior models provide further useful constraints on mass and inferences on composition[23]. Assuming Kepler-20 e and Kepler-20 f are rocky bodies comprised of iron and silicates, and considering the uncertainty on their radii, the planet masses are constrained to be $0.39 M_\oplus < M_p < 1.67 M_\oplus$ for Kepler-20 e, and $0.66 M_\oplus < M_p < 3.04 M_\oplus$ for Kepler-20 f. The lower and upper mass bounds are set by a homogeneous silicate composition and by the densest composition from a model of planet formation with collisional mantle stripping[24]. The planet equilibrium temperatures assume an Earth-like Bond albedo of 0.3, isotropic redistribution of heat for reradiation, and a circular orbit. The errors in these quantities reflect only the uncertainty due to the stellar luminosity.

To establish the planetary nature of these signals with confidence we must establish that the planet hypothesis is much more likely than that of a false positive. For this we used the BLENDER procedure[7–9], a technique used previously to validate the three smallest known exoplanets, Kepler-9 d (ref. 8), Kepler-10 b (ref. 3), and CoRoT-7 b (ref. 10). The latter two were also independently confirmed with Doppler studies[3,11]. We used BLENDER to identify the allowed range of properties of blends that yield transit light curves matching the photometry of Kepler-20 e and Kepler-20 f. We varied as free parameters the brightness and spectral type (of the stars) or the size (for the planetary companions), the impact parameter, the eccentricity and the longitude of periastron. We simulated large numbers of these scenarios and compared the resulting light curves with the observations. We ruled out fits significantly worse (at the 3σ level, or greater) than that of a true transiting planet around the target, and we tabulated all remaining scenarios that were consistent with the Kepler light curves.

We assessed the frequency of blend scenarios through a Monte Carlo experiment in which we randomly drew 8×10^5 background main-sequence stars from a Galactic structure model[12] in a one-square-degree area around the target, and assigned them each a stellar or planetary transiting companion based on the known properties of eclipsing binaries[13] and the size distribution of planet candidates as determined from the Kepler mission itself[14]. We counted how many satisfy the constraints from BLENDER as well as observational constraints from our high-resolution imaging observations and centroid motion analysis[4], and made use of estimates of the frequencies of larger transiting planets and eclipsing binaries (see Fig. 2). In this way we estimated a blend frequency of background stars transited by larger planets of 2.1×10^{-7} and a blend frequency of background eclipsing binaries of 3.1×10^{-8}, yielding a total of 2.4×10^{-7} for Kepler-20 e. Similarly, $4.5 \times 10^{-7} + 1.26 \times 10^{-6}$ yields a total blend frequency of 1.7×10^{-6} for Kepler-20 f.

Another type of false positive consists of a planet transiting another star physically associated with the target star. To assess their frequency we simulated 10^6 such companions in randomly oriented orbits around the target, based on known distributions of periods, masses

and eccentricities of binary stars[13]. We excluded those that would have been detected in our high-resolution imaging or that would have an overall colour inconsistent with the observed colour of the target, measured between the Sloan r band (12.423 ± 0.017; ref. 14) and the Warm Spitzer 4.5-μm band (10.85 ± 0.02; ref. 4). We used BLENDER to determine the range of permitted sizes for the planets as a function of stellar mass, and to each we assigned an eccentricity drawn from the known distribution for close-in exoplanets[15]. The frequency of blends of this kind is 5.0×10^{-7} for Kepler-20 e, and 3.5×10^{-6} for Kepler-20 f. Summing the contributions of background stars and physically bound stars, we find a total blend frequency of 7.4×10^{-7} for Kepler-20 e and 5.2×10^{-6} for Kepler-20 f.

We estimated the a priori chance that Kepler-20 has a planet of a similar size as implied by the signal using a 3σ criterion as in BLENDER, by calculating the fraction of Kepler objects of interest in the appropriate size range. We counted 102 planet candidates in the radius range allowed by the photometry of Kepler-20 e, and 228 for Kepler-20 f. We made the assumption that only 10% of them are planets (which is conservative in comparison to other estimates of the false positive rate that are an order of magnitude larger[16]). From numerical simulations, we determined the fraction of the 190,186 Kepler targets for which planets of the size of Kepler-20 e and Kepler-20 f could have been detected (17.4% and 16.0%, respectively), using actual noise levels. We then calculated the planet priors (the a priori chance of a planet) to be $(102 \times 10\%)/(190,186 \times 17.4\%) = 3.1 \times 10^{-4}$ for Kepler-20 e, and $(228 \times 10\%)/(190,186 \times 16.0\%) = 7.5 \times 10^{-4}$ for Kepler-20 f. These priors ignore the fact that Kepler-20 is more likely to have a transiting planet at the periods of Kepler-20 e and Kepler-20 f than a random Kepler target, because the star is already known to have three other transiting planets, and multi-planet systems tend to be coplanar[17]. When accounting for this using the procedure described for the validation of Kepler-18 d (ref. 18), we find that the flatness of the system increases the transit probability from 7.7% to 63% for Kepler-20 e, and from 3.7% to 35% for Kepler-20 f. With this coplanarity boost, the planet priors increase to 2.5×10^{-3} for Kepler-20 e and 7.1×10^{-3} for Kepler-20 f. Comparing this with the total
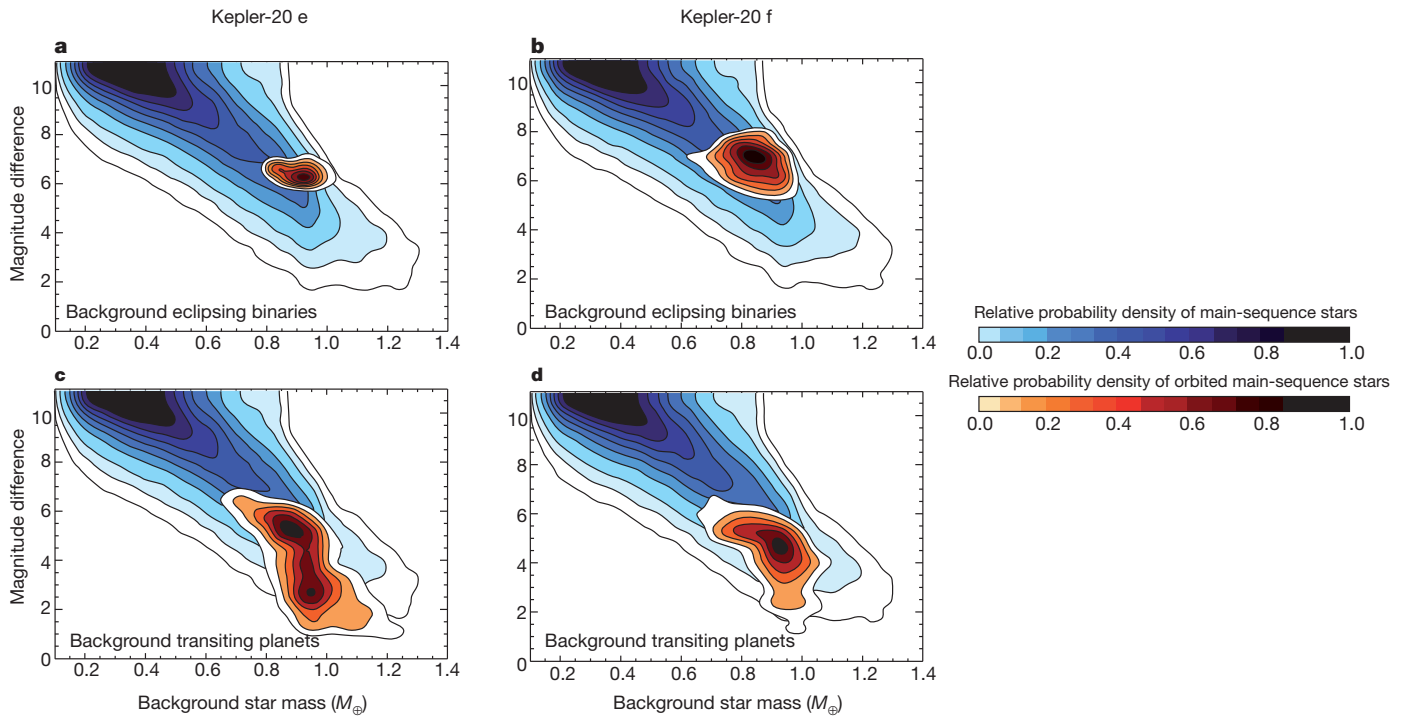
Figure 2 | Density map of stars in the background of Kepler-20. The blue-shaded contours correspond to main-sequence star counts from the Besancon model in the vicinity of Kepler-20, as a function of stellar mass and magnitude difference in the Kepler passband compared to Kepler-20. The red-shaded contours represent the fractions of those stars orbited by another smaller star (**a** and **c**) or by a planet (**b** and **d**) with sizes such that the resulting light curves mimic the transit signals for Kepler-20 e and Kepler-20 f. The displacement of the blue and red contours in magnitude and spectral type results in very small fractions of the simulated background stars being viable false positives for Kepler-20 e (1.6% when transited by a planet, and 0.1% when transited by a smaller star). We obtained similar results for Kepler-20 f (2.1% when transited by a planet, and 3.1% when transited by a smaller star). Most of these background stars have masses (spectral types) near that of the target, and are two to seven magnitudes fainter. The above fractions are further reduced because background stars able to match the signals but that are also bright enough and at large enough angular separation from the target would have been detected in our imaging observations and/or centroid motion analysis. Finally, to obtain the blend frequencies we scaled these estimates to account for the fraction of background stars expected to have transiting planets (1.29%, the ratio between the number of Kepler objects of interest and the total number of Kepler targets[25]) or stellar companions (0.79% based on the statistics of detached eclipsing binaries in the Kepler field[26]). We examined non-main-sequence stars as alternatives to either object of the blend eclipsing pair, but found that they either do not reproduce the observed transit shape well enough, or are much less common (<1%) than main-sequence blends.
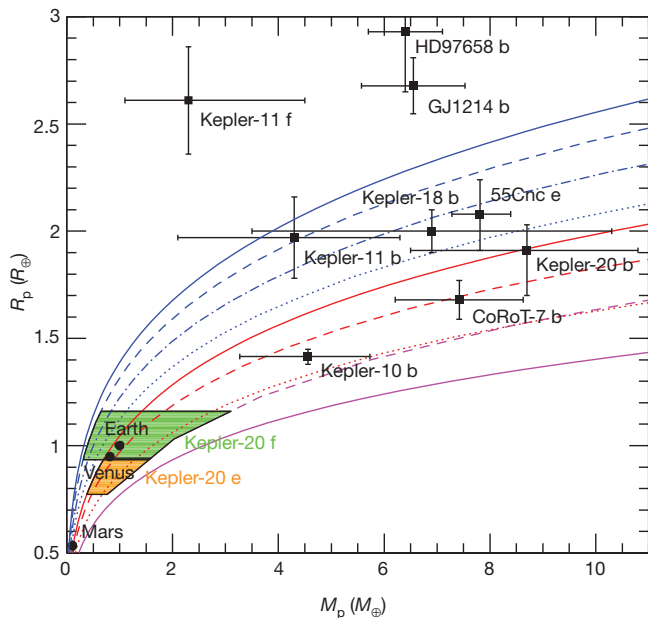

Figure 3 | Mass versus radius relation for small planets. Kepler-20 e and Kepler-20 f theoretical mass and observed radius ranges (1σ) are plotted as orange- and green-shaded areas, while the other transiting planets with dynamically determined masses are plotted in black, with 1σ error bars. The curves are theoretical constant-temperature mass–radius relations[27]. The solid lines are homogeneous compositions: water ice (solid blue), $MgSiO_3$ perovskite (solid red), and iron (magenta). The non-solid lines are mass–radius relations for differentiated planets: 75% water ice, 22% silicate shell and 3% iron core (dashed blue); Ganymede-like with 45% water ice, 48.5% silicate shell and 6.5% iron core (dot-dashed blue); 25% water ice, 52.5% silicate shell and 22.5% iron core (dotted blue); approximately Earth-like with 67.5% silicate mantle and 32.5% iron core (dashed red); and Mercury-like with 30% silicate mantle and 70% iron core (dotted red). The dashed magenta curve corresponds to the density limit from a formation model[24]. The minimum density for Kepler-20 e corresponds to a 100% silicate composition, because this highly irradiated small planet could not keep a water reservoir. The minimum density for Kepler-20 f follows the 75% water-ice composition, representative of the maximum water content of comet-like mix of primordial material in our Solar System[28].

blend frequencies, we find that the hypothesis of an Earth-size planet for Kepler-20 e is 3,400 times more likely than that of a false positive, and 1,370 times for Kepler-20 f. Both of these odds ratios are sufficiently large to validate these objects with very high confidence as Earth-size exoplanets.

With measured radii close to that of the Earth, Kepler-20 e and Kepler-20 f could have bulk compositions similar to Earth's (approximately 32% iron core, 68% silicate mantle by mass; see Fig. 3), although in the absence of a measured mass the composition cannot be determined unambiguously. We infer that the two planets almost certainly do not have a hydrogen-dominated gas layer, because this would readily be lost to atmospheric escape owing to their small sizes and high equilibrium temperatures. A planet with several per cent water content by mass surrounding a rocky interior is a possibility for Kepler-20 f, but not for Kepler-20 e. If the planets formed beyond the snowline from a comet-like mix of primordial material and then migrated closer to the star, Kepler-20 f could retain its water reservoir for several billion years in its current orbit, but the more highly irradiated Kepler-20 e would probably lose its water reservoir to extreme-ultraviolet-driven escape within a few hundred million years[19]. In this scenario, Kepler-20 f could develop a thick vapour atmosphere with a mass of $0.05M_\oplus$ that would protect the planet surface from further vaporization[20]. From the theoretical mass estimates in Table 1, we infer the semi-amplitude of the stellar radial velocity to be between $15\,\mathrm{cm\,s^{-1}}$ and $62\,\mathrm{cm\,s^{-1}}$ for Kepler-20 e and between $17\,\mathrm{cm\,s^{-1}}$ and $77\,\mathrm{cm\,s^{-1}}$ for Kepler-20 f. Such signals could potentially be detectable in the next few years, and would constrain the composition of the two planets.

1. Latham, D. W., Stefanik, R. P., Mazeh, T., Mayor, M. & Burki, G. The unseen companion of HD114762—a probable brown dwarf. *Nature* **339**, 38–40 (1989).
2. Mayor, M. & Queloz, D. A Jupiter-mass companion to a solar-type star. *Nature* **378**, 355–359 (1995).
3. Batalha, N. M. et al. Kepler's first rocky planet: Kepler-10b. *Astrophys. J.* **729**, 27 (2011).
4. Gautier, T. N. et al. A Sun-like star with three sub-Neptune exoplanets and two Earth-size candidates. *Astrophys. J.* (submitted).
5. Koch, D. G. et al. Kepler mission design, realized photometric performance, and early science. *Astrophys. J.* **713**, L79–L86 (2010).
6. Bryson, S. T. et al. The Kepler pixel response function. *Astrophys. J.* **713**, L97–L102 (2010).
7. Torres, G., Konacki, M., Sasselov, D. D. & Jha, S. Testing blend scenarios for extrasolar transiting planet candidates. I. OGLE-TR-33: a false positive. *Astrophys. J.* **614**, 979–989 (2004).
8. Torres, G. et al. Modeling Kepler transit light curves as false positives: rejection of blend scenarios for Kepler-9, and validation of Kepler-9 d, a super-earth-size planet in a multiple system. *Astrophys. J.* **727**, 24 (2011).
9. Fressin, F. et al. Kepler-10c, a 2.2-Earth radius transiting planet in a multiple system. *Astrophys. J. Suppl.* **197**, 5 (2011).
10. Fressin, F. et al. Spitzer Infrared observations and independent validation of the transiting Super-Earth CoRoT-7 b. *Astrophys. J.* **669**, 1279–1297 (2011).
11. Queloz, D. et al. The CoRoT-7 planetary system: two orbiting super-Earths. *Astron. Astrophys.* **506**, 303–319 (2009).
12. Robin, A. C., Derrière, S. & Picaud, S. A synthetic view on structure and evolution of the Milky Way. *Astron. Astrophys.* **409**, 523–540 (2003).
13. Raghavan, D. et al. A survey of stellar families: multiplicity of solar-type stars. *Astrophys. J. Suppl.* **190**, 1–42 (2010).
14. Brown, T. M., Latham, D. W., Everett, M. E. & Esquerdo, G. A. Kepler input catalog: photometric calibration and stellar classification. *Astrophys. J.* **142**, 112 (2011).
15. Schneider, J., Dedieu, C., Le Sidaner, P., Savalle, R. & Zolotukhin, I. Defining and cataloging exoplanets: the exoplanet.eu database. *Astron. Astrophys.* **532**, A79 (2011).
16. Morton, T. D. & Johnson, J. A. On the low false positive probabilities of Kepler planet candidates. *Astrophys. J.* **738**, 170 (2011).
17. Lissauer, J. J. et al. Architecture and dynamics of Kepler's candidate multiple planet systems. *Astrophys. J., Suppl.* **197**, 8 (2011).
18. Cochran, W. D. et al. Kepler 18-b, c, and d: a system of three planets confirmed by transit timing variations, lightcurve validation, Spitzer photometry and radial velocity measurements. *Astrophys. J. Suppl.* **197**, 7 (2011).
19. Selsis, F. et al. Could we identify hot ocean-planets with CoRoT, Kepler and Doppler velocimetry? *Icarus* **191**, 453–468 (2007).
20. Kuchner, M. J. Volatile-rich Earth-mass planets in the habitable zone. *Astrophys. J.* **596**, L105–L108 (2003).
21. Valenti, J. A. & Piskunov, N. Spectroscopy made easy: a new tool for fitting observations with synthetic spectra. *Astron. Astrophys.* **118** (Suppl.), 595–603 (1996).
22. Yi, S. et al. Toward better age estimates for stellar populations: the Y^2 isochrones for solar mixture. *Astrophys. J., Suppl.* **136**, 417–437 (2001).
23. Rogers, L. A., Bodenheimer, P., Lissauer, J. J. & Seager, S. Formation and structure of low-density exo-Neptunes. *Astrophys. J.* **738**, 59 (2011).
24. Marcus, R. A., Sasselov, D., Hernquist, L. & Stewart, S. T. Minimum radii of super-Earths: constraints from giant impacts. *Astrophys. J.* **712**, L73–L76 (2010).
25. Borucki, W. J. et al. Characteristics of planetary candidates observed by Kepler. II. Analysis of the first four months of data. *Astrophys. J.* **736**, 19 (2011).
26. Slawson, R. W. et al. Kepler eclipsing binary stars. II. 2165 eclipsing binaries in the second data release. *Astrophys. J.* **142**, 160 (2011).
27. Seager, S., Kuchner, M., Hier-Majumder, C. A. & Militzer, B. Mass-radius relationships for solid exoplanets. *Astrophys. J.* **669**, 1279–1297 (2007).
28. Marcus, R. A., Sasselov, D., Stewart, S. T. & Hernquist, L. Water/icy super-Earths: giant impacts and maximum water content. *Astrophys. J.* **719**, L45–L49 (2010).

LETTER

Electromagnetically induced transparency with resonant nuclei in a cavity

Ralf Röhlsberger[1], Hans-Christian Wille[1], Kai Schlage[1] & Balaram Sahoo[1]

The manipulation of light–matter interactions by quantum control of atomic levels has had a profound impact on optical sciences. Such manipulation has many applications, including nonlinear optics at the few-photon level[1–3], slow light[4,5], lasing without inversion[6–8] and optical quantum information processing[9,10]. The critical underlying technique is electromagnetically induced transparency, in which quantum interference between transitions in multilevel atoms[11–15] renders an opaque medium transparent near an atomic resonance. With the advent of high-brilliance, accelerator-driven light sources such as storage rings or X-ray lasers, it has become attractive to extend the techniques of optical quantum control to the X-ray regime[16,17]. Here we demonstrate electromagnetically induced transparency in the regime of hard X-rays, using the 14.4-kiloelectronvolt nuclear resonance of the Mössbauer isotope iron-57 (a two-level system). We exploit cooperative emission from ensembles of the nuclei, which are embedded in a low-finesse cavity and excited by synchrotron radiation. The spatial modulation of the photonic density of states in a cavity mode leads to the coexistence of superradiant and subradiant states of nuclei, respectively located at an antinode and a node of the cavity field. This scheme causes the nuclei to behave as effective three-level systems, with two degenerate levels in the excited state (one of which can be considered metastable). The radiative coupling of the nuclear ensembles by the cavity field establishes the atomic coherence necessary for the cancellation of resonant absorption. Because this technique does not require atomic systems with a metastable level, electromagnetically induced transparency and its applications can be transferred to the regime of nuclear resonances, establishing the field of nuclear quantum optics.

The basic requirement to observe electromagnetically induced transparency (EIT) is a three-level system represented by the ground state, $|1\rangle$, and two upper states, $|2\rangle$ and $|3\rangle$, with respective energies E_2 and $E_3 > E_2$, where a strong laser field with Rabi frequency Ω_C induces an atomic coherence between states $|2\rangle$ and $|3\rangle$. This leads to a Fano-type quantum interference[18] if a (weak) probe laser field is tuned across the resonant transition $|1\rangle \rightarrow |3\rangle$, rendering the medium almost transparent in a narrow window around the exact resonance frequency. The degree of transparency is limited by the dephasing of the atomic coherence caused by the decay of state $|2\rangle$. Thus, maximum transparency is observed if $|2\rangle$ can be considered metastable, that is, if it has a decay width, γ_2, that is negligibly small relative to the coherent decay width, γ_3, of the state $|3\rangle$. Quantitatively, the spectral response of an ensemble of N atoms under these conditions can be described in terms of the linear susceptibility, χ:

$$\chi(\Delta) = \frac{g^2 N(i\Delta + \gamma_2)}{(i\Delta + \gamma_2)(i\Delta + \gamma_3) + |\Omega_C|^2} \qquad (1)$$

Here Δ is the detuning of the probe field from the exact resonance and g is the atom–field coupling constant[10]. The susceptibility approaches zero at exact resonance ($\Delta = 0$) as $\gamma_2 \rightarrow 0$. This is the phenomenon of EIT.

Here we extend the concept of EIT into the regime of hard X-rays by using the Mössbauer isotope ^{57}Fe, which is a two-level system (neglecting

the nuclear hyperfine interaction) with a transition energy of 14.4 keV and a natural linewidth of $\Gamma_0 = 4.7$ neV. It is not immediately clear how to achieve EIT without a proper nuclear three-level system. In fact, nuclear three-level systems with a metastable level together with a properly synchronized two-colour X-ray/X-ray or X-ray/light source are not available for use to establish conventional EIT schemes in the nuclear regime. Therefore, the possibility of using a single field of hard X-rays for EIT with a nuclear two-level system is highly desirable. This would open the way to exploring quantum optical concepts and non-linear effects at very short wavelengths, which is particularly appealing in view of the X-ray laser sources in development. The key to the realization of nuclear EIT described here is cooperative emission from ensembles of Mössbauer nuclei that are properly placed in a planar cavity for hard X-rays.

The physics of cooperative emission from atoms in cavities has many interesting phenomena even in the linear regime where the atom–cavity interaction can be treated in the weak-coupling limit, which is typically the case at X-ray wavelengths. Owing to the high resonant cross-section of ^{57}Fe, its 14.4-keV transition is a two-level system well suited to such studies. This isotope was recently used to study superradiant emission and the collective Lamb shift for a single ensemble of atoms located at an antinode of the field within a planar cavity[19,20]. Figure 1a shows the energy spectrum of the reflectivity of such a cavity excited in its third-order mode at a grazing angle of $\varphi = 3.5$ mrad. The spectrum, calculated using a transfer matrix algorithm for resonant X-ray scattering from layered media[21], shows the superradiant enhancement of the decay width, Γ_N, together with the collective Lamb shift, L_N.

A qualitatively new situation is encountered when two resonant ^{57}Fe layers instead of one are placed in a cavity. A pronounced dip in the spectral response appears when one of the ^{57}Fe layers is placed at a node of the standing wave in the cavity and the other is placed at an antinode (Fig. 1b). This dip is very reminiscent of the transparency dip observed in EIT. The appearance of this feature sensitively depends on the separation and the location of the two resonant layers within the cavity. For example, EIT completely vanishes if the two layers are arranged in the sequence antinode–node instead of node–antinode as seen from the top surface of the cavity (Fig. 1c). To investigate this effect quantitatively, we performed a perturbation expansion of the cavity reflectivity in powers of the grazing-incidence nuclear resonant scattering amplitude, $f_N(\varphi)$, for a single Lorentzian resonance line, $f_N = f_0(\varphi)/(x - i)$, where f_0 is the scattering amplitude at resonance (Supplementary Information, section C), $x = \Delta/\gamma_0$ and $\gamma_0 = \Gamma_0/2$. We obtain the following expression for the reflected amplitude as function of energy detuning, Δ (details of the derivation are given in Supplementary Information, sections A and C):

$$R(\Delta)$$

$$= \frac{d_2 f_0 \gamma_0 E_{2-+}(i\Delta + \gamma_0)}{(i\Delta + \gamma_0)(i\Delta + \gamma_0[1 + d_2 f_0 E_{2--}]) + d_1 d_2 f_0^2 \gamma_0^2 E_{2-+} E_{1+-}} \qquad (2)$$

Here the quantities E_{2-+}, E_{2--} and E_{1+-} are elements of the transfer matrices that describe the propagation of the photon fields in the
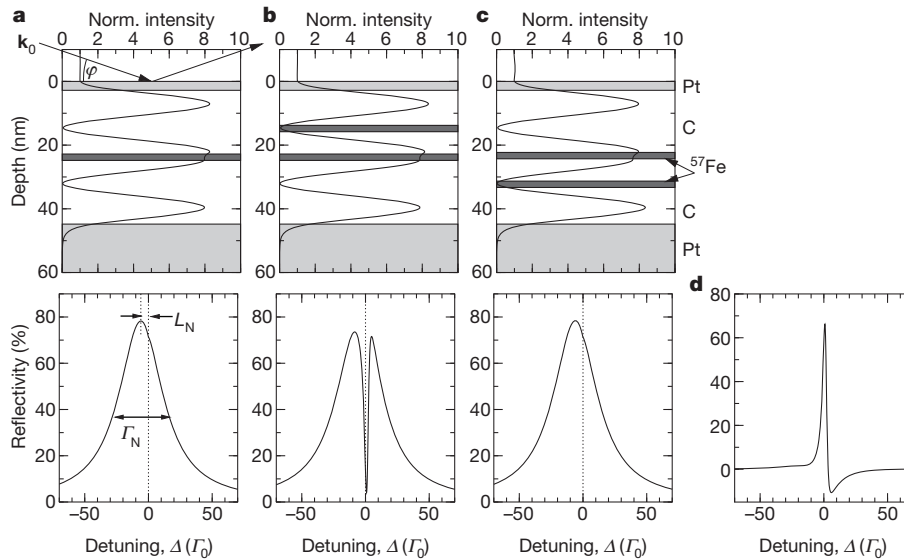
Figure 1 | Calculated reflectivity spectra of different cavity configurations. Sample geometry (top row) of planar cavities for X-rays, containing 2-nm-thick layers of ^{57}Fe nuclei (dark grey), and spectrally resolved reflectivity (bottom row) around the 14.4-keV nuclear resonance energy. The cavities are excited in the third-order mode under grazing angle $\varphi = 3.5$ mrad. The graphs in the top row show the standing wave intensity of the electromagnetic field in the cavities. **a**, For a single layer in the centre of the cavity, the collective decay width, Γ_N, is broadened owing to superradiant enhancement and shows a shift,

the collective Lamb shift L_N (refs 19, 20). **b**, If the cavity contains two ^{57}Fe layers placed respectively at a node and an antinode of the standing wave field, we observe a pronounced dip in the spectral response, indicative of an EIT window. **c**, The transparency window vanishes if the ^{57}Fe layers are arranged in the sequence antinode–node as viewed from the top of the planar cavity. **d**, Subtracting the energy spectrum in **b** from that in **a** reveals an asymmetric line profile resembling that of a Fano resonance.
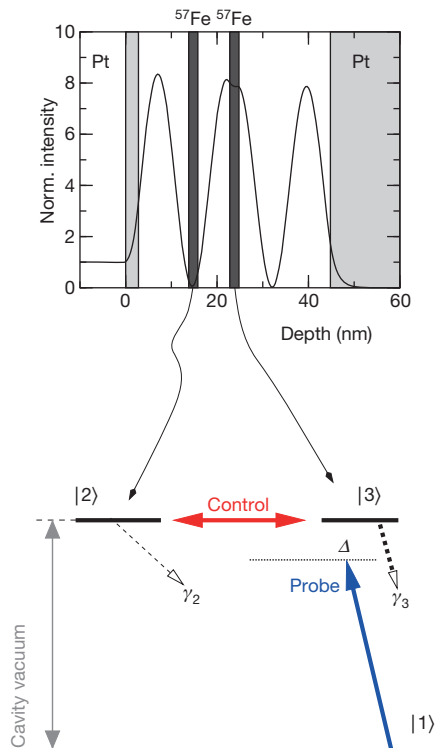

Figure 2 | Effective EIT level scheme of the nuclei in the cavity. Two ensembles of resonant ^{57}Fe atoms are respectively located at a node and an antinode of a low-finesse cavity. Owing to the significantly different photonic density of states of the atoms at the antinode of the cavity relative to that of those at the node, the radiative widths of the two ensembles differ considerably ($\gamma_2 \ll \gamma_3$), such that $|2\rangle$ can effectively be considered a metastable state. The cavity field thus causes each nucleus to act as a three-level system. The two upper states are coupled through their common ground state by the vacuum field of the cavity, effectively establishing the control field between them. We note that the cavity width (~ 100 eV) is much larger than the decay width of the nuclei ($\gamma_3 \approx 250$ neV).

transmitted (+) and reflected (−) directions in the unperturbed cavity, and d_1 and d_2 are the respective thicknesses of the two resonant layers. Equation (2) is basically identical to equation (1) for the complex susceptibility in EIT with $\gamma_2 = \gamma_0$, $\gamma_3 = \gamma_0(1 + d_2 f_0 E_{2--})$ and $\Omega_C^2 = d_1 d_2 (f_0 \gamma_0)^2 E_{2-+} E_{1+-}$. This result admits the following interpretation: the two atomic ensembles, one at the node of the standing wave field and one at the antinode, experience two significantly different photonic densities of states, leading to two different collective decay rates, γ_2 and γ_3. This effectively converts the nuclei in the cavity into three-level systems with two degenerate upper levels represented by the states $|2\rangle$ and $|3\rangle$ in the level scheme sketched in Fig. 2. The term $d f_0 \gamma_0 \sqrt{E_{2-+} + E_{1+-}}$, where $d = \sqrt{d_1 d_2}$, takes the role of the Rabi frequency, Ω_C, of the EIT control field. Here it scales with the two transfer matrix elements E_{2-+} and E_{1+-}, which are proportional to the two counter-propagating fields in the cavity at the position of the two resonant layers. That is, here the control field with Rabi frequency Ω_C arises from the radiative coupling of the two resonant layers (Fig. 3). The two excited states, $|2\rangle$ and $|3\rangle$, are coupled through their common ground state, $|1\rangle$, by the vacuum field of the cavity, which effectively establishes a control field between the two upper states. This scheme bears some resemblance to the recently reported effect of vacuum-induced transparency, whereby the probe field generates its own control field[22]. Because $\gamma_3 \gg \gamma_2$, the probe field dominantly couples the ground state, $|1\rangle$, to the excited state $|3\rangle$. The resulting arrangement of levels and their coupling in Fig. 2 closely resembles a Λ-type level scheme.

Cooperative emission is critical to EIT in this system. Whereas one of the atomic ensembles undergoes single-photon superradiant enhancement leading to a decay width of $\Gamma_N = 2\gamma_3 = d_2 f_0 \mathrm{Re}[E_{2--}]\Gamma_0$ and a collective Lamb shift of $L_N = -d_2 f_0 \mathrm{Im}[E_{2--}]\Gamma_0/2$, the decay width, $2\gamma_2$, of the other 'subradiant' ensemble is given by just the natural linewidth, Γ_0, such that $\gamma_3 \approx 50\gamma_2$ in the example shown in Fig. 1b. Thus, in the presence of a strong superradiant enhancement of state $|3\rangle$, state $|2\rangle$ is relatively long lived and therefore can be considered metastable. This is an important condition if a pronounced EIT effect is to be observed. If $\gamma_2 = \gamma_3$, the response of the system given by equation (1) is merely a sum of two Lorentzians without any destructive
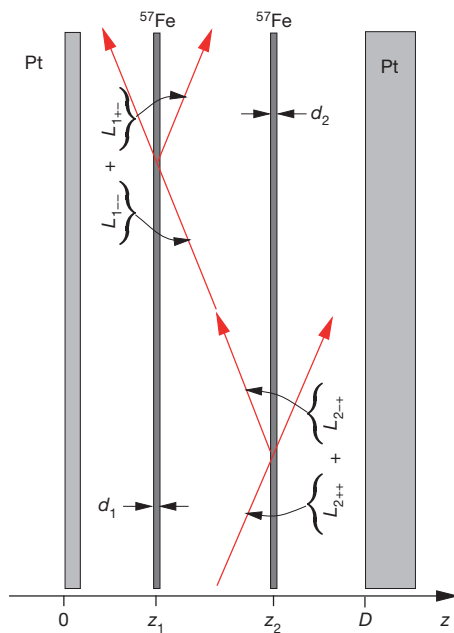
Figure 3 | Origin of the coherent control field in the cavity. The Rabi frequency of the control field is given by $\Omega_C = idf_0\gamma_0(L_{2++} + L_{2-+})(L_{1--} + L_{1+-})$ with $d = \sqrt{d_1 d_2}$ (Supplementary Information, equation (27)). The graphical representation of the scattering amplitudes, $L_{n\pm\pm}$ and $L_{n\pm\mp}$, for the fields propagating in the transmitted (+) and reflected (−) directions at the positions of the resonant layers supports the interpretation that Ω_C arises from the radiative coupling between the two resonant layers in the cavity.

interference between them. In an earlier investigation of nuclear resonant EIT, a reduction in absorption of a few per cent was observed on the basis of nuclear level anticrossing in a $FeCO_3$ crystal[23–25] with $\gamma_3 \approx 2\gamma_2$.

The magnitude of Ω_C relative to the decay rates γ_2 and γ_3 determines the emergence of Fano interference as the basic signature of EIT. When the control field is resonantly applied to the $|2\rangle \rightarrow |3\rangle$ transition, the excited state splits into the dressed states $|\pm\rangle = (|2\rangle \mp |3\rangle)/\sqrt{2}$, which are separated by an energy Ω_C. If this splitting is smaller than the linewidth γ_3, Fano-type interference[18] occurs between the two indistinguishable quantum mechanical paths between the dressed states and the ground state. Conversely, if Ω_C is much larger than γ_3 then Fano interference is negligible and two well-separated Lorentzians are obtained in the spectral response, equivalent to an Autler–Townes doublet[26] in atomic absorption profiles for strong driving fields. Thus, for EIT, Ω_C must not exceed γ_3. However, Ω_C has to be large enough to overcome decoherence effects, such as non-radiative decay, that result in a non-zero value of γ_2. Evaluation of equation (1) at resonance ($\Delta = 0$) reveals that EIT is still observable even for non-zero values of γ_2 if $|\Omega_C|^2$ is greater than $\gamma_2\gamma_3$. Overall, we find that the criterion $\gamma_3^2 > |\Omega_C|^2 > \gamma_2\gamma_3$ must be obeyed to obtain a pronounced EIT effect. The right-hand side of this inequality is to a very good approximation equivalent to $df_0|E_{1+-}E_{2-+}/E_{2--}| > 1$. A numerical investigation reveals that $|E_{1+-}E_{2-+}/E_{2--}|$ is of order one. With $f_0(\varphi) = 2.3 \text{ nm}^{-1}$ for metallic ^{57}Fe and $\varphi = 3.5$ mrad, this condition is fulfilled for layer thicknesses $d \geq 2$ nm, as used in the calculations the results of which are shown in Fig. 1. In the example in Fig. 1b, we find that $\hbar\Omega_C \approx 10\Gamma_0 \ll \hbar\gamma_3 \approx 50\Gamma_0$, where $\hbar$ is Planck's constant divided by 2π, such that the left-hand side of the EIT criterion above is also satisfied. To analyse the spectral shape of the transparency window, we subtract the spectrum in Fig. 1b from that of a single layer at an antinode (Fig. 1a). We obtain a spectral shape (Fig. 1d) that is indicative of a Fano resonance profile. If levels $|2\rangle$ and $|3\rangle$ were energetically degenerate, the transparency window would be of Lorentzian shape[27]. In our system, however, the degeneracy is lifted

by the collective Lamb shift, $L_N \ll \gamma_3$, leading to the asymmetry of the transparency window.

As emphasized above and illustrated in Fig. 1, the EIT effect is not symmetric with respect to the sequence of the resonant layers at the nodes and antinodes of the cavity field. To understand this, we use the definitions of E_{1+-} and E_{2-+} (Supplementary Information, section A) to write $\Omega_C = df_0\Gamma_0(L_{1--} + L_{1+-})(L_{2++} + L_{2-+})$. The quantities $L_{n\mu\nu}$ ($\mu, \nu \in \{+, -\}$) are the amplitudes for the scattering of fields from the incoming direction (ν) into the outgoing direction (μ) after interaction with the atomic ensembles located at the positions z_n ($n \in \{1, 2\}$) in the cavity. The graphical representation of these amplitudes (Fig. 3) supports the interpretation that Ω_C is determined by the radiative coupling between the two ensembles in the cavity. It can be shown (Supplementary Information, section A) that $L_{n++} + L_{n-+}$ vanishes at the nodes of the cavity field and that $L_{n--} + L_{n+-}$ does not. Therefore, if the second layer is located at a node of the field, $\Omega_C = 0$ and there is no EIT effect. This means that the effective magnitude of Ω_C can be controlled not only by the thicknesses of the resonant layers but also by their placement within the wave field in the cavity.

To verify EIT experimentally, we prepared two planar X-ray cavities consisting of Pt(3 nm)/C(38 nm)/Pt(10 nm) sandwich structures, each containing two ^{57}Fe layers occupying a node and, respectively, an antinode of the cavity field in the sequences shown in Fig. 1b, c. To achieve a maximum EIT effect without perturbing the cavity field too much, we chose each ^{57}Fe layer to be 3 nm thick. At this thickness, the ^{57}Fe layers order ferromagnetically with the magnetization confined to the plane of the films. The magnetic hyperfine interaction lifts the degeneracy of the nuclear magnetic sublevels, leading to four allowed dipole transitions for the given scattering geometry, where the magnetization is aligned parallel to the wavevector of the incident photons, $\mathbf{k}_0$. Despite the magnetic level splitting, the basic physics of the EIT effect as discussed above remains unaffected because the separation of the lines is large enough to consider each of the four resonance lines separately.

The experiments were performed at the PETRA III synchrotron radiation source (DESY, Hamburg) using the method of nuclear resonant scattering (Supplementary Information, section D). This technique relies on the pulsed broadband excitation of nuclear levels followed by the time-resolved detection of coherently scattered, delayed photons that are emitted on a timescale of $\tau = \hbar/\Gamma_0$ after resonant excitation[28]. To determine the energy spectrum of the cavity reflectivity from the time-resolved data, we used a technique based on stroboscopic detection of the delayed response of the sample after passing through a resonant energy analyser[29] (Supplementary Information, section E). As an analyser, we used a 1-μm-thick stainless steel foil with a 95% isotopic enrichment in ^{57}Fe, providing a single-line transmission function with a spectral width of $\sim$10 neV. The foil was mounted on a Doppler drive that provided a periodic energy detuning, Δ, as a function of which single photons were counted using a fast avalanche photodiode detector. This set-up is similar to that used for detection of the collective Lamb shift[19]. Single-photon events were registered by recording their arrival time after excitation together with the velocity of the Doppler drive at the moment of detection. From data sets of about 10^7 such events, we extracted the reflectivity, $|R(\Delta)|$, of the samples as a function of Δ by applying the procedures described in ref 29. The results are shown in Fig. 4 for the samples with the ^{57}Fe layers arranged in the sequence node–antinode (sample A) and antinode–node (sample B). The red solid lines are simulated spectra based on the structural data obtained from X-ray reflectivity and nuclear hyperfine interaction parameters obtained from conversion electron Mössbauer spectroscopy, using no adjustable parameters.

Sample A shows evidence of EIT, as predicted. Its spectral response has transparency dips that are particularly pronounced at the outer (and strongest) resonance lines of the hyperfine-split spectrum, at detunings of $\Delta = \pm51\Gamma_0$ (Fig. 4a–d, vertical dashed lines). The EIT
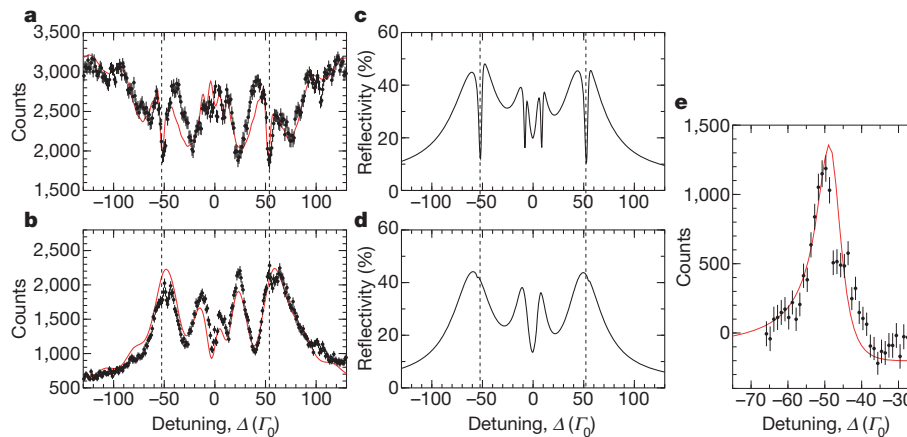
Figure 4 | Observation of nuclear resonant EIT. a, b, Measured spectral response (reflectivity) of sample A (**a**) and sample B (**b**). Sample A shows evidence for EIT, as predicted, with two strong transparency dips at Doppler detunings of $\Delta = -51\Gamma_0$ and $51\Gamma_0$ (dashed vertical lines). The EIT effect vanishes in sample B. Solid red lines are calculations taking into account the stroboscopic detection procedure applied here[29]. The difference in the baselines of the experimental spectra of the two samples at large detunings is a feature of the stroboscopic detection technique. **c, d,** The central areas of the measured spectra ($\Delta = -70\Gamma_0$ to $70\Gamma_0$), however, closely resemble the calculated spectra of sample A (**c**) and sample B (**d**). **e,** Difference spectrum of the measured spectra in **a** and **b** in the Δ/Γ_0 range between -70 and -30. The solid red line is a guide to the eye based on a Fano resonance line shape. Error bars are estimated from photon counting statistics.

dips in the centre are less resolved because they are much sharper than the outer ones and thus exceed the resolution limit of the stroboscopic detection method. In sample B, the EIT vanishes owing to the reversal of the sequence of the two layers in the cavity field. As a result, the spectral response is dominated by superradiantly broadened lines without transparency dips. Owing to the stroboscopic detection process, the off-resonance baseline is located at different levels in the two spectra. This is fully described by the simulation and does not affect the spectral signature of EIT observed here. To characterize the spectral shape of the measured EIT window, we subtracted the spectrum of sample A from that of sample B over an energy range from $\Delta/\Gamma_0 = -70$ to -30. The spectral profile (Fig. 4e) shows an asymmetric shape typical for a Fano resonance, as indicated by the red solid curve. Its width, of $\sim 10\Gamma_0$, is consistent with the value of $\hbar\Omega_C$ estimated above.

The depth of the strong transparency dips of sample A corresponds to a reflectivity that is reduced to a level of $|R|^2 = 0.10$ relative to $|R|^2 = 0.45$ in sample B. The degree of transparency would be much larger if the nuclear resonance were a single line, as illustrated in Fig. 1, where $|R|^2 = 0.03$ in the transparency window. In forthcoming experiments, it should be possible also to detect the transmitted field that leaves the cavity after propagating along its axis. This brings the production of slow light in this spectral regime into reach. We estimate group velocities in the range of $30\,\mathrm{m\,s}^{-1}$ (Supplementary Information, section B). To produce such group velocities, it is necessary to adjust the width of the transparency window to allow propagation of transform-limited pulses along the axis of the cavity. The width can be controlled by adjusting the thickness, d_1, of the layer located at the node of the cavity field or by placing several layers in the multiple nodes of high-order cavity modes. Moreover, we predict that drastic changes will occur in the linear refractive index at resonance if the positions of the resonant layers in the cavity are changed by only 1 nm. This effect is analogous to the cross-phase modulation due to giant Kerr nonlinearity[2] as it makes it possible to control, for example, the phase accumulation of the probe field on propagation along the cavity. Self-phase modulation of the probe field under nuclear EIT conditions might become relevant at intensities achieved at X-ray laser facilities in the near future (Supplementary Information, section B).

We emphasize that this technique, based on coherent light scattering, can be generally applied to any ensembles of resonant emitters (for example atoms, ions or quantum dots) properly placed in optical cavities. Although in our experiment the EIT effect relies on the spatial modulation and interference of single-photon cooperative emission from two ensembles of many emitters in a cavity field, the same effect can be expected for two single emitters that are sufficiently well confined in a cavity or microcavity. With a large Purcell enhancement of the spontaneous emission from an antinode of the cavity field and inhibited emission from a node, the condition for our cavity EIT scheme will be equally well fulfilled for single emitters as for many. The fact that this scheme works with two-level systems extends EIT and its applications to systems that do not have a metastable level, facilitating the transfer of EIT and its applications to the nuclear realm. Moreover, the radiative coupling of separated atomic ensembles in a cavity provides a very sensitive tool to probe the properties of cooperative emission such as superradiance and the collective Lamb shift.

1. Harris, S. E., Field, J. E. & Imamoğlu, A. Nonlinear optical processes using electromagnetically induced transparency. *Phys. Rev. Lett.* **64,** 1107–1110 (1990).
2. Schmidt, H. & Imamoğlu, A. Giant Kerr nonlinearity obtained by electromagnetically induced transparency. *Opt. Lett.* **21,** 1936–1938 (1996).
3. Harris, S. E. & Hau, L. V. Nonlinear optics at low light levels. *Phys. Rev. Lett.* **82,** 4611–4614 (1999).
4. Hau, L. V., Harris, S. E., Dutton, Z. & Behroozi, C. H. Light speed reduction to 17 metres per second in an ultracold atomic gas. *Nature* **397,** 594–598 (1999).
5. Kash, M. M. *et al.* Ultraslow group velocity and enhanced nonlinear optical effects in a coherently driven hot atomic gas. *Phys. Rev. Lett.* **82,** 5229–5232 (1999).
6. Zibrov, A. S. *et al.* Experimental demonstration of laser oscillation without population inversion via quantum interference in Rb. *Phys. Rev. Lett.* **75,** 1499–1502 (1995).
7. Padmabandu, G. G. *et al.* Laser oscillation without population inversion in a sodium atomic beam. *Phys. Rev. Lett.* **76,** 2053–2056 (1996).
8. Scully, M. O. & Fleischhauer, M. Lasers without inversion. *Science* **263,** 337–338 (1994).
9. Liu, C., Dutton, Z., Behrozzi, C. H. & Hau, L. V. Observation of coherent optical information storage in an atomic medium using halted light pulses. *Nature* **409,** 490–493 (2001).
10. Lukin, M. D. Trapping and manipulating photon states in atomic ensembles. *Rev. Mod. Phys.* **75,** 457–472 (2003).
11. Hänsch, T., Keil, R., Schabert, A., Schmelzer, Ch & Toschek, P. Interaction of laser light waves by dynamic stark splitting. *Z. Phys.* **226,** 293–296 (1969).
12. Kocharovskaya, O. A. & Khanin, Ya. I. Coherent amplification of an ultrashort pulse in a three-level medium without a population inversion. *JETP Lett.* **48,** 630–634 (1988).
13. Boller, K. J., Imamoğlu, A. & Harris, S. E. Observation of electromagnetically induced transparency. *Phys. Rev. Lett.* **66,** 2593–2596 (1991).
14. Lukin, M. D. & Imamoğlu, A. Controlling photons using electromagnetically induced transparency. *Nature* **413,** 273–276 (2001).
15. Fleischhauer, M., Imamoğlu, A. & Marangos, J. P. Electromagnetically induced transparency: optics in coherent media. *Rev. Mod. Phys.* **77,** 633–673 (2005).
16. Buth, C., Santra, R. & Young, L. Electromagnetically induced transparency for X-rays. *Phys. Rev. Lett.* **98,** 253001 (2007).

17. Glover, T. E. *et al.* Controlling X-rays with light. *Nature Phys.* **6,** 69–74 (2010).
18. Fano, U. Effects of configuration interaction on intensities and phase shifts. *Phys. Rev.* **124,** 1866–1878 (1961).
19. Röhlsberger, R., Schlage, K., Sahoo, B., Couet, S. & Rüffer, R. Collective Lamb shift in single- photon superradiance. *Science* **328,** 1248–1251 (2010).
20. Röhlsberger, R. The collective Lamb shift in nuclear γ-ray superradiance. *J. Mod. Opt.* **57,** 1979–1992 (2010).
21. Röhlsberger, R., Klein, T., Schlage, K., Leupold, O. & Rüffer, R. Coherent X-ray scattering from ultrathin probe layers. *Phys. Rev. B* **69,** 235412 (2004).
22. Tanji-Suzuki, H., Chen, W., Landig, R., Simon, J. & Vuletic, V. Vacuum-induced transparency. *Science* **333,** 1266–1269 (2011).
23. Coussement, R. *et al.* Controlling absorption of gamma radiation via nuclear level anticrossing. *Phys. Rev. Lett.* **89,** 107601 (2002).
24. Gheysen, S. & Odeurs, J. Quantum mechanical study of resonant scattering in a nuclear Λ scheme. *Phys. Rev. B* **74,** 155443 (2006).
25. Gheysen, S. & Odeurs, J. Nuclear level mixing-induced interference in $FeCO_3$. *J. Phys. Condens. Matter* **20,** 485214 (2008).
26. Autler, S. H. & Townes, C. H. Stark effect in rapidly varying fields. *Phys. Rev.* **100,** 703–722 (1955).
27. Anisimov, P. M., Dowling, J. P. & Sanders, B. Objectively discerning Autler-Townes splitting from electromagnetically induced transparency. *Phys. Rev. Lett.* **107,** 163604 (2011).
28. Röhlsberger, R. *Nuclear Condensed Matter Physics with Synchrotron Radiation* (Springer Tracts Mod. Phys. 208, Springer, 2004).
29. Callens, R. *et al.* Principles of stroboscopic detection of nuclear forwardscattered synchrotron radiation. *Phys. Rev. B* **67,** 104423 (2003).

LETTER

Thresholdless nanoscale coaxial lasers

M. Khajavikhan[1], A. Simic[1]*, M. Katz[1]*, J. H. Lee[1]†, B. Slutsky[1], A. Mizrahi[1], V. Lomakin[1] & Y. Fainman[1]

The effects of cavity quantum electrodynamics (QED), caused by the interaction of matter and the electromagnetic field in sub-wavelength resonant structures, have been the subject of intense research in recent years[1]. The generation of coherent radiation by subwavelength resonant structures has attracted considerable interest, not only as a means of exploring the QED effects that emerge at small volume, but also for its potential in applications ranging from on-chip optical communication to ultrahigh-resolution and high-throughput imaging, sensing and spectroscopy. One such strand of research is aimed at developing the 'ultimate' nanolaser: a scalable, low-threshold, efficient source of radiation that operates at room temperature and occupies a small volume on a chip[2]. Different resonators have been proposed for the realization of such a nanolaser—microdisk[3] and photonic band-gap[4] resonators, and, more recently, metallic[5,6], metallo-dielectric[7–10] and plasmonic[11,12] resonators. But progress towards realizing the ultimate nanolaser has been hindered by the lack of a systematic approach to scaling down the size of the laser cavity without significantly increasing the threshold power required for lasing. Here we describe a family of coaxial nanostructured cavities that potentially solve the resonator scalability challenge by means of their geometry and metal composition. Using these coaxial nanocavities, we demonstrate the smallest room-temperature, continuous-wave telecommunications-frequency laser to date. In addition, by further modifying the design of these coaxial nanocavities, we achieve thresholdless lasing with a broadband gain medium. In addition to enabling laser applications, these nanoscale resonators should provide a powerful platform for the development of other QED devices and metamaterials in which atom–field interactions generate new functionalities[13,14].

The miniaturization of laser resonators using dielectric or metallic material structures faces two challenges: (1) the (eigen-)mode scalability, implying the existence of a self-sustained electromagnetic field regardless of the cavity size, and (2) a relationship between optical gain and cavity loss which results in a large and/or unattainable lasing threshold as the volume of the resonator is reduced[15]. Here we propose and demonstrate a new approach to nano-cavity design that resolves both challenges: first, subwavelength-size nano-cavities with modes far smaller than the operating wavelength are realized by designing a plasmonic coaxial resonator that supports the cut-off-free transverse electromagnetic (TEM) mode; second, the high lasing threshold for small resonators is reduced by utilizing cavity QED effects, causing high coupling of spontaneous emission into the lasing mode[16,17]. When fully exploited, this approach can completely eliminate the threshold constraint by reaching so-called thresholdless lasing, which occurs when every photon emitted by the gain medium is funnelled into the lasing mode[16,17].

The coaxial laser cavity is shown in Fig. 1a. At the heart of the cavity lies a coaxial waveguide that supports plasmonic modes and is composed of a metallic rod enclosed by a metal-coated semiconductor ring[18,19]. The impedance mismatch between a free-standing coaxial waveguide and free space creates a resonator. However, our design

uses additional metal coverage on top of the device and thin, low-index dielectric plugs of silicon dioxide (SiO_2) at the top end of the coaxial waveguide and air at the bottom end to improve the mode confinement. The role of the top SiO_2 plug is to prevent the formation of undesirable plasmonic modes at the top interface, between the metal and the gain medium. The lower air plug is used to allow pump energy into the cavity and also to couple out the light generated in the coaxial resonator. The metal in the sidewalls of the coaxial cavity is placed in direct contact with the semiconductor to ensure the support of plasmonic modes, providing a large overlap between the modes of the resonator and the emitters distributed in the volume of the gain medium. In addition, the metallic coating serves as a heat sink that facilitates room-temperature and continuous-wave operation.

To reduce the lasing threshold, the coaxial structures are designed to maximize the benefits from the modification of the spontaneous emission due to the cavity QED effects[16,17]. Because of their small size, the modal content of the nanoscale coaxial cavities is sparse, which is a key requirement to obtain high spontaneous emission coupling into the lasing mode of the resonator. Their modal content can be further modified by tailoring the geometry, that is, the radius of the core, the width of the ring, and the height of the gain medium and the low-index plugs. Note that the number of modes supported by the
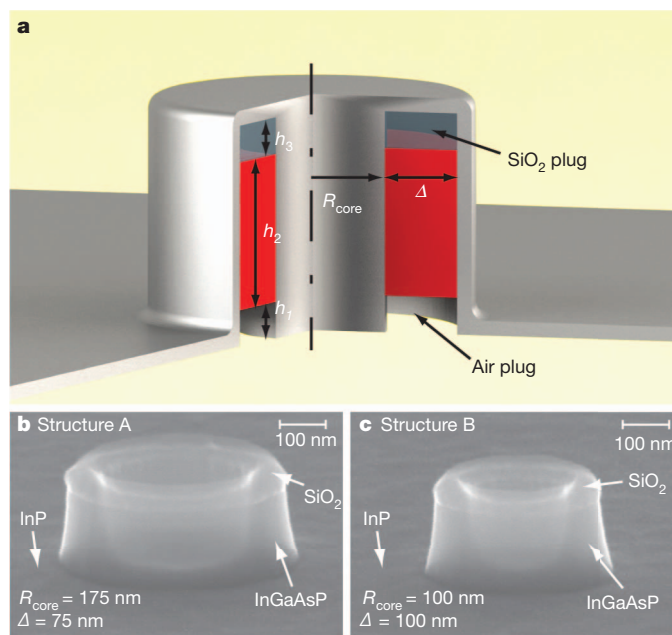

Figure 1 | Nanoscale coaxial laser cavity. a, Diagram of a coaxial laser cavity; the gain medium is shown in red. See main text for description of nomenclature. **b, c,** Scanning electron microscope images of the constituent rings in structure A and structure B, respectively. A side view of the rings comprising the coaxial structures is seen; the rings consist of SiO_2 on top, and a quantum-well gain region underneath. See main text for details.

[1]Department of Electrical and Computer Engineering, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0407, USA. †Present address: Oracle Labs, 9515 Town Centre Drive, San Diego, California 92121, USA.
*These authors contributed equally to this work.

resonator that can participate in the lasing process is ultimately limited to those that occur at frequencies that coincide with the gain bandwidth of the semiconductor gain material. In this work we use a semiconductor gain medium composed of six quantum wells of $In_{x=0.56}Ga_{1-x}As_{y=0.938}P_{1-y}$ (10 nm thick)/$In_{x=0.734}Ga_{1-x}As_{y=0.57}P_{1-y}$ (20 nm thick), resulting in a gain bandwidth that spans frequencies corresponding to wavelengths in vacuum from 1.26 μm to 1.59 μm at room temperature (295 K), and from 1.27 μm to 1.53 μm at a temperature of 4.5 K (ref. 20).

We consider two different geometries of the structure shown in Fig. 1a. The first, referred to as structure A, has an inner core radius of $R_{core} = 175$ nm, a gain-medium ring with a thickness of $\Delta = 75$ nm, a lower plug height of $h_1 = 20$ nm, a quantum-wells height of 200 nm covered by a 10-nm overlayer of InP, resulting in a total gain-medium height of $h_2 = 210$ nm, and an upper plug height of $h_3 = 30$ nm. The second, structure B, is smaller in diameter, having $R_{core} = 100$ nm and $\Delta = 100$ nm. The heights of the plugs and gain medium are identical to those of structure A. Figure 1b and c shows scanning electron microscope images of the constituent rings in structure A and structure B, respectively. The two structures are fabricated using standard nanofabrication techniques. Additional details of the fabrication procedure are provided in Supplementary Information part 1.

Figure 2 shows the modal content of the two structures at a temperature of 4.5 K, modelled using the three-dimensional finite element method (FEM) eigenfrequency solver in the radio-frequency package of COMSOL Multiphysics. Figure 2a shows that for structure A the fundamental TEM-like mode and the two degenerate HE_{11} modes are supported by the resonator and fall within the gain bandwidth of the gain material. This simulation is also repeated for structure A with room-temperature material parameters (see Supplementary Fig. 1), showing that for structure A at room temperature, the two degenerate HE_{11} modes are red-shifted to 1,400 nm, and exhibit a reduced quality factor of $Q \approx 35$, compared to $Q \approx 47$ at 4.5 K. The TEM-like mode is red-shifted to 1,520 nm with $Q \approx 53$, compared to $Q \approx 120$ at 4.5 K. All cavity quality factors are at transparency, meaning that the imaginary part of the gain medium's permittivity is set to zero in the calculations. The simulations are performed with nominal values for the permittivity of the active medium and metal at 4.5 K and at room temperature (see Supplementary Information part 2). A discussion on deviations of the material properties from nominal values, as well as additional technical details about FEM simulations, are provided in Supplementary Information part 3.

Structure B, shown in Fig. 2b, supports only the fundamental TEM-like mode at a temperature of 4.5 K. The quality factor $Q \approx 265$ for this mode is higher than that of structure A. In general, the metal coating and the small aperture of the nanoscale coaxial cavity inhibit the gain emitters from coupling into the continuum of the free-space radiation modes[21]. Hence, the single-mode cavity of structure B exhibits a very high spontaneous emission coupling factor ($\beta \approx 0.99$), approaching the condition for an ideal thresholdless laser[16,17]. The spontaneous emission factor is calculated by placing randomly oriented and randomly positioned dipoles in the active area of the cavity, and then computing their emitted power at different wavelengths. The β-factor is given by the emitted power that spectrally coincides with the lasing mode, divided by the total emitted power[22].

Characterization of the nanoscale coaxial lasers was performed under optical pumping with a $\lambda = 1,064$ nm laser pump beam in continuous wave and pulsed regime. Additional details on the measurement system are provided in Supplementary Information part 4. Excitation of the cavity modes is confirmed by the measurements of the far-field emission from the devices. The mode profiles are given in Supplementary Information part 5.

Figure 3 shows the emission characteristics of the nanoscale coaxial laser of structure A operating at 4.5 K (light–light curve, Fig. 3a; spectral evolution, Fig. 3b; linewidth, Fig. 3c) and at room temperature (light–light curve, Fig. 3d; spectral evolution, Fig. 3e; linewidth, Fig. 3f). The light–light curves of Fig. 3a and d show standard laser action behaviour, where spontaneous emission dominates at lower pump powers (referred to as the photoluminescence region), and stimulated emission is dominant at higher pump powers (referred to as the lasing region). The photoluminescence and lasing regions are connected through a pronounced transient region, referred to as amplified spontaneous emission (ASE). The evolution of the spectrum shown in Fig. 3b and e also confirms these three regimes of operation. The spectral profiles at low pump powers reflect the modification of the spontaneous emission spectrum by the cavity resonances depicted in Fig. 2a. The linewidth of the lasers shown in Fig. 3c and f narrows with the inverse of the output power at lower pump levels (the solid trend line). This is in agreement with the well-known Schawlow–Townes formula for lasers operating below threshold[23]. Around threshold, in semiconductor lasers the rapid increase of the coupling between the gain coefficient and the refractive index of the gain medium slows down the narrowing of the linewidth, until charge carrier pinning resumes the modified Schawlow–Townes inverse power narrowing rate[24,25]. In practice, only a few semiconductor lasers are shown to have above-threshold linewidth behaviour that follows the modified Schawlow–Townes formula. In most reported lasers, the linewidth behaviour differs distinctly from the inverse power narrowing rate. The mechanisms affecting the above-threshold linewidth, especially for lasers with high spontaneous emission coupling to the lasing mode, are still a subject of research[26–28]. Supplementary Information part 6 contains detailed diagrams of emission properties for the lasers reported above.
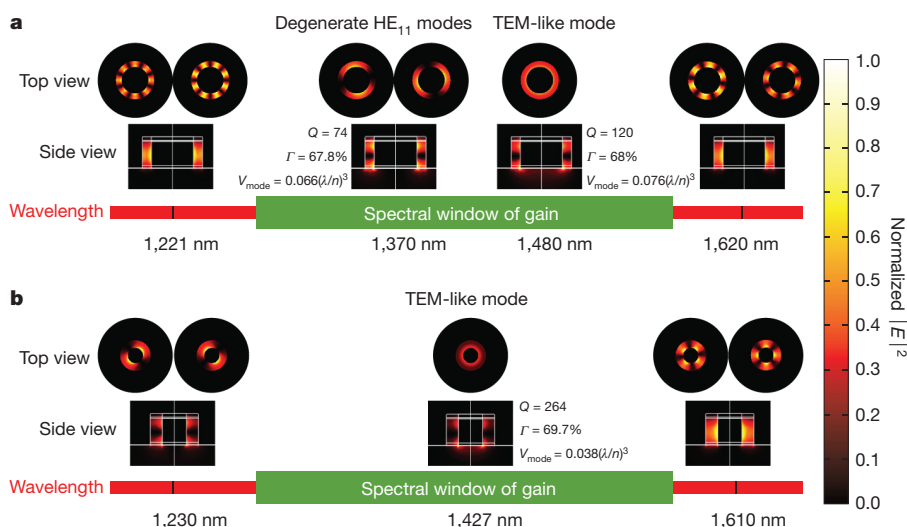

Figure 2 | Simulation of the electromagnetic properties of nanoscale coaxial cavities. a, The modal spectrum of the cavity of structure A at a temperature of 4.5 K. **b**, As **a** but for structure B. Q, quality factor; Γ, factor giving extent of energy confinement to the semiconductor region[30]; V_{mode}, the effective modal volume[30]. The colour bar shows normalized $|E|^2$, where E is the electric field intensity. Nominal permittivity values are used in this simulation. (See Supplementary Information parts 2 and 3 for nominal permittivities and the deviation of the permittivities from the nominal values, respectively.)
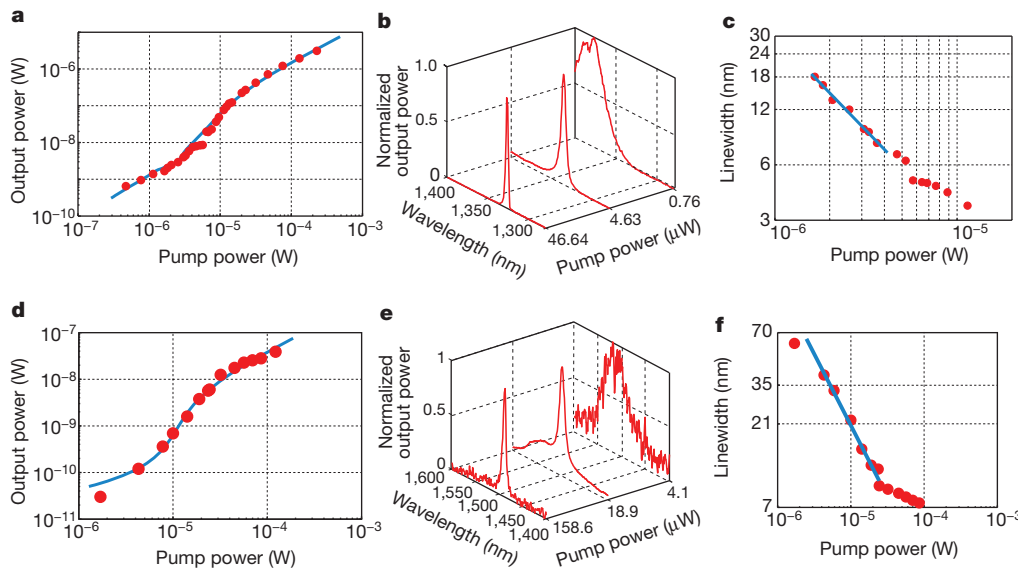
Figure 3 | Optical characterization of nanoscale coaxial cavities of structure A at 4.5 K and room temperature, showing lasing. a–c, At 4.5 K; **d–f,** at room temperature. Shown are light–light curves (**a, d**), spectral evolution diagrams for lasers with threshold (**b, e**), and linewidth versus pump power (**c, f**). The pump power is calculated as the fraction of the power incident on the laser aperture. The solid curves in **a** and **d** are the best fit of the rate-equation model. The solid lines in **c** and **f** show the inverse power narrowing rate of the linewidth. The resolution of the monochromator was set to 3.3 nm.

A rate-equation model is adopted to study the dynamics of the photon–carriers interaction in the laser cavities. Details of the rate-equation model are provided in Supplementary Information part 7. The light–light curves obtained from the rate-equation model for the laser of structure A are shown as solid blue lines in Fig. 3a and d. For the laser operating at 4.5 K, by fitting the rate-equation model to the experimental data, we found that almost 20% of the spontaneous emission is coupled to the lasing mode, which is assumed to be the mode with the highest quality factor (TEM-like mode). This assumption is validated by examining the far-field radiation pattern and the polarization state of the output beam (see Supplementary Information part 5). At room temperature, the surface and Auger non-radiative recombination processes dominate. As the carriers are lost through non-radiative channels, the ASE kink of the laser is more pronounced, and, as expected, the laser threshold shifts to higher pump powers.

Next, we examine the emission characteristics of structure B. According to the electromagnetic analysis (Fig. 2b), this structure is expected to operate as a thresholdless laser, as only one non-degenerate mode resides within the gain medium's emission bandwidth. The emission characteristics of structure B at 4.5 K are shown in Fig. 4. More detailed diagrams of emission properties for this laser are also given in Supplementary Information part 6. The light–light curve of Fig. 4a, which follows a straight line with no pronounced kink, agrees with the thresholdless lasing hypothesis. The thresholdless behaviour is further manifested in the spectral evolution, seen in Fig. 4b, where a single narrow, Lorentzian-like emission is obtained over the entire five-orders-of-magnitude range of pump power. This range spans

from the first signal detected above the detection system noise floor at 720 pW pump power, to the highest pump power of more than 100 µW. Because the homogeneously broadened linewidth of the gain medium is larger than the linewidth of the observed emission, the emission profile is attributed to the cavity mode. The measured linewidth at low pump power ($\Delta\lambda_{FWHM} \approx 5$ nm), which agrees with the cavity Q-factor of the TEM-like mode at transparency, as well as the radiation pattern reported in Supplementary Information part 5, confirm the electromagnetic simulation given in Fig. 2b.

The assertion that the device indeed reaches lasing is further substantiated by careful study of the linewidth behaviour. At low pump levels, the linewidth depicted in Fig. 4c is almost constant, and does not narrow with output power, implying that the linewidth shows no subthreshold behaviour[23,25]. The lack of variation of linewidth with pump power is most likely to be the result of the increasing gain–index coupling, which is a well-known around-threshold behaviour in semiconductor lasers[24,25]. Another indication, and more decisive proof that structure B does not exhibit subthreshold behaviour, is that the linewidth narrowing above the 100 nW pump power level does not follow the inverse power narrowing rate that is clearly identified in structure A. The observed narrowing rate for this laser is attributed to the carrier-pinning effect, as further corroborated by the results of the rate-equations model for the carrier density presented in Supplementary Fig. 13. To the best of our knowledge, this linewidth behaviour, though predicted in theory[26–28], has never been reported in any laser, and is unique to our thresholdless laser. The light–light curve obtained from the rate-equation model for the laser of structure B at 4.5 K is shown by

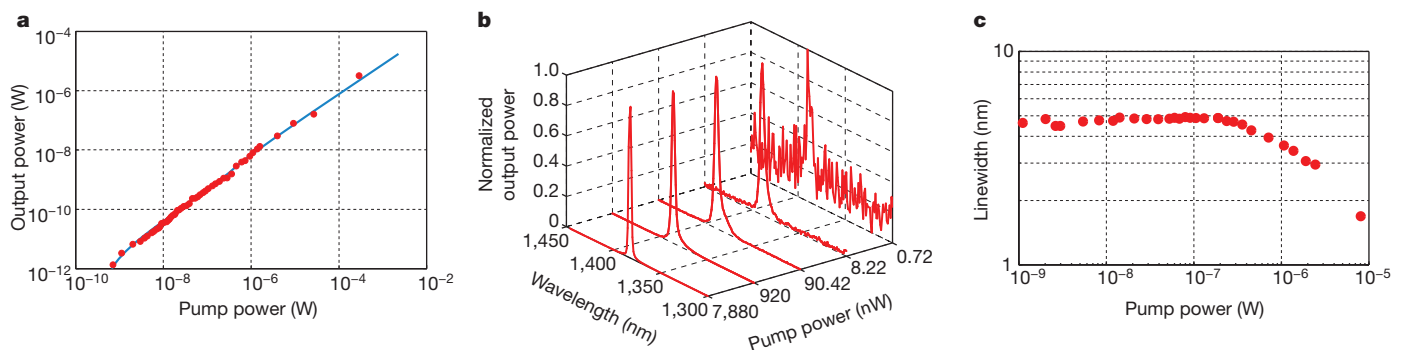**Figure 4 | Optical characterization of nanoscale coaxial cavities of structure B at 4.5 K, showing thresholdless lasing. a,** Light–light curve; **b,** spectral evolution; and **c,** linewidth evolution. The pump power is calculated as in Fig. 3;

the solid curve in **a** is the best fit of the rate-equation model. The resolution of the monochromator was set to 1.6 nm.

the solid blue line in Fig. 4a. The best fit of our rate-equation model to the experimental data is achieved if 95% of the spontaneous emission is coupled to the lasing mode ($\beta = 0.95$). The deviation from $\beta = 0.99$ predicted by the electromagnetic simulation can be attributed to other non-radiative recombination processes that have not been considered in the rate-equation model, and to the spectral shift of the mode at higher pump levels that causes variations in the available gain for the mode. In summary, all the experimental observations, including output spectrum and beam profile, electromagnetic simulations, rate-equation model, and comparison with the non-thresholdless lasers, suggest thresholdless lasing as the only plausible hypothesis that satisfactorily explains all aspects of the emission of the light-emitting device based on structure B at 4.5 K.

The thresholdless lasing in nanoscale coaxial cavities clearly differs from the state-of-the-art, high-quality-factor, photonic-bandgap structures[29]. In the latter, near-thresholdless lasing is achieved in a quantum dot gain-medium system with spectrally narrow band emission, and relies extensively on tuning of the cavity mode to the centre of the quantum dot emission spectrum[29]. In the former, thresholdless lasing in a broadband gain medium is achieved with a low-quality-factor, single-mode metal cavity. Smaller size, straightforward fabrication procedure, and better thermal properties are just a few of the advantages of nanoscale coaxial cavities for the realization of thresholdless lasing.

In conclusion, with nanoscale coaxial structures, we have successfully demonstrated room-temperature, continuous-wave lasing, as well as low-temperature thresholdless lasing in a spectrally broadband semiconductor gain medium. Owing to the fundamental TEM-like mode with no cut-off, these cavities support ultra-small modes, offer large mode-emitter overlap that results in optimal utilization of the pump power, and provide multifold scalability. Further developments towards electrical pumping of thresholdless nanoscale coaxial lasers that can operate at room temperature are in progress.

The implications of our work are threefold. First, the demonstrated nanoscale coaxial lasers have a great potential for future nano-photonic circuits on a chip. Second, thresholdless operation and scalability provide the first systematic approach toward the realization of QED objects and functionalities, specifically the realization of quantum metamaterials. Last, this new family of resonators paves the way to in-depth study of the unexplored physics of emitter–field interaction, photon statistics, and carrier dynamics in ultra-small metallic structures.

METHODS SUMMARY

Device fabrication. The devices are fabricated on an InP wafer, with 300 μm InP substrate, 200 nm total height of quantum wells, and covered by a 10-nm-thick InP over-layer. Hydrogen silsesquioxane (HSQ) is used as a negative tone resist, on which rings with different inner radii and widths are written by electron beam exposure. The exposed HSQ serves as a mask for the subsequent reactive ion etching (RIE) process that utilizes H_2:CH_4:Ar plasma to remove InGaAsP and InP. The wafer is cleaned with oxygen plasma, and an alloy of silver and aluminium (98%Ag+2%Al) is deposited, using electron-beam evaporation. The sample is mounted on a silicon wafer with silver epoxy, and is dipped in hydrochloric acid to remove the InP substrate.

Material constants. For the finite element method simulation of devices operating at 4.5 K, we used $\varepsilon_{silver} = -120.43 - 0.03073i$ for silver permittivity, $\varepsilon_g = 11.15$ for gain-medium permittivity, $\varepsilon_{InP} = 9.49$ for InP permittivity, $\varepsilon_{SiO2} = 2.1$ for SiO2 permittivity, and $\varepsilon_{air} = 1$ for air permittivity. For room temperature, the permittivities are the same as at $T = 4.5$ K, except $\varepsilon_{silver} = -120.43 - 3.073i$, $\varepsilon_g = 11.56$ and $\varepsilon_{InP} = 9.86$.

Device measurement. The devices are optically pumped with a 1,064-nm laser beam, focused to an area of ~64 μm² on the sample surface. A microscope objective with a numerical aperture of 0.4 is used to focus the pump beam and to collect the output emission. The devices are examined under both continuous wave (CW) and pulsed mode pumping (12-ns pulse width at 300-kHz repetition rate) conditions. Output spectra were obtained using a monochromator with a resolution set at 3.3 nm. When necessary, the linewidth is measured with monochromator resolution set to 1.65 nm and 0.67 nm. The cryogenic measurements were obtained by placing the devices in a continuous-flow microscopy cryostat, and then cooling them with liquid helium to a temperature of 4.5 K.

1. Berman, P. (ed.) *Cavity Quantum Electrodynamics* (Academic, 1994).
2. Noda, S. Seeking the ultimate nanolaser. *Science* **314**, 260–261 (2006).
3. McCall, S. L., Levi, A. F. J., Slusher, R. E., Pearton, S. J. & Logan, R. A. Whispering-gallery mode microdisk lasers. *Appl. Phys. Lett.* **60**, 289–291 (1992).
4. Painter, O. *et al.* Two-dimensional photonic band-gap defect mode laser. *Science* **284**, 1819–1821 (1999).
5. Hill, M. T. *et al.* Lasing in metallic-coated nano-cavities. *Nature Photon.* **1**, 589–594 (2007).
6. Walther, C., Scalari, G., Amanti, M. I., Beck, M. & Faist, J. Microcavity laser oscillating in a circuit-based resonator. *Science* **327**, 1495–1497 (2010).
7. Mizrahi, A. *et al.* Low threshold gain metal coated laser nanoresonators. *Opt. Lett.* **33**, 1261–1263 (2008).
8. Nezhad, M. P. *et al.* Room-temperature subwavelength metallo-dielectric lasers. *Nature Photon.* **4**, 395–399 (2010).
9. Yu, K., Lakhani, A. & Wu, M. C. Subwavelength metal-optic semiconductor nanopatch lasers. *Opt. Express* **18**, 8790–8799 (2010).
10. Ding, Q., Mizrahi, A., Fainman Y. & Lomakin, V. Dielectric shielded nanoscale patch laser resonators. *Opt. Lett.* **36**, 1812–1814 (2011).
11. Noginov, M. A. *et al.* Demonstration of a spaser-based nanolaser. *Nature* **460**, 1110–1112 (2009).
12. Oulton, R. F. *et al.* Plasmon lasers at deep subwavelength scale. *Nature* **461**, 629–632 (2009).
13. Burgos, S. P., deWaele, R., Polman, A. & Atwater, H. A. A single-layer wide-angle negative-index metamaterial at visible frequencies. *Nature Mater.* **9**, 407–412 (2010).
14. Jacob, Z. & Shalaev, V. M. Plasmonics goes quantum. *Science* **334**, 463–464 (2011).
15. Vahala, K. J. Optical microcavities. *Nature* **424**, 839–846 (2003).
16. Yokoyama, H. Physics and device applications of optical microcavities. *Science* **256**, 66–70 (1992).
17. Bjork, G. & Yamamoto, Y. Analysis of semiconductor microcavity lasers using rate equations. *IEEE J. Quantum Electron.* **27**, 2386–2396 (1991).
18. Baida, F. I., Belkhir, A. & Van Labeke, D. Subwavelength metallic coaxial waveguides in the optical range: Role of the plasmonic modes. *Phys. Rev. B* **74**, 205419 (2006).
19. Feigenbaum, E. & Orenstein, M. Ultrasmall volume plasmons, yet with complete retardation effects. *Phys. Rev. Lett.* **101**, 163902 (2008).
20. Benzaquen, R. *et al.* Alloy broadening in photoluminescence spectra of $Ga_xIn_{1-x}As_yP_{1-y}$ lattice matched to InP. *J. Appl. Phys.* **75**, 2633–2639 (1994).
21. Bayer, M. *et al.* Inhibition and enhancement of the spontaneous emission of quantum dots in structured microresonators. *Phys. Rev. Lett.* **86**, 3168–3171 (2001).
22. Vuckovic, J., Painter, O., Xu, Y., Yariv, A. & Scherer, A. Finite-difference time-domain calculation of the spontaneous emission coupling factor in optical microcavities. *IEEE J. Quantum Electron.* **35**, 1168–1175 (1999).
23. Schawlow, A. L. & Townes, C. H. Infrared and optical masers. *Phys. Rev.* **112**, 1940–1949 (1958).
24. Henry, C. Theory of the linewidth of semiconductor lasers. *IEEE J. Quantum Electron.* **18**, 259–264 (1982).
25. Björk, G., Karlsson, A. & Yamamoto, Y. On the linewidth of microcavity lasers. *Appl. Phys. Lett.* **60**, 304–306 (1992).
26. Rice, P. R. & Carmichael, H. J. Photon statistics of a cavity-QED laser: a comment on the laser-phase-transition analogy. *Phys. Rev. A* **50**, 4318–4329 (1994).
27. Pedrotti, L. M., Sokol, M. & Rice, P. R. Linewidth of four-level microcavity lasers. *Phys. Rev. A* **59**, 2295–2301 (1999).
28. Roy-Choudhury, K. & Levi, A. F. J. Quantum fluctuations and saturable absorption in mesoscale lasers. *Phys. Rev. A* **83**, 043827 (2011).
29. Strauf, S. *et al.* Self-tuned quantum dot gain in photonic crystal lasers. *Phys. Rev. Lett.* **96**, 127404 (2006).
30. Chang, S. W. & Chuang, S. L. Fundamental formulation for plasmonic nanolasers. *IEEE J. Quantum Electron.* **45**, 1014–1023 (2009).

LETTER

Supercontinent cycles and the calculation of absolute palaeolongitude in deep time

Ross N. Mitchell[1], Taylor M. Kilian[1] & David A. D. Evans[1]

Traditional models of the supercontinent cycle predict that the next supercontinent—'Amasia'—will form either where Pangaea rifted (the 'introversion'[1] model) or on the opposite side of the world (the 'extroversion'[2–4] models). Here, by contrast, we develop an 'orthoversion'[5] model whereby a succeeding supercontinent forms 90° away, within the great circle of subduction encircling its relict predecessor. A supercontinent aggregates over a mantle downwelling but then influences global-scale mantle convection to create an upwelling under the landmass[6]. We calculate the minimum moment of inertia about which oscillatory true polar wander occurs owing to the prolate shape of the non-hydrostatic Earth[5,7]. By fitting great circles to each supercontinent's true polar wander legacy, we determine that the arc distances between successive supercontinent centres (the axes of the respective minimum moments of inertia) are 88° for Nuna to Rodinia and 87° for Rodinia to Pangaea—as predicted by the orthoversion model. Supercontinent centres can be located back into Precambrian time, providing fixed points for the calculation of absolute palaeolongitude over billion-year timescales. Palaeogeographic reconstructions additionally constrained in palaeolongitude will provide increasingly accurate estimates of ancient plate motions and palaeobiogeographic affinities.

Two hypotheses have been proposed for the organizing pattern of successive supercontinents. 'Introversion' is the model whereby the relatively young, interior ocean stops spreading and closes such that a successor supercontinent forms where its predecessor was located[1]. 'Extroversion' is the model in which the relatively old, exterior ocean closes completely, such that a successor supercontinent forms in the hemisphere opposite to that of its predecessor[2–4]. A third model,

which we call 'orthoversion', predicts that a successor supercontinent forms in the downwelling girdle of subduction orthogonal to the centroid of its predecessor[5]. Hypothetical predictions for each model type can be considered for the future Asia-centred supercontinent, 'Amasia'[8], relative to the location of Pangaea in a deep mantle reference frame. (Amasia will merge the Americas with Asia, including the forward-extrapolated northward motions of Africa and Australia, and possibly include Antarctica.) According to the introversion model, the comparatively young Atlantic Ocean will close and Amasia will be centred more or less where Pangaea was (Fig. 1a). According to the extroversion model, the comparatively old Pacific Ocean will close and Amasia will be centred on the opposite side of the world from Pangaea (Fig. 1b). Finally, according to our orthoversion model, the Americas will remain in the Pacific 'ring of fire' girdle of post-Pangaean subduction, closing the Arctic Ocean and Caribbean Sea (Fig. 1c).

If any one model can be empirically demonstrated, then not only can we speculatively forecast where and how Amasia will form, but also we can extrapolate palaeogeography, including the historically elusive palaeolongitude, backwards in Earth history, from supercontinent to supercontinent. Using our orthoversion model, we find that Pangaea orthoverted from Rodinia, and Rodinia orthoverted from Nuna. Extrapolating this model into the future, Amasia should be centred within Pangaea's subduction girdle. Orthoversion helps to resolve the problems of the popular introversion and extroversion models, which have led to a "fundamental disconnection … between the geologic evidence for supercontinent formation, and the models purported to explain their assembly"[9].

Figure 1 | Supercontinent cycle hypotheses. Predicted locations of the future supercontinent Amasia, according to three possible models of the supercontinent cycle: **a**, introversion; **b**, extroversion, and **c**, orthoversion. The labelled centres of Pangaea and Rodinia are the conjectured locations of each supercontinent's I_{min} (Fig. 2). Yellow equatorial circles represent supercontinent-induced mantle upwellings, and the orthogonal blue great circle swath represents Pangaea's subduction girdle (as in Fig. 3). In **c**, Amasia could be centred anywhere along Pangaea's subduction girdle. Red arrows indicate where ocean basins would close according to each model. Continents are shown in present-day coordinates.

[1]Yale University, 210 Whitney Avenue, New Haven, Connecticut 06511, USA.

Several independent methods are available to estimate the centre of supercontinent Pangaea (see Methods). However, only the palaeomagnetic identification of ancient true polar wander (TPW)—the rotation of solid Earth about the equatorial minimum moment of inertia[10], I_{min}—allows us to measure the angle between successive supercontinents in deep time. For a long-lived (more than a hundred million years) prolate Earth for which the intermediate and maximum moments of inertia are subequal and prone to interchange, I_{min} provides a quantitative datum for the mantle convective planform beneath a supercontinent's centre (see Methods). Recently, Steinberger and Torsvik[7] determined the palaeolongitude of the Pangaean I_{min} by identifying four Mesozoic TPW oscillatory swings about nearly the same axis near central Africa.

We fitted great circles, and their orthogonal axes defining I_{min}, to the TPW-rich portion (260–90 million years (Myr) ago) of the global palaeomagnetic apparent polar wander (APW) path in a South African reference frame[7] (Fig. 2). One great-circle fit to all the 260–90 Myr-old poles does not convey the true azimuths of the individual TPW swings, which are subparallel. Because plate motions as well as the TPW signal are included in the APW paths, continental drift relative to the stationary I_{min} will appear in a continental reference frame as the TPW great-circle segments shifting with age. We therefore fitted two great circles, one to 260–220-Myr-old poles (20° N, 349° E, $A_{95} = 3°$ (error), $N = 5$ (sample)) and one to 200–90-Myr-old poles (10° N, 001° E, $A_{95} = 4°$, $N = 12$; light and dark blue, respectively, in Fig. 2). The two great circles are proposed to be caused by oscillations about the same Pangaean I_{min} axis, but are distinct geographically owing to the movement of the South African reference frame relative to the stable I_{min}.

Before these Africa-centred TPW rotations, Gondwanaland experienced early Palaeozoic oscillatory rotations around a distinctly different axis. Instead of the African region swiveling in azimuth in constantly tropical latitudes as described above, early Palaeozoic rotations involved rapid changes of palaeolatitude for the African and South American regions of the large continent. These rotations, about an axis near the Australian sector of Gondwanaland, have also been attributed to TPW[11,12] and closely match the motions produced by migrations of ice centres across the drifting supercontinent[13]. Continuing backwards in time into the Ediacaran period, additional large-magnitude rotations recorded in the Australian palaeomagnetic database[14] suggest similar TPW-dominated kinematics for that time.

In South African coordinates, the earliest Palaeozoic I_{min} (−30° N, 075° E, $A_{95} = 12°$, $N = 14$) plots near the reconstructed Australian sector of the supercontinent (Fig. 2a). The results of this calculation are similar whether only Cambrian data are considered, or only Ediacaran data, or the combined Ediacaran–Cambrian data set. The angular distance between the successive I_{min} locations from 550–490 Myr ago and 260–220 Myr ago, in the same reference frame, is 83 ± 15 degrees. In principle, this could represent the steady drift of Gondwanaland over the mantle, including a stationary I_{min} throughout Phanerozoic time[15,16]. However, the rapidity of the shift between 370 and 260 Myr ago would suggest rates of motion averaging about $10 \ cm \ yr^{-1}$ (see the 'alternative animation' in the Supplementary Information), which would be unusual for a continent of that size[17]. We consider it more likely that I_{min} shifted substantially relative to Gondwanaland, owing to the post-Pangaean mantle axis being created in a position orthogonal to that of its post-Rodinian predecessor. The orthoversion model of supercontinent succession neatly explains this result.

With its perimeter surrounded by Neoproterozoic rifted to passive margins, Laurentia occupies a central place in most Rodinia reconstructions[18,19], akin to Africa in Pangaea. The orthoversion model predicts that soon after Rodinia's assembly, the young supercontinent's centre could have experienced oscillatory TPW with large changes in palaeolatitude around an axis corresponding to the preceding Nuna supercontinent's[20] convection-driven I_{min}. This would be

Figure 2 | Supercontinent centres. a, Successive post-Rodinia (orange and green open ellipses) and post-Pangaea (light-blue and dark-blue open ellipses) I_{min} axes are calculated as the poles to great-circle fits of palaeomagnetic poles from Australia at 650–560 Myr ago (orange solid ellipses) and Gondwanaland from 550–490 Myr ago (light-green solid ellipses), and the global running-mean apparent polar wander path for 260–220 Myr ago (light-blue solid ellipses) and 210–90 Myr ago (dark-blue solid ellipses)[24]. Dark-green poles for later Palaeozoic time from Gondwanaland are displayed but not included in any mean calculation (see text for discussion). **b,** Successive post-Nuna (red open ellipse) and post-Rodinia (orange open ellipse) I_{min} axes. Filled red ellipses are poles for Laurentia from 1,165–1,015 Myr ago. Filled orange ellipses are poles for Laurentia rotated from Svalbard at around 800 Myr ago[18] (see Methods section for discussion of rotation). All ellipses are projections of cones of 95% confidence. Pole information is listed in Supplementary Table 1, statistical parameters are detailed in Supplementary Table 2, and a version of this figure with the poles numbered to give a sense of age order is provided in Supplementary Fig. 1.

followed by mantle convective reorganization to a new super-continent-centred I_{min} location with swivel-like TPW oscillations at constantly tropical latitudes during Rodinia break-up. The palaeomagnetic record for Laurentia during Meso–Neoproterozoic time, the crossover interval between Nuna and Rodinia TPW legacies[5], shows this pattern.

The Rodinia interval is marked by rapid, oscillatory continental motions that have been interpreted as TPW during the time of supercontinent amalgamation at around 1,100–1,000 Myr ago[5,21] and break-up at around 800 Myr ago[18,22]. We fitted two great circles, for 1,165–1,015 Myr ago[18] ($-28°$ N, $263°$ E, $A_{95} = 6°$, $N = 19$) and for three approximately 800-Myr-old poles from Svalbard[18,23] rotated to Laurentia ($59°$ N, $293°$ E, $A_{95} = 17°$, $N = 3$). The angle between the two successive I_{min} axes of Nuna and Rodinia is $89° \pm 23°$ (Fig. 2b).

As with Gondwanaland in Palaeozoic–Mesozoic time, the large shift in location of I_{min} relative to Laurentia can be interpreted in principle as the motion of the Rodinia supercontinent over a single, long-lived, mantle-stationary inertial axis. However, the orthoversion model also neatly explains the data, invoking a post-Rodinia I_{min} axis created 90° away from that of its predecessor, Nuna. Given that the two kinematically quantifiable supercontinental transitions (Nuna to Rodinia, and Rodinia to Pangaea) are both characterized by nearly ideal 90° shifts in location of the I_{min} axes, we conclude that orthoversion is the most parsimonious model for supercontinental cyclicity through the past billion years.

The orthoversion model of the supercontinent cycle makes palaeogeographic predictions deep into Earth history, from supercontinent to supercontinent, that include historically elusive absolute palaeolongitude constraints. Continents can be reconstructed latitudinally and longitudinally relative to supercontinent centres, as determined by a supercontinent's TPW legacy (fixed I_{min}). Figure 3 consists of five global maps in 200-million-year intervals, including simplified schematic mantle convection planforms through time. Given that the actual measured angles between supercontinent centres are within a few degrees of 90°, we choose to assume in our reconstructions that successive supercontinents are ideally orthogonal, that is, successive I_{min} axes are offset by exactly 90° in palaeolongitude. Such ideality conforms to the self-organizing behaviour of mantle convection towards predominantly degree-2 spherical harmonics[6].

Absolute reconstructions are provided back to 500 Myr ago in an animation in the Supplementary Information and back to 800 Myr ago for select continents in Fig. 3. Reconstructions from 320 Myr ago to the present are identical to the configurations of ref. 24 except that our solutions additionally track the drift of South Africa relative to the long-lived Pangaean I_{min}. By aligning successive I_{min} axes from historical Pangaean TPW (260–90 Myr ago[24]) with the post-Pangaean I_{min} currently at 0° N, 010° E (ref. 25), we reconstruct all continents including South Africa with respect to present-day (or 'absolute') latitude and longitude coordinates.

Before 320 Myr ago, absolute reconstructions are limited to those continents for which TPW segments have been identified. The post-Rodinian I_{min} axis (650–490 Myr old) is ideally shifted in our model to a location 90° in longitude from the post-Pangaean axis (Fig. 3). Of the two possible orthoverted equatorial axes, 100° E and 80° W, we choose 100° E to minimize plate-tectonic drift rates of large, continent-bearing plates. We note that early Palaeozoic kimberlites and large igneous provinces, particularly widespread in Siberia and Australia, reconstruct within or near the idealized I_{min} circles in our model, consistent with their derivation from plume-generating zones in the deep mantle[16]. At 600 Myr ago, just before Gondwanaland assembly, Australia is reconstructed relative to the I_{min} axis according to its palaeomagnetic data (Supplementary Table 2), as is Laurentia, assuming that TPW is responsible for the bulk of variance in its palaeomagnetic poles[26]. At 800 Myr ago, we reconstructed Rodinia, according to ref. 19, around Laurentia, which is fixed to the I_{min} axis by its restored Svalbard palaeomagnetic data.

Figure 3 | Absolute palaeogeographic maps. Since 260 Myr ago, each I_{min} about which TPW occurred is pinned at 0° N, 10° E. Before 260 Myr ago, continents are rotated in palaeolongitude such that the Rodinian I_{min} is ideally 'orthoverted' at 0° N, 100° E. Yellow equatorial circles represent supercontinent-induced mantle upwellings (not showing the antipodal upwellings such as under the Pacific Ocean), and orthogonal blue great-circle swaths represent subduction girdles (as in Fig. 1). See text for details and Supplementary Tables 3, 4 and 5 for absolute reconstruction parameters. An animation for the past 500 million years is also included in the Supplementary Information.

If a supercontinent-induced two-cell mantle topology[6] drives the supercontinent cycle by orthoversion, can we predict plate motions during supercontinental transitions? Generally, orthoversion may not be expected to disaggregate a supercontinent entirely in order to form its successor because the new centroid is only a half-hemisphere away (as opposed to the extroversion model, for example). It would not be possible to predict, however, which newly rifted continent that had been peripheral to the predecessor would become the central

nucleation point for the succeeding supercontinent. The orthoversion pattern is perhaps best embodied in the present tectonic transition from Pangaea to Amasia, in which rifted fragments of Gondwanaland are reassembling in Eurasia[27]: most recently India and Arabia, imminently Africa, more distantly Australia, and possibly Antarctica (Fig. 1c). In particular, the most distant continent, Australia, advanced eastward only to the circum-Pangaean subduction girdle[28] before turning northward and accelerating towards Asia[24].

Two related implications of the orthoversion model of the supercontinent cycle concern mantle convection. First, orthoversion provides the missing geodynamic model to explain the enigmatic closure of the early Palaeozoic Rheic–Iapetus ocean system and thus resolve the Pangaea "conundrum"[9]: the Rheic–Iapetus oceanic tract originated about 90° away from Rodinia's centroid (Fig. 3) and was thus destined for continent–continent collision and a central position in Pangaea, irrespective of its young age. One can regard the Indian Ocean as a present-day Iapetus–Rheic-like young oceanic system that opens and closes in a single hemisphere, as the ring of subduction around the rifting supercontinent prevents the Indian Ocean from widening further. Rifted terranes, like Avalonia and Carolinia in the Iapetus–Rheic oceanic system, and India and the multitude of other Eurasian blocks in the Tethys-Indian oceanic system, traverse the young ocean system only to reassemble in the broad subduction girdle inherited from the Pangaean two-cell convective planform[6].

Second, the orthoversion model implies that the antipodal upwellings underneath the African and Pacific plates today have existed only since the creation of Pangaea[29,30], not earlier[15,16]. Reorganization of global mantle convection by only 90° every 700 million years is a slow enough process to distinguish long-lived geochemical tracers from separate reservoirs in mantle-derived basalts[15,31,32] and also to accommodate the observed sizes of African and Pacific large low-velocity provinces[33] in the context of reasonable amounts of entrainment by normal rates of whole-mantle convection through hundreds of millions of years.

METHODS SUMMARY

For each of the six time intervals, I_{min} is calculated as the pole to the best-fitted great circle to a swath of palaeomagnetic poles (Fig. 2; Supplementary Table 1) relative to a given reference frame during proposed intervals of TPW: 200–90 Myr relative to South Africa, 260–220 Myr relative to South Africa, 550–490 Myr relative to South Africa, 650–560 Myr relative to South Africa, 805–790 Myr relative to Laurentia, and 1,165–1,015 Myr relative to Laurentia (Supplementary Table 2). Reconstructions from 260 Myr ago to the present are taken from ref. 24, modified slightly in absolute palaeolongitude such that the I_{min} axes align with 010° E (ref. 25) throughout each of the two TPW-defined time intervals (Supplementary Table 3). For 500–370 Myr ago, the four continents Gondwanaland (reconstructed in Supplementary Table 4), Siberia, Baltica and Laurentia are constrained in palaeolatitude according to palaeomagnetic poles (Supplementary Table 1) in 20-Myr running-mean APW paths from various summary models (Supplementary Table 5), and constrained in palaeolongitude according to our idealized orthoversion model of early Palaeozoic TPW around 100° E (including an additional proposed TPW oscillation at 450–375 Myr ago[12]). The animation in the Supplementary Information from 500 Myr ago to the present uses our global rotation model, which incorporates kinematic interpolations seeking to minimize rates of absolute motions between TPW-defined intervals, while also minimizing areas of cratonic overlap and conforming to the global tectonic record. This model is formatted (Supplementary Table 6) for the GPlates freeware (www.gplates.org), which provides continuous kinematic interpolation shown at two-million-year intervals (animation in the Supplementary Information). The 800-Myr and 600-Myr reconstructions (Fig. 3) are slightly modified from ref. 19.

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

1. Nance, R. D., Worsley, T. R. & Moody, J. B. The supercontinent cycle. *Sci. Am.* **259**, 72–79 (1988).
2. Hartnady, C. J. H. About turn for supercontinents. *Nature* **352**, 476–478 (1991).
3. Hatton, C. J. The superocean cycle. *S. Afr. J. Geol.* **100**, 301–310 (1997).
4. Veevers, J. J., Walter, M. R. & Scheibner, E. Neoproterozoic tectonics of Australia—Antarctica and Laurentia and the 560 Ma birth of the Pacific Ocean reflect the 400 m.y. Pangean supercycle. *J. Geol.* **105**, 225–242 (1997).
5. Evans, D. A. D. True polar wander and supercontinents. *Tectonophysics* **362**, 303–320 (2003).
6. Zhong, S. J., Zhang, N., Li, Z. X. & Roberts, J. H. Supercontinent cycles, true polar wander, and very long-wavelength mantle convection. *Earth Planet. Sci. Lett.* **261**, 551–564 (2007).
7. Steinberger, B. & Torsvik, T. H. Absolute plate motions and true polar wander in the absence of hotspot tracks. *Nature* **452**, 620–623 (2008).
8. Hoffman, P. F. in *Earth Structure: An Introduction to Structural Geology and Tectonics* (eds van der Pluijm, B. & Marshak, S.) 459–464 (McGraw-Hill, 1997).
9. Murphy, J. B., Nance, R. D. & Cawood, P. A. Contrasting modes of supercontinent formation and the conundrum of Pangea. *Gondwana Res.* **15**, 408–420 (2009).
10. Gold, T. Instability of the Earth's axis of rotation. *Nature* **175**, 526–529 (1955).
11. Mitchell, R. N., Evans, D. A. D. & Kilian, T. M. Rapid Early Cambrian rotation of Gondwana. *Geology* **38**, 755–758 (2010).
12. Van der Voo, R. True polar wander during the middle Paleozoic? *Earth Planet. Sci. Lett.* **122**, 239–243 (1994).
13. Caputo, M. V. & Crowell, J. C. Migration of glacial centers across Gondwana during Paleozoic Era. *Geol. Soc. Am. Bull.* **96**, 1020–1036 (1985).
14. Schmidt, P. W. & Williams, G. E. Ediacaran palaeomagnetism and apparent polar wander path for Australia: no large true polar wander. *Geophys. J. Int.* **182**, 711–726 (2010).
15. Burke, K., Steinberger, B., Torsvik, T. & Smethurst, M. Plume generation zones at the margins of large low shear velocity provinces on the core–mantle boundary. *Earth Planet. Sci. Lett.* **265**, 49–60 (2008).
16. Torsvik, T., Burke, K., Steinberger, B., Webb, S. J. & Ashwal, L. D. Diamonds sampled by plumes from the core-mantle boundary. *Nature* **466**, 352–355 (2010).
17. Forsyth, D. & Uyeda, S. On the relative importance of the driving forces of plate motion. *Geophys. J. Int.* **43**, 163–200 (1975).
18. Maloof, A. C. *et al.* Combined paleomagnetic, isotopic, and stratigraphic evidence for true polar wander from the Neoproterozoic Akademikerbreen Group, Svalbard, Norway. *Geol. Soc. Am. Bull.* **118**, 1099–1124 (2006).
19. Li, Z. X. *et al.* Assembly, configuration, and break-up history of Rodinia: a synthesis. *Precambr. Res.* **160**, 179–210 (2008).
20. Evans, D. A. D. & Mitchell, R. N. Assembly and breakup of the core of Paleoproterozoic-Mesoproterozoic supercontinent Nuna. *Geology* **39**, 443–446 (2011).
21. Swanson-Hysell, N. L., Maloof, A. C., Weiss, B. P. & Evans, D. A. D. No asymmetry in geomagnetic reversals recorded by 1.1-billion-year-old Keweenawan basalts. *Nat. Geosci.* **2**, 713–717 (2009).
22. Li, Z. X., Evans, D. A. D. & Zhang, S. A 90 degrees spin on Rodinia: possible causal links between the Neoproterozoic supercontinent, superplume, true polar wander and low-latitude glaciation. *Earth Planet. Sci. Lett.* **220**, 409–421 (2004).
23. Macdonald, F. A. *et al.* Calibrating the Cryogenian. *Science* **327**, 1241–1243 (2010).
24. Torsvik, T. H., Müller, R. D., Van der Voo, R., Steinberger, B. & Gaina, C. Global plate motion frames: toward a unified model. *Rev. Geophys.* **46**, http://dx.doi.org/10.1029/2007RG000227 (2008).
25. Pavoni, N. Present true polar wander in the frame of the Geotectonic reference system. *Swiss J. Geosci.* **101**, 629–636 (2008).
26. McCausland, P. J. A., Hankard, F., Van der Voo, R. & Hall, C. M. Ediacaran paleogeography of Laurentia: paleomagnetism and 40Ar-39Ar geochronology of the 583 Ma Baie des Moutons syenite, Quebec. *Precambr. Res.* **187**, 58–78 (2011).
27. Sengör, A. M. C. Mid-Mesozoic closure of Permo-Triassic Tethys and its implications. *Nature* **279**, 590–593 (1979).
28. Collins, W. J. Slab pull, mantle convection, and Pangaean assembly and dispersal. *Earth Planet. Sci. Lett.* **205**, 225–237 (2003).
29. Evans, D. A. D. Proposal with a ring of diamonds. *Nature* **466**, 326–327 (2010).
30. Li, Z. X. & Zhong, S. Supercontinent-superplume coupling, true polar wander and plume mobility: plate dominance in whole-mantle tectonics. *Phys. Earth Planet. Inter.* **176**, 143–156 (2009).
31. Deschamps, F., Kaminski, E. & Tackley, P. J. A deep mantle origin for the primitive signature of ocean island basalt. *Nature Geosci.* **4**, 879–882 (2011).
32. Gonnermann, H. M. & Mukhopadhyay, S. Preserving noble gases in a convecting mantle. *Nature* **459**, 560–563 (2009).
33. Garnero, E. J., Lay, T. & McNamara, A. K. Implications of lower mantle structural heterogeneity for existence and nature of whole mantle plumes. *Geol. Soc. Am. Spec. Pap.* **430**, 79–101 (2007).

Author Contributions R.N.M. developed the conceptual idea for the study, D.A.D.E. assembled input data and developed statistical methods, and T.M.K. executed calculations and generated reconstructions and animations. All authors contributed to discussions and the writing of the manuscript.

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

METHODS

To test between introversion (0°), extroversion (180°), or orthoversion (90°) models of the supercontinent cycle, one must quantitatively determine the centres of supercontinents and measure the angular distance between successive supercontinent centres. Pangaea's centre can be determined by post-Pangaean seafloor-spreading reconstructions of continents and large igneous provinces, allowing for precise determination of the supercontinent's centre of mass[24] or four oscillatory rotations shared by all continents about an equatorial Euler pole (which is I_{min} according to the TPW hypothesis) in the aftermath of Pangaea[7]; this I_{min} axis closely coincides with two antipodal large low-shear-wave-velocity provinces at the core–mantle boundary underneath Africa and the Pacific imaged by present-day seismic tomography[34].

For each of the six time intervals, I_{min} is calculated as the pole to the best-fit great-circle to a swath of palaeomagnetic poles (Fig. 2; Supplementary Table 1) relative to a given reference frame during proposed intervals of TPW: 200–90 Myr relative to South Africa, 260–220 Myr relative to South Africa, 550–490 relative to South Africa, 650–560 Myr relative to South Africa, 805–790 Myr relative to Laurentia, and 1,165–1,015 Myr relative to Laurentia (Supplementary Table 2). We limit our calculation of Rodinia's I_{min} to the 650–360-Myr APW path for Gondwanaland and Australia alone before the Early Cambrian period (Fig. 2a). Poles from Gondwanaland are rotated into South African coordinates (Supplementary Table 4). The Rodinian I_{min} for Laurentia (Fig. 2b) is affected by the rotation of Svalbard to Laurentia[18,35] but our results do not change significantly if geologically reasonable juxtapositions are considered. Confidence limits on the poles to great circles (Supplementary Table 1) are calculated using the software package of ref. 36 employing two alternative statistical methods[37,38]. For the 800-Myr I_{min} calculation, the method of ref. 37 cannot be used for $N < 4$ and so we use the mean angular deviation method of ref. 38, which probably overestimates error. Errors on and angular distances between successive I_{min} axes were calculated by numerical bootstrap methods following ref. 39.

Reconstructions from 260 Myr ago to the present are taken from ref. 24, modified slightly in absolute palaeolongitude such that the I_{min} axes align with 010° E (ref. 25) throughout each of the two TPW-defined time intervals (Supplementary Table 3). For 500–370 Myr ago, the four continents Gondwanaland (reconstructed in Supplementary Table 4), Siberia, Baltica and Laurentia are constrained in palaeolatitude according to palaeomagnetic poles (Supplementary Table 1) in 20-Myr running-mean APW paths from various summary models (Supplementary Table 5), and in palaeolongitude according to our idealized orthoversion model of early Palaeozoic TPW around 100° E (including an additional proposed TPW oscillation at 450–375 Myr ago[12]. The animation from 500 Myr ago to the present (in the Supplementary Information) uses our global rotation model, which incorporates kinematic interpolations seeking to minimize rates of absolute motions between TPW-defined intervals, while also minimizing areas of cratonic overlap and conforming to the global tectonic record. Minor problems with overlapping plates in Pangaea, which have generated discussion on non-dipole geomagnetic field behaviour[40] are taken 'as is' from the smoothed pole paths without correction. This model is formatted (Supplementary Table 6) for the GPlates freeware (www.gplates.org) that provides continuous kinematic interpolation shown at two-million-year intervals (animation in the Supplementary Information). An alternative animation in the Supplementary Information demonstrates that the solution where I_{min} is held constant, prompted by the model where geoid highs are stable through time[16], involves much more east–west motion than the orthoversion model. The fixed I_{min} solution requires about 60° of east–west motion over 60 Myr.

Kinematics before 500 Myr ago are more speculative, and the illustrated snapshots at 600 and 800 Myr ago (Fig. 3) are merely indicative of plausible global palaeogeographies. The 800- and 600-Myr-ago reconstructions are slightly modified from ref. 19. Aside from Gondwanaland and Laurentia at 600 Myr ago, other cratons are reconstructed according to a recent model for Rodinia break-up[41].

34. Becker, T. W. & Boschi, L. A comparison of tomographic and geodynamic mantle models. *Geochem. Geophys. Geosyst.* **3**, http://dx.doi.org/10.1029/2001GC000168 (2002).

35. Evans, D. A. D. The paleomagnetically viable, long-lived and all-inclusive Rodinia supercontinent reconstruction. *Geol. Soc. Lond. Spec. Publ.* **327**, 371–404 (2009).

36. Cogne, J. P. PaleoMac; a Macintosh™ application for treating paleomagnetic data and making plate reconstructions. *Geochem. Geophys. Geosyst.* **4**, 1007 (2003).

37. Mardia, K. V. & Gadsden, R. J. A small circle of best fit for spherical data and areas of vulcanism. *J. R. Statist. Soc. C* **26**, 238–245 (1977).

38. Kirschvink, J. L. The least-squares line and plane and the analysis of palaeomagnetic data. *Geophys. J. R. Astron. Soc.* **62**, 699–718 (1980).

39. McFadden, P. L. A new fold test for palaeomagnetic studies. *Geophys. J. Int.* **103**, 163–169 (1990).

40. Torsvik, T. & Van der Voo, R. Refining Gondwana and Pangea palaeogeography: estimates of Phanerozoic non-dipole (octupole) fields. *Geophys. J. Int.* **151**, 771–794 (2002).

41. Li, Z. X. & Evans, D. A. D. Late Neoproterozoic 40° intraplate rotation within Australia allows for a tighter-fitting and longer-lasting Rodinia. *Geology* **39**, 39–42 (2011).

LETTER

Genetic contributions to stability and change in intelligence from childhood to old age

Ian J. Deary[1,2]*, Jian Yang[3]*, Gail Davies[1,2], Sarah E. Harris[2,4], Albert Tenesa[4,5], David Liewald[1,2], Michelle Luciano[1,2], Lorna M. Lopez[1,2], Alan J. Gow[1,2], Janie Corley[1], Paul Redmond[1], Helen C. Fox[6], Suzanne J. Rowe[5], Paul Haggarty[7], Geraldine McNeill[6], Michael E. Goddard[8], David J. Porteous[2,4], Lawrence J. Whalley[6], John M. Starr[2,9] & Peter M. Visscher[2,3,10,11]*

Understanding the determinants of healthy mental ageing is a priority for society today[1,2]. So far, we know that intelligence differences show high stability from childhood to old age[3,4] and there are estimates of the genetic contribution to intelligence at different ages[5,6]. However, attempts to discover whether genetic causes contribute to differences in cognitive ageing have been relatively uninformative[7–10]. Here we provide an estimate of the genetic and environmental contributions to stability and change in intelligence across most of the human lifetime. We used genome-wide single nucleotide polymorphism (SNP) data from 1,940 unrelated individuals whose intelligence was measured in childhood (age 11 years) and again in old age (age 65, 70 or 79 years)[11,12]. We use a statistical method that allows genetic (co)variance to be estimated from SNP data on unrelated individuals[13–17]. We estimate that causal genetic variants in linkage disequilibrium with common SNPs account for 0.24 of the variation in cognitive ability change from childhood to old age. Using bivariate analysis, we estimate a genetic correlation between intelligence at age 11 years and in old age of 0.62. These estimates, derived from rarely available data on lifetime cognitive measures, warrant the search for genetic causes of cognitive stability and change.

General cognitive ability (also known as general intelligence, or g[18]) is an important human trait. It shows consistent and strong associations with important life outcomes such as educational and occupational success, social mobility, health, illness and survival[18]. Maintaining good general cognitive ability in old age is associated with better physical health and the ability to carry out everyday tasks[19,20]. Intelligence differences are highly heritable from adolescence, and through adulthood to old age[5,6]. Long-term follow-up studies have shown that about half of the phenotypic variance in general intelligence in old age is accounted for by its measure in childhood[3,4]. The corollary of this is that there are systematic changes through the life course in the rank order of intelligence between people; that is, some people's intelligence ages better than others. The determinants of stability and change in intelligence across the human life course are being sought, and candidate determinants include a wide range of genetic and environmental factors[1,5,7,19,21,22]. There have been longitudinal studies within childhood/adolescence, middle adulthood and old age, but none that stretches from childhood to old age with the same individuals (to our knowledge). Until now, the proportion of the variance in lifetime cognitive stability and change explained by genetic and environmental causes has been almost unknown. Apart from a small contribution from variation in the *APOE* gene, suggested individual genetic contributions to stability and change in intelligence across the

life course are largely unreplicated[22]. Therefore, an important novel contribution would be to partition the covariance between intelligence scores at either end of the human life course into genetic and environmental causes. To address this, the present study applies a new analytical method[13–17] to genome-wide association data from human participants with general cognitive ability test scores in childhood and again in old age.

Participants were members of the Aberdeen Birth Cohort 1936 (ABC1936) and the Lothian Birth Cohorts of 1921 and 1936 (LBC1921, LBC1936)[11,12,17]. They are community-dwelling, surviving members of the Scottish Mental Surveys of 1932 (the 1921-born individuals) and 1947 (the 1936-born individuals), in which they took a well-validated test of general intelligence (Moray House Test) at a mean age of 11 years. They were traced and re-tested again in old age on a large number of medical and psychosocial factors for studies of healthy mental and physical ageing. Here, we use cognitive ability test data from childhood and from the first occasion of testing in old age for each subject. For all three cohorts, cognitive ability in old age was measured using the first unrotated principal component from a number of diverse cognitive tests. Additionally, the LBC1921 and LBC1936 cohorts re-took the Moray House Test in old age. Thus, the present study partitions into genetic and environmental causes the variance in stability and change in general intelligence over a period of between 54 and 68 years. Testing for 599,011 SNPs was performed on the Illumina610-Quadv1 chip (Illumina); the genotyping of the samples in this study was described previously[17] and quality control is described in Methods Summary.

To estimate additive genetic and environmental contributions to variation in cognitive ageing we used genotype information from 536,295 genome-wide autosomal SNPs. The method used here is a multivariate extension of our recently developed method, which allows the estimation of distant relationships between conventionally unrelated individuals from the SNP data and correlates genome-wide SNP similarity with phenotypic similarity[13,15]. A detailed description of the overall approach and statistical methods is given in Supplementary Fig. 1 and the Supplementary Note. We used a linear mixed model to estimate variance components. The methodology for the estimation of genetic variation from population samples was described previously and has been applied to continuous traits, including height, body-mass index and cognitive ability[13,15–17], and to disease[23]. The method is analogous to a pedigree analysis, with the important difference that we estimate distant relatedness from SNP markers. Because the relationships are estimated from common SNP markers, phenotypic variance explained by such estimated relationships is due to linkage

[1]Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. [2]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. [3]Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4006, Australia. [4]Medical Genetics Section, Molecular Medicine Centre, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh EH4 2XU, UK. [5]The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Roslin, Edinburgh EH25 9RG, UK. [6]Institute of Applied Health Sciences, Foresterhill, Aberdeen AB25 2ZD, UK. [7]Nutrition and Epigenetics Group, Rowett Institute of Nutrition and Health, University of Aberdeen, Greenburn Road, Bucksburn, Aberdeen AB21 9SB, UK. [8]Department of Food and Agricultural Systems, University of Melbourne, Parkville, Victoria 3011, Australia and Victorian Department of Primary Industries, Bundoora, Victoria 3083, Australia. [9]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. [10]University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland 4102, Australia. [11]The Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia.
*These authors contributed equally to this work.

disequilibrium between the genotyped markers and unknown causal variants[13,14,21]. The method estimates genetic variation from SNPs that are in linkage disequilibrium with unknown causal variants, and so provides a lower limit of the total narrow sense heritability because additive variation due to variants that are not in linkage disequilibrium with the genotyped SNPs is not captured.

We first performed a univariate analysis of cognitive ageing (Supplementary Note), which we had defined previously as intelligence scores in old age phenotypically adjusted for intelligence at childhood, by fitting the Moray House Test of intelligence at age 11 as a linear covariate[24]. We estimated that 0.24 (standard error 0.20) of phenotypic variance in cognitive ageing was accounted for by the SNP-based similarity matrix. We next conducted a bivariate genetic analysis of intelligence scores early and later in life, to partition the observed phenotypic covariance in intelligence measured in childhood and old age into genetic and environmental sources of variation. Information on the environmental correlation comes from the comparison of the two phenotypes within individuals whereas the genetic correlation is inferred from between-individual comparisons of the two phenotypes (Supplementary Note). That is, the analysis can inform us about genetic and environmental contributions to stability and change in intelligence across the life course. The phenotypic correlation between Moray House Test intelligence at age 11 and the general intelligence component in old age was 0.63 (standard error 0.02) (Table 1). The bivariate analysis resulted in estimates of the proportion of phenotypic variation explained by all SNPs for cognition, as follows: 0.48 (standard error 0.18) at age 11; and 0.28 (standard error 0.18) at age 65, 70 or 79 (referred to hereafter as 65–79). The genetic correlation between these two traits was 0.62 (standard error 0.22), and the environmental correlation was 0.65 (standard error 0.12). From the results of the bivariate analyses we can make a prediction about the proportion of phenotypic variance explained by the SNPs for cognition at 65–79 years given the phenotype at age 11 years. This provided a prediction of 0.21 (standard error 0.20), which is consistent with the actual estimate of 0.24 (standard error 0.20) from the univariate analysis (Supplementary Table 1), suggesting that the bivariate normal distribution assumption underlying the bivariate analysis is reasonable. Hence, the results from the bivariate analysis contain the full description of the genetic and environmental relationships between cognition at childhood, cognition at old age, and cognitive change. We re-ran this model with different cut-offs for relatedness (Supplementary Table 2). The estimates are very similar but with, as expected, larger standard errors for more stringent cut-offs, which result in a smaller sample size. This shows that the results are not driven by unusually high correlations for a few close relatives.

In the present analyses we did not adopt the usual procedure of dividing the parameter estimates by the standard errors to obtain test statistics and accompanying P values, because the standard errors were derived from a first-order Taylor series of the logarithm of the likelihood about the parameter estimates[25] and these can be biased for modest sample sizes. A more appropriate procedure is to use the likelihood-ratio test statistic to test the hypotheses that the genetic correlation coefficient is zero (no genetic correlation) or 1 (perfect genetic correlation). When using a likelihood-ratio test, the estimated genetic correlation coefficient of 0.62 has a borderline significant difference from zero (likelihood-ratio test statistic = 2.56, $P = 0.055$, one-sided test) (Supplementary Fig. 2), and does not differ significantly from 1. This was tested by fitting a repeatability model (which implies a genetic correlation of 1.0 and the same heritability of repeat observations) that has three fewer parameters than the full bivariate model. It resulted in a very similar value of the maximum log-likelihood value; the likelihood-ratio test statistic was 5.6 ($P = 0.133$, 3 degrees of freedom) (Supplementary Table 3).

LBC1921 and LBC1936 had the same Moray House Test administered at age 11 and again in old age. The bivariate analyses were repeated, therefore, using the same test of intelligence in childhood and old age in this subsample of the cohorts. The phenotypic correlation between Moray House Test intelligence at age 11 and in old age was 0.68 (standard error 0.01) (Table 1). The bivariate analysis resulted in estimates of the proportion of phenotypic variation explained by all SNPs for the Moray House Test, as follows: 0.30 (standard error 0.23) at age 11; and 0.29 (standard error 0.22) at age 70–79. The genetic correlation between these two traits was 0.80 (standard error 0.27). When using a likelihood-ratio test, the estimated genetic correlation coefficient of 0.80 is not significantly different from zero (likelihood-ratio test statistic = 1.51, $P = 0.11$). The environmental correlation between these two traits was 0.63 (standard error 0.13). From the results of the bivariate analyses we can make a prediction of the proportion of phenotypic variance explained by the SNPs for the Moray House Test at 70–79 years conditional on the phenotype at age 11 years. This results in an estimate of 0.074 (standard error 0.24) (Supplementary Table 4). Although the standard errors of the estimates are larger because a smaller data set was used, the results are similar to those using the full data and it appears that the choice of phenotype at old age (Moray House Test or a linear combination of a number of tests) has not led to a bias in inference. The estimates suggest that cognition early and late in life are similar traits, with possibly some genetic variation for cognitive change.

Using population-based genetic analyses, we have quantified, for the first time, the genetic and environmental contribution to stability and change in intelligence differences for most of the human lifespan. Genetic factors seem to contribute much to the stability of intelligence differences across the majority of the human lifespan. We provide a lower limit of the narrow sense heritability of lifetime cognitive ageing. The point estimate using a general cognitive ability component in old age is 0.24, albeit with a large standard error (0.20). We describe the estimate as a lower limit because the methods used in the present study allow us only to estimate the proportion of the genetic variation contributing to cognitive ageing that is captured by genetic variants in linkage disequilibrium with common SNPs; this will be lower than the total narrow sense heritability. We do not have a good estimate of the total amount of additive genetic variation for cognitive ageing, and so we cannot easily quantify any heritability that is missing from our estimate. Some of the possible genetic contribution we have found to cognitive change might be attributable to developmental change between age 11 and young adulthood. However, the large phenotypic correlation between age 11 and old-age intelligence, and the fact that heritability estimates of general intelligence by age 11 are at about adult levels[5], lead us to posit that most of the genetic variation we have found is a contribution to ageing-related cognitive changes. The estimate of the genetic contribution to lifetime cognitive change was lower when, for a subsample, the same test was used in childhood and old age.

The bivariate analysis conducted here quantifies how differences in intelligence early and late in life are attributable to environmental

Table 1 | Bivariate analysis of intelligence at age 11 and at age 65–79

	Using general intelligence component in old age		Using Moray House Test in old age	
	Estimate	Standard error*	Estimate	Standard error*
h_1^2	0.478	0.177	0.298	0.229
h_2^2	0.280	0.177	0.289	0.221
r_G	0.623	0.218	0.798	0.266
r_e	0.652	0.125	0.630	0.132
r_P	0.627	0.015	0.680	0.014

Where h_1^2 and h_2^2 are variance explained by all SNPs for intelligence at age 11 and old age, respectively; r_G is genetic correlation; r_e is residual correlation; r_P is phenotypic correlation. A total of 1,940 unrelated individuals were included with the general intelligence component phenotype data at childhood (1,830) or old age (1,839) (1,729 individuals had both phenotypes). Of the 1,515 LBC1921 and LBC1936 individuals, there were 1,391 with genetic information and Moray House Test scores both at age 11 and in old age.

* The standard errors are estimated from a first-order Taylor series expansion about the estimated maximum likelihood values and may be biased downwards[25]. For testing hypotheses we have used the likelihood-ratio test statistic, which is more accurate.

or genetic factors. A genetic correlation of zero would imply that intelligence early and late in life are entirely separate traits genetically, and that variation in the change in intelligence from childhood to old age is partly genetic and a function of the heritability of intelligence early and late in life. At the other extreme, a genetic correlation of one implies that the two traits have the same genetic determinants, so that any variation in the change in intelligence between the two stages in life is purely environmental. At conventional levels of significance we could not rule out either a genetic correlation of zero or one; however, our estimates suggest that genetics and environment could each contribute substantially to the covariance between intelligence at age 11 and old age, and that genetic factors might have a role in cognitive change between the two stages of the life course.

The samples studied here comprise the birth cohorts' survivors, those healthy enough to take part in the studies, and people with less cognitive decline. Therefore, we considered whether our estimate of genetic variation at older ages may be biased downwards because of censoring. From life tables officially published by the Scottish Government based on census data, we estimate that the individuals in our oldest sample who were born in 1921 and alive at age 11 are among the ~50% that were still alive at the time of sample collection. We know that lower childhood cognitive ability per se is associated with premature mortality[26], which, of course, our analyses adjust for, as specified in the models. However, because there is a paucity of data about genetic influences on lifetime cognitive change, we have limited information with regard to how these might affect life expectancy. The only way to know across the lifespan would have been if all children (that is, the ones who survived to older ages—whom we know about—and the ones who did not) had been genotyped in 1947. For non-normative (that is, pathological) cognitive change, there are genetic risk factors associated with younger-onset Alzheimer's disease that result in premature mortality, but such strongly heritable disease is rare and the genes do not seem to affect normative cognitive ageing in those aged 70 years and over[22]. Hence, this is not a concern with regard to our analyses. *APOE* ε4 is a well-known risk factor for non-normative cognitive decline, but any differential effect on survival occurs later in life, and is thus unlikely to have resulted in attrition in our cohort. Moreover, *APOE* is in Hardy–Weinberg equilibrium in even our oldest samples[24], supporting this inference. Other known genetic risk factors for Alzheimer's disease have a very small effect on the risk of disease[27]. Hence, a priori, we have nothing to suggest anything but a largely neutral effect of genes that influence cognitive ageing on survival. However, if there is an effect, the example of cognition[26] (by contrast with cognitive change) would suggest that this would be negative, which would somewhat reduce genetic variation in cognitive change across the lifespan among the survivors.

Until now, studies aimed at finding genetic contributions to cognitive ageing have offered little information. They use too-short follow-up periods, thereby providing too small an amount of cognitive change[7,22]. Cognitive assessments tend to be made only within old age, even though cognitive ageing occurs from young adulthood onwards. They are largely based on behavioural data in twin samples rather than information on DNA variation. The present study is unusual and valuable in capturing over half a century of cognitive stability and change and examining its causes. The results here provide estimates for the genetic and environmental contributions to cognitive stability and change across most of the human lifespan. Even with almost 2,000 individuals, the study's power was insufficient to achieve conventional levels of significance for the estimates. Our emphasis here has not been on the traditional significance thresholds for *P* values per se, but in trying to partition variance in cognitive ability into environmental and genetic causes. The phenotypes available here are rare, and so these point estimates are useful to guide future research. The present findings render attractive a search for genetic mechanisms of cognitive change across the life course. They also suggest the importance of environmental contributions to lifetime cognitive change.

METHODS SUMMARY

Subjects. Recruitment, phenotyping and genotyping of the samples were described previously[11,12,17]. The mental test at age 11 was a Moray House Test[11,12]. In old age, general intelligence was derived using principal components analysis of a number of mental tests and saving scores on the first unrotated principal component (Supplementary Note). In old age, the assessments of general intelligence were made at ages as follows: ABC1936, 64.6 years (standard deviation 0.9); LBC1936, 69.5 (standard deviation 0.8); LBC1921, 79.1 (standard deviation 0.6). The LBC1921 and LBC1936 samples, but not the ABC1936, had repeat testing of the Moray House Test (already taken at age 11 years) at 79.1 and 69.5 years, respectively. After applying the genome-wide complex trait analysis method[13,15], the distribution of inferred relationships in the samples was as shown in Supplementary Fig. 3. We removed one of each pair of individuals whose estimated genetic relatedness was >0.2. We retained 1,940 individuals with childhood or old-age phenotype data (1,729 individuals had both): ABC1936, 425; LBC1921, 512; and LBC1936, 1,003. Of the 1,515 LBC1921 and LBC1936 individuals, there were 1,391 with genetic information and Moray House Test scores at age 11 and in old age.

Genotyping quality control. Quality control procedures were performed per SNP and per sample. Individuals were excluded from further analysis if genetic and reported gender did not agree. Samples with a call rate ≤ 0.95, and those showing evidence of non-European descent by multidimensional scaling, were removed[17]. SNPs were included in the analyses if they met the following conditions: call rate ≥ 0.98, minor allele frequency ≥ 0.01, and Hardy–Weinberg equilibrium test with $P \geq 0.001$. After these quality control stages, 1,948 samples remained (ABC1936, $N = 426$; LBC1921, $N = 517$; LBC1936, $N = 1,005$), and 536,295 autosomal SNPs were included in the analysis.

1. Plassman, B. L., Williams, J. W., Burke, J. R., Holsinger, T. & Benjamin, S. Systematic review: factors associated with risk for and possible prevention of decline in later life. *Ann. Intern. Med.* **153,** 182–193 (2010).
2. Brayne, C. The elephant in the room—healthy brains in later life, epidemiology and public health. *Nature Rev. Neurosci.* **8,** 233–239 (2007).
3. Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R. & Starr, J. M. The stability of individual differences in mental ability from childhood to old age: follow-up of the 1932 Scottish Mental Survey. *Intelligence* **28,** 49–55 (2000).
4. Gow, A. J. et al. Stability and change in intelligence from age 11 to ages 70, 79, and 87: the Lothian Birth Cohorts of 1921 and 1936. *Psychol. Aging* **26,** 232–240 (2011).
5. Deary, I. J., Johnson, W. & Houlihan, L. M. Genetic foundations of human intelligence. *Hum. Genet.* **126,** 215–232 (2009).
6. Deary, I. J., Penke, L. & Johnson, W. The neuroscience of human intelligence differences. *Nature Rev. Neurosci.* **11,** 201–211 (2010).
7. Lee, T., Henry, J. D., Trollor, J. N. & Sachdev, P. S. Genetic influences on cognitive functions in the elderly: a selective review of twin studies. *Brain Res. Rev.* **64,** 1–13 (2010).
8. Reynolds, C. A. et al. Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood. *Dev. Psychol.* **41,** 3–16 (2005).
9. Finkel, D., Reynolds, C. A., McArdle, J. J., Hamagami, F. & Pedersen, N. L. Genetic variance in processing speed drives variation in aging of spatial and memory abilities. *Dev. Psychol.* **45,** 820–834 (2009).
10. McGue, M. & Christensen, K. Social activity and healthy aging: a study of aging Danish twins. *Twin Res. Hum. Genet.* **10,** 255–265 (2007).
11. Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J. & Fox, H. C. The impact of childhood intelligence in later life: following up the Scottish Mental Surveys of 1932 and 1947. *J. Pers. Soc. Psychol.* **86,** 130–147 (2004).
12. Deary, I. J. et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr.* **7,** 28 (2007).
13. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nature Genet.* **42,** 565–569 (2010).
14. Visscher, P. M., Yang, J. & Goddard, M. E. A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). *Twin Res. Hum. Genet.* **13,** 517–524 (2010).
15. Yang, J., Lee, H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).
16. Yang, J. et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genet.* **43,** 519–525 (2011).
17. Davies, G. et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol. Psychiatry* **16,** 996–1005 (2011).
18. Deary, I. J. Intelligence. *Annu. Rev. Psychol.* **63,** 453–482 (2012).
19. Deary, I. J. et al. Age-associated cognitive decline. *Br. Med. Bull.* **92,** 135–152 (2009).
20. Tucker-Drob, E. M. Neurocognitive functions and everyday functions change together in old age. *Neuropsychology* **25,** 368–377 (2011).
21. Powell, J. E., Visscher, P. M. & Goddard, M. E. Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Rev. Genet.* **11,** 800–805 (2010).

22. Harris, S. E. & Deary, I. J. The genetics of cognitive ability and cognitive ageing in healthy older people. *Trends Cogn. Sci.* **15,** 388–394 (2011).
23. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 294–305 (2011).
24. Deary, I. J. *et al.* Cognitive change and the *APOE* ε4 allele. *Nature* **418,** 932 (2002).
25. Gilmour, A. R., Thompson, R. & Cullis, B. R. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51,** 1440–1450 (1995).
26. Calvin, C. M. *et al.* Intelligence in youth and all-cause mortality: systematic review with meta-analysis. *Int. J. Epidemiol.* **40,** 626–644 (2011).
27. Hollingworth, P. *et al.* Common variants at *ABCA7, MS4A6A/MS4A4E, EPHA1, CD33,* and *CD2AP* are associated with Alzheimer's disease. *Nature Genet.* **43,** 429–435 (2011).

Author Contributions I.J.D. and P.M.V. designed the study. J.Y. and P.M.V. performed statistical analyses, with I.J.D., M.E.G., A.T. and S.J.R. contributing to discussions regarding analyses. G.D., S.E.H., D.L., A.T., M.L. and L.M.L. performed quality control analyses and prepared data. S.E.H., M.L., L.M.L., A.J.G., J.C., P.R., H.C.F., S.J.R., P.H., L.J.W., G.M., D.J.P., J.M.S. and I.J.D. contributed genotype and phenotype data. I.J.D., P.M.V. and J.Y. contributed to writing the paper and Supplementary Information. All authors contributed to revising the paper and Supplementary Information.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to I.J.D. (i.deary@ed.ac.uk) or P.M.V. (peter.visscher@uq.edu.au).

LETTER

Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells

Mason A. Israel[1,2], Shauna H. Yuan[1,3], Cedric Bardy[4], Sol M. Reyna[1,2], Yangling Mu[4], Cheryl Herrera[1], Michael P. Hefferan[5], Sebastiaan Van Gorp[6], Kristopher L. Nazor[7], Francesca S. Boscolo[8], Christian T. Carson[9], Louise C. Laurent[8], Martin Marsala[5,10], Fred H. Gage[4], Anne M. Remes[11], Edward H. Koo[3] & Lawrence S. B. Goldstein[1,3]

Our understanding of Alzheimer's disease pathogenesis is currently limited by difficulties in obtaining live neurons from patients and the inability to model the sporadic form of the disease. It may be possible to overcome these challenges by reprogramming primary cells from patients into induced pluripotent stem cells (iPSCs). Here we reprogrammed primary fibroblasts from two patients with familial Alzheimer's disease, both caused by a duplication of the amyloid-β precursor protein gene[1] (*APP*; termed APP[Dp]), two with sporadic Alzheimer's disease (termed sAD1, sAD2) and two non-demented control individuals into iPSC lines. Neurons from differentiated cultures were purified with fluorescence-activated cell sorting and characterized. Purified cultures contained more than 90% neurons, clustered with fetal brain messenger RNA samples by microarray criteria, and could form functional synaptic contacts. Virtually all cells exhibited normal electrophysiological activity. Relative to controls, iPSC-derived, purified neurons from the two APP[Dp] patients and patient sAD2 exhibited significantly higher levels of the pathological markers amyloid-β(1–40), phospho-tau(Thr 231) and active glycogen synthase kinase-3β (aGSK-3β). Neurons from APP[Dp] and sAD2 patients also accumulated large RAB5-positive early endosomes compared to controls. Treatment of purified neurons with β-secretase inhibitors, but not γ-secretase inhibitors, caused significant reductions in phospho-Tau(Thr 231) and aGSK-3β levels. These results suggest a direct relationship between APP proteolytic processing, but not amyloid-β, in GSK-3β activation and tau phosphorylation in human neurons. Additionally, we observed that neurons with the genome of one sAD patient exhibited the phenotypes seen in familial Alzheimer's disease samples. More generally, we demonstrate that iPSC technology can be used to observe phenotypes relevant to Alzheimer's disease, even though it can take decades for overt disease to manifest in patients.

Alzheimer's disease is a common neurodegenerative disorder, defined post mortem by the increased presence of amyloid plaques and neurofibrillary tangles in the brain[2]. Amyloid plaques are extracellular deposits consisting primarily of amyloid-β peptides, and neurofibrillary tangles are intraneuronal aggregations of hyperphosphorylated tau, a microtubule-associated protein involved in microtubule stabilization[3]. The causative relationship between amyloid plaque/amyloid-β and tau pathologies is unclear in humans. Although the vast majority of Alzheimer's disease is apparently sporadic with significant non-Mendelian genetic contributions[4], analyses of cellular and animal models of rare, dominantly inherited familial forms of Alzheimer's disease have driven most ideas about disease mechanisms. These rare cases have mutations or a duplication of *APP*, which encodes the amyloid-β precursor protein, or mutations in the presenilin genes, which encode proteolytic enzymes that cleave APP into amyloid-β

and other fragments. Mouse models that overexpress familial Alzheimer's disease mutations develop extensive plaque deposition and amyloid-associated pathology, but neurofibrillary tangles and significant neuronal loss are conspicuously absent[5,6]. Fetal human cortical cultures have also been used to study the APP–tau relationship. For example, cortical cultures treated with 20 μM amyloid-β have elevated phosphorylated tau (p-tau)[7]. However, it is still unclear whether physiologically relevant levels of amyloid-β directly cause elevated p-tau and which kinases are directly involved in this aberrant phosphorylation. Additionally, experimental approaches using fetal human neurons are hindered by limited availability of samples and unknown genetic backgrounds. The recent developments in iPSCs and induced neurons have allowed investigation of phenotypes of neurological diseases *in vitro*[8,9,10]. However, not all diseases have been successfully modelled using iPSCs[11], and it is unclear whether iPSCs can be used to study sporadic forms of disease.

Here we report the derivation and neuronal differentiation of iPSCs from patients with familial and sporadic Alzheimer's disease, as well as from non-demented, age-matched controls. Using purified human neurons we probe three key questions concerning Alzheimer's disease: (1) can iPSC technology be used to observe phenotypes of patients with Alzheimer's disease, even though it can take decades for overt disease to manifest; (2) is there a causative relationship between APP processing and tau phosphorylation; and (3) can neurons with the genome of a sAD patient exhibit phenotypes seen in familial Alzheimer's disease samples? Supplementary Fig. 1 summarizes the experimental approach and findings.

We characterized *APP* metabolism in fibroblasts before reprogramming to iPSCs (Supplementary Fig. 2). *APP* expression and amyloid-β secretion were quantified in early-passage primary fibroblasts from two non-demented control (NDC) individuals, two sAD patients and two APP[Dp] patients (Table 1). The presence of the genomic duplication was confirmed in fibroblasts. Relative to NDC and sAD cells, APP[Dp] fibroblasts expressed higher levels of *APP* mRNA and secreted 1.5- to twofold higher amounts of amyloid-β(1–40) peptides into culture media compared to NDC cells. We did not detect significant increases in amyloid-β(1–42/1–40) or amyloid-β(1–38/1–40) in patient samples versus controls.

We generated iPSC lines by transducing fibroblasts with retroviruses encoding *OCT4*, *SOX2*, *KLF4*, *c-MYC* and, in one-third of cultures, *EGFP*. Each of the six individuals was represented by three clonal iPSC lines. All 18 iPSC lines maintained embryonic stem (ES)-cell-like morphology, expressed the pluripotency-associated proteins NANOG and TRA1-81, maintained euploid karyotypes, expressed endogenous locus-derived SOX2, repressed retroviral transgenes, and could differentiate into cells of ectodermal, mesodermal and

[1]Howard Hughes Medical Institute and Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093, USA. [2]Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, California 92093, USA. [3]Department of Neurosciences, University of California, San Diego, La Jolla, California, USA. [4]The Salk Institute for Biological Studies, La Jolla, California 92037, USA. [5]Department of Anesthesiology, University of California, San Diego, La Jolla, California 92093, USA. [6]Department of Anesthesiology, Maastricht University Medical Center, Maastricht 6202 AZ, Netherlands. [7]Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA. [8]Department of Reproductive Medicine, University of California, San Diego, La Jolla, California 92093, USA. [9]BD Biosciences, La Jolla, California 92037, USA. [10]Institute of Neurobiology, Slovak Academy of Sciences, Kosice SK-04001, Slovakia. [11]Department of Clinical Medicine, Neurology and Clinical Research Center, University of Oulu, Oulu FIN-90015, Finland.

Table 1 | Summary of patient information

Code	Diagnosis	Gender	Family history	Age at onset	Age at biopsy	MMSE at biopsy	*APOE*
NDC1	Non-demented control	M	Possible	N/A	86	30	2-3
NDC2	Non-demented control	M	N	N/A	86	30	3-3
sAD1	Sporadic AD	F	N	78	83	4	3-3
sAD2	Sporadic AD	M	N	78	83	18	3-3
APP^{Dp}1	Familial AD, APP duplication	M	Y	46	51	21	3-3
APP^{Dp}2	Familial AD, APP duplication	F	Y	53	60	17	3-3

MMSE, mini mental state examination (perfect score = 30). AD, Alzheimer's disease.

endodermal lineages *in vitro* (Fig. 1a–d, Supplementary Figs 3a–e and 4). All lines tested (one per individual) formed teratomas when injected into nude rats (Supplementary Fig. 5). Supplementary Table 1 provides details of each iPSC line.

Variability in differentiation efficiency exists between pluripotent cell lines. To analyse variability in our iPSC lines, we used a fluorescence-activated cell sorting (FACS)-based method of neuronal differentiation and purification (summarized in Supplementary Fig. 6), based on work described previously[12]. Briefly, the 18 iPSC lines were first differentiated into cultures containing neural rosettes (Supplementary Fig. 3f). From these cultures, neural progenitor cells (NPCs) were purified and the efficiency of NPC formation was assessed by CD184$^+$CD15$^+$CD44$^-$CD271$^-$ immunoreactivity. These FACS-purified NPCs maintained expression

Figure 1 | Generation of iPSC lines and purified neurons from APP^{Dp}, sAD and NDC fibroblasts. a, b, iPSC lines express NANOG and TRA1-81. **c, d**, iPSC-derived, FACS-purified NPCs express SOX2 and nestin. **e–h**, iPSC-derived, FACS-purified neurons express MAP2 and βIII-tubulin. Scale bars in **a–h**, 50 μm. **i**, Representative action potentials in response to somatic current injections. Data from iPSC line APP^{Dp}2.2. **j**, Spontaneous synaptic activity was detected (voltage clamp recording at the reversal potential of sodium (0 mV)) and reversibly blocked by GABA$_A$ receptor antagonist SR95531 (10 μM). Each panel represents ~4 min continuous recordings separated in 25 sweeps (grey traces) and superimposed for clarity. Black traces represent a single sweep. Data from iPSC line NDC2.1. **k, l**, No significant difference was seen between NDCs and any patient's cultures in the ability of iPSCs to generate NPCs at day 11 ($P = 0.08$, $n = 9$), or the ability of NPCs to form neurons at 3 weeks ($P = 0.82$, $n = 9$). Error bars indicate s.e.m.

of NPC-associated markers, such as SOX2 and nestin, over multiple passages (Fig. 1c, d). NPCs were differentiated for 3 weeks into heterogeneous cultures containing neurons (Supplementary Fig. 3g, h). *APP* copy number was faithfully maintained in differentiated cultures (Supplementary Fig. 3i). From these cultures, neurons were purified to near homogeneity, and the efficiency of neuron generation was assessed by CD24$^+$CD184$^-$CD44$^-$ immunoreactivity. No significant differences between any of the individuals in the efficiency of NPC or neuronal differentiation were detected (Fig. 1k, l).

Although we observed variability in differentiation among lines from each individual, the extent of inter-individual variation was less than observed intra-individual variability. These results suggest that any observed biochemical aberrations in neurons, if present in multiple lines derived from the same patient, are probably caused by features of that patient's genotype. Purified neurons were plated at a density of 2×10^5 cells per well of a 96-well plate and cultured for an additional 5 days. More than 90% of cells in these cultures were neurons, as judged by the presence of βIII-tubulin$^+$, MAP2$^+$ projections (Fig. 1e–h). Genome-wide mRNA expression profiles of five representative purified neuronal cultures were compared to the parental iPSC lines and samples from fetal brain, heart, liver and lung (Supplementary Fig. 7 and Supplementary Table 2). Unsupervised hierarchical clustering analysis revealed that purified neurons most closely resembled fetal brain samples, in part due to a global upregulation of neuronal genes. Interestingly, the largest difference between fetal brain samples and purified neurons was downregulation in purified neurons of the hippo signalling cascade (~6.1 fold), which regulates proliferation of cells such as NPCs and glia[13,14].

We determined multiple electrophysiological properties of purified neurons to assess passive membrane properties and synaptic connectivity (Fig. 1i, j, Supplementary Table 3 and Supplementary Fig. 8). Notably, virtually all neurons tested generated voltage-dependent action potentials and currents (Fig. 1i), which were blocked by tetrodotoxin (Supplementary Fig. 8). Transient bath application of ionotropic receptor agonists (25 μM muscimol or 10 μM AMPA) evoked transient currents, showing that purified neurons expressed functional GABA and AMPA receptors, respectively (Supplementary Table 3). To determine whether neurons were also able to form functional synaptic contacts, we analysed continuous whole-cell voltage clamp recordings. We detected spontaneous inhibitory and/or excitatory synaptic currents in a subset of cells (~40%). Analysis of the kinetics of those events combined with reversible blockade using GABA$_A$ or AMPA receptor antagonists demonstrated that the neurons not only fire action potentials but also made functional synaptic contacts (Supplementary Table 3). The electrophysiological results were supported by analysis of expression of protein markers of glutamatergic and GABAergic neuronal subtypes (VGluT1 and GABA, respectively), which were detected by immunofluorescence, with approximately 15% of cells staining brightly for VGluT1 and 8% for GABA, and most remaining neurons staining dimly for one of the markers (Supplementary Fig. 9a). RNAs indicative of glutamatergic, GABAergic and cholinergic subtypes (that is, *VGLUT1*, *GAD67* and *CHAT*, respectively) were detected by quantitative polymerase chain reaction (qPCR). Importantly, no significant differences in neuronal subtypes were detected between patients and controls (Supplementary Fig. 9b–f).

Elevated or altered secretion of amyloid-β peptides by fibroblasts is a feature common to all familial Alzheimer's disease mutations
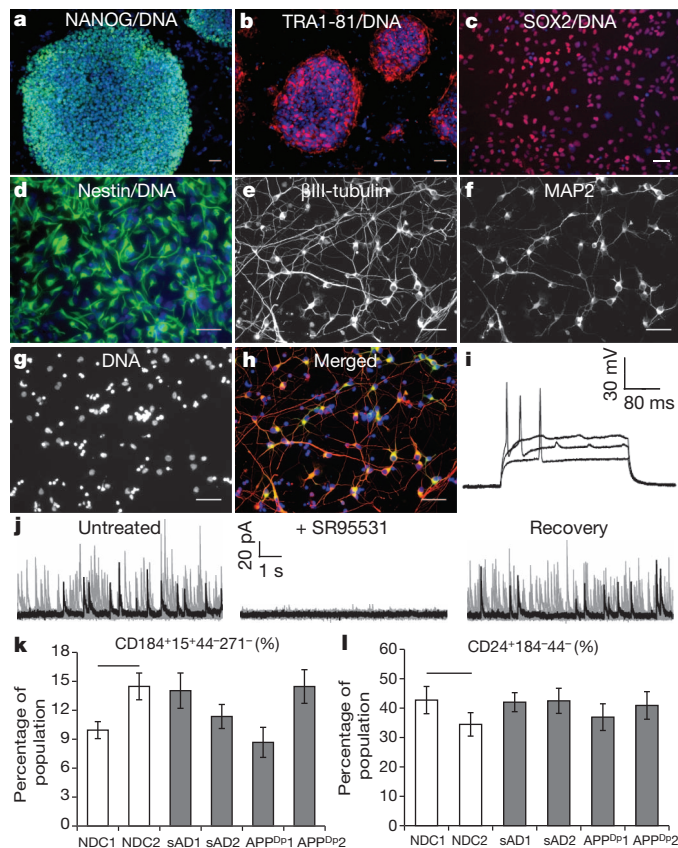
identified so far[15,16]. It is not known if iPSC-derived neurons from familial Alzheimer's disease patients maintain the elevated amyloid-β production seen in the parental fibroblasts. In sAD fibroblasts and other peripheral cells, APP expression and amyloid-β secretion are not consistently altered[17]. To determine if iPSC-derived neurons from APP[Dp] and sAD patients exhibit elevated amyloid-β secretion, amyloid-β levels in neuron-conditioned media were measured and normalized to total protein levels of cell lysates. Purified neurons from patients APP[Dp]1 and APP[Dp]2, each represented by three independently derived iPSC lines, secreted significantly higher levels of amyloid-β(1–40) compared to mean NDC levels (Fig. 2a). Neurons from patient sAD2 also had significantly higher amyloid-β(1–40) levels compared to NDC neurons, even though no difference was observed between the fibroblasts of sAD2 and NDC individuals. We found that amyloid-β(1–42) and amyloid-β(1–38) levels in these purified neuronal cultures were often below the detection range of our assay, owing to the relatively small number of neurons purified. By cell type, neurons exhibited a larger difference in amyloid-β levels between APP[Dp] and NDC than fibroblasts, further suggesting that fibroblasts are not fully predictive of neuronal phenotypes (Fig. 2b).

Genetic evidence implicates altered or elevated APP processing and amyloid-β levels as the driving agents behind familial Alzheimer's disease[2] and, because of identical neuropathology, sporadic Alzheimer's disease. However, tau, although not genetically linked to Alzheimer's disease, forms neurofibrillary tangles, which correlate better with disease severity than plaque numbers[18]. The mechanism by which altered APP processing might cause elevated p-tau and neurofibrillary tangle pathology is unclear. Tau phosphorylation at Thr 231, one of several tau phosphoepitopes, regulates microtubule stability[19] and correlates with both neurofibrillary tangle number and degree of cognitive decline[20,21]. To determine if tau phosphorylation at Thr 231 is elevated in APP[Dp] and sAD neurons, we measured the amount of p-tau(Thr 231) relative to total tau levels in lysates from purified neurons from three iPSC lines from each of the NDC, sAD and APP[Dp] patients. Neurons from both APP[Dp] patients had significantly higher p-tau/total tau ratios than neurons from NDC lines (Fig. 2c). p-Tau/total tau in the two sAD patients mirrored the amyloid-β findings: no difference was observed between sAD1 and NDC neurons whereas sAD2 neurons had significantly increased p-tau/total tau.

Tau can be phosphorylated by multiple kinases. The kinase GSK-3β can phosphorylate tau at Thr 231 *in vitro* and co-localizes with neurofibrillary tangles and pre-tangle phosphorylated tau in sAD postmortem neurons[22]. GSK-3β is thought to be constitutively active but
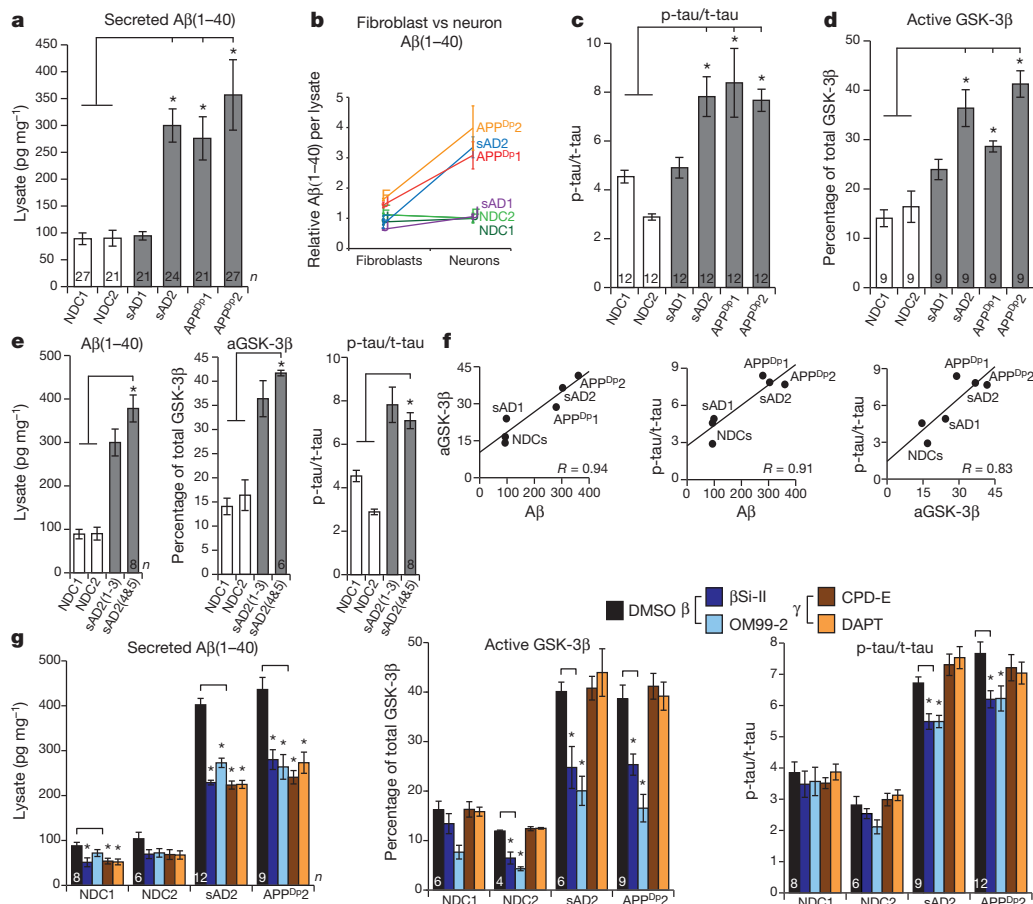

Figure 2 | Increased amyloid-β, p-tau and aGSK-3β in sAD2 and APP[Dp] neuronal cultures. **a,** Purified neurons from sAD2, APP[Dp]1 and APP[Dp]2 secrete increased amyloid-β(1–40) (Aβ(1–40)) compared to NDC samples (*P* = 0.0012, 0.0014 and <0.0001, respectively). **b,** Amyloid-β differences between patients and controls are larger in neurons versus fibroblasts. Data sets are relative to NDC mean. **c, d,** Neurons from sAD2, APP[Dp]1 and APP[Dp]2 have increased aGSK-3β (percentage non-phospho-Ser 9) and p-tau/total tau (p-tau/t-tau) compared to NDC samples (aGSK-3β, *P* < 0.0001, 0.0005 and 0.0001; p-tau/total tau, *P* < 0.0001, 0.0001 and 0.0002). In **a–d**, *n* values on graphs indicate the number of biological replicates per patient, contributed equally by three iPSC lines. **e,** sAD2 findings verified in two additional iPSC lines (sAD2.4 and sAD2.5). sAD2(1-3) indicates findings from initial sAD2

iPSC lines. For amyloid-β, aGSK-3β and p-tau/total tau, sAD2 remained significantly higher than controls (*P* < 0.0001). No significant difference was found between original and secondary sAD2 lines (*P* = 0.14, 0.44, 0.63). **f,** Strong positive correlations between amyloid-β(1–40), aGSK-3β and p-tau/total tau in purified neurons. Pearson *R* = 0.94, 0.91 and 0.83, respectively. **g,** Twenty-four hour treatment with β- and γ-secretase inhibitors reduced secreted amyloid-β(1–40) compared to control treatment. β-Secretase inhibitors partially rescued aGSK-3β and p-tau/total tau in sAD2 and APP[Dp]2 neurons (*P* < 0.01 for aGSK-3β, *P* < 0.03 for p-tau). γ-Secretase inhibition did not significantly affect aGSK-3β and p-tau/total tau. In **g**, number of treatment sets is indicated on the graph (*n*), NDCs are represented by two iPSC lines each and sAD2 and APP[Dp]2 are represented by three. Error bars indicate s.e.m.

is inactivated when phosphorylated at Ser 9 (ref. 23). To determine if iPSC-derived neurons with elevated p-tau have increased GSK-3β activity, the proportion of aGSK-3β in purified neurons was calculated by measuring the amount of GSK-3β lacking phosphorylation at Ser 9 relative to total GSK-3β levels. We observed that neurons from patients APP^{Dp}1, APP^{Dp}2 and sAD2 had significantly higher aGSK-3β than NDC neurons (Fig. 2d). The amyloid-β, GSK-3β and tau findings of sAD2 were verified by analysing an additional two iPSC lines (sAD2.4 and sAD2.5; characterization in Supplementary Fig. 10), and we observed that levels remained consistently elevated (Fig. 2e). Results are detailed per patient in Supplementary Table 4a, per cell line in Supplementary Fig. 11, and per cell culture in Supplementary Table 5.

Although amyloid-β, p-tau and GSK-3β clearly have roles in Alzheimer's disease pathogenesis, their relationship is unclear. We observed that iPSC-derived neurons exhibited strong or very strong correlations between amyloid-β(1–40), p-tau/total tau and aGSK-3β levels (Fig. 2f and Supplementary Table 4b). We reasoned that if APP proteolytic products, such as amyloid-β or carboxy-terminal fragments (CTFs), have a causative role in p-tau and aGSK-3β elevation, then inhibiting γ- or β-secretase activity could reduce p-tau and aGSK-3β. We treated purified neurons from NDC1, NDC2, sAD2 and APP^{Dp}2 (2–3 iPSC lines each) with γ-secretase inhibitors (CPD-E and DAPT) or β-secretase inhibitors (βSi-II and OM99-2) for 24 h and measured amyloid-β, GSK-3β and p-tau/total tau compared to vehicle-treated samples. All inhibitors reduced amyloid-β(1–40) by similar levels (32–45% in patient samples) (Fig. 2g). Intriguingly, for both sAD2 and APP^{Dp}2 neurons, we observed that β-secretase inhibitors significantly reduced aGSK-3β and p-tau/total tau (Fig. 2g, and shown per iPSC line in Supplementary Fig. 12). Neither γ-secretase inhibitor significantly differed from control-treated samples for aGSK-3β levels and p-tau/total tau.

We extended phenotypic characterization of sAD2 and APP^{Dp} by analysing endosomal and synaptic markers in FACS-purified neurons co-cultured with astrocytes for 12 days. Accumulation of large RAB5$^+$ early endosomes in neurons has been observed in autopsies from sporadic Alzheimer's disease and some forms of familial Alzheimer's disease[24,25]. As β-secretase is localized to endosomes and has an acidic pH optimum, it has been proposed that early endosomes potentially mediate the effects of APP processing on downstream pathologies such as increased p-tau, neurofibrillary tangles, synaptic loss and apoptosis[26]; however, these hypotheses have been difficult to test directly without live, patient-specific neurons. To determine if early endosome phenotypes are present in iPSC-derived neurons from Alzheimer's disease patients, purified neurons from NDC1, NDC2, sAD2 and APP^{Dp}2 (two iPSC lines each) co-cultured with astrocytes were harvested and large and very large Rab5$^+$ early endosomes (1–2.1 μm^3 and 2.1–7 μm^3) in neuronal soma were counted. Whereas

control neurons generally had few Rab5$^+$ structures >1 μm^3, neurons from both sAD2 and APP^{Dp}2 frequently had Rab5$^+$ early endosomes highly similar in volume, morphology and localization to what has been observed in autopsy samples (Fig. 3a–c). When compared, the neurons from both sAD2 and APP^{Dp}2 had significantly increased numbers of both large and very large early endosomes relative to controls (Fig. 3d). We sought to determine if neuronal cultures from sAD2 and APP^{Dp}2 also contained reduced levels of the presynaptic marker synapsin I. In Alzheimer's disease autopsies, synaptic loss is one of the strongest pathological correlates with dementia severity, and in regions of the brain affected by Alzheimer's disease, the presynaptic marker synapsin I is decreased in patients versus controls[27,28]. To analyse synapsin I levels in iPSC-derived neurons co-cultured with astrocytes, we quantified synapsin I$^+$ puncta on MAP2$^+$ dendrites (Fig. 3e). We found no significant difference between NDCs and either sAD2 or APP^{Dp}2 in the number of puncta per μm dendrite (Fig. 3f). Extended culture periods may be required to study Alzheimer's disease-associated loss of synaptic proteins.

The results of this study provide strong evidence that iPSC technology can be used in concert with post-mortem samples and animal models to study early pathogenesis and drug response in sporadic and familial Alzheimer's disease. In purified, electrophysiologically active neurons from one sporadic Alzheimer's disease and two APP^{Dp} patients, each represented by at least three clonally derived iPSC lines, we observed significantly increased levels of three major biochemical markers of Alzheimer's disease: amyloid-β(1–40), aGSK-3β and p-tau/total tau. Increased sAD2 amyloid-β levels were not observed in the parental fibroblasts, suggesting a cell-type-specific phenotype. Among the individuals in this study, not only did strong correlations exist between amyloid-β(1–40), p-tau/total tau and aGSK-3β, but both p-tau/total tau and aGSK-3β levels were also partially rescued in neurons from sAD2 and APP^{Dp} following treatment with β-secretase inhibitors, suggesting that the APP processing pathway has a causative role in tau Thr 231 phosphorylation in human neurons. Because γ-secretase inhibition did not cause a significant effect, products of APP processing other than amyloid-β may have a role in induction of GSK-3β activity and p-tau. One potential culprit is the β-CTF, the levels of which correlate with axonopathies in mouse models harbouring APP duplications[29] and mediate early endosome accumulation in human Down's syndrome fibroblasts[30]. The observation that neurons from patients sAD2 and APP^{Dp}2 have early endosome phenotypes raises the question of how aberrant early endosomes relate to other phenotypes of Alzheimer's disease, such as axonopathies, synaptic loss and cell death, in human neurons. Neurons and synapses rely heavily on endocytic pathways, and thus iPSC technology can now be used to study the role of this dynamic process in live patient-specific neurons. One point of caution is that it is possible that the cultures of purified neurons that we
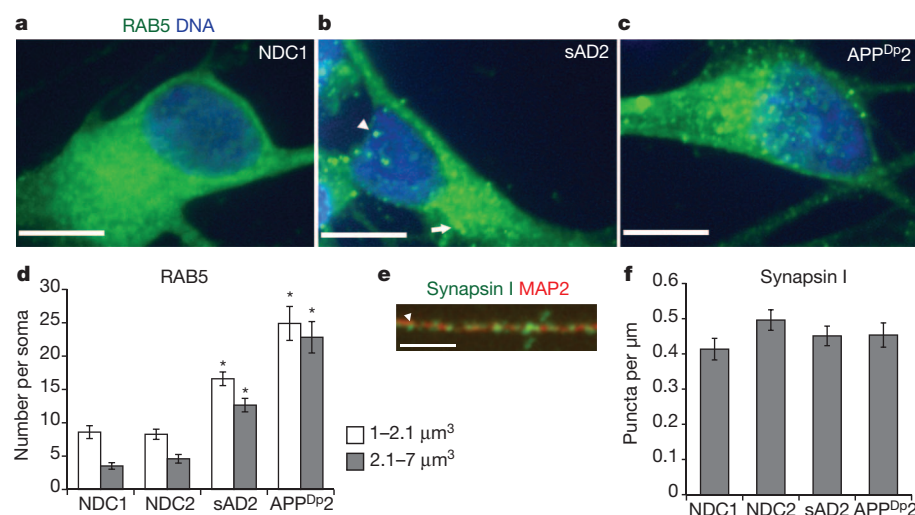

Figure 3 | Analysis of early endosome and synapsin levels in purified neurons co-cultured with astrocytes. a–c, Extended focus images of Rab5-stained neuronal soma from NDC1, sAD2 and APP^{Dp}2. Arrowhead in b marks a 1–2.1 μm^3 early endosome, and the arrow marks a 2.1–7 μm^3 early endosome. Scale bars, 10 μm. d, Neurons from sAD2 and APP^{Dp}2 have significantly increased numbers of large and very large early endosomes compared to NDC neurons ($P < 0.0001$, $n = 40$ neurons from two iPSC lines per individual). e, Representative image of synapsin I (green) on a MAP2$^+$ dendrite (red). Arrowhead marks a synapsin I$^+$ punctum. Scale bar, 3 μm. f, No significant difference between patients and controls in the number of synapsin$^+$ structures per μm dendrite ($P = 1.00$, $n = 40$ dendrites from two iPSC lines per individual). Neurons were scored blinded to genotype. Error bars indicate s.e.m.

generated and studied may not have been fully mature, as they lacked repetitive action potentials and had limited spontaneous activity. Although some types of mature neurons also have these properties, it is conceivable that the phenotypes we observed might be modified by duration of *in vitro* culture. In this context, while there is debate about when Alzheimer's disease phenotypes initiate, evidence exists that Alzheimer's disease-like pathology can occur in Down's syndrome fetuses as early as 28 weeks of gestation[24].

Our finding that the genome of patient sAD2, but not patient sAD1, generates significant Alzheimer's disease phenotypes in purified neurons has important implications. First, this finding suggests that an unknown frequency of sporadic Alzheimer's disease patients will have genomes that generate strong neuronal phenotypes. The frequency of such genomes in the sporadic Alzheimer's disease population cannot be determined from the small sample size we report and will require a larger sample size to ask how frequent such genomes are in the clinical population diagnosed with sporadic Alzheimer's disease. Second, the genome of sAD2 clearly harbours one or more variants that generate Alzheimer's disease phenotypes, which can thus be elucidated by future molecular genetic studies. Third, we speculate that sporadic Alzheimer's disease might be sub-divided depending on whether neurons themselves are altered, as in the case of sAD2, as opposed to other cell types such as astrocytes, which could be altered in other cases, for example, sAD1. Thus, future iPSC studies examining larger numbers of patients and controls have the potential to provide great insight into the mechanisms behind the observed heterogeneity in sporadic Alzheimer's disease pathogenesis, the role of different cell types, patient-specific drug responses, and prospective diagnostics.

METHODS SUMMARY

iPSC generation and differentiation. Primary fibroblast cultures were established from dermal punch biopsies taken from individuals following informed consent and Institutional Review Board approval. To generate iPSCs, fibroblasts were transduced with MMLV vectors containing the complementary DNAs for *OCT4*, *SOX2*, *KLF4*, *c-MYC* and ± *EGFP*. IPSC-derived NPCs were differentiated for 3 weeks, neurons were purified by FACS, and amyloid-β, p-tau/total tau and aGSK-3β were measured on purified control and mutant neurons from multiple lines cultured in parallel for an additional 5 days by multi-spot electrochemiluminescence assays (Meso Scale Diagnostics). Early endosomes were analysed by confocal microscopy on purified neurons co-cultured with human astrocytes (Lonza) for 12 days. To ensure reproducible and consistent data, we found that it is important to differentiate and evaluate neurons from full sets of mutant and control iPSC lines together.

Statistics. $P < 0.05$ was considered statistically significant. Individuals were statistically compared to the total NDC pool by Tukey's test. Drug responses were compared to controls by Dunnett's method. N values signify the total number of separate cultures analysed, with each iPSC line contributing equally to the total.

1. Rovelet-Lecrux, A. *et al.* APP locus duplication in a Finnish family with dementia and intracerebral haemorrhage. *J. Neurol. Neurosurg. Psychiatry* **78**, 1158–1159 (2007).
2. Tanzi, R. E. & Bertram, L. Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. *Cell* **120**, 545–555 (2005).
3. Ballatore, C., Lee, V. M. Y. & Trojanowski, J. Q. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nature Rev. Neurosci.* **8**, 663–672 (2007).
4. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63**, 168–174 (2006).
5. Games, D. *et al.* Alzheimer-type neuropathology in transgenic mice overexpressing V717F β-amyloid precursor protein. *Nature* **373**, 523–527 (1995).
6. Roberson, E. D. *et al.* Reducing endogenous tau ameliorates amyloid β-induced deficits in an Alzheimer's disease mouse model. *Science* **316**, 750–754 (2007).
7. Busciglio, J., Lorenzo, A., Yeh, J. & Yankner, B. A. β-Amyloid fibrils induce tau phosphorylation and loss of microtubule binding. *Neuron* **14**, 879–888 (1995).
8. Ebert, A. D. *et al.* Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* **457**, 277–280 (2009).
9. Nguyen, H. N. *et al.* LRRK2 mutant iPSC-derived DA neurons demonstrate increased susceptibility to oxidative stress. *Cell Stem Cell* **8**, 267–280 (2011).
10. Qiang, L. *et al.* Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons. *Cell* **3**, 359–371 (2011).
11. Urbach, A., Bar-Nur, O., Daley, G. Q. & Benvenisty, N. Differential modeling of fragile X syndrome by human embryonic stem cells and induced pluripotent stem cells. *Cell Stem Cell* **6**, 407–411 (2010).
12. Yuan, S. H. *et al.* Cell-surface marker signatures for the isolation of neural stem cells, glia and neurons derived from human pluripotent stem cells. *PLoS ONE* **6**, e17540 (2011).
13. Cao, X., Pfaff, S. L. & Gage, F. H. YAP regulates neural progenitor cell number via the TEA domain transcription factor. *Genes Dev.* **22**, 3320–3334 (2008).
14. Reddy, B. V. & Irvine, K. D. Regulation of *Drosophila* glial cell proliferation by Merlin-Hippo signaling. *Development* **138**, 5201–5212 (2011).
15. Citron, M. *et al.* Excessive production of amyloid β-protein by peripheral cells of symptomatic and presymptomatic patients carrying the Swedish familial Alzheimer disease mutation. *Proc. Natl Acad. Sci. USA* **91**, 11993–11997 (1994).
16. Scheuner, D. *et al.* Secreted amyloid β-protein similar to that in the senile plaques of Alzheimer's disease is increased *in vivo* by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nature Med.* **2**, 864–870 (1996).
17. Gasparini, L. *et al.* Peripheral markers in testing pathophysiological hypotheses and diagnosing Alzheimer's disease. *FASEB J.* **12**, 17–34 (1998).
18. Arriagada, P. V., Growdon, J. H., Hedley-Whyte, E. T. & Hyman, B. T. Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. *Neurology* **42**, 631 (1992).
19. Cho, J.-H. & Johnson, G. V. W. Primed phosphorylation of tau at Thr 231 by glycogen synthase kinase 3β (GSK3β) plays a critical role in regulating tau's ability to bind and stabilize microtubules. *J. Neurochem.* **88**, 349–358 (2004).
20. Buerger, K. *et al.* CSF tau protein phosphorylated at threonine 231 correlates with cognitive decline in MCI subjects. *Neurology* **59**, 627–629 (2002).
21. Buerger, K. *et al.* CSF phosphorylated tau protein correlates with neocortical neurofibrillary pathology in Alzheimer's disease. *Brain* **129**, 3035–3041 (2006).
22. Cho, J. & Johnson, G. Glycogen synthase kinase 3β phosphorylates tau at both primed and unprimed sites. *J. Biol. Chem.* **278**, 187–193 (2003).
23. Dajani, R. *et al.* Crystal structure of glycogen synthase kinase 3β: structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell* **105**, 721–732 (2001).
24. Cataldo, A. M. *et al.* Endocytic pathway abnormalities precede amyloid β deposition in sporadic Alzheimer's disease and Down syndrome: differential effects of APOE genotype and presenilin mutations. *Am. J. Pathol.* **157**, 277–286 (2000).
25. Cataldo, A. *et al.* Endocytic disturbances distinguish among subtypes of Alzheimer's disease and related disorders. *Ann. Neurol.* **50**, 661–665 (2001).
26. Nixon, R. A. Endosome function and dysfunction in Alzheimer's disease and other neurodegenerative diseases. *Neurobiol. Aging* **26**, 373–382 (2005).
27. Hamos, J. E., DeGennaro, L. J. & Drachman, D. A. Synaptic loss in Alzheimer's disease and other dementias. *Neurology* **39**, 355–361 (1989).
28. Qin, S., Hu, X. Y., Xu, H. & Zhou, J. N. Regional alteration of synapsin I in the hippocampal formation of Alzheimer's disease patients. *Acta Neuropathol.* **107**, 209–215 (2004).
29. Salehi, A. *et al.* Increased *App* expression in a mouse model of Down's syndrome disrupts NGF transport and causes cholinergic neuron degeneration. *Neuron* **51**, 29–42 (2006).
30. Jiang, Y. *et al.* Alzheimer's-related endosome dysfunction in Down syndrome is Aβ-independent but requires APP and is reversed by BACE-1 inhibition. *Proc. Natl Acad. Sci. USA* **107**, 1630–1635 (2010).

METHODS

Patients and fibroblast derivation. NDC and sAD individuals were enrolled in the longitudinal study at the UCSD Alzheimer's Disease Research Center. APP^{Dp} individuals were patients at the Department of Clinical Medicine, Neurology, Oulu University Hospital, Oulu, Finland. For all individuals, dermal punch biopsies were taken following informed consent and Institutional Review Board approval. Primary fibroblast cultures were established from biopsies using established methods[31]. Fibroblasts were cultured in DMEM containing 15% FBS, L-glutamine and penicillin/streptomycin.

iPSC generation and expansion. iPSCs were generated as described[32], with the following modifications. The cDNAs for *OCT4*, *SOX2*, *KLF4*, *c-MYC* and *EGFP* were cloned into pCX4 vectors[33] and vectors were packaged into VSVG-pseudotyped retroviruses. For each patient, three independent viral transductions were performed. Three wells, each containing ~1×10^5 fibroblasts, were transduced with retroviruses. On days 2–8, 2 mM valproic acid was added to cultures. Potential iPSC colonies were picked at ~3 weeks and transferred to 96-well plates containing irradiated mouse embryonic fibroblasts (MEFs). For passaging, cells were dissociated with TrypLE (Invitrogen). The efficiency of potential iPSC colony formation was roughly 100 colonies per 1×10^5 fibroblasts at 3 weeks. The efficiency of successful establishment of a stable iPSC line from an initial colony was roughly 10%.

Karyotype analysis and pluripotency assays. Karyotype analysis was performed by Cell Line Genetics. IPSCs were assayed for teratoma formation by injections into spinal cords of athymic rats, as previously described[34], with the following modifications: cells were dissociated with Accutase (Innovative Cell Technologies) and passed through a 100-μm mesh filter before injections and each animal received 10 injections of roughly 10,000 cells. For *in vitro* pluripotency assays, iPSC cultures were dissociated with dispase, and embryoid bodies were generated in low-attachment plates in media containing 15% fetal bovine serum (FBS). After 7 days, cultures were plated on Matrigel (BD Biosciences)-coated glass coverslips and cultured for 7 days.

Genotyping and qPCR. To determine *APP* copy number, genomic DNA was isolated from fibroblasts or differentiated NPCs. qPCR was performed using FastStart Universal SYBR Green Master Mix (Roche) and analysed on an Applied Biosystems 7300 Real Time PCR System using the $\Delta\Delta C_T$ method. *APP* levels were normalized to mean β-globin/albumin. To compare RNA levels between samples, RNA was purified (PARIS kit, Ambion), DNase treated (Ambion) and reverse transcribed (Superscript II, Invitrogen). For transgene expression, primers detected a sequence common to all transgenes, and expression was normalized to the housekeeping gene *NONO*. PCR to detect endogenous *SOX2* expression was performed using Qiagen HotStarTaq and primers previously described[35].

Immunocytochemistry and microscopy. Cells were fixed in 4% paraformaldehyde, permeabilized with buffer containing TritonX-100 and stained with primary and secondary antibodies (see below). Samples, except for early endosome studies, were imaged on a Nikon TE2000-U inverted microscope and acquired using Metamorph software (Molecular Devices). ImageJ software (National Institutes of Health) was used to pseudo-colour images, adjust contrast, and add scale bars. Endosomes and synapses were imaged on a PerkinElmer UltraVIEW VoX microscope with a ×60 objective and a z-step of 0.5 μm. Quantifications were done blinded to genotype.

Antibodies. The antibodies used for FACS purification of cells were TRA1-81-APC, CD184-APC, CD15-FITC, CD24-PECy7, CD44-PE and CD271-PE (all from BD Biosciences) and were used at a concentration of one test per 1×10^6 cells. The following antibodies were from Millipore: AFP (mouse 1:1,000), APP^{FL} (22C11, ms 1:1,000), SMA (ms 1:50), SOX2 (rb 1:2,000), synapsin I (rb 1:500); from Sigma: GABA (rb 1:200), MAP2a/b (ms 1:500), tau total (rb 1:500), tau Thr 231 (rb 1:150). Other vendors: APP^{CTF} (Zymed rb 1:500), GAPDH (Ambion ms 1:250), VGluT1 (Synaptic Systems rb 1:200), NANOG (Santa-Cruz rb 1:200), nestin (Santa-Cruz rb 1:50), RAB5A (Santa-Cruz rb 1:50), synuclein (BD ms 1:500), tau PHF1 (Pierce ms 1:500), βIII-tubulin (Covance ms 1:2,000), βIII-tubulin (Covance rabbit 1:1,000) anti-rabbit Alexa Fluor 488 (Invitrogen 1:200) and anti-mouse Alexa Fluor 568 (Invitrogen 1:200).

Neuronal differentiation and FACS. To ensure reproducible and consistent data, we found that it is important to differentiate and evaluate neurons from full sets of mutant and control iPSC lines together. Differentiation to NPCs and neurons was performed as previously described[12]. 3×10^5 FACS-purified TRA1-81$^+$ cells were seeded onto 3×10 cm plates that were seeded the previous day with 5×10^5 PA6 cells[36]. At day 11, cells were dissociated with Accutase and ~5×10^5 CD184$^+$CD15$^+$CD44$^-$CD271$^-$ NPCs were FACS-purified and plated onto poly-ornithine/laminin-coated plates and cultured with bFGF. At passage 7, NPCs were differentiated with BDNF, GDNF and cAMP. After 3 weeks of differentiation, cells were dissociated with Accutase and CD24$^+$CD184$^-$CD44$^-$ cells were purified. FACS was performed with a FACSAria II (BD Biosciences) and

analysed with FloJo (Tree Star). Differentiation methods are also summarized in Supplementary Fig. 6.

Gene expression profiling. Total RNA was extracted from collected sample pellets (Ambion mirVana; Applied Biosystems) according to the manufacturer's protocol. RNA quantity (Qubit RNA BR Assay Kits; Invitrogen) and quality (RNA6000 Nano Kit; Agilent) was determined to be optimal for each sample before further processing. 200 ng RNA per sample was amplified using the Illumina Total Prep RNA Amplification Kit according to manufacturer's protocol and quantified as above. 750 ng biotinylated RNA per sample was hybridized to Illumina HT-12v4 Expression BeadChips, scanned with an Illumina iScan Bead Array Scanner, and quality controlled in GenomeStudio and the lumi bioconductor package. All RNA processing and microarray hybridizations were performed in house according to manufacturer's protocols. In GenomeStudio, probes were filtered for those detected with a *P* value of 0.01 in at least one sample and exported for normalization in R. Raw probe expression values were transformed and normalized using the robust spline normalization (RSN) as implemented in the lumi R/Bioconductor package. Probes were further filtered for those having a minimum value of 150 in at least two samples and a minimum difference between any two samples (maximum value minus minimum value) of at least 150.

Electrophysiology methods. Electrophysiology was performed on purified neurons, 5 days after FACS. For whole-cell patch-clamp recordings, individual coverslips were transferred into a heated recording chamber and continuously perfused (1 ml min^{-1}) with artificial cerebrospinal fluid (ACSF) bubbled with a mixture of CO$_2$ (5%) and O$_2$ (95%) and maintained at 25 °C. The ACSF contained 124 mM NaCl, 3 mM KCl, 1.3 mM MgSO$_4$, 26 mM NaHCO$_3$, 1.25 mM NaHPO$_4$, 20 mM glucose and 2 mM CaCl$_2$ (all chemicals from Sigma). For targeted whole-cell recordings, we used a ×40 water-immersion objective, differential interference contrast filters (all Olympus), a digital camera (Rolera XR -Qimaging), a halogen (Olympus), a digidata 1440A/ Multiclamp 700B and Clampex 10.3 (Molecular devices). The resistance of the patch electrodes was between 3–5 MOhm. Patch electrodes were filled with two different internal solutions both containing 4 mM NaCl, 10 mM Na-HEPES, 10 mM D-glucose, nucleotides (0.3 mM GTP, 2 mM Mg-ATP, 0.2 mM cAMP) 0.15% biocytin and 0.06% rhodamine. For current-clamp experiments, the internal solution also contained 130 mM K-gluconate, 6 mM KCl and 0.2 mM K-EGTA; in all other experiments, it contained instead: 126 mM Cs-gluconate, 6 mM CsCl and 0.2 mM Cs-EGTA. The pH and osmolarity of the internal solution were close to physiological conditions (pH 7.3, 290–300 mOsm). Data were all corrected for liquid junction potentials (10 mV). Electrode capacitances were compensated on-line in cell-attached mode (~7 pF). Recordings were low-pass filtered at 2 kHz, digitized, and sampled at intervals of 50 μs (20 kHz). To control the quality and the stability of the recordings throughout the experiments, access resistance, capacitance and membrane resistance were continuously monitored on-line and recorded. The access resistance of the cells in our sample was 21 ± 1 MOhm. Electrophysiological statistical analysis was assisted with Clampfit 10.3, Igor Pro 6, Prism 5 and Microsoft Excel. Mean ± standard error of the mean were reported.

Amyloid-β, p-tau/total tau and aGSK-3β measurements. FACS-purified neurons were plated at 2×10^5 per well of a 96-well plate. Cells were cultured for an additional 5 days with a full media change on day 3. Amyloid-β was measured with MSD Human (6E10) Abeta3-Plex Kits (Meso Scale Discovery). p-tau/total tau was measured with a MSD Phospho(Thr231)/Total Tau Kit. aGSK-3β was measured with MSD Phospho/Total GSK-3b Duplex Kit. Fibroblast and neuronal amyloid-β levels were normalized to total protein levels determined by BCA assay (Thermo Scientific). aGSK-3β (the per cent of unphosphorylated GSK-3β at Ser 9) was calculated by manufacturer's recommendations: $(1 - (2 \times \text{phospho signal})/(\text{phospho signal} + \text{total signal})) \times 100$.

Inhibitor treatments. CPD-E (Compound-E) and DAPT were used at a final concentration of 200 nM. βSi-II (β-secretase inhibitor II) and OM99-2 were used at 10 μM and 750 nM, respectively. 1 μl of inhibitor or vehicle was added to the existing culture media of parallel cultures on day 4 and cultures were harvested on day 5. All inhibitors were from EMD Chemicals and were dissolved in DMSO.

Endosomal analysis. 1.5×10^5 per FACS-purified neurons were plated per well of a 96-well plate that was seeded the previous day with 5,000 human astrocytes (Lonza). After 12 days of culture, cultures were stained for RAB5 and βIII-tubulin and imaged on a PerkinElmer UltraVIEW VoX microscope with a ×60 objective and a z-step of 0.5 μm. Quantification was performed blinded to genotype with Volocity software (PerkinElmer) on βIII-tubulin$^+$ cells only.

Statistics. Data were analysed using JMP software (SAS Institute). $P < 0.05$ was considered statistically significant. Individuals were statistically compared to the total NDC pool by ANOVA followed by Tukey's test. *N* values signify the total number of separate cultures analysed, with each iPSC line contributing equally to the total. Drug responses were compared to controls by Dunnett's method. Correlations were determined by calculating Pearson coefficients (*R*).

31. Takashima, A. in *Current Protocols in Cell Biology* Ch. 2 12 (John Wiley & Sons, 2001).
32. Park, I.-H., Lerou, P. H., Zhao, R., Huo, H. & Daley, G. Q. Generation of human-induced pluripotent stem cells. *Nature Protocols* **3,** 1180–1186 (2008).
33. Akagi, T., Sasai, K. & Hanafusa, H. Refractory nature of normal human diploid fibroblasts with respect to oncogene-mediated transformation. *Proc. Natl Acad. Sci. USA* **100,** 13567–13572 (2003).
34. Emre, N. *et al.* The ROCK inhibitor Y-27632 improves recovery of human embryonic stem cells after fluorescence-activated cell sorting with multiple cell surface markers. *PLoS ONE* **5,** e12148 (2010).
35. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
36. Kawasaki, H. *et al.* Induction of midbrain dopaminergic neurons from ES cells by stromal cell derived inducing activity. *Neuron* **28,** 31–40 (2000).

LETTER

Enhancer decommissioning by LSD1 during embryonic stem cell differentiation

Warren A. Whyte[1,2]*, Steve Bilodeau[1]*, David A. Orlando[1], Heather A. Hoke[1,2], Garrett M. Frampton[1,2], Charles T. Foster[3,4], Shaun M. Cowley[4] & Richard A. Young[1,2]

Transcription factors and chromatin modifiers are important in the programming and reprogramming of cellular states during development[1,2]. Transcription factors bind to enhancer elements and recruit coactivators and chromatin-modifying enzymes to facilitate transcription initiation[3,4]. During differentiation a subset of these enhancers must be silenced, but the mechanisms underlying enhancer silencing are poorly understood. Here we show that the histone demethylase lysine-specific demethylase 1 (LSD1; ref. 5), which demethylates histone H3 on Lys 4 or Lys 9 (H3K4/K9), is essential in decommissioning enhancers during the differentiation of mouse embryonic stem cells (ESCs). LSD1 occupies enhancers of active genes that are critical for control of the state of ESCs. However, LSD1 is not essential for the maintenance of ESC identity. Instead, ESCs lacking LSD1 activity fail to differentiate fully, and ESC-specific enhancers fail to undergo the histone demethylation events associated with differentiation. At active enhancers, LSD1 is a component of the NuRD (nucleosome remodelling and histone deacetylase) complex, which contains additional subunits that are necessary for ESC differentiation. We propose that the LSD1–NuRD complex decommissions enhancers of the pluripotency program during differentiation, which is essential for the complete shutdown of the ESC gene expression program and the transition to new cell states.

The histone H3K4/K9 demethylase LSD1 (also known as KDM1A) is one of the chromatin regulators that have been implicated in the control of early embryogenesis[6–8]. Loss of LSD1 leads to embryonic lethality, and ESCs lacking LSD1 function fail to differentiate into embryoid bodies[6–8]. These results suggest that LSD1 contributes to changes in chromatin that are critical to the differentiation of ESCs, but the role of LSD1 in this process is not yet understood. To investigate the function of LSD1 in ESCs, we first identified the sites it occupies in the genome by using chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP-Seq; Fig. 1 and Supplementary Fig. 1). The results revealed that LSD1 occupies the enhancers and core promoters of a substantial population of actively transcribed and bivalent genes (Fig. 1a, b and Supplementary Table 1). Inspection of individual gene tracks showed that LSD1 occupies well-characterized enhancer regions together with the ESC master transcription factors Oct4, Sox2 and Nanog and the Mediator coactivator (Fig. 1b and Supplementary Fig. 1). Loci bound by Oct4, Sox2 and Nanog are generally associated with Mediator and p300 coactivators and have enhancer activity[9,10]. A global view of enhancer regions occupied by Oct4, Sox2, Nanog and Mediator confirmed that 97% of the 3,838 high-confidence enhancers were also occupied by LSD1 ($P < 10^{-9}$) (Fig. 1c and Supplementary Table 2). This is consistent with evidence that LSD1 can interact with Oct4 (refs 11, 12). LSD1 signals were also observed at core promoter regions with RNA polymerase II (Pol II) and TATA-binding protein (TBP; Fig. 1d). The

density of LSD1 signals at enhancers was higher than at core promoters ($P < 10^{-16}$; Supplementary Fig. 1), indicating that LSD1 is associated predominantly with the enhancers of actively transcribed genes in ESCs.

It was striking to find that LSD1 is associated with active genes in ESCs because previous studies have shown that LSD1 is not essential for the maintenance of ESC state but is required for normal differentiation[6–8]. We used an ESC differentiation assay to further investigate the involvement of LSD1 in cell state transitions (Fig. 2a, b). Prolonged depletion of Oct4 in ZHBTc4 ESCs with doxycycline causes loss of pluripotency and differentiation into trophectoderm[13]. As expected, loss of Oct4 expression led to a rapid loss of ESC morphology and a marked decrease in the levels of SSEA-1 and alkaline phosphatase, two markers of ESCs (Fig. 2c and Supplementary Fig. 2). When these ESCs were treated with the LSD1 inhibitor tranylcypromine (TCP) during Oct4 depletion, they failed to undergo the morphological changes associated with differentiation of ESCs (Fig. 2c). Instead, the TCP-treated cells formed small colonies resembling those of untreated ESCs and maintained expression of SSEA-1 and alkaline phosphatase (Fig. 2c and Supplementary Fig. 3). Very similar results were partly obtained in LSD1 knockout ESCs (Supplementary Figs 4 and 5) and in cells treated with another LSD1 inhibitor, pargyline, or a short hairpin RNA against LSD1 (Supplementary Figs 2 and 3). LSD1 inhibition also caused an increase in cell death during differentiation, as has been observed with cells lacking LSD1 in other assays[7,8]. These results suggest that LSD1 may be required for ESCs to silence the ESC gene expression program completely.

Further analysis of ESCs that were forced to differentiate in the absence of LSD1 activity confirmed that these cells failed to make a complete transition from the ESC gene expression program; although key genes of the trophectoderm gene expression program were activated, including *Cdx2* and *Esx1* (ref. 14), there was incomplete repression of many ESC genes, including *Sox2* and *Fbox15* (Fig. 2d). A global analysis confirmed that a set of genes neighbouring LSD1-occupied enhancers in ESCs are repressed during differentiation and that the repression of this set of genes is partly relieved in the presence of TCP (Fig. 2e and Supplementary Table 3). Similar results were obtained with LSD1 knockout cells (Supplementary Figs 4 and 5) and with cells treated with either pargyline or a short hairpin RNA against LSD1 (Supplementary Fig. 3). These results indicate that the trophectoderm differentiation program can be induced in cells lacking LSD1 function, but the ESC program is not fully silenced in these cells.

To gain further insight into the role of LSD1 in ESC differentiation, we investigated whether LSD1 is associated with previously described complexes, including NuRD, cofactor of REST (CoREST), and the androgen receptor/oestrogen receptor complexes[8,15–17]. We first studied whether the LSD1 found at Oct4-occupied genes is a component of NuRD, because Oct4 and Nanog have been reported to interact with several components of NuRD[11,12,18]. ChIP-Seq experiments

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Department of Molecular Biology, Adolf-Butenandt Institut, Ludwig-Maximilians-Universität München, 80336 Munich, Germany. [4]Department of Biochemistry, University of Leicester, Leicester LE1 9HN, UK.
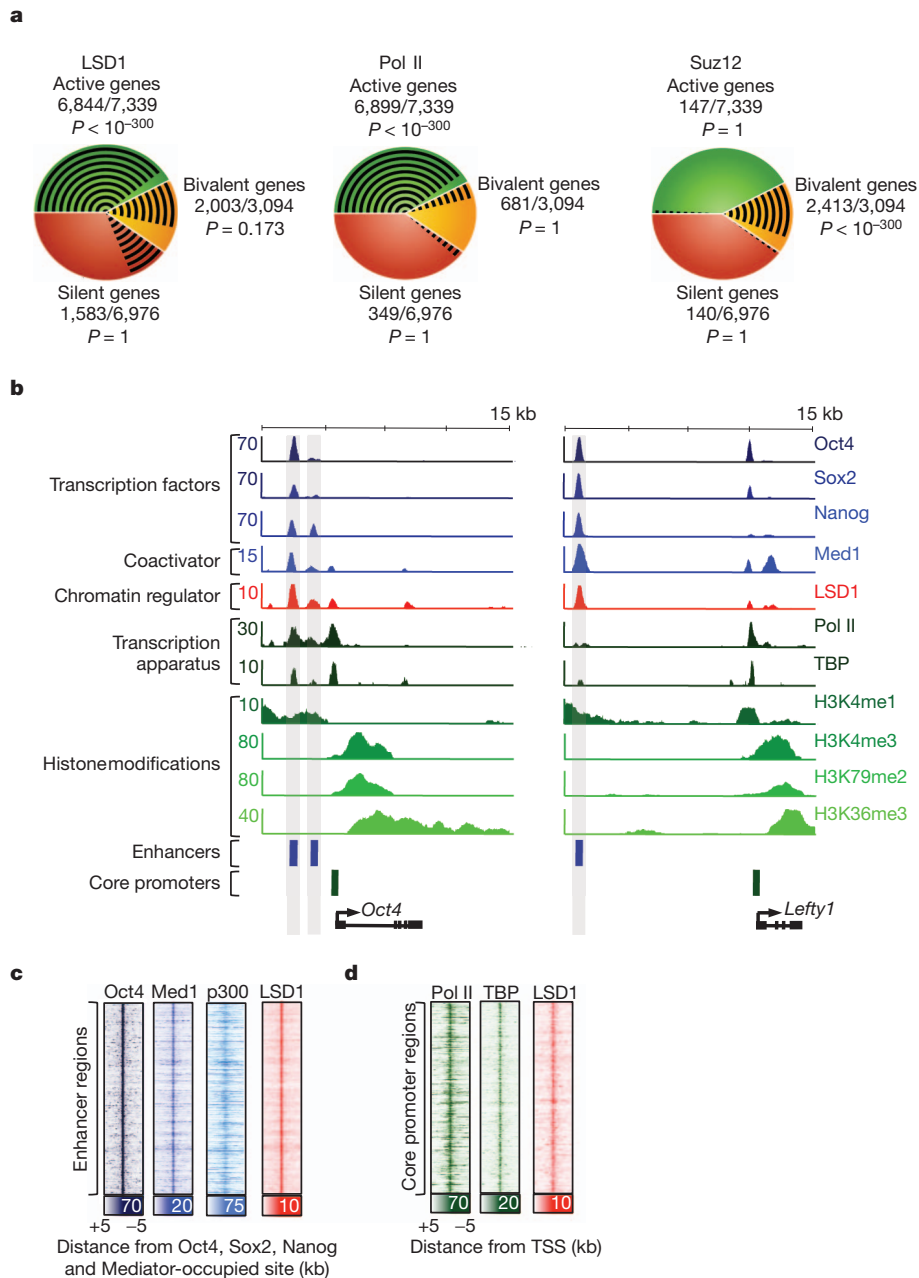*These authors contributed equally to this work.

Figure 1 | LSD1 is associated with enhancer and core promoter regions of active genes in ESCs. **a**, LSD1 occupies a substantial population of actively transcribed genes in murine ESCs. The pie charts depict active (green), bivalent (yellow) and silent (red) genes, and the proportion (black lines) occupied by LSD1, Pol II or the Polycomb protein Suz12 (Supplementary Table 1 and Supplementary Information). The numbers of genes bound and the total number of genes in each of the active, bivalent and silent classes are shown. LSD1 ChIP-Seq data are from combined biological replicates using an antibody specific for LSD1 as determined by knockdown experiments (Supplementary Fig. 1). The *P* value for each category was determined by a hypergeometric test. **b**, LSD1 occupies enhancers and core promoter regions of actively transcribed genes. Shown are ChIP-Seq binding profiles (reads per million) for ESC transcription factors (Oct4, Sox2, Nanog), coactivator (Med1), chromatin regulator (LSD1), the transcriptional apparatus (Pol II, TBP) and histone modifications (H3K4me1, H3K4me3, H3K79me2, H3K36me3) at the *Oct4*

(*Pou5f1*) and *Lefty1* loci in ESCs, with the *y*-axis floor set to 1. Gene models and previously described enhancer regions[27–29] are shown below the binding profiles. **c**, LSD1 occupies enhancer sites. A density map is shown of ChIP-Seq data at Oct4, Sox2, Nanog and Med1 co-occupied enhancer regions. Data are shown for an ESC transcription factor (Oct4), coactivators (Med1 and p300) and a chromatin regulator (LSD1) in ESCs. Enhancers were defined as Oct4, Sox2, Nanog and Mediator co-occupied regions. More than 96% of the 3,838 high-confidence enhancers were co-occupied by LSD1 ($P < 10^{-9}$). Colour scale indicates ChIP-seq signal in reads per million. **d**, LSD1 occupies core promoter sites. Shown is a density map of ChIP-Seq data at transcriptional start sites (TSSs) of genes neighbouring the 3,838 previously defined enhancers (**c**). Data are shown for components of the transcription apparatus (Pol II and TBP) and the chromatin regulator LSD1 in ESCs. Core promoters were defined as the closest TSS from each enhancer. Colour scale indicates ChIP-Seq signal in reads per million.

confirmed that NuRD subunits Mi-2β, HDAC1 and HDAC2 together occupy sites with LSD1 at enhancers ($P < 10^{-9}$; Fig. 3 and Supplementary Table 1). Immunoprecipitation of LSD1 confirmed its association with Mi-2β, HDAC1 and HDAC2 (Fig. 3b, c). We then investigated whether LSD1 is associated with CoREST; ChIP-Seq data revealed that

a minor fraction of LSD1 occupies sites together with CoREST and REST (2% and 6%, respectively) (Supplementary Fig. 6 and Supplementary Table 1). As expected, LSD1–REST sites were frequently found associated with neuronal genes (Supplementary Fig. 7 and Supplementary Table 4). Immunoprecipitation experiments confirmed that LSD1 is
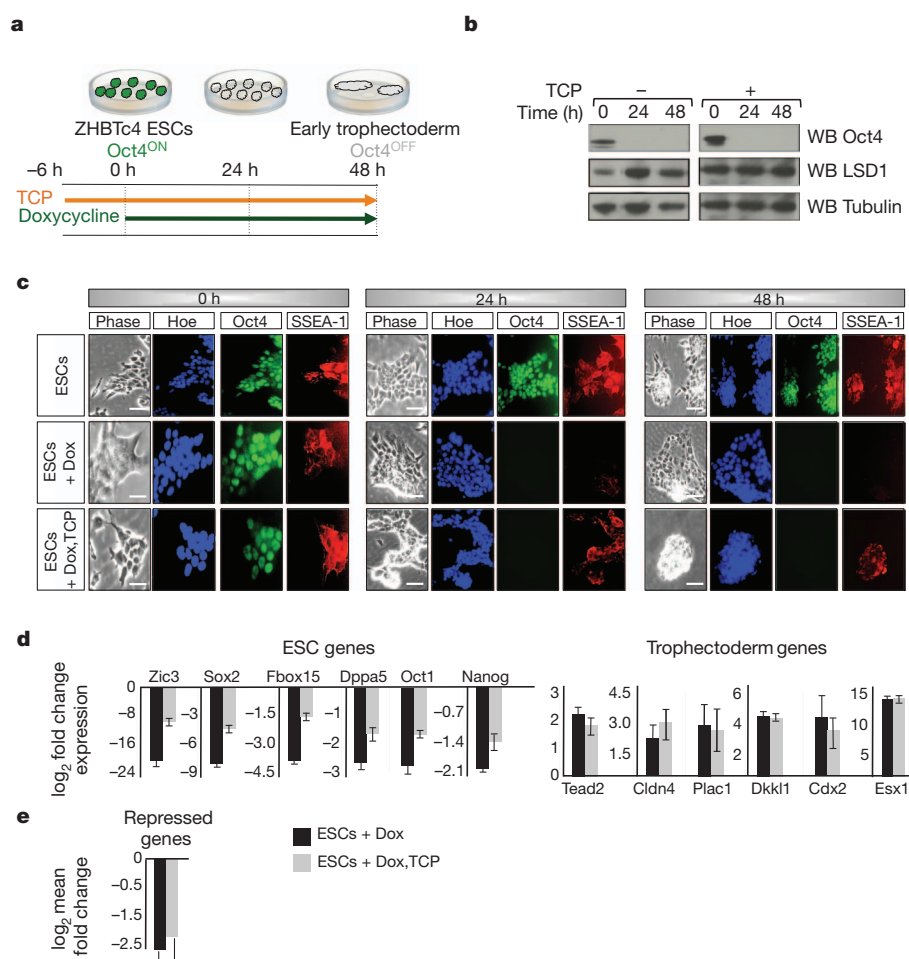
Figure 2 | LSD1 inhibition results in incomplete silencing of ESC genes during differentiation. a, Schematic representation of trophectoderm differentiation assay using the doxycycline-inducible Oct4 shutdown murine ESC line ZHBTc4. Treatment with doxycycline for 48 h leads to depletion of Oct4 and early trophectoderm specification. Cells were treated with dimethylsulphoxide (DMSO; control) or the LSD1 inhibitor TCP for 6 h before 2 μg ml^{-1} doxycycline was added for a further 24 or 48 h. **b**, Treatment of ZHBTc4 ESCs with doxycycline leads to loss of Oct4 proteins. Oct4 and LSD1 protein levels in nuclear extracts determined by western blotting (WB) before and after treatment of ZHBTc4 ESCs with 2 μg ml^{-1} doxycycline. Tubulin served as loading control. **c**, Doxycycline (Dox)-treated cells treated with TCP maintained SSEA-1 cell surface marker expression. Cells were stained for DNA (Hoechst; Hoe), Oct4 and SSEA-1. Scale bar, 100 μm. **d**, Expression of selected ESC and trophectodermal genes 48 h after Oct4 depletion in dimethylsulphoxide-treated and TCP-treated cells (black and grey bars, respectively). Treatment with TCP partly relieved repression of ESC genes but did not affect upregulation of trophectodermal genes. Error bars show s.d. from biological replicates. **e**, Genes neighbouring LSD1-occupied enhancers are less downregulated during ESC differentiation after TCP treatment. Shown is the mean fold change in expression of the 630 downregulated (at least 1.25-fold; $P < 0.01$) genes nearest LSD1-occupied enhancers (Fig. 1c) during differentiation of TCP-treated and untreated control cells. Alleviation of repression is significantly higher (asterisk, $P < 0.005$) for LSD1 enhancer-bound repressed genes than for all repressed genes.

associated with CoREST (Fig. 3b, c). Androgen receptor and oestrogen receptor are not expressed in ESCs, as indicated by the lack of histone H3K79me2 and H3K36me3 (modifications associated with transcriptional elongation) at the genes encoding these proteins (Supplementary Table 1). Further examination of the ChIP-Seq data revealed that enhancers were significantly more likely to be occupied by the LSD1 and NuRD proteins than by REST and CoREST ($P < 10^{-9}$) (Fig. 3d and Supplementary Fig. 8). Multiple components of NuRD are dispensable for ESC state but are required for normal differentiation[6,19–21]. ESCs with decreased levels of the core NuRD ATPase Mi-2β failed to differentiate properly and partly maintained expression of SSEA-1, alkaline phosphatase and ESC genes (Supplementary Fig. 9), which are the same phenotypes as those we observed with decreased levels of LSD1. These results indicate that LSD1 at enhancers is associated with a NuRD complex that is essential for normal cell state transitions.

Nucleosomes with histone H3K4me1 are commonly found at enhancers of active genes and are a substrate for LSD1 (refs 5, 22). If LSD1-dependent H3K4me1 demethylase activity is involved in enhancer silencing during ESC differentiation, LSD1 inhibition should cause the retention of H3K4me1 levels at active ESC enhancers when differentiation is induced. During trophectoderm differentiation with control ESCs, we found decreased levels of p300 and H3K27ac at a set of active ESC enhancers, suggesting that these enhancers were being silenced (Supplementary Fig. 10). The levels of H3K4me1 at enhancers were also decreased, as seen for example at *Lefty1* (Fig. 4a and Supplementary Table 5), whereas the levels of H3K4me1 increased at newly active trophectoderm genes such as *Gata2* (Fig. 4b). In contrast, H3K4me1 signals were higher at LSD1-occupied enhancers in differentiating ESCs treated with TCP than in control cells, including *Lefty1* and *Sox2* (Fig. 4a, c). Most enhancers (1,722 of 2,755) that were occupied by LSD1 and that experienced decreased levels of H3K4me1 during differentiation retained H3K4me1 in TCP-treated ESCs, in contrast to untreated control differentiating ESCs (Fig. 4d, e). These results are consistent with the model that LSD1 demethylates H3K4me1 at the enhancers of ESC-specific genes during differentiation and that this activity is essential to fully repress the genes associated with these enhancers.

Our results indicate that an LSD1–NuRD complex is required for silencing of ESC enhancers during differentiation, which is essential
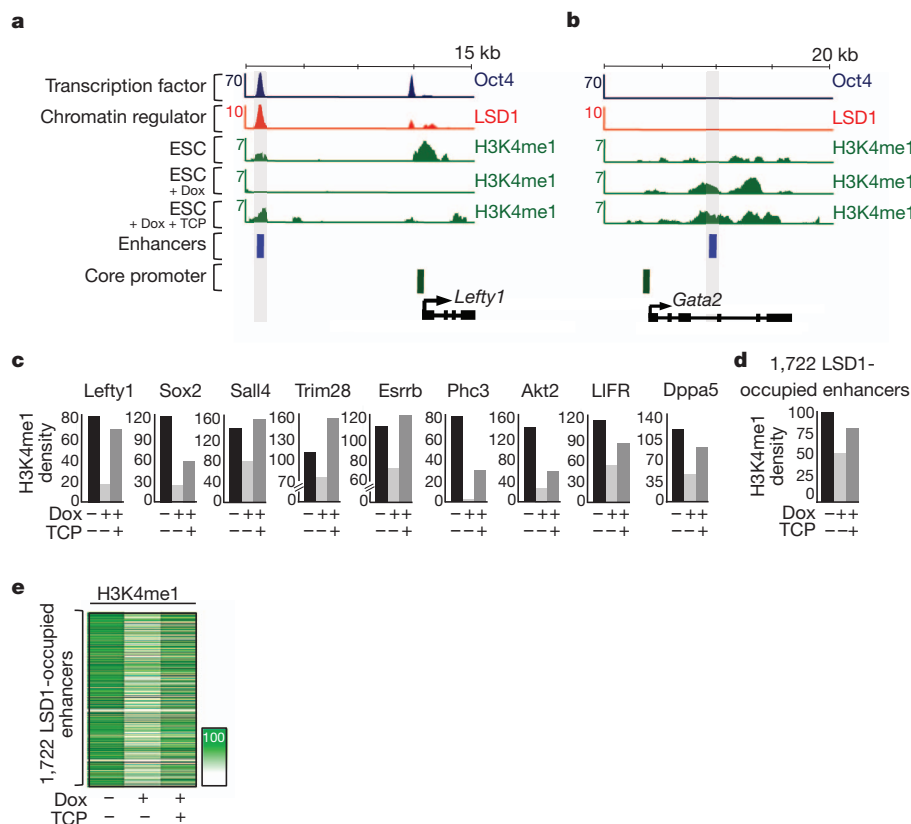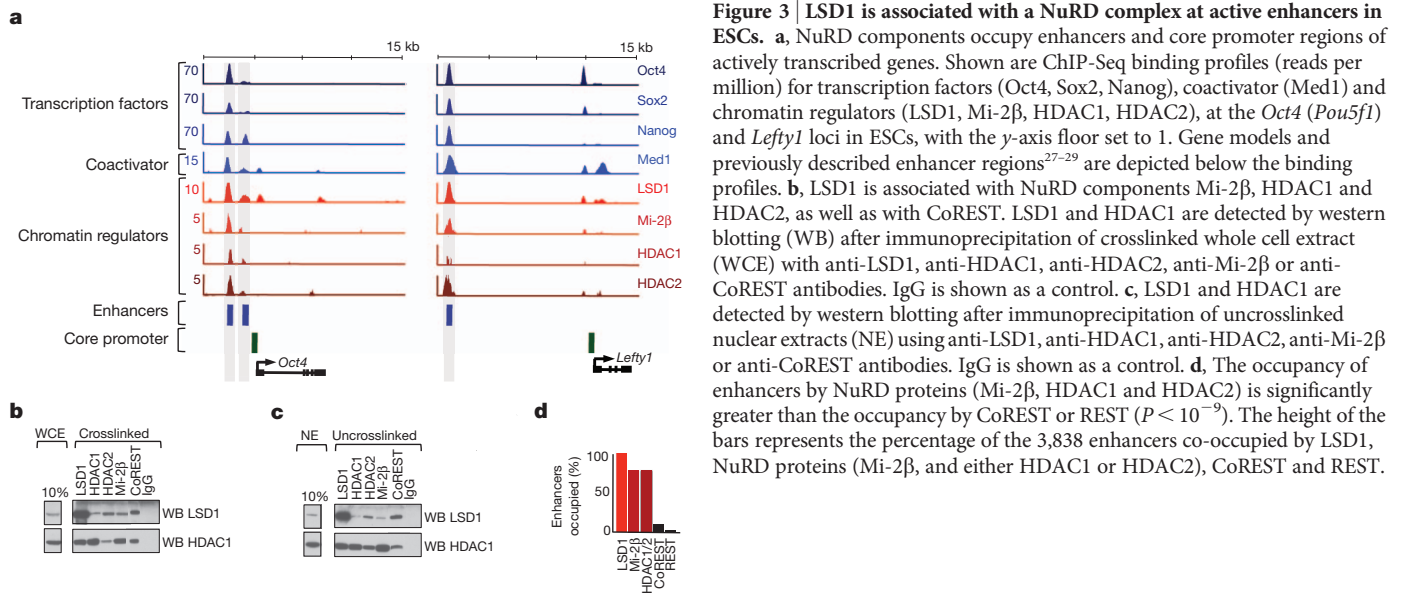
a

b   c   d

Figure 3 | LSD1 is associated with a NuRD complex at active enhancers in ESCs. a, NuRD components occupy enhancers and core promoter regions of actively transcribed genes. Shown are ChIP-Seq binding profiles (reads per million) for transcription factors (Oct4, Sox2, Nanog), coactivator (Med1) and chromatin regulators (LSD1, Mi-2β, HDAC1, HDAC2), at the *Oct4* (*Pou5f1*) and *Lefty1* loci in ESCs, with the *y*-axis floor set to 1. Gene models and previously described enhancer regions[27–29] are depicted below the binding profiles. **b**, LSD1 is associated with NuRD components Mi-2β, HDAC1 and HDAC2, as well as with CoREST. LSD1 and HDAC1 are detected by western blotting (WB) after immunoprecipitation of crosslinked whole cell extract (WCE) with anti-LSD1, anti-HDAC1, anti-HDAC2, anti-Mi-2β or anti-CoREST antibodies. IgG is shown as a control. **c**, LSD1 and HDAC1 are detected by western blotting after immunoprecipitation of uncrosslinked nuclear extracts (NE) using anti-LSD1, anti-HDAC1, anti-HDAC2, anti-Mi-2β or anti-CoREST antibodies. IgG is shown as a control. **d**, The occupancy of enhancers by NuRD proteins (Mi-2β, HDAC1 and HDAC2) is significantly greater than the occupancy by CoREST or REST ($P < 10^{-9}$). The height of the bars represents the percentage of the 3,838 enhancers co-occupied by LSD1, NuRD proteins (Mi-2β, and either HDAC1 or HDAC2), CoREST and REST.

Figure 4 | LSD1 is required for H3K4me1 removal at ESC enhancers.
a, H3K4me1 levels are decreased at LSD1-occupied enhancers during ESC differentiation, and this effect is partly blocked on treatment with TCP. Dox, doxycycline. **b**, Treatment with TCP does not affect the increase in H3K4me1 levels at trophectodermal genes during differentiation. Shown are ChIP-Seq binding profiles (reads per million) for Oct4 and LSD1 at the *Lefty1* and *Gata2* loci in ESCs. Below these profiles, histone H3K4me1 levels are shown for ZHBTc4 control ESCs, cells treated with doxycycline for 48 h to repress Oct4 and induce differentiation (ESCs + Dox), and ESCs treated with doxycycline and TCP (ESCs + Dox + TCP). For appropriate normalization, ChIP-Seq data for histone H3K4me1 is shown as rank normalized reads per million with the *y*-axis floor set to 1 (Supplementary Information). Gene models and previously described enhancer regions[29,30] are depicted below the binding profiles. **c**, Sum

of the normalized H3K4me1 density ±250 nucleotides surrounding LSD1-occupied enhancer regions before and during trophectoderm differentiation in the presence or absence of TCP. The associated genes were identified on the basis of their proximity to the LSD1-occupied enhancers. **d**, Sum of the normalized H3K4me1 density ±250 nucleotides surrounding 1,722 LSD1-occupied enhancers before and during differentiation in the presence or absence of TCP. Of the 2,755 LSD1-occupied enhancers with decreased levels of H3K4me1 on differentiation, 63% (1,722) had higher H3K4me1 levels after TCP treatment ($P < 10^{-16}$). **e**, Heat map displaying the sum of the normalized H3K4me1 density ±250 nucleotides surrounding the 1,722 LSD1-occupied enhancers that retained H3K4me1 in TCP-treated ESCs compared with untreated control differentiating ESCs. Colour scale indicates ChIP-Seq signal in normalized reads per million.

for complete shutdown of the ESC gene expression program and the transition to new cell states. These results, together with those of previous studies on NuRD function[18,21,23,24], suggest the following model for LSD1–NuRD in enhancer decommissioning. LSD1–NuRD complexes occupy Oct4-regulated active enhancers in ESCs but do not substantially demethylate histone H3K4 because the H3K4 demethylase activity of LSD1 is inhibited in the presence of acetylated histones[23,24]. Enhancers occupied by Oct4, Sox2 and Nanog are also occupied by the HAT p300 and nucleosomes with acetylated histones (Supplementary Fig. 10)[10]. Thus, as long as the enhancer-bound transcription factors recruit HATs to enhancers, the net effect of having both HATs and NuRD-associated HDACs present is to have sufficient levels of acetylated histones to suppress LSD1 demethylase activity. During ESC differentiation, the levels of Oct4 and p300 are decreased, thus decreasing the level of acetylated histones, which in turn permits the demethylation of H3K4 by LSD1. Consistent with this model, we find that the shutdown of Oct4 leads to decreased levels of p300 and histone H3K27ac at enhancers that are occupied by Oct4 and LSD1 (Supplementary Figs 10 and 11), and this is coincident with decreased levels of methylated H3K4 (Fig. 4 and Supplementary Figs 12 and 13). This model would explain why key components of LSD1–NuRD complexes are not essential for the maintenance of ESC state but are essential for normal differentiation, when the active enhancers must be silenced. Additional HATs expressed in ESCs may also contribute to the dynamic balance of nucleosome acetylation. Future biochemical analysis of HAT, HDAC and demethylase complexes at enhancers will be valuable for testing this model and for further understanding how enhancers are regulated during differentiation.

We conclude that LSD1–NuRD complexes present at active promoters in ESCs are essential for normal differentiation, when the active enhancers must be silenced. Given that there is evidence that LSD1 is required for differentiation of multiple cell types[6,25,26], LSD1 is likely to be generally involved in enhancer silencing during differentiation. The ESC gene expression program can be maintained in the absence of many other chromatin regulators[2], and it is possible that some of these also have key functions in the transition from one transcriptional program to another during differentiation.

METHODS SUMMARY

ESC culture conditions. ESCs were grown on irradiated murine embryonic fibroblasts (MEFs) and passaged as described previously[9]. In drug treatment experiments, ESCs were split off MEFs and treated with 1 mM TCP or 3 mM pargyline to inhibit LSD1 activity. Lentiviral constructs were purchased from Open Biosystems and produced according to the *Trans*-lentiviral shRNA Packaging System (catalogue no. TLP4614).

Differentiation assay, immunofluorescence, and alkaline phosphatase staining. ZHBTc4 ESCs were split off MEFs in ESC medium containing 2 µg ml^{-1} doxycycline to decrease Oct4 expression levels. For immunofluorescence, ESCs were crosslinked, blocked and permeabilized before incubation with anti-Oct4 (Santa Cruz, sc-9081x; 1:200 dilution) or anti-SSEA1 (mc-480, Developmental Studies Hybridoma Bank; 1:20 dilution) antibodies. Alexa-conjugated secondary antibodies were used for detection. Staining of ESCs for alkaline phosphatase was achieved with the Alkaline Phosphatase Detection Kit (Millipore, SCR004). Cells were harvested at indicated time points for ChIP-Seq, quantitative polymerase chain reaction or expression array analyses.

ChIP-Seq. Chromatin immunoprecipitations (ChIPs) were performed and analysed as described previously[9]. The following antibodies were used: anti-LSD1 (Abcam, ab17721), anti-Mi-2b (Abcam, ab72418), anti-HDAC1 (Abcam, ab7028), anti-HDAC2 (Abcam, ab7029), anti-REST (Millipore, 07-579), anti-CoREST (Abcam, ab32631), anti-H3K4me1 (Abcam, ab8895), anti-p300 (Santa-Cruz, sc-584) and anti-H3K27Ac (Abcam, ab4729).

For ChIP-Seq analyses, reads were aligned with Bowtie and analysed as described in Supplementary Information.

1. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).
2. Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
3. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription *in vivo*. *Nature* **461**, 186–192 (2009).
4. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
5. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
6. Wang, J. *et al.* Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* **446**, 882–887 (2007).
7. Wang, J. *et al.* The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nature Genet.* **41**, 125–129 (2009).
8. Foster, C. T. *et al.* Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability. *Mol. Cell. Biol.* **30**, 4851–4863 (2010).
9. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
10. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
11. Pardo, M. *et al.* An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* **6**, 382–395 (2010).
12. van den Berg, D. L. *et al.* An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**, 369–381 (2010).
13. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24**, 372–376 (2000).
14. Rossant, J. & Cross, J. C. Placental development: lessons from mouse mutants. *Nature Rev. Genet.* **2**, 538–548 (2001).
15. Metzger, E. *et al.* LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436–439 (2005).
16. Shi, Y. J. *et al.* Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell* **19**, 857–864 (2005).
17. Wang, Y. *et al.* LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell* **138**, 660–672 (2009).
18. Liang, J. *et al.* Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nature Cell Biol.* **10**, 731–739 (2008).
19. Dovey, O. M., Foster, C. T. & Cowley, S. M. Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proc. Natl Acad. Sci. USA* **107**, 8242–8247 (2010).
20. Kaji, K. *et al.* The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nature Cell Biol.* **8**, 285–292 (2006).
21. Scimone, M. L., Meisel, J. & Reddien, P. W. The Mi-2-like *Smed-CHD4* gene is required for stem cell differentiation in the planarian *Schmidtea mediterranea*. *Development* **137**, 1231–1241 (2010).
22. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
23. Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E. & Mattevi, A. Human histone demethylase LSD1 reads the histone code. *J. Biol. Chem.* **280**, 41360–41365 (2005).
24. Lee, M. G. *et al.* Functional interplay between histone demethylase and deacetylase enzymes. *Mol. Cell. Biol.* **26**, 6395–6402 (2006).
25. Musri, M. M. *et al.* Histone demethylase LSD1 regulates adipogenesis. *J. Biol. Chem.* **285**, 30034–30041 (2010).
26. Choi, J. *et al.* Histone demethylase LSD1 is required to induce skeletal muscle differentiation by regulating myogenic factors. *Biochem. Biophys. Res. Commun.* **401**, 327–332 (2010).
27. Okumura-Nakanishi, S., Saito, M., Niwa, H. & Ishikawa, F. Oct-3/4 and Sox2 regulate *Oct-3/4* gene in embryonic stem cells. *J. Biol. Chem.* **280**, 5307–5317 (2005).
28. Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996).
29. Nakatake, Y. *et al.* Klf4 cooperates with Oct3/4 and Sox2 to activate the *Lefty1* core promoter in embryonic stem cells. *Mol. Cell. Biol.* **26**, 7772–7782 (2006).
30. Ray, S. *et al.* Context-dependent function of regulatory elements and a switch in chromatin occupancy between GATA3 and GATA2 regulate Gata2 transcription during trophoblast differentiation. *J. Biol. Chem.* **284**, 4978–4988 (2009).

Author Contributions W.A.W., S.B., H.A.H., C.T.F., S.M.C and R.A.Y. designed, conducted and interpreted the ChIP-Seq, immunofluorescence and expression experiments. W.A.W., D.A.O. and G.M.F. performed data analysis. The manuscript was written by S.B., W.A.W. D.A.O., H.A.H., G.M.F. and R.A.Y.

LETTER

Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma

Jeremy Schwartzentruber[1]*, Andrey Korshunov[2]*, Xiao-Yang Liu[3]*, David T. W. Jones[4], Elke Pfaff[4], Karine Jacob[3], Dominik Sturm[4], Adam M. Fontebasso[3], Dong-Anh Khuong Quang[3], Martje Tönjes[5], Volker Hovestadt[5], Steffen Albrecht[6], Marcel Kool[4], Andre Nantel[7], Carolin Konermann[8], Anders Lindroth[8], Natalie Jäger[9], Tobias Rausch[10], Marina Ryzhova[11], Jan O. Korbel[10], Thomas Hielscher[12], Peter Hauser[13], Miklos Garami[13], Almos Klekner[14], Laszlo Bognar[14], Martin Ebinger[15], Martin U. Schuhmann[16], Wolfram Scheurlen[17], Arnulf Pekrun[18], Michael C. Frühwald[19], Wolfgang Roggendorf[20], Christoph Kramm[21], Matthias Dürken[22], Jeffrey Atkinson[23], Pierre Lepage[1], Alexandre Montpetit[1], Magdalena Zakrzewska[24], Krzysztof Zakrzewski[25], Pawel P. Liberski[24], Zhifeng Dong[26], Peter Siegel[26], Andreas E. Kulozik[27], Marc Zapatka[5], Abhijit Guha[28], David Malkin[29], Jörg Felsberg[30], Guido Reifenberger[30], Andreas von Deimling[2,31], Koichi Ichimura[32], V. Peter Collins[32], Hendrik Witt[4,27], Till Milde[27,33], Olaf Witt[27,33], Cindy Zhang[28], Pedro Castelo-Branco[28], Peter Lichter[5], Damien Faury[3], Uri Tabori[28,29], Christoph Plass[8], Jacek Majewski[3], Stefan M. Pfister[4,27] & Nada Jabado[3,34]

Glioblastoma multiforme (GBM) is a lethal brain tumour in adults and children. However, DNA copy number and gene expression signatures indicate differences between adult and paediatric cases[1–4]. To explore the genetic events underlying this distinction, we sequenced the exomes of 48 paediatric GBM samples. Somatic mutations in the H3.3-ATRX-DAXX chromatin remodelling pathway were identified in 44% of tumours (21/48). Recurrent mutations in *H3F3A*, which encodes the replication-independent histone 3 variant H3.3, were observed in 31% of tumours, and led to amino acid substitutions at two critical positions within the histone tail (K27M, G34R/G34V) involved in key regulatory post-translational modifications. Mutations in *ATRX* (α-thalassaemia/mental retardation syndrome X-linked)[5] and *DAXX* (death-domain associated protein), encoding two subunits of a chromatin remodelling complex required for H3.3 incorporation at pericentric heterochromatin and telomeres[6,7], were identified in 31% of samples overall, and in 100% of tumours harbouring a G34R or G34V H3.3 mutation. Somatic *TP53* mutations were identified in 54% of all cases, and in 86% of samples with *H3F3A* and/or *ATRX* mutations. Screening of a large cohort of gliomas of various grades and histologies ($n = 784$) showed *H3F3A* mutations to be specific to GBM and highly prevalent in children and young adults. Furthermore, the presence of *H3F3A/ATRX-DAXX/TP53* mutations was strongly associated with alternative lengthening of telomeres and specific gene expression profiles. This is, to our knowledge, the first report to highlight recurrent mutations in a regulatory histone in humans, and our data suggest that defects of the chromatin architecture underlie paediatric and young adult GBM pathogenesis.

Brain tumours are currently the leading cause of cancer-related mortality and morbidity in children. Glioblastoma multiforme (GBM) is a highly aggressive brain tumour and the first cancer to be comprehensively profiled by The Cancer Genome Atlas (TCGA) consortium. Whereas GBM is less common in the paediatric setting than in adults, affected children show dismal outcomes similar to adult patients, and the vast majority will die within a few years of diagnosis despite aggressive therapeutic approaches. Tumours arise *de novo* (primary GBM) and are morphologically indistinguishable from their adult counterparts. A number of comprehensive studies have identified transcriptome-based subgroups and indicator mutations in adult GBM, and have thus enabled its molecular sub-classification[8–11]. In contrast, although we and others have demonstrated the presence of distinct molecular subsets of childhood GBM and described different genetic alterations compared to adult cases, the paediatric disease remains understudied[1–4,12]. There is currently insufficient information to improve disease management, and because conventional treatments universally fail, there is a crucial need to identify relevant targets for the design of new therapeutic agents.

To decipher the molecular pathogenesis of paediatric GBM, we undertook a comprehensive mutation analysis in protein-coding genes by performing whole-exome sequencing (WES) on 48 well-characterized paediatric GBMs, including 6 patients for whom we had matched non-tumour (germline) DNA. Samples from the tumour core containing more than 90% neoplastic tissue were collected from patients aged between 3 and 20 years (Supplementary Table 1). Coding regions of the genome were enriched by capture with the Illumina TruSeq kit and sequenced with 100-base-pairs paired-end reads on an Illumina HiSeq
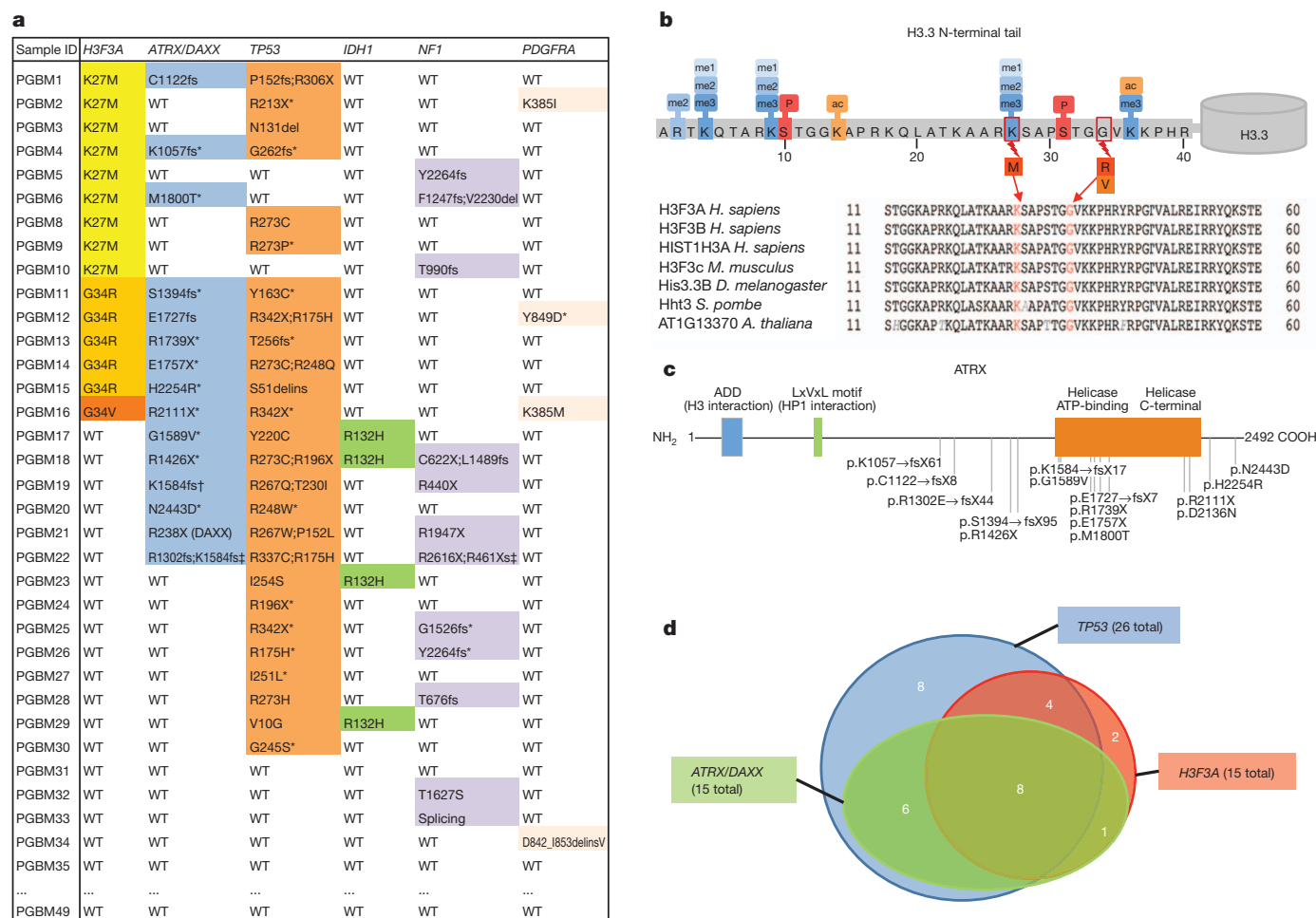
[1]McGill University and Genome Quebec Innovation Centre, Montreal, Quebec H3A 1A4, Canada. [2]Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [3]Departments of Human Genetics and Experimental Medicine, McGill University, Montreal, H3Z 2Z3, Canada. [4]Division of Pediatric Neuro-oncology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [5]Division of Molecular Genetics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [6]Department of Pathology, Montreal Children's Hospital, McGill University Health Center, Montreal, Quebec H3H 1P3, Canada. [7]Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec H4P 2R2, Canada. [8]Division of Epigenomics and Cancer Risk Factors, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [9]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [10]European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany. [11]Burdenko Neurosurgical Institute, Department of Neuropathology, 125047 Moscow, Russia. [12]Division of Biostatistics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [13]2nd Department of Paediatrics, Semmelweis University, Budapest, H-1094 Hungary. [14]Department of Neurosurgery, Medical and Health Science Center, University of Debrecen, Debrecen, H-4032 Hungary. [15]Department of Hematology and Oncology, Children's University Hospital, 72076 Tübingen, Germany. [16]Department of Neurosurgery, University Hospital, 72076 Tübingen, Germany. [17]Cnopfsche Kinderklinik, Nuremberg Children's Hospital, 90419 Nuremberg, Germany. [18]Prof. Hess Children's Hospital, 28177 Bremen, Germany. [19]Klinikum Augsburg, Augsburg Children's Hospital, 86156 Augsburg, Germany. [20]Department of Neuropathology, Institute of Pathology, University of Würzburg, 97080 Würzburg, Germany. [21]University Children's Hospital, 06097 Halle, Germany. [22]University Children's Hospital, 68167 Mannheim, Germany. [23]Department of Surgery, Division of Neurosurgery, Montreal Children's Hospital, McGill University Health Center, Montreal, Quebec H3H 1P3, Canada. [24]Department of Molecular Pathology and Neuropathology, Medical University of Lodz, 92-216 Poland. [25]Department of Neurosurgery, Polish Mother's Memorial Hospital Research Institute, Lodz, 93-338 Poland. [26]Rosalind and Morris Goodman Cancer Research Centre, McGill University, Montreal, Quebec H3A 1A3, Canada. [27]Department of Paediatric Oncology, Hematology and Immunology, University Hospital Heidelberg, 69120 Heidelberg, Germany. [28]The Arthur and Sonia Labatt Brain Tumor Research Centre, Hospital for Sick Children Research Institute, University of Toronto, Toronto, Ontario M5G1X8, Canada. [29]Division of Hematology/Oncology, Department of Paediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario M5G 1X8, Canada. [30]Department of Neuropathology, Heinrich-Heine-University, 40225 Düsseldorf, Germany. [31]Department of Neuropathology, Heidelberg University Hospital, 69120 Heidelberg Germany. [32]Division of Molecular Histopathology, Department of Pathology, University of Cambridge, CB2 0QQ, UK. [33]Clinical Cooperation Unit Paediatric Oncology, German Cancer Research Center, 69120 Heidelberg, Germany. [34]Department of Paediatrics, McGill University and the McGill University Health Center Research Institute, Montreal, Quebec H3Z 2Z3, Canada.
*These authors contributed equally to this work.

2000 platform (Supplementary Methods). The median coverage of each base in the targeted regions was 61-fold, and 91% of the bases were represented by at least 10 reads (Supplementary Table 2). We identified 87 somatic mutations in 80 genes among the 6 tumours for which we had matched constitutive DNA. The mutation count per tumour ranged from 3 to 31, with a mean of 15 (Supplementary Table 3). This is much lower than the rate observed using Sanger sequencing in other solid tumours including adult GBM[11], but somewhat higher than in another paediatric brain tumour, medulloblastoma[13] (Supplementary Table 4). Relevant mutations (as defined below) were validated by Sanger sequencing.

Initially, we focused on the distribution of somatic, non-silent protein-coding mutations in the six tumours with matched germline DNA. Four samples had recurrent heterozygous mutations in *H3F3A*, which encodes the replication-independent histone variant H3.3. Both mutations were single-nucleotide variants (SNVs), in two samples changing lysine 27 to methionine (K27M), and in two samples changing glycine 34 to arginine (G34R) (Fig. 1a and Supplementary Table 3). These mutations are particularly interesting because histone

genes are highly conserved throughout eukaryotes (Fig. 1b), and to our knowledge no human disorders have specifically been associated with mutations in histones, including H3.3. Both mutations are at or very near positions in the amino-terminal tail of the protein that undergo important post-translational modifications associated with either transcriptional repression (K27) or activation (K36) (Fig. 1b). All four samples additionally harboured mutations in *ATRX*, which encodes a member of a transcription/chromatin remodelling complex required for the incorporation of H3.3 at pericentric heterochromatin and at telomeres, as well as at several transcription factor binding sites[7,14–17]. We extended our WES analysis to 42 additional tumour samples and focused on *ATRX* and *H3F3A*, as well as *DAXX* (because the gene product heterodimerizes with ATRX and participates in H3.3 recruitment to DNA[6,7]). A total of 15 samples had heterozygous H3.3 mutations (9 K27M, 5 G34R, 1 G34V) and 14 samples had a mutation in *ATRX*, including frameshift insertions/deletions (6 samples), gains of a stop codon (4 samples), and missense SNVs (4 samples). Nearly all of the *ATRX* mutations occurred either within the carboxy-terminal helicase domain or led to truncation of the protein upstream of this

Sample ID	H3F3A	ATRX/DAXX	TP53	IDH1	NF1	PDGFRA
PGBM1	K27M	C1122fs	P152fs;R306X	WT	WT	WT
PGBM2	K27M	WT	R213X*	WT	WT	K385I
PGBM3	K27M	WT	N131del	WT	WT	WT
PGBM4	K27M	K1057fs*	G262fs*	WT	WT	WT
PGBM5	K27M	WT	WT	WT	Y2264fs	WT
PGBM6	K27M	M1800T*	WT	WT	F1247fs;V2230del	WT
PGBM8	K27M	WT	R273C	WT	WT	WT
PGBM9	K27M	WT	R273P*	WT	WT	WT
PGBM10	K27M	WT	WT	WT	T990fs	WT
PGBM11	G34R	S1394fs*	Y163C*	WT	WT	WT
PGBM12	G34R	E1727fs	R342X;R175H	WT	WT	Y849D*
PGBM13	G34R	R1739X*	T256fs*	WT	WT	WT
PGBM14	G34R	E1757X*	R273C;R248Q	WT	WT	WT
PGBM15	G34R	H2254R*	S51delins	WT	WT	WT
PGBM16	G34V	R2111X*	R342X*	WT	WT	K385M
PGBM17	WT	G1589V*	Y220C	R132H	WT	WT
PGBM18	WT	R1426X*	R273C;R196X	R132H	C622X;L1489fs	WT
PGBM19	WT	K1584fs†	R267Q;T230I	WT	R440X	WT
PGBM20	WT	N2443D*	R248X*	WT	WT	WT
PGBM21	WT	R238X (DAXX)	R267W;P152L	WT	R1947X	WT
PGBM22	WT	R1302fs;K1584fs‡	R337C;R175H	WT	R2616X;R461Xs‡	WT
PGBM23	WT	WT	I254S	R132H	WT	WT
PGBM24	WT	WT	R196X*	WT	WT	WT
PGBM25	WT	WT	R342X*	WT	G1526fs*	WT
PGBM26	WT	WT	R175H*	WT	Y2264fs*	WT
PGBM27	WT	WT	I251L*	WT	WT	WT
PGBM28	WT	WT	R273H	WT	T676fs	WT
PGBM29	WT	WT	V10G	R132H	WT	WT
PGBM30	WT	WT	G245S*	WT	WT	WT
PGBM31	WT	WT	WT	WT	WT	WT
PGBM32	WT	WT	WT	WT	T1627S	WT
PGBM33	WT	WT	WT	WT	Splicing	WT
PGBM34	WT	WT	WT	WT	WT	D842_I853delinsV
PGBM35	WT	WT	WT	WT	WT	WT
...	...	...	...	...	...	...
PGBM49	WT	WT	WT	WT	WT	WT

*Homozygous mutations.
†Sample PGBM19 additionally has a DAXX mutation C629Sfs, whereas PGBM21 has no ATRX mutation but has the DAXX mutation shown.
‡Sample PGBM22 has a third ATRX mutation, p.D2136N, and a third NF1 mutation, p.A887T.

Figure 1 | Most frequent mutations in paediatric GBM. a, Most frequent somatic mutations in 48 paediatric glioblastoma tumours. Mutations identified in genes listed in this table were confirmed by Sanger sequencing, and were not present in dbSNP nor in the 1000 Genomes data set (October 2011), except for the *TP53* SNP at R273, which is associated with cancer. Detailed description of the mutations in affected samples is provided in Supplementary Table 5. **b**, Three recurrent non-synonymous single nucleotide variants (SNVs) were observed in *H3F3A*. The K27M, G34R and G34V mutations are shown in the

context of the common post-translational modifications of the H3.3 N-terminal tail, which regulates the histone code. H3.3 has 136 amino acids, and is highly conserved across species from mammals to plants, including the residues subject to mutation in paediatric GBM (see multiple alignment of amino acids 11 to 60). **c**, Schematic of the mutations observed in ATRX in the 48 WES samples. **d**, Schematic of the overlap between mutations affecting *ATRX-DAXX*, *H3F3A* and *TP53*. Eight samples had all three mutations.

domain (Fig. 1c). Mutations were accompanied by an absence of detectable ATRX protein by immunohistochemistry in samples for which paraffin material was available (Supplementary Fig. 1). Two samples had heterozygous *DAXX* mutations, simultaneously with an *ATRX* mutation in one sample (Fig. 1a and Supplementary Table 3). Overall, 21 of 48 samples (44%) had a mutation in at least one of these three genes. Notably, we also identified *TP53* mutations in 26 samples (25 somatic, 1 germline in PGBM26), which overlapped significantly with samples that had *ATRX*, *DAXX* and/or *H3F3A* mutations (18/21 cases, 86%, Fig. 1d; $P = 1.1 \times 10^{-4}$, permutation test). A list of all mutations discovered by WES in selected genes associated with GBM is given in Supplementary Table 5.

H3F3A, *ATRX* or *DAXX* were not part of the 600 genes sequenced by The Cancer Genome Atlas (TCGA) glioblastoma project[10,18], and no *H3F3A* mutations were identified in 22 adult GBM samples sequenced previously[11]. To investigate whether *H3F3A* mutations are specific to GBM and/or paediatric disease, we sequenced this gene in 784 glioma samples from all grades and histological diagnoses

across the entire age range (Fig. 2a). H3.3 mutations were highly specific to GBM and were much more prevalent in the paediatric setting (32/90, 36%), although they also occurred rarely in young adults with GBM (11/318, 3%) (Fig. 2b). K27M-H3.3 mutations occurred mainly in younger patients (median age 11 years, range 5–29) and thalamic GBM (Supplementary Table 1), whereas G34R- or G34V-H3.3 mutations occurred in older patients (median age 20 years, range 9–42) and in tumours of the cerebral hemispheres (Fig. 2b). Further comparison of our data set with adult GBM databases[8,10,11,18,19] indicated limited overlap in frequently mutated genes between paediatric GBM and any of the four previously described adult GBM subtypes[18] (Fig. 2c, Supplementary Fig. 2 and Supplementary Table 6).

Somatic mutations in *ATRX* and *DAXX* have recently been reported in a large proportion (43%) of pancreatic neuroendocrine tumours (PanNETs), a rare form of pancreatic cancer with a 10-year overall survival of ∼40%, and no reported association with *TP53* or *H3F3A* mutations[19]. A follow-up study found *ATRX* mutations in a series of
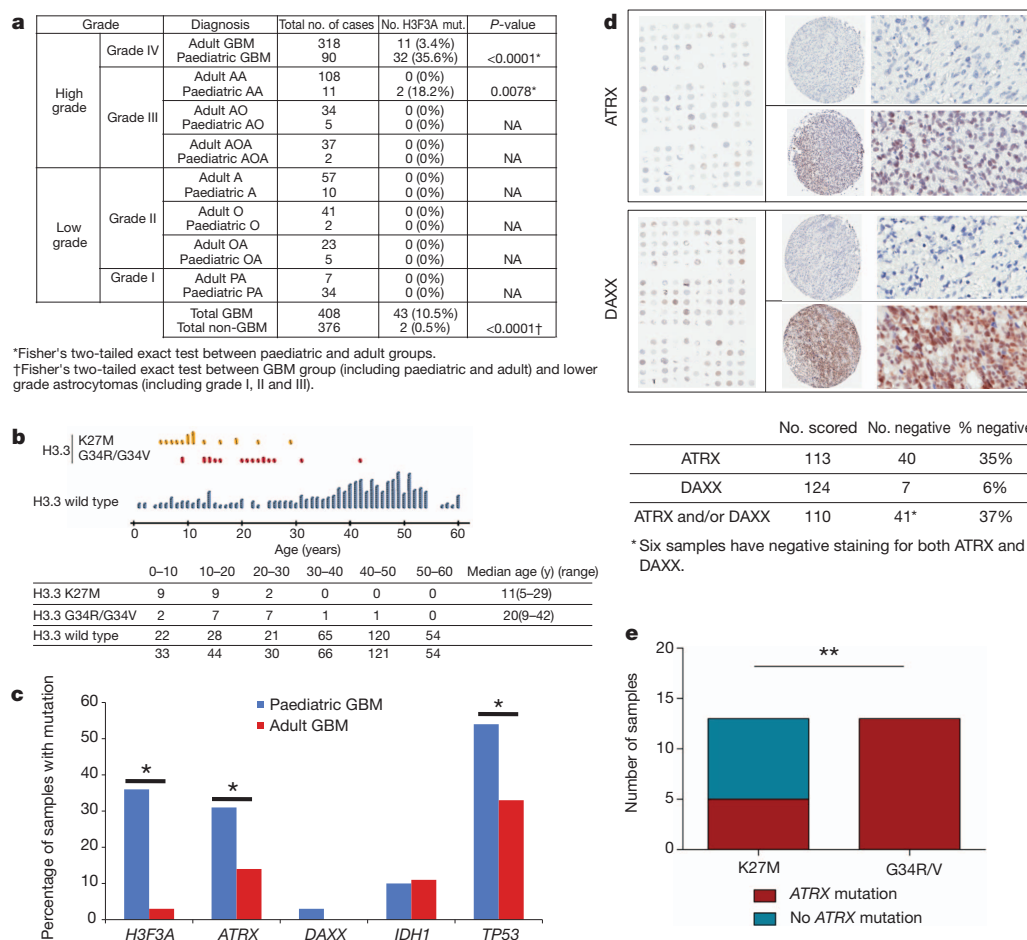

Figure 2 | **Mutations in *H3F3A*, *ATRX* and *DAXX* distinguish paediatric from adult GBM. a**, H3F3A mutations in a set of 784 gliomas from all ages and grades. *H3F3A* mutations are exclusive to high-grade tumours and the vast majority occur in glioblastoma (GBM) and in the paediatric setting. A, diffuse astrocytoma grade II; AA, anaplastic astrocytoma; AO, anaplastic oligodendroglioma; AOA, anaplastic oligoastrocytoma; O, oligodendroglioma; OA, oligoastrocytoma; PA, pilocytic astrocytoma. **b**, H3.3 mutations are specific to paediatric and young adult glioblastoma (GBM). K27M-H3.3 mutations occur mainly in younger patients (median age 11 years) and G34R/V-H3.3 mutations occur in older children and young adults (median age 20 years). No H3.3 mutations were identified in older patients with GBM. **c**, Comparison of the most frequently mutated genes in paediatric and adult GBM shows that *H3F3A*, *ATRX* and *DAXX* mutations are largely specific to paediatric disease. Except for similarities in the mutation rate for *TP53* and

PDGFRA with the previously identified proneural adult GBM subgroup, the rate and type of genes mutated were distinct between paediatric and adult GBM whatever the molecular subgroup[27] (Supplementary Fig. 2). Data for adult GBM regarding other genes included in the table was compiled from refs 11 and 18. **d**, ATRX and DAXX immunohistochemical staining of a paediatric GBM tissue microarray (TMA) comprising 124 samples. View of the TMA slide and an example of a negative and of a positive core at high magnification to show specific nuclear staining (or lack thereof) for DAXX and ATRX. No gender bias for ATRX loss was observed. Overall survival and progression-free survival were similar in patients with and without loss of ATRX and/or DAXX (data not shown). **e**, Differential association of K27M and G34R/V H3F3A mutations with ATRX mutations. G34R/V-H3.3 mutations were always associated with ATRX mutations (two-sided Fisher's exact test, *P* = 0.0016), whereas a nonsignificant overlap was observed for K27M.

cancers, including GBM, where *ATRX* (but not *DAXX*) mutations were identified in 3/21 paediatric GBMs (14%) and 8/122 adult GBMs (7%)[20]. To evaluate further the prevalence of *ATRX* and *DAXX* mutations in paediatric GBM, we performed immunostaining for these proteins on a well-characterized tissue microarray (TMA) with samples from 124 paediatric GBM patients. Lack of immuno-positivity for ATRX was seen in 35% of cases (40/113 scored, 22 females and 18 males) and for DAXX in 6% (7/124 scored) (Fig. 2d and Supplementary Fig. 1). Overall, 37% of samples had lost nuclear expression of either factor, corroborating our WES findings. Strikingly, *ATRX-DAXX* mutations (as assessed by direct sequencing or loss of protein expression) were found in 100% of G34-H3.3 mutant cases in the larger cohort of GBMs (13/13) where sufficient material was available ($P = 1.4 \times 10^{-8}$, permutation test). The overlap of *ATRX* mutations with K27M-H3.3-mutated samples was not significant in either the exome data set (3/9 samples, $P = 0.58$) or the full set of GBM screened (5/13, $P = 0.40$) (Fig. 2e).

The histone code—post-translational modifications of specific histone residues—regulates virtually all processes that act on or depend on DNA, including replication and repair, regulation of gene expression, and maintenance of centromeres and telomeres[21]. Accordingly, although recurrent histone mutations have not previously been reported in cancer, mutations in genes affecting histone post-translational modifications are increasingly described[22]. H3.3 is a universal, replication-independent histone predominantly incorporated into transcription sites and telomeric regions, and associated with active and open chromatin (reviewed in ref. 23). This role is conserved in the single histone H3 present in yeast, indicating its importance throughout evolution. It functions as a neutral replacement histone, but also participates in the epigenetic transmission of active chromatin states and is associated with chromatin assembly factors in large-scale replication-independent chromatin remodelling events[23].

The non-random recurrence of the exact same mutation in different tumours, and the absence of truncating mutations, indicate that
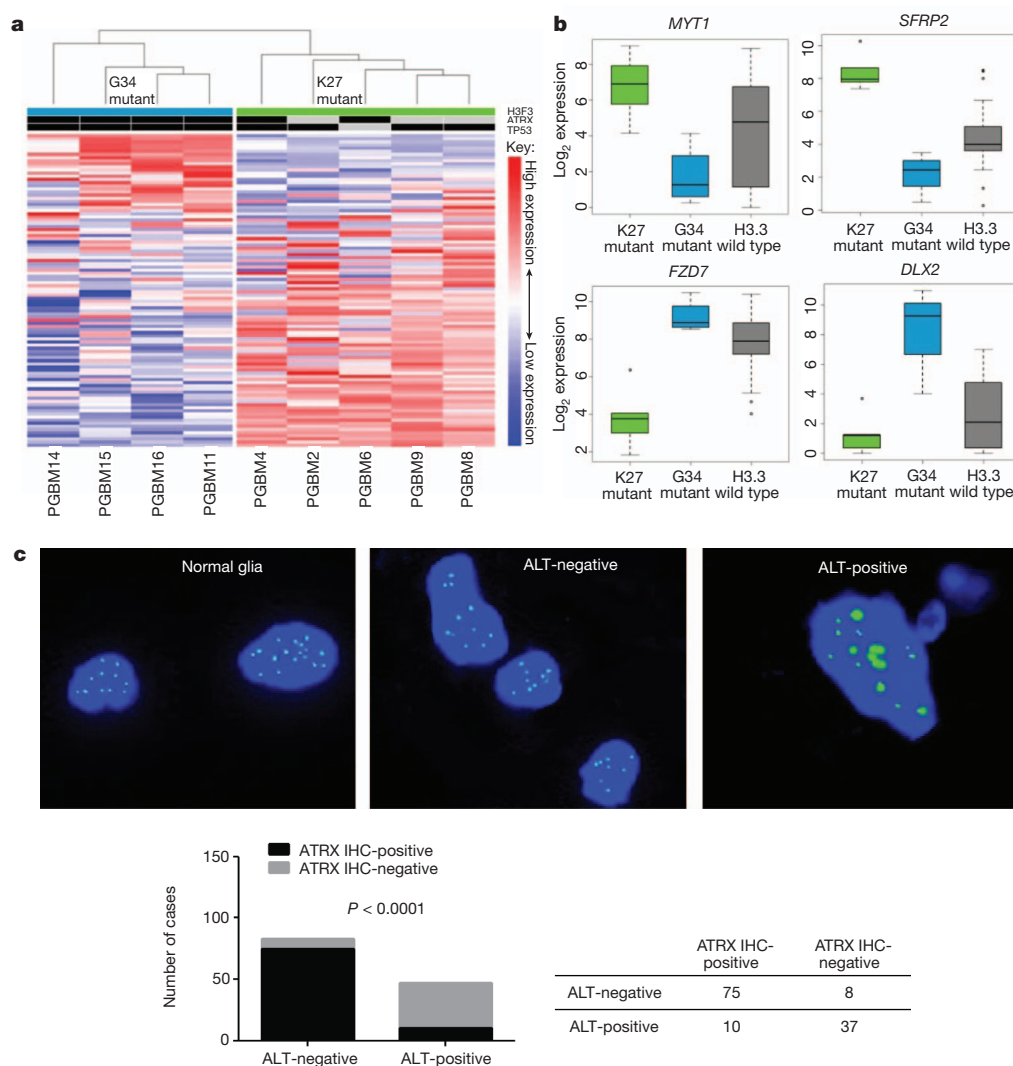

Figure 3 | H3F3A mutation variants show distinct expression profiles and are associated with alternative lengthening of telomeres. a, Unsupervised hierarchical clustering of differentially expressed genes in 27 of the 48 GBM samples analysed by whole-exome sequencing shows that samples with K27M and G34R/V H3.3 have specific gene expression profiles. Clustering was based on the top 100 genes by standard deviation from autosomal genes detected as present in >10% of samples (see also Supplementary Fig. 3). **b**, Genes involved in development and differentiation show H3.3 mutation-specific expression patterns. Expression levels of developmental-related genes including *DLX2*, *SFRP2*, *FZD7* and *MYT1* are distinct among H3.3-K27 mutant and H3.3-G34 mutants following gene expression profiling (see also Supplementary Table 7). **c**, Alternate lengthening of telomere (ALT) is associated with the presence of mutant H3F3A/ATRX in a tissue microarray (TMA) comprising 124 paediatric GBM samples. We assessed ALT using telomere-specific FISH (shown here and in Supplementary Fig. 4) on the paediatric TMA we investigated for ATRX expression (Fig. 2d) and using telomere-specific Southern blotting of high molecular weight genomic DNA (data not shown). Fisher's exact test was used to identify any association relationship. Representative images of ALT-positive and -negative staining of a paediatric GBM tissue microarray and a control brain are provided.

H3F3A mutations are most probably gain-of-function events. Lysine 27 is a critical residue of histone 3 and its variants, and methylation at this position (H3K27me), which may be mimicked by the terminal CH₃ of methionine substituted at this residue, is commonly associated with transcriptional repression[24]. In contrast, H3K36 methylation or acetylation typically promotes gene transcription[25,26]. Thus, although their morphological phenotype is very similar (K27M and G34R/V mutant tumours are histologically indistinguishable), the two H3.3 variants are expected to act through a different set of genes. This indeed seems to be the case when looking at expression profiles of GBMs harbouring these two mutations. Unsupervised hierarchical clustering of gene expression from 27 of the WES cohort samples for which sufficient RNA was available revealed a clear separation in the expression of K27M versus G34R/V mutant samples (Supplementary Fig. 3). Further analysis of just those samples harbouring an *H3F3A* mutation additionally showed a clear distinction in the expression pattern of these two variants (Fig. 3a and Supplementary Table 7). Among these differentially expressed genes were several linked to brain development that showed a clear mutation-specific expression pattern when comparing both between K27 and G34 mutants and with H3.3 wild-type GBMs, including *DLX2*, *SFRP2*, *FZD7* and *MYT1* (Fig. 3b). We also identified increased levels of H3K36 trimethylation in cells carrying the G34V-H3.3 mutation in one sample for which we had available material (PGBM14) compared to other cells, potentially supporting this hypothesis (Supplementary Fig. 5).

ATRX loss, frequently observed in this study, has recently been shown to be associated with alternative lengthening of telomeres (ALT) in PanNETs and GBMs[20]. We performed telomere-specific fluorescence *in situ* hybridization (FISH) on the samples with K27M or G34R/V mutations identified by WES for which we had slides available (Supplementary Fig. 4) and on the paediatric GBM TMA (Fig. 3c). These experiments showed that ALT is strongly correlated with ATRX loss (37/47 samples with ALT showed ATRX loss, $P < 0.001$). However, some samples with nuclear ATRX staining still showed ALT, indicating that additional defects may also account for elongated telomeres in GBM. The presence of ALT was best explained by the simultaneous presence of *ATRX/H3F3A/TP53* mutations ($P = 0.0002$, Fisher's exact test). Tumours without *ATRX/H3F3A/TP53* mutations almost invariably showed shorter telomeres than are observed with ALT, as seen in telomerase-positive gliomas[27].

Genetic stability was also assessed through evaluating DNA copy number aberrations (CNAs) in 31 of the 48 tumours using Illumina SNP arrays containing ~2.5 million oligonucleotides (Supplementary Tables 1, 8, 9). Loss of heterozygosity (whole chromosome changes, broad and focal heterozygous deletions, Supplementary Table 9) was common in paediatric GBM samples, as we have previously reported[4], and the focal gains and losses we identified in our study showed a high degree of overlap with other published paediatric data sets[3,21]. The number of CNAs per tumour was higher in samples with *H3F3A/ATRX-DAXX/TP53* mutations (Supplementary Fig. 6).

Recurrent point mutations in *IDH1* (mainly R132H) are gain of function mutations commonly identified in secondary GBM and the lower-grade tumours from which they develop (86–98% of these astrocytomas), and typically occur in younger adults[11,28]. Strikingly, *IDH1* and *H3F3A* mutations were mutually exclusive in our sequencing cohort ($P = 1.6 \times 10^{-4}$). Neomorphic enzyme activity resulting from *IDH1* mutation leads to the production of high quantities of the onco-metabolite 2-hydroxyglutarate (2-HG)[29]. Increased 2-HG inhibits histone demethylases, specifically inducing increased methylation of both H3K27 and H3K36[29,30], the two residues affected directly (K27) or indirectly (K36) by the mutations in *H3F3A* uncovered in this study. Furthermore, overlap of *H3F3A* and *TP53* mutations in children with GBM (all of the G34R/V and 82% of K27M mutants also harbour *TP53* mutations) mirrors the large overlap of *IDH1* mutations with *TP53* mutations in the proneural adult GBM

sub-group[18]. Thus, mutations which directly (*H3F3A*), or indirectly (*IDH1*) affect the methylation of H3.3 K27 or H3.3 K36, in combination with *TP53* mutations, characterize the pathogenesis of paediatric and young adult GBM.

Our data indicate a central role of H3.3/ATRX-DAXX perturbation in paediatric GBM. Mutant H3.3 recruitment would occur across the genome and induce abnormal patterns of chromatin remodelling to yield distinct gene expression profiles for the K27 and G34 mutations. Additional loss of ATRX may act to reduce H3.3 incorporation at a subset of genes important in oncogenesis, preventing mutant H3.3 from altering their transcription. ATRX loss will also impair H3.3 loading at telomeres and disrupt their heterochromatic state, facilitating alternative lengthening of telomeres (ALT). Our findings provide an intriguing example of the interplay of genetic and epigenetic events in driving cancer, indicate a new mechanism through which these epigenetic alterations are brought about (mutation of key residues in a regulatory histone), and provide a rationale for targeting the chromatin remodelling machinery in this deadly paediatric cancer.

METHODS SUMMARY

All samples were obtained with informed consent after approval of the Institutional Review Board of the respective hospitals they were treated in and were independently reviewed by senior paediatric neuropathologists (S.A., A.K.) according to the World Health Organization guidelines. Standard manufacturer protocols were used to perform target capture with the Illumina TruSeq exome enrichment kit and sequencing of 100 bp paired end reads on Illumina HiSeq. We generated approximately 10 gigabases of sequence for each subject such that >90% of the coding bases of the exome defined by the consensus coding sequence (CCDS) project were covered by at least 10 reads. We removed adaptor sequences and quality trimmed reads using the Fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and then used a custom script to ensure that only read pairs with both mates present were subsequently used. A complete description of the materials and methods is provided in the Supplementary Information.

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

1. Faury, D. *et al.* Molecular profiling identifies prognostic subgroups of pediatric glioblastoma and shows increased YB-1 expression in tumors. *J. Clin. Oncol.* **25**, 1196–1208 (2007).
2. Haque, T. *et al.* Gene expression profiling from formalin-fixed paraffin-embedded tumors of pediatric glioblastoma. *Clin. Cancer Res.* **13**, 6284–6292 (2007).
3. Paugh, B. S. *et al.* Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. *J. Clin. Oncol.* **28**, 3061–3068 (2010).
4. Qu, H. Q. *et al.* Genome-wide profiling using single-nucleotide polymorphism arrays identifies novel chromosomal imbalances in pediatric glioblastomas. *Neuro-oncol.* **12**, 153–163 (2010).
5. Villard, L. *et al.* XNP mutation in a large family with Juberg-Marsidi syndrome. *Nature Genet.* **12**, 359–360 (1996).
6. Dhayalan, A. *et al.* The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9. *Hum. Mol. Genet.* **20**, 2195–2203 (2011).
7. Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C. & Allis, C. D. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc. Natl Acad. Sci. USA* **107**, 14075–14080 (2010).
8. Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010).
9. Phillips, H. S. *et al.* Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157–173 (2006).
10. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008) Medline CrssRef.
11. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
12. Bax, D. A. *et al.* A distinct spectrum of copy number aberrations in pediatric high-grade gliomas. *Clin. Cancer Res.* **16**, 3368–3377 (2010).
13. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331**, 435–439 (2011).
14. Iwase, S. *et al.* ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nature Struct. Mol. Biol.* **18**, 769–776 (2011).

15. Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140,** 678–691 (2010).
16. Wong, L. H. *et al.* ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res.* **20,** 351–360 (2010).
17. De La Fuente, R., Viveiros, M. M., Wigglesworth, K. & Eppig, J. J. ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. *Dev. Biol.* **272,** 1–14 (2004).
18. Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA, IDH1, EGFR,* and *NF1. Cancer Cell* **17,** 98–110 (2010).
19. Jiao, Y. *et al.* DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science* **331,** 1199–1203 (2011).
20. Heaphy, C. M. *et al.* Altered telomeres in tumors with *ATRX* and *DAXX* mutations. *Science* **333,** 425 (2011).
21. Chi, P., Allis, C. D. & Wang, G. G. Covalent histone modifications–miswritten, misinterpreted and mis-erased in human cancers. *Nature Rev. Cancer* **10,** 457–469 (2010).
22. Füllgrabe, J., Kavanagh, E. & Joseph, B. Histone onco-modifications. *Oncogene* **30,** 3391–3403 (2011).
23. Talbert, P. B. & Henikoff, S. Histone variants–ancient wrap artists of the epigenome. *Nature Rev. Mol. Cell Biol.* **11,** 264–275 (2010).
24. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
25. Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genet.* **41,** 376–381 (2009).
26. Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **27,** 406–420 (2008).
27. Hakin-Smith, V. *et al.* Alternative lengthening of telomeres and survival in patients with glioblastoma multiforme. *Lancet* **361,** 836–838 (2003).
28. Yan, H. *et al. IDH1* and *IDH2* mutations in gliomas. *N. Engl. J. Med.* **360,** 765–773 (2009).
29. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462,** 739–744 (2009).
30. Cervera, A. M., Bayley, J. P., Devilee, P. & McCreath, K. J. Inhibition of succinate dehydrogenase dysregulates histone modification in mammalian cells. *Mol. Cancer* **8,** 89 (2009).

METHODS

Sample characteristics and pathological review. All samples were obtained with informed consent after approval of the Institutional Review Board of the respective hospitals they were treated in and were independently reviewed by senior paediatric neuropathologists (S.A., A.K.) according to the WHO guidelines. Forty-eight paediatric grade IV astrocytomas (glioblastoma GBM) patients between the age of 3 and 20 years were included in the study. Clinical characteristics of patients are summarized in Supplementary Table 1. Samples were taken at the time of the first surgery, before further treatment as needed. Tissues were obtained from the London/Ontario Tumour Bank of the Pediatric Cooperative Health Tissue Network, the Montreal Children's Hospital and from collaborators in Hungary and Germany. Seven hundred and eighty-four glioma samples from all grades and histological diagnoses across the entire age range in this study were obtained from collaborators across Europe and North America.

Alignment and variant calling for whole-exome sequencing. We followed standard manufacturer protocols to perform target capture with the Illumina TruSeq exome enrichment kit and sequencing of 100 bp paired end reads on Illumina HiSeq. We generated approximately 10 Gb of sequence for each subject such that >90% of the coding bases of the exome defined by the consensus coding sequence (CCDS) project were covered by at least 10 reads. We removed adaptor sequences and quality trimmed reads using the Fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and then used a custom script to ensure that only read pairs with both mates present were subsequently used. Reads were aligned to hg19 with BWA[31], and duplicate reads were marked using Picard (http://picard.sourceforge.net/) and excluded from downstream analyses. Single nucleotide variants (SNVs) and short insertions and deletions (indels) were called using samtools (http://samtools.sourceforge.net/) pileup and varFilter[32] with the base alignment quality (BAQ) adjustment disabled, and were then quality filtered to require at least 20% of reads supporting the variant call. Variants were annotated using both ANNOVAR[33] and custom scripts to identify whether they affected protein coding sequence, and whether they had previously been seen in dbSNP131, the 1000 Genomes data set (October 2011), or in approximately 160 exomes previously sequenced at our centre.

Somatic mutation identification for whole-exome sequencing. A variant called in a tumour was considered to be a candidate somatic mutation if the matched normal sample had at least 10 reads covering this position and had zero variant reads, and the variant was not reported in dbSNP131 or the 1000 Genomes data set (October 2011). For the resulting 117 candidate somatic mutations, we manually examined the alignment of each to check for sequencing artefacts and alignment errors. Fifteen variants were easily identified as sequence-specific error artefacts commonly seen shortly downstream of GGC sequences on Illumina sequencers[34]. Once genes of interest were identified (*H3F3A*, *ATRX*, *DAXX*, *TP53*, *NF1*), we examined positions in these genes in the 34 tumour samples where less than 20% of the reads supported the variant. This identified only two additional variants,

both in sample PGBM19 where there were low read counts for frameshift insertions in both *ATRX* (6/32 reads) and *DAXX* (8/47 reads).

Immunohistochemistry and immunoblotting. Formalin-fixed, paraffin-embedded sections of paediatric GBM and TMA (4 μm) were immunohistochemically stained for ATRX and DAXX proteins. Unstained sections were subjected to antigen retrieval in 10 mM citrate buffer (pH 6.0) for 10 min at sub-boiling temperatures. Individual slides were incubated overnight at 4 °C with rabbit anti-ATRX (1:750 dilution, Sigma, catalogue no. HPA001906) or rabbit anti-DAXX (1:100 dilution, Sigma, catalogue no. HPA008736) antibodies. After incubation with the primary antibody, secondary biotin-conjugated donkey anti-rabbit antibodies (Jackson) were applied for 30 min. After washing with PBS, slides were developed with diaminobenzidine (Dako) as the chromogen. All slides were counterstained using Harris haematoxylin. The criterion for positive staining was described previously[20]. Immunohistochemistry staining on TMA was scored by three individuals independently, including a pathologist. To test the level of mono-, di- and trimethylated H3 at position K36, cell lysates from tumour cells were analysed by western blot. Antibodies against H3K36me3 (Abcam, catalogue no. ab9050), H3K36me2 (Abcam, catalogue no. ab9049), H3K36me1 (Abcam, catalogue no. ab9048) and H3.3 (Abcam, catalogue no. ab97968) were used, with conditions suggested by the manufacturer.

Gene expression profiling. Total RNA from frozen samples was hybridized to Affymetrix-HG-U133 plus 2.0 gene chips (Affymetrix). Array quality assurance was determined using β-actin and GAPDH 3′/5′ ratio, as recommended by the manufacturer.

Genome-wide SNP array. DNA from 31 of the 48 paediatric GBM tumours analysed by whole-exome sequencing was hybridized to Illumina Human Omni 2.5M Single Nucleotide Polymorphism (SNP) arrays, according to the manufacturer's protocol. Copy number alterations were analysed using Illumina GenomeStudio Data Analysis Software (Illumina) as previously described[35]. Statistical analysis of Fisher's exact test was performed using GraphPad Prism software.

Telomere specific fluorescence *in situ* hybridization (FISH). Telomere-specific FISH was done using a standard formalin-fixed paraffin-embedded FISH protocol (as described in ref. 20), using a FITC peptide nucleic acid telomere probe from Dako.

31. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
32. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
33. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).
34. Nakamura, K. *et al.* Sequence-specific error profile of Illumina sequences. *Nucleic Acids Res.* **39,** e90 (2011).
35. Peiffer, D. A. *et al.* High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* **16,** 1136–1148 (2006).

LETTER

High-throughput decoding of antitrypanosomal drug efficacy and resistance

Sam Alsford[1], Sabine Eckert[2]†, Nicola Baker[1], Lucy Glover[1], Alejandro Sanchez-Flores[2], Ka Fai Leung[3], Daniel J. Turner[2]†, Mark C. Field[3], Matthew Berriman[2] & David Horn[1]

The concept of disease-specific chemotherapy was developed a century ago. Dyes and arsenical compounds that displayed selectivity against trypanosomes were central to this work[1,2], and the drugs that emerged remain in use for treating human African trypanosomiasis (HAT)[3]. The importance of understanding the mechanisms underlying selective drug action and resistance for the development of improved HAT therapies has been recognized, but these mechanisms have remained largely unknown. Here we use all five current HAT drugs for genome-scale RNA interference target sequencing (RIT-seq) screens in *Trypanosoma brucei*, revealing the transporters, organelles, enzymes and metabolic pathways that function to facilitate antitrypanosomal drug action. RIT-seq profiling identifies both known drug importers[4,5] and the only known pro-drug activator[6], and links more than fifty additional genes to drug action. A bloodstream stage-specific invariant surface glycoprotein (ISG75) family mediates suramin uptake, and the AP1 adaptin complex, lysosomal proteases and major lysosomal transmembrane protein, as well as spermidine and *N*-acetylglucosamine biosynthesis, all contribute to suramin action. Further screens link ubiquinone availability to nitro-drug action, plasma membrane P-type H$^+$-ATPases to pentamidine action, and trypanothione and several putative kinases to melarsoprol action. We also demonstrate a major role for aquaglyceroporins in pentamidine and melarsoprol cross-resistance. These advances in our understanding of mechanisms of antitrypanosomal drug efficacy and resistance will aid the rational design of new therapies and help to combat drug resistance, and provide unprecedented molecular insight into the mode of action of antitrypanosomal drugs.

African trypanosomes are transmitted by the tsetse insect vector and circulate in the bloodstream and tissue fluids of their mammalian hosts. These protozoan parasites cause HAT, also known as sleeping sickness, and the livestock disease known as Nagana. HAT is typically fatal if there is no chemotherapeutic intervention. The public health situation has improved recently with increased monitoring and chemotherapy averting more than 1.3 million disability-adjusted life years (DALYs) in the year 2000 and the estimated number of cases at less than 70,000 in 2006 (ref. 7). However, therapies have many problems, including severe toxicity and increasing resistance, which is a major concern owing to the absence of a vaccine or therapeutic alternatives[3]. The current HAT therapies are pentamidine or suramin, which are only suitable for the first stage of the disease before central nervous system involvement, and eflornithine, nifurtimox or melarsoprol for advanced disease[3] (Supplementary Table 1). All of these drugs were developed well before the advent of molecular, target-based therapy and, with the exception of eflornithine, they elicit their antitrypanosomal effects by disrupting unknown targets. HAT treatment failure rates were reported to be increasing for suramin, when this drug was still in use in West Africa in the 1950s[8], and melarsoprol treatment failure is a current and increasing problem[9].

We used genome-scale tetracycline-inducible RNA interference (RNAi) library screens in *T. brucei* to identify the genes that contribute to drug action. In these screens, replicating cells only persist in an otherwise toxic environment if knockdown confers a selective advantage (Fig. 1a); note that knockdown is not expected to identify drug targets. The RNAi library consists of ~750,000 clones, each transformed with one RNAi construct, and represents >99% of the approximately 7,500 non-redundant *T. brucei* gene set. Because each gene is identified by an average of approximately five different RNAi sequences, true leads can be identified with high confidence and potential off-target false leads can be minimized (see Supplementary Methods). Screens were performed using all current HAT drugs and each yielded a population of cells displaying an inducible drug resistance phenotype after eight or fourteen days of selection (Fig. 1b and Supplementary Fig. 1). Genomic DNA from these cells was subjected to RIT-seq[10] to create profiles of RNAi targets associated with increased resistance and to identify the genes that contribute to drug susceptibility. Genome-wide association maps show read density for 7,435 *T. brucei* genes (Fig. 1c). We defined genes with 'primary signatures' as those associated with two or more independent RIT-seq tags, each with a read density of >99; the screens yielded 55 of these signatures (Fig. 1c; see Supplementary Methods and Supplementary Data 1. Previous work linked the P2 adenosine transporter 1 (AT1) to melarsoprol uptake[4,11–13], an amino acid transporter family member (AAT6) to eflornithine uptake[5,13,14] and a nitroreductase (NTR) to nifurtimox activation[6,14]. Each of these genes is identified on the appropriate genome-wide association map (Fig. 1c), providing validation for our screens and indicating excellent genome-scale coverage in the RNAi library. Selected read-density signatures that establish new genetic links to drug susceptibility are shown in Fig. 1d.

The known eflornithine transporter is the only primary signature from the eflornithine screen. By contrast, the suramin screen revealed 28 genes associated with primary signatures (Fig. 1c and Supplementary Data 1). Suramin, which has been used for HAT therapy since the 1920s[15], is a colourless sulphated naphthylamine related to trypan red. Because this drug has a strong negative charge, it cannot cross lipid membranes by passive diffusion. Genes that are linked to the action of suramin encode ISG75, the function of which is unknown[16], four lysosomal proteins (the cathepsin L (CatL) and CBP1 peptidases, p67 and Golgi/lysosomal protein 1 (GLP1)), all four subunits of the adaptin complex (AP1), which are involved in endosomal, clathrin-mediated trafficking, and multiple spermidine and *N*-acetylglucosamine biosynthetic enzymes (Supplementary Fig. 2 and Supplementary Data 1).

Eight of these genes were selected for further analysis. We assembled multiple independent inducible RNAi strains for each gene and confirmed that knockdown (Fig. 2a and Supplementary Fig. 3) increased suramin resistance in every case (Fig. 2b and Supplementary Fig. 4). We then determined subcellular localization for the putative major facilitator superfamily transporter (MFST); the tandem of three closely

[1]London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK. [2]The Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK. [3]Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK. †Present address: Oxford Nanopore Technologies, 4 Robert Robinson Avenue, Oxford, OX4 4GA, UK.
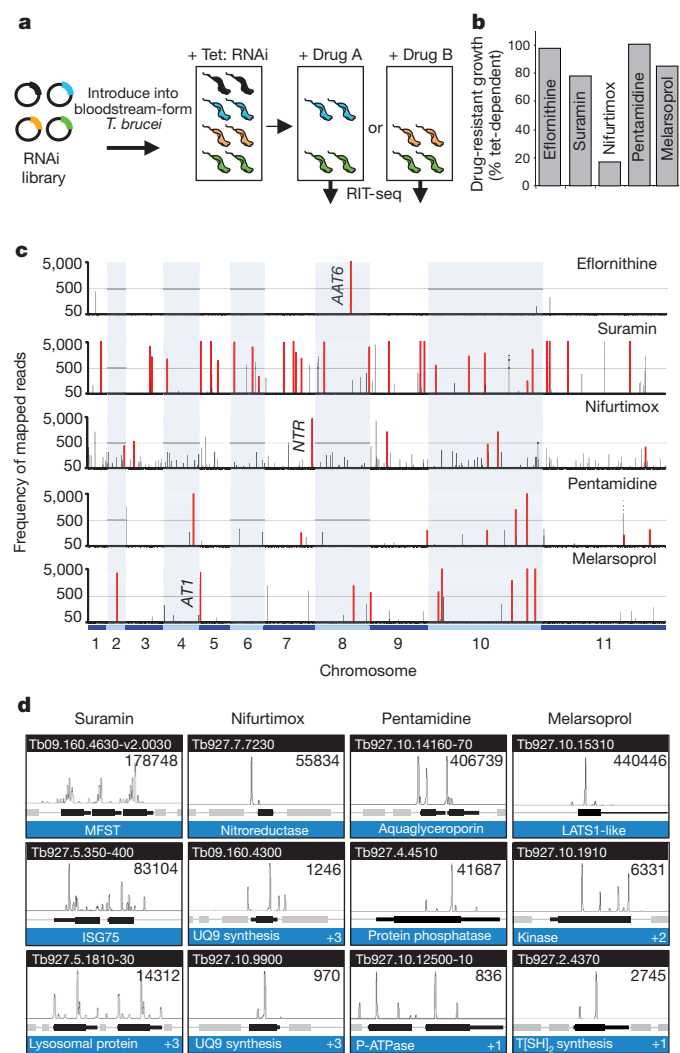
Figure 1 | Identification of drug efficacy determinants in *T. brucei*. a, A schematic showing the RNAi library screening approach. The expected outcomes are given for RNAi targets that fail to affect drug resistance (black), increase resistance to drug A (blue), drug B (orange) or both (green). **b**, Each screen yielded a population displaying tetracycline (Tet)-inducible (RNAi-dependent) drug-resistance (see Supplementary Fig. 1). The plot indicates the proportion of the resistance phenotype that is tetracycline inducible. **c**, Genome-wide RIT-seq profiles. Each map represents a non-redundant set of 7,435 protein-coding sequences. Red bars represent 'primary' read-density signatures. Black bars represent all other signatures of >50 reads (see Supplementary Data 1). All three expected 'hits', *AAT6*, *AT1* and *NTR*, are indicated. **d**, Selected signatures. Each peak represents a unique RIT-seq tag. '+', numbers of additional genes identified in each category. See Supplementary Fig. 2 for details and additional signatures.

related *MFST* genes gave the strongest read-density signature in the suramin screen and the greatest half-maximum effective concentration (EC_{50}) increase (> tenfold) following knockdown (Fig. 2b). In contrast to a putative ubiquitin hydrolase (UBH1) identified by the screen, MFST and a member of the endomembrane EMP70 family partitioned into the *T. brucei* membrane fraction, as expected (Fig. 2c), and MFST localized to the lysosome as did the major lysosomal type I membrane glycoprotein, p67 (ref. 17), which was also identified in the screen (Fig. 2d). Because ISG75 trafficking is ubiquitin dependent[18], we investigated whether UBH1 influenced ISG75 expression. UBH1 knockdown reduced ISG75 but not ISG65 expression (Fig. 2e), suggesting that de-ubiquitination by UBH1 specifically affects ISG75 copy number; clearly this mimics the direct effect of RNAi against ISG75. A vacuolar protein sorting factor, Vps5, which positively controls ISG75 expression[19], and a second putative ubiquitin hydrolase, were
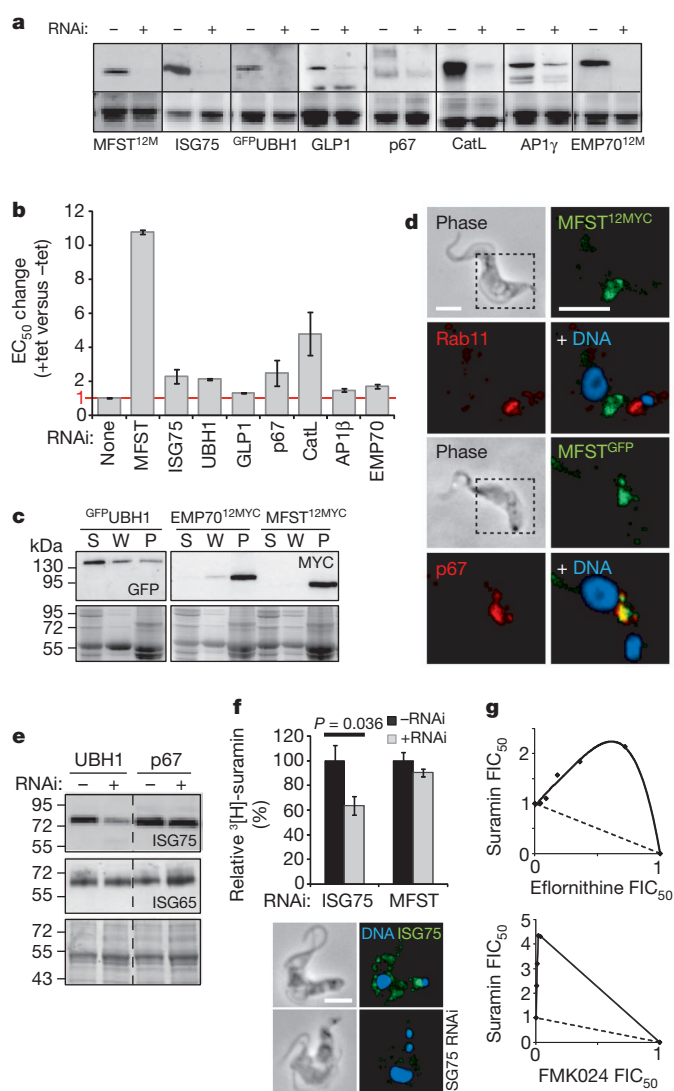
Figure 2 | A network of proteins link ISG75, endocytosis and lysosomal functions to suramin action. a, Western blots demonstrate knockdown; Coomassie stains serve as loading controls. Tags, green fluorescent protein (GFP) and 12×MYC epitope (12M). See Supplementary Fig. 3 for growth curves. **b**, Endosomal and lysosomal factors and ISG75 contribute to suramin action. Error bars, s.d. from independent RNAi strains (see Supplementary Fig. 4). **c**, MFST and EMP70 are membrane associated. The western blots show supernatant (S), wash (W) and pellet (P; membrane fraction). **d**, MFST co-localizes with lysosomal protein, p67, but not recycling endosomes (Rab11). Dashed boxes, areas magnified in fluorescent images. **e**, Knockdown of UBH1 specifically decreases ISG75 expression. **f**, ISG75 mediates suramin binding. Error bars, s.d. from duplicate experiments. *P* value from Student's *t*-test. ISG75 knockdown is shown. Scale bar, 5 μm. **g**, The CatL–CatB, and ODC inhibitors FMK024 and eflornithine, respectively, antagonize suramin action. Isobolograms showing 50% fractional inhibitory concentrations (FICs). The solid lines indicate antagonism. The dashed lines indicate expected outcomes for no interaction.

also identified by the screen (see Supplementary Fig. 2 and Supplementary Data 1), suggesting that ISG75 copy number is highly connected to suramin resistance. To investigate whether ISG75 contributes to suramin binding, we performed whole-cell binding assays using ³[H]-labelled suramin. Cells that were depleted for ISG75 displayed significantly and specifically reduced suramin binding (Fig. 2f).

We observed a greater than fourfold increase in EC_{50} after knockdown of the CatL-like protease known as brucipain, another abundant lysosomal protein[20], and an orthogonal assay using a dual-specificity CatL–CatB inhibitor revealed inhibitor antagonism (Fig. 2g), indicating that protease activity enhances suramin toxicity. Taken together, the

results demonstrate a central role for lysosomal functions in suramin action. As four enzymes that are involved in spermidine biosynthesis, including ornithine decarboxylase (ODC), were linked to suramin action (Supplementary Data 1), we used eflornithine to specifically inhibit ODC, which again revealed inhibitor antagonism (Fig. 2g; Supplementary Table 1). Thus, ODC activity enhances suramin toxicity, probably through spermidine biosynthesis. Suramin endocytosis[21] and intralysosomal accumulation[22] have previously been demonstrated in *T. brucei* and an acquired suramin resistance phenotype was stable in bloodstream stage *T. brucei* but was not expressed in the insect stage[23]. The RIT-seq profile reported here, bloodstream-stage-specific expression of ISG75[16] and strong downregulation of endocytic and lysosomal activities in the insect stage[24], are all consistent with stage-specific, intralysosomal accumulation of suramin.

Work with dyes and arsenicals revealed the first examples of resistance to chemotherapy a century ago and, based on cross-resistance, it was deduced that there are shared mechanisms contributing to the action of certain 'parasitotropic' compounds[1]. Among current HAT therapies, cross-resistance has been documented only for melarsoprol and pentamidine[9], but our understanding of the mechanism remains incomplete. Both drugs enter trypanosomes through the P2 AT1 but additional, dual-specificity transporters are predicted[9]. To identify cross-resistance mechanisms, we analysed all pair-wise comparisons among our screens (Fig. 3a). A single robust signature emerged, implicating two closely related aquaglyceroporins (AQPs)[25] in melarsoprol and pentamidine cross-resistance. To directly test the role of the AQPs, we generated a strain that was deficient in *aqp2* and *aqp3* (*aqp2/aqp3*-null strain) (Fig. 3b). The EC_{50} was increased more than 2-fold and 15-fold for melarsoprol and pentamidine, respectively, in *aqp2/aqp3*-null cells compared to wild-type cells (Fig. 3c). Our favoured hypothesis involves regulation of dual-specificity transporters by AQPs.

The nifurtimox, pentamidine and melarsoprol screens yielded eight, nine and nine genes associated with primary signatures, respectively. The major primary signature in the nifurtimox profile identified the mitochondrial, flavin-dependent nitroreductase that activates this class of nitro pro-drugs[6]. We also identified the putative flavokinase that converts riboflavin to FMN, an essential nitroreductase cofactor[6]. Four additional signatures identified genes that encode proteins linked to ubiquinone biosynthesis (Supplementary Fig. 2 and Supplementary Data 1), in support of the hypothesis that nitroreductase, like NADH dehydrogenases, transfers electrons from NADH to ubiquinone to generate ubiquinol[6]. We assembled RNAi strains for one of these factors and demonstrated that knockdown increased the EC_{50} for nifurtimox by approximately 1.5-fold (Supplementary Fig. 5). Thus, six gene signatures support a dominant role for nitroreductase in nifurtimox activation and suggest that this is dependent upon the availability of the FMN cofactor and the natural substrate.

Pentamidine is an aromatic diamidine, a nucleic acid binding drug that accumulates to millimolar concentrations and collapses trypanosome mitochondrial membrane potential[26]. Two primary signatures from the pentamidine screen identify genes encoding P-type ATPases (Supplementary Fig. 2 and Supplementary Data 1), and one of these represents the plasma membrane H^+-ATPases, HA1, HA2 and HA3 (ref. 27). We assembled RNAi strains for these ATPases and demonstrated that knockdown increased the EC_{50} for pentamidine more than eightfold (Supplementary Fig. 5), suggesting that an HA1–3 dependent proton motive force is required to drive pentamidine uptake. We used a similar approach to demonstrate a greater than twofold increase in the EC_{50} for pentamidine following knockdown of a putative protein phosphatase (Supplementary Fig. 5).

Melarsoprol acts primarily by forming a stable adduct with trypanothione, known as Mel T[28], but whether this adduct reduces or increases toxicity has remained unclear. The melarsoprol screen identified a link to trypanothione synthase and trypanothione reductase (Supplementary Fig. 2 and Supplementary Data 1), suggesting that the Mel T adduct is toxic. Three other primary signatures identified an
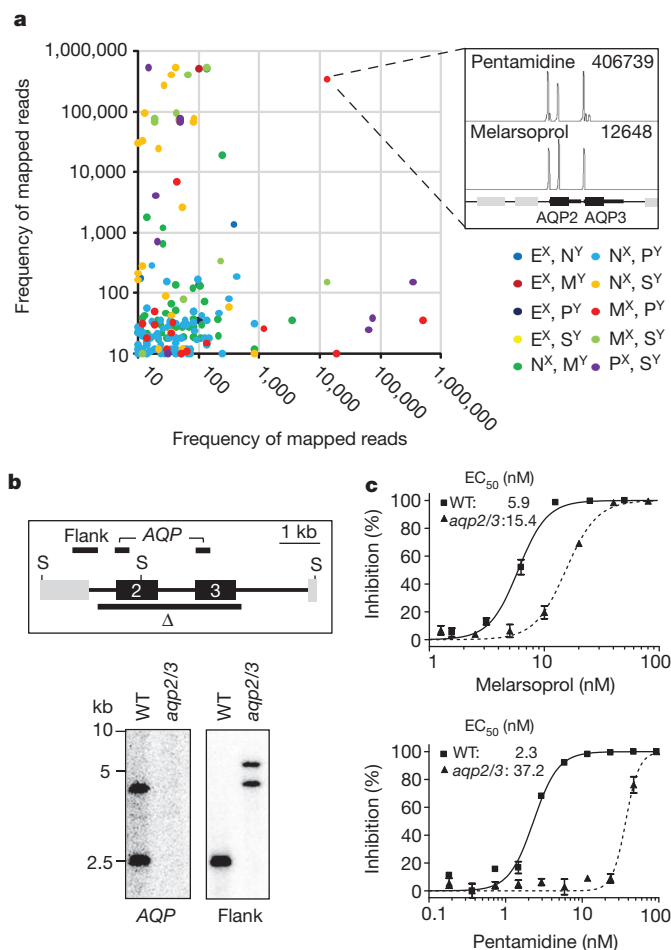

Figure 3 | *aqp2/aqp3*-null cells are melarsoprol, pentamidine cross-resistant. a, Analysis of read density for all (74,350) possible pair-wise comparisons of a non-redundant *T. brucei* gene set. E, eflornithine; M, melarsoprol; N, nifurtimox; P, pentamidine; S, suramin; X and Y, axes representing each data set. The box on the right shows the read-density signatures for this locus (Tb927.10.14160-70). **b**, *AQP2* and *AQP3* knockout was confirmed by Southern blot analysis. Δ, the region deleted; S, *SacII*; WT, wild type. Bars indicate probes. **c**, EC_{50} analysis indicates melarsoprol, pentamidine cross-resistance in *aqp2/aqp3*-null cells. Error bars, s.d. from triplicate assays and independent null strains.

over-representation ($P = 2.3 \times 10^{-9}$, χ^2 test) of putative protein kinases (Supplementary Fig. 2 and Supplementary Data 1), and another signature identified a gene encoding a highly phosphorylated protein related to the amino-terminal segment of the large tumour suppressor, LATS1 (see Supplementary Fig. 2a). We used independent strains to confirm that LATS1-like knockdown increased the EC_{50} for melarsoprol by approximately 1.5-fold (Supplementary Fig. 5). On the basis of these signatures, we suggest a role for a signalling cascade in melarsoprol susceptibility. Our findings are summarized in Fig. 4. In particular, we propose that suramin uptake occurs through ISG75-mediated endocytosis (Fig. 4a). Metabolic pathways that contribute to suramin or nifurtimox action are detailed in Fig. 4b.

All but one of the current HAT drugs was developed in the absence of an understanding of the chemical–biological relationships underlying toxicity or selectivity. Our RIT-seq profiles revealed more than 50 *T. brucei* genes that enhance drug susceptibility, unearthing interactions that are largely inaccessible using other approaches. Notably, the knockdown approach and the sensitivity of RIT-seq allow access to essential proteins, complexes and pathways such as H^+-ATPase, the adaptin complex and spermidine biosynthesis. Our results also show the utility of drugs as molecular probes for functional networks. In particular, the findings highlight factors that contribute to drug
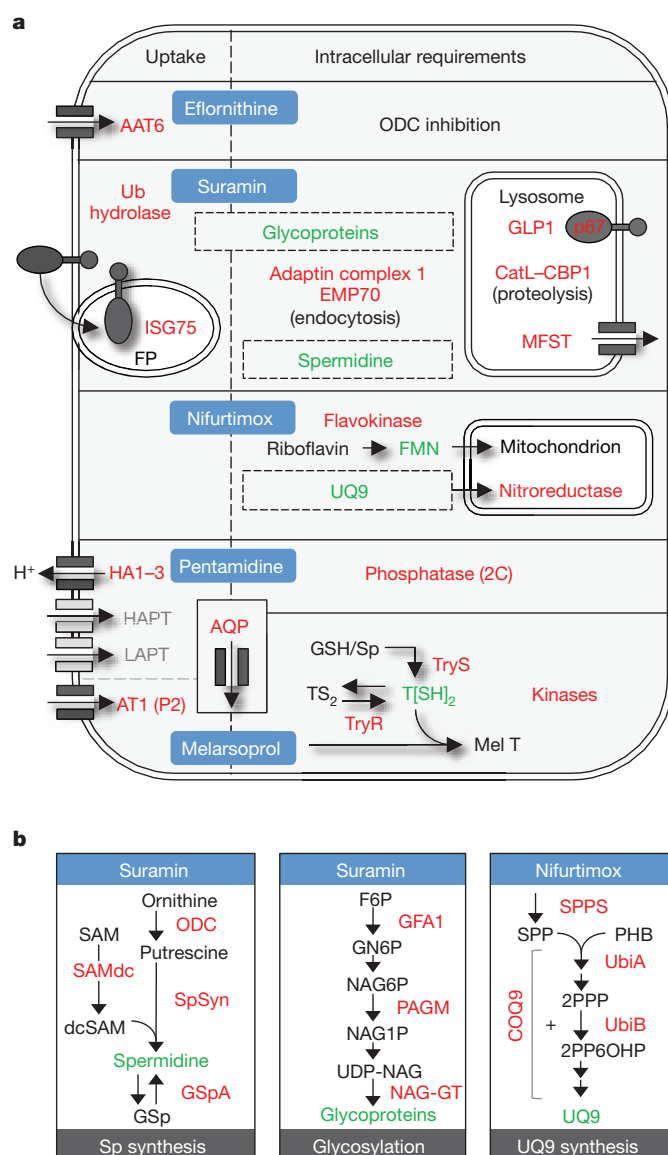
Figure 4 | Determinants of drug efficacy in African trypanosomes.
a, b, Proteins (red) and metabolites (green) that are linked to drug action. a, A schematic summarizing the findings from the RIT-seq screens. In the case of suramin, we propose that ISG75 binds the drug at the cell surface. ISG75 trafficking then delivers the complex, through the flagellar pocket (FP), to the endosomal system, leading to accumulation in the lysosome where the drug is liberated by proteases. The MFST may deliver the drug to the cytosol. HAPT, high-affinity pentamidine transporter; LAPT, low-affinity pentamidine transporter; TS_2, oxidised trypanothione; $T[SH]_2$, reduced trypanothione; UQ9, ubiquinone 9. b, Biosynthetic pathways that are linked to drug action. See Supplementary Data 1 for definitions and further details.

accumulation or the generation of toxic metabolites, features that could be exploited to deliver or generate novel toxins. Additionally, absence or loss of function could explain innate or acquired resistance; suramin resistance or melarsoprol and pentamidine cross-resistance may be due to reduced MFST or AQP expression, respectively (for examples, see Supplementary Fig. 6). These advances in our understanding of drug–trypanosome interactions will facilitate rational approaches to the design of more efficacious and durable therapies, and will be useful for monitoring the emergence and spread of resistance.

METHODS SUMMARY

Assembly of the bloodstream-form *T. brucei* RNAi library and RIT-seq were reported previously[10]. Briefly, a tetracycline-inducible RNAi plasmid library,

containing randomly sheared genomic fragments (with a mean fragment size of ~600 bp) under the control of head-to-head, tetracycline-inducible phage T7 promoters[29], was targeted to a single genomic locus that had been validated for robust expression[30]. The long double-stranded RNAs (dsRNAs) that were generated in the presence of tetracycline are processed to produce a pool of short interfering RNAs that programme the endogenous RNAi machinery to mediate sequence-specific destruction of the cognate messenger RNA. For this study, the library was grown under inducing conditions with drug selection, and genomic DNA was isolated from surviving populations. For RIT-seq profiling, adaptor-ligated sequencing libraries were prepared from each genomic DNA sample and used to amplify DNA fragments containing RNAi cassette-insert junctions in semi-specific PCR reactions; one primer was specific for the RNAi vector and the other for the Illumina adaptor. Size-selected DNA was sequenced with 76 cycle runs on an Illumina GAII. Sequencing reads containing a nine-base RNAi cassette-insert junction sequence were then mapped to the *T. brucei* reference genome. In cases in which loss of function increases drug tolerance, RNAi-target sequence representation is increased relative to the otherwise susceptible population, revealing 'hot spots'. Thus, RNAi target fragments serve as templates for the production of dsRNA and also provide unique sequence identifiers for each clonal population.

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

1. Ehrlich, P. Address in pathology, on chemiotherapy: delivered before the seventeenth international congress of medicine. *BMJ* **2**, 353–359 (1913).
2. Williamson, J. in *The African Trypanosomiases* (ed. Mulligan, H. W.) (Allen and Unwin, 1970).
3. Fairlamb, A. H. Chemotherapy of human African trypanosomiasis: current and future prospects. *Trends Parasitol.* **19**, 488–494 (2003).
4. Mäser, P., Sutterlin, C., Kralli, A. & Kaminsky, R. A nucleoside transporter from *Trypanosoma brucei* involved in drug resistance. *Science* **285**, 242–244 (1999).
5. Vincent, I. M. *et al.* A molecular mechanism for eflornithine resistance in African trypanosomes. *PLoS Pathog.* **6**, e1001204 (2010).
6. Wilkinson, S. R., Taylor, M. C., Horn, D., Kelly, J. M. & Cheeseman, I. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. *Proc. Natl Acad. Sci. USA* **105**, 5022–5027 (2008).
7. Fèvre, E. M., Wissmann, B. V., Welburn, S. C. & Lutumba, P. The burden of human African trypanosomiasis. *PLoS Negl. Trop. Dis.* **2**, e333 (2008).
8. Pépin, J. & Milord, F. The treatment of human African trypanosomiasis. *Adv. Parasitol.* **33**, 1–47 (1994).
9. de Koning, H. P. Ever-increasing complexities of diamidine and arsenical crossresistance in African trypanosomes. *Trends Parasitol.* **24**, 345–349 (2008).
10. Alsford, S. *et al.* High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res.* **21**, 915–924 (2011).
11. Carter, N. S. & Fairlamb, A. H. Arsenical-resistant trypanosomes lack an unusual adenosine transporter. *Nature* **361**, 173–176 (1993).
12. Matovu, E. *et al.* Mechanisms of arsenical and diamidine uptake and resistance in *Trypanosoma brucei*. *Eukaryot. Cell* **2**, 1003–1008 (2003).
13. Schumann Burkard, G., Jutzi, P. & Roditi, I. Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. *Mol. Biochem. Parasitol.* **175**, 91–94 (2011).
14. Baker, N., Alsford, S. & Horn, D. Genome-wide RNAi screens in African trypanosomes identify the nifurtimox activator NTR and the eflornithine transporter AAT6. *Mol. Biochem. Parasitol.* **176**, 55–57 (2011).
15. Steverding, D. The development of drugs for treatment of sleeping sickness: a historical review. *Parasit. Vectors* **3**, 15 (2010).
16. Overath, P., Chaudhri, M., Steverding, D. & Ziegelbauer, K. Invariant surface proteins in bloodstream forms of *Trypanosoma brucei*. *Parasitol. Today* **10**, 53–58 (1994).
17. Peck, R. F. *et al.* The LAMP-like protein p67 plays an essential role in the lysosome of African trypanosomes. *Mol. Microbiol.* **68**, 933–946 (2008).
18. Leung, K. F., Riley, F. S., Carrington, M. & Field, M. C. Ubiquitylation and developmental regulation of invariant surface protein expression in trypanosomes. *Eukaryot. Cell* **10**, 916–931 (2011).
19. Koumandou, V. L. *et al.* Evolutionary reconstruction of the retromer complex and its function in *Trypanosoma brucei*. *J. Cell Sci.* **124**, 1496–1509 (2011).
20. Caffrey, C. R. *et al.* Active site mapping, biochemical properties and subcellular localization of rhodesain, the major cysteine protease of *Trypanosoma brucei rhodesiense*. *Mol. Biochem. Parasitol.* **118**, 61–73 (2001).
21. Fairlamb, A. H. & Bowman, I. B. Uptake of the trypanocidal drug suramin by bloodstream forms of *Trypanosoma brucei* and its effect on respiration and growth rate *in vivo*. *Mol. Biochem. Parasitol.* **1**, 315–333 (1980).
22. Vansterkenburg, E. L. *et al.* The uptake of the trypanocidal drug suramin in combination with low-density lipoproteins by *Trypanosoma brucei* and its possible mode of action. *Acta Trop.* **54**, 237–250 (1993).

23. Scott, A. G., Tait, A. & Turner, C. M. Characterisation of cloned lines of *Trypanosoma brucei* expressing stable resistance to MelCy and suramin. *Acta Trop.* **60,** 251–262 (1996).

24. Natesan, S. K., Peacock, L., Matthews, K., Gibson, W. & Field, M. C. Activation of endocytosis as an adaptation to the mammalian host by trypanosomes. *Eukaryot. Cell* **6,** 2029–2037 (2007).

25. Uzcategui, N. L. *et al.* Cloning, heterologous expression, and characterization of three aquaglyceroporins from *Trypanosoma brucei. J. Biol. Chem.* **279,** 42669–42676 (2004).

26. Lanteri, C. A., Tidwell, R. R. & Meshnick, S. R. The mitochondrion is a site of trypanocidal action of the aromatic diamidine DB75 in bloodstream forms of *Trypanosoma brucei. Antimicrob. Agents Chemother.* **52,** 875–882 (2008).

27. Luo, S., Fang, J. & Docampo, R. Molecular characterization of *Trypanosoma brucei* P-type H$^+$-ATPases. *J. Biol. Chem.* **281,** 21963–21973 (2006).

28. Fairlamb, A. H., Henderson, G. B. & Cerami, A. Trypanothione is the primary target for arsenical drugs against African trypanosomes. *Proc. Natl Acad. Sci. USA* **86,** 2607–2611 (1989).

29. Morris, J. C., Wang, Z., Drew, M. E. & Englund, P. T. Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library. *EMBO J.* **21,** 4429–4438 (2002).

30. Alsford, S., Kawahara, T., Glover, L. & Horn, D. Tagging a *T. brucei RRNA* locus improves stable transfection efficiency and circumvents inducible expression position effects. *Mol. Biochem. Parasitol.* **144,** 142–148 (2005).

Author Contributions S.A., N.B., L.G. and K.F.L. carried out the *T. brucei* manipulation and analyses, S.E., A.S.-F. and D.J.T. carried out the Illumina sequencing and mapping, D.H. coordinated the study and S.A., M.C.F., M.B. and D.H. wrote the paper.

Author Information Sequence data from this study have been submitted to the European Nucleotide Archive at http://www.ebi.ac.uk/ena under accession number ERA071064. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.H. (david.horn@lshtm.ac.uk).

METHODS

T. brucei growth and drug selection. The bloodstream-form _T. brucei_ MITat 1.2 clone 221a RNAi library[10] was derived using the randomly sheared genomic fragment (with mean fragment length ~600 bp) RNAi plasmid library[29]. The _T. brucei_ RNAi library and 2T1 cells were maintained as described[29,30]. For selective screens, the RNAi library, maintained throughout at $>5 \times 10^6$ cells, was induced with tetracycline (1 μg ml^{-1}) for 24 h and then grown in medium containing tetracycline, plus each HAT drug at $0.5 \times EC_{50}$ to $3.5 \times EC_{50}$ (Supplementary Table 1 and Supplementary Fig. 1). All drug stocks were in dimethylsulphoxide.

RIT-seq. Selected populations from each screen were assessed for tetracycline-dependent drug resistance. The RNAi target fragments provide unique identifiers for each clone in the population. As a quality-control step, PCR amplification, agarose gel fractionation and Sanger sequencing of the eluted products were performed as described[14], and followed with RIT-seq analysis[10]. All nine genes that were identified by Sanger sequencing were associated with high-density Illumina read-counts (13,000 to 528,000; see Supplementary Data 1a). Briefly, we ran 76-cycle sequencing on an Illumina GAII; this generates sequence tags derived from the ends of the RNAi target fragments. Only sequences containing a terminal RNAi-vector junction sequence (GCCTCGCGA) were mapped to the _T. brucei_ 927 reference genome[31] using the SSAHA sequence alignment algorithm[32]. After mapping, for each protein coding sequence (CDS) in each experiment, we obtained a count of reads mapping; all genes associated with >9 reads are detailed in Supplementary Data 1b. We also browsed all read-density plots in Artemis[33] for signatures that fell outside of CDSs to generate the full non-redundant 'hit list' detailed in Supplementary Data 1a.

Read-density signatures. Genome coverage in the current RNAi library represents >99% of all genes, with 5 RNAi targets per gene on average[10]; shorter genes are expected to be represented by fewer RNAi targets. Our screens yielded 5–59 genes (0.07–0.8%) with a $^{>99}$RIT-seq tag (a tag with a read density of >99; the eflornithine screen yielded 5, the suramin screen 59 the nifurtimox screen 54, the pentamidine screen 17 and the melarsoprol screen 19). In each screen, at least one gene was associated with a $^{>50,000}$RIT-seq tag (Supplementary Data 1a). From this set, we derived 55 genes with 'primary signatures', those associated with two or more $^{>99}$RIT-seq tags. If these tags were randomly distributed, we would expect a single primary signature from 300 screens using eflornithine or from two screens using suramin, assigning a high degree of confidence to the vast majority of observed primary signatures (Supplementary Data 1a). The nifurtimox output is unusual compared to the other outputs and may reflect drug-mediated mutagenesis[34]; for example, inactivating mutations within _NTR_ may prolong the survival of clones carrying unrelated RNAi targets. However, even limited tetracycline-regulated drug resistance (Fig. 1b) and a high number of sequence tags in the nifurtimox screening profile (Supplementary Data 1 and Fig. 1c) had little impact on primary signature confidence. Many of the 130 genes that are associated with 'secondary signatures' in Supplementary Data 1a may also reflect mechanisms of drug action, but here we only considered seven of these genes that were linked to a common function with a primary hit (Supplementary Fig. 2). We observe that, on average, 3.5 tags per gene are associated with the 24 primary, single copy genes that are shown in Supplementary Fig. 2. Minimal library propagation could explain a modest reduction in coverage but we suggest that reduced coverage in the current RIT-seq outputs is primarily explained by major fitness defects following knockdown.

Plasmid construction and strain assembly. The _AQP_ locus was disrupted by replacement of a 4,772-bp (_AQP2_ and _AQP3_) fragment with _NPT_ and _BLA_ selectable markers (the _T. brucei_ genome is diploid). Gene-specific RNAi fragments of 400–600 bp or 200 bp, to facilitate moderate knockdown in the case of the known essential gene p67 (ref. 17), were amplified using PCR primers designed using RNAit[35] and cloned into pRPaiSL for the generation of stem-loop, 'hairpin' dsRNA as the trigger for RNAi[36]. We used a long, 400–600-bp RNAi target fragment for CatL because RNAi previously produced no growth defect[37]. However, cells retained 35% CatL activity in that study[37], probably explaining why we see a major growth defect when expressing a more potent stem-loop

dsRNA (Supplementary Fig. 3). For epitope tagging at native loci, C-terminal fragments, or an N-terminal fragment (UBH1), were amplified and cloned in pNATx^{TAG} and pNAT^{TAG}x (ref. 36), respectively. Constructs were introduced into 2T1 cells as described[30]. Full oligonucleotide details are available on request.

Strain analysis. Cumulative growth curves were generated from cultures seeded at 10^5 cells ml^{-1}, counted on a haemocytometer and diluted back to 10^5 cells ml^{-1} as necessary. For EC$_{50}$ assays, RNAi strains were pre-induced for 72 h in 1 μg ml^{-1} tetracycline, except CatL and AP1β, which were pre-induced for 24 h at 2.5 and 1 ng ml^{-1}, respectively. Isobolograms were generated using a checkerboard assay as described[38]; FMK024 (_N_-morpholineurea-phenylalanyl-homophenylalanylfluoromethyl ketone; Sigma) is an irreversible, dual-specificity inhibitor of CatL and CatB. All EC$_{50}$ assays were carried out using alamarBlue as described[14,39]. Southern blotting was carried out according to standard procedures[40]. Subcellular fractionation by hypotonic lysis was carried out as described[41]. All protein samples were stored in the presence of a protease inhibitor cocktail (Roche) and were not boiled. Whole-cell lysates and hypotonic lysis fractions were separated by SDS–PAGE using standard protocols[40]. Immunofluorescence was carried out as previously described[10]. We used specific antisera to detect ISG75 (ref. 42), p67 (ref. 43), CatL[17], GLP1 (ref. 44) and AP1γ (ref. 45), and anti-MYC or anti-GFP antisera were used to detect tagged versions of MFST, UBH1 and EMP70. To assess suramin binding, cells were collected at mid-log phase and resuspended at 10^7 ml^{-1} in 35 nM ^{3}[H]-suramin (Hartmann Analytic; pre-incubated for 16 h in complete HMI11) at 37 °C. Cells were washed in ice-cold PBS, resuspended in 100 μl Optiphase Supermix scintillant (Perkin Elmer) and ^{3}[H]-suramin incorporation quantified using a 1450 Microbeta scintillation counter (Perkin Elmer).

31. Berriman, M. _et al._ The genome of the African trypanosome _Trypanosoma brucei. Science_ **309,** 416–422 (2005).
32. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. _Genome Res._ **11,** 1725–1729 (2001).
33. Carver, T. _et al._ Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. _Bioinformatics_ **24,** 2672–2676 (2008).
34. Buschini, A. _et al._ Genotoxicity reevaluation of three commercial nitroheterocyclic drugs: nifurtimox, benznidazole, and metronidazole. _J. Parasitol. Res._ **2009,** 463575 (2009).
35. Redmond, S., Vadivelu, J. & Field, M. C. RNAit: an automated web-based tool for the selection of RNAi targets in _Trypanosoma brucei. Mol. Biochem. Parasitol._ **128,** 115–118 (2003).
36. Alsford, S. & Horn, D. Single-locus targeting constructs for reliable regulated RNAi and transgene expression in _Trypanosoma brucei. Mol. Biochem. Parasitol._ **161,** 76–79 (2008).
37. Mackey, Z. B., O'Brien, T. C., Greenbaum, D. C., Blank, R. B. & McKerrow, J. H. A cathepsin B-like protease is required for host protein degradation in _Trypanosoma brucei. J. Biol. Chem._ **279,** 48426–48433 (2004).
38. Singh, P. K., Tack, B. F., McCray, P. B. Jr & Welsh, M. J. Synergistic and additive killing by antimicrobial factors found in human airway surface liquid. _Am. J. Physiol. Lung Cell. Mol. Physiol._ **279,** L799–L805 (2000).
39. Räz, B., Iten, M., Grether-Buhler, Y., Kaminsky, R. & Brun, R. The Alamar Blue assay to determine drug sensitivity of African trypanosomes (_T.b. rhodesiense_ and _T.b. gambiense_) _in vitro. Acta Trop._ **68,** 139–147 (1997).
40. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (eds) _Current Protocols in Molecular Biology_ (John Wiley and Sons, 1998).
41. Leung, K. F., Dacks, J. B. & Field, M. C. Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. _Traffic_ **9,** 1698–1716 (2008).
42. Ziegelbauer, K. & Overath, P. Organization of two invariant surface glycoproteins in the surface coat of _Trypanosoma brucei. Infect. Immun._ **61,** 4540–4545 (1993).
43. Kelley, R. J., Brickman, M. J. & Balber, A. E. Processing and transport of a lysosomal membrane glycoprotein is developmentally regulated in African trypanosomes. _Mol. Biochem. Parasitol._ **74,** 167–178 (1995).
44. Lingnau, A., Zufferey, R., Lingnau, M. & Russell, D. G. Characterization of tGLP-1, a Golgi and lysosome-associated, transmembrane glycoprotein of African trypanosomes. _J. Cell Sci._ **112,** 3061–3070 (1999).
45. Allen, C. L., Liao, D., Chung, W. L. & Field, M. C. Dileucine signal-dependent and AP-1-independent targeting of a lysosomal glycoprotein in _Trypanosoma brucei. Mol. Biochem. Parasitol._ **156,** 175–190 (2007).

LETTER

G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody

Tomoya Hino[1,2]†*, Takatoshi Arakawa[1,2]*, Hiroko Iwanari[3]*, Takami Yurugi-Kobayashi[1,2], Chiyo Ikeda-Suno[1,2], Yoshiko Nakada-Nakura[3,4], Osamu Kusano-Arai[3,5], Simone Weyand[1,6,7,8], Tatsuro Shimamura[1,2], Norimichi Nomura[1,2], Alexander D. Cameron[1,6,7,8], Takuya Kobayashi[1,2,9], Takao Hamakubo[3], So Iwata[1,2,6,7,8,10] & Takeshi Murata[1,2,10,11]

G-protein-coupled receptors are the largest class of cell-surface receptors, and these membrane proteins exist in equilibrium between inactive and active states[1–13]. Conformational changes induced by extracellular ligands binding to G-protein-coupled receptors result in a cellular response through the activation of G proteins. The A_{2A} adenosine receptor ($A_{2A}AR$) is responsible for regulating blood flow to the cardiac muscle and is important in the regulation of glutamate and dopamine release in the brain[14]. Here we report the raising of a mouse monoclonal antibody against human $A_{2A}AR$ that prevents agonist but not antagonist binding to the extracellular ligand-binding pocket, and describe the structure of $A_{2A}AR$ in complex with the antibody Fab fragment (Fab2838). This structure reveals that Fab2838 recognizes the intracellular surface of $A_{2A}AR$ and that its complementarity-determining region, CDR-H3, penetrates into the receptor. CDR-H3 is located in a similar position to the G-protein carboxy-terminal fragment in the active opsin structure[1] and to CDR-3 of the nanobody in the active β_2-adrenergic receptor structure[2], but locks $A_{2A}AR$ in an inactive conformation. These results suggest a new strategy to modulate the activity of G-protein-coupled receptors.

The structures of G-protein-coupled receptors (GPCRs) in an inactive conformation solved recently[3–12] greatly advance our understanding of the molecular signalling mechanisms of the receptors. The first details of GPCR activation were provided by the structure of bovine opsin in an active conformation complexed with a G-protein C-terminal peptide[1] (GαCT). Most recently, determination has been made of the crystal structures of β_2 adenosine receptor (β_2AR) in an active state with a camelid antibody fragment[2] (nanobody Nb80) and with a heterotrimeric G_s protein[13]. In these structures, the complementarity-determining region (CDR-3) of Nb80 and the C-terminal α-helix of α subunit ($G\alpha_s$) of G_s protein were located in the same pocket as was GαCT in the opsin structure. Nb80 and G_s protein change the conformational equilibrium of β_2AR toward the active state in a similar manner, thereby substantially increase their agonist affinities[2,13].

$A_{2A}AR$ is responsible for regulating blood flow to the cardiac muscle and is important in the regulation of glutamate and dopamine release in the brain[14]. Caffeine is a well-known antagonist of this receptor. Strong epidemiological evidence indicates that coffee drinkers have a lower risk of Parkinson's disease[15]. The structure of $A_{2A}AR$ has been reported[9,16] as a complex with both an antagonist (ZM241385) and an agonist (UK-432097). These structures reveal the molecular framework of the receptor; however, in both cases the intracellular loop 3

(ICL3), critical for G-protein binding, has been replaced by T4 lysozyme (T4L).

Here we report the crystal structure of $A_{2A}AR$ with complete ICL3 in complex with a mouse monoclonal-antibody Fab fragment, Fab2838. $A_{2A}AR$ was expressed in *Pichia pastoris* and the antibody was raised against the purified receptor with antagonist (ZM241385) bound using the conventional mouse hybridoma system combined with improved immunization and screening methods (Methods). Fab2838, a Fab fragment generated from one (IgG2838) of the obtained antibodies, completely inhibited binding of the agonist [³H]-5′-*N*-ethylcarboxamido adenosine ([³H]-NECA) but did not affect binding of the antagonist [³H]-ZM241385 (Fig. 1a, d and Supplementary Fig. 2). The results were confirmed by competition binding assays (Supplementary Discussion and Fig. 1). These findings suggest that Fab2838 induces an inactive conformation (that is, one to which agonist cannot bind) of the $A_{2A}AR$ ligand-binding pocket without blocking the ligand-binding site.

We crystallized $A_{2A}AR$ with Fab2838 in the presence of ZM241385 and solved the structure at a resolution of 2.7 Å (Supplementary Table 2). Because the occupancy of ZM241385 in the structure was low for unknown reasons, we repeated the experiments and obtained a higher occupancy structure at 3.1 Å (Supplementary Table 2 and Supplementary Figs 3 and 4). Except for the occupancy of the ligand, the two structures are almost identical (root mean squared deviation of Cα, 0.57 Å) (Supplementary Table 2). ZM241385 occupies the ligand-binding pocket on the extracellular side by making hydrophobic interactions with Phe 168[5.29] and Ile 274[7.39] and hydrogen bonds with Asn 253[6.55] as observed in the $A_{2A}AR$-T4L structure (Supplementary Fig. 4) (superscripts indicate residue numbers as per the Ballesteros–Weinstein scheme[17]). Although the overall structure of $A_{2A}AR$ in the $A_{2A}AR$–Fab2838 complex is similar to that of $A_{2A}AR$-T4L (Protein Data Bank code, 3EML; root mean squared deviation of Cα, 0.85 Å), there is a major difference around the intracellular portions of helices V and VI; these are connected by ICL3, which in $A_{2A}AR$-T4L is replaced with T4L (Supplementary Fig. 5). In our structure, ICL3 forms two regular helices—effectively continuations of helices V and VI, respectively—connected by a short turn (Supplementary Fig. 6a).

The $A_{2A}AR$–Fab2838 structure has a modified 'ionic lock' where Glu 228[6.30] (helix VI) and Arg 102[3.50] of the D/ERY motif (helix III) interact through a water molecule (W1; Fig. 2c, d). In the inactive bovine rhodopsin structure, the equivalent residues form a direct salt bridge[3] (Supplementary Fig. 7). Arg 102[3.50] of $A_{2A}AR$–Fab2838 forms salt bridges or hydrogen bonds with Asp 101[3.49] and Tyr 112 in ICL2
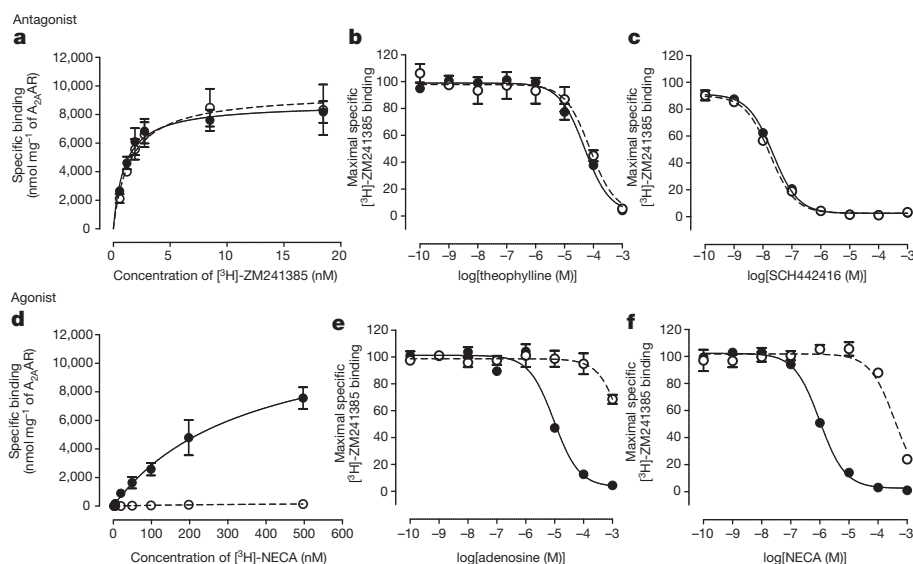
Figure 1 | Effect of Fab2838 on $A_{2A}AR$–ligand binding. a, Saturation binding curves for an antagonist [^{3}H]-ZM241385 binding to $A_{2A}AR$ with (open circles) or without (filled circles) Fab2838. **b, c,** Inhibition of [^{3}H]-ZM241385 binding by the antagonists theophylline (**b**) and SCH442416 (**c**) with (open circles) and without (filled circles) Fab2838. The binding of [^{3}H]-ZM241385 in the absence of a competitor was set at 100%. **d,** As in **a,** but for the agonist [^{3}H]-NECA. **e, f,** As in **c** and **d,** but for the agonists adenosine (**e**) and NECA (**f**). All data are the mean ± s.e.m. of three independent experiments performed in duplicate.

and with Thr $41^{2.39}$ as observed in the $A_{2A}AR$-T4L structure (Supplementary Fig. 5b). Because of the insertion of the water molecule, Glu $228^{6.30}$ shifts towards the cytoplasmic space, as compared with the equivalent residue in rhodopsin (Glu $247^{6.30}$), resulting in the formation of a salt bridge with Arg 220 in the short helical turn of ICL3. This

interaction may be important in the formation of the helical structure in ICL3. The ionic lock has not been observed in the crystal structures of other inactive GPCRs[6–11], including $A_{2A}AR$-T4L, except for the D3 dopamine receptor[12]. This may be because the ICL3 loops in the other structures were modified to stabilize the protein. While this paper was under review, the crystal structures of thermostabilized $A_{2A}AR$ mutants with native ICL3 were published[18,19]. The antagonist-bound inactive structures have the ionic lock[19]. Thus, the ionic lock of $A_{2A}AR$ seems to stabilize the inactive conformation of the protein, which is why the receptor has a low basal activity.

Fab2838 binds on the intracellular side of the receptor (Fig. 2a). CDR-H3 of Fab2838 is unusually long and penetrates a pocket formed by helices II, III, VI and VII (Fig. 2b). CDR-H3 interacts with the surrounding helices by forming six hydrogen bonds and eight van der Waals contacts (Fig. 2c, d). The most extensive interactions are with helix II (mainly through hydrogen bonds) and helix VI (mainly through van der Waals contacts). In addition, a hydrogen bond network including two water molecules is observed between CDR-H3 and helices III and VI (Fig. 2c, d). This hydrogen bond network together with the van der Waals interactions seem to stabilize the modified ionic lock interaction between Glu $228^{6.30}$ (helix VI) and Arg $102^{3.50}$ (helix III) discussed above. Other complementarity-determining regions further stabilize the $A_{2A}AR$–Fab2838 complex by forming 14 hydrogen bonds with helices VI and VIII and ICL1, ICL2 and ICL3 (Fig. 2b).

Figure 2 | Structure of the $A_{2A}AR$–Fab2838 complex. a, Overall structure viewed parallel to the membrane. $A_{2A}AR$ and the Fab light (Fab(L)) and heavy (Fab(H)) chains are shown in blue-grey, cyan and magenta, respectively. The three disulphide bonds in the extracellular loops (ECLs) are represented by yellow sticks. The bound antagonist ZM241385 in the ligand-binding pocket is shown as a space-filling model. The CDRs of Fab2838 are coloured as follows: CDR-H1, yellow; CDR-H2, orange; CDR-H3, red; CDR-L1, green; CDR-L2, purple; CDR-L3, marine blue. EXT, extracellular; IN, intracellular. **b,** Surface representation of the interface between $A_{2A}AR$ (top) and Fab2838 (bottom). Relative to **a,** $A_{2A}AR$ has been rotated 90° around a horizontal axis, whereas Fab2838 is shown in the same orientation. **c,** View of the interface between $A_{2A}AR$ (green residues) and CDR-H3 (orange residues). The main chain of $A_{2A}AR$ is shown as ribbon representation as in **a.** Red spheres show the positions of water molecules. Red dotted lines indicate hydrogen-bond interactions. **d,** Schematic representation of the interface between $A_{2A}AR$ and CDR-H3.

The extensive interactions explain the high affinity of Fab2838 (dissociation constant, $K_d = 4.4$ nM) (Supplementary Fig. 8).

The binding site of Fab2838 CDR-H3 in $A_{2A}AR$ is similar to those of Nb80 CDR-3 in β_2AR[2] and GαCT in opsin[1]. A critical difference is that Fab2838 stabilizes an inactive conformation whereas the others recognize active conformations of the receptors. These structures are compared in Fig. 3. In the opsin structure, GαCT, which forms a short α-helix, fits into a large pocket formed by helices II, III, V, VI and VII interacting with the Arg residue of the D/ERY motif in helix III (Fig. 3, left panels). CDR-3 of Nb80 in the β_2AR structure binds in a similar position to GαCT although CDR-3 forms a β-hairpin[1] (Fig. 3, middle panels). CDR-H3 of Fab2838 also forms a β-hairpin but induces a differently shaped binding pocket (Fig. 3c). In the β_2AR structure, CDR-3 of Nb80 is positioned between helices III and VI, whereas in the $A_{2A}AR$ structure CDR-H3 of Fab2838 is ~6 Å closer to helices II and VII (Fig. 3b and Supplementary Fig. 9). This allows the close association of helices III and VI and the formation of the modified ionic lock between Arg 102[3.50] in helix III and Glu 228[6.30] in helix VI, consequently stabilizing the inactive conformation. In the β_2AR/G_s-protein complex structure, the C-terminal α-helix (α5) of Gα_s also

binds in a similar position to CDR-H3[13] (Supplementary Fig. 10). The conformational changes of α5 together with the Gα_s amino-terminal region induced by the activated receptor has been proposed to result in a nucleotide exchange from GDP to GTP in Gα_s and to subsequent dissociation of the subunit from the receptor[20]. Thus, the binding pocket formed by helices II, III, VI and VII seems to be the key site for the signal transfer between GPCRs and G proteins.

A possible inactivation mechanism of $A_{2A}AR$ by Fab2838 is summarized as follows. Agonist binding induces large displacements of the intracellular ends of helices III, VI and VII[16,18], which are essential to form the G-protein binding pocket[13,20] (Supplementary Fig. 1). This indicates that the signal from the ligand-binding pocket is transferred through these helices and the conformations of the two pockets are strongly coupled. Our agonist- and antagonist-binding experiments indicate that this coupling also allows signal transfer in the reverse direction, from the G-protein-binding pocket to the ligand-binding pocket (Fig. 1). CDR-H3 of Fab2838 locks the positions of helices III, VI and VII from the cytoplasmic side, leading to an inactive conformation of the extracellular ligand-binding pocket to which agonists cannot bind, probably because of the rearrangement of the

Figure 3 | Comparison of the structures of the opsin–GαCT, β₂AR–Nb80 and A₂AR–Fab2838 complexes. Left, middle and right panels show the structures of an active form of opsin (green) in complex with GαCT (yellow), an active form of β₂AR (brown) bound agonist BI-167107 in complex with Nb80 CDR-3 (blue) and an inactive form of A₂AR (blue-grey) bound antagonist ZM241385 in complex with Fab2838 CDR-H3 (red). **a**, Views parallel to the membrane. Bound ligands are shown as stick models in β₂AR and A₂AR. The residues involved in the ionic lock formation are also shown. Nitrogen and oxygen atoms are coloured blue and red, respectively. **b**, Cytoplasmic views of the complexes. **c**, Surface representations of cytoplasmic surfaces of the receptors. Surfaces within 4 Å of GαCT, CDR-3 or CDR-H3 are coloured red.

side chains at the bottom of the ligand-binding pocket including Trp 246[6.48], which is the toggle switch for activation (Supplementary Figs 1 and 11). A similar conceptual model of β_2AR activation has been reported[21]. In the case of β-adrenergic receptors, the conformations of the ligand- and G-protein-binding pockets are less strongly coupled, as demonstrated in the structures of β_1AR–agonist complexes[22] and the β_2AR/irreversible-agonist complex[23]. This may be because the A_{2A}AR and β_1AR or β_2AR agonists interact with different helices in the binding pockets (Supplementary Discussion).

Antibody fragments (and nanobodies) such as Nb80 and Fab2838, which recognize conformational epitopes of GPCRs, have great potential for GPCR studies *in vitro* and *in vivo*. Although antibodies recognizing the intracellular surface are not suitable for direct therapeutic use, the CDR structures should provide useful information for the design of peptides or small-molecule compounds against their clearly defined pockets to control the activation states of GPCRs. The antibody fragments will also be useful tools to study ligand-binding kinetics of GPCRs because they can separate ligand binding from equilibrium shifts between different activation states of the receptors. Our approach based on the conventional mouse hybridoma system allows us to raise antibodies against various receptors in three to four months using standard laboratory equipment.

METHODS SUMMARY

Expression and purification. $A_{2A}AR^{N154Q}$ (residues 1–316) was expressed in *P. pastoris* as described previously[24] and purified as described in Methods.

Antibody generation. MRL/lpr mice were immunized with the purified A_{2A}AR with the antagonist ZM241385. Antibodies were raised to recognize conformational epitopes of A_{2A}AR using the conventional mouse hybridoma system[25] in combination with new screening methods as described in Methods. The Fab fragments were obtained by papain cleavage and purified by anion exchange column chromatography.

Crystallographic data collection and structure determination. Purified A_{2A}AR was mixed with the Fab fragment and the A_{2A}AR–Fab complex was purified twice by gel filtration chromatography. Crystals were grown by vapour diffusion under the conditions described in Methods. Diffraction data were collected from a single cryo-cooled crystal on beamline I24 at the Diamond Light Source, UK. The structures were solved by molecular replacement using the receptor from the A_{2A}AR-T4L structure (PDB code, 3EML) and an antibody Fab fragment structure (PDB code, 1P7K) as search models. Data collection and refinement statistics are summarized in Supplementary Table 2.

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

1. Scheerer, P. et al. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008).
2. Rasmussen, S. G. F. et al. Structure of a nanobody-stabilized active state of the β2 adrenoceptor. *Nature* **469**, 175–180 (2011).
3. Palczewski, K. et al. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289**, 739–745 (2000).
4. Shimamura, T. et al. Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. *J. Biol. Chem.* **283**, 17753–17756 (2008).
5. Murakami, M. & Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **453**, 363–367 (2008).
6. Warne, T. et al. Structure of a β1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008).
7. Rasmussen, S. G. F. et al. Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).
8. Cherezov, V. et al. High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).
9. Jaakola, V.-P. et al. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **322**, 1211–1217 (2008).
10. Wu, B. et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066–1071 (2010).
11. Shimamura, T. et al. Structure of the human histamine H1 receptor complex with doxepin. *Nature* **475**, 65–70 (2011).
12. Chien, E. Y. T. et al. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330**, 1091–1095 (2010).
13. Rasmussen, S. G. F. et al. Crystal structure of the β2 adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).
14. Fredholm, B. B., Chen, J. F., Masino, S. A. & Vaugeois, J. M. Actions of adenosine at its receptors in the CNS: insights from knockouts and drugs. *Annu. Rev. Pharmacol. Toxicol.* **45**, 385–412 (2005).
15. Müller, C. E. & Jacobson, K. A. Recent developments in adenosine receptor ligands and their potential as novel drugs. *Biochim. Biophys. Acta* **1808**, 1290–1308 (2011).
16. Xu, F. et al. Structure of an agonist-bound human A2A adenosine receptor. *Science* **332**, 322–327 (2011).
17. Ballesteros, J. A&. Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
18. Lebon, G. et al. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **474**, 521–525 (2011).
19. Doré, A. S. et al. Structure of the adenosine A2A receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **19**, 1283–1293 (2011).
20. Chung, K. Y. et al. Conformational changes in the G protein Gs induced by the β2 adrenergic receptor. *Nature* **477**, 611–615 (2011).
21. Yao, X. J. et al. The effect of ligand efficacy on the formation and stability of a GPCR-G protein complex. *Proc. Natl Acad. Sci. USA* **106**, 9501–9506 (2009).
22. Warne, T. et al. The structural basis for agonist and partial agonist action on a β1-adrenergic receptor. *Nature* **469**, 241–244 (2011).
23. Rosenbaum, D. M. et al. Structure and function of an irreversible agonist-β2 adrenoceptor complex. *Nature* **469**, 236–240 (2011).
24. Yurugi-Kobayashi, T. et al. Comparison of functional non-glycosylated GPCRs expression in Pichia pastoris. *Biochem. Biophys. Res. Commun.* **380**, 271–276 (2009).
25. Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975).

Author Contributions S.I. and T.M. designed the original research project. T.Y.-K. and T.K. established the A2AAR expression and purification protocols. T. Hino and C.I.-S. expressed, purified and characterized the receptor. H.I., Y.N.-N., O.K.-A. and T. Hamakubo performed the immunization, selection and isolation of antibodies. T. Hino, T.A. and C.I.-S. purified and characterized antibodies. N.N. sequenced antibodies. T. Hino, T.A. and T.S. purified and crystallized the receptor/Fab-fragment complex. S.W., A.D.C. and S.I. performed data collection. T. Hino solved and refined the structure. T. Hino, S.I. and T.M. wrote the manuscript and all authors provide editorial input. The project was managed by T.K., T. Hamakubo, S.I. and T.M.

Author Information Atomic coordinates and structure factors for the A2AAR–Fab structure have been deposited in the Protein Data Bank under the accession codes 3VG9 (2.7 Å) and 3VGA (3.1 Å). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.M. (t.murata@faculty.chiba-u.jp) or S.I. (so_iwata@mac.com).

METHODS

Construction of A$_{2A}$AR expression vectors for *Pichia pastoris*. The coding sequence of A$_{2A}$AR from residues 1 to 316 including the N-terminal α-factor, FLAG-tag sequence and C-terminal 10×His-tag was synthesized by optimization of codon usage for *P. pastoris* (Takara Bio). In the construct, Asn 154 was also replaced by Gln to eliminate N-linked glycosylation. The DNA fragment was inserted into the multiple cloning site of the pPIC9K vector, and the linearized vector was transformed into the *P. pastoris* strain SMD1163 (Invitrogen) as described previously[24]. The transformed cells were stored as glycerol stocks at −80 °C.

Expression and purification of A$_{2A}$AR. A$_{2A}$AR was expressed in *P. pastoris* as described previously[24]. Cells were suspended in buffer A (50 mM sodium phosphate, 100 mM NaCl, 5% glycerol, 2 mM EDTA, protease inhibitor cocktail (Roche); pH 7.4) and disrupted with glass beads (0.5 mm; Biospec) by vigorous agitation with a conventional orbital shaker at 350 r.p.m. for 2 h at 4 °C. Following removal of unbroken cells and cell debris at 10,000g, membranes were isolated by ultracentrifugation at 100,000g for 45 min. Membranes were resuspended in buffer B (20 mM HEPES, 500 mM NaCl, 30% glycerol, EDTA-free protease inhibitor cocktail (Roche); pH 7.0) and solubilized using 1% *n*-dodecyl β-D-maltoside (DDM; Anatrace) containing 0.2% cholesterol hemisuccinate (CHS; Sigma) in the presence of 4 mM theophylline (antagonist) for 1–2 h at 4 °C. After ultracentrifugation, the supernatant was supplemented with solid imidazole to a final concentration of 40 mM and incubated overnight with a TALON immobilized metal ion affinity chromatography resin (Clonetech) at 4 °C with gentle rotation (1 ml of TALON resin per 150 mg of total protein). The resin was washed with buffer C (20 mM HEPES, 250 mM NaCl, 10% glycerol, protease inhibitor cocktail, 0.05% DDM, 0.01% CHS; pH 7.0) containing 20 mM imidazole, and the bound A$_{2A}$AR was eluted with buffer C containing 300 mM imidazole. The purified sample was incubated overnight with ConA resin at 4 °C to remove contaminating glycosylated proteins and was collected in the flow-through fraction. The final purified sample was dialysed against buffer C and concentrated to approximately 20 mg ml^{-1} by ultrafiltration (ULTRA-4 100 K, Millipore).

Construction, expression, and purification of A$_{2A}$AR-T4L. A$_{2A}$AR-T4L is a variant of A$_{2A}$AR in which the ICL3 region is replaced with a bacteriophage T4 lysozyme (T4L): Asn 2 to Tyr 161 of T4L were inserted between Leu 208 and Arg 222 within the ICL3 region, replacing residues Lys 209 to Ala 221. A$_{2A}$AR-T4L was expressed in *P. pastoris* and purified as described above.

Antibody generation. All animal experiments described in this study conformed to the guidelines outlined in the Guide for the Care and Use of Laboratory Animals of Japan and were approved by the University of Tokyo Animal Care Committee (approval no. RAC07101).

To raise antibodies against conformational epitopes of A$_{2A}$AR, we modified existing protocols for immunization and screening of mouse monoclonal antibodies. A detailed description of these modified protocols will be published elsewhere. Briefly, MRL/lpr mice were immunized with 0.1 mg purified A$_{2A}$AR antagonist ZM241385 complex three times at two-week intervals. The immunized mice were killed and single-cell suspensions were prepared from their spleens. These cells were fused with NS-1 myeloma cells using polyethylene glycol (PEG) according to conventional methods[25].

To screen antibodies that specifically recognize native receptors, we developed a novel ELISA method using proteoliposomes. For 'liposome-ELISA', we used purified A$_{2A}$AR reconstituted into liposomes containing biotinyl phosphatidylethanolamine (Avanti) to maintain the protein in its native conformation and effectively immobilize liposomes onto Streptavidin-coated plates (Nunc). To eliminate antibodies recognizing flexible loops, N (and C) termini or unstructured regions of A$_{2A}$AR, we performed ELISA using A$_{2A}$AR denatured with 1% sodium dodecyl sulphate. Denatured ELISA-negative cells were collected and evaluated using a BIAcore T100 (GE Healthcare) as described below. The selected cells were isolated by limiting dilution to establish monoclonal hybridoma cell lines producing antibodies against A$_{2A}$AR.

For large-scale antibody production, the monoclonal hybridoma cells were inoculated into BALB/c athymic nude mice. Immunoglobulin-G was collected from mouse ascites by precipitating twice with 50% ammonium sulphate and purified using Melon Gel (Thermo) according to the manufacturer's protocol. Fab fragments were obtained by proteolytic cleavage of immunoglobulin-G with papain (Worthington) and purified by anion exchange column chromatography (DEAE 5-PW, TOSOH). The sequences of Fab fragments were determined according to the standard 5′-RACE method using total RNA isolated from hybridoma cells.

Binding assay by surface plasmon resonance. The BIAcore T100 system and reagents, including sensor chips and amine coupling kit, were obtained from GE Healthcare. Monoclonal anti-mouse Fc antibody (200 μg ml^{-1}; Millipore) was immobilized on a CM5 sensor chip using the amine coupling kit according to the manufacturer's instructions. Antibodies in hybridoma culture supernatants (50 μl) or purified monoclonal antibodies (50 μg ml^{-1}) were tightly trapped by the

Fc antibody fixed on the sensor chip. The antibodies bound tightly enough not to be released from the surface when it was washed with buffer D (20 mM HEPES, 100 mM NaCl, 0.05% DDM, 0.01% CHS; pH 7.0). Purified A$_{2A}$AR (or A$_{2A}$AR-T4L) was passed over the surface and the specific binding was monitored for 2 min at 20 °C. Subsequently, the sensor surface was washed with buffer D and the dissociation was monitored for 6 min at 20 °C. Association and dissociation rate constants (k_{on} and k_{off}) were determined using a curve-fitting protocol as implemented in the BIAEVALUATION software (version 1.1, GE Healthcare) based on the Langmuir isotherm model assuming 1:1 binding stoichiometry.

Ligand binding assays. Ligand binding assays were performed using radioligands of the antagonist [^{3}H]-ZM241385 and the agonist [^{3}H]-NECA (GE Healthcare). For single-point binding assays, 5 nM [^{3}H]-ZM241385 and 5 μM [^{3}H]-NECA were incubated in 50 μl of buffer D containing 5 or, respectively, 50 nM purified A$_{2A}$AR with or without 500 nM antibody for 1 h on ice. For saturation-binding assays, varying concentrations of [^{3}H]-ZM241385 and [^{3}H]-NECA were incubated in 50 μl of buffer D containing 5 or, respectively, 50 nM purified A$_{2A}$AR with or without 500 nM antibody (Fab2838) for 1 h on ice. Receptor-bound ligands were separated by gel filtration[26] and radioactivity was measured using a LS6500 scintillation counter (Beckman). Data were analysed by a nonlinear-regression-fitting program using the GraphPad PRISM software. Competition assays with antagonists (SCH442416, theophylline) and agonists (NECA, adenosine) were performed in the presence of 1.0 nM [^{3}H]-ZM241385 for A$_{2A}$AR or 1.5 nM [^{3}H]-ZM241385 for A$_{2A}$AR-Fab (corresponding to the respective K_d values).

Purification and crystallization of the A$_{2A}$AR–Fab complex. Purified A$_{2A}$AR and the Fab fragments were mixed in a 1:1.2 molar ratio and were incubated on ice for 1 h. The mixture was loaded onto a Superdex 10/300 column (GE Healthcare) equilibrated with buffer C and eluted using the same buffer. The gel filtration step was repeated twice to ensure successful crystallization of the A$_{2A}$AR–Fab complex. Fractions containing the complex were concentrated to approximately 20 mg ml^{-1} by ultrafiltration (ULTRA-4 100 K, Millipore). Initial crystals were obtained using MemGold (Molecular Dimensions). After optimization, well-diffracting crystals were obtained in hanging drops by vapour diffusion at 20 °C with the protein solution containing 0.3–0.6% octylthioglucoside and the reservoir solution (1 μl) containing 30% PEG400, 0.1 M MES (pH 6.5) and 0.2 M MgCl$_2$. Crystals appeared after one day and grew to maximum dimensions in one week before being flash-frozen and stored in liquid nitrogen.

Data collection and structure determination. Diffraction data were collected from single cryo-cooled crystals (100 K) on beamline I24 at Diamond Light Source, UK, using a 10-μm focused beam (wavelength, 0.9795 Å) and a PILATUS 6M detector (Dectris). Data were processed using MOSFLM and SCALA from the CCP4 program suite[27]. The structure was initially solved using the data at 2.7 Å. Molecular replacement was carried out with PHASER[28] using the receptor from the A$_{2A}$AR-T4L fusion structure (PDB code, 3EML) and an antibody fragment (PDB code, 1P7K) as search models. Iterative cycles of model building and structure refinement were performed using COOT[29], REFMAC5[30] and phenix.refine in the PHENIX program package[31]. The final model from this refinement was used as the initial model for refinement against the data at 3.1 Å. The refinement was carried out as above. Model validation was performed using PROCHECK[32] and MOLPROBITY[33]. The resulting crystallographic and refinement statistics are summarized in Supplementary Table 2. Disordered region of A$_{2A}$AR was predicted by the RONN program[34]. Figures were prepared using PYMOL[35].

26. Warne, T., Chirnside, J. & Schertler, G. F. X. Expression and purification of truncated, non-glycosylated turkey β-adrenergic receptors for crystallization. *Biochim. Biophys. Acta* **1610,** 133–140 (2003).
27. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
28. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
29. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
30. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
31. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr. D* **61,** 850–855 (2005).
32. Laskowski, R. A., MacArthur, M. W. & Thornton, J. M. Validation of protein models derived from experiment. *Curr. Opin. Struct. Biol.* **8,** 631–639 (1998).
33. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
34. Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21,** 3369–3376 (2005).
35. DeLano, W. L. The PyMOL Molecular Graphics System ⟨http://www.pymol.org⟩ (2002).

LETTER

doi:10.1038/nature10752

Gated regulation of CRAC channel ion selectivity by STIM1

Beth A. McNally[1], Agila Somasundaram[1], Megumi Yamashita[1] & Murali Prakriya[1]

Two defining functional features of ion channels are ion selectivity and channel gating. Ion selectivity is generally considered an immutable property of the open channel structure, whereas gating involves transitions between open and closed channel states, typically without changes in ion selectivity[1]. In store-operated Ca^{2+} release-activated Ca^{2+} (CRAC) channels, the molecular mechanism of channel gating by the CRAC channel activator, stromal interaction molecule 1 (STIM1), remains unknown. CRAC channels are distinguished by a very high Ca^{2+} selectivity and are instrumental in generating sustained intracellular calcium concentration elevations that are necessary for gene expression and effector function in many eukaryotic cells[2]. Here we probe the central features of the STIM1 gating mechanism in the human CRAC channel protein, ORAI1, and identify V102, a residue located in the extracellular region of the pore, as a candidate for the channel gate. Mutations at V102 produce constitutively active CRAC channels that are open even in the absence of STIM1. Unexpectedly, although STIM1-free V102 mutant channels are not Ca^{2+}-selective, their Ca^{2+} selectivity is dose-dependently boosted by interactions with STIM1. Similar enhancement of Ca^{2+} selectivity is also seen in wild-type ORAI1 channels by increasing the number of STIM1 activation domains that are directly tethered to ORAI1 channels, or by increasing the relative expression of full-length STIM1. Thus, exquisite Ca^{2+} selectivity is not an intrinsic property of CRAC channels but rather a tuneable feature that is bestowed on otherwise non-selective ORAI1 channels by STIM1. Our results demonstrate that STIM1-mediated gating of CRAC channels occurs through an unusual mechanism in which permeation and gating are closely coupled.

Functional CRAC channels are tetramers of ORAI1 subunits[3–5], with the pore flanked by residues of the first transmembrane domain (TM1) of each subunit[6,7] (Fig. 1a). To localize the gate region that governs STIM1-dependent activation, we mutated individual pore-lining residues to Cys and analysed state-dependent differences in the sensitivity of mutant channels to methanethiosulphonate (MTS) Cys-reactive reagents[6,8]. Because the unusually narrow CRAC channel pore[9,10] prevents entry of the relatively large MTS reagents[6], we performed these studies in the E106D ORAI1 mutant, which has a wider pore yet maintains store-dependent activation[10]. With this as a background, several TM1 pore-lining residues, including V102C and G98C, became accessible to the small MTS reagent, MTSEA, with G98C showing particularly strong sensitivity to this reagent (Fig. 1 and Supplementary Fig. 1). Inhibition of G98C by MTSEA could be protected by La^{3+} (Fig. 1b, d), which blocks CRAC channels by binding to residues in the outer vestibule[6], and this is consistent with modification of G98C occurring from within the pore.

To determine differences in MTSEA accessibility between closed and open channels, we quantified the relief of MTSEA blockade that was elicited by the reducing agent bis(2-mercaptoethyl)sulphone (BMS)[6]. Resting cells were exposed to MTSEA for 100–120 s and subsequently, CRAC current (I_{CRAC}) was activated by passive store depletion (Fig. 1c). These experiments indicated that modification of G98C was profoundly state-dependent, with no modification occurring in closed channels (Fig. 1c, e). By contrast, D110C, a pore-lining residue located in the outer vestibule, was modified to similar extents in both closed and open states (Fig. 1e). The most straightforward explanation for this result is that the closed channel conformation prevents access of MTSEA to G98C, suggesting that the gate is located externally to G98 but below D110. As the key pore-lining TM1 residues in this region are E106 and V102, the gated access of G98C implicates these residues as potential candidates for the gate. E106 controls Ca^{2+} selectivity[11–13] and is not thought to regulate store-operated gating[10], leaving V102 as the most promising residue for further study.

Previous reports suggest that V102 is located very close to the central symmetry axis of the channel[6,7], that is, in a narrow constriction of the pore. If V102 is a component of the gating mechanism, mutations at this locus would be predicted to destabilize channel gating. Consistent with this possibility, a Cys mutation of V102 eliminated store-dependent gating. Cells expressing V102C ORAI1 and STIM1 displayed a large standing I_{CRAC} after whole-cell break-in (Fig. 2a). Moreover, resting cells exhibited constitutive Ca^{2+} entry and activation of the Ca^{2+}-dependent nuclear factor of activated T cell (NFAT) transcription factor (Supplementary Fig. 2), indicating that V102C ORAI1 channels are constitutively active.

Several lines of evidence indicated that the constitutive activation of V102C ORAI1 is STIM1-independent. Large La^{3+}-sensitive standing currents were seen in cells expressing V102C ORAI1 alone (Fig. 2b). Furthermore, Ca^{2+} imaging and NFAT activation experiments revealed constitutive Ca^{2+} entry in these cells (Fig. 2c and Supplementary Fig. 2c). Recent evidence indicates that STIM1 drives the redistribution of ORAI1 into discrete puncta after endoplasmic reticulum (ER) store depletion[2]. However, when expressed alone, V102C ORAI1 remained diffusely distributed in the plasma membrane (Fig. 2d). Moreover, I_{CRAC} in these cells did not show Ca^{2+}-dependent fast inactivation (CDI) (Supplementary Fig. 3). Because puncta formation and CDI require STIM1 (ref. 14–16), these results indicate that when overexpressed alone, the mutant channels are functionally free of STIM1. Consistent with this interpretation, knockdown of endogenous STIM1 in HEK293 cells did not affect the constitutive V102C current (Supplementary Table 1). However, when V102C ORAI1 was co-expressed with STIM1, puncta formation, interaction with STIM1, and CDI were indistinguishable from the behaviour of wild-type ORAI1 (Fig. 2d and Supplementary Figs 3 and 4). Additional analyses indicated that introducing the mutations E106A or R91W, which abrogate store-operated ORAI1 activity[11,12,17], strongly diminished V102C ORAI1 currents (Supplementary Fig. 5a), indicating that these residues are essential for both store-operated and constitutive activation modes of ORAI1. Mutation of the equivalent residue in Orai3 (V77C) also resulted in a STIM1-independent activation phenotype similar to that seen in V102C ORAI1 (Supplementary Fig. 6). Together, these results indicate that the V102C mutation destabilizes the channel gate, resulting in STIM1-independent constitutive ORAI1 activation.

Many ion channels including nicotinic acetylcholine receptors and the mechanosensitive channel MscL are reported to use hydrophobic

[1]Department of Molecular Pharmacology and Biological Chemistry, Northwestern University School of Medicine, 303 E Chicago Avenue, Ward 8-296, Chicago, Illinois 60611, USA.
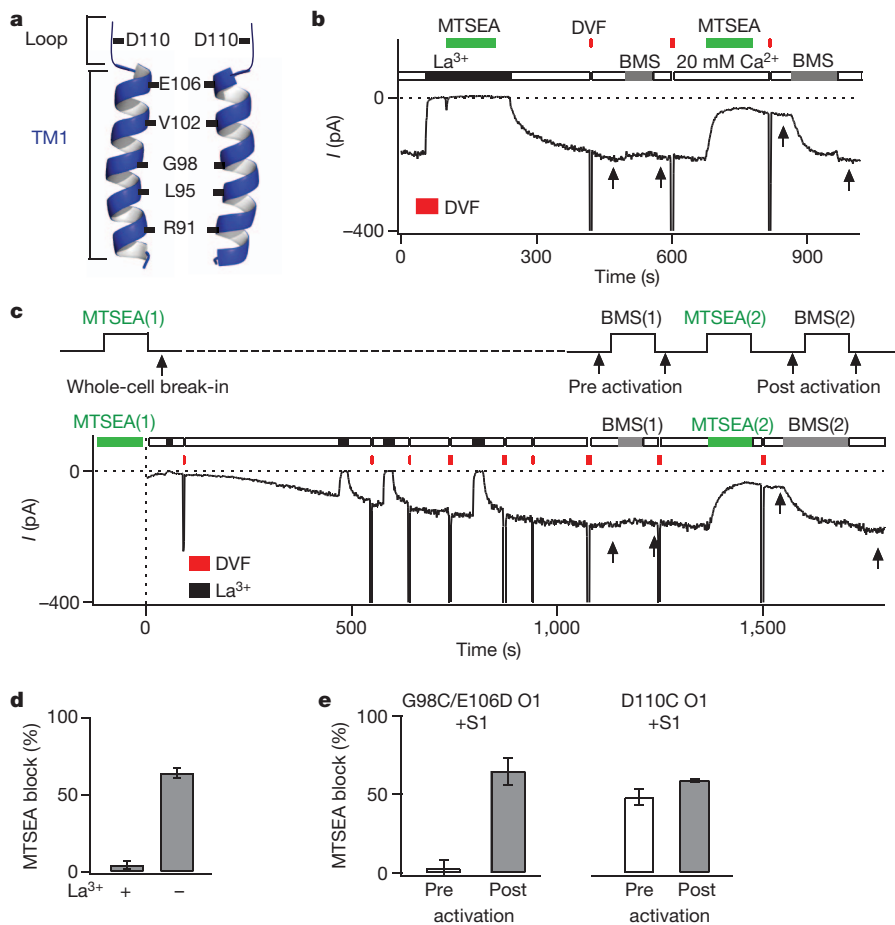
9 FEBRUARY 2012 | VOL 482 | NATURE | 241

©2012 Macmillan Publishers Limited. All rights reserved

Figure 1 | State-dependent accessibility of pore-lining residues localizes the activation gate to the extracellular TM1 region. a, Schematic representation of the key pore-lining residues in ORAI1 (refs 6, 7). **b**, MTSEA modification of G98C is protected by La^{3+}. A HEK293 cell co-expressing G98C/E106D ORAI1 (O1) and STIM1 was exposed to two applications of MTSEA (100 µM), the first in the presence of La^{3+} (100 µM) and the second after washout of La^{3+}. Periodic applications of a divalent free (DVF, red rectangles) solution facilitated washout of La^{3+}. MTSEA inhibition was quantified by the relief of block induced by BMS (5 mM) (arrows). **c**, State-dependent modification of G98C. MTSEA (200 µM) was applied for 120 s to resting cells and then washed off. After whole-cell break-in, I_{CRAC} was activated by passive store depletion by dialysing in BAPTA. BMS was applied to examine relief from MTSEA blockade (arrows). A second application of MTSEA and BMS provides a measure of blockade in open channels. A DVF solution was periodically applied to monitor Na^{+}–I_{CRAC}. **d**, Summary of MTSEA blockade of open G98C/E106D ORAI1 in the presence and absence of La^{3+}. **e**, Summary of blockade of G98C/E106D and D110C ORAI1 by MTSEA in closed and open channels. Values are mean ± s.e.m.

Figure 2 | Mutations at V102 cause STIM1-independent constitutive ORAI1 activation. a, Time course of the development of I_{CRAC} in cells expressing wild-type (WT) or V102C ORAI1 and STIM1 after whole-cell break-in. Intracellular Ca^{2+} stores were depleted by dialysing cells with 8 mM BAPTA. **b**, V102C ORAI1 currents are constitutively active in the absence of STIM1 co-expression. **c**, Intracellular calcium concentration $[Ca^{2+}]_I$ measurements in HEK293 cells expressing the indicated ORAI1 constructs in the absence of STIM1. UT, untransfected. **d**, Localization of V102C ORAI1–cyan fluorescent protein (CFP) before, and after, ER Ca^{2+} store depletion in the absence (left) or presence (right) of STIM1–yellow fluorescent protein (YFP). **e**, Mutational analysis of V102. Normalized current densities of V102 substitutions plotted against the solvation energies of the substituted amino acids[30] in the presence or absence of STIM1 co-expression. Currents were normalized to the mutant that yielded the maximal current density for each condition (Ala for STIM1-free cells and Ile in STIM1-co-expressing cells). Green, mutants that yield large constitutively active currents in the absence of STIM1; red, mutants that are not constitutively active but that require STIM1 for activation. TG, thapsigargin.

residues (Leu, Val and Ile) as gates to inhibit the flux of ions[18–20]. To explore the possibility that V102 comprises a hydrophobic gate in ORAI1, we investigated the side-chain dependence at this position for constitutive activation. We observed constitutive activity with several mildly hydrophobic and polar substitutions, including Cys, Gly, Ala, Ser and Thr (Fig. 2e). Conversely, substitutions to the highly hydrophobic amino acids Leu, Ile and Met resulted in only STIM1-dependent activation, as seen in wild-type ORAI1. Large hydrophobic residues such as Trp, Tyr and Phe attenuated both the constitutive and STIM1-induced currents, probably because of pore occlusion, as expected for a position that is nestled in a narrow region of the pore[6,7] (Supplementary Fig. 5b). Substitutions to extremely polar residues such as Glu, Asp, Lys and Arg resulted in non-functional channels with or without STIM1, probably owing to secondary effects of these mutations on the nearby selectivity filter at E106. Despite these deviations, however, the overall pattern is consistent with the hypothesis that V102 comprises a hydrophobic gate, with less hydrophobic substitutions producing a leaky gate.

CRAC channels are extraordinarily Ca^{2+} selective and poorly permeable to the large monovalent cation Cs^+. However, the ion selectivity of STIM1-free V102C ORAI1 channels differed from wild-type ORAI1 channels in both respects. STIM1-free V102C ORAI1 channels showed significantly lower Ca^{2+} permeability ($P<0.0001$), as demonstrated by the left-shifted reversal potentials of mutant Ca^{2+} currents (Fig. 3a). Consistent with this interpretation, replacement of extracellular Na^+ with $NMDG^+$, an impermeant ion, revealed significant Na^+ conduction in these channels ($P<0.0001$) (Fig. 3b). Direct estimates of

fractional Ca^{2+} currents using fluo-4 indicated that V102C ORAI1 conducts only 36% of the Ca^{2+} carried by wild-type ORAI1 in 20 mM Ca^{2+} (Supplementary Fig. 7a). In addition, unlike wild-type channels, STIM1-free V102C channels were highly permeable to the large monovalent cation Cs^+ (Fig. 3a–c and Supplementary Table 1).

Unexpectedly, co-expressing exogenous STIM1 together with V102C ORAI1 enhanced the Ca^{2+} permeability and lowered the Cs^+ permeability of V102C ORAI1, effectively correcting its aberrant ion selectivity (Fig. 3a–c). STIM1 also modified permeation of V102C ORAI1 for Ba^{2+} and Sr^{2+} (Supplementary Fig. 7b). Modification of ion selectivity by STIM1 was not unique to V102C ORAI1 but occurred in all constitutively active V102X mutants (Supplementary Table 1). These changes in ion selectivity required direct STIM1–ORAI1 interactions, as modification of V102C ORAI1 ion selectivity was nullified in the V102C/L276D ORAI1 double mutant (Fig. 3d), in which STIM1–ORAI1 binding was impaired (Supplementary Fig. 4b, c)[21]. STIM1-free and STIM1-bound V102C channels also displayed very different minimal pore widths (Fig. 3e), directly demonstrating that STIM1 alters the pore structure of V102C channels. Further analyses, using concatenated ORAI1 dimers (see Supplementary Methods) indicated that the subunit stoichiometry of STIM1-free and STIM1-bound mutant channels are identical, arguing that their distinct permeation properties are due to different pore structures of fully assembled, tetrameric channels rather than different subunit stoichiometries (Supplementary Fig. 8). Collectively, these results indicate that STIM1 binding modifies the structural features of the mutant channel

Figure 3 | STIM1 regulates ion selectivity of constitutively active V102C ORAI1 channels. a, Current–voltage (I–V) relationships of V102C ORAI1 currents in 20 mM Ca^{2+} and DVF Ringer's solutions. Arrows emphasize the reversal potential (V_{rev}) in each case. The bar graphs summarize the V_{rev} in the presence or absence of STIM1. Values are mean ± s.e.m. siSTIM1, short interfering RNA targeting STIM1. **b**, Effects of substituting extracellular Na^+ with $NMDG^+$ on V102C ORAI1 currents in the absence or presence of STIM1. **c**, Effects of replacing the standard extracellular Ringer's solution with Na^+- or Cs^+-based DVF solutions. In the absence of STIM1, large Cs^+ currents are seen

in V102C ORAI1 channels. By contrast, no Cs^+ conduction is observed in the presence of STIM1. **d**, I–V relationship of currents in the V102C/L276D ORAI1 double mutant in the presence or absence of STIM1. The bar graphs summarize the V_{rev} values of this mutant in the presence or absence of STIM1. Values are mean ± s.e.m. **e**, Relative permeabilities of V102C ORAI1 channels to different organic monovalent cations plotted against the size of each cation (test ion P_X and Na^+ P_{Na}). Dotted lines are fits to the hydrodynamic relationship. Values of d_{pore} (the apparent width of the pore) estimated from the fits are 4.9 Å for V102C ORAI1 + STIM1 channels and 6.9 Å for V102C ORAI1 channels.

pore, bestowing permeation properties that are associated with CRAC channels.

The normalization of ion selectivity of mutant channels by STIM1 suggests that their altered ion selectivity in the absence of STIM1 is not merely a byproduct of the mutations, but that it is indicative of a native intermediate activation state that is not readily seen in wild-type channels owing to a closed gate. Recent studies indicate that ORAI1 channels are activated in a nonlinear and cooperative manner by STIM1, with maximal channel activation requiring binding of eight STIM1 molecules per channel[15,22,23]. We reasoned that if modification of ion selectivity is coupled to the stoichiometry of STIM1 binding, then partial activation of wild-type ORAI1 by a sub-saturating concentration of STIM1 may lead to incomplete normalization of ion selectivity, revealing an intermediate activation state. To test this hypothesis, we used constructs in which wild-type ORAI1 was tethered to either one or two functional STIM1 (S or SS) domains (resulting in four or eight S domains per channel, respectively), as recently described[23]. We found that wild-type ORAI1 channels tethered to one S domain per subunit produced currents that were smaller and displayed diminished CDI compared to ORAI1–SS channels (Supplementary Fig. 9a), as expected from the known requirement of STIM1 for CRAC channel activation and inactivation[15,22,23]. Importantly, reversal potential measurements and ion substitution experiments indicated that unlike ORAI1–SS channels, ORAI1–S channels exhibited diminished Ca^{2+} and enhanced Cs^+ selectivity (Fig. 4a,b). Similar alterations in ion selectivity were also seen when wild-type ORAI1 was expressed with limiting concentrations of full-length STIM1 (Supplementary Fig. 9b). These results support the hypothesis that the V102 mutations stabilize an intermediate channel activation state. To gauge the dose dependence of ion selectivity modulation by STIM1, we examined reversal potentials of V102C and wild-type channels tagged to zero, one or two S domains (Fig. 4c, d). Despite the different starting functional states of wild-type (closed) and V102C (open) channels, STIM1 caused similar,

dose-dependent alterations in ion selectivity in both cases, while concomitantly enhancing activation of wild-type channels (Fig. 4d). Thus, ORAI1 channel activation and changes in ion selectivity probably result from the same underlying energetic changes that are driven by STIM1 binding.

Collectively, our results show that mutations at V102 cause constitutive activation of ORAI1 channels through a mechanism that probably involves destabilization of the channel gate at V102. This disposition of the STIM1 activation gate, in the extracellular region of the pore close to the selectivity filter (E106), is markedly different from the familiar structural designs in K^+ channels and by extension, voltage-gated Ca^{2+} (Ca_v) channels, which are constructed with the gate located at the cytoplasmic end of the pore[24]. We exploited the constitutive channel activity that resulted from mutations in the putative gate to identify an unusual ion channel gating mode in which STIM1 regulates ion selectivity and the pore architecture of CRAC channels. Activation by STIM1 bestows several key distinctive characteristics that are associated with CRAC channels including high Ca^{2+} selectivity, low Cs^+ permeability and a narrow pore to otherwise non-selective ORAI1 channels. Although the underlying mechanism remains to be established, the close proximity of the putative STIM1 activation gate (V102) to the selectivity filter (E106) probably contributes to the tight coupling of permeation and gating during channel activation. The altered ion selectivity of ORAI1 channels when STIM1 is limiting is reminiscent of the ion selectivity of ORAI1 and ORAI3 channels directly activated by the compound 2-APB, which exhibit lower Ca^{2+} selectivity and higher Cs^+ selectivity than STIM1-activated ORAI channels[25–28]. These findings indicate that the exquisite Ca^{2+} selectivity of CRAC channels is not an intrinsic and immutable property of ORAI1 but is instead uniquely manifested only in response to STIM1 gating. Given the emerging evidence that indicates that ORAI1 can be activated in a STIM1-independent manner by other cellular activators[29], these results raise the possibility that activation of

Figure 4 | STIM1 dose-dependently modulates the ion selectivity of ORAI1 channels. a, The addition of S domains to wild-type ORAI1 produces a rightward shift in the V_{rev} of I_{CRAC}. I–V relationships of wild-type ORAI1 channels that are tagged to either one or two S domains in the 20 mM Ca^{2+} and DVF Ringer's solutions. The bar graphs summarize the V_{rev} in each solution. Values are mean ± s.e.m. The top and bottom traces for each condition are from the same cell. b, Effects of substituting extracellular Na^+ with an impermant ion, $NMDG^+$. Removal of Na^+ diminishes the inward current in ORAI1–S channels, but not ORAI1–SS channels. c, I–V relationships and reversal potentials of V102C, V102C–S and V102C–SS channels. Values are mean ± s.e.m. Increasing the number of S domains to the V102C monomer causes a progressive rightward shift in the V_{rev} of I_{CRAC}. d, A plot of V102C ORAI1 V_{rev} (in the 20 mM Ca^{2+} Ringer's solution) against current density of wild-type ORAI1 channels tagged to zero, one or two S domains.

highly Ca^{2+} selective or non-selective ORAI1 currents may be a general mechanism for cells to tune Ca^{2+} and Na^+ entry through ORAI1 channels depending on the nature of the upstream activation signal.

METHODS SUMMARY

I_{CRAC} was recorded in the standard whole-cell patch-clamp configuration in HEK293 cells transfected with the indicated ORAI1 mutants, which were cloned into a bicistronic expression vector that co-expressed GFP. For recording I_{CRAC}, the membrane potential was hyperpolarized from $+30$ mV (holding) to -100 mV (100 ms) and then ramped from -100 mV to $+100$ mV (100 ms). MTS reagents were added to a 20 mM Ringer's or divalent-free Ringer's solution at the indicated concentrations. Second-order rate constants of MTSEA blockade were determined at a constant holding potential of -80 mV, as previously described[6].

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

1. Olcese, R. And yet it moves: conformational states of the Ca^{2+} channel pore. *J. Gen. Physiol.* **129,** 457–459 (2007).
2. Hogan, P. G., Lewis, R. S. & Rao, A. Molecular basis of calcium signaling in lymphocytes: STIM and ORAI. *Annu. Rev. Immunol.* **28,** 491–533 (2010).
3. Mignen, O., Thompson, J. L. & Shuttleworth, T. J. Orai1 subunit stoichiometry of the mammalian CRAC channel pore. *J. Physiol.* **586,** 419–425 (2008).
4. Penna, A. *et al.* The CRAC channel consists of a tetramer formed by Stim-induced dimerization of Orai dimers. *Nature* **456,** 116–120 (2008).
5. Madl, J. *et al.* Resting-state Orai1 diffuses as homotetramer in the plasma membrane of live mammalian cells. *J. Biol. Chem.* **285,** 41135–41142 (2010).
6. McNally, B. A., Yamashita, M., Engh, A. & Prakriya, M. Structural determinants of ion permeation in CRAC channels. *Proc. Natl Acad. Sci. USA* **106,** 22516–22521 (2009).
7. Zhou, Y., Ramachandran, S., Oh-Hora, M., Rao, A. & Hogan, P. G. Pore architecture of the ORAI1 store-operated calcium channel. *Proc. Natl Acad. Sci. USA* **107,** 4896–4901 (2010).
8. Karlin, A. & Akabas, M. H. Substituted-cysteine accessibility method. *Methods Enzymol.* **293,** 123–145 (1998).
9. Prakriya, M. & Lewis, R. S. Regulation of CRAC channel activity by recruitment of silent channels to a high open-probability gating mode. *J. Gen. Physiol.* **128,** 373–386 (2006).
10. Yamashita, M., Navarro-Borelly, L., McNally, B. A. & Prakriya, M. Orai1 mutations alter ion permeation and Ca^{2+}-dependent inactivation of CRAC channels: evidence for coupling of permeation and gating. *J. Gen. Physiol.* **130,** 525–540 (2007).
11. Prakriya, M. *et al.* Orai1 is an essential pore subunit of the CRAC channel. *Nature* **443,** 230–233 (2006).
12. Yeromin, A. V. *et al.* Molecular identification of the CRAC channel by altered ion selectivity in a mutant of Orai. *Nature* **443,** 226–229 (2006).
13. Vig, M. *et al.* CRACM1 multimers form the ion-selective pore of the CRAC channel. *Curr. Biol.* **16,** 2073–2079 (2006).
14. Mullins, F. M., Park, C. Y., Dolmetsch, R. E. & Lewis, R. S. STIM1 and calmodulin interact with Orai1 to induce Ca^{2+}-dependent inactivation of CRAC channels. *Proc. Natl Acad. Sci. USA* **106,** 15495–15500 (2009).
15. Scrimgeour, N., Litjens, T., Ma, L., Barritt, G. J. & Rychkov, G. Y. Properties of Orai1 mediated store-operated current depend on the expression levels of STIM1 and Orai1 proteins. *J. Physiol.* **587,** 2903–2913 (2009).
16. Derler, I. *et al.* A Ca^{2+} release-activated Ca^{2+} (CRAC) modulatory domain (CMD) within STIM1 mediates fast Ca^{2+}-dependent inactivation of ORAI1 channels. *J. Biol. Chem.* **284,** 24933–24938 (2009).
17. Feske, S. *et al.* A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441,** 179–185 (2006).
18. Doyle, D. A. Structural changes during ion channel gating. *Trends Neurosci.* **27,** 298–302 (2004).
19. Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423,** 949–955 (2003).
20. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T. & Rees, D. C. Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. *Science* **282,** 2220–2226 (1998).
21. Navarro-Borelly, L. *et al.* STIM1-Orai1 interactions and Orai1 conformational changes revealed by live-cell FRET microscopy. *J. Physiol.* **586,** 5383–5401 (2008).
22. Hoover, P. J. & Lewis, R. S. Stoichiometric requirements for trapping and gating of Ca^{2+} release-activated Ca^{2+} (CRAC) channels by stromal interaction molecule 1 (STIM1). *Proc. Natl Acad. Sci. USA* **108,** 13299–13304 (2011).
23. Li, Z. *et al.* Graded activation of CRAC channel by binding of different numbers of STIM1 to Orai1 subunits. *Cell Res.* **21,** 305–315 (2011).
24. Yellen, G. The voltage-gated potassium channels and their relatives. *Nature* **419,** 35–42 (2002).
25. Peinelt, C., Lis, A., Beck, A., Fleig, A. & Penner, R. 2-Aminoethoxydiphenyl borate directly facilitates and indirectly inhibits STIM1-dependent gating of CRAC channels. *J. Physiol.* **586,** 3061–3073 (2008).
26. Zhang, S. L. *et al.* Store-dependent and -independent modes regulating Ca^{2+} release-activated Ca^{2+} channel activity of human Orai1 and Orai3. *J. Biol. Chem.* **283,** 17662–17671 (2008).
27. Schindl, R. *et al.* 2-aminoethoxydiphenyl borate alters selectivity of Orai3 channels by increasing their pore size. *J. Biol. Chem.* **283,** 20261–20267 (2008).
28. Yamashita, M., Somasundaram, A. & Prakriya, M. Competitive modulation of CRAC channel gating by STIM1 and 2-aminoethyldiphenyl borate (2-APB). *J. Biol. Chem.* **286,** 9429–9442 (2011).
29. Feng, M. *et al.* Store-independent activation of Orai1 by SPCA2 in mammary tumors. *Cell* **143,** 84–98 (2010).
30. Radzicka, A., Pedersen, L. & Wolfenden, R. Influences of solvent water on protein folding: free energies of solvation of cis and trans peptides are nearly identical. *Biochemistry* **27,** 4538–4541 (1988).

Author Contributions B.A.M. generated, expressed and functionally characterized the ORAI mutants by patch-clamp electrophysiology with help from M.Y. The Ca^{2+} imaging, FRET imaging, TIRF microscopy and western blot analysis of ORAI1 mutants were performed by A.S. B.A.M, A.S. and M.P. contributed to the writing and editing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.P. (m-prakriya@northwestern.edu).

METHODS

Cells. HEK293 cells were maintained in suspension in a medium containing CD293 supplemented with 4 mM GlutaMAX (Invitrogen) at 37 °C, 5% CO$_2$. For imaging and electrophysiology, cells were plated and adhered to poly-L-lysine-coated coverslips at the time of passage, and grown in a medium containing 44% DMEM (Mediatech), 44% Ham's F12 (Mediatech), 10% fetal calf serum (HyClone), 2 mM glutamine, 50 U ml^{-1} penicillin and 50 μg ml^{-1} streptomycin.

Plasmids and transfections. Cys mutations that were used for characterization using electrophysiology were engineered into the previously described C-terminal Myc-tagged ORAI1 construct (MO7O ORAI1) in the bicistronic expression vector pMSCV-CITE-eGFP-PGK-Puro[11]. For generating the tandem dimers, the basic 'building blocks' (monomeric ORAI1 (MO7O) and the Bluescript SK$^+$ vector with ORAI1 attached to a linker) were obtained from T. Shuttleworth and constructed as previously described[3]. The orientation and number of subunits in the final constructs were confirmed by restriction enzyme analysis and western blot analysis. The ORAI1–S–eGFP and ORAI1–SS–eGFP constructs were kind gifts of T. Xu. The ORAI1–CFP, STIM1–YFP, and CFP–ORAI3 plasmids have been previously described[21,28]. Site-directed mutagenesis to generate ORAI1 mutations was performed using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) according to manufacturer's instructions and the results were confirmed by DNA sequencing. For electrophysiology, the indicated ORAI1 constructs were transfected into HEK293 cells either alone or together with a construct expressing unlabelled STIM1 (pCMV6-XL5, Origene Technologies). Cells were transfected with the indicated STIM1 and/or ORAI1 complementary DNA (ratio of 10:1 by mass for electrophysiology) using TransPass D2 (NEB Labs) and studied 24 h later. Cells that were transfected with siSTIM1 (Ambion) were studied 72 h following transfection.

Electrophysiology. Currents were recorded in the standard whole-cell configuration at room temperature on an Axopatch 200B amplifier (Molecular Devices) interfaced to an ITC-18 input–output board (Instrutech). Routines developed by R. S. Lewis and M. Prakriya on the Igor Pro software (Wavemetrics) were used for stimulation, data acquisition and data analysis. Data are corrected for the liquid junction potential of the pipette solution relative to Ringer's in the bath (-10 mV). The holding potential was $+30$ mV. The standard voltage stimulus consisted of a 100-ms step hyperpolarization to –100 mV followed by a 100-ms ramp depolarization from –100 to $+100$ mV, applied at 1-s intervals. Unless noted otherwise, the peak currents during the –100 mV pulse were measured for data analysis. For examining Ca^{2+}-dependent fast inactivation, the voltage protocol consisted of a 300-ms step decrease to –100 mV, applied at 2-s intervals. Unless otherwise indicated, I_{CRAC} was activated by passive depletion of intracellular Ca^{2+} stores by internal dialysis of 8 mM BAPTA through the pipette solution. To prevent complications arising from the changing membrane potential in the standard step–ramp voltage protocol, rate constants of blockade by MTS reagents were determined at a constant potential of –80 mV by acquiring 200-ms sweeps of current at 4 Hz. All currents were acquired at 5 kHz and low-pass filtered with a 1-kHz Bessel filter built into the amplifier. All data were corrected for liquid junction potential of the pipette solution and for leak currents collected in 50–150 μM LaCl$_3$.

MTS reagent protocol. The protocol for analysis of state-dependent modification of ORAI1 bearing Cys mutants is described in Fig. 1c. For the D110C mutant, this protocol was slightly modified to adjust for the formation of spontaneous disulphide bonds in this mutant[6]. The protocol included an additional application of the reducing agent BMS 90–120 s before the first MTSEA application to remove pre-existing disulphide bonds, as described previously[6].

Solutions. The standard extracellular Ringer's solution contained 130 mM NaCl, 4.5 mM KCl, 20 mM CaCl$_2$, 10 mM tetraethylammonium chloride (TEA-Cl), 10 mM D-glucose, and 5 mM Na-HEPES (pH 7.4). The DVF Ringer's solution contained 150 mM NaCl, 10 mM HEDTA, 1 mM EDTA, 10 mM TEA-Cl and 5 mM HEPES (pH 7.4). The 110-mM Ca^{2+} solution contained 110 mM CaCl$_2$, 10 mM D-Glucose and 5 mM HEPES (pH 7.4). The standard internal (pipette) solution contained 135 mM Cs-aspartate, 8 mM MgCl$_2$, 8 mM BAPTA and 10 mM Cs HEPES (pH 7.2). In experiments examining CDI, BAPTA was replaced with EGTA to accentuate CDI. In these experiments, the internal solution contained 125 mM Cs-aspartate, 10 mM EGTA, 3 mM MgCl$_2$, 8 mM NaCl and 10 Cs HEPES (pH 7.2). Stock solutions of MTS reagents (Toronto Research Chemicals) were prepared as previously described[6].

Data analysis. Reversal potentials were measured from the average of several leak-subtracted sweeps (4–6 sweeps) in each cell. Measurements were taken from 6–15 cells per mutant per condition. In cases in which the $I–V$ curve asymptotically approached the x axis at very positive membrane potentials with no clear reversal (for example, in wild-type ORAI1-expressing cells), the reversal potential was assigned as $+80$ mV. The MTSEA reaction rate constant was estimated from single exponential fits to the current decline after MTSEA application. The apparent second-order modification rate constant k_{on} was calculated from the relationship:

$$k_{on} = \frac{1}{\tau[\text{MTS}]}$$

where [MTS] is the concentration of the MTS reagent.

Relative permeabilities were calculated from changes in the reversal potential using the Goldman–Hodgkin–Katz voltage equation:

$$\frac{P_x}{P_{Na}} = \frac{[\text{Na}]_o}{[\text{X}]_o} e^{\Delta E_{rev} F/RT}$$

where R, T and F have their usual meanings, P_X and P_{Na} are the permeabilities of the test ion and Na$^+$, respectively, [X] and [Na] are the ionic concentrations, and ΔE_{rev} is the shift in reversal potential when the test cation is exchanged for Na$^+$. To estimate the minimal width of ORAI1 channels, the relative channel permeabilities for a series of organic monovalent cations of increasing size were examined as described before[10]. The cations used were ammonium (3.2 Å), methylammonium (3.78 Å), dimethylammonium (4.6 Å) and trimethylammonium (5.34 Å). These experiments were carried out in buffered Ca^{2+}-free solutions to avoid the potent blocking effects of Ca^{2+} ions on monovalent I_{CRAC}. The data were fitted to the hydrodynamic relationship[10]:

$$\frac{P_X}{P_{Na}} = k\left(1 - \frac{d_{ion}}{d_{pore}}\right)^2$$

where d_{ion} is the diameter of the tested ion and d_{pore} is the apparent width of the pore.

All data were corrected for leak currents collected in 20 mM Ca^{2+} + 50–150 μM La^{3+}. All curve fitting was done by least-squares methods using built-in functions in Igor Pro 5.0.

Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy

Guangbo Chen[1,2], William D. Bradford[1], Chris W. Seidel[1] & Rong Li[1,2]

Aneuploidy—the state of having uneven numbers of chromosomes—is a hallmark of cancer[1] and a feature identified in yeast from diverse habitats[2–5]. Recent studies have shown that aneuploidy is a form of large-effect mutation that is able to confer adaptive phenotypes under diverse stress conditions[2,6]. Here we investigate whether pleiotropic stress could induce aneuploidy in budding yeast (*Saccharomyces cerevisae*). We show that whereas diverse stress conditions can induce an increase in chromosome instability, proteotoxic stress, caused by transient Hsp90 (also known as Hsp82 or Hsc82) inhibition or heat shock, markedly increased chromosome instability to produce a cell population with high karyotype diversity. The induced chromosome instability is linked to an evolutionarily conserved role for the Hsp90 chaperone complex in kinetochore assembly[7,8]. Continued growth in the presence of an Hsp90 inhibitor resulted in the emergence of drug-resistant colonies with chromosome XV gain. This drug-resistance phenotype is a quantitative trait involving copy number increases of at least two genes located on chromosome XV. Short-term exposure to Hsp90 stress potentiated fast adaptation to un-related cytotoxic compounds by means of different aneuploid chromosome stoichiometries. These findings demonstrate that aneuploidy is a form of stress-inducible mutation in eukaryotes, capable of fuelling rapid phenotypic evolution and drug resistance, and reveal a new role for Hsp90 in regulating the emergence of adaptive traits under stress.

How cells maintain stable phenotypes and yet can adapt to diverse stress conditions through heritable change is a question with broad implications in evolution and disease progression. In prokaryotes, although the genome is propagated with high fidelity under normal conditions, extensive studies have demonstrated that different modes of genetic variation can be directly induced by stress, fuelling stress adaptation[9]. Recent work has revealed that one form of adaptive mutation in eukaryotic cells is the alteration of chromosome copy number, or aneuploidy[2,6,10]. Aneuploid yeast have been observed in diverse laboratory[2], industrial[3–5] and natural[4] environments. Aneuploidy leads to expression changes of many genes at levels that largely scale with gene copy number changes, bringing about marked phenotypic variation in a karyotype-specific manner under diverse growth conditions[10]. These findings suggest that to maintain phenotypic stability, karyotype stability must be ensured, and indeed intricate mechanisms have evolved to achieve highly accurate chromosome segregation and to prevent chromosome instability (CIN) during mitotic proliferation. Furthermore, as aneuploids are often at a growth disadvantage compared to euploids under stress-free conditions[10,11], the pre-existing karyotype diversity in a euploid population is likely to be limited for rapid adaptation when exposed to stressful environments. This raises the question of whether the cellular mechanisms ensuring chromosome transmission fidelity may be relaxed under stress, thus allowing the emergence of karyotypic diversity to fuel rapid cellular adaptation.

To test whether stress conditions in general could increase the rate of whole chromosomal instability, we exposed haploid yeast cells to chemicals inducing various types of pleiotropic stress (Supplementary Table 1) for 12–14 h and quantified chromosome loss rate by using the selection-neutral, chromosome-fragment-based colony colour assay (Fig. 1a and Supplementary Fig. 2; Supplementary Information)[12]. This initial screen revealed that many stress conditions, including hydrogen peroxide (oxidative stress), cycloheximide (translational stress), tunicamycin (endoplasmic reticulum stress), and so on, elevated the chromosome loss rate to a level similar to that caused by benomyl, a microtubule inhibitor (Fig. 1a). Surprisingly, radicicol, an Hsp90 inhibitor[13], was by far the most effective CIN inducer: the chromosome loss rate (7.4×10^{-2} per cell division) was hundreds of
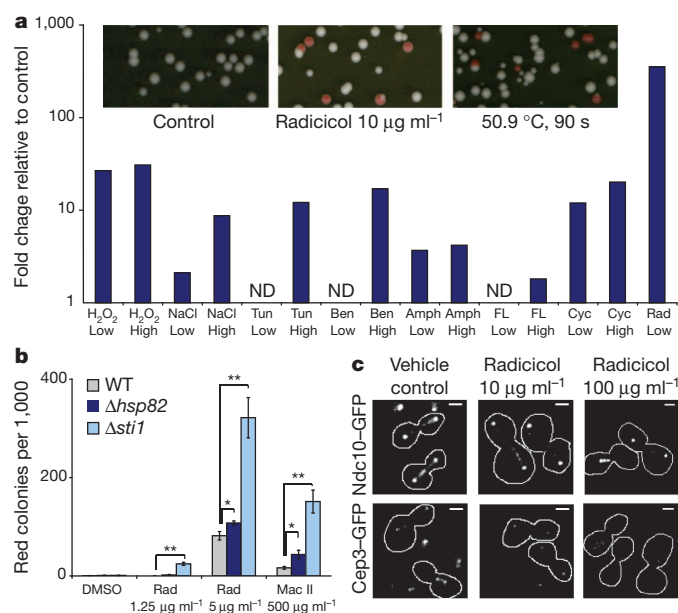

Figure 1 | Diverse stress conditions, especially Hsp90 inhibition, induce chromosomal instability. a, Top, colony appearance on yeast extract/peptone/dextrose (YPD) plates after cells were exposed to no stress (left), 16 h of $10 \, \mu g \, ml^{-1}$ radicicol treatment (middle) or 90 s heat shock at 50.9 °C. White colony colour indicates retention of the chromosome fragment; red indicates chromosome fragment loss. Bottom, chromosome fragment loss rates during exposure to diverse stress conditions were inferred from red colony frequencies normalized to that of the vehicle control population. Amph, amphotericin B; Ben, benomyl; Cyc, cycloheximide; Fl, fluconazole; Rad, radicicol; Tun, tunicamycin. ND, increase not detected over control. See Supplementary Figs 2, 3 and Supplementary Information for details. **b**, Deletion of *HSP82* or *STI1* sensitized the CIN-inducing effect of radicicol and macbecin II. Red colony frequencies normalized to that of wild-type (WT) dimethylsulphoxide (DMSO) control were averaged among 4 replicates, shown with standard error of the mean (s.e.m.). *$P < 0.05$; **$P < 0.01$, two-tailed *t*-test. **c**, Representative images showing kinetochore localization of Ndc10–GFP and Cep3–GFP under different conditions as indicated. Radicicol diminished Cep3–GFP localization at the kinetochore. Scale bar, 2 μm. See Supplementary Fig. 5 for additional images and quantification.

[1]Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, Missouri 64110, USA. [2]Department of Molecular and Integrative Physiology, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160, USA.

times above the control (2×10^{-4} per cell division), even at a radicicol concentration ($10\,\mu g\,ml^{-1}$ or $27\,\mu M$) with only a minor effect on growth (Fig. 1a and Supplementary Fig. 3). Quantitative polymerase chain reaction (qPCR) confirmed that red colonies induced by radicicol had lost the whole chromosome fragment (Supplementary Fig. 4a). Two of the thirteen tested red colonies were confirmed to have also gained chromosome (Chr) X or Chr XI (Supplementary Fig. 4b, c).

A similar aneuploidy-inducing effect was also observed with macbecin II, a structurally distinct Hsp90 inhibitor (Fig. 1b)[14]. Deletion of one copy of the Hsp90 genes, *HSP82*, led to enhanced chromosome fragment loss compared to the wild type in the presence of radicicol or macbecin II (Fig. 1b). Interestingly, deletion of *STI1*, the yeast homologue of mammalian *Hop* and a co-chaperone of Hsp90, resulted in significantly elevated CIN even at a concentration of radicicol too low to induce CIN on its own (Fig. 1b and Supplementary Fig. 5a). Heat is a common environmental stress known to tax Hsp90 function[15]. Heat shock for 90 s at 50.9 °C induced subsequent chromosome fragment loss at a rate comparable to that by pharmacological inhibition of Hsp90 (Fig. 1a). These results confirmed that Hsp90 stress is a potent inducer of aneuploidy. Hsp90 chaperone complexes are crucial facilitators of many cellular functions[16]. Previous biochemical studies suggested that Hsp90 is important for the activation of Ctf13 and assembly of the centromeric DNA binding factor 3 (CBF3) inner kinetochore complex[7]. Most CBF3 complex components, as well as the two co-chaperones involved in Ctf13 activation, showed haploinsufficiency towards radicicol (Supplementary Fig. 5b). Radicicol disrupted the kinetochore localization of Cep3 but had less effect on Ndc10 (also known as Cbf2), thus altering the stoichiometry of CBF3 complex at the kinetochore (Fig. 1c and Supplementary Fig. 5c, d, e). In addition to the CBF3 complex, Hsp90

interacts with several other pathways that could affect chromosome transmission fidelity, including the spindle assembly checkpoint[17] (see later).

Hsp90 taxation has previous been proposed to affect evolution by releasing phenotypic variation from pre-stored genetic diversity in the population and by transposon mobilization[16,18]. It is unclear whether Hsp90 inhibition also promotes adaptation by means of the induction of aneuploidy. As a first test, a diploid strain was grown in the presence of a high concentration of radicicol and the three largest radicicol-resistant (Rad^r) colonies were selected and reconfirmed (Fig. 2a, Supplementary Fig. 6 and Supplementary Information). Karyotyping revealed that all three Rad^r colonies were aneuploid with a dominant karyotype feature: all three Rad^r colonies, which adapted independently, contained one or two additional copies of Chr XV (Fig. 2a). A haploid Chr XV disomy strain, generated by genetic manipulation[11], also showed strong resistance to radicicol (Fig. 2b). A previous genome-wide screen identified a set of genes exhibiting haploinsufficiency towards macbecin II, among which two of the top genes are located on Chr XV: *STI1* and *PDR5*, a pleiotropic drug pump[17]. We deleted a single copy of the *STI1* or *PDR5* gene from Rad^r colony 3, trisomy for Chr XV. Growth measurements showed that either deletion abolished more than 50% of the growth rate gained by Chr XV trisomy over diploid in the presence of radicicol (Fig. 2c). A single copy of *STI1* and/or *PDR5* was then introduced into the parental diploid strain. An extra copy of each gene mildly but significantly increased radicicol resistance, but their combination markedly improved radicicol resistance (Fig. 2c and Supplementary Fig. 7). These results indicate that Chr XV gain directly confers radicicol resistance by increasing the copy number of *STI1* and *PDR5*, and possibly other genes carried on this chromosome (for example, *SGT1*).

Figure 2 | Aneuploidy is the predominant genetic change conferring adaptation to radicicol. a, Left, re-plating and growth of control (Ctrl) or three adapted Rad^r strains on $100\,\mu g\,ml^{-1}$ radicicol plates after 3 days incubation. See the experimental scheme in Supplementary Fig. 6. Right, all 3 re-confirmed Rad^r colonies were aneuploids with different levels of Chr XV gain. Intensity log$_2$ ratios over euploid are shown. Repetitive elements are shown as vertical

lines. **b,** Haploid Chr XV disomy generated by genetic manipulation shows higher growth rate than euploid in radicicol. OD, optical density (measured by Tecan M200Pro). **c,** Increased gene dosages of *STI1* and *PDR5* encoded on Chr XV are partially required and sufficient for radicicol resistance. The maximum growth rates were averaged for 4 replicates and normalized to diploid, shown with s.e.m.

We next investigated whether the karyotype diversity produced by Hsp90 stress-induced CIN could fuel adaptation to various other stress conditions. A karyotypically mosaic yeast cell population (~1/3 of the population were aneuploid with different karyotypes; Supplementary Fig. 8c) was generated by growing a diploid strain under moderate Hsp90 stress (20 μg ml^{-1} radicicol) for 2 days. This population was then tested for enhanced adaptability towards other stress conditions, including the presence of growth inhibiting concentrations of fluconazole,
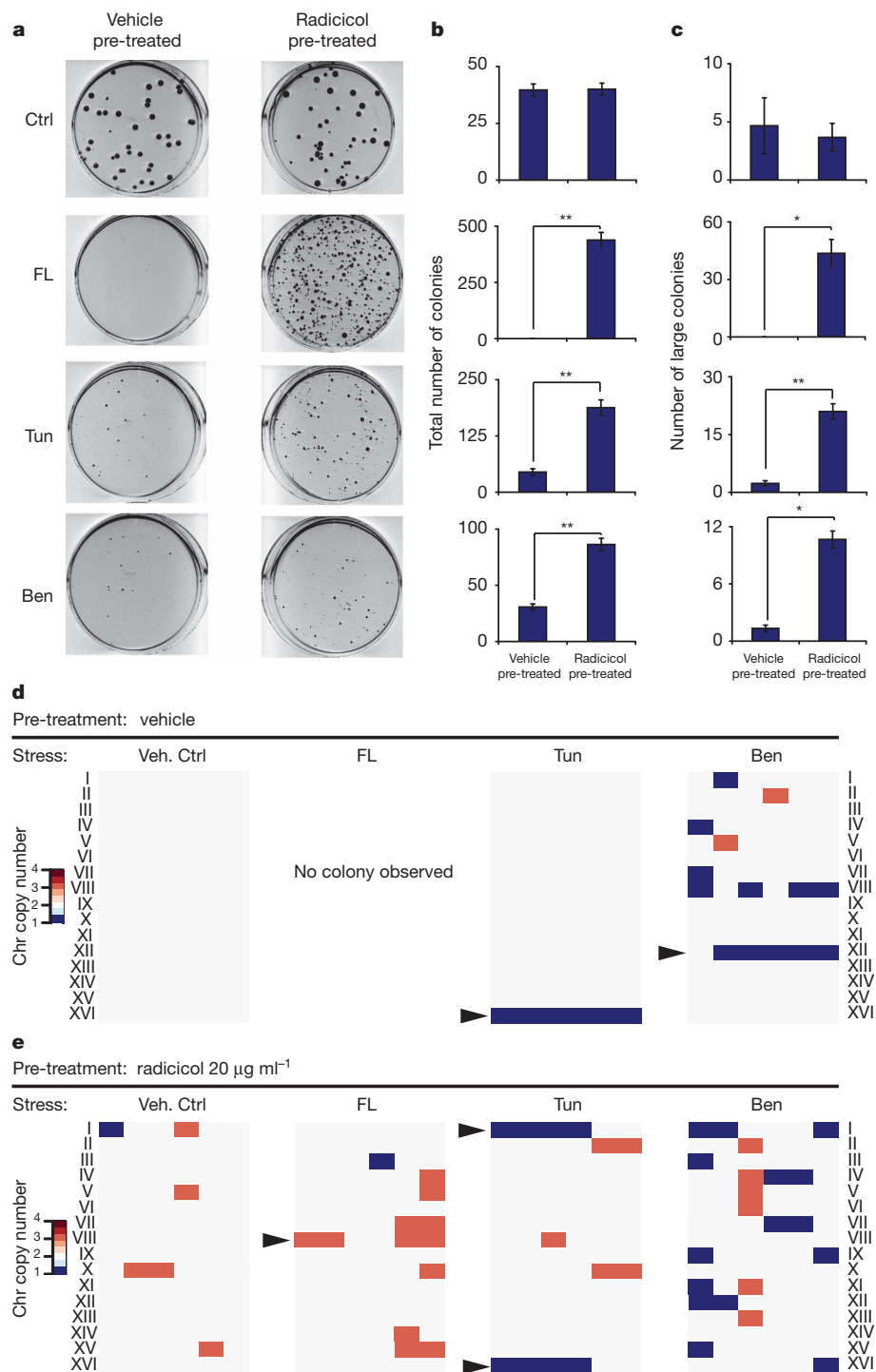

Figure 3 | Prior Hsp90 inhibition potentiates adaptation to other stress conditions through divergent aneuploid karyotypes. a, Plates of vehicle pre-treated group and radicicol pre-treated groups on different media as indicated. Approximately forty cells were plated on DMSO (Ctrl); ~40,000 cells were plated onto each drug plate. Ben, 30 μg ml^{-1} benomyl; FL, 32 μg ml^{-1} fluconazole; Tun, 2.5 μg ml^{-1} tunicamycin. **b**, Quantification of the number of viable colonies. Data are shown as mean ± s.e.m. from triplicate experiments. **c**, The sizes of all colonies (including both radicicol pre-treated and vehicle pre-treated groups) grown on each type of plates were measured. The distributions

of the top 10% largest colonies between the two groups are shown. *$P < 0.05$; **$P < 0.01$, two-tailed paired t-test. **d**, The karyotypes of 6 vehicle pre-treated colonies from 3 replicate experiments of each type as determined by qPCR[10]. **e**, The karyotypes of 6 independent radicicol pre-treated colonies from 3 replicate experiments of each type determined by qPCR. Arrowheads point to aneuploid chromosomes whose gain or loss frequency among resistant colonies was significantly higher than the starting populations ($P < 0.01$, Mantel–Haenszel tests, including data on Supplementary Fig. 8e).

tunicamycin or benomyl, over a control homogeneous euploid population (see experimental scheme in Supplementary Fig. 8a). The radicicol pre-treated population did not show any growth advantage over the control diploid (vehicle pre-treated) population on drug-free plates (Fig. 3a). However, on each of the different drug-containing plates, the radicicol pre-treated populations demonstrated markedly enhanced colony viability and increased frequency to form large drug-resistant colonies compared to the vehicle pre-treated population (Fig. 3a–c).

Twenty-one colonies were picked from the vehicle control plates bearing the radicicol pre-treated population, and out of these 12 were aneuploid, whereas none (0/9) from the control plate bearing the vehicle pre-treated population were aneuploid (Fig. 3d, e and Supplementary Fig. 8d, e). The vast majority (17/18) of the large colonies karyotyped from the drug plates bearing the radicicol pre-treated population were aneuploid (Fig. 3e). The drug-resistant colonies from the vehicle pre-treated population were also aneuploid (Fig. 3d). Importantly, the aneuploid colonies resistant to the same drug showed obvious karyotypic commonalities and tended to cluster together on the basis of karyotype similarity (Fig. 3e and Supplementary Fig. 9). For example, four of the five aneuploid colonies from the fluconazole plates karyotyped gained an extra copy of Chr VIII, which carries *ERG11*, encoding an ergostrol biosynthetic enzyme known to confer fluconazole resistance in *Candida albicans*[19]. Losing a copy of Chr XVI is a predominant karyotype change among the tunicamycin-resistant colonies (seen in 10/12 karyotyped colonies; Fig. 3d). Of the 12 benomyl-resistant colonies, 10 demonstrated karyotype clustering with 6 of them losing one Chr XII, but it appears that more than one karyotypic pattern could confer benomyl resistance. This, however, is consistent with our previous observation of phenotypic convergence of distinct karyotypic patterns[10]. All the above common karyotype features were significantly (Mantel–Haenszel tests) enriched in drug-resistant colonies but not the starting radicicol pre-treated population before selection on drug plates (Fig. 3d, e and Supplementary Fig. 8e), suggesting an association of specific karyotypes with resistance to certain drugs.

To assess further the selective advantage of aneuploidy and karyotype dynamics under varying stress levels, two Chr XVI monosomy colonies (Parent A and Parent B) from a tunicamycin plate were streaked on drug-free plates. Colonies of two distinct sizes emerged, with the small ones being predominant (Fig. 4a). Karyotyping showed that the small colonies represented Chr XVI monosomy, whereas the rare large

colonies had gained back the missing Chr XVI and returned to diploid (Fig. 4b and Supplementary Fig. 10a). Tunicamycin resistance was tightly linked to Chr XVI monosomy: all of the small colonies were tunicamycin resistant whereas the growth of the big colonies was abolished by tunicamycin (Fig. 4c and Supplementary Fig. 10b–d). This result shows that an adapted aneuploid population also has the potential to return to a euploid state when the stress condition is attenuated, suggesting that aneuploidy is not only a readily accessible mutation with large phenotypic impacts but that it is also reversible.

Taken together, the above results demonstrated that stress-induced CIN, leading to aneuploidy, is a mechanism of stress-induced mutagenesis in eukaryotes with high adaptive value to diverse perturbations (Supplementary Fig. 1). Hsp90 inhibition is by far the most potent inducer of aneuploidy among the stress conditions tested. This may be due to a broad but critical involvement of Hsp90 in pathways governing chromosome transmission fidelity and cell division[17]. For example, the mitotic checkpoint gene *MAD2* is a genetic interaction hub sensitive to Hsp90 perturbation[17]. *MAD2* deletion was also sufficient to lead to the rapid emergence of fluconazole-resistant colonies bearing an extra copy of Chr VIII (Supplementary Fig. 11). As Mad2 requires the CBF3 complex for its activity at the kinetochore[20], the exceptionally high-level CIN induced by Hsp90 inhibitors may be explained by a combined effect of interference with both kinetochore assembly and the checkpoint monitoring spindle defects. It is presently unknown whether the other stress conditions induce CIN through similar or different cellular targets.

The Hsp90 chaperone complex specializes in modulating the stability and function of many important regulatory and structural proteins[17]. As a result, Hsp90 acts as a capacitor facilitating evolutionary adaptation by unleashing the effects of pre-existing mutations when Hsp90 activity is taxed under mild stress[15,16]. Strong Hsp90 inhibition also induces phenotypic variation through transposon activation in *Drosophila*[18]. The results presented in this work reveal a new role for Hsp90 in adaptive evolution—as the guardian of chromosomal stability, the inhibition of which could trigger *de novo* karyotypic diversity, leading to rapid adaptation through aneuploidy. We note that our observed induction of aneuploidy required more potent Hsp90 inhibition than that required to reveal phenotypic effects of pre-existing mutations[15]. As the function of the Hsp90 chaperone complex in kinetochore assembly is conserved in mammalian species[8,21], Hsp90 stress-induced aneuploidy may be a mechanism of cellular adaptation affecting a wide range of organisms.

Figure 4 | Karyotype requirement and dynamics associated with tunicamycin resistance. a, b, Chr XVI monosomy (small colonies (arrow)) is unstable and produces large euploid progenies (arrowhead). Shown are representative images of the colonies (observed after 3 days growth on YPD) (**a**) and karyotypes of the Parent A and progeny colonies (A1–3) determined by qPCR (**b**). V-tun-1a, vehicle-pre-treated and tunicamycin-resistant 1a. **c**, Chr XVI monosomy progenies (A1 and A2) but not euploid progeny (A3) showed tunicamycin resistance. Note that the size difference between small and large colonies on control plates was no longer apparent after 7 days growth. See Supplementary Fig. 10 for data on Parent B.

METHODS SUMMARY

Yeast strains are listed in Supplementary Table 2. Standard genetic techniques were used for yeast strain construction. All deletions were verified by genomic PCR, and all aneuploid transformants were re-karyotyped by qPCR, and those retaining the original karyotype were used for experiments. Yeast qPCR karyotyping was performed as previously described[10]. Briefly, the chromosome copy number was inferred from qPCR with sets of primers located on peri-centrimeric regions. Array-based comparative genomic hybridization (aCGH) was performed on a home-made spot array.

A detailed description of all methods is provided in Supplementary Information.

1. Holland, A. J. & Cleveland, D. W. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nature Rev. Mol. Cell Biol.* **10,** 478–487 (2009).
2. Pavelka, N., Rancati, G. & Li, R. Dr Jekyll and Mr Hyde: role of aneuploidy in cellular adaptation and cancer. *Curr. Opin. Cell Biol.* **22,** 809–815 (2010).
3. Infante, J. J., Dombek, K. M., Rebordinos, L., Cantoral, J. M. & Young, E. T. Genome-wide amplifications caused by chromosomal rearrangements play a major role in the adaptive evolution of natural yeast. *Genetics* **165,** 1745–1759 (2003).
4. Kvitek, D. J., Will, J. L. & Gasch, A. P. Variations in stress sensitivity and genomic expression in diverse *S. cerevisiae* isolates. *PLoS Genet.* **4,** e1000223 (2008).
5. Borneman, A. R. et al. Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet.* **7,** e1001287 (2011).
6. Selmecki, A., Forche, A. & Berman, J. Genomic plasticity of the human fungal pathogen *Candida albicans*. *Eukaryot. Cell* **9,** 991–1008 (2010).
7. Stemmann, O., Neidig, A., Köcher, T., Wilm, M. & Lechner, J. Hsp90 enables Ctf13p/Skp1p to nucleate the budding yeast kinetochore. *Proc. Natl Acad. Sci. USA* **99,** 8585–8590 (2002).
8. Davies, A. E. & Kaplan, K. B. Hsp90–Sgt1 and Skp1 target human Mis12 complexes to ensure efficient formation of kinetochore–microtubule binding sites. *J. Cell Biol.* **189,** 261–274 (2010).
9. Rosenberg, S. M. Evolving responsively: adaptive mutation. *Nature Rev. Genet.* **2,** 504–515 (2001).
10. Pavelka, N. et al. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468,** 321–325 (2010).
11. Torres, E. M. et al. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317,** 916–924 (2007).
12. Spencer, F., Gerring, S. L., Connelly, C. & Hieter, P. Mitotic chromosome transmission fidelity mutants in *Saccharomyces cerevisiae*. *Genetics* **124,** 237–249 (1990).
13. Sharma, S. V., Agatsuma, T. & Nakano, H. Targeting of the protein chaperone, HSP90, by the transformation suppressing agent, radicicol. *Oncogene* **16,** 2639–2645 (1998).
14. Bohen, S. P. Genetic and biochemical Analysis of p23 and ansamycin antibiotics in the function of Hsp90-dependent signaling proteins. *Mol. Cell. Biol.* **18,** 3330–3339 (1998).
15. Jarosz, D. F. & Lindquist, S. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. *Science* **330,** 1820–1824 (2010).
16. Taipale, M., Jarosz, D. F. & Lindquist, S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nature Rev. Mol. Cell Biol.* **11,** 515–528 (2010).
17. McClellan, A. J. et al. Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell* **131,** 121–135 (2007).
18. Specchia, V. et al. Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* **463,** 662–665 (2010).
19. Selmecki, A., Gerami-Nejad, M., Paulson, C., Forche, A. & Berman, J. An isochromosome confers drug resistance *in vivo* by amplification of two genes, *ERG11* and *TAC1*. *Mol. Microbiol.* **68,** 624–641 (2008).
20. Gardner, R. D. et al. The spindle checkpoint of the yeast *Saccharomyces cerevisiae* requires kinetochore function and maps to the CBF3 domain. *Genetics* **157,** 1493–1502 (2001).
21. Niikura, Y. et al. 17-AAG, an Hsp90 inhibitor, causes kinetochore defects: a novel mechanism by which 17-AAG inhibits cell proliferation. *Oncogene* **25,** 4133–4146 (2006).

Author Contributions G.C. and R.L. designed and G.C. performed the experiments. W.D.B performed qPCR karyotyping. C.W.S. performed aCGH data analysis. G.C. and R.L. prepared the manuscript. R.L. conceived and supervised the project. All authors read and agreed with the content of the paper.

Author Information aCGH data are deposited at Gene Expression Omnibus under accession number GSE34087. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.L. (rli@stowers.org).

LETTER

Functional dissection of lysine deacetylases reveals that HDAC1 and p300 regulate AMPK

Yu-yi Lin[1,2,3], Samara Kiihl[4]*, Yasir Suhail[5]*, Shang-Yun Liu[1], Yi-hsuan Chou[1], Zheng Kuang[6,7], Jin-ying Lu[2,8], Chin Ni Khor[9], Chi-Long Lin[9], Joel S. Bader[5], Rafael Irizarry[4] & Jef D. Boeke[6,7]

First identified as histone-modifying proteins, lysine acetyltransferases (KATs) and deacetylases (KDACs) antagonize each other through modification of the side chains of lysine residues in histone proteins[1]. Acetylation of many non-histone proteins involved in chromatin, metabolism or cytoskeleton regulation were further identified in eukaryotic organisms[2–6], but the corresponding enzymes and substrate-specific functions of the modifications are unclear. Moreover, mechanisms underlying functional specificity of individual KDACs[7] remain enigmatic, and the substrate spectra of each KDAC lack comprehensive definition. Here we dissect the functional specificity of 12 critical human KDACs using a genome-wide synthetic lethality screen[8–13] in cultured human cells. The genetic interaction profiles revealed enzyme–substrate relationships between individual KDACs and many important substrates governing a wide array of biological processes including metabolism, development and cell cycle progression. We further confirmed that acetylation and deacetylation of the catalytic subunit of the adenosine monophosphate-activated protein kinase (AMPK), a critical cellular energy-sensing protein kinase complex, is controlled by the opposing catalytic activities of HDAC1 and p300. Deacetylation of AMPK enhances physical interaction with the upstream kinase LKB1, leading to AMPK phosphorylation and activation, and resulting in lipid breakdown in human liver cells. These findings provide new insights into previously underappreciated metabolic regulatory roles of HDAC1 in coordinating nutrient availability and cellular responses upstream of AMPK, and demonstrate the importance of high-throughput genetic interaction profiling to elucidate functional specificity and critical substrates of individual human KDACs potentially valuable for therapeutic applications.

To study the functional specificity of individual KDACs, we developed a genome-wide genetic interaction profiling technology in cultured human cells by RNA interference (RNAi) using a pooled human short hairpin RNA (shRNA) library from The RNAi Consortium (TRC)[12], and complexity deconvolution using a half-hairpin microarray (Fig. 1a). Microarray performance was evaluated (Supplementary Fig. 1a–d), and correlations between technical (Supplementary Fig. 1e) and biological replicates (Supplementary Fig. 1f) confirmed high methodological reproducibility.

In the screen, we used stable polyclonal HCT116 cells expressing shRNAs targeting firefly luciferase (shLuciferase) as a control. We checked the knockdown efficiency of individual shRNAs for 12 human KDACs (*HDAC1–4*, *HDAC6–9*, *SIRT1–3* and *SIRT5*) by immunoblotting (Supplementary Fig. 2a) or quantitative polymerase chain reaction (PCR) (Supplementary Fig. 2b), and generated stable polyclonal query cell lines expressing two shRNAs with the highest knockdown efficiency for each KDAC. The other six KDACs (*HDAC5*,
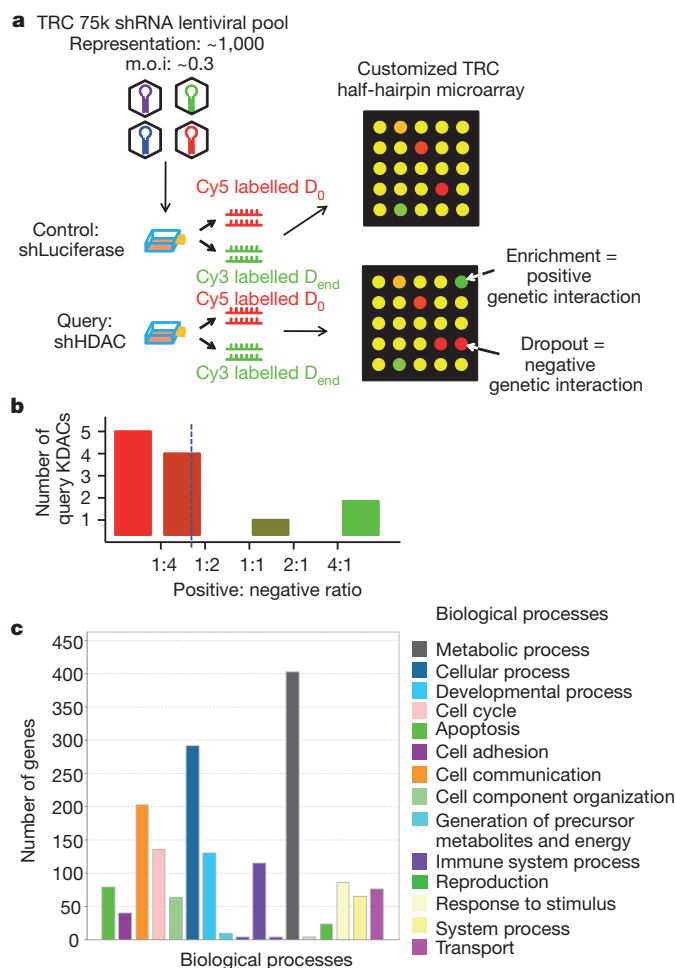

Figure 1 | Overview of human KDAC genetic interaction screen. a, Scheme of pooled shRNA-based primary screen. Selectively depleted and enriched shRNA clones in query KDAC knockdown cells indicate synthetic lethal (negative/aggravating) and rescue (positive/alleviating) interactions, respectively. **b**, Ratio of positive to negative genetic interactions for each query KDAC varies across the genome. The blue dashed line indicates the average ratio of all KDAC genetic interactions (approximately 1:2.6). **c**, Functional classification of validated KDAC genetic interaction partner genes based on GO biological process annotations. *P* values indicate significant enrichment for genes in corresponding biological processes; metabolic process, $P = 4.88 \times 10^{-22}$; cellular process, $P = 3.68 \times 10^{-16}$; developmental process, $P = 3.18 \times 10^{-8}$; cell cycle, $P = 9.45 \times 10^{-8}$.

[1]Institute of Biochemistry and Molecular Biology, College of Medicine, National Taiwan University, Taipei 100, Taiwan. [2]Department of Internal Medicine, National Taiwan University Hospital, Taipei 100, Taiwan. [3]Department of Oncology, National Taiwan University Hospital, Taipei 100, Taiwan. [4]Department of Biostatistics, Johns Hopkins Bloomberg, School of Public Health, Baltimore, Maryland 21231, USA. [5]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA. [6]Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [7]The High Throughput Biology Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [8]Department of Laboratory Medicine, National Taiwan University Hospital, Taipei 100, Taiwan. [9]National RNAi Core Facility, Institute of Molecular Biology, Academia Sinica, Taipei 115, Taiwan.
*These authors contributed equally to this work.

HDAC10 and *11*, *SIRT4* and *SIRT6* and *7*) were not tested owing to unsatisfactory knockdown efficiency with available shRNAs. After transduction by TRC shRNA lentiviral pools, 'benchmark samples' and 'end samples' were harvested before and after puromycin selection, respectively. A half-hairpin barcode library of each sample was recovered and hybridized to the microarray (Fig. 1a). Genes for which multiple shRNAs exceeded the threshold in either of two query lines for that KDAC were further validated by a cell viability assay (Supplementary Fig. 3). Eight-hundred and seventy-eight genetic interactions of human KDACs were validated from 6,307 candidates (Supplementary Tables 1 and 2). Query KDACs have mostly negative genetic interactions, excepting *HDAC6*, *SIRT3* and *SIRT5* (Fig. 1b and Supplementary Table 2), with an average positive to negative ratio of approximately 1:2.6, similar to observations in other human genes (1:3.8) and yeast genes (1:5.5)[14].

We arranged query KDAC genes by hierarchical clustering of genetic interaction pattern similarities and observed that KDACs of the same class co-clustered (Supplementary Fig. 4). Consistent with sharing common genetic interactions and biological functions, we also observed frequent aggravating interactions between same-class KDACs (Supplementary Fig. 5), including *HDAC1–HDAC2*, as previously shown[15], and four newly identified pairs (*HDAC3–HDAC8*, *HDAC4–HDAC5*, *SIRT1–SIRT2* and *SIRT3–SIRT5*). In contrast, alleviating interactions often exist between KATs and KDACs, suggesting that cells need to maintain homeostatic protein acetylation levels for viability, similar to observations in yeast[16]. This finding is

also consistent with the alleviating interactions between class I KDACs and *ACLY* (ATP-citrate lyase), the main source of intracellular acetyl-CoA, which controls KAT activity in human cells[6]. Functional classification by Gene Ontology (GO) annotation analysis revealed that several biological processes including metabolism, cell cycle and development are enriched among 615 genetic interaction partners (Fig. 1c and Supplementary Table 3). We also observed enrichment of co-repressors (Supplementary Table 4), consistent with crucial functions of KDACs in transcriptional regulation extending beyond histones. Interestingly, genes with predominantly negative interactions tend to be required for normal cell cycle progression in yeast[17], similar to these findings.

Beyond functional redundancy, distinct genetic interaction profiles also reveal functional hierarchies such as specific enzyme–substrate relationships. Consistent with this principle, we observed substantial overlap of the interaction profiles between knockdowns and catalytically defective (H199F) HDAC1 (ref. 18; Supplementary Table 5), and also significant enrichment of coexistent protein–protein interactions between KDACs and their interaction partners (Supplementary Table 6). Using a manually curated data set of human acetylated proteins[3–5], we observed significant enrichment of acetylation among KDAC genetic interaction partners (Supplementary Table 7), prompting us to wonder whether these interaction partners are substrates of the corresponding query KDACs. *In vitro* and *in vivo* deacetylation assays confirmed many such enzyme–substrate relationships (28/50 or 56%; Supplementary Fig. 6 and Supplementary Table 8) but not others

Figure 2 | Negative genetic interactions and enzyme–substrate relationship between HDAC1 and PRKAA1. a, Synthetic lethality was observed in HCT116 cells. Double KD, double knockdown. m.o.i., multiplicity of infection. Results are presented with error bars indicating ± 1 s.e.m. from three biological replicates. Significance was tested by a quasi-Poisson model. ***$P < 0.001$. **b, c**, *In vivo* acetylation assays show that endogenous PRKAA1 protein is deacetylated by HDAC1 and acetylated by p300 in HCT116 (**b**) and HepG2 (**c**) cells. Ac-K, acetyl-lysine; IP, immunoprecipitate; WCE, whole cell extract.

d, *In vivo* acetylation changes in GST–PRKAA1 in response to *HDAC1* and *p300* knockdown. **e**, Sequential mutation of three acetylable lysine residues (K40, K42 and K80; or K31, K33 and K71 in another reading frame) to arginine progressively diminished the acetylation signal of GST–PRKAA1. 1KR, K80R; 2KR, K40/42R; 3KR, K40/42/80R. **f**, A conventional histone acetyltransferase activity assay revealed significantly diminished *in vitro* p300 acetylation of PRKAA1(3KR). **g**, PRKAA1 is deacetylated *in vitro* by HDAC1 (activity inhibited by TSA), but not SIRT1.

(Supplementary Figs 7–10 and Supplementary Table 8). Most of the validated substrates (22/26 or 84.6%) have metabolic functions in maintaining macromolecular homeostasis.

Differentiating HDAC1 from HDAC2 biochemically is challenging owing to extensive sequence homology and frequent co-membership in protein complexes[19]. Using a functional genomics approach we efficiently identified HDAC1-specific substrates and functions. Nearly all AMPK subunit isoforms are negative genetic interacting partners of *HDAC1* but not *HDAC2* (Supplementary Table 2). To further investigate biological significance, we confirmed the negative genetic interaction between *PRKAA1* and *HDAC1* in HCT116 (Fig. 2a), HepG2 (Supplementary Fig. 11a) and IMR-90 cells (Supplementary Fig. 11b). Two efficient shRNAs each were selected for *HDAC1* (Supplementary Fig. 2a and 12a) and *PRKAA1* (Supplementary Fig. 12b). Previously identified as an acetylated protein[3], we investigated the *in vivo* acetylation status of endogenous PRKAA1 in HCT116 cells. Consistent with a genetic interaction between *PRKAA1* and *HDAC1* but not the other KDACs examined, we uncovered a substantial increase in the fraction of endogenous PRKAA1 acetylated only in *HDAC1* knockdown cells (Fig. 2b). A similar approach was applied to identify p300 as the probable major acetyltransferase for PRKAA1 in HCT116 cells (Fig. 2b and Supplementary Fig. 12c). This finding is consistent with alleviating genetic interactions between counteracting *p300* and *HDAC1* (Supplementary Fig. 5 and Supplementary Table 2).

The enzyme–substrate relationships were also conserved in HepG2 cells (Fig. 2c) and IMR-90 cells (Supplementary Fig. 13a). The significant increase of endogenous PRKAA1 acetylation in HepG2 cells treated with class I/II KDAC inhibitor trichostatin A (TSA) but not class III KDAC inhibitor nicotinamide (Supplementary Fig. 13b) is consistent with the specific enzyme–substrate relationship between HDAC1 and PRKAA1.

Three potential acetylation sites in PRKAA1—K40, K42 and K80— have been identified by tandem mass spectrometry[3]. To examine their physiological relevance, we introduced GST-tagged wild-type *PRKAA1* into HepG2 cells. The recombinant GST–PRKAA1 co-immunoprecipitated with endogenous AMPK regulatory β (PRKAB1) and γ (PRKAG1) subunits (Supplementary Fig. 14), supporting the formation of a fully functional AMPK complex in these cells. The acetylation status of GST–PRKAA1 changed in parallel to that of endogenous PRKAA1 on knockdown of *HDAC1* or *p300* (Fig. 2d), and its acetylation decreased incrementally as the three lysine residues were sequentially mutated to arginine to mimic constitutive deacetylation (Fig. 2e). *In vitro* acetylation revealed that p300-dependent acetylation of PRKAA1 from HepG2 cells largely required these three lysine residues (Fig. 2f and Supplementary Fig. 15). Moreover, the affinity-purified HDAC1 complex effectively deacetylated purified PRKAA1 *in vitro*, an activity inhibited by TSA, whereas purified SIRT1 did not (Fig. 2g). The three acetylable lysine residues are
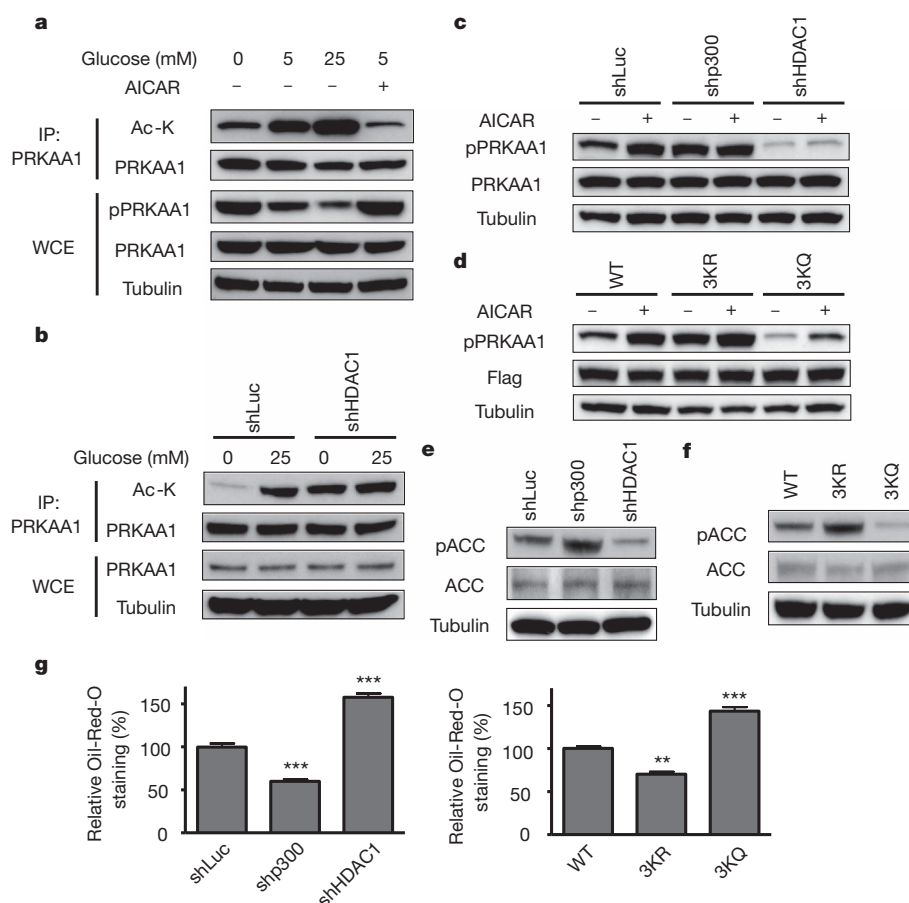

Figure 3 | Deacetylation of PRKAA1 increases its phosphorylation and activity. a, Acetylation and phosphorylation signal of endogenous PRKAA1 upon different glucose concentration or AICAR treatment (2 mM) in HepG2 cells. pPRKAA1, phosphorylated PRKAA1. **b**, Knockdown of *HDAC1* preserves the acetylation signal of endogenous PRKAA1 upon glucose deprivation. **c**, Knockdown of *p300* or *HDAC1* changes basal and responsive levels of phosphorylation of endogenous PRKAA1. **d**, Deacetylation (3KR) and acetylation (3KQ) mimics of PRKAA1 increase and decrease basal and responsive level of phosphorylation, respectively. **e**, Consistent with changes of PRKAA1 phosphorylation and activity, perturbations that decrease (*p300* knockdown) or increase (*HDAC1* knockdown) acetylation cause increases and decreases of ACC phosphorylation (pACC), respectively. **f**, Deacetylation (3KR) and acetylation (3KQ) mimics of PRKAA1 increase and decrease ACC phosphorylation, respectively. WT, wild type. **g**, Acetylation of PRKAA1 regulates intracellular lipid droplet abundance assessed by Oil-Red-O staining. Error bars indicate ± 1 s.e.m. from three biological replicates. Significance was tested by Student's *t*-test. **P < 0.01; ***P < 0.001.

completely conserved in human PRKAA2, the other AMPK catalytic subunit homologue. We evaluated the *in vivo* acetylation of endogenous PRKAA2 in HepG2 cells and discovered similar changes in acetylation on *HDAC1* or *p300* knockdown (Supplementary Fig. 16a). In contrast, the acetylation status of PRKAG1 and PRKAG2 was unaffected by *HDAC1* knockdown (Supplementary Fig. 16b).

AMPK is both an important sensor and a regulator of energy homeostasis, maintaining the balance of ATP consumption and production in eukaryotic cells[20]. AMPK is activated when intracellular energy status is compromised by metabolic stress that increases AMP/ATP and ADP/ATP ratios[21,22]. In energy-deprived conditions, a crucial threonine residue in the activation loop of the catalytic subunit (T183 of PRKAA1 or T172 of PRKAA2) is phosphorylated by upstream kinases, and the high concentration of AMP and ADP allosterically binds the γ subunit and protects AMPK against dephosphorylation[21–25]. The activated AMPK in turn phosphorylates various downstream effector proteins to switch on catabolic pathways, enhance transcription of stress-response genes and reduce protein synthesis[20,26]. To investigate whether acetylation of the AMPK catalytic subunit regulates its activity, we examined the correlation between PRKAA1 acetylation and phosphorylation of the critical threonine residue using immunoblotting. Consistent with previous findings, PRKAA1 phosphorylation increased upon energy deprivation, which was achieved by lowering glucose concentration or by adding the AMP analogue 5-aminoimidazole-4-carboxamide riboside (AICAR) (Fig. 3a). PRKAA1 acetylation and phosphorylation were oppositely regulated, suggesting that acetylation might be negatively correlated with AMPK activity (Fig. 3a). Moreover, the decrease in endogenous PRKAA1 acetylation seen on glucose deprivation is largely reverted in *HDAC1* knockdown (Fig. 3b). Basal levels of endogenous PRKAA1 phosphorylation in 5 mM glucose increased and diminished markedly upon knockdown of *p300* and *HDAC1*, respectively, whereas enhancement of phosphorylation reactive to AICAR treatment was greatly damped in both knockdowns (Fig. 3c). The changes in basal and AICAR-reactive phosphorylation discovered in acetylation (3KQ) and deacetylation (3KR) mimics of Flag-tagged PRKAA1 in HepG2 cells were similar to those seen in response to *HDAC1* and *p300* knockdown, respectively (Fig. 3d), suggesting that acetylation of these three lysine residues critically modulates enzyme activation by phosphorylation.

AMPK phosphorylates and inactivates acetyl-CoA carboxylase (ACC) to shut down fatty acid synthesis and enhance fatty acid oxidation[20]. Using ACC phosphorylation as an intracellular indicator of AMPK enzymatic activity, the observed increases and decreases of ACC phosphorylation were consistent with the proposed trend of AMPK activity change in response to *p300* and *HDAC1* knockdown, respectively (Fig. 3e). We further showed that regulation of ACC phosphorylation is also controlled mainly by acetylation of the three critical PRKAA1 lysines (Fig. 3f and Supplementary Fig. 17a, b). Consistent with a negative impact of PRKAA1 acetylation on ACC phosphorylation, intracellular lipid droplet content dropped and rose in low and high acetylation conditions, respectively (Fig. 3g).

PRKAA1 phosphorylation is reduced and unresponsive to changes in PRKAA1 acetylation in HepG2 cells with knockdown of *LKB1* (Fig. 4a and Supplementary Fig. 18a) or in HeLa cells lacking *LKB1* expression (Supplementary Fig. 18b), suggesting that the inverse levels of PRKAA1 acetylation and phosphorylation depend on *LKB1*. A possible mechanism underlying this observation is the direct regulation of the physical interaction between PRKAA1 and LKB1 by lysine acetylation. We examined binding between recombinant or endogenous PRKAA1 and LKB1 by co-immunoprecipitation and observed enhanced and weakened binding in conditions of low (*p300* knockdown or *PRKAA1(3KR)* mutant) and high (*HDAC1* knockdown or *PRKAA1(3KQ)* mutant) PRKAA1 acetylation, respectively (Fig. 4b, c). Consistently, acetylation of purified PRKAA1 also negatively controlled its phosphorylation by LKB1 *in vitro* (Fig. 4d and
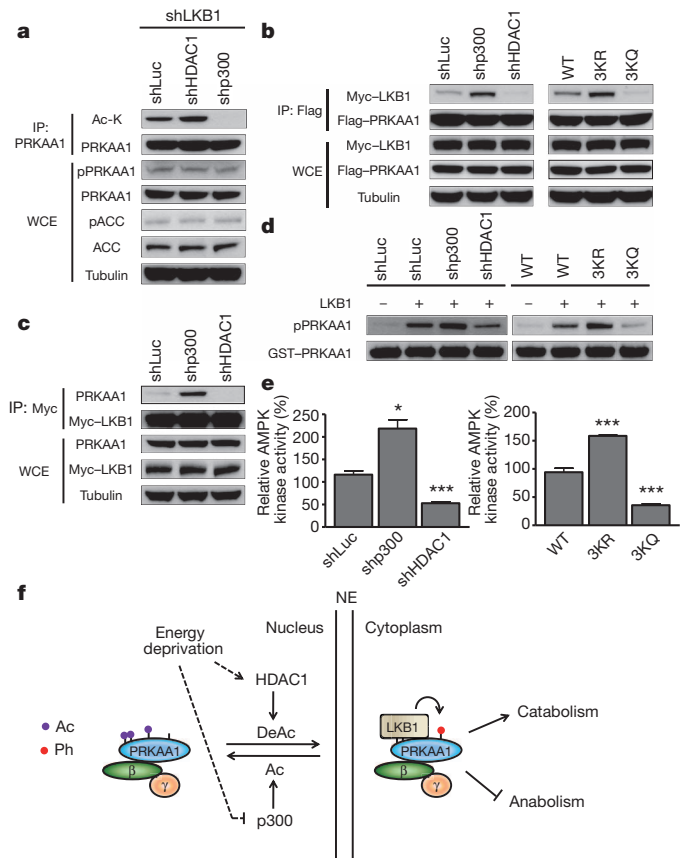

Figure 4 | Deacetylation of PRKAA1 specifically enhances its physical interaction with LKB1 kinase. a, Phosphorylation of endogenous PRKAA1 is reduced and unresponsive to acetylation status upon knockdown of *LKB1*. **b**, **c**, Acetylation of PRKAA1 inhibits the physical interaction between LKB1 and recombinant PRKAA1 (**b**) or endogenous PRKAA1 (**c**) as assessed by co-immunoprecipitation. **d**, Acetylation of PRKAA1 regulates phosphorylation by LKB1 *in vitro*. **e**, Acetylation of PRKAA1 regulates its kinase activity. Error bars indicate ± 1 s.e.m. from three biological replicates. Significance was tested by Student's *t*-test. ***$P < 0.001$. **f**, Schematic model for crosstalk between acetylation and phosphorylation of AMPK catalytic subunit PRKAA1 governed by counteracting HDAC1 and p300. Nucleocytoplasmic translocation of AMPK may be required to approach p300 and HDAC1 in the nucleus and LKB1 in the cytoplasm. Ac, acetylation; Ph, phosphorylation; NE, nuclear envelope. Solid lines indicate the paths implicated in previous and present studies; dashed lines indicate those paths hypothesized as part of this study.

Supplementary Fig. 18c), and its own kinase activity (Fig. 4e). These findings suggest that acetylation on PRKAA1 leads to inhibition of its physical interaction with LKB1, and subsequent phosphorylation and activation of itself and downstream effector proteins. Therefore, HDAC1 serves as the critical metabolic regulator to govern deacetylation and the subsequent activation of AMPK, which adaptively turns on catabolic processes and switches off anabolic pathways in human liver cells upon energy deprivation (Fig. 4f).

With the increasing use of KDAC inhibitors for the treatment of neoplastic and neurodegenerative diseases[27,28], and also the generation of induced pluripotent stem cells[29], it is critical to understand the molecular mechanisms underlying these effects. Despite the potential limitation in terms of cell line and phenotype specificity, the genome-wide genetic interaction profiling of human KDACs described here helps identify a multitude of specific substrates of individual KDACs. We further report important metabolism-regulating roles of HDAC1 to govern crosstalk between acetylation and phosphorylation of the AMPK catalytic subunit by controlling its physical interaction with the upstream kinase LKB1 that modulates AMPK activity and thus lipid

metabolism in human liver cells. Identifying the enzyme–substrate relationships of individual KDACs and understanding the 'induced essentiality' of genes upon KDAC knockdown, combined with the recent development of selective KDAC inhibitors[30], should pave the way for future molecular targeted therapy through inhibition of specific KDACs.

METHODS SUMMARY

All human cell lines were obtained from the American Type Culture Collection unless mentioned otherwise. The knockdown efficiency of query KDAC shRNAs was assayed using immunoblotting and quantitative PCR; the two shRNAs with maximal knockdown effect for each query were used to generate stable polyclonal query cell lines. The primary screen was performed on a custom half-hairpin microarray as previously described[12] with further optimization. Genetic interactions between target and query genes were identified using normalized $\log_2$ ratios of Cy5 to Cy3 signal intensities of the benchmark and final samples. Genes for which multiple shRNAs exceeded the threshold were further validated by a cell viability assay in 96-well format. Enzyme–substrate relationships of genetic interacting partners and their query KDACs were confirmed by *in vitro* and *in vivo* deacetylation assays. The effects of (de)acetylation conducted by counteracting HDAC1 and p300 on PRKAA1 phosphorylation and activation were demonstrated using biochemical experiments. For more details on experimental procedures and data analysis see Methods.

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

1. Yang, X. J. & Seto, E. The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nature Rev. Mol. Cell Biol.* **9**, 206–218 (2008).
2. Lin, Y. Y. *et al.* Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. *Cell* **136**, 1073–1084 (2009).
3. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840 (2009).
4. Kim, S. C. *et al.* Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell* **23**, 607–618 (2006).
5. Zhao, S. *et al.* Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000–1004 (2010).
6. Wellen, K. E. *et al.* ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* **324**, 1076–1080 (2009).
7. Haberland, M., Montgomery, R. L. & Olson, E. N. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nature Rev. Genet.* **10**, 32–42 (2009).
8. Scholl, C. *et al.* Synthetic lethal interaction between oncogenic *KRAS* dependency and *STK33* suppression in human cancer cells. *Cell* **137**, 821–834 (2009).
9. Luo, J. *et al.* A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835–848 (2009).
10. Silva, J. M. *et al.* Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science* **319**, 617–620 (2008).
11. Schlabach, M. R. *et al.* Cancer proliferation gene discovery through functional genomics. *Science* **319**, 620–624 (2008).
12. Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proc. Natl Acad. Sci. USA* **105**, 20380–20385 (2008).
13. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic *KRAS*-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
14. Stark, C. *et al.* The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.* **39**, D698–D704 (2011).
15. Haberland, M., Johnson, A., Mokalled, M. H., Montgomery, R. L. & Olson, E. N. Genetic dissection of histone deacetylase requirement in tumor cells. *Proc. Natl Acad. Sci. USA* **106**, 7751–7755 (2009).
16. Lin, Y. Y. *et al.* A comprehensive synthetic genetic interaction network governing yeast histone acetylation and deacetylation. *Genes Dev.* **22**, 2062–2074 (2008).
17. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
18. Hassig, C. A. *et al.* A role for histone deacetylase activity in HDAC1-mediated transcriptional repression. *Proc. Natl Acad. Sci. USA* **95**, 3519–3524 (1998).
19. Grozinger, C. M. & Schreiber, S. L. Deacetylase enzymes: biological functions and the use of small-molecule inhibitors. *Chem. Biol.* **9**, 3–16 (2002).
20. Hardie, D. G. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. *Nature Rev. Mol. Cell Biol.* **8**, 774–785 (2007).
21. Xiao, B. *et al.* Structure of mammalian AMPK and its regulation by ADP. *Nature* **472**, 230–233 (2011).
22. Oakhill, J. S. *et al.* AMPK is a direct adenylate charge-regulated protein kinase. *Science* **332**, 1433–1435 (2011).
23. Hawley, S. A. *et al.* Characterization of the AMP-activated protein kinase kinase from rat liver and identification of threonine 172 as the major site at which it phosphorylates AMP-activated protein kinase. *J. Biol. Chem.* **271**, 27879–27887 (1996).
24. Hawley, S. A. *et al.* Complexes between the LKB1 tumor suppressor, STRADα/β and MO25α/β are upstream kinases in the AMP-activated protein kinase cascade. *J. Biol.* **2**, 28 (2003).
25. Suter, M. *et al.* Dissecting the role of 5'-AMP for allosteric stimulation, activation, and deactivation of AMP-activated protein kinase. *J. Biol. Chem.* **281**, 32207–32216 (2006).
26. Bungard, D. *et al.* Signaling kinase AMPK activates stress-promoted transcription via histone H2B phosphorylation. *Science* **329**, 1201–1205 (2010).
27. Bolden, J. E., Peart, M. J. & Johnstone, R. W. Anticancer activities of histone deacetylase inhibitors. *Nature Rev. Drug Discov.* **5**, 769–784 (2006).
28. Kazantsev, A. G. & Thompson, L. M. Therapeutic application of histone deacetylase inhibitors for central nervous system disorders. *Nature Rev. Drug Discov.* **7**, 854–868 (2008).
29. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnol.* **26**, 795–797 (2008).
30. Bantscheff, M. *et al.* Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nature Biotechnol.* **29**, 255–265 (2011).

Author Contributions Y.-y.L. and J.D.B. designed the experiments, with help from J.-y.L. Y.-y.L. and Y.-h.C. performed the primary screens, with help from C.N.K. and C.-L.L. Validation and further functional categorization experiments were performed by Y.-y.L., S.-Y.L. and Z.K. S.K. and Y.S. performed computational analyses, supervised by R.I. and J.S.B., respectively. Y.-y.L., J.-y.L. and J.D.B. wrote the manuscript. All authors discussed results and edited the manuscript.

METHODS

Mammalian cell culture and treatment. All human cell lines were obtained from the American Type Culture Collection unless mentioned otherwise. HCT116, HEK 293T and HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM; GIBCO) containing 10% fetal bovine serum (FBS; GIBCO), 100 units ml^{-1} penicillin and 100 µg ml^{-1} streptomycin at 37 °C in a humidified atmosphere containing 5% CO_2. HepG2 and IMR-90 cells were cultured in minimum essential medium (MEM; GIBCO). The shRNAs (all obtained from TRC shRNA library) targeting GFP, firefly luciferase (both as control) or candidate genetic interaction partner genes of query KDACs were used for transfection of the packaging HEK 293T cells with helper vectors (psPAX2 and pMD2.G, Addgene) using Fugene 6 transfection reagent (Roche) according to the manufacturer's instructions. Medium containing lentiviral particles was harvested, filtered, aliquoted and stored at −80 °C. These viruses were used to transduce 10^6 cells in the presence of 8 µg ml^{-1} polybrene (Sigma-Aldrich). Transduced cells were selected in appropriate medium containing puromycin (Sigma-Aldrich). Knockdown efficiency was assayed using immunoblotting and quantitative PCR. The two shRNAs with the greatest knockdown effect for each query were cloned from pLKO.1 vector into the NdeI-BamHI sites of pLKO.1-hPGK-Neo vector (Sigma-Aldrich), and then packaged into lentivirus to transduce cells followed by selection in appropriate medium containing geneticin (GIBCO) to generate stable polyclonal query cell lines. The RNAi Consortium Numbers (TRCNs) and sequences of the shRNAs used to generate stable cell lines for primary screen and further characterization experiments in this study are: shLuciferase, CGCTGAGTACTTCGAAATGTC; shGFP, TACAACAGCCACA ACGTCTAT; shHDAC1#1 (TRCN0000004814), CGTTCTTAACTTTGAAC CATA; shHDAC1#5 (TRCN0000004818), GCTGCTCAACTATGGTCTCTA; shHDAC2#1 (TRCN0000004819), CAGTCTCACCAATTTCAGAAA; shHDAC2#3 (TRCN0000004821), GCCTATTATCTCAAAGGTGAT; shHDAC3#2 (TRCN00 00004825), CCTTCCACAAATACGGAAATT; shHDAC3#3 (TRCN0000004826), GCACCCAATGAGTTCTATGAT; shHDAC4#1 (TRCN0000004829), CGACT CATCTTGTAGCTTATT; shHDAC4#4 (TRCN0000004832), GCCAAAGAT GACTTCCCTCTT; shHDAC6#1 (TRCN0000004839), CATCCCATCCTGAA TATCCTT; shHDAC6#5 (TRCN0000004843), CCTCACTGATCAGGCCAT ATT; shHDAC7#2 (TRCN0000004845), GCCAGCAAGATCCTCATTGTA; shHDAC7#5 (TRCN0000004848), GCTACCATGTTTCTGCCAAAT; shHDAC8#2 (TRCN0000004850), GCATTCTTTGATTGAAGCATA; shHDAC8#3 (TRCN0 000004851), GCGTATTCTCTACGTGGATTT; shHDAC9#1 (TRCN000000 4854), CCTAGAATCTTTGTGAGGTTT; shHDAC9#5 (TRCN0000004858), GCAAAGATAGAGGACGAGAAA; shSIRT1#1 (TRCN0000018979), GCAAAG CCTTTCTGAATCTAT; shSIRT1#3 (TRCN0000018981), GCGGGAATCCAAA GGATAATT; shSIRT2#4 (TRCN0000040221), GCCAACCATCTGTCACTA CTT; shSIRT2#5 (TRCN0000040222), GCTAAGCTGGATGAAAGAGAA; shSIRT3#1 (TRCN0000038889), CCCAACGTCACTCACTACTTT; shSIRT3#4 (TRCN0000038892), GTGGGTGCTTCAAGTGTTGTT; shSIRT5#2 (TRCN00 00018544), GAGTCCAATTTGTCCAGCTTT; shSIRT5#4 (TRCN0000018546), CGTCCACACGAAACCAGATTT; shPRKAA1#1 (TRCN0000000857), GCATA ATAAGTCACAGCCAAA; shPRKAA1#2 (TRCN0000000859), CCTGGAAG TCACACAATAGAA; shPRKAA2#1 (TRCN0000002172), GCTGTGTTTATC GCCCAATTT; shp300#1 (TRCN0000009882), CAGACAAGCTTTGGCCATG GTA; shp300#2 (TRCN0000039883), CCTCACTTTATGGAAGAGTTA; shLKB1#1 (TRCN0000000407), GAGTGTGCGGTCAATATTTAT; shLKB1#2 (TRCN0000000408), GCCAACGTGAAGAAGGAAATT; shLKB1#3 (TRCN00 00000409), GATCCTCAAGAAGAAGAAGTT.

Construction of a customized half-hairpin microarray. A customized microarray was designed to contain replicated probes complementary to the 23-nucleotide target sequence that included the specific 21-nucleotide sense-strand sequence of each shRNA along with one nucleotide immediately flanking both 5′ (a C nucleotide) and 3′ (a G nucleotide) ends. The probes also contained a stretch of 60-nucleotide linker sequence to attach the slide surface, and were randomly distributed across the array. The customized microarray slides were synthesized by Agilent at a density of 4 × 180,000 (4 subarrays containing 180k probes each) and are publicly available with design AMADID 024081.

shRNA lentivirus pool transduction (primary screen). High-titre (>10^9 infectious units per ml) TRC human genome-wide shRNA lentivirus pools were acquired from Sigma-Aldrich and from the National RNAi Core Facility at Academia Sinica (AS RNAi core). Large-scale transductions were performed as previously described, with optimization[12]. A stable luciferase-shRNA-expressing HCT116 cell line was transduced in parallel with the stable query knockdown cells as control. 7.5 × 10^7 target cells (1,000× coverage) for each experiment were resuspended in 24 ml of DMEM containing 0.8 mg ml^{-1} geneticin and 8 µg ml^{-1} polybrene. The genome-wide lentivirus pool was added in an appropriate volume to achieve a m.o.i. of 0.3 according to the titre of transducing virus reported by the provider. This cell-virus mixture was then evenly split across one 12-well tissue culture plate for a spin transduction (centrifugation at 930g for 2 h at 30 °C). After spin transduction, the supernatants were aspirated and replaced by 2 ml fresh medium containing geneticin. The transduced cells were cultured overnight, and then cells of the entire plate were trypsinized and pooled, resuspended in 30 ml of fresh medium containing geneticin and transferred into one T225 flask. At day 3 after transduction, three quarters of the cells were taken from each flask as an initial 'benchmark sample'. The rest of each population was selected with puromycin to remove untransduced cells and propagated for an additional 18 doublings before the 'end sample' was taken.

shRNA half-hairpin probe production. Genomic DNA was purified from harvested cells according to the QIAamp Blood Maxi Kit protocol (Qiagen). The shRNA full hairpin coding sequence containing the 5′ end of the puromycin resistance marker gene was PCR amplified (with the following program: 94 °C for 5 min, 15 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 1.5 min, and a final step of 72 °C for 10 min) in a 600 µl solution containing ~80 µg of genomic DNA template, 200 µM dNTPs, 1 µM for each PCR primer (sequence, forward: 5′-TTCACCGAGGGCCTATTTCCCATG-3′, reverse: 5′-CGTGAGGAAGAGT TCTTGCAGCTC-3′), 5% DMSO, 1× ExTaq buffer, and 3 µl ExTaq (Takara). PCR products were purified using a MinElute PCR Purification kit (Qiagen). For each screen, the shRNA coding regions of the benchmark and end samples were further amplified in a 600 µl solution containing 3 µl of the puromycin marker-enriched amplicon, 200 µM dNTPs, 1× ExTaq buffer, 6 µl ExTaq (Takara), and labelled with Cy5 and Cy3 dyes (PCR primer sequence, forward: 5′-[Cy5/Cy3]-AATGGACTATCATATGCTTACCGTAACTTGAA-3′, reverse: 5′-TGTGGA TGAATACTGCCATTTGTCTCGAGGTC-3′), respectively, with the following PCR program: 95 °C for 5 min, 35 cycles of 94 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min, and a final step of 72 °C for 10 min. Immediately after the first round of PCR amplification, reaction volumes were doubled with the addition of PCR mixture without DNA template, and subsequently amplified with the following program: 95 °C for 7 min, 55 °C for 2 min, 72 °C for 60 min. Amplified full-hairpin DNA was further digested overnight into half-hairpins using XhoI, and the resultant half-hairpin probes were gel purified using a QIAquick Gel Purification kit (Qiagen).

Half-hairpin probe microarray hybridizations. A hybridization mixture for each sub-array was prepared as below. 500 ng each of Cy5- and Cy3-labelled, gel-purified half-hairpin probes were mixed together with 16.5 nmol of blocking oligonucleotide (blocking oligonucleotide sequence: 5′-GTCCTTTCCACAAGA TATATAAAGCCAAAATCGAAATA-3′). The mixture was denatured by heating to 95 °C for 5 min and transferred to ice for 5 min. Hybridization solution was added to the mixture containing a final concentration of 1× hybridization buffer (1 M NaCl, 100 mM Tris-HCl, pH 7.5, 0.5% Triton X-100) in a final volume of 110 µl; 100 µl of this was loaded to each sub-array and hybridized at 44 °C in a hybridization oven (Agilent) for 16 to approximately 20 h. The microarray was washed once in wash buffer I (6× SSPE: 0.9 M NaCl, 60 mM NaH$_2$PO$_4$, 6 mM EDTA, 0.05% Triton X-100), once in wash buffer II (1× SSPE: 150 mM NaCl, 10 mM NaH$_2$PO$_4$, 1 mM EDTA), spin dried and scanned using a G2565CA microarray scanner (Agilent).

Statistical analysis of microarray data. Microarray images were processed using Agilent Feature Extraction Software 10.7 (Agilent), and further statistical analysis was performed using customized software written in R. The resultant feature signal intensity data sets were normalized using a loess model without background subtraction to calculate $\log_2$(Cy5/Cy3) for the shRNA probes within each array, which represent the relative abundance changes of each shRNA between the initial and end time points. For each array, median and robust variance (mad) of $\log_2$(Cy5/Cy3) were computed. The $\log_2$ ratio from each shRNA in the array was standardized by subtracting the median and dividing the result by the robust variance of the array. To control for probe effects, we then subtracted the standardized $\log_2$ ratio of the control sample from that of the query sample. For each shRNA, the Z-score of the $\log_2$ ratio of the normalized Cy5 to Cy3 signal intensities was computed, and Z-score difference was calculated by subtracting the Z-score of the control sample from that of the query sample. Z-score differences larger than 1.5 and less than −1.5 were used as arbitrary thresholds to define candidate negative and positive genetic interactions, respectively. Genes with multiple shRNAs that met or surpassed these criteria were further confirmed by a cell viability assay described below.

Cell viability assay. Seven-hundred and fifty control luciferase shRNA cells or stable KDAC knockdown query cells were transduced in triplicate in 96-well plates with lentiviruses of target gene shRNAs or GFP and luciferase shRNA (as control) at a m.o.i. of ~40. Three shRNAs giving the best knockdown efficiency according to the TRC database were selected for each target gene. After transduction, the cells were selected with geneticin and puromycin. On day 7, viable cell number was measured using CellTiter-Glo reagent (Promega) according to the manufacturer's

instructions on an Infinite F500 microplate reader (Tecan), and the relative viability of any mutant in each plate was normalized against the viability of luciferase shRNA-transduced cells. Normalized viability ratios were obtained for each target gene (W_{QT}/W_T) and GFP control (W_Q/W_G), respectively. Here W_{QT} is the normalized fitness of query cells with target gene knockdown and represents the effect of double knockdown of both the query and target gene; W_T is the normalized fitness of control cells with target gene knockdown and represents the effect of single target gene knockdown; W_Q is the normalized fitness of query cells under shGFP treatment and represents the effect of single query gene knockdown; W_G is the normalized fitness of control cells under shGFP treatment and normally close to 1. These ratios were log-transformed and fit to a linear model to estimate the mean difference in viability ratios between knockdowns of target genes and GFP control in the query KDAC cell line compared to the control luciferase shRNA cell line. A mean viability ratio difference (epistasis coefficient) greater than 0.12 or smaller than −0.12 with the associated P value < 0.05 computed with one-way analysis of variance (one-way ANOVA) was assigned to the validated synthetic rescue interaction and synthetic lethality interaction, respectively. The threshold was chosen based on a stringent cut-off used in a recently published large-scale yeast genetic interaction database[17].

Hierarchical clustering analysis. Hierarchical clustering was performed using Cluster 3.0[31] for both queries and targets. Agglomerative hierarchical clustering builds clusters in a bottom-up fashion. We used the Pearson correlation coefficient to quantify similarity between the genetic interaction profiles of two query genes. When two clusters being joined contain m_1 and m_2 query genes respectively, the similarity score between them could be defined as a function of similarity scores between their individual components. We used the average linkage method, which defined similarity between two clusters as the average of $(m_1 \times m_2)/2$ pair-wise similarity scores between components of the two clusters.

GO enrichment analysis. The functional association of genetic interaction targets of query KDAC genes by GO enrichment analysis in the 'biological process' category was assessed using Protein Analysis Through Evolutionary Relationships (PANTHER)[32,33] and Funcassociated 2.0[34].

Enrichment analysis of co-repressor and acetylation among targets. Statistical significance (that is, P values) of co-repressor and acetylation enrichment among validated targets was calculated based on Fisher's exact test.

Enrichment analysis of co-occurring interactions. Among the total gene set under consideration (the KDAC queries and all tested targets), we assembled a set of known physical interactions by accumulating all physical interactions from BioGRID[14] and the Michigan Molecular Interactions databases[35]. This provided us with a set of 94,475 physical interactions. P values were calculated assuming a random graph model where physical interactions are randomly assigned while keeping the number of physical interactions for each gene the same as those actually observed in our compiled network. Thus, the probability of a physical interaction between two genes i and j is $\frac{d_i d_j}{2E_{tot}}$ where d_i and d_j represent the number of physical interactions observed for the genes and E_{tot} is the total number of observed physical interactions. If we consider that a query KDAC q with the validated set of genetic interaction targets G_q, the expected number of physical interactions between the query and its genetic interaction targets i is $\lambda = \sum_{i \in G_q} \frac{d_q d_i}{2E_{tot}}$. The P value is calculated with a Poisson model as $\sum_{k \geq \lambda_q} \frac{\lambda_q^k e^{-\lambda_q}}{\Gamma(k+1)}$.

Purification of GST-tagged proteins. Cells grown in ten 15-cm dishes were transiently transfected with vectors containing the GST-tagged substrate construct and then harvested. Whole-cell extracts were obtained by breaking cells in lysis buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1 mM EDTA, 0.1% v/v Triton X-100, 1 mM DTT, 1 mM PMSF, 5 μM pepstatin A, 1 μM MG-132, and EDTA-free complete protease inhibitor mix (Roche)), and incubated with 3 ml of glutathione sepharose beads (GE Healthcare) at 4 °C for 1 h with head-to-head rotation (~10 r.p.m.). After binding, the glutathione sepharose beads were washed four times with 10 ml of wash buffer I (500 mM NaCl, 50 mM HEPES, pH 7.0, 1 mM EDTA, 1 mM EGTA, 0.1% Triton X-100) followed by being washed four times with 10 ml of wash buffer II (50 mM NaCl, 50 mM HEPES, pH 7.0, 10% v/v glycerol, 0.1% Triton X-100, 10 mM NaOH). The washed beads were incubated with 3 ml of elute buffer (50 mM NaCl, 50 mM HEPES, pH 7.0, 25% v/v glycerol, 10 mM glutathione, pH 7.0) at 4 °C for 1 h with rotation. The eluate was collected by gravity flow or centrifugation, and the eluted protein was concentrated to a final concentration of 0.1 to approximately 0.5 μg μl^{-1} by ultrafiltration with Vivaspin 500 concentration columns (Sartorius). The final protein concentration was determined by a Nanodrop analyser using the A_{280}.

Purification of Flag-tagged protein complexes. The purification of Flag-tagged proteins was performed as described[36] with minor adjustments. Protein concentration of the extracts was determined at 280 nm on a Nanodrop analyser and 2 mg

protein was used for immunoprecipitation in IP150 buffer (25 mM Tris pH 7.5, 150 mM NaCl, 1.5 mM DTT, 10% glycerol, 0.5% v/v NP-40) supplemented with protease inhibitors (for HDAC1 purification) as well as KDAC and phosphatase inhibitors (for PRKAA1 and LKB1 purification). An equimolar amount of STRADA expression plasmid was concomitantly transfected with Flag-tagged LKB1. Flag-tagged proteins were then captured with M2 anti-Flag antibody-conjugated agarose (Sigma-Aldrich). The immunoprecipitate was then subjected to immunoblotting to detect the signal of phosphorylation, acetylation and binding proteins. For HDAC1, AMPK or LKB1 complex purification, the immuno-precipitate was eluted with 50 μg ml^{-1} of Flag peptide solution. The eluate was spin-concentrated and the final protein concentration was determined by a Nanodrop analyser based on the A_{280}.

In vivo acetylation assay. To detect acetylation of endogenous proteins, the protein of interest was immunoprecipitated from whole-cell extracts, and the acetylation signal of the immunoprecipitate was assessed by immunoblotting.

In vitro acetylation assay. In vitro HAT reactions were performed for 1 h at 30 °C in a 25 μl reaction mixture containing ~5 μg of GST-tagged PRKAA1 protein (wild type or 3KR), 0.25 μCi of [^{3}H]-acetyl-CoA (GE healthcare, 250 μCi ml^{-1}, 3.4 Ci mmol^{-1}), 100 μM TSA, 5 mM nicotinamide, 5 mM PMSF and 5 mM DTT in pCAB buffer (50 mM HEPES/NaOH, pH 7.9, 0.1 mM EDTA, 50 μg ml^{-1} BSA), and 1 μg of purified p300 protein (Enzo Life Sciences). The acetylated species were then analysed by scintillation counting or immunoblotting.

In vitro deacetylation assay. Flag-tagged HDAC1 was purified as described from *HDAC2* knockdown cells to minimize effects of co-purified HDAC2 proteins, and purified SIRT1 was acquired from Enzo Life Sciences. About 5 μg of GST-tagged substrate proteins purified from cells with stable knockdown of the corresponding KDAC gene were subjected to 0.5 to 1 μg purified KDACs at 30 °C for 1 h. 1 mM NAD$^+$ was added as a cofactor in the SIRT1 reactions. The residual acetylation signals of the substrate proteins were analysed by immunoblotting.

In vitro kinase assay. Phosphorylation of AMPK by LKB1 was performed for 30 min at 30 °C in a 25 μl reaction mixture containing ~5 μg GST-tagged PRKAA1 protein in kinase buffer (50 mM Tris-HCl, pH 7.5, 10 mM MgCl$_2$, 1 mM DTT, 100 μM ATP), and 0.5 μg of purified LKB1 (wild type or kinase dead). The phosphorylation signals of the substrate proteins were analysed by immunoblotting.

AMPK kinase activity assay. AMPK activity was assessed using the CycLex AMPK Kinase Assay Kit (CycLex) according to the manufacturer's instructions. Briefly, immunoprecipitated Flag-tagged AMPK was added to a plate precoated with a substrate peptide corresponding to mouse insulin receptor substrate-1 (IRS-1) and incubated for 30 min at 30 °C. Kinase activity was measured spectro-photometrically at 450 nm to monitor the level of phosphorylation of serine 789 in IRS-1 peptide.

Immunoblotting. WCEs were denatured in boiling SDS sample buffer, resolved by SDS–PAGE, transferred to nitrocellulose or PVDF membranes and probed with specific primary antibodies: anti-HA (F-7), sc-7392, Santa Cruz; anti-α-tubulin, T5168, Sigma-Aldrich; anti-GST, AB3282, Millipore; anti-Flag, F3165, Sigma-Aldrich; anti-acetyl-lysine, 05-515, Millipore; anti-acetyl-lysine, ICP0380, Immunechem; anti-HDAC1, ab7028, Abcam; anti-HDAC2, ab7029, Abcam; anti-HDAC3, ab7030, Abcam; anti-HDAC4, SA-404, BioMol, anti-HDAC6, 07-732, Millipore; anti-HDAC7, 07-937, Millipore; anti-HDAC8 (E-5), sc-17778, Santa Cruz, anti-SIRT1 (H300), sc-15404, Santa Cruz; anti-SIRT2, 09-843, Millipore; anti-SIRT3 (C73E3), 2627S, Cell Signaling; anti-PRKAA1, ab32047, Abcam; anti-PRKAA2, GTX103487, GeneTex; anti-PRKAA P-T172, 2535, Cell Signaling; anti-PRKAG1, GTX101661, GeneTex; anti-PRKAG2, GTX114178, GeneTex; anti-ACC, 04-322, Millipore; anti-P-ACC, 07-303, Millipore; anti-p300, 05-257, Millipore; anti-LKB1, ab15095, Abcam.

Real-time PCR with reverse transcription. Total RNA was extracted from one 10-cm dish of ~95% confluent cells using TRIzol (Invitrogen) according to the manufacturer's protocol. Complementary DNA (cDNA) was synthesized from 400 ng of DNA-free total RNA using SuperScript^{TM} III reverse transcriptase and random hexamer primers (Invitrogen), and then used for PCR with reverse transcription (RT–PCR) using SYBR Green with gene-specific primers on a 7900 Real-Time PCR System (Applied Biosystems). The relative mRNA amount of target genes transcribed was quantified by comparing the fluorescence of their PCR products with the fluorescence of *ACTB* as the reference (ΔC_T), and the difference between the two ΔC_T values ($\Delta\Delta C_T = \Delta C_T(WT) - \Delta C_T(mutant)$) was calculated to determine the effect of knockdown on the mRNA level of target genes. All RT–PCR experiments were performed using three biological replicates.

Oil-Red-O stain. Oil-Red-O staining of HepG2 cells was performed as previously described[37] with optimization. Cells were washed with ice-cold 1×PBS, fixed with 10% formalin for 60 min, and stained with Oil-Red-O working solution (1.8 mg ml^{-1} of Oil-Red-O in 6:4 isopropanol:water solution) for 60 min at 25 °C. After staining,

cells were washed with water to remove any remaining dye. For quantification of Oil-Red-O staining, the cell-retained dye was extracted by isopropanol and the content was measured spectrophotometrically at 500 nm.

Software. Microarray data were analysed by R version 2.10.0. Hierarchical clustering results were visualized by Java Treeview version 1.1.0[38]. Networks were created with Cytoscape version 2.4.1. Statistical analysis was performed and plotted using GraphPad Prism 4 (GraphPad).

31. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95,** 14863–14868 (1998).
32. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13,** 2129–2141 (2003).
33. Mi, H. *et al.* PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.* **38,** D204–D210 (2010).
34. Berriz, G. F., Beaver, J. E., Cenik, C., Tasan, M. & Roth, F. P. Next generation software for functional trend analysis. *Bioinformatics* **25,** 3043–3044 (2009).
35. Jayapandian, M. *et al.* Michigan Molecular Interactions (MiMI): putting the jigsaw puzzle together. *Nucleic Acids Res.* **35,** D566–D571 (2007).
36. Nakatani, Y. & Ogryzko, V. Immunoaffinity purification of mammalian protein complexes. *Methods Enzymol.* **370,** 430–444 (2003).
37. Hwang, J. T. *et al.* Genistein, EGCG, and capsaicin inhibit adipocyte differentiation process via activating AMP-activated protein kinase. *Biochem. Biophys. Res. Commun.* **338,** 694–699 (2005).
38. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20,** 3246–3248 (2004).

THE CHANGES THAT COUNT

As more mutations are found across the genome, geneticists are focusing on learning which ones are likely to cause human disease, and how.

ALENGO/ISTOCKPHOTO

BY MONYA BAKER

Even before the first draft of the human genome was complete, researchers knew that one genome wouldn't be enough. They needed sequence data from many individuals to reveal the mutations that make people different and sometimes make them ill. Now, tens of thousands of people have had their genomes fully or partially sequenced. Each person's genome contains an average of more than 3 million variants, or differences from the reference genome. A partial sequence, focusing on the 1.5% of the genome that codes for proteins, usually has about 20,000.

For the most part, scientists don't know what those variants do. "The ultimate goal is to sequence a person's genome and make credible predictions just given the list of variants," says Greg Cooper, a genomicist at the Hudson-Alpha Institute for Biotechnology in Huntsville, Alabama. "We're a really long way from that."

Scientists have sorted through the most common variants, using genome-wide association studies to learn which occur more often in people with disease, but these variants tend to have small effects, with the biology behind those effects largely unknown. And as techniques that use sequencing to identify genetic variation become cheaper and more reliable,

more rare variants are being uncovered. That is changing the questions that researchers are asking, says David Goldstein, director of the Center for Human Genome Variation at Duke University in Durham, North Carolina. "The field will transition from doing primarily association work to figuring out what implicated variants do biologically."

Disparate strands of research are coming together to do exactly that. A host of increasingly sophisticated algorithms predict whether a mutation is likely to change the function of a protein, or alter its expression. Sequencing data from an increasing number of species and larger human populations are revealing ▶

which variants can be tolerated by evolution and exist in healthy individuals. Huge research projects are assigning putative functions to sequences throughout the genome and allowing researchers to improve their hypotheses about variants. And for regions with known function, new techniques can use yeast and bacteria to assess the effects of hundreds of potential mammalian variants in a single experiment.

ALIGNMENTS AND ALGORITHMS

Many bioinformatics tools rely on evolution to rate how likely a variant is to be harmful. Most focus on identifying the 'non-synonymous' mutations that alter the amino acids that make up the proteins for which genes code. It is expected that the more species have evolved with a certain amino acid in a certain place, the more likely a change is to be harmful. "The idea is that evolution has tested it and that's why you don't see that mutation," says Pauline Ng, a genomicist at the Genome Institute of Singapore. Ng co-wrote an algorithm called SIFT (sorting intolerant from tolerant; http://sift-dna.org), one of the first programs for predicting the effects of protein changes and still one of the most popular. It was originally designed to evaluate one gene at a time, but Ng has updated the protocol to accommodate genomic data files produced by sequencing analyses.

The algorithm first identifies mutations that affect highly conserved amino acids, then predicts whether a particular change is likely to be harmful. To train it for such assessments, Ng used published data that assessed amino-acid changes in a well-studied bacterial protein. That showed how often a change from one particular amino acid to another altered protein function. When researchers run SIFT on their sequencing data, the algorithm uses evolutionary conservation and patterns inferred from that original data set to evaluate whether mutated human proteins are likely to behave in similar ways to their non-mutated counterparts.

Another popular algorithm is Poly-Phen (prediction of functional effects of human non-synonymous single-nucleotide polymorphisms; http://genetics.bwh.harvard.edu/pph2), which was co-written by Shamil Sunyaev, a geneticist at Harvard Medical School in Boston, Massachusetts. This algorithm, too, uses evolutionary data in its predictions, but it also incorporates biochemical predictors of stability and spatial structure. Sunyaev trained it using single-gene mutations that are known to cause diseases, reasoning that they did so by disabling proteins.

Stephanie Hicks and Marek Kimmel, statisticians at Rice University in Houston, Texas, were part of a team that evaluated[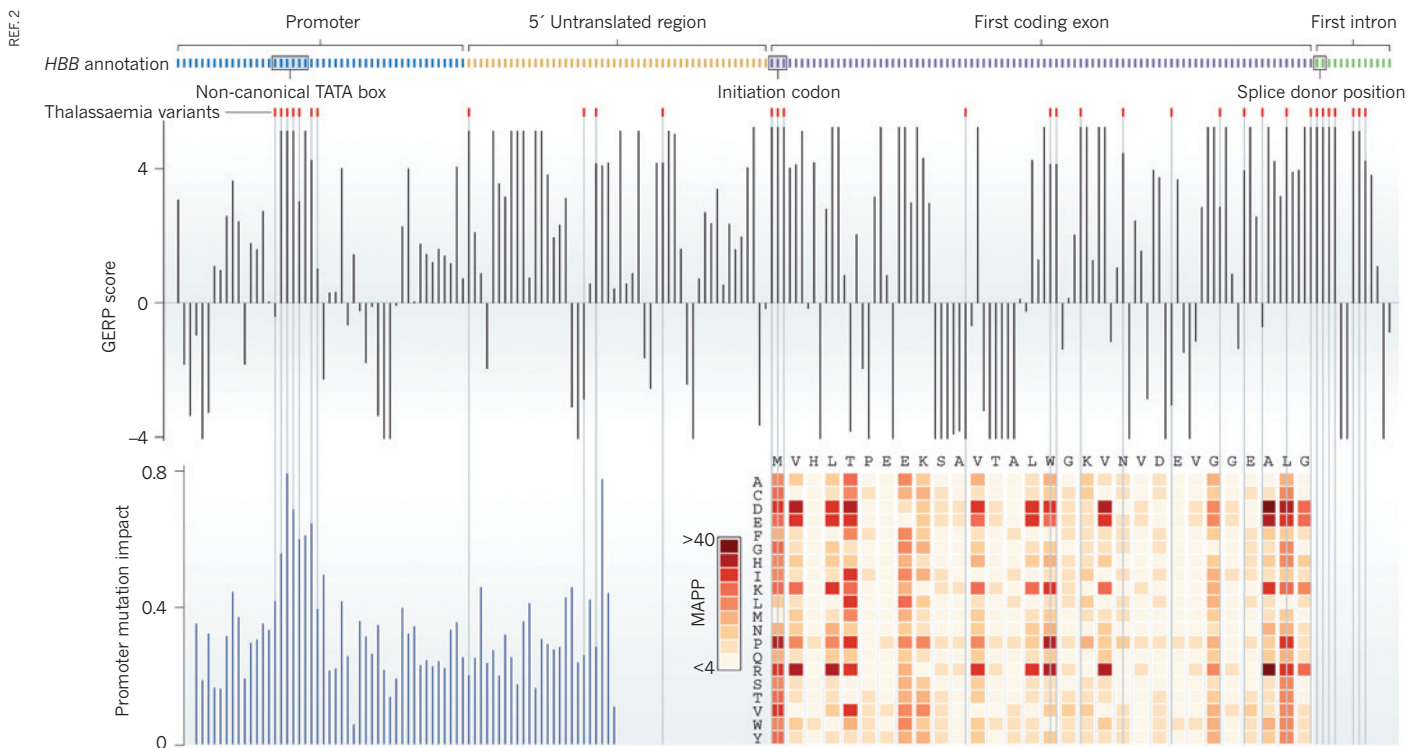1] the abilities of 4 popular algorithms to predict the effects of 267 well-understood 'missense mutations', which swap one amino acid for another. The algorithms all had accuracies of about 80%. However, even when working from the same 'alignment data' — comparisons of protein sequences — the algorithms made different predictions about the same set of proteins. And Kimmel cautions that algorithms may perform less effectively with mutations that aren't well-known.

Even if algorithms were 100% accurate, knowing that a variant causes a protein to lose function is a very long way from knowing whether it contributes to disease, says Sunyaev. The effects of loss-of-function mutations can be surprisingly minimal, buffered by redundancies in cellular machinery. Algorithms alone are certainly not good enough for clinical diagnostics, he says, and he frets that some clinicians are starting to take an interest in these scores. "This is how I lose sleep at night."

MORE THAN MISSENSE

Even if their predictions were perfect, algorithms that focus on protein sequences would miss many variants that potentially cause disease. Evolutionary analyses indicate that natural selection has conserved five times more base pairs that don't code for proteins than ones that do, which implies that these sequences have some sort of function, even if that is not yet obvious — and mutations in these genomic regions could therefore have a biological effect.

Researchers have now introduced computational tools that use evolution to rank variants in non-coding regions[2]. These include GERP

> *"The field will transition from association work to figuring out what variants do biologically."*
> David Goldstein

Analysis of variants in coding and non-coding regions of part of a haemoglobin gene (*HBB*). The variants marked in red cause the blood disorder thalassaemia.
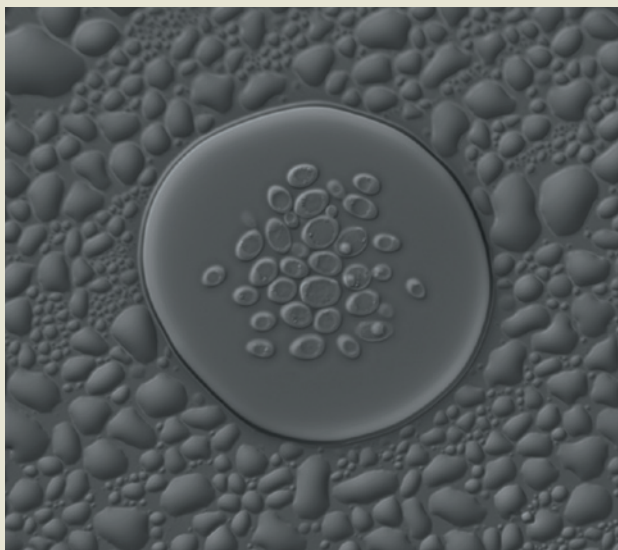
Variants in context

The key to human individuality may not be genetic variants so much as the interactions between them. Consider this example: in 2002, a knockout library of more than 5,000 genes from yeast (*Saccharomyces cerevisiae*) found about 1,000 that the microbes literally can't live without[10]. In 2010, Charles Boone, a molecular geneticist at the University of Toronto, Canada, and Gerald Fink and David Gifford, molecular geneticists at the Broad Institute of the Massachusetts Insitute of Technology and Harvard in Cambridge, made a similar library[11] using a second strain of the same species. Startlingly, dozens of 'essential genes' were unique to one strain or the other. And the strains are about as similar to each other genetically as individual humans.

The implications of such studies are frightening, says David Goldstein, director of the Center for Human Genome Variation at Duke University in Durham, North Carolina. "It means that the whole concept of whether variants are pathogenic is not well formulated. When we consider pathogenicity at the level of the individual, we don't always know what we're talking about."

Indeed, researchers have very strong ideas about what contributes to pathogenicity at the population level, but don't necessarily know how to translate them to the individual. Last year, a team at the University of Geneva, Switzerland, characterized[12] an often-overlooked type of interaction using established cell lines for many individuals and data from the international 1000 Genomes Project. "We asked: 'If there was a deleterious coding variant, how likely is it that there is a regulatory variant modifying that effect?'" says Stephen Montgomery, a member of the team and now a geneticist at Stanford University in California. The answer is, pretty likely. For nearly half of the coding variants that the researchers examined, they also found at least one individual who expressed the gene at atypical levels, a situation that

Yeast cells can be used to demonstrate how the effects of one genetic variant depend on those of other variants.

could decrease or increase levels of a pathogenic protein and perhaps affect the course of disease.

Researchers at the University of Nottingham, UK, and the Wellcome Trust Sanger Institute in Hinxton, UK, mated a heat-tolerant yeast strain that normally grows on tree bark with a heat-sensitive strain used to make palm wine[13]. They bred the progeny for 12 generations, giving variants on the same chromosome many chances to shuffle and reassort. That helped them to pinpoint loci — defined regions on a chromosome — that contain variants that help yeast to survive. They found around 20 loci containing variants that boost yeast's ability to withstand heat — but surprisingly, one-third of these loci originated in the heat-sensitive strain. "Once you put those mutations in a random background, then you can see their positive effect," says Leopold Parts, first author of the study and now a postdoc at the University of Toronto.

Leonid Kruglyak, a geneticist at Princeton University in New Jersey, has found a way to combine high-throughput genotyping with yeast mating to work out how many spots on a genome contribute to a trait[14]. He says that attributing disease heritability to multiple common variants that each have small effects just doesn't add up. "If you project from the numbers that are being reported," he says, "you end up with preposterous numbers, multiple variants for every single gene in the genome." It will take empirical work to learn the relative importance of common variants, rare variants and the interactions between them, he says.

The problem is that biological experiments are set up to get information about averages, not individuals, says Ben Lehner, a systems biologist at the Center for Genome Regulation in Barcelona, Spain, who is studying how yeast-sequencing data can be used to predict phenotypes. "We talk about the typical effect of an allele in the population, but that is not useful if you want to find out what that means for an individual," he says. **M.B.**

(genomic evolutionary rate profiling; http://mendel.stanford.edu/SidowLab/downloads/gerp) and phastCons (phylogenetic analysis with space/time models, conservationl; http://compgen.bscb.cornell.edu/phast). Like algorithms that assess protein-coding genes, they evaluate variants on the basis of how often the sequence changes between species. However, because non-coding regions evolve very quickly, sequences can be compared only among mammals. "Even if you go to chickens, nearly all the non-coding stuff won't align," says Cooper, who co-wrote GERP.

And it is not always clear what the rankings mean. Because non-coding regions do not have a corresponding protein, rules regarding amino-acid changes are irrelevant, and there are no data sets appropriate for training such algorithms. "The evolutionary data we do have are informative, but it's early days, so you have to take them with a grain of salt," says Arend Sidow, a genomicist at Stanford University in California, who co-wrote GERP and other predictive algorithms. But algorithms for non-coding sequences can provide evidence that a mutation has an impact by looking at conservation, says Sidow. For example, if a child with a rare disease has an unknown mutation not shared by his or her healthy parents, a score indicating that the mutation is in an evolutionarily conserved region would encourage researchers to examine it more carefully in follow-up experiments.

Alternatively, researchers can consider the results of human-sequencing experiments. One algorithm, VAAST (variant annotation, analysis and search tool; www.yandell-lab.org/software/vaast.html), received a lot of attention last year when researchers used it[3] on just two newly sequenced genomes to pinpoint the mutation that causes Ogden syndrome, a fatal condition linked to the X chromosome in males. The algorithm was also able to re-identify single genes already known to cause some conditions and implicated in more complex diseases[4].

VAAST was developed by Mark Yandell, a geneticist at the University of Utah in Salt Lake City, and Martin Reese, chief executive of genetic-analysis company Omicia in Emeryville, California. It is different from other predictive algorithms that focus on protein-coding and non-coding regions, says Yandell. "Instead of saying, is this conserved?' The algorithm asks, 'How often do we see humans with these variants?'" Unlike many other algorithms, which score each variant as 'probably harmful' or 'probably benign', VAAST provides a ranked list of which variants are most likely to contribute to disease.

"The idea is that evolution has tested it and that's why you don't see that mutation."
Pauline Ng

The algorithm integrates many sources of information: whether a variant has been observed in healthy individuals; whether it occurs in a known functional region; and, for protein-coding variants, what its functional impact is expected to be. When working out whether a single gene is likely to contribute to a condition, it also looks at all the variants that occur in that gene throughout the surveyed population. "You dump all the variants for each gene into a bucket and then see which bucket has the most likely damaging variants. That goes to the top of the list," says Yandell. Future iterations of the algorithm, he says, will consider variants in genes that are associated with common biological pathways.

VAAST is just one in a wave of algorithms to incorporate human-sequencing data. Another is ANNOVAR (http://www.openbioinformatics.org/annovar), which was developed at the Children's Hospital of Philadelphia in Pennsylvania. Knome, a genetic-analysis company in Cambridge, Massachusetts, provides informatics and services for interpreting genomes, and Softgenetics in State College, Pennsylvania, and GenomeQuest, in Westborough, Massachusetts, pluck out variants that might affect patients' health.

But the results of such algorithms can't be trusted without further verification. Predictive algorithms can tell researchers which variants should be flagged up for follow-up studies, but not which ones cause disease, says Cooper. "The best we can do computationally is to prioritize things. It's still going to be a lot of work to nail it."

And there are few ways to assess predictive algorithms, particularly those that go beyond evaluating missense mutations, says John Moult, a bioinformatician at the University of Maryland in Rockville. Moult is one of the co-organizers of the Critical Assessment of Genome Interpretation, a contest in which bioinformatics teams compete to predict a phenotype — an organism's characteristics — from genetic data. Of 13 teams that competed last year, only 2 tried to predict how nucleotide sequences might affect gene expression and splicing.

But the field is still young, says Moult. For algorithms to improve, researchers will need more data — and the data are coming, he says. Not only are more genomes being sequenced, but researchers are working out protocols to share data without compromising patient privacy. Last year, the contest could provide data for only ten whole genomes. This year, Moult expects data for 500.

EXPERIMENTS REQUIRED

Laboratory experiments are essential for verifying the effects of variants, but with so many new variants cropping up, there is currently no way to test them all. "What we need are functional approaches that have a bit of the feel of genomics," says Goldstein. "They need to be scalable; they need to be applied if not to every variant, at least to an awful lot of variants." In particular, Goldstein wants to know whether a variant associated with a gene affects RNA splicing or transcription rates. To find out, he is collecting genome-wide gene-expression data alongside sequencing data. That allows him to find out whether genetic variants correlate with changes in messenger RNA. "It's an affordable additional expense," he says.

Other researchers are developing high-throughput techniques for testing protein variants. Just changing one amino acid at a time, a protein containing 1,000 amino acids would have 19,000 variants. In the past, variants had to be tested individually or in small batches, limiting assays to a few hundred. New methods allow the testing of hundreds of thousands at a time.

Stan Fields, a molecular geneticist at the University of Washington in Seattle, is designing assays that exploit the basic principle of natural selection. He places many variants of a protein-coding gene into viruses or cells that depend on the protein variants that they produce to grow and reproduce, allowing him to interrogate characteristics such as the protein's stability, structure, enzymatic activity and interaction with other proteins. Sequencing can log which variants become more common and which become less so over several generations. "You can come up with all sorts of assays," says Fields, "and the answer comes down to a simple sequence run."

With his postdoc Doug Fowler, Fields has demonstrated[5] that this approach, called deep mutational scanning, can be used to assess the binding activity of hundreds of thousands of variants of the WW domain, a stretch of 40 amino acids that is found in many human proteins and is often important in protein–protein interactions. Fields and Fowler are working out ways to analyse the residues that contribute to protein function, and so learn about general principles of protein design. Fowler is also using the technique to assess which mutations confer drug resistance on Src-kinase, an enzyme implicated in cancer.

It should be possible eventually to assess all the single-amino-acid mutations that could occur in important genes, says Fields. "Then if someone shows up with any mutation, you can say: 'Looking at that particular protein activity, we know what the mutation means.'"

Last year, Dan Bolon, a protein biochemist at the University of Massachusetts Medical School in Worcester, described[6] a similar approach, which he calls EMPIRIC (extremely methodical and parallel investigation of randomized individual codons). He and his colleagues used this technique to test every possible point mutation in a short stretch of Hsp90, a protein that is necessary for yeast growth. The team examined some 500 genetic changes that collectively encoded 180 protein variants. After growing yeast for several generations, Bolon could see which variants enabled the fastest growth, by measuring which showed up the most often in sequencing data. Previous approaches would have required one-by-one testing, but Bolon's method evaluated all the variants at once. An experiment that would normally have taken years was completed in days.

Bolon found that about 15% of amino-acid substitutions that never occurred in evolution grew just as well in his experiments as the wild type, perhaps because effects of those substitutions were too small to matter over the tested time frame, or were irrelevant under the test conditions. Evolution eventually removes both lethal and slightly deleterious variants, but a variant that has an effect only over many generations might make little difference to an individual.

"The best we can do computationally is to prioritize things. It's still going to be a lot of work to nail it."
Greg Cooper

As well as providing direct information on particular proteins, such subtle analyses could be used to train algorithms and improve their accuracy, says Peter Good, programme director for genome informatics at the US National Human Genome Research Institute (NHGRI) in Bethesda, Maryland.

Both Bolon and Fields expect rapid increases in the number and complexity of variants that can be assessed. Bolon is able to vary 100 amino acids at once, the entire length of some small proteins. Already, he can imagine testing all protein variants within small viral genomes. "The ability to look at systematic libraries

across an entire genome is just very exciting in terms of understanding the raw evolutionary basis for an entire organism," he says.

Such sequencing approaches can also be applied to regulatory elements. Knowing that a mutation changes a transcription-factor binding site says nothing about how it will affect the binding of the gene-activating protein, says Gary Stormo, a molecular biologist at Washington University School of Medicine in St Louis, Missouri. The protein may bind just as well as without the mutation; hardly at all; slightly worse; or even slightly better. So Stormo has created experimental systems that link transcription-factor binding to cell proliferation. The cells that grow best are those that contain the best-binding DNA, and next-generation sequencing is allowing a more systematic exploration of more variants than ever before. Only two or three years ago, scientists would manually pick 20–50 of the fastest-growing colonies to examine, says Stormo. "We now just scrape the whole plate. You can get millions of examples in a single experiment." Even better, with that many samples, researchers can derive quantitative data, and so show how much better the best-binding sites are.

However, *in vitro* results are far from perfect in predicting *in vivo* binding, says Stormo. "Some of the best sites won't be bound, and there will be binding to other places that you wouldn't expect." The good thing is that differences observed between test tubes and living cells indicate interesting biology. "That tells you we're missing a lot of information, and that's what we want to figure out," says Stormo (see 'Variants in context').

DECODING REGULATORY ELEMENTS

Before they can work out what a variant might do, researchers need to learn whether it occurs in an active part of the genome. Several genome-wide studies are providing crucial clues. The NHGRI's ENCODE project (Encyclopedia of DNA Elements) hopes to map and annotate all functional elements in the genome, and the International Cancer Genome Consortium is mapping genomic changes in cancer. The International Human Epigenome Consortium and the US National Institutes of Health's Roadmap Epigenomics Mapping Consortium are studying features such as DNA methylation and other modifications across the genome in many types of cell, and so are showing which regions of the genome might be functional in particular tissues. Annotation alone will not demonstrate that a variant is pathogenic, but the information can help researchers to design the right experiment, says Good. "The question is knowing why it's pathogenic, that's where the annotation helps you. It's a big difference to say, 'this variant affects a protein-coding region or a promoter active in particular cell types.'"

In work[7] funded by these consortia, Manolis Kellis, a computational biologist at the

Massachusetts Institute of Technology (MIT), along with Bradley Bernstein, a pathologist at Harvard Medical School, and their colleagues, mapped 'chromatin states' — sets of chemical modifications to DNA and DNA-binding proteins that distinguish genomic regions. The location of these states varies across different cell types and is correlated with gene expression. By comparing chromatin states on gene promoters, enhancers and other regulatory regions with data on gene expression, the researchers linked regulatory elements to target genes.

The team then cross-referenced chromatin states with variants that had been associated with specific diseases. This revealed patterns that made sense: for example, variants that had been statistically associated with leukaemia occurred in what chromatin states revealed to be enhancer regions active in leukaemia cells. Similarly, variants thought to affect lipid and triglyceride levels in blood were found in regulatory elements active in liver cells.

> *"When we are talking about synonymous changes, we can no longer think of them as neutral."*
> Manolis Kellis

Other mapping projects rely on comparative genomics. Last year, researchers based at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, completed whole-genome sequencing of 20 mammalian species, then analysed[8] these sequences along with those of 9 other mammals that had already been sequenced. This revealed more than 3.5 million evolutionarily constrained elements in the human genome, up from a few hundred thousand that had been previously identified. Still, only about 60% of these could be assigned any putative function. Most of the new elements were located either between genes or in non-coding parts of genes.

Furthermore, even nucleotides in protein-coding genes that would not alter amino acids were under evolutionary constraint, and further analysis suggests that these sites affect RNA-transcript processing, microRNA binding and how chromatin states are established[9]. "When we are talking about synonymous changes, we can no longer think of them as neutral," says Kellis, who was part of the study.

And more regulatory elements are being revealed. Scores of researchers have noticed that non-conserved areas of the genome have activities associated with function. Many such regions are transcribed; others host various DNA-binding proteins. One-half to one-third of 'biochemically active' elements are unique to humans, says Ewan Birney, a bioinformatician at the European Bioinformatics Institute

in Hinxton, UK. When the number of these active, non-coding elements was first discovered, their activity was dismissed as an experimental artefact, then discounted as irrelevant noise. But unpublished work shows that many of these regions are in fact evolutionarily conserved in the human population, presumably because they have a function that helps individuals to survive and reproduce.

Of course, changes to evolutionarily conserved sequences do not necessarily contribute to disease, says Birney. But researchers should start thinking about what variation in regulatory regions might do. Six months ago, the Variant Effect Predictor (VEP) tool went live on the Ensembl Genome Browser (www.ensembl.org), which brings together information from several databases, including human-sequencing projects and chromatin signatures across cell types. The tool shows, for example, whether a mutation affects a site that binds known transcription factors.

Other tools are also coming online. Michael Snyder, a geneticist at Stanford, is developing RegulomeDB (www.regulomedb.org), which identifies binding sites and other elements in non-coding DNA. This January, Kellis introduced HaploReg (www.broadinstitute.org/mammals/haploreg/haploreg.php), which brings together data from chromatin-mapping and comparative-genomics studies. Researchers can enter common variants and see whether they fall in a highly conserved region, disrupt a regulatory motif or are associated with a regulatory element in a particular cell type. It provides the same information for common variants that tend to be inherited along with the ones entered.

This is just the beginning of efforts to assign functions to the millions of DNA variants. In time, says Kellis, it will help researchers to pin down the mechanisms that cause disease. "The marriage of human genetics and functional genomics can deliver what the original plan of the human genome promised to medicine." ∎

Monya Baker *is technology editor for* Nature *and* Nature Methods.

1. Hicks, S., Wheeler, D. A., Plon, S. E. & Kimmel, M. *Hum. Mutat.* **32,** 661–668 (2011).
2. Cooper, G. M. & Shendure, J. *Nature Rev. Genet.* **12,** 628–640 (2011).
3. Rope, A. F. *et al. Am. J. Hum. Genet.* **89,** 28–43 (2011).
4. Yandell, M. *et al. Genome Res.* **21,** 1529–1542 (2011).
5. Fowler, D. M. *et al. Nature Meth.* **7,** 741–746 (2010).
6. Hietpas, R. T., Jensen, J. D. & Bolon, D. N. A. *Proc. Natl Acad. Sci. USA* **108,** 7896–7901 (2011).
7. Ernst, J. *et al. Nature* **473,** 43–49 (2011).
8. Lindblad-Toh, K. *et al. Nature* **478,** 476–482 (2011).
9. Lin, M. F. *et al. Genome Res.* **21,** 1916–1928 (2011).
10. Giaver, G. *Nature* **418,** 387–391 (2002).
11. Dowell, R. D. *et al. Science* **328,** 469 (2010).
12. Lappalainen, T., Montgomery, S. B., Nica, A. C. & Dermitzakis, E. T. *Am. J. Hum. Genet.* **89,** 459–463. (2011).
13. Parts, L. *et al. Genome Res.* **21,** 1131–1138 (2011).
14. Ehrenreich, I. M. *et al. Nature* **464,** 1039–1042 (2010).

CAREERS

D. BISHOP

BY ERIKA CHECK HAYDEN

David Alexander's job didn't exist ten years ago. He works for Pacific Biosciences in Menlo Park, California, writing software that can analyse the data generated by DNA polymerase enzymes, which sequence DNA in real time. A decade ago, it took scientists weeks to sequence DNA, one base at a time, using a seemingly endless series of reactions. Back then, they also thought that they would be able to find the roots of major diseases just by identifying the common genetic variants shared by affected individuals.

Both the technology and the hypotheses have changed greatly since then. In the mid- to late-2000s, while Alexander was working towards his PhD, scientists were using genome-wide association studies (GWAS) — searching genomes for known genetic variants that are shared by people with a particular disease or trait. But by the time he graduated, last June, GWAS had mostly been superseded by techniques that sequence entire genomes. The machines designed to do this sequencing are pouring out huge amounts of data, thereby creating a huge need for mathematics and statistics experts. So Alexander, and many others working on statistical genetics, now have many more opportunities. "Scientifically, there are much richer questions to ask, and there are still a lot of deep discoveries to be made; it's an interesting time," he says. His career track reveals just how much opportunities in the field have changed.

CAREER VARIATION

It was not for a lack of trying that GWAS didn't pan out. The completion of the Human Genome Project in 2003 spurred major funders from around the world to invest millions of dollars to build an international haplotype map, a catalogue of all the common human variants at single bases, called single nucleotide polymorphisms (SNPs), to be used in GWAS. The SNP map should have helped researchers to identify genes that are associated with disease. But instead, it showed that SNPs don't account for much of the heritability of disease.

Researchers now think that many rare variants play a part in causing disease, but rare variants are much harder to find than the common SNPs. As a result, statistical geneticists are now mining sequence data for directly ▶

BIOSTATISTICS

Revealing analysis

As the challenges of analysing genomic data evolve, statistical expertise has become more valuable than ever.

causative mutations, rather than for SNPs. And geneticists are starting to combine data from different types of studies, using a method called integrative genomics — for instance, studying combinations of SNPs, the protein-coding genes surveyed in exome studies, epigenetic factors (heritable information not found in the DNA sequence), gene-expression factors and environmental interactions. "This field has ballooned and changed to a ridiculous degree in the past ten years, because there have been multiple waves of technological revolution," says Gilean McVean, a statistical geneticist at the University of Oxford, UK. "As genomics becomes a much more integrated part of health care, things are going to change again and new opportunities will open up, so it's a good time to be a statistical geneticist."

BAG OF TRICKS

Statisticians will be kept busy for years by the problems raised by analysing these huge data sets. They will need to find the best ways to grapple with studies that combine multiple methods, each of which yield millions of data points. The challenge is to find true associations within the huge volumes of data without getting duped by the errors that tend to affect data sets of this magnitude, says Lucia Hindorff, an epidemiologist at the US National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. "The answers aren't straightforward," she says. "That's one of the reasons why statisticians have a lot of work to do." And statistical geneticists are needed at universities, at genome centres and in industry alike.

However, a survey of statistical geneticists by a working group from US National Institutes of Health in Bethesda has suggested that trainers are having difficulty recruiting enough qualified trainees into their programmes. Alexander Wilson, head of genometrics at the NHGRI, who organized the survey, says that although the

"There are much richer questions to ask, and there are still a lot of deep discoveries to be made."
David Alexander

number of genetic variants available to be analysed has grown significantly since the 1980s, the number of people available to analyse them has remained relatively constant. According to Suzanne Leal, a genetic epidemiologist at Baylor College of Medicine in Houston, Texas, many biologists eschew significant statistics training. And because only a handful of statistical geneticists are trained each year, "these positions are difficult to fill", says Michael Boehnke of the University of Michigan in Ann Arbor. So, although job demand outstrips supply in many fields, the market remains promising for statistics specialists, not least because they can help funding agencies to make good on their research investments.

And unlike other fields, many academic jobs in statistical genetics require only a doctoral degree, so PhD holders don't tend to find themselves stuck on an extended treadmill of multiple postdoc positions. "You're going to have many job opportunities; it's not like with other biological sciences where you do six or seven years of postdocs," Leal says. "You can do a two-year postdoc and then go on to a faculty position if you're any good."

With the plummeting cost of equipment, sequencing is becoming more feasible for many labs. However, the analytical problems are becoming so complex and expensive that disease-focused centres are starting to create joint analysis positions with larger hubs of genome expertise.

"Biology is now a science in which large data sets are central, but bioinformatics and statistical genetics are getting to a point where there are many specialized roles — data handling, processing, quality control, interpreting — that cannot all be done well by one person," says McVean. Analysts working on moving genomics technologies into health care at the University of Oxford's Biomedical Research Centre, for instance, are made honorary members of a bioinformatics and statistical genetics core at the Wellcome Trust Centre for Human Genetics in Oxford, run by McVean. They have access to the pipelines for sequencing data as well as to bioinformatics and statistical genetics expertise, but are funded separately from the centre.

Although statisticians in these positions can expect to have their own students and develop new methods, the roles are more inherently collaborative than many academic jobs, says McVean. "It's not the traditional academic route of going off to form your own little group and working in isolation, but rather going off to support diverse groups in a centre," he says. He is preparing to recruit for similar positions at the Ludwig Institute for Cancer Research and the Kennedy Institute of Rheumatology, both in Oxford. Both institutions, says McVean, would find it difficult to amass the personnel needed for independent, dedicated bioinformatics support.

Increased competition between new sequencing technologies — and companies hoping to make sense of the data — also means opportunities for computational and statistical experts in genetics in industry. Companies such as Pacific Biosciences, Illumina in San Diego, California, and Life Technologies in Carlsbad, California, are developing new methods for sequencing and need people who can come up with ways to analyse the new forms of data that will be produced.

Another track, which might be called clinical genomics, is relatively small, but growing. Companies in this field are developing ways to interpret individuals' genomic data for either medical or drug-discovery purposes, and are looking for individuals with a suite of talents. For instance, Omicia, based in the San Francisco Bay area of California, is developing a platform to help physicians and clinical labs to interpret genomic data. In just the past few months, it has hired three people: a Silicon Valley engineer who specializes in quick analyses of large data sets; an application engineer to help the company develop interfaces that are fast and easy for customers to use; and a medical researcher who has a bachelor's degree in genetics and hopes to attend medical school. Omicia's chief executive and co-founder, Martin Reese, says that the company is looking to hire more people in these specialities, especially analysts.

"This field has ballooned and changed to a ridiculous degree in the past ten years."
Gilean McVean

Rowan Chapman, a partner at Mohr Davidow, a venture-capital firm in Menlo Park that funds companies such as Pacific Biosciences, says that the firms are always looking for analysis experts. "There's a massive amount of data being generated, particularly by next-generation sequencing platforms, and the cost of the analysis is now greater than the cost of the data generation," she says. "Finding the right people to analyse those data is a challenge."

STRONG BACKGROUND

Succeeding in statistical genetics requires a good grounding in both statistics and genetics, which can be gained through academic work as part of any doctoral programme that allows students to take classes in both disciplines. But two other skills are increasingly necessary: expertise in computer-programming languages designed to aid manipulation of large data sets, such as R, Perl or Python, and the ability to use these languages to analyse large amounts of data quickly. Expertise in distributed computing and writing code for various operating systems is particularly desirable.

Most researchers say that these skills can be gained through hands-on experience working with large data sets, or during doctoral or postdoctoral work on a specific project. And that work doesn't have to be in biology. Stefano Lise, an analyst recently hired by the Oxford

R. C. GRIFFITHS

D. ADELE

Biomedical Research Centre, did his undergraduate, graduate and postdoctoral work in physics before switching to bio-informatics and next-generation sequencing; and McVean sees many recruits enter the field from banking and finance.

Statistician Yun Li joined the faculty of the University of North Carolina in Chapel Hill after earning her doctoral degree in biostatistics at the University of Michigan in 2009. In her undergraduate degree, Li had minored in computer science; she then earned a master's in statistics before starting her doctorate. While working on her PhD, Li developed data-analysis methods for the 1000 Genomes Project, a multi-national study in which more than 1,000 individuals' genomes are being sequenced. She says that the hands-on experience working with what she calls "dirty" data — raw data whose characteristics and limitations have not been fully explored by researchers — has been invaluable in her current position.

"A typical genetic study nowadays will need to analyse millions or tens of millions of variants in tens of thousands of individuals," says Li, who is now developing ways to work with large data sets and applying these and other methods to disease-focused studies. "This entails skills both to identify problems — which is important because many issues are typically not defined for data from cutting-edge research — and to solve problems."

H. ZHOU

"A typical genetic study nowadays will need to analyse millions of variants."
Yun Li

Whether trainees are interested in an academic or industrial job, it is computer-science skills that will help them to secure it. By far the most successful candidates are those who can not only write software, but also work with distributed computing systems, and computer operating systems such as Linux and Unix, say those in the field. "The more you understand software and computer science, the better off you are; writing software is 90% of what we're doing," says Alexander.

For a field that is likely to continue its rapid change, the only sure thing is that data sets will continue to get bigger, and those who know how to handle them will be in high demand. ∎

Erika Check Hayden *reports for* Nature *from San Francisco.*

COLUMN
Friend or foe?

It is difficult to balance the benefits of collaboration and competition, argues **Lydia Murray**.

As a PhD student, learning to navigate the murky waters of collaboration and competition is pretty confusing. I recently attended my first conference — and never mind the name badges, I wanted to tattoo 'FRIEND' or 'FOE' on people's foreheads. Given that a researcher's publications are often months, if not years, behind their current lab work, it is hard to discover who is working on what. Knowing when to share unpublished ideas and when to practise your poker face can be a nightmare for an early-career scientist.
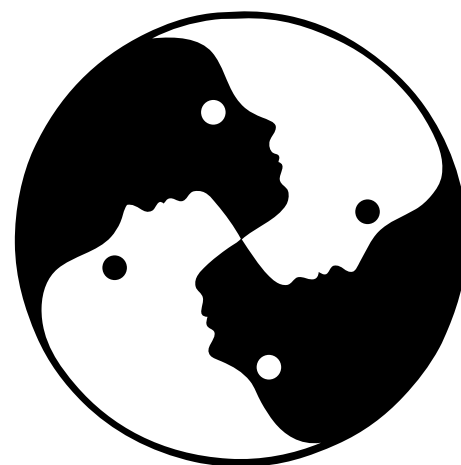
Why is it so hard? One reason is that science is a truly integrated discipline: completely independent fields are rare. As multiple groups generate data around the world, hypotheses evolve, and the direction of a scientist's research can change. One group's work might bleed into another's field of interest. So when two labs find their investigations becoming a bit too close for comfort, how do they decide whether to collaborate or compete?

Collaborations can be brilliant. Bringing together different skills and expertise offers fresh insight into old challenges and opens up new avenues of research. However, sharing a research theme does not always result in happy scientist families. Competition can overshadow the collaborative spirit and hinder progress.

Of course, competition is essential to science. It can stimulate motivation and productivity for labs addressing the same questions with conflicting hypotheses: the opportunity to deliver a scientific 'I told you so' is an appealing incentive. Healthy rivalry keeps fields exciting and ensures that all angles of research questions are considered.

However, when different groups are testing the same hypothesis, the contest is often simply a race to publication. The group that wins increases its citation number and strengthens its reputation. But does this justify the duplicated data, man hours and, potentially, taxpayers' money? In the current economic climate, I find it hard to understand how this style of competition remains prevalent.

There is at least one intermediate path between collaboration and competition: labs can coordinate publications. Instead of rushing through projects in parallel, they can agree to submit simultaneously and address a complementary range of questions. Without the time

J. TOE/SHUTTERSTOCK

pressure, compromises in research quality are reduced. Ultimately, the journal audience can read a far more comprehensive story.

But many labs continue to jealously guard their progress and sacrifice paper quality for personal recognition. Should such egotism be acceptable in science, the main aims of which are, ideally, discovery and innovation, rather than accolades for its practitioners? As a young researcher, I am puzzled that a community reliant on integrity and transparency is tolerant of lies and misdirection in the publications race.

That said, I'm not sure it would be prudent to advise young scientists always to speak freely at conferences and discard the poker face. Unless every person in the room does the same thing, you will eventually get scooped. As physicist Max Planck once wrote, "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it". Young scientists will have a crucial role in establishing a culture of greater cooperation amid a global scientific enterprise increasingly populated with far-flung collaborations. But we also need to recognize the importance of a bit of competition — and the reality that researchers will probably always be on the lookout for both friend and foe. ∎

Lydia Murray *is a PhD student in the department of medicine, veterinary and life sciences at the University of Glasgow, UK.*

Biomedical Research Centre, did his undergraduate, graduate and postdoctoral work in physics before switching to bio-informatics and next-generation sequencing; and McVean sees many recruits enter the field from banking and finance.

Statistician Yun Li joined the faculty of the University of North Carolina in Chapel Hill after earning her doctoral degree in biostatistics at the University of Michigan in 2009. In her undergraduate degree, Li had minored in computer science; she then earned a master's in statistics before starting her doctorate. While working on her PhD, Li developed data-analysis methods for the 1000 Genomes Project, a multinational study in which more than 1,000 individuals' genomes are being sequenced. She says that the hands-on experience working with what she calls "dirty" data — raw data whose characteristics and limitations have not been fully explored by researchers — has been invaluable in her current position.

"A typical genetic study nowadays will need to analyse millions or tens of millions of variants in tens of thousands of individuals," says Li, who is now developing ways to work with large data sets and applying these and other methods to disease-focused studies. "This entails skills both to identify problems — which is important because many issues are typically not defined for data from cutting-edge research — and to solve problems."

H. ZHOU

"A typical genetic study nowadays will need to analyse millions of variants."
Yun Li

Whether trainees are interested in an academic or industrial job, it is computer-science skills that will help them to secure it. By far the most successful candidates are those who can not only write software, but also work with distributed computing systems, and computer operating systems such as Linux and Unix, say those in the field. "The more you understand software and computer science, the better off you are; writing software is 90% of what we're doing," says Alexander.

For a field that is likely to continue its rapid change, the only sure thing is that data sets will continue to get bigger, and those who know how to handle them will be in high demand. ■

Erika Check Hayden *reports for* Nature *from San Francisco.*

COLUMN
Friend or foe?

It is difficult to balance the benefits of collaboration and competition, argues **Lydia Murray**.

As a PhD student, learning to navigate the murky waters of collaboration and competition is pretty confusing. I recently attended my first conference — and never mind the name badges, I wanted to tattoo 'FRIEND' or 'FOE' on people's foreheads. Given that a researcher's publications are often months, if not years, behind their current lab work, it is hard to discover who is working on what. Knowing when to share unpublished ideas and when to practise your poker face can be a nightmare for an early-career scientist.
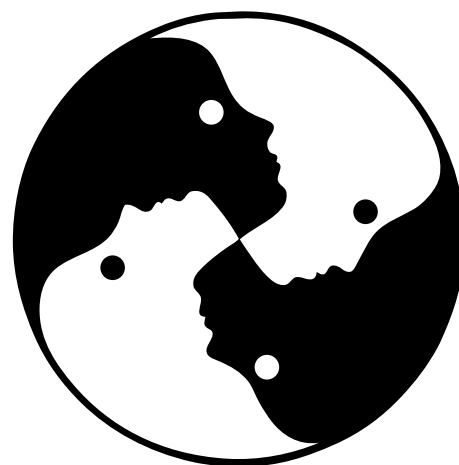
Why is it so hard? One reason is that science is a truly integrated discipline: completely independent fields are rare. As multiple groups generate data around the world, hypotheses evolve, and the direction of a scientist's research can change. One group's work might bleed into another's field of interest. So when two labs find their investigations becoming a bit too close for comfort, how do they decide whether to collaborate or compete?

Collaborations can be brilliant. Bringing together different skills and expertise offers fresh insight into old challenges and opens up new avenues of research. However, sharing a research theme does not always result in happy scientist families. Competition can overshadow the collaborative spirit and hinder progress.

Of course, competition is essential to science. It can stimulate motivation and productivity for labs addressing the same questions with conflicting hypotheses: the opportunity to deliver a scientific 'I told you so' is an appealing incentive. Healthy rivalry keeps fields exciting and ensures that all angles of research questions are considered.

However, when different groups are testing the same hypothesis, the contest is often simply a race to publication. The group that wins increases its citation number and strengthens its reputation. But does this justify the duplicated data, man hours and, potentially, taxpayers' money? In the current economic climate, I find it hard to understand how this style of competition remains prevalent.

There is at least one intermediate path between collaboration and competition: labs can coordinate publications. Instead of rushing through projects in parallel, they can agree to submit simultaneously and address a complementary range of questions. Without the time

J. TOE/SHUTTERSTOCK

pressure, compromises in research quality are reduced. Ultimately, the journal audience can read a far more comprehensive story.

But many labs continue to jealously guard their progress and sacrifice paper quality for personal recognition. Should such egotism be acceptable in science, the main aims of which are, ideally, discovery and innovation, rather than accolades for its practitioners? As a young researcher, I am puzzled that a community reliant on integrity and transparency is tolerant of lies and misdirection in the publications race.

That said, I'm not sure it would be prudent to advise young scientists always to speak freely at conferences and discard the poker face. Unless every person in the room does the same thing, you will eventually get scooped. As physicist Max Planck once wrote, "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it". Young scientists will have a crucial role in establishing a culture of greater cooperation amid a global scientific enterprise increasingly populated with far-flung collaborations. But we also need to recognize the importance of a bit of competition — and the reality that researchers will probably always be on the lookout for both friend and foe. ■

Lydia Murray *is a PhD student in the department of medicine, veterinary and life sciences at the University of Glasgow, UK.*

THE INTERRUPTION

Grinding halt.

BY PATRICIA FRONEK

"Let me out … pleeease … let me out." The fading voice from the ladies' room behind Colin was becoming less frequent, like a deflating balloon.

"Hey you … psst … get over here. C'mon fella, I can hear you." Fingers summoned Colin from the gap under the door marked 'Security Personnel Only'. He slid to the floor and sat, back against the wall.

"What's going on out there now?" The voice was muffled through the door.

Colin looked around at the decimated airport — perhaps desiccated would be more appropriate. The life had certainly been sucked out of it. Robotic staff made occasional whirring noises as if they were about to spring into action, only to remain frozen mid-task. People tried to sleep despite the din and suffocating heat. In the play area, fathers exchanged I-know-how-you-feel looks as they paced their now irritable children around. At knee height, the children eyeballed one another as they walked past. Their mothers, too lethargic to chat, sat together. Colin's eyes scanned the perimeter.

"Well, there's a big guy in the corner at the food dispensary. He keeps trying to open the hatch. Nothing's coming out until The Network's back, that's for sure."

The Zdevice3 was dead. No money, no food, no ordering, no directions, no communication, no help — in fact no nothing. With all the doors immobilized, everyone was stuck wherever they were when The Network went down. Restless, Colin fiddled with the device in his pocket — nothing.

The big guy in the corner was crying now. Stuffing tissues into his mouth, he chewed slowly as tears of frustration poured down his face. Behind him, people pressed against the viewing pane and talked in hushed tones.

"I think the plane's still circling," Colin said. "The Network had better kick in soon or it'll be bye-bye birdie." The small craft was in an endless loop circling the airport.

The man behind the door was silent for a moment. "Remember when we had *human* pilots?"

Colin put his head in his hands and let the disembodied voice ramble. He looked over at the tube that ran from the hotels to the airport. A capsule was blocking the access door — not far enough out to prize the doors open but close enough for the passenger to see and to be seen from the airport lounge. The woman in the capsule had long since stopped signalling for help and was now asleep, face pressed up against the window. Lipstick smeared her cheeks.

"So, you arriving or leaving?"

"Arriving," said Colin preoccupied. "And you?"

"Waiting for the boss. He's on that plane. I work for the company that runs The Net-

work." The voice under the door chuckled. "Heads will roll over this one!" After a long pause: "My wife is with him."

Oh brother! Colin really didn't want to know about this one.

"Hey mate," he said, "I didn't mean what I said before about the plane … Anyway, I thought this could never happen?"

"The failure? It shouldn't! It has to be local."

"What if it's not? We've been here for hours already."

"Then we're in big trouble — they'll sue the pants off us — lost productivity and all that."

"What about security? How *secure* are we *really*?"

"Everything'll shut down. Nothing to worry about. The systems are impenetrable — defence, the international banking system, stock market … the lot. Impossible to penetrate."

Colin wasn't so sure …

A commotion at the viewing pane drew his attention. Whispers turned into gasps. A crowd formed to watch the increasingly drunken movement of the small aircraft.

"What's going on?"

Colin didn't answer. The woman trapped in the toilets began to yell again. Colin stood up. The device in his pocket vibrated. Like a single organism, commuters reached for their devices and read the now lit screens.

The Network has now been restored. We apologize for the interruption.

Doors opened. Robotic staff reanimated. Orbital vacuum cleaners were spat from the walls to clean up the detritus that had accumulated while they were out of action. People picked themselves up and returned to their business as though nothing had happened. The food hatch opened. Transfer capsules, one after the other, were ejected from the tube. Dishevelled passengers spewed out, falling over one another. The aircraft regained altitude and prepared for landing — for real this time.

The owner of the voice behind the door, a little worse for wear, stepped out rubbing his bald head and looked around. After a moment, he shrugged and walked towards the arrivals area. He had no idea who Colin was. Colin was indistinguishable from the rest, simply one of the crowd, and Colin had no desire to make it otherwise. There was one more thing to do. Colin entered a code into his device, picked up his bag and walked slowly out of the airport.

The device vibrated a second time.

Your funds have been received. Transaction complete. Have a nice day!

Colin slowly smiled — the smile of a man who had just come into a lot of money. A very, very rich man. His stiff demeanour gave no hint of his inner excitement. He loved The Network, no longer impenetrable. ∎

Patricia Fronek *is a senior lecturer in the School of Human Services and Social Work at Griffith University, Gold Coast Campus, Australia, and a member of the Population and Social Health Research Program, Griffith Health Institute, Griffith University.*

JACEY

LETTER

Cancer exome analysis reveals a T–cell–dependent mechanism of cancer immunoediting

Hirokazu Matsushita[1]†*, Matthew D. Vesely[1]*, Daniel C. Koboldt[2], Charles G. Rickert[1], Ravindra Uppaluri[3], Vincent J. Magrini[2,4], Cora D. Arthur[1], J. Michael White[1], Yee–Shiuan Chen[1], Lauren K. Shea[1], Jasreet Hundal[2], Michael C. Wendl[2,4], Ryan Demeter[2], Todd Wylie[2], James P. Allison[5,6], Mark J. Smyth[7,8], Lloyd J. Old[9], Elaine R. Mardis[2,4] & Robert D. Schreiber[1]

Cancer immunoediting, the process by which the immune system controls tumour outgrowth and shapes tumour immunogenicity, is comprised of three phases: elimination, equilibrium and escape[1–5]. Although many immune components that participate in this process are known, its underlying mechanisms remain poorly defined. A central tenet of cancer immunoediting is that T-cell recognition of tumour antigens drives the immunological destruction or sculpting of a developing cancer. However, our current understanding of tumour antigens comes largely from analyses of cancers that develop in immunocompetent hosts and thus may have already been edited. Little is known about the antigens expressed in nascent tumour cells, whether they are sufficient to induce protective antitumour immune responses or whether their expression is modulated by the immune system. Here, using massively parallel sequencing, we characterize expressed mutations in highly immunogenic methylcholanthrene-induced sarcomas derived from immunodeficient $Rag2^{-/-}$ mice that phenotypically resemble nascent primary tumour cells[1,3,5]. Using class I prediction algorithms, we identify mutant spectrin-β2 as a potential rejection antigen of the d42m1 sarcoma and validate this prediction by conventional antigen expression cloning and detection. We also demonstrate that cancer immunoediting of d42m1 occurs via a T-cell-dependent immunoselection process that promotes outgrowth of pre-existing tumour cell clones lacking highly antigenic mutant spectrin-β2 and other potential strong antigens. These results demonstrate that the strong immunogenicity of an unedited tumour can be ascribed to expression of highly antigenic mutant proteins and show that outgrowth of tumour cells that lack these strong antigens via a T-cell-dependent immunoselection process represents one mechanism of cancer immunoediting.

For this study, we chose two representative, highly immunogenic, unedited methylcholanthrene (MCA)-induced sarcoma cell lines, d42m1 and H31m1, derived from immunodeficient $Rag2^{-/-}$ mice[1]. Both grow progressively when transplanted orthotopically into $Rag2^{-/-}$ mice, but are rejected when transplanted into naive wild-type mice (Supplementary Figs 1 and 2). Using a modified form of exome sequencing involving complementary DNA (cDNA) capture by mouse exome probes and Illumina deep sequencing (that is, cDNA capture sequencing or cDNA CapSeq), we identified 3,737 somatic, non-synonymous mutations in d42m1 cells (3,398 missense, 221 nonsense, 2 nonstop and 116 splice site mutations) and 2,677 non-synonymous mutations in H31m1 cells (2,391 missense, 160 nonsense, 3 nonstop and 123 splice site mutations) (Fig. 1a and Supplementary Fig. 3 and Supplementary Table 1). The mutations in each cell line

were largely distinct—d42m1 and H31m1 share only 119 identical missense mutations (Fig. 1b and Supplementary Table 2)—a result that potentially explains the unique antigenicity of each cell line (Supplementary Fig. 4). Although d42m1 and H31m1 display mutations in known cancer genes[6], the functional effects of these novel mutations remain undefined. Nevertheless, both tumours have cancer-causing mutations in Kras (codon 12) and Trp53 that are frequently observed in human and mouse cancers[7–9] (Supplementary Table 3). The mutation calls were confirmed by independent Roche/454 pyrosequencing of 22 genes using tumour genomic DNA and by documenting their absence in normal cells from the same mouse that developed the tumour (Supplementary Table 4).

Comparing cDNA CapSeq data of d42m1 and H31m1 cells to human cancer genomes[10–17] revealed two similarities. First, 46–47% of mutations in d42m1 and H31m1 are C/A or G/T transversions, which represent chemical-carcinogen signatures[7,13,14] similar to those of lung cancers from smokers (44–46%) but not seen in human cancers induced by other mechanisms (8–16%) (Fig. 1c). Second, the mutation rates of d42m1 and H31m1 are about tenfold higher than those of lung cancers from smokers, but within threefold of hypermutator smoker lung cancers with mutations in DNA repair pathway genes (Fig. 1d). Interestingly, d42m1 and H31m1 also show mutations in DNA repair genes (Supplementary Table 3), although these novel mutations have not been functionally characterized. Thus, mouse MCA-induced sarcomas have qualitative and quantitative genomic similarities to carcinogen-induced human cancers.

When parental d42m1 sarcoma cells were transplanted into naive wild-type mice, approximately 20% of recipients developed escape tumours (Supplementary Fig. 5a, c). Cell lines made from three escape tumours (d42m1-es1, d42m1-es2 and d42m1-es3) formed progressively growing sarcomas when transplanted into naive wild-type recipients (Fig. 2a). In contrast, parental d42m1 tumour cells passaged through $Rag2^{-/-}$ mice maintained high immunogenicity (Supplementary Fig. 5b, d). Additional analyses revealed that whereas eight of ten clones of d42m1 were rejected in wild-type mice, two clones (d42m1-T3 and d42m1-T10) grew with kinetics similar to d42m1 escape tumours (Fig. 2a and Supplementary Fig. 6). Thus, the d42m1 cell line consists mostly, but not entirely, of highly immunogenic clones and undergoes immunoediting in wild-type mice. cDNA CapSeq of parental d42m1 cells, clones and escape tumours revealed that all expressed similar numbers of mutations (Supplementary Fig. 7a and Supplementary Table 1) and phylogenetic analysis revealed that all d42m1-derived cells were genomically related to one another but distinct from H31m1 and normal fibroblasts (Supplementary

[1]Department of Pathology and Immunology, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. [2]The Genome Institute, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. [3]Department of Otolaryngology, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. [4]Department of Genetics, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. [5]Ludwig Center for Cancer Immunotherapy, Department of Immunology, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. [6]Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. [7]Cancer Immunology Program, Peter MacCallum Cancer Centre, East Melbourne, 3002 Victoria, Australia. [8]Department of Pathology, University of Melbourne, Parkville, 2010 Victoria, Australia. [9]New York Branch of The Ludwig Institute for Cancer Research at Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. †Present address: Department of Immunotherapeutics (Medinet), The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan.
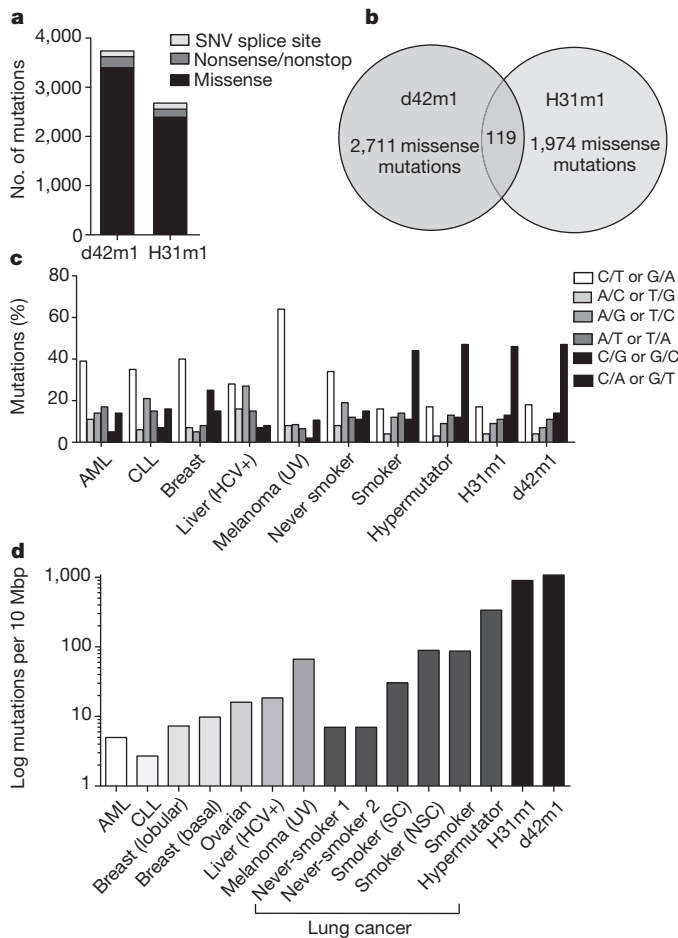*These authors contributed equally to this work.

Figure 1 | Unedited MCA-induced sarcomas d42m1 and H31m1 genomically resemble carcinogen-induced human cancers. a, Number of non-synonymous mutations in d42m1 and H31m1 tumour cells as detected by cDNA CapSeq. SNV, single nucleotide variant. **b,** Missense mutations compared between d42m1 and H31m1 that had at least 20× sequencing coverage. **c,** Spectrum of DNA nucleotide substitutions detected in d42m1 and H31m1 as compared to previously generated data from human cancers including acute myelogenous leukaemia[10] (AML), chronic lymphocytic leukaemia[16] (CLL), breast cancer (breast lobular[12], breast basal[11]), ovarian cancer (E. R. Mardis *et al.*, manuscript in preparation), liver cancer (hepatitis C virus (HCV)-positive)[15], melanoma (ultraviolet (UV)-induced)[17] and lung cancers (non-small cell (NSC)[13], small cell (SC)[14], never-smoker, smoker and hypermutator (E. R. Mardis *et al.*, manuscript in preparation). **d,** Mutation rates for d42m1, H31m1 and human cancers described in **c** including tumours from never-smoker 1 (bronchioloalveolar carcinoma) and never-smoker 2 (lung adenocarcinoma).

Fig. 7b). However, regressor clones clustered more closely to parental d42m1 cells whereas progressor clones clustered more closely to cells from escape tumours. Thus, the d42m1 tumour cell line consists of a related, but heterogeneous population of tumour cells.

Tumour-specific mutant proteins presented on mouse or human MHC class I molecules are known to represent one class of tumour-specific antigens for CD8[+] T cells[18,19]. Therefore, we used *in silico* analysis[20] to assess the theoretical capacities of missense mutations from d42m1-related tumour cells to bind MHC class I proteins. Each d42m1-related cell type expressed many potential high-affinity (half-maximum inhibitory concentration (IC_{50}) < 50 nM; affinity value ($1/IC_{50} \times 100$) > 2) epitopes that could bind to H-2D[b] or H-2K[b] (Fig. 2b). Of these, 39–42 were expressed only in the regressor subset of d42m1-related cells (7–9 for H-2D[b], 30–35 for H-2K[b]), including 31 expressed in all regressor cells (Supplementary Table 5). Thus, ~1% of the missense mutations in d42m1 are selectively expressed in rejectable d42m1 clones.

Whereas parental and regressor d42m1 cells stimulated interferon-γ (IFN-γ) release *in vitro* when incubated with a specific CD8[+] cytotoxic T lymphocyte (CTL) clone (C3) derived from a wild-type mouse that had rejected parental d42m1 tumour cells (Fig. 3a, b), progressor d42m1 clones, cells from escape tumours or unrelated MCA sarcomas did not. This result demonstrated that all regressor d42m1 tumour cells share a mutation that forms the epitope recognized by C3 CTLs. As recognition of d42m1 regressor cells by C3 CTLs is restricted by H-2D[b] (Fig. 3c), we postulated that an R913L mutation in spectrin-β2 produced the most likely target for C3 CTLs because its expression was restricted to d42m1 regressor clones and it formed an epitope that showed high-affinity binding potential to H-2D[b] in contrast to the wild-type sequence predicted to bind with low affinity (Fig. 3d and Supplementary Table 5).

To verify the importance of mutant spectrin-β2 on d42m1 antigenicity, we independently identified the tumour antigen recognized by the C3 CTL clone using a T-cell-based expression cloning approach[21]. After three screening rounds, a single positive cDNA was identified encoding a sequence identical to the R913L spectrin-β2 mutant (Fig. 3e). Thus, conventional antigen expression cloning identified the same mutation predicted by the genomic sequencing.

Mutation-specific real-time quantitative polymerase chain reaction with reverse transcription (qRT–PCR) revealed the presence of mutant spectrin-β2 messenger RNA in parental d42m1 tumour cells and regressor d42m1 clones, but not in progressor d42m1 clones or escape tumours (Fig. 3f), nor in normal tissue of the mouse from which the d42m1 tumour was derived (Supplementary Table 4 and Supplementary Fig. 8). Additionally, C3 CTLs discriminated between mutant and wild-type spectrin-β2 peptide sequences when presented on an unrelated H-2D[b]-expressing cell line (Fig. 3g). Whereas the mutant (VAVVNQIAL; underline letter indicates the site of mutation) peptide stimulated C3 CTLs in a dose-dependent manner, the wild-type (VAVVNQIAR) peptide did not, even when added in 1,000-fold excess. Using labelled H-2D[b] tetramers generated with mutant peptide, mutant spectrin-β2-specific CD8[+] T cells accumulated over time in parental d42m1 tumours developing *in vivo* and draining lymph nodes before tumour rejection (Fig. 4a, b). In contrast, no mutant spectrin-β2-specific CD8[+] T cells were detected in progressively growing escape tumours or draining lymph nodes. These data demonstrate that mutant spectrin-β2 expressed selectively in a high proportion of unedited d42m1 tumour cells evokes a T-cell response in naive wild-type mice that promotes the elimination of antigen-expressing tumour cells.

To test whether expression of mutant spectrin-β2 was sufficient to drive rejection of d42m1 tumour cells, we enforced expression of either mutant or wild-type spectrin-β2 in d42m1-es3 cells that lack this mutation (Supplementary Fig. 9a) and followed their growth in wild-type mice. Whereas d42m1-es3 tumour cell clones transduced with either control retrovirus or retrovirus encoding wild-type spectrin-β2 (WT.1 and WT.3) grew progressively with growth kinetics similar to unmanipulated d42m1-es3 cells, d42m1-es3 clones expressing mutant spectrin-β2 (mu.6 and mu.14) were rejected in wild-type mice, but not in *Rag2*[−/−] mice (Fig. 4c and Supplementary Fig. 9b, c, d). CD8[+] T cells specific for mutant spectrin-β2 did not infiltrate d42m1-es3 tumours expressing wild-type spectrin-β2 (WT.3), but were present in d42m1-es3 tumours expressing mutant spectrin-β2 (mu.14) that were rejected in wild-type mice (Fig. 4d). Thus, mutant spectrin-β2 is indeed a major rejection antigen of d42m1 sarcoma cells and d42m1 escape from immune control is the consequence of outgrowth of d42m1 clones that lack expression of dominant rejection antigens.

The possibility that the lack of dominant rejection antigen(s) in a small subset of d42m1 cells was due to epigenetic silencing was ruled out because no spectrin-β2 mutation was (1) found by sequencing genomic DNA from progressor d42m1 clones or escape tumours (Supplementary Table 4) or (2) expressed in d42m1 progressor clones or escape tumours after treatment with inhibitors of methyltransferases and histone deacetylases (Supplementary Fig. 10). We therefore
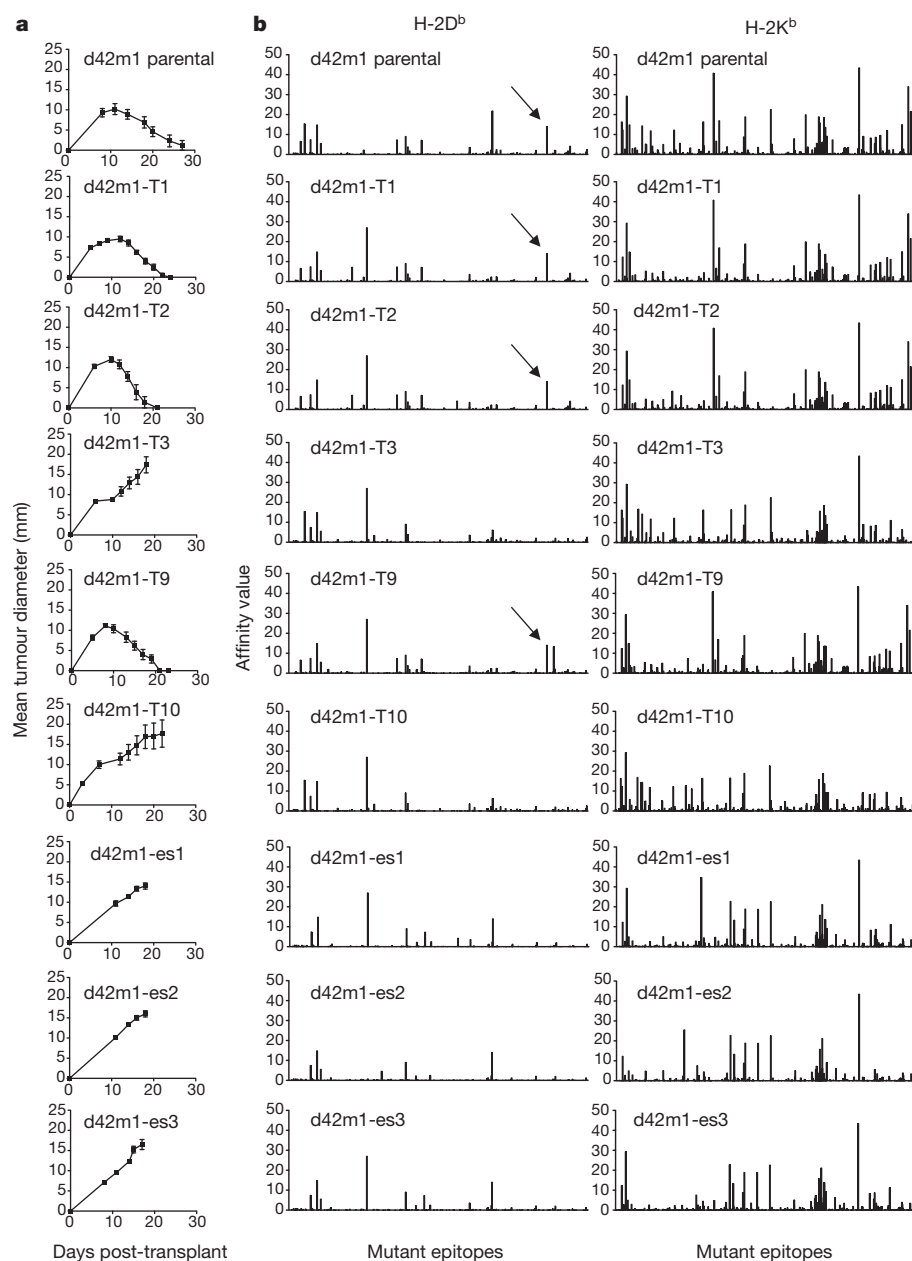
Figure 2 | Affinity value profiles of predicted MHC class I epitopes from tumour-specific mutations. a, Growth of d42m1 parental cells, a representative sample of tumour clones, and three escape tumours following transplantation into wild-type mice ($n = 5$, squares). Data are presented as average tumour diameter ± s.e.m. and are representative of three independent experiments. **b**, Missense mutations for each d42m1-related tumour examined in **a** were analysed for potential MHC class I neoepitopes that bind either H-2D^b or H-2K^b. Predicted epitope binding affinities were ultimately expressed as affinity values (1/$IC_{50} \times 100$). Arrows indicate H-2D^b epitopes created by the R913L spectrin-β2 mutant.

asked whether T-cell-dependent immunoselection explained the outgrowth of escape tumours. Specifically, we examined the *in vivo* growth behaviour of a tumour cell mixture containing a vast majority of highly immunogenic, mutant spectrin-β2$^+$ d42m1-T2 cells and a minority of mutant spectrin-β2$^-$ d42m1-T3 progressor cells. To distinguish between the two cell types, we labelled d42m1-T2 with red fluorescent protein (RFP) (modified to eliminate class I epitopes) and d42m1-T3 with green fluorescent protein (GFP) and documented that the labelling did not alter their *in vivo* growth characteristics. We found that we could recapitulate the tumour growth phenotype of parental d42m1 at a ratio of 95% d42m1-T2 cells to 5% d42m1-T3 cells (Fig. 4e). At this ratio, 100% of *Rag2*$^{-/-}$ mice and wild-type mice depleted of either CD4$^+$ or CD8$^+$ T cells developed progressively growing tumours (Fig. 4f). In contrast, 5/20 (25%) wild-type mice injected with the tumour cell mixture developed escape tumours, a result that recapitulated the behaviour of parental d42m1. Tumours harvested from *Rag2*$^{-/-}$ mice were comprised of 84% d42m1-T2 cells and 14% d42m1-T3 cells (Fig. 4h) and expressed mutant spectrin-β2 (Fig. 4g), that is, they resembled the initial 95:5 cell mixture. In contrast, tumours that grew out in wild-type mice consisted of 98%

d42m1-T3 tumour cells and lacked mutant spectrin-β2 (Fig. 4g, h). Thus, d42m1 escape tumours develop as a consequence of T-cell-dependent immunoselection favouring the outgrowth of tumour cells that lack major rejection antigens.

This report shows that the combination of cancer exome sequencing and *in silico* epitope prediction algorithms can identify highly immunogenic, tumour-specific mutational antigens in unedited carcinogen-induced cancers that serve as targets for the elimination phase of cancer immunoediting. To our knowledge, this is the first study to use a genomics approach to experimentally identify a tumour antigen, to specifically identify an antigen from an unedited tumour and to demonstrate that T-cell-dependent immunoselection is a mechanism underlying the outgrowth of tumour cells that lack strong rejection antigens. This mechanism most likely also produces other types of escape tumours, such as those that develop inactivating mutations in antigen presentation genes (for example, those encoding MHC class I proteins), which are frequently observed in clinically apparent human cancers[22,23]. Developing carcinogen-induced tumours (for example, mouse MCA sarcomas or human smoker lung cancers) may be the preferred targets of cancer immunoediting because they express the
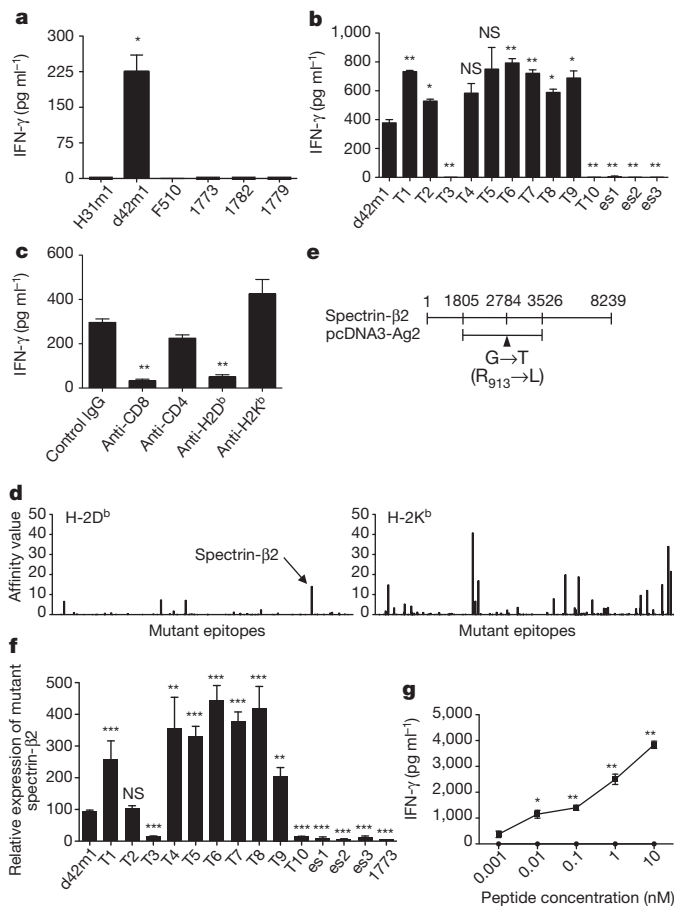
Figure 3 | Identification of mutant spectrin-β2 as an authentic antigen of an unedited tumour. a, b, IFN-γ release by C3 CTLs following co-culture with different unedited sarcomas (a) or d42m1-related tumours (b). c, IFN-γ release by C3 CTLs is inhibited by monoclonal antibodies that block CD8 and H-2D[b], but not CD4 or H-2K[b]. d, MHC class I epitopes predicted to be shared in all of the regressor d42m1 tumours, but not in progressor d42m1 tumours. e, Representation of the cDNA clone that stimulated C3 CTLs encoding the spectrin-β2 R913L mutation. f, qRT–PCR for mutant spectrin-β2 in d42m1-related tumours and 1773. g, IFN-γ release by C3 CTLs incubated with COS-D[b] cells pulsed with wild-type (circles) or mutant (squares) spectrin-β2 peptides. Data are representative of three independent experiments. Samples were compared in b, f to d42m1 using an unpaired, two-tailed Student's t test ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$; NS, not significant).

greatest number of mutations that might function as neoantigens. However, as ~1% of the mutations in d42m1 are selectively expressed in regressor tumour clones, it is possible that spontaneous tumours arising by other means that harbour as few as 100–200 mutations could still be susceptible to immunological sculpting as they develop. In this regard it is significant that, as documented in a complementary study reported in this issue[24], oncogene-induced primary sarcomas engineered to express a strong model antigen can also undergo T-cell-dependent immunoediting, resulting in the outgrowth of tumours that escape immune control. It will be interesting in the future to compare the effects of immunity on the antigenic profiles of oncogene- versus carcinogen-induced tumours.

The immunodominance of mutant spectrin-β2 in driving tumour rejection in many ways resembles that of certain viral antigens[25] and is probably due to the presence in d42m1 of four copies of chromosome 11, each of which carries the spectrin-β2 gene, thereby producing a highly abundant neoepitope that binds to H-2D[b] 750-fold stronger than that of the wild-type sequence. More work is needed to determine which of the other mutations, if any, selectively expressed in d42m1 regressors function as rejection antigens. Immunoepitope analysis of parental
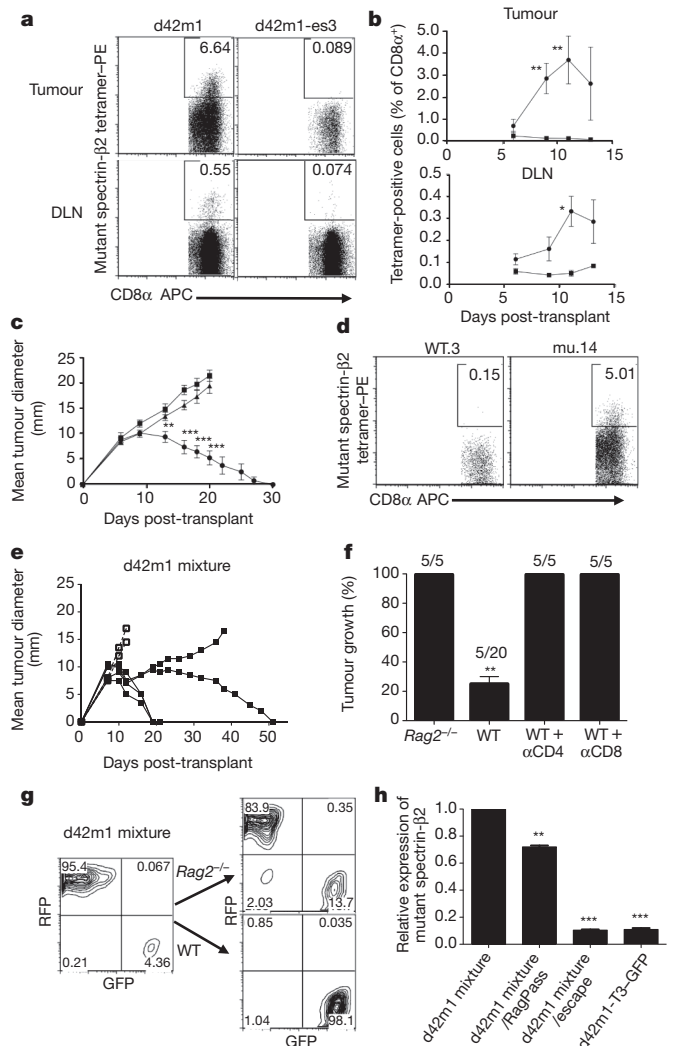

Figure 4 | Mutant spectrin-β2 is a major rejection antigen of d42m1. a, Mutant spectrin-β2-specific CD8[+] T cells were detected by tetramer staining in tumours and draining lymph nodes (DLNs) from mice challenged with d42m1 parental cells, but not d42m1-es3 cells on day 11 post-transplant. APC, allophycocyanin; PE, phycoerythrin. b, Quantification and kinetics of mutant spectrin-β2 tetramer staining in mice challenged with d42m1 parental cells ($n = 3$, circles) or d42m1-es3 cells ($n = 3$, squares). c, Growth of d42m1-es3 tumour cell clones transduced with wild-type ($n = 5$, squares) or mutant spectrin-β2 ($n = 5$, circles) and control d42m1-es3 cells ($n = 5$, triangles) after transplantation (1×10^6 cells) into wild-type mice. Data are presented as average tumour diameter ± s.e.m. d, d42m1-es3 tumours reconstituted with wild-type (WT.3) or mutant spectrin-β2 (mu.14) were harvested at day 11 and CD8[+] T cells were stained with mutant spectrin-β2 tetramers. e, Growth of a mixture of d42m1-T2–RFP (95%) and of d42m1-T3–GFP (5%) after transplantation (1×10^6 total cells) into wild-type ($n = 5$, solid lines, closed squares) or $Rag2^{-/-}$ ($n = 2$, dashed lines, open squares) mice. f, Tumour outgrowth in $Rag2^{-/-}$ or wild-type (WT) mice treated or untreated with monoclonal antibodies that deplete CD4[+] or CD8[+] T cells after challenge with 1×10^6 cells of a d42m1 mixture (95% d42m1-T2–RFP and 5% d42m1-T3–GFP). Data are presented as per cent tumour positive mice from 2–4 independent experiments ($n = 2$–5 mice per group). g, h, GFP and RFP expression (g) and mutant spectrin-β2 expression (h) were analysed in the d42m1-T2–RFP/d42m1-T3–GFP tumour cell mixture before injection and from tumours that grew out in $Rag2^{-/-}$ mice (RagPass) or escaped in wild-type mice by flow cytometry (g) or qRT–PCR (h). Data are representative of two independent experiments. Samples were compared using an unpaired, two-tailed Student's t-test ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$; NS, not significant).

H31m1 reveals that it expresses multiple potential strong neoantigens (19 potential strong binders to H-2D[b] and 58 to H-2K[b]) (Supplementary Fig. 11a) and induces both H-2D[b]- and H-2K[b]-restricted

CD8[+] T-cell responses during rejection (Supplementary Fig. 11b). This result suggests that H31m1 shows an even more complex antigenicity than d42m1 and probably explains why H31m1 never produces escape tumours in wild-type mice (Supplementary Fig. 11c).

Chemically induced tumours have had a critical role in the history of tumour immunology, providing the first unequivocal demonstration of tumour-specific antigens[26,27] and, subsequently, the first evidence of cancer immunoediting[1-3]. It is therefore significant that this same model has now provided new insights into the antigenic targets of cancer immunoediting and some of the key molecular mechanisms that drive the process. Although more work is needed to determine whether and how frequently this process occurs during development of spontaneous and carcinogen-induced human cancers, it is tempting to speculate that a genomics approach to tumour antigen identification could, in the future, facilitate the development of individualized cancer immunotherapies directed at tumour-specific—rather than cancer-associated—antigens.

METHODS SUMMARY

d42m1 and H31m1 MCA-induced sarcomas were generated in male 129/Sv $Rag2^{-/-}$ mice as previously described[1]. Total RNA was isolated from low-passage MCA-induced sarcoma cell lines and skin fibroblasts from male 129/Sv $Rag2^{-/-}$ mice using the RNeasy Mini kit (Qiagen) and cDNA was prepared using oligo (dT) primers and SuperScript II Reverse Transcriptase (Invitrogen). Illumina libraries prepared with this cDNA were hybridized to biotinylated Agilent mouse exome probes. Library components were captured using strepavidin-coated magnetic beads (DynaBeads), PCR amplified and sequenced using an Illumina GAIIx analyser (cDNA CapSeq). Putative somatic mutations were identified using VarScan 2 (v.2.2.4). Missense mutations were analysed for potential neoepitope binding to MHC class I using an algorithm[20] available at Immune Epitope Database and Analysis Resource (http://www.immuneepitope.org) and were expressed as affinity values (reciprocal of the predicted IC_{50} multiplied by 100).

All tumour cell lines were injected subcutaneously in the flank of naïve syngeneic male mice (1×10^6 cells). Ten d42m1 tumour cell clones were isolated from the parental cell line by limiting dilution. Escape tumours of d42m1 were harvested from tumours growing in wild-type mice and cell lines were produced. To generate the C3 d42m1-specific CTL clone, splenocytes from a mouse that rejected d42m1 were harvested, stimulated with parental d42m1 target cells pre-treated with $100\,U\,ml^{-1}$ IFN-γ for 48 h and irradiated with 100 Gy and cloned by limiting dilution. To clone the antigen recognized by the C3 CTL clone, a d42m1 cDNA library was cloned into pcDNA3 (Invitrogen), transfected into COS cells expressing mouse H-2D[b], and screened for C3 reactivity by IFN-γ ELISA (eBioscience). Mutant spectrin-β2 expression was detected by qRT–PCR using mutation-specific primers. H-2D[b] tetramers were generated with 905–913 mutant spectrin-β2 peptides by the NIH Tetramer Facility (Emory).

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

1. Shankaran, V. et al. IFNγ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature 410, 1107–1111 (2001).
2. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. Nature Immunol. 3, 991–998 (2002).
3. Koebel, C. M. et al. Adaptive immunity maintains occult cancer in an equilibrium state. Nature 450, 903–907 (2007).
4. Vesely, M. D., Kershaw, M. H., Schreiber, R. D. & Smyth, M. J. Natural innate and adaptive immunity to cancer. Annu. Rev. Immunol. 29, 235–271 (2011).
5. Schreiber, R. D., Old, L. J. & Smyth, M. J. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science 331, 1565–1570 (2011).
6. Futreal, P. A. et al. A census of human cancer genes. Nature Rev. Cancer 4, 177–183 (2004).
7. Chen, A. C. & Herschman, H. R. Tumorigenic methylcholanthrene transformants of C3H/10T1/2 cells have a common nucleotide alteration in the c-Ki-ras gene. Proc. Natl Acad. Sci. USA 86, 1608–1611 (1989).
8. Tuveson, D. A. et al. Endogenous oncogenic K-ras(G12D) stimulates proliferation and widespread neoplastic and developmental defects. Cancer Cell 5, 375–387 (2004).
9. Kirsch, D. G. et al. A spatially and temporally restricted mouse model of soft tissue sarcoma. Nature Med. 13, 992–997 (2007).
10. Ley, T. J. et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 456, 66–72 (2008).
11. Ding, L. et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 464, 999–1005 (2010).
12. Shah, S. P. et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 461, 809–813 (2009).
13. Lee, W. et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature 465, 473–477 (2010).
14. Pleasance, E. D. et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 463, 184–190 (2010).
15. Totoki, Y. et al. High-resolution characterization of a hepatocellular carcinoma genome. Nature Genet. 43, 464–469 (2011).
16. Puente, X. S. et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature 475, 101–105 (2011).
17. Pleasance, E. D. et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 463, 191–196 (2010).
18. Boon, T., Coulie, P. G., Van den Eynde, B. J. & van der Bruggen, P. Human T cell responses against melanoma. Annu. Rev. Immunol. 24, 175–208 (2006).
19. Segal, N. H. et al. Epitope landscape in breast and colorectal cancer. Cancer Res. 68, 889–892 (2008).
20. Nielsen, M. et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 12, 1007–1017 (2003).
21. van der Bruggen, P. et al. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. Science 254, 1643–1647 (1991).
22. Khong, H. T. & Restifo, N. P. Natural selection of tumor variants in the generation of "tumor escape" phenotypes. Nature Immunol. 3, 999–1005 (2002).
23. Dunn, G. P., Old, L. J. & Schreiber, R. D. The three Es of cancer immunoediting. Annu. Rev. Immunol. 22, 329–360 (2004).
24. DuPage, M., Mazumdar, C., Schmidt, L. M., Cheung, A. F. & Jacks, T. Expression of tumour-specific antigens underlies cancer immunoediting. Nature doi:10.1038/nature10803 (this issue).
25. Yewdell, J. W. Confronting complexity: real-world immunodominance in antiviral CD8[+] T cell responses. Immunity 25, 533–543 (2006).
26. Prehn, R. T. & Main, J. M. Immunity to methylcholanthrene-induced sarcomas. J. Natl. Cancer Inst. 18, 769–778 (1957).
27. Old, L. J. & Boyse, E. A. Immunology of experimental tumors. Annu. Rev. Med. 15, 167–186 (1964).

METHODS

Mice. *Ifngr1*$^{-/-}$ mice[28] and *Ifnar1*$^{-/-}$ mice[29] on a 129/Sv background were originally provided by M. Aguet and were bred in our specific pathogen-free animal facility. Wild-type and *Rag2*$^{-/-}$ mice were purchased from Taconic Farms. All mice were male and on a 129/Sv background and were housed in our specific pathogen-free animal facility. For all experiments, male mice were 8–12 weeks of age and studies were performed in accordance with procedures approved by the AAALAC accredited Animal Studies Committee of Washington University in St. Louis.

Tumour transplantation. MCA-induced sarcomas used in this study were generated in male 129/Sv strain wild-type or *Rag2*$^{-/-}$ mice and banked as low-passage tumour cells as previously described[1]. Tumour cells derived from frozen stocks were propagated *in vitro* in RPMI media (Hyclone) supplemented with 10% FCS (Hyclone) and injected subcutaneously in 150 µl of endotoxin-free PBS into the flanks of recipient mice. Tumour cells were >90% viable at the time of injection as assessed by trypan blue exclusion and tumour size was quantified as the average of two perpendicular diameters. For antibody depletion studies, 250 µg of control IgG (PIP), anti-CD4 (GK1.5) or anti-CD8α (YTS169.4) were injected intraperitoneally into mice at day −1 and every 7 days thereafter.

Isolation of normal skin fibroblasts. Skin fibroblasts were isolated from three independent male 129/Sv *Rag2*$^{-/-}$ pups by harvesting skin and incubating in 0.25% trypsin (Hyclone) at 37 °C for 30 min before washing in DMEM media (Hyclone). After washing, chunks of skin were filtered to achieve single-cell suspensions and cultured *in vitro* with DMEM media. After three passages, skin fibroblasts were harvested to isolate genomic DNA and total RNA.

Extraction of genomic or complementary DNA. Genomic DNA from sarcoma cells and normal skin fibroblasts was extracted using DNeasy Blood & Tissue Kit (Qiagen). For cDNA isolation, total RNA from sarcoma cells and normal skin fibroblasts was isolated using RNeasy Mini kit (Qiagen) and cDNA was synthesized using oligo (dT) primers and SuperScript II Reverse Transcriptase (Invitrogen).

cDNA CapSeq. cDNA samples from each tumour (100 ng) were constructed into Illumina libraries according to the manufacturer's protocol (Illumina) with the following modifications. First, cDNA was fragmented using Covaris S2 DNA Sonicator (Covaris) in 1× end-repair buffer followed by the direct addition of the enzyme repair cocktail (Lucigen). Fragment sizes ranged between 100–500 bp. Second, Illumina adaptor-ligated DNA was amplified in four 50 µl PCRs for five cycles using 4 µl adaptor-ligated cDNA, 2× Phusion Master Mix and 250 nM forward and reverse primers, 5′-AATGATACGGCGACCACCGAGATCTAC ACTCTTTCCCTACACGACGCTCTTCCGATC and 5′- CAAGCAGAAGACG GCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC, respectively. Third, Solid Phase Reversible Immobilization (SPRI) bead cleanup was used to purify the PCR-amplified library and to select for 300–500 bp fragments. Five-hundred nanograms of the size-fractionated Illumina library were hybridized with the Agilent mouse exome reagent. After hybridization at 65 °C for 24 h, we added 50 µl of DynaBeads M-270 streptavidin-coated paramagnetic beads (10 mg ml^{-1}) to selectively remove the biotinylated Agilent probes and hybridized cDNA library fragments. The beads were washed according to manufacturer's protocol (Agilent) and the captured library fragments were released into solution using 50 µl of 0.125 N NaOH and neutralized with an equal volume of neutralization buffer (Agilent). The recovered fragments then were PCR amplified according to the manufacturer's protocol using 11 cycles in the PCR. Illumina library quantification was completed using the KAPA SYBR FAST qPCR Kit (KAPA Biosystems). The qPCR result was used to determine the quantity of library necessary to produce 180,000 clusters on a single lane of the Illumina GAIIx. One lane of 100 bp paired-end data was generated for each captured sample (as cDNA was used as the source for sequencing, we refer to this process as cDNA Capture Sequencing or cDNA CapSeq). Illumina reads were aligned to the NCBI build 37 (Mm9) mouse reference sequence using BWA[30] v.0.5.5 (with −q 5 soft trimming). Alignments from multiple lanes for the same sample were merged together using SAMtools r599, and duplicates were marked using Picard v.1.29.

Mutation detection and annotation. Putative somatic mutations were identified using VarScan 2 (v.2.2.4)[31] with the parameters '−min-coverage 3−min-var-freq 0.08−p-value 0.10−somatic-p-value 0.05−strand-filter 1' and specifying a minimum mapping quality of 10. Variants whose supporting reads exhibited read position bias (average read position <10 or >90), strand bias (>99% of reads on one strand), or mapping quality (score difference >30, or mismatch quality sum difference >100) relative to reference supporting reads were removed as probable false positives. We also required that the variant allele be present in at least 10% of tumour reads and no more than 5% of normal reads. The SNVs meeting these criteria were annotated using an internal database of GenBank/Ensembl transcripts (v58_73k). In the event that a variant was annotated using multiple transcripts, the annotation of most severe effect was used. Non-silent coding mutations (missense, nonsense/nonstop or splice site) were prioritized for downstream analysis.

Mutation rate and overlap comparisons. Mutation rates were estimated for each tumour sample using the number of putative 'tier 1' SNVs (missense, nonsense/nonstop, splice site, silent or noncoding RNA). To account for variability in coverage between samples, the SNV count for each tumour sample (S) was divided by a coverage factor (F), computed as the fraction of all tier 1 SNVs identified in any tumour sample ($n = 16,991$) that were covered by at least four reads in a given sample. For example, in the d42m1 parental sample, 15,852 of 16,991 tier 1 SNV positions were covered, for a coverage factor of 93.30%. The number of coverage-adjusted mutations in each sample was divided by the total size of tier 1 space in the mouse genome (43.884 Mbp) to determine the number of coding mutations per megabase (R).

$$R = (S/F) / (43.884 \text{ Mbp})$$

For the mutation overlap comparisons and relatedness-to-parental-tumour analysis, only high-confidence missense mutations were used (that is, 20× or above). A mutation was considered 'shared' between two samples if both samples had a predicted mutation at the same genomic position. For the comparison of mutated genes between d42m1 and H31m1 parental lines, a gene was considered 'shared' if both d42m1 and H31m1 samples had a predicted missense mutation in that gene, even if the mutations did not occur at the same position.

Roche/454 sequencing and validation. PCR primers were designed for 11 SNVs predicted to be somatic in d42m1 tumour samples, as well as 11 control sites that were H31m1-specific, low-confidence, or removed by the false-positive filter. All 22 SNVs were PCR amplified individually in 11 samples (SK1.1, d42m1, H31m1, T2, T3, T5, T9, T10, es1, es2 and es3) using MID-tailed primers to enable sample identification. PCR products were pooled together before sequencing on a quarter run of the Roche/454 Titanium platform. Read sequences and quality scores were extracted from 454 data files using *sffinfo* (454 proprietary software) then aligned to the mouse build 37 reference sequence (Mm19) using SSAHA2 v.2.5.3[32] with the SAM output option. Alignments were imported to BAM format and a 'pileup' assembly file generated using SAMtools v.0.1.18[33]. The average 454 sequence depth for targeted positions was 1,216× per sample. Validation read counts and allele frequencies in each sample at each variant position were determined using the *pileup2cns* command of VarScan v.2.2.7[31]. At least 20 reads with base quality of 20 or higher were required to confirm or refute a variant. 454 sequencing data and the primers used are presented in Supplementary Table 4.

3730 sequencing and validation. Eight SNVs predicted to be somatic were selected for validation by PCR and 3730 sequencing in flow-sorted CD45$^+$ and CD45$^-$ cells from the original d42m1 tumour. Genomic DNA and cDNA from CD45$^-$ (tumour) cells, and cDNA from CD45$^+$ (normal immune) cells were used for PCR amplification and then PCR products were sequenced individually on ABI 3730 using universal primers. Manual review was performed using amplicon-based assembly in the Integrative Genomics Viewer (IGV)[34] to determine the somatic status for each site. Data are presented in Supplementary Table 4.

MHC class I epitope prediction. All missense mutations for each d42m1-related tumour or H31m1 were analysed for the potential to form MHC class I neoepitopes that bind to either H-2D^b or H-2K^b molecules. The artificial neural network (ANN) algorithm provided by the Immune Epitope Database and Analysis Resource (http://www.immuneepitope.org) was used to predict epitope binding affinities[20] and the results were ultimately expressed as affinity values ($1/IC_{50} \times 100$). Predicted strong affinity epitopes expressed in d42m1 regressor tumours are listed in Supplementary Table 5.

Phylogenetic analysis of tumour samples. Sequencing data from normal *Rag2*$^{-/-}$ fibroblasts, d42m1 parental cells, d42m1 regressor clones, d42m1 progressor clones, d42m1 escape tumours and H31m1 tumour cells were compared using PHYLogeny Inference Package (PHYLIP)[35] to generate a phylogenetic tree displaying the relatedness of each sample.

Antibodies. Anti-H-2K^b (B8-24-3) and anti-H-2D^b (B22/249) monoclonal antibodies were provided by T. H. Hansen (Washington University School of Medicine). Anti-CD4 (GK1.5), anti-CD8α (YTS169.4) monoclonal antibodies and control immunoglobulin (PIP, a monoclonal antibody specific for bacterial glutathione *S*-transferase) were produced from hybridoma supernatants and purified in endotoxin-free form by Protein G affinity chromatography (Leinco Technologies). Purified Rat IgG was purchased from Sigma (St. Louis). CD45–FITC, CD45–PE, CD8–APC and purified anti-CD16/32 were purchased from BioLegend.

cDNA library construction and screening. To generate a d42m1 tumour cell cDNA library, mRNA was isolated from parental d42m1 tumour cells using a QuickPrep mRNA Purification kit (Amersham), converted into cDNA using SuperScript II First Strand Synthesis System (Invitrogen) and inserted into the EcoRI site of the expression vector pcDNA3 (Invitrogen). The cDNA library was divided into pools of 100 bacterial colonies with 200–300 ng of DNA from each

pool transfected into 2.5×10^4 monkey COS cells engineered to ectopically express mouse H-2D^b (COS-D^b) cells using Lipofectamine 2000. After 48 h, 5×10^3 C3 CTL cells were added, and supernatants were assayed for IFN-γ release 24 h later by ELISA. A single positive cDNA clone was isolated after screening 120,000 cDNA colonies. The putative H-2D^b-binding peptide VAVVNQIAL was predicted using the algorithm available at the Immune Epitope Database and Analysis Resource, http://www.immuneepitope.org/. The peptides were produced by P. Allen and S. Horvath (Washington University School of Medicine).

Expression vectors. Full-length cDNA encoding wild-type spectrin-β2 and mutant spectrin-β2 were cloned from parental d42m1 tumour cells by RT–PCR using primer pairs 5′-TGAGACAGTCAAGATGACGACCACGGTAGCCACA-3′ and 5′-CGGGACAACAGGGAAGTTCACTTCTTCTTGCCGA-3′. Wild-type and mutant spectrin-β2 cDNA were subcloned from the TOPO-XL vector (Invitrogen) into the retrovirus (RV)–GFP vector[36]. To generate the RV–RFP vector, full-length cDNA encoding RFP was cloned from the pTurboRFP-C vector (Evrogen) by RT–PCR using primer pairs 5′- ATCTCAGAATTCATGAGC GAGCTGATCAAGGA-3′ and 5′-ATCTCAGGATCCTTATCTGTGCCCCA GTTTGCTAG-3′. RFP cDNA was then cloned into the RV vector. To remove candidate T-cell epitopes in RFP, the nucleotide A was replaced by G at position 334 in the cDNA, resulting in amino acid substitution N112D. Coding sequences of the constructs were verified by DNA sequencing (Big Dye method; Applied Biosciences). The dominant-negative version of the IFNGR1 subunit (IFNGR1ΔIC) was expressed into H31m1 and d42m1 tumour cells as previously described[37].

Establishment of CTL lines and clones. To generate the d42m1-specific C3 CTL clone, wild-type mice were injected with 1×10^6 parental d42m1 tumour cells. Fourteen days later, the spleen was harvested from a mouse that rejected the tumour and a CTL line was established by stimulating 40×10^6 splenocytes with 2×10^6 parental d42m1 tumour cells pre-treated for 48 h with 100 U ml^{-1} of recombinant murine IFN-γ and irradiated (100 Gy). After CD8$^+$ T-cell purification using magnetic beads (Miltenyi Biotec) and limiting dilution, the CTL clone C3 was obtained.

Measurement of IFN-γ production. To generate target cells, tumour cells were treated with 100 U ml^{-1} IFN-γ for 48 h and irradiated with 100 Gy before use. The C3 CTL clone was co-cultured at the indicated ratios with target tumour cells (10,000 or 5,000 cells) in 96-well round-bottomed plates overnight. IFN-γ in supernatants was quantified using an IFN-γ ELISA kit (eBioscience). For blocking assays, 10 µg ml^{-1} of anti-CD8 (YTS-169.4), anti-CD4 (GK1.5) or control immunoglobulin (PIP) were added to the cell culture of effector (C3 CTL clone) and target cells (tumours).

Cytotoxicity assay. To generate target cells, tumour cells were treated with 100 U ml^{-1} rMuIFN-γ for 48 h before use. One million tumour cells were labelled with 25 µCi of Na$_2$^{51}CrO4 (PerkinElmer) for 90 min at 37 °C, washed and 10,000 cells seeded per well in 96-well round-bottom plates. The C3 CTL clone was co-cultured with the tumour target cells at the indicated effector/target cell ratios and incubated for 4 h at 37 °C in 5% CO$_2$. Radioactivity was detected in the supernatants and per cent specific killing was defined as (experimental condition c.p.m. − spontaneous c.p.m.) / (maximal (detergent) c.p.m. − spontaneous c.p.m.) × 100. Data points were obtained in duplicate.

Fluorescence-activated cell sorting analysis. For flow cytometry, cells were stained for 20 min at 4 °C with 500 ng of Fc block (anti-CD16/32) and 200 ng of CD45, CD4 or CD8α in 100 µl of staining buffer (PBS with 1% FCS and 0.05% NaN$_3$ (Sigma)). Propidium iodide (PI) (Sigma) was added at 1 µg ml^{-1} immediately before FACS analysis. For quantitative analysis of tumour-infiltrating lymphocytes/leukocytes (TIL) and lymph node populations, a CD45$^+$ PI$^-$ gate was used and gated events were collected on a FACSCalibur (BD Biosciences) and analysed using FloJo software.

Tumour, draining lymph node and spleen harvest. After tumour cell transplantation, established tumours were excised from mice, minced and treated with 1 mg ml^{-1} type IA collagenase (Sigma) in HBSS (Hyclone) for 2 h at room temperature (22 °C). The ipsilateral inguinal tumour draining lymph nodes and spleen were also harvested and crushed between two glass slides and vigorously resuspended to make single-cell suspensions.

Tetramers. H-2D^b tetramers conjugated to PE were prepared with mutant spectrin-β2 peptides and produced by the NIH Tetramer Core Facility (Emory University).

Mutation-specific RT–PCR and real-time RT–PCR. Total RNA from tumour cells was isolated by RNeasy Mini kit (Qiagen) and cDNA was synthesized from the total RNA using oligo (dT) primers and SuperScript II Reverse Transcriptase (Invitrogen). Real-time PCR specific for wild-type spectrin-β2, mutant spectrin-β2 and GAPDH using the SYBR Green Mastermix kit (Applied Biosystems) were performed on ABI 7000. The primer sequences for used for mutant spectrin-β2 are 5′-GGTGAACCAGATTGCACT-3′ and 5′-TGTCCACCAGTTCTCTGAACT-3′.

Detection of mutation in spectrin-β2 cDNA. The point mutation in the spectrin-β2 gene creates a PstI restriction site (CGGCAG to CTGCAG, underlined letters indicate the site of mutation). To amplify spectrin-β2 cDNA we used a forward primer (ACCCTGGCCCTGTACAAGAT) and reverse primer (TAGACTCGATGACCTTGGTCT). The PCR conditions used were 94 °C for 2 min, followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s. The PCR products were digested for 2 h at 37 °C with PstI restriction enzyme, which cleaved mutant spectrin-β2, but not wild-type spectrin-β2, and generates a 200 bp fragment from cDNA. The products were resolved by electrophoresis on a 1.2% agarose gel and visualized by ethidium bromide staining.

Isolation of non-transformed cells from d42m1 biopsy. A frozen d42m1 tumour biopsy from the original d42m1 tumour was thawed and treated with 1 mg ml^{-1} type IA collagenase (Sigma) in HBSS for 2 h at room temperature. After filtration, single-cell suspensions were stained for 20 min at 4 °C with 500 ng of Fc block (anti-CD16/32) and 200 ng of CD45-PE in 100 µl of staining buffer. Propidium iodide was added at 1 µg ml^{-1} immediately before sorting. A CD45$^+$ PI$^-$ gate was used and the top 15% and the bottom 15% of gated events were collected using a FACSAria II (BD Biosciences). Sorted CD45$^+$ cells (host leukocytes) and CD45$^-$ cells (primary d42m1 tumour cells) were collected and genomic DNA as well as RNA was isolated to synthesize cDNA for 3730 sequencing to validate that the mutation calls detected by Illumina were somatic and tumour specific.

Statistical analysis. Samples were compared using an unpaired, two-tailed Student's t-test, unless specified.

28. Huang, S. et al. Immune response in mice that lack the interferon-γ receptor. Science **259**, 1742–1745 (1993).
29. Muller, U. et al. Functional role of type I and type II interferons in antiviral defense. Science **264**, 1918–1921 (1994).
30. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**, 1754–1760 (2009).
31. Koboldt, D. C. et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics **25**, 2283–2285 (2009).
32. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. Genome Res. **11**, 1725–1729 (2001).
33. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**, 2078–2079 (2009).
34. Robinson, J. T. et al. Integrative genomics viewer. Nature Biotechnol. **29**, 24–26 (2011).
35. Felsenstein, J. Phylogeny Inference Package. Cladistics **5**, 164–166 (1989).
36. Ranganath, S. et al. GATA-3-dependent enhancer activity in IL-4 gene regulation. J. Immunol. **161**, 3822–3826 (1998).
37. Dighe, A. S., Richards, E., Old, L. J. & Schreiber, R. D. Enhanced in vivo growth and resistance to rejection of tumor cells expressing dominant negative IFNγ receptors. Immunity **1**, 447–456 (1994).

LETTER

Extrathymically generated regulatory T cells control mucosal T_H2 inflammation

Steven Z. Josefowicz[1,2]*, Rachel E. Niec[1]*, Hye Young Kim[3], Piper Treuting[4], Takatoshi Chinen[1,5], Ye Zheng[6], Dale T. Umetsu[3] & Alexander Y. Rudensky[1]

A balance between pro- and anti-inflammatory mechanisms at mucosal interfaces, which are sites of constitutive exposure to microbes and non-microbial foreign substances, allows for efficient protection against pathogens yet prevents adverse inflammatory responses associated with allergy, asthma and intestinal inflammation[1]. Regulatory T (T_{reg}) cells prevent systemic and tissue-specific autoimmunity and inflammatory lesions at mucosal interfaces. These cells are generated in the thymus (tT_{reg} cells) and in the periphery (induced (i)T_{reg} cells), and their dual origin implies a division of labour between tT_{reg} and iT_{reg} cells in immune homeostasis. Here we show that a highly selective blockage in differentiation of iT_{reg} cells in mice did not lead to unprovoked multi-organ autoimmunity, exacerbation of induced tissue-specific autoimmune pathology, or increased pro-inflammatory responses of T helper 1 (T_H1) and T_H17 cells. However, mice deficient in iT_{reg} cells spontaneously developed pronounced T_H2-type pathologies at mucosal sites—in the gastrointestinal tract and lungs—with hallmarks of allergic inflammation and asthma. Furthermore, iT_{reg}-cell deficiency altered gut microbial communities. These results suggest that whereas T_{reg} cells generated in the thymus appear sufficient for control of systemic and tissue-specific autoimmunity, extrathymic differentiation of T_{reg} cells affects commensal microbiota composition and serves a distinct, essential function in restraint of allergic-type inflammation at mucosal interfaces.

Exquisitely balanced control mechanisms operating at mucosal sites are able to accommodate potent immune defences and the need to prevent tissue damage resulting from inflammatory responses caused by commensal microorganisms, food and environmental antigens, allergens, and noxious substances[1].

Prominent among multiple regulatory lymphoid and myeloid cell subsets operating at environmental interfaces are Foxp3+ T_{reg} cells. Genetic deficiency in Foxp3 (forkhead box P3, a key transcription factor specifying T_{reg} cell differentiation) leads to paucity of Foxp3+ T_{reg} cells and consequent generalized lympho- and myelo-proliferative syndrome, featuring sharply augmented serum IgE levels, production of T_H1, T_H2 and T_H17 cytokines, and widespread tissue inflammation[2]. Foxp3 can be induced in thymocytes in response to T-cell receptor (TCR) and CD28 stimulation, and IL-2. In addition, Foxp3 can be upregulated upon TCR stimulation of mature peripheral CD4+ T cells in the presence of tumour growth factor β (TGFβ) in a manner dependent on an intronic *Foxp3* enhancer CNS1 (refs 3–5). Inflammatory cytokines and potent co-stimulatory signals antagonize the peripheral induction of Foxp3, and retinoic acid augments Foxp3 induction through mitigating inflammatory cytokine production and through cell intrinsic mechanisms[1,6–8]. Although differing in their sites of generation, tT_{reg} and iT_{reg} cells are comingled in the secondary lymphoid organs and non-lymphoid tissues once mature, and their relative contributions to the total population of T_{reg} cells and their specific roles in control of various aspects of immune homeostasis and microbial colonization in normal animals has remained unexplored.

Our recent investigation[5] showed that CNS1, which contains binding sites for transcription factors (NFAT, Smad3 and RAR/RXR) downstream of three signalling pathways implicated in iT_{reg} cell generation[4,8] (Supplementary Fig. 1), is critical for TGFβ-dependent induction of Foxp3, but has no apparent role in tT_{reg} differentiation or maintenance of Foxp3 expression. This observation suggested that CNS1 activity represents a dedicated genetic determinant for the differentiation of iT_{reg} cells, and its deficiency in mice provides a unique means to evaluate the function of these cells *in vivo*. Our initial characterization of CNS1− mice and littermates maintained on a 129/B6 genetic background failed to reveal disease phenotypes. Because mixed genetic backgrounds frequently mask adverse phenotypes or make them highly variable, to understand iT_{reg} function *in vivo* we back-crossed CNS1 mice onto the B6 background (Supplementary Fig. 2).

First, we sought to ascertain that on the B6 genetic background CNS1 is dispensable for tT_{reg} cell generation but critical for generation of iT_{reg} cells. Two recent studies established a role for TGFβ signalling in tT_{reg} cell differentiation in neonates[9,10]. Thus, to exclude the possibility that CNS1 deficiency adversely affects generation of Foxp3+ T cells in the neonatal thymus, we examined the Foxp3+ T_{reg} cell population in heterozygous female $CNS1^{WT/-}$ mice. As Foxp3 is encoded on the X chromosome and is subject to random X-chromosome inactivation, characterization of female $CNS1^{WT/-}$ mice allows for comparison of CNS1− and $CNS1^{WT}$ T_{reg} cells in a competitive environment. In neonatal female $CNS1^{WT/-}$ mice, CNS1− cells constituted, on average, one-half of the thymic Foxp3+ cell population (Fig. 1a). Additionally, neonatal CNS1− hemizygous and control males harboured comparable numbers of Foxp3+ thymocytes (Supplementary Fig. 3). Therefore, tT_{reg} differentiation is independent of CNS1. In contrast, CNS1− naive CD4 T cells showed severely impaired induction of Foxp3 *in vitro* (Fig. 1b). Analyses of heterozygous female $CNS1^{WT/-}$ mice and transfer of CNS1− or $CNS1^{WT}$ T_{reg} cells into lymphopenic recipients demonstrated that the ability of T_{reg} cells to accumulate and proliferate in various tissues was unperturbed in the absence of CNS1 (Supplementary Fig. 4). Furthermore, CNS1 deficiency did not affect suppressor activity of tT_{reg} cells (assessed using *in vitro* suppression assays and adoptive transfers of Foxp3-deficient effector T cells with predominantly tT_{reg}-containing Foxp3+ cells isolated from 4-week-old CNS1− and $CNS1^{WT}$ mice into lymphopenic recipients (Supplementary Fig. 5)). Likewise, CNS1 ablation did not negatively affect maintenance of Foxp3 expression and overall function of NFAT, TGFβ and retinoic acid signalling pathways in these cells (Supplementary Fig. 5 and data not shown). To assess how the deficiency in iT_{reg} cell generation affects the size of the peripheral T_{reg} cell compartment, we analysed T_{reg} cell frequencies in various tissues throughout the lifespan of mice. CNS1− mice failed to exhibit a progressive age-dependent increase in Foxp3+
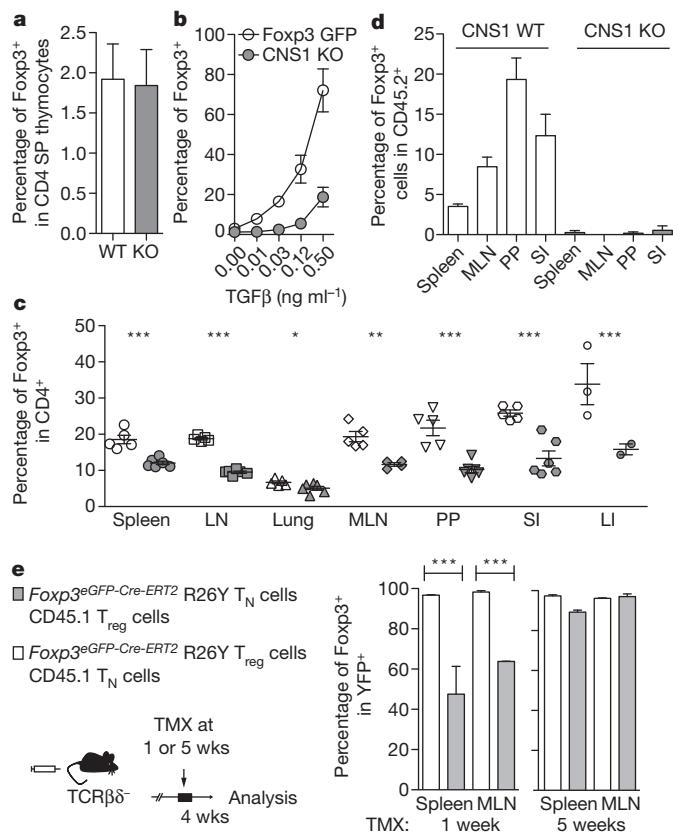
Figure 1 | Impaired iT$_{reg}$ cell generation and altered composition of the peripheral T$_{reg}$ cell population in CNS1-deficient mice. a, Relative contribution of CNS1$^-$ (GFP$^+$) and CNS1^{WT} (GFP$^-$) cells to the Foxp3$^+$ thymocyte subset in 4-day-old CNS1$^{WT/-}$ female mice. SP, single positive. **b**, Induction of Foxp3 in Foxp3$^-$ T$_N$ (naive) cells FACS sorted from CNS1$^-$ (knockout, KO) or Foxp3^{GFP} mice stimulated in vitro with TGFβ, IL-2 anti-CD3 and anti-CD28. **c**, Percentage of Foxp3$^+$ cells (of CD4$^+$) in the spleen, lymph node (LN), mesenteric lymph nodes (MLN), Peyer's patches (PP) and cells from the small and large intestine lamina propria (SI and LI) of 6–9 month old CNS1$^-$ or control mice. **d**, Percentage of transferred (CD45.2$^+$) CNS1$^-$ or CNS1$^+$ CD25$^-$CD44^{low}CD45.2$^+$OTII$^+$ cells that induced Foxp3 following administration of OVA in water for 6 days. **e**, Stability of Foxp3 expression in iT$_{reg}$ cells. FACS sorted GFP$^+$ or GFP$^-$ cells from Foxp3$^{eGFP-Cre-ERT2}$ mice were transferred with GFP$^-$ or GFP$^+$ cells, respectively, from CD45.1 Foxp3^{GFP} mice into TCRβδ-deficient recipients. Mice received tamoxifen (TMX) at 1 (left) or 5 weeks (right) after transfer and stability of Foxp3 expression among YFP-labelled cells was assessed after 4 weeks. All data are representative of two or more independent experiments with $n \geq 3$. Error bars, s.d.; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, as calculated by Students' t-test.

cell frequencies observed in wild-type littermates (Fig. 1c and Supplementary Fig. 6). By 6–8 months of age, CNS1$^-$ mice contained markedly fewer Foxp3$^+$ cells in comparison to control animals, with most prominent differences in mesenteric lymph nodes, Peyer's patches, and small and large intestine lamina propria, sites known to support iT$_{reg}$ cell generation[11]. This trend was not the result of expression of a Foxp3–GFP fusion protein in CNS1$^-$ mice, because age-matched CNS1^{WT} Foxp3–GFP and littermate control CNS1^{WT} mice expressing unmodified Foxp3 protein exhibited similar age-dependent increases in T$_{reg}$ cell frequencies (Supplementary Fig. 6).

To assess the extent of impairment of peripheral generation of T$_{reg}$ cells in vivo, we examined Foxp3 induction in antigen-specific naive T cells upon exposure to ingested 'non-self' antigen[12]. Ovalbumin (OVA)-specific OT-II$^+$ TCR-transgenic Foxp3$^-$ (GFP$^-$) T$_{reg}$ cells from CNS1$^-$ or Foxp3^{GFP} mice were transferred into CD45.1$^+$ lymphoreplete recipients followed by ad libitum administration of OVA in drinking water. We failed to detect Foxp3 induction in CNS1-deficient cells, whereas up to 20% of transferred OT-II T cells

from control Foxp3^{GFP} mice induced Foxp3 upon exposure to cognate antigen in the intestinal tract (Fig. 1d and Supplementary Fig. 7). These results were in agreement with a marked impairment in Foxp3 induction in polyclonal CNS1-deficient Foxp3$^-$ T cells in vitro, which was most severe at lower, more physiologically relevant concentrations of TGFβ (Fig. 1b). Together these data indicate that iT$_{reg}$ cells have a stringent requirement for CNS1 for their differentiation.

Recent studies showed a limited TCR-dependent clonal niche for tT$_{reg}$ cell differentiation and peripheral maintenance[13–15]. The sustained numerical impairment in the peripheral T$_{reg}$ cell populations in CNS1-deficient mice suggests that tT$_{reg}$ cells fail to fill the 'void' in the peripheral T$_{reg}$ cell pool, left by iT$_{reg}$ cell deficiency. This observation combined with largely non-overlapping TCR repertoires of tT$_{reg}$ and iT$_{reg}$ cells suggests that iT$_{reg}$ and tT$_{reg}$ cells occupy distinct 'niches'[16]. To test this notion we co-transferred CNS1$^-$ (tT$_{reg}$ cells) or CNS1^{WT} T$_{reg}$ cells (iT$_{reg}$ + tT$_{reg}$) from aged mice with CNS1-sufficient naive CD45.1$^+$Foxp3$^-$CD4$^+$ T cells into lymphopenic recipients. We observed more efficient Foxp3 induction in CD45.1$^+$CD4$^+$ T cells upon co-transfer with CNS1$^-$ T$_{reg}$ cells (tT$_{reg}$ cells), indicating that in lymphopenic recipients the de novo generation of iT$_{reg}$ cells is markedly more efficient in the absence of pre-existing iT$_{reg}$ cells (Supplementary Fig. 8). These data also imply the existence of a stable iT$_{reg}$ cell subset in normal mice. However, the dynamics and stability of Foxp3 expression has been a controversial issue, with a number of studies favouring unstable Foxp3 expression in iT$_{reg}$ cells[17–19]. Thus, we next employed genetic fate mapping using inducible Cre recombinase expressed in a T$_{reg}$-specific manner (Foxp3$^{EGFP-CRE-ERT2}$) and a Rosa26–YFP recombination reporter allele (R26Y)[20] to determine if iT$_{reg}$ cells generated in vivo are able to acquire stable Foxp3 expression and, thus, have the capacity to contribute to the stable T$_{reg}$ cell compartment.

Double-sorted naive CD45.2$^+$Foxp3$^-$YFP$^-$ CD4 T cells from Foxp3$^{EGFP-CRE-ERT2}$ R26Y mice were transferred together with congenically marked CD45.1$^-$Foxp3$^+$ T$_{reg}$ cells into lymphopenic recipient mice. Foxp3 expression within the population of tagged YFP$^+$ cells generated from YFP$^-$Foxp3$^-$ precursors was assessed four weeks after treatment of recipient mice with tamoxifen, which was administered early (one week) and late (five weeks) following cell transfer. Approximately half of the newly generated YFP-tagged iT$_{reg}$ cells lost Foxp3 expression, whereas 'mature' iT$_{reg}$ cells tagged at a later time point displayed remarkable stability (>90% Foxp3$^+$ cells among YFP$^+$ cells), comparable to that of transferred peripheral T$_{reg}$ cells (Fig. 1e and Supplementary Fig. 9). Together these data indicate that iT$_{reg}$ cells have a stringent requirement for CNS1 for their differentiation, accumulate throughout life, and occupy a sizable fraction of the stable peripheral T$_{reg}$ cell compartment.

CNS1$^-$ mice on the B6 genetic background displayed neither early-nor late-onset systemic autoimmunity nor spontaneous widespread tissue lesions nor severe morbidity associated with systemic T$_{reg}$ cell deprivation (data not shown). However, it was possible that iT$_{reg}$ cell deficiency may exacerbate initial or late stages of provoked tissue-specific autoimmune pathology directed against a self-antigen. To address this question, we induced experimental autoimmune encephalomyelitis (EAE) in CNS1-deficient or littermate control mice through immunization with myelin oligodendrocyte glycoprotein (MOG) peptide. The onset, severity and remission of disease were indistinguishable, and no detectable differences were observed in T$_{reg}$ cell subsets in the brain in these two groups of mice (Supplementary Fig. 10). Although it will be important to evaluate the role of iT$_{reg}$ cells in additional models of induced autoimmunity, these results indicate that tT$_{reg}$ cells are largely sufficient for control of tolerance to self-antigens and that the distinct functional role of iT$_{reg}$ cells might be to control inflammation at mucosal surfaces, which are sites of preponderant exposure to non-self substances. This notion is consistent with data indicating that tT$_{reg}$ cells arise from a subset of thymocytes, which exhibit TCR with an increased affinity for self-antigens yet insufficient for negative selection[10,21], whereas iT$_{reg}$ cells are efficiently generated

upon TCR engagement with a high affinity cognate ligand under subimmunogenic conditions[22,23].

The absence of iT$_{reg}$ cell induction in response to oral antigen in CNS1$^-$ mice suggested that the immune balance in the gastrointestinal tract might be impaired owing to deficiency in gut antigen-specific iT$_{reg}$ cells. Indeed, while IL-17 and IFN-γ production by CD4$^+$ T cells was unaffected by iT$_{reg}$ deficiency in CNS1$^-$ mice (Supplementary Fig. 11), we observed markedly augmented production of the T$_H$2 cytokines, IL-4, IL-5 and IL-13, by CD4$^+$ T cells, especially in the mesenteric lymph nodes, Peyer's patches and intestinal lamina propria (Fig. 2a and Supplementary Fig. 12). Furthermore, the vast majority of CD4$^+$ T cells in the lamina propria of CNS1$^-$ mice expressed high amounts of Gata3, a key T$_H$2 differentiation factor. Increases in Gata3$^+$CD4$^+$ T cells were observed not only in gastrointestinal tract tissues in CNS1$^-$ mice but also in other lymphoid tissues, albeit to a lesser extent (Fig. 2b and Supplementary Fig. 12). Consistent with the sharply augmented T$_H$2 responses at mucosal sites, CNS1$^-$ mice exhibited increased frequencies of germinal centre B cells (Fas$^+$GL7$^+$) in the Peyer's patches, but not in the spleen or peripheral lymph nodes (Supplementary Fig. 13), and spontaneous increases in serum levels of IgE and IgA, but not in other Ig isotypes (Fig. 2c, and data not shown).

The dysregulated T$_H$2 responses were associated with a decreased body weight (Fig. 3a and Supplementary Fig. 2) and distinct highly penetrant pathology throughout the gastrointestinal tract (Fig. 3b and Supplementary Fig. 14): all CNS1$^-$ mice (12/12) and no CNS1^{WT} control littermates (0/6) were affected by gastritis and plasmacytic enteritis characterized by increased frequencies of plasma cells in the intestinal lamina propria and other associated lesions such as crypt abscesses. Accordingly, serum antibodies in CNS1$^-$ mice exhibited reactivity against antigens of the small and large intestine, pancreas and chow (Supplementary Fig. 13). Notably, the pathology observed in the gastrointestinal tissue of CNS1$^-$ mice was markedly diminished upon B-cell depletion, but was not ameliorated by administration of IL-4 neutralizing antibody (Supplementary Fig. 15). The inflammatory

features and lesions observed in CNS1$^-$ mice were consistent with allergic T$_H$2-type intestinal disease (Fig. 3).

One possible explanation for the pronounced T$_H$2 responses and intestinal pathology associated with iT$_{reg}$ cell deficiency is simply a numerical decrease in T$_{reg}$ cells. However, we consider this possibility unlikely, because graded depletion of Foxp3$^+$ T$_{reg}$ cells in Foxp3^{DTR} mice upon administration of titrated amounts of diphtheria toxin resulting in T$_{reg}$ frequencies similar to those observed in CNS1$^-$ mice revealed augmented T$_H$1 and T$_H$17, but not T$_H$2, responses[24]. Alternatively, certain qualitative features of iT$_{reg}$ cells could allow them to efficiently limit T$_H$2 inflammation in the gut. Recent studies suggested that some of the transcriptional regulators involved in a particular type of effector T-cell response facilitate the ability of T$_{reg}$ cells to suppress those responses[25–27]. Thus, we explored the expression of T$_H$2-associated transcription factor Gata3 in T$_{reg}$ cells in CNS1$^-$ and CNS1^{WT} mice. In contrast to a sharp increase in Gata3 expression in effector T cells (Fig. 2b and Supplementary Fig. 12), we found its expression markedly diminished in T$_{reg}$ cells in CNS1$^-$ mice (Fig. 3c and Supplementary Fig. 12). Notably, ablation of a conditional Gata3 allele in T$_{reg}$ cells leads to T$_{reg}$ cell dysfunction[28,29] and marked augmentation of T$_H$2 cytokine production by CD4$^+$ T cells (D. Rudra, R.E.N. and A.Y.R., manuscript in preparation). We hypothesized that increased Gata3 expression in iT$_{reg}$ cells reflects their activation state upon TCR ligation by high affinity ligands in the gut rather than an intrinsic feature of iT$_{reg}$ cells. In support of this idea, we found that both CNS1$^-$ and control T$_{reg}$ cells stimulated in vitro through the TCR and IL-2 receptor exhibited similarly robust Gata3 induction (Supplementary Fig. 12). Thus, we suggest that increased Gata3 expression in iT$_{reg}$ cells, a likely consequence of their generation in response to high affinity TCR ligands present in the gut, endows these cells with the capacity to efficiently control spontaneous mucosal T$_H$2 inflammation.

Certain commensal bacteria increase the frequencies of T$_{reg}$ cells in the gut and provide antigens recognized by a considerable proportion of iT$_{reg}$ TCR[1,16]. In addition to TCR ligands the gut microbial community
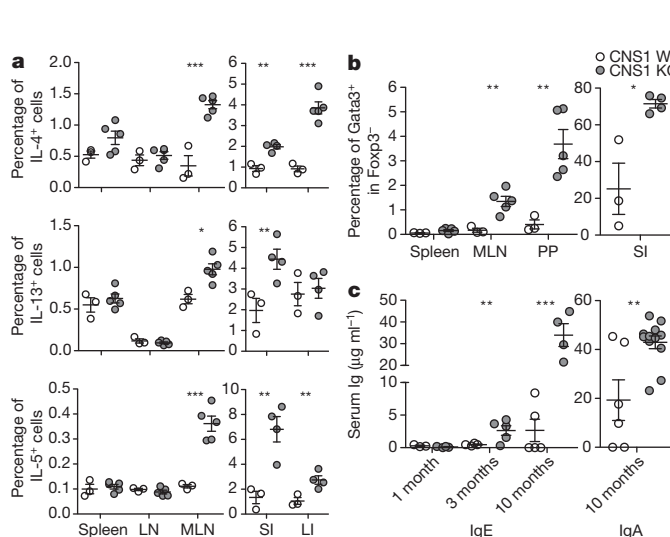

Figure 2 | Paucity of iT$_{reg}$ cells results in T$_H$2 inflammation in the gastrointestinal tract. a, Percentage of CD4$^+$ cells producing IL-4 (top), IL-13 (middle) and IL-5 (bottom) in 3-month-old mice. Left, spleen, peripheral lymph nodes (LN) and mesenteric lymph nodes (MLN); right, lamina propria of small and large intestine (SI and LI, respectively). b, Percentage of Foxp3$^-$ CD4$^+$ cells that were Gata3$^+$ in 3-month old mice (PP, Peyer's patches). c, Concentration of IgE and IgA in serum, determined by enzyme linked immunosorbent assay (ELISA) at 1, 3 and 10 months. All data are representative of three or more independent experiments with ≥3 mice per group. Error bars, s.d.; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, as calculated by Students' t-test.

Figure 3 | iT$_{reg}$ cell deficiency leads to T$_H$2 type gastrointestinal pathology and altered microbial communities. a, Body weights of 9–12 (left) or 2.5-month-old individually housed (right) CNS1$^-$ (KO) and littermate control (WT) mice ($n ≥ 12$). b, Plasmacytic enteritis (arrowhead) in CNS1-deficient mice revealed by haematoxylin and eosin staining of small intestine from 9–12-month-old CNS1$^-$ (bottom and right) and littermate control mice (top). An early crypt abscess is indicated (asterisk). Data are representative of ≥20 mice analysed. c, Percentage of Foxp3$^+$ CD4$^+$ cells expressing Gata3$^+$ in 3-month-old mice. d, Percentage of total 16S rRNA gene sequences of the Firmicutes and Bacteroidetes phyla in stool from individually housed CNS1$^-$ ($n = 9$) and WT ($n = 6$) littermate mice. All data are representative of three or more independent experiments with ≥3 mice per group. Error bars show s.d. (a, c) or s.e.m. (d). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, as calculated by Student's t-test. Scale bars, 150 μm.

Figure 4 | Unprovoked asthma-like airway pathology in CNS1-deficient mice. a, Representative haematoxylin and eosin-stained lung sections from CNS1$^-$ (top) and WT (bottom) mice. The CNS1$^-$ lung has marked peribronchiolar inflammation (arrowhead). The reduced lumen (L) contains mucus produced by the hyperplastic respiratory epithelium (E). Arrows indicate reactive (top) and normal (bottom) endothelium. Bottom right hand corner insets are higher magnification of boxed regions and bar indicates smooth muscle thickness. Top right inset (KO) demonstrates eosinophilic crystals. Asterisk marks acidophilic macrophages. **b**, Periodic acid Schiff with Alcian Blue staining highlighting mucus-producing goblet cells (dark blue-purple) **c**, Trichrome staining illustrating lung fibrosis (blue staining). **d**, Arginase-1 staining of lungs from CNS1$^-$ and WT mice. A indicates airway; an acidophilic crystal is marked by the arrowhead. **e**, Chitinase 3-like 3 (Chi3l3) staining of lungs from CNS1$^-$ and WT mice at 10× magnification (top) and 60× magnification of lungs from CNS1$^-$ mice demonstrating robust Chi3l3 expression within acidophilic macrophages (bottom). **f**, Lung resistance (left) and compliance (right) of CNS1$^-$ and WT littermate control mice after exposure to methacholine. Data representative of two independent experiments with ≥4 mice per group. Error bars, s.d.; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, as calculated by Students' t-test. Scale bars, 100 μm.
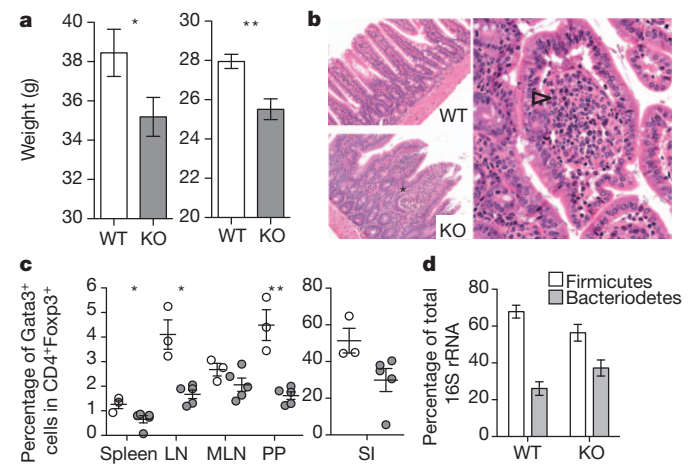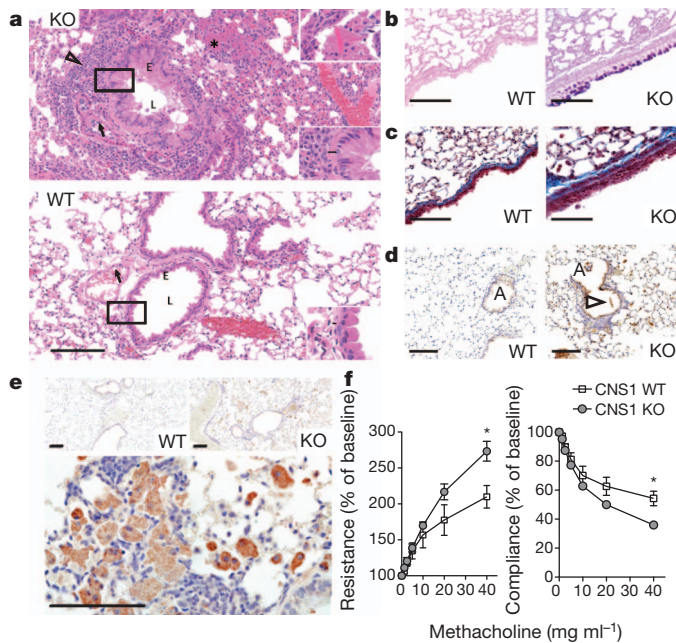
also contributes to the local cytokine environment, which facilitates iT$_{reg}$ cell differentiation and maintenance in the gut[1]. These observations raise a question as to whether iT$_{reg}$ cells, in turn, influence composition of the commensal microbiota. To address this question, we sequenced 16S ribosomal RNA coding genes from bacterial contents of stool samples isolated from CNS1$^-$ and CNS1^{WT} littermates, which were housed individually for 5 weeks after weaning. Phylogenetic analysis revealed distinct gut microbial communities in CNS1$^-$ mice, with statistically significant enrichment of the candidate phylum TM7 and the genus Bacteroidetes *Alistipes* (Supplementary Fig. 16), and an overall decrease in the ratio of Firmicutes to Bacteroidetes (2.60 in wild-type and 1.51 in knockout) (Fig. 3d). Interestingly, an opposite trend in the Firmicutes/Bacteroidetes ratio was correlated with obesity[30], suggesting the possibility that alterations in energy harvest and metabolism (caused by inflammation or microbe-dependent effects on energy balance) could account for the decreased weight observed in iT$_{reg}$ cell deficient mice. Thus, iT$_{reg}$ cells help maintain a 'normal' microbial community in the gut, probably through exerting control over T$_H$2 mucosal inflammation.

These observations raised the question of whether the altered microbiota, rather than iT$_{reg}$ deficiency, was the direct cause of observed T$_H$2 inflammation. To equalize gut microbiota, CNS1$^-$ and littermate controls were treated with antibiotics (metronidazole and ciprofloxacin) for 4 weeks. Despite indistinguishable microbial communities,

antibiotic treatment did not lead to a decrease in Gata3 expression or Th2 cytokine production by effector T cells in CNS1$^-$ mice, and characteristic histopathologic features were maintained (Supplementary Fig. 17). Furthermore, iT$_{reg}$ cell sufficient germ-free mice colonized with CNS1$^-$ or control microbiota exhibited a similar spectrum of T$_H$1, T$_H$2 and T$_H$17 cytokine production and eventual normalization of microbiota (Supplementary Fig. 18 and data not shown). These results suggest that iT$_{reg}$ deficiency results in immune dysregulation and T$_H$2 inflammation in the gut with subsequent perturbation of the microbial community.

According to the notion of specialized iT$_{reg}$ cell function in suppression of T$_H$2 responses at mucosal sites, one would expect to observe T$_H$2-type pathology in the lungs of CNS1$^-$ mice, despite an only modest ~20–25% decrease in numbers of T$_{reg}$ cells in this tissue compared to littermate controls (Fig. 1c). Indeed, we discovered that CNS1$^-$ mice suffer from spontaneous T$_H$2-type airway inflammation (Fig. 4 and Supplementary Fig. 19). The lungs of CNS1$^-$ mice were characterized by increased infiltration by lymphocytes, plasma cells and macrophages, and by moderate neutrophil infiltration (Fig. 4). The consistent features of the chronic inflammatory airway disease observed in CNS1$^-$ mice include lymphocytic infiltration, narrowed airway lumen (Fig. 4a), increased goblet cells and mucus production (Fig. 4a and b), smooth muscle hyperplasia, and fibrosis (Fig. 4c). Notably, 9/12 CNS1$^-$ and 0/6 CNS1^{WT} mice developed acidophilic macrophage pneumonia (AMP) with characteristic increases in acidophilic macrophages and both intracellular and extracellular chitinase 3-like 3 crystals (Chi3l3, formerly Ym1), analogous to Charcott-Lyden crystals found in asthmatic patients (Fig. 4a and e). In addition, the prominent presence of alternatively activated macrophages in the lungs of CNS1$^-$ mice was confirmed by morphology and expression of arginase 1 in addition to Chi3l3 (Fig. 4d and Supplementary Fig. 20). Furthermore, both young (6–8 week old) and aged (20 week old) CNS1$^-$ mice exhibited airway hyper-responsiveness accompanied by AMP, perivascular, peribronchiolar and intramucosal inflammation, bronchial epithelial hyperplasia, and airway narrowing (Fig. 4f and Supplementary Fig. 21). These spontaneous lesions are especially striking considering the T$_H$2-resistant, T$_H$1-prone C57BL/6 genetic background of CNS1$^-$ mice. The lung pathology in CNS1$^-$ mice reflects the hallmark features of chronic allergic inflammation and asthma.

Our results demonstrate that T$_{reg}$ cells of thymic and extrathymic origin have distinct mechanistic requirements for differentiation and exert specialized functions in immune homeostasis. The restriction of lesions to mucosal tissues in iT$_{reg}$ deficient mice implies that under steady state conditions T$_{reg}$ cells generated in the thymus are largely sufficient for control of most immune responses to self-antigens. These findings suggest that in normal animals, T$_{reg}$ cells generated extrathymically in a CNS1-dependent manner play a non-redundant role in control of mucosal allergic Th2 inflammation and asthma.

METHODS SUMMARY

The generation of the following mouse strains has been previously described[5,20]: CNS1$^-$ (*Foxp3*$^{\Delta CNS1}$), *Foxp3*^{GFP} and *Foxp3*$^{eGFP-Cre-ERT2}$ R26Y. *Rag1$^-$* mice were purchased from The Jackson Laboratory, and CD45.1 B6 and *Tcrb/Tcrd$^-$* mice, along with above strains were maintained in the Sloan Kettering Institute Research Laboratories animal facility in accordance with institutional regulations. Tissues for histologic analysis were fixed in 10% phosphate-buffered formalin and processed routinely for staining. *In vitro* induction assays were performed with 5×10^4 Foxp3–GFP$^-$ CD4$^+$ T cells and 5 μg ml^{-1} of anti-CD3 and anti-CD28 antibody, 100 U ml^{-1} IL-2, in 96-well, flat-bottom plates. For *in vitro* and transfer experiments, CD4$^+$ T cells were pre-enriched using mouse CD4 Dynabeads (L3T4, Invitrogen) and FACS sorted on an LSR-II (BD Biosciences). Intracellular staining for IL-4 used Cytofix/Cytoperm (BD Biosciences), and staining for other cytokines, Foxp3 and Gata3, used the Foxp3 staining kit (eBiosciences). For measurement of AHR, mice were anaesthetized with pentobarbitol and AHR was assessed by invasive measurement of airway resistance using modified version of a described method (Buxco Electronics). 16S rRNA sequencing was performed on a 454 GS

FLX Titanium pyrosequencing platform following the Roche 454 recommended procedures.

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

1. Maloy, K. J. & Powrie, F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature* **474,** 298–306 (2011).
2. Sakaguchi, S., Yamaguchi, T., Nomura, T. & Ono, M. Regulatory T cells and immune tolerance. *Cell* **133,** 775–787 (2008).
3. Chen, W. *et al.* Conversion of peripheral CD4⁺CD25⁻ naive T cells to CD4⁺CD25⁺ regulatory T cells by TGF-β induction of transcription factor Foxp3. *J. Exp. Med.* **198,** 1875–1886 (2003).
4. Tone, Y. *et al.* Smad3 and NFAT cooperate to induce Foxp3 expression through its enhancer. *Nature Immunol.* **9,** 194–202 (2008).
5. Zheng, Y. *et al.* Role of conserved non-coding DNA elements in the *Foxp3* gene in regulatory T-cell fate. *Nature* **463,** 808–812 (2010).
6. Hill, J. A. *et al.* Retinoic acid enhances Foxp3 induction indirectly by relieving inhibition from CD4⁺CD44ʰⁱ cells. *Immunity* **29,** 758–770 (2008).
7. Nolting, J. *et al.* Retinoic acid can enhance conversion of naive into regulatory T cells independently of secreted cytokines. *J. Exp. Med.* **206,** 2131–2139 (2009).
8. Xu, L. *et al.* Positive and negative transcriptional regulation of the *Foxp3* gene is mediated by access and binding of the Smad3 protein to enhancer I. *Immunity* **33,** 313–325 (2010).
9. Liu, Y. *et al.* A critical function for TGF-β signaling in the development of natural CD4⁺CD25⁺Foxp3⁺ regulatory T cells. *Nature Immunol.* **9,** 632–640 (2008).
10. Ouyang, W., Beckett, O., Ma, Q. & Li, M. O. Transforming growth factor-β signaling curbs thymic negative selection promoting regulatory T cell development. *Immunity* **32,** 642–653 (2010).
11. Curotto de Lafaille, M. A. & Lafaille, J. J. Natural and adaptive foxp3⁺ regulatory T cells: more of the same or a division of labor? *Immunity* **30,** 626–635 (2009).
12. Mucida, D. *et al.* Oral tolerance in the absence of naturally occurring Tregs. *J. Clin. Invest.* **115,** 1923–1933 (2005).
13. Bautista, J. L. *et al.* Intraclonal competition limits the fate determination of regulatory T cells in the thymus. *Nature Immunol.* **10,** 610–617 (2009).
14. Leung, M. W., Shen, S. & Lafaille, J. J. TCR-dependent differentiation of thymic Foxp3⁺ cells is limited to small clonal sizes. *J. Exp. Med.* **206,** 2121–2130 (2009).
15. Moran, A. E. *et al.* T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. *J. Exp. Med.* **208,** 1279–1289 (2011).
16. Lathrop, S. K. *et al.* Peripheral education of the immune system by colonic commensal microbiota. *Nature* **478,** 250–254 (2011).
17. Komatsu, N. *et al.* Heterogeneity of natural Foxp3⁺ T cells: a committed regulatory T-cell lineage and an uncommitted minor population retaining plasticity. *Proc. Natl Acad. Sci. USA* **106,** 1903–1908 (2009).
18. Hori, S. Regulatory T cell plasticity: beyond the controversies. *Trends Immunol.* **32,** 295–300 (2011).
19. Zhou, X. *et al.* Instability of the transcription factor Foxp3 leads to the generation of pathogenic memory T cells *in vivo. Nature Immunol.* **10,** 1000–1007 (2009).
20. Rubtsov, Y. P. *et al.* Stability of the regulatory T cell lineage *in vivo. Science* **329,** 1667–1671 (2010).
21. Josefowicz, S. Z. & Rudensky, A. Control of regulatory T cell lineage commitment and maintenance. *Immunity* **30,** 616–625 (2009).
22. Gottschalk, R. A., Corse, E. & Allison, J. P. TCR ligand density and affinity determine peripheral induction of Foxp3 *in vivo. J. Exp. Med.* **207,** 1701–1711 (2010).
23. Kretschmer, K. *et al.* Inducing and expanding regulatory T cell populations by foreign antigen. *Nature Immunol.* **6,** 1219–1227 (2005).
24. Tian, L. *et al.* Foxp3 regulatory T cells exert asymmetric control over murine helper responses by inducing Th2 cell apoptosis. *Blood* **118,** 1845–1853 (2011).
25. Chaudhry, A. *et al.* CD4⁺ regulatory T cells control T$_H$17 responses in a Stat3-dependent manner. *Science* **326,** 986–991 (2009).
26. Koch, M. A. *et al.* The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. *Nature Immunol.* **10,** 595–602 (2009).
27. Zheng, Y. *et al.* Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T$_H$2 responses. *Nature* **458,** 351–356 (2009).
28. Wang, Y., Su, M. A. & Wan, Y. Y. An essential role of the transcription factor GATA-3 for the function of regulatory T cells. *Immunity* **35,** 337–348 (2011).
29. Wohlfert, E. A. *et al.* GATA3 controls Foxp3⁺ regulatory T cell fate during inflammation in mice. *J. Clin. Invest.* **121,** 4503–4515 (2011).
30. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).

Author Contributions S.Z.J., R.E.N. and H.Y.K. performed experiments and analysed data, with assistance from T.C. for tissue Ig ELISA experiments, and P.T. for immunohistochemistry and histopathology analysis. D.T.U., S.Z.J., R.E.N., H.Y.K, and A.Y.R designed and interpreted AHR experiments. Y.Z. generated CNS1⁻ mice. S.Z.J., R.E.N. and A.Y.R. designed experiments and wrote the paper.

METHODS

Mice. The generation of the following mouse strains has been previously described[5,20]: CNS1⁻ ($Foxp3^{\Delta CNS1}$), $Foxp3^{GFP}$ and $Foxp3^{eGFP-Cre-ERT2}$ R26Y. $Rag1^-$ mice were purchased from The Jackson Laboratory, and CD45.1 B6 and $Tcrb/Tcrd^{-/-}$ mice, along with above strains were maintained in the Sloan Kettering Institute Research Laboratories animal facility in accordance with institutional regulations. Mice were killed by CO_2 asphyxiation. EAE was induced and scored as previously described[31]. For antibiotic treatment, CNS1-deficient and sufficient mice were treated with $1\,g\,l^{-1}$ metronidazole (Sigma-Aldrich) and $0.2\,g\,l^{-1}$ ciprofloxacin (ENZO Life Sciences International) dissolved in drinking water for 4 weeks. Mouse anti-CD20[8] (MB20-11, provided by T. Tedder) and anti-IL-4 (11b.11, NCI-Frederick) were administered weekly as intraperitoneal injections of 50 μg or 5 μg, respectively, for 3 weeks.

Cell isolation, transfer and FACS staining. For *in vitro* and *in vivo* transfer experiments, CD4⁺ T cells were pre-enriched using mouse CD4 Dynabeads (L3T4, Invitrogen) and FACS sorted on an LSR-II (BD Biosciences). Intracellular staining for IL-4 used Cytofix/Cytoperm following treatment with Golgi-Stop (BD Biosciences), and staining for other cytokines (following treatment with Golgi-Plug, BD Biosciences) and Foxp3 and Gata3 used the Foxp3 staining kit (eBiosciences).

In vitro assays. *In vitro* induction assays were performed with 5×10^4 Foxp3–GFP⁻ CD4⁺ T cells and $5\,\mu g\,ml^{-1}$ of anti-CD3 and anti-CD28 antibody, $100\,U\,ml^{-1}$ IL-2, in 96-well, flat-bottom plates. For *in vitro* suppression assays, 4×10^4 CD4⁺Foxp3⁻CD62L^high naive T cells FACS purified from WT mice were cultured with graded numbers of CD4⁺Foxp3⁺ T_{reg} cells FACS purified from $Foxp3^{\Delta CNS1}$ or $Foxp3^{gfp}$ mice in the presence of 10^5 irradiated T cell-depleted splenocytes and $1\,\mu g\,ml^{-1}$ anti-CD3 antibody in a 96-well round-bottom plate for 80 h. Cell proliferation was assessed by [³H]thymidine incorporation during the final 8 h of culture.

Histology and immunohistochemistry. Necropsies were performed, and sections of pancreas, stomach, heart, lungs, kidney, external ear and haired skin were fixed in 10% phosphate-buffered formalin. Tissues were processed routinely for staining with haematoxylin and eosin, periodic acid Schiff with Alcian blue or Masson Trichrome if indicated. Slides were examined by an American Board of Veterinary Practitioners-certified veterinary pathologist blinded to genotypes. Morphological diagnoses were applied for all tissues. Immunohistochemical staining was performed by the University of Washington Histology and Imaging Core using standard protocols with a Leica Bond Automated Immunostainer. Primary antibodies: goat anti-mouse chitinase 3-like 3/ECF-L (YM1) (R&D systems, cat. no. AF2446, lot no. UNU01), $0.2\,\mu g\,ml^{-1}$; rabbit polyclonal anti iNOS/NOS II, NT (Millipore, cat. no. 06-573), $1\,\mu g\,ml^{-1}$; rabbit polyclonal anti arginase 1 (H-52) (Santa Cruz, cat. no. sc-20150, lot no. K0807), $0.2\,\mu g\,ml^{-1}$. Isotype controls were used at the same concentration as the primary antibody with all antibodies run with Lecia Bond reagents and Bond Polymer Refine (DAB) detection with haematoxylin counter stain.

Histology inflammation scoring. 0, None; 1, focal or multifocal mild perivascular accumulations with minimal extension into surrounding adventia or parenchyma; 2, multifocal mild or focal moderate perivascular accumulations with mild extension into surrounding parenchyma or mild to moderate parenchymal accumulations; 3, grade 2 plus mild inflammation-associated parenchymal lesions such as loss or degeneration of cells; 4, grade 2 plus moderate to severe inflammation-associated parenchymal lesions. Inflammation in the gastrointestinal tract was scored as described previously[32].

Airway hyperresponsiveness measurements. For measurement of AHR, mice were anaesthetized with pentobarbitol (7.5–10 mg per mouse) and AHR was assessed by invasive measurement of airway resistance using modified version of a described method (Buxco Electronics). Mice were ventilated at a tidal volume of 0.2 ml with the use of a ventilator (Harvard Apparatus) and frequency was set around 150 Hz. Baseline pulmonary mechanics and responses to ventilated saline (0.9% NaCl) were measured, and lung resistance (R_L) was measured in response to increasing doses ($0.125–40\,mg\,ml^{-1}$) of acetyl-β-methylcholine chloride (methacholine; MCh) (Sigma-Aldrich). The three values of R_L obtained after each dose of methacholine were averaged to obtain the final values for each dose. Results are expressed as percentage of increase of saline-baseline. Following measurement of AHR, mouse tracheas were cannulated and the lungs were lavaged twice with 1 ml of PBS 2% FCS and the fluids were pooled. Cells in the lavage fluid were counted using a haemocytometer, and BAL cell differential counts were determined on slide preparations stained with DiffQuik. At least 200 cells were differentiated on stained slides by light microscopy using conventional morphological criteria. For some experiments, BAL for each mouse or grouped BAL was stained and analysed by flow cytometry.

Stool sample collection. Fresh stool samples were induced directly into sterile collection tubes from live CNS1⁻ and control mice and snap frozen before preparation of material for sequencing (see below).

DNA extraction. DNA extraction was performed on each fecal specimen using phenol-chloroform extraction with mechanical disruption based on a previously described protocol[33] Briefly, an aliquot (~500 mg) of each sample was suspended in a solution containing 500 μl of extraction buffer (200 mM Tris, pH 8.0; 200 mM NaCl; and 20 mM EDTA), 210 μl of 20% SDS, 500 μl of phenol/chloroform/isoamyl alcohol (25:24:1), and 500 μl of 0.1-mm-diameter zirconia/silica beads (BioSpec Products). Microbial cells were lysed by mechanical disruption with a bead beater (BioSpec Products) for 2 min, after which two rounds of phenol/chloroform/isoamyl alcohol extraction were performed. DNA was precipitated with ethanol and resuspended in 50 μl of nuclease-free water. DNA was subjected to additional purification with the QIAamp DNA Mini Kit (Qiagen).

PCR amplification and sequencing. For each sample, three replicate 25 μl PCR amplifications were performed, each containing 5 ng of purified DNA, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 1.25 U Platinum Taq DNA polymerase, 2.5 μl of 10× PCR buffer, and 0.2 μM each of broad-range bacterial forward and reverse primers as described previously[34], flanking the V1–V3 variable region. The primers were modified to include adaptor sequences required for 454 sequencing, with the addition of a unique 6–8 base barcode in the reverse primer. The forward primer (5′-<u>CCTATCCCCTGTGTGCCTTGGCAGTCTC</u>*AGAGTTTGATCCTGGCTCAG*-3′) consisted of the 454 Lib-L primer B (underlined) and the broad-range universal bacterial primer 8F (italics); the reverse primer (5′-<u>CCATCTCATCCCTGCGTGTCTCCGACTCAG</u>NNNNNNN*ATTACCGCGGCTGCTGG*-3′) consisted of the 454 Lib-L primer A, barcode (NNNNNNN), and the broad-range primer 534R (italics). The cycling conditions were: 94 °C for 3 min, then 25 cycles of 94 °C for 30 s, 56 °C for 30 s, and 72 °C for 1 min. The three replicate PCR products were pooled and subsequently purified using the Qiaquick PCR Purification Kit (Qiagen). The purified PCR products were sequenced unidirectionally on a 454 GS FLX Titanium pyrosequencing platform following the Roche 454 recommended procedures.

Sequence processing and analysis. Sequences were converted to standard FASTA format using Vendor 454 software. Sequences shorter than 200 base pairs (bp), containing undetermined bases or homopolymer stretches longer than 8 bp, or failing to align with the V1–V3 region were excluded from the analysis. Using the 454 base quality scores, which range from 0 to 40 (0 being an ambiguous base), sequences were trimmed using a sliding-window technique, such that the minimum average quality score over a window of 50 bases never dropped below 35. Sequences were trimmed from the 3′-end until this criterion was met. Sequences were aligned to the V1–V3 region of the 16S gene, using as template the SILVA reference alignment[35] and the Needleman-Wunsch algorithm with default scoring options. Potentially chimaeric sequences were removed using the chimaera uchime program[36]. Sequences were grouped into operational taxonomic units (OTUs) using the average neighbour algorithm. Sequences with distance-based similarity of 97% or greater were assigned to the same OTU. For each fecal sample, OTU-based microbial diversity was estimated by calculating the Shannon diversity index[37]. Phylogenetic classification to genus level was performed for each sequence, using the Bayesian classifier algorithm described by Wang and colleagues, using a database of known 16S sequences generated by the Ribosomal Database Project (RDP)[38]. For each experiment, data were analysed on each taxon level individually. The count data was rescaled using DESeq R package[39]. Bacteria with less than 10 mean count in both conditions were removed from further analysis and bacteria with statistically significant differences between two conditions (for example, WT and KO), were determined using binomial test (from DESeq package). Bacteria with fold-change greater than two and FDR = 0.05 were declared significant.

31. Stromnes, I. M. & Goverman, J. M. Active induction of experimental allergic encephalomyelitis. *Nature Protocols* **1,** 1810–1819 (2006).
32. Burich, A. *et al.* Helicobacter-induced inflammatory bowel disease in IL-10- and T cell-deficient mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* **281,** G764–G778 (2001).
33. Ubeda, C. *et al.* Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *J. Clin. Invest.* **120,** 4332–4341 (2010).
34. Nossa, C. W. *et al.* Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World J. Gastroenterol.* **16,** 4135–4144 (2010).
35. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).
36. Edgar, R. C. *et al.* UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).
37. Magurran, A. E. *Measuring Biological Diversity* (Blackwell, 2004).
38. Wang, Q. *et al.* Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–5267 (2007).
39. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).

LETTER

Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search

Anthony L. Forget[1,2] & Stephen C. Kowalczykowski[1,2]

DNA breaks can be repaired with high fidelity by homologous recombination. A ubiquitous protein that is essential for this DNA template-directed repair is RecA[1]. After resection of broken DNA to produce single-stranded DNA (ssDNA), RecA assembles on this ssDNA into a filament with the unique capacity to search and find DNA sequences in double-stranded DNA (dsDNA) that are homologous to the ssDNA. This homology search is vital to recombinational DNA repair, and results in homologous pairing and exchange of DNA strands. Homologous pairing involves DNA sequence-specific target location by the RecA–ssDNA complex. Despite decades of study, the mechanism of this enigmatic search process remains unknown. RecA is a DNA-dependent ATPase, but ATP hydrolysis is not required for DNA pairing and strand exchange[2,3], eliminating active search processes. Using dual optical trapping to manipulate DNA, and single-molecule fluorescence microscopy to image DNA pairing, we demonstrate that both the three-dimensional conformational state of the dsDNA target and the length of the homologous RecA–ssDNA filament have important roles in the homology search. We discovered that as the end-to-end distance of the target dsDNA molecule is increased, constraining the available three-dimensional (3D) conformations of the molecule, the rate of homologous pairing decreases. Conversely, when the length of the ssDNA in the nucleoprotein filament is increased, homology is found faster. We propose a model for the DNA homology search process termed 'intersegmental contact sampling', in which the intrinsic multivalent nature of the RecA nucleoprotein filament is used to search DNA sequence space within 3D domains of DNA, exploiting multiple weak contacts to rapidly search for homology. Our findings highlight the importance of the 3D conformational dynamics of DNA, reveal a previously unknown facet of the homology search, and provide insight into the mechanism of DNA target location by this member of a universal family of proteins.

The mechanism by which the RecA family of DNA strand exchange proteins (which include T4 UvsX, archaeal RadA and eukaryotic Rad51) locate DNA sequence identity is unknown. Ensemble studies have constrained possible mechanisms by establishing that ATP hydrolysis is not needed[3,4] and 1D sliding is not operative[5]. Consequently, the manner by which the RecA nucleoprotein filament promotes the efficient, rapid and accurate search for homology has remained undefined for decades[6]. Single-molecule methods have the potential to provide new insight into this long-standing question. In fact, magnetic tweezer experiments showed that the endpoint of homologous pairing can be detected as a change in the length of a single dsDNA target molecule[7,8]. However, the mechanism by which homology was found and DNA pairing occurred was not shown. Therefore, we sought to directly observe the manner by which RecA nucleoprotein filaments locate their homologous target in dsDNA.

Initially we attempted to directly observe fluorescent RecA nucleoprotein filaments interacting with bacteriophage λ dsDNA in real time by using total internal reflected fluorescence microscopy (TIRFM)[9]. Fully homologous fluorescent ssDNA that was complementary to three different loci of λ DNA (Fig. 1A) was generated by incorporation of 5-(3-aminoallyl) dUTP into ssDNA using polymerase chain reaction (PCR), followed with covalent attachment of ATTO565 (Supplementary Methods). RecA nucleoprotein filaments were assembled on these fluorescent ssDNA substrates in ensemble reactions containing ssDNA-binding protein (SSB) and the non-hydrolysable ATP analogue, ATPγS (5′-O-3′-thiotriphosphate)[4]. ATPγS was used to maintain the filament in its active form, eliminate filament disassembly and prevent dissociation of DNA pairing products[7,10–12]. Using

A, 0 bp   λ dsDNA molecule   48,502 bp

ssDNA pairing partners:
162   1,762   430 → ssDNA length (nt)
12,449   14,648   24,003 → Center of position on λ DNA (bp)
0.26   0.30   0.49 → Relative distance on λ DNA

B, Syringe pump

DNA pairing product with 430-nt ssDNA

a Ensemble reaction: DNA pairing product extended on surface

b Doubly attached DNA: DNA pairing in absence or presence of flow

Flow

c Singly attached DNA: DNA pairing in absence of flow; DNA extended by flow after pairing

Flow

d Doubly attached DNA with proximal ends: DNA pairing in absence of flow; DNA extended by flow after pairing
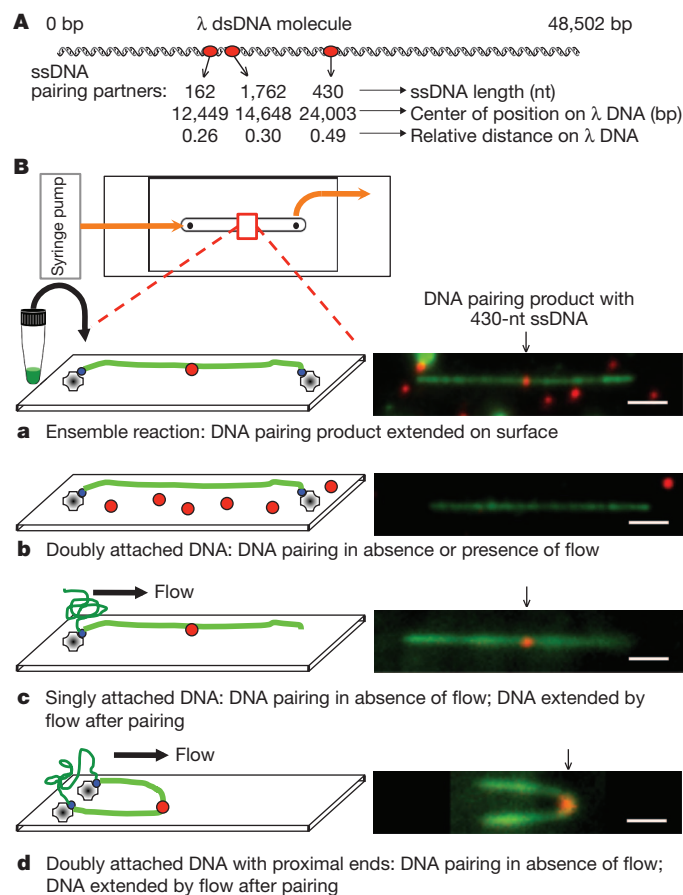
Figure 1 | DNA pairing by RecA, imaged using single-molecule TIRFM, indicates that the three-dimensional conformation of target dsDNA is important in the homology search. **A**, DNA substrates. **B**, DNA pairing between λ DNA (green) and RecA filament assembled on 430-nucleotide (nt) ssDNA (red). The ensemble reaction was examined by TIRFM (**B**, **a**). In *in situ* reactions dsDNA was attached before pairing; doubly attached extended DNA (**B**, **b**), singly attached DNA (**B**, **c**) and doubly attached DNA with ends in proximity (**B**, **d**). Homologously paired products were observed in **B**, **c** and **B**, **d** when DNA was relaxed by stopping flow and then extended by flow for visualization. Scale bars, 2.4 μm.

[1]Department of Microbiology, University of California, Davis, California 95616-8665, USA. [2]Department of Molecular and Cellular Biology, University of California, Davis, California 95616-8665, USA.

biochemical assays, we confirmed that the fluorescent ssDNA that was generated by this procedure was functional for RecA-mediated DNA pairing (Supplementary Fig. 1). The λ dsDNA, biotinylated at each end, was attached under flow to the interior surface of a single-channel microfluidic device (flowcell) (Fig. 1B). Owing to sequential attachment of each end to the streptavidin-coated surface, most DNA molecules were extended to nearly (~80%) B-form length, and extension could be maintained in the absence of flow (Fig. 1B, a, b).

To confirm DNA pairing at the homologous λ DNA target site, reactions were conducted under ensemble conditions, and products were extended on the surface of a flowcell for analysis by single-molecule, two-colour TIRFM; dsDNA was imaged by YOYO1 binding (green) and ssDNA by ATTO565 (red). DNA pairing products were observed; the sites of interaction coincided with the region of homology within the λ DNA molecule (Fig. 1B, a). For the 430-nucleotide ssDNA, all bound fluorescent ssDNA RecA filaments were at the homologous locus (observed fractional distance 0.51 ± 0.02; $n = 21$; Supplementary Fig. 2).

Next, we attempted to detect homologous pairing in real time using single-molecule TIRFM. Preformed RecA nucleoprotein filaments were introduced into a flowcell to which λ DNA molecules were tethered, buffer flow was stopped, and the reaction was monitored in real time (Fig. 1B, b). Although the dsDNA was readily visible, we failed to observe any interaction between the fluorescent nucleoprotein filaments and extended λ DNA, even for reaction periods longer than 1 h. However, we noticed that in addition to the desired doubly tethered extended λ DNA molecules, some DNA molecules were attached only by one end (Fig. 1B, c). When flow was stopped to score pairing with the doubly tethered λ DNA molecules, these singly tethered molecules relaxed to a randomly coiled state. Unexpectedly, when these unconstrained DNA molecules were subsequently re-extended by buffer flow, 80% ($n = 20$) revealed a stable pairing product (Fig. 1B, c). This finding suggested that either a free DNA end or a random coiled DNA was needed for pairing. In the same field of view, there were also λ DNA molecules that had both ends attached, but at a relatively close end-to-end distance (Fig. 1B, d). When the flow was stopped, we observed that these molecules also participated in homologous pairing during the time that flow was off, demonstrating that a free DNA end was not required. These unanticipated results revealed that DNA pairing did not occur on DNA that was extended to near its entropic elastic limit, and suggested that the DNA homology search required the 3D states that are accessible in randomly coiled DNA. Collectively, they suggested that a coiled conformation of the target dsDNA is crucial.

To address this possibility, we developed an alternative single-molecule imaging strategy that permitted reproducible measurement of the effects of dsDNA conformational structure, unperturbed by flow, on the DNA homology search process. This method uses a specialized flowcell (Fig. 2A), two optical laser traps operated in position-clamp mode, epifluorescent detection, fluorescent RecA–ssDNA filaments and a λ DNA dumbbell (a single λ DNA molecule with a 1-μm polystyrene bead attached at each end[13] (Supplementary Methods)). The DNA pairing assay was performed in situ using the dsDNA dumbbell target, and the dual optical trap configuration was used to reliably vary the end-to-end distance of the dsDNA. The flowcell has four channels and a flow-free reservoir. Movement of DNA dumbbells between channels of the flowcell was accomplished through stage translation, and manipulation of optical traps relative to one another was accomplished using a steering mirror controlling one of the traps. Each experiment (Fig. 2B and Supplementary Movie 1) consisted of six steps: first, in channel one, a streptavidin-coated bead was trapped in each of the two optical traps (Fig. 2B, a); second, the beads were moved to channel two to capture a λ dsDNA molecule (biotinylated on both ends and stained with YOYO1) on one bead (Fig. 2B, b); third, the beads were moved into channel three, and by independent steering of a trap, the distal end of the DNA was attached
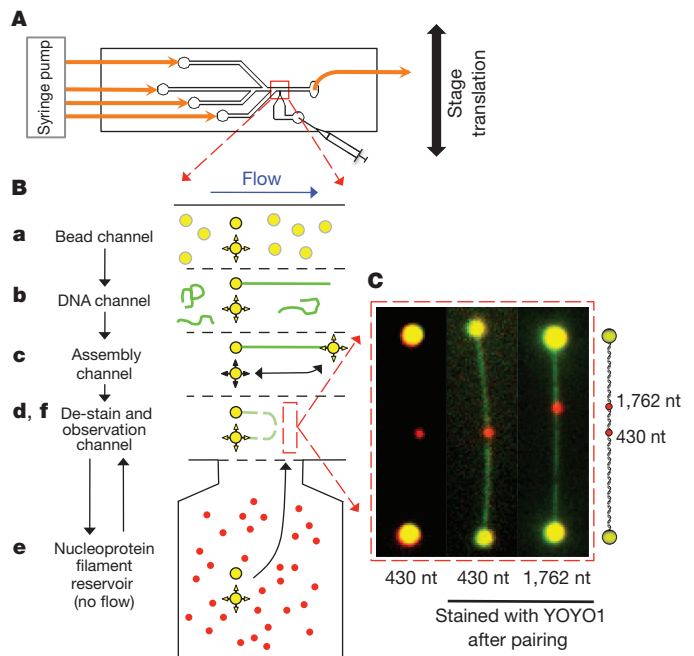

Figure 2 | Visualization of RecA-promoted DNA pairing with an individual optically trapped DNA dumbbell, imaged by epifluorescence. A, Four-channel flowcell with a flow-free reservoir. **B,** DNA dumbbell assembly and RecA-pairing reaction: two beads (yellow) are trapped (**B, a**); a λ DNA molecule (green) is captured on one bead (**B, b**); the free DNA end is captured with the second bead using a steerable optical trap (**B, c**); the centre-to-centre bead distance is set and YOYO1 is removed (**B, d**, de-stain); the DNA dumbbell is incubated in reservoir with RecA nucleoprotein filaments (red) (**B, e**) and DNA is extended to visualize products (**B, f**). **C,** Images of pairing products with 430- and 1,762-nucleotide nucleoprotein filaments.

to the second bead (Fig. 2B, c); fourth, the DNA-dumbbell was moved to the dye-free channel for de-staining, and the end-to-end distance was fixed (Fig. 2B, d); fifth, the DNA-dumbbell was moved to the flow-free reservoir containing the fluorescent ssDNA-RecA filaments (Fig. 2B, e); and sixth, after a defined incubation time, the DNA dumbbell was moved back to channel four, which is free of nucleoprotein filaments, extended to its contour length (~16 μm) and examined for DNA pairing products (Fig. 2B, f).

Shown in Fig. 2C are representative products of reactions in which the DNA dumbbells were initially held at a centre-to-centre bead distance of 2 μm and incubated for 2 min in the reservoir that contained RecA nucleoprotein filaments. For the two homologous ssDNA nucleoprotein filaments shown (430 nucleotides and 1,762 nucleotides), the pairing is clearly at the homologous locus. For a 2 min incubation with dsDNA at a bead-to-bead distance of 2 μm and the 430-nucleotide substrate, 90% of the dsDNA molecules ($n = 29$) contained a nucleoprotein filament stably bound to the expected region of homology (Fig. 3a). To determine the effect of end-to-end distance (that is, 3D conformation) on the RecA-mediated DNA pairing reaction, the reactions were performed at increasing bead separations (Fig. 3a). As the bead distance was increased from 2 μm to 8 μm, the efficiency of DNA pairing decreased to near zero, extrapolating to zero at ~9 μm; for comparison, in the TIRFM experiments in which no DNA pairing was detected in situ, the DNA end-to-end distance was ~13 μm.

We compared the time course of homologous pairing for fixed centre-to-centre bead distances of 2 μm and 6 μm (Fig. 3b) to determine the effect of decreasing DNA conformational states on the rate of the reaction. For the 2 μm separation, the rate of DNA pairing increased with a half-time of ~30 s and approached a yield of 100%. When the separation was increased to 6 μm, the rate slowed fourfold to a half-time of ~125 s, but nevertheless approached a yield of 100%
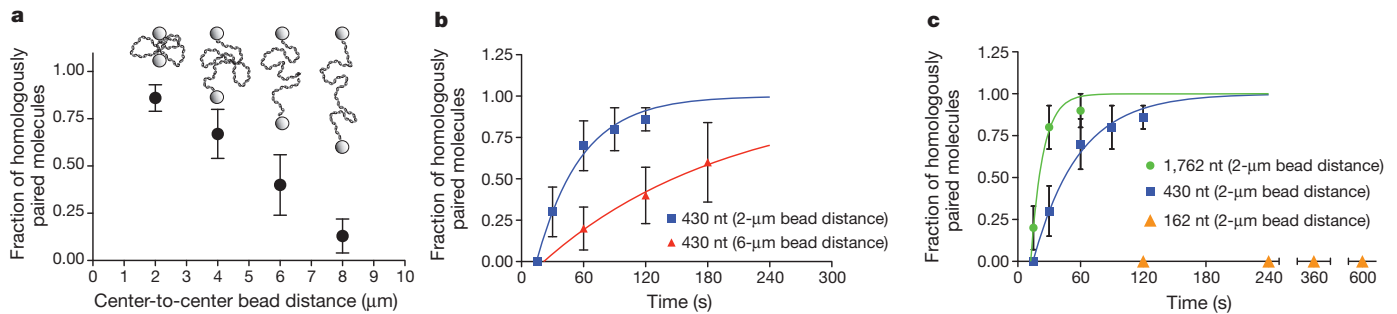
Figure 3 | DNA three-dimensional conformation and nucleoprotein filament length contribute to the homology search. a, Effect of DNA end-to-end distance; 430-nucleotide substrate (2 min). Error bars, s.e.m. from multiple experiments ($n = 10$ to 29). **b**, Time course; 430-nucleotide substrate: 2-μm (squares) and 6-μm (triangles) bead separation; respective pairing rates, 0.023

(± 0.002) s^{-1} ($n \geq 10$) and 0.0056 (± 0.0006) s^{-1} ($n \geq 5$). **c**, Effect of ssDNA length; 162 nucleotides (triangles; $n = 5, 6, 4$ and 2 at the times indicated), 430 nucleotides (squares; same data as Fig. 3b; $n \geq 10$) and 1,762 nucleotides (circles; $n \geq 10$); error bars, s.e.m.; 2-μm separation; respective rates: zero, 0.023 (± 0.002) s^{-1}, and 0.086 (± 0.026) s^{-1}.

(Fig. 3b). To establish the kinetic reaction order, we conducted single-molecule DNA pairing assays as a function of RecA nucleoprotein filament concentration (Supplementary Fig. 3). The reaction rate was independent of nucleoprotein filament concentration, showing that DNA pairing under these conditions is not diffusion limited, but that it is limited instead by a rate-determining unimolecular step as in the ensemble studies[14]. However, the pairing rate was dependent on dsDNA conformation and therefore was not dependent on the sequence recognition step itself.

To understand the nature of the complex that limits the rate of DNA pairing, we varied the length of RecA nucleoprotein filaments. Shown in Fig. 3c is a comparison of the time courses for 162-, 430- and 1,762-nucleotide nucleoprotein filaments. Increasing the ssDNA length approximately fourfold, from 430 to 1,762 nucleotides, increased the observed rate of pairing approximately 3.8-fold. However, when the length of the ssDNA was decreased to 162 nucleotides, we did not observe any stably bound homologously paired products after incubations for 10 min at the closest bead-to-bead distance possible (2 μm), despite this substrate being active in ensemble DNA pairing reactions (Supplementary Fig. 2). We conclude that the length of the RecA nucleoprotein filament is a crucial factor in the rate-limiting step of homologous pairing.

In addition to the anticipated stable, homologously paired end products, short-lived non-homologous interactions were observed (Fig. 4a). These events, which occurred outside of the homologous regions, were relatively unstable and dissociated during the movement of the molecule from the reservoir to the observation channel, during the separation of beads or after the λ DNA molecule was extended (Supplementary Movie 2). These heterologous events lasted no more than a few tens of seconds and never persisted on a timescale of minutes. When the molecules from the 2-μm data set were analysed, 22% of the reactions with the 430-nucleotide ssDNA and 40% of reactions with the 1,762-nucleotide ssDNA had these unstable heterologously paired intermediates (Fig. 4b), and for the 162-nucleotide ssDNA, only 1 heterologously bound filament was seen out of 28 molecules.

Some intermediates of the pairing process had a second filament bound non-specifically to spatially separated regions of the λ DNA molecule. For such a heterologously bound nucleoprotein filament, when the relaxed DNA molecule was moved into the observation channel and the beads were separated for observation, the existence of a loop could be inferred from a sudden recoil of the homologously paired spot. As the beads were separated, the weaker of the two heterologous interactions was released, and there was a simultaneous movement ('jump') of the fluorescence at the homologous pairing locus (Fig. 4a and Supplementary Movie 3) resulting from the release of DNA that was constrained in the loop. Approximately 12% ($n = 50$) of the DNA dumbbells showed loop release events for the 430-nucleotide nucleoprotein filament and, consistent with expectations,

when the length of the nucleoprotein filament was increased to 1,762 nucleotides, the number of molecules with transient loop structures increased to 47% ($n = 30$) (Fig. 4c).

Our results clearly establish that both the 3D conformation of dsDNA and the length of the nucleoprotein filament are important determinants of the rate for DNA homologous pairing. These findings lead us to propose a model termed 'intersegmental contact sampling' to describe the search for homology by a RecA nucleoprotein filament (Fig. 4d). One of the key features of the model is that the RecA nucleoprotein filament has a polyvalent interaction surface that is capable of binding simultaneously and non-specifically, but weakly, with non-contiguous segments of dsDNA. The second related feature of this
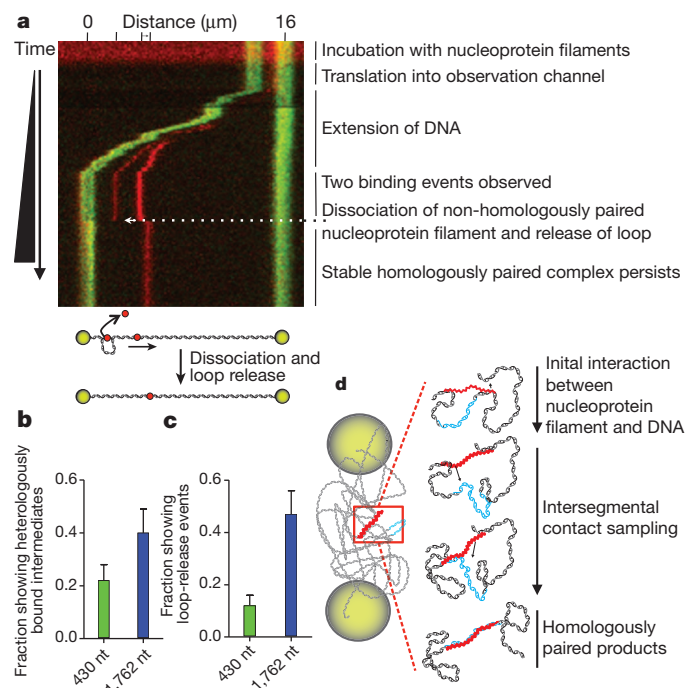

Figure 4 | RecA nucleoprotein filaments exhibit transient non-homologous interactions and loop-release events. a, Kymograph of DNA dumbbell during bead separation (Fig. 2B, f). Distance scale (top) and tick marks show positions of beads (green) and nucleoprotein filaments (red); illustration depicts dissociation of heterologously bound filament. **b, c**, Fraction of dsDNA dumbbells with non-homologously bound intermediates (**b**) and loop release events (**c**); 430-nucleotide (blue) and 1,762-nucleotide (green) filaments; $n = 50$ and 30, respectively. Error bars, s.e.m. **d**, Model for RecA homology search by intersegmental contact sampling; for simplicity, only two simultaneous points of interaction are depicted.

model is that 3D conformational entropy of the dsDNA greatly enhances the probability that DNA sequence homology will be found through iterated homology sampling, using multiple weak contacts, by this polyvalent filament. This model is compatible both with our key experimental findings, which we expect would apply to the search in the presence of ATP as well, and with the involvement of heterologously bound intermediates that have been inferred from biochemical studies[15,16]. Our data show that dsDNA extended to near contour length fails to produce homologously paired products. This observation provides an explanation for the observation that the formation of stable DNA pairing products in single-molecule studies using magnetic tweezers required negative plectonemic supercoils in the DNA target[7,8]. By contrast, when a ssDNA–RecA filament was extended to near its contour length, homologous pairing with fully homologous coiled dsDNA occurred[7], which is compatible with our finding that the coiled structure of dsDNA is essential to the homology search. Here we established that as the end-to-end distance of the dsDNA was decreased, allowing it to assume a more random coil-like 3D conformation, the rate of DNA pairing increased because the local DNA concentration increases, and the likelihood that DNA segments will be in close proximity also greatly increases. The increased local DNA concentration results in a greater statistical probability that a single nucleoprotein filament can simultaneously interact with and sample multiple regions of the same DNA molecule. This, in turn, is manifest as a kinetically more efficient homology sampling process. In further support of the intersegmental contact sampling model, when the length of the ssDNA in the nucleoprotein filament is increased, the observed rate of pairing, as well as the number of nucleoprotein filaments with multiple, transient, heterologous intersegmental interactions is increased. This shows that longer nucleoprotein filaments can simultaneously and independently sample more segments of the target dsDNA than shorter nucleoprotein filaments. Kinetically, our findings are consistent with the following two-step scheme:

$$\text{NPF} + \text{dsDNA} \underset{}{\overset{K_{\text{het}}}{\rightleftharpoons}} \underset{\text{(heterologously bound)}}{\text{NPF—dsDNA}} \overset{k_{\text{s}}}{\to} \underset{\text{(homologously paired)}}{\text{NPF—dsDNA}}$$

where K_{het} is the equilibrium constant for the binding of a RecA nucleoprotein filament (NPF) to heterologous dsDNA (the kinetic steps comprising K_{het} are rapid compared to k_{s}) and k_{s} is the rate-limiting unimolecular rate constant for intersegmental homology searching step within the dsDNA molecule or domain. In general, this kinetic formalism predicts a hyperbolic dependence of homologous pairing on the component concentrations unless the equilibrium constant for formation of the heterologous complex is large; when this is case, the observed rate is defined by the first-order rate constant, k_{s}. Given that the rate of target location is independent of nucleoprotein filament concentration, this implies that the heterologously bound complex is saturated at a filament concentration of 100 pM (Supplementary Fig. 3), placing a limit on the apparent equilibrium dissociation constant of <10 pM (that is, $K_{\text{het}} > 10^{11}$ M^{-1}). In the context of this kinetic model, values for k_{s} are defined by the experiments in Fig. 3b, c, which show that the rate of the intersegmental homology search decreases fourfold when the DNA end-to-end distance increases from 1 μm to 5 μm and increases approximately fourfold when the ssDNA length increases approximately fourfold. The correlation of rate with the length of ssDNA suggests that the intradomainal search is enhanced proportionately by the increase in either heterologous contacts or the reach of the longer ssDNA. In many regards, the homology search by RecA has parallels to target location by sequence-specific DNA-binding proteins, with the notable exception that the specificity of the RecA filament is determined by the sequence of the associated ssDNA. Seminal work on the DNA target selection by transcriptional regulatory proteins identified sliding, hopping and intersegmental transfer as potentially facilitating mechanisms[17,18]. Here we have established intersegmental transfer as the operative pathway used by RecA to find DNA sequence homology;

this behaviour is distinct from the sliding and hopping used to enhance the rate of target location by most regulatory proteins, which are typically univalent or bivalent with regard to site binding[18]. Our approach now provides a framework for future studies on the previously mysterious homology search by recombination proteins. It is applicable to studies of more complex systems such as eukaryotic Rad51, as it can provide insight into the function of the many accessory proteins that enhance DNA pairing[9]. Finally, the imaging strategy and flow-free cell design can easily be adapted to visualize target location and mechanism of processes as diverse as DNA replication and repair, RNA interference, transcription and protein translation, in which the 3D conformations of nucleic acids are undoubtedly important.

METHODS SUMMARY

RecA and SSB were purified as described[19,20]. Fluorescent ssDNA was prepared as detailed in the Supplementary Information. Nucleoprotein filaments were formed as described[4] in SM buffer (25 mM Tris acetate (Tris-OAc) (pH 7.5), 1 mM DTT and 4 mM Mg(OAc)$_2$), SSB (at a ratio of 1 SSB monomer to 11 nucleotides), 2 nM molecules fluorescent ssDNA, and 1 mM ATPγS were incubated for 10 min at 37 °C; RecA was added at 1 monomer per 1.7 nucleotides, and incubated 1 h. Nucleoprotein filaments were diluted to 0.2 nM before use.

For DNA pairing using TIRFM, biotinylated λ DNA (1 pM, molecules) in SM2 (SM with 50 mM NaCl) was bound to the flowcell and then washed to remove free DNA, and to attach the second DNA end. Reactions were started by addition of 0.2 nM nucleoprotein filaments. For ensemble experiments visualized by TIRFM, nucleoprotein filaments and λ DNA were incubated for 1 h (162-nucleotide substrate) or 30 min (430-nucleotide substrate) at 37 °C.

Visualization of RecA-mediated pairing with individual DNA dumbbells was performed at 37 °C. The flowcell was treated for 1 h with BSA (1 mg ml^{-1}) in SM3 (50 mM Tris-OAc (pH 8.2), 50 mM DTT, 1 mM Mg(OAc)$_2$ and 15% sucrose). Biotinylated λ DNA and buffers were pumped into the flowcell at a linear flow rate of $\sim$100 μm s^{-1}. Channels contained SM3, 18 fM streptavidin-coated polystyrene beads (1 μm, Bangs Laboratories) and 5 nM YOYO-1 (Invitrogen) (Fig. 2B, a); SM3, 100 nM YOYO1 and 10 pM (molecules) biotinylated λ DNA (Fig. 2B, b); SM3 (Fig. 2B, c); SM and 15% sucrose (Fig. 2B, d, f). The reaction reservoir contained 0.2 nM nucleoprotein filaments in SM with 15% sucrose and 0.5 mM ATPγS (Fig. 2B, e).

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

1. Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehrauer, W. M. Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.* **58**, 401–465 (1994).
2. Menetski, J. P. & Kowalczykowski, S. C. Interaction of recA protein with single-stranded DNA. Quantitative aspects of binding affinity modulation by nucleotide cofactors. *J. Mol. Biol.* **181**, 281–295 (1985).
3. Kowalczykowski, S. C. & Krupp, R. A. DNA-strand exchange promoted by RecA protein in the absence of ATP: implications for the mechanism of energy transduction in protein-promoted nucleic acid transactions. *Proc. Natl Acad. Sci. USA* **92**, 3478–3482 (1995).
4. Menetski, J. P., Bear, D. G. & Kowalczykowski, S. C. Stable DNA heteroduplex formation catalyzed by the *Escherichia coli* RecA protein in the absence of ATP hydrolysis. *Proc. Natl Acad. Sci. USA* **87**, 21–25 (1990).
5. Adzuma, K. No sliding during homology search by RecA protein. *J. Biol. Chem.* **273**, 31565–31573 (1998).
6. Kowalczykowski, S. C. Biochemistry of genetic recombination: energetics and mechanism of DNA strand exchange. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 539–575 (1991).
7. Fulconis, R., Miné, J., Bancaud, A., Dutreix, M. & Viovy, J. L. Mechanism of RecA-mediated homologous recombination revisited by single molecule nanomanipulation. *EMBO J.* **25**, 4293–4304 (2006).
8. van der Heijden, T. *et al.* Homologous recombination in real time: DNA strand exchange by RecA. *Mol. Cell* **30**, 530–538 (2008).
9. Forget, A. L. & Kowalczykowski, S. C. Single-molecule imaging brings Rad51 nucleoprotein filaments into focus. *Trends Cell Biol.* **20**, 269–276 (2010).
10. McEntee, K., Weinstock, G. M. & Lehman, I. R. Binding of the recA protein of *Escherichia coli* to single- and double-stranded DNA. *J. Biol. Chem.* **256**, 8835–8844 (1981).
11. Honigberg, S. M., Gonda, D. K., Flory, J. & Radding, C. M. The pairing activity of stable nucleoprotein filaments made from recA protein, single-stranded DNA, and adenosine 5′-(gamma-thio)triphosphate. *J. Biol. Chem.* **260**, 11845–11851 (1985).

12. Galletto, R., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. Direct observation of individual RecA filaments assembling on single DNA molecules. *Nature* **443,** 875–878 (2006).
13. van Mameren, J. *et al.* Counting RAD51 proteins disassembling from nucleoprotein filaments under tension. *Nature* **457,** 745–748 (2009).
14. Julin, D. A., Riddles, P. W. & Lehman, I. R. On the mechanism of pairing of single- and double-stranded DNA molecules by the recA and single-stranded DNA-binding proteins of *Escherichia coli. J. Biol. Chem.* **261,** 1025–1030 (1986).
15. Gonda, D. K. & Radding, C. M. By searching processively recA protein pairs DNA molecules that share a limited stretch of homology. *Cell* **34,** 647–654 (1983).
16. Tsang, S. S., Chow, S. A. & Radding, C. M. Networks of DNA and recA protein are intermediates in homologous pairing. *Biochemistry* **24,** 3226–3232 (1985).
17. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20,** 6929–6948 (1981).
18. Berg, O. G. in *The Biology of Nonspecific DNA Protein Interactions* (ed. Revzin, A.) 71–85 (CRC, 1990).
19. Mirshad, J. K. & Kowalczykowski, S. C. Biochemical characterization of a mutant RecA protein altered in DNA-binding loop 1. *Biochemistry* **42,** 5945–5954 (2003).
20. Harmon, F. G. & Kowalczykowski, S. C. RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. *Genes Dev.* **12,** 1134–1144 (1998).

METHODS

Microscope. The instrument that was developed was based on an Eclipse TE2000-U inverted microscope with a total internal reflected fluorescence (TIRF) attachment (Nikon) using a CFI Plan Apo TIRF 100×, 1.45 numerical aperture oil-immersed objective. Infrared laser trapping, operated in position-clamp mode, was achieved almost exactly as previously described[21] with the addition of a polarizer (Newport) to split the beam and generate two traps, and a steering mirror (Newport) to control the x–y position of one of the beams. Fluorescence of the sample in TIRF mode was achieved by excitation using a Cyan 488-nm laser (Picarro) or a 561-nm laser (Cobolt). Epifluorescence illumination was achieved with an X-Cite 120-W mercury vapour lamp (Lumen Dynamics). The fluorescence emission was directed through a polychroic mirror (centre wavelength 515 nm, bandwidth 30 nm; and centre wavelength 600, bandwidth 40 nm; Chroma). Light was guided into a Dual-View apparatus (Optical Insights) where the green and red components were spatially separated (dichroic 565dxcr, emission HQ515/30 nm and HQ600/40 nm, Chroma). Movies were captured on a DU-897E iXon CCD camera (Andor, 100-ms exposure) and processed using IQ imaging software (Andor).

Biotinylated λ duplex DNA. Multiple biotin moieties were incorporated into both ends of bacteriophage λ DNA (NEB) by an end-filling reaction. A 30-μl reaction contained 1× NEB buffer number 2, 33 μM each of dATP, dTTP, dCTP and biotin-11-dGTP (Perkin Elmer), 5 μg λ DNA and 5 units of Klenow exo$^-$ (NEB). The reaction was incubated for 15 min at 25 °C then terminated by the addition EDTA to a final concentration of 10 mM and heat inactivation of Klenow at 75 °C for 20 min. The reaction was then diluted to a 100-μl final volume with Nanopure water (Millipore) and passed through an S-400 spin column (GE Healthcare) equilibrated with TE buffer (10 mM Tris-HCl (pH 7.5) and 1 mM EDTA).

Fluorescent ssDNA substrates. DNA primer sequences that were used to amplify defined regions of λ DNA by PCR are the following, for: an 87-bp product for D-loop assay with pUC19 supercoiled DNA: forward primer 5′-biotin-CGACGGCCAGTGAATTCCCCGA-3′, reverse primer 5′-TTACGCCAAGCTT ACTCGGGAAACAT-3′; a 162-bp product (identical to λ DNA between base pairs 12,368–12,529): forward primer 5′-biotin-TAACGTCATGTCAGAGCAGAAAA AG-3′, reverse primer 5′-GCAATACCATCAAAGGTCTGCGTG-3′; a 430-bp product (identical to λ DNA between base pairs 23,788–24,217): forward primer 5′-biotin-ACTGTTCTTGCGGTTTGGAGG-3′, reverse primer 5′-CTATCGGA AGTTCACCAGCCAG-3′; and a 1,762-bp product (identical to λ DNA between base pairs 13,767–15,528): forward primer 5′-biotin-GGATGCGGTGAACTTCGT CAAC-3′, reverse primer 5′-CCCCTTACTGCTTCCTTTACCC-3′.

PCR reactions contained 1× ThermoPol buffer (NEB), 0.2 mM dATP, 0.2 mM dCTP, 0.2 mM dGTP, 0.1 mM dTTP, 0.2 mM 5-(3-aminoallyl) dUTP (Fermentas), 0.25 ng μl^{-1} λ DNA (NEB) (pUC19 for a 87-nucleotide substrate), 0.5 μM each primer and 0.05 U μl^{-1} Vent exo$^-$ polymerase (NEB). The thermocycler (iCycler, Bio-Rad) program involved initial denaturation at 95 °C for 2 min, 30 cycles of a denaturation phase at 95 °C for 30 s, an annealing phase at 60.6, 63, 62.2 or 59.4 °C for 30 s, for 87-, 162-, 430- or 1762-nucleotide products, respectively, and an extension phase at 72 °C for 0.25, 0.25, 1 and 5 min for 87-, 162-, 430- and 1762-nucleotide products, respectively. The final PCR step was extension at 72 °C for 5 min. The reactions were then processed with a QIAquick PCR purification kit (Qiagen). Following purification, the DNA was ethanol-precipitated at −20 °C. To fluorescently label the PCR products, a 20-μl reaction containing 10–20 μg of PCR-generated DNA containing amine-modified nucleotides, 200 mM sodium bicarbonate (pH 9.0) and 5 mM ATTO565 NHS-ester (ATTO-TEC GmbH) was incubated for 1–2 h at 25 °C while protected from light. Alexa Fluor 488 succinimidyl ester (Invitrogen) was used to label the 87-nucleotide substrate used in the D-loop assay. Following incubation, 180 μl Nanopure water was added and a QIAquick PCR purification kit (Qiagen) was used to remove free label. Purified labelled DNA was stored at 4 °C until the strand-separation step. Alkali denaturation in combination with the single 5′-biotin incorporated from the forward primer in the PCR reaction was used to produce ssDNA from the fluorescently labelled duplex PCR product as follows: 800 μl avidin-agarose (400 μl settled gel; Thermo Scientific) was prepared in a 1.5-ml Eppendorf tube using centrifugation to pellet agarose. All centrifugation steps were performed using a bench-top centrifuge at 4,524g for 1 min. The resin was pelleted and washed three times with 1 ml binding and wash buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA and 150 mM NaCl). Fluorescently labelled biotinylated dsDNA (~10–20 μg, from the PCR reaction above) was diluted to 1 ml with binding and wash buffer. The diluted DNA was added to the prepared avidin-agarose, and mixed end-over-end for 1 h while protected from light. The agarose and bound DNA were pelleted by centrifugation and washed three times with 1 ml binding and wash buffer to remove unbound DNA. The ssDNA was eluted by alkali denaturation of the dsDNA, by addition of 200 μl of 0.15 M NaOH to the pelleted agarose and mixing

end-over-end for 10 min to release the non-biotinylated strand. The slurry was transferred to an empty micro-spin column (Bio-Rad) and centrifuged at 4,700g to recover the eluted ssDNA. A Microspin S-400 column (GE Healthcare) was used to exchange the ssDNA into the TE buffer. Samples of each fraction were analysed by polyacrylamide or agarose gel electrophoresis. Fractions containing ssDNA were pooled, purified and concentrated with QIAquick PCR purification kit (Qiagen). The DNA concentration was determined using an extinction coefficient of 8,919 M^{-1} cm^{-1} at 260 nm, taking into account a correction factor of 0.34 for absorbance at 260 nm by the dye. The dye concentration was determined using an extinction coefficient of 120,000 M^{-1} cm^{-1} at 563 nm.

Flowcell fabrication. Channels and holes were etched by CO_2 laser into glass slides (Fisher Scientific 25 × 75 × 1 mm) covered with an adhesive abrasive blasting mask (Epilog) using a 30 W Mini-24 Laser Engraver (Epilog Lasers). Following the engraving step, the slides were blasted using 220 grit silicon carbide (Electro Abrasives) to remove residual laser-ablated glass from the channels. A cover glass (Corning No. 1, 24 × 60 mm) was attached with ultraviolet Optical adhesive number 74 (Norland Products) applied through capillary action. The adhesive was cured by placing the flowcell 30 cm from a 100 W HBO lamp (Zeiss) for 20 min followed by a final heat curing at 70 °C for 12 h. PEEK tubing with 0.5 mm inner diameter (Upchurch Scientific) was inserted into each of the etched holes to create inlet and outlet connection ports using 5 min Epoxy (Devcon).

Surface preparation of single-channel flowcell for TIRFM experiments. The surface modification procedure was done at 25 °C. The flowcells were cleaned with 1 M NaOH for 30–60 min, and washed twice with 1 ml Nanopure water and then with 1 ml of buffer (25 mM Tris-OAc (pH 7.5), 50 mM NaCl). 1 mg ml^{-1} biotinylated BSA (Thermo Scientific) in the above buffer was then incubated in the flowcell for 5 min and then washed with 1 ml of buffer. After this, 0.1 mg ml^{-1} streptavidin (Promega) in buffer was incubated in the flowcell for 5 min then washed with 1 ml of buffer. Finally, the flowcell was blocked with 1.5 mg ml^{-1} Roche Blocking Reagent (Roche) in buffer for 30–60 min and washed with 1 ml buffer. The prepared flowcell was then mounted on the microscope and attached to the syringe pump (KD Scientific).

D-loop assay. RecA and SSB were purified as previously described[19,20]. The AlexaFluor 488-labelled 87-nucleotide ssDNA substrate was prepared as described above. A 10-μl reaction containing 25 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 1 mM DTT, 2 mM ATPγS, 100 μg ml^{-1} BSA, 4.5 μM RecA and 105 nM fluorescently labelled 87-nucleotide ssDNA was incubated for 8 min at 37 °C. The reaction was started with the addition of 35 nM supercoiled DNA (pUC19) and incubated at 37 °C for 20 min. The reaction was stopped by mixing with 5 μl of stop solution (4.8% SDS, 7 mg ml^{-1} proteinase K) and incubating for 10 min at 37 °C. Products were resolved by electrophoresis in a 1% ultrapure agarose gel (Invitrogen) using TAE (40 mM Tris, 20 mM acetic acid and 1 mM EDTA) at 100 V until the bromophenol blue had migrated 4 cm. The gel was imaged and analysed with a STORM scanner and Image Quant software (Molecular Dynamics). The efficiency of the reaction was calculated as the fraction of ssDNA that formed D-loops multiplied by three to correct for the threefold molar excess of ssDNA relative to supercoiled pUC19 in the reaction.

Single-molecule DNA pairing experiments. Nucleoprotein filaments were formed essentially as described previously[4] in SM buffer (25 mM Tris-OAc (pH 7.5), 1 mM DTT and 4 mM Mg(OAc)₂); SSB (at a ratio of 1 SSB monomer to 11 nucleotides), 2 nM molecules fluorescent ssDNA and 1 mM ATPγS were incubated for 10 min at 37 °C. RecA was added at a ratio of 1 monomer to 1.7 nucleotides and incubated for 1 h. Nucleoprotein filaments were then diluted tenfold to a final concentration of 0.2 nM in buffer before introduction into the flowcell. In the DNA pairing experiments using TIRFM, biotinylated λ DNA (1 pM, molecules) in SM2 buffer (SM and 50 mM NaCl) was introduced into the flowcell and allowed to bind for several minutes. The flowcell was then washed with 500 μl SM2 buffer to remove free DNA as well as to extend and attach the second end of the λ DNA molecules. The reaction was started by the addition of 0.2 nM nucleoprotein filaments in SM2 buffer. For ensemble experiments visualized by TIRFM, the nucleoprotein filaments and λ DNA were incubated for 1 h (162-nucleotide substrate) or 30 min (430 nucleotide substrate) at 37 °C before visualization in a single-channel flowcell.

Visualization of RecA-mediated pairing with individual DNA dumbbells was performed at 37 °C. The flowcell surface was treated for 1 h with BSA (1 mg ml^{-1}) in single-molecule (SM3) buffer (50 mM Tris-OAc (pH 8.2), 50 mM DTT, 1 mM Mg(OAc)₂ and 15% sucrose). Biotinylated λ DNA and buffers were pumped at a linear flow rate of ~100 μm s^{-1} into the flowcell. The channels contained SM3 buffer, 18 fM streptavidin-coated polystyrene beads (1 μm; Bangs Laboratories) and 5 nM YOYO-1 (Invitrogen) (Fig. 2B, a); SM3 buffer, 100 nM YOYO1, and 10 pM (molecules) biotinylated λ DNA (Fig. 2B, b); SM3 buffer (Fig. 2B, c); SM buffer and 15% sucrose (Fig. 2B, d, f). The reaction reservoir contained 0.2 nM nucleoprotein filaments in SM with 15% sucrose and 0.5 mM ATPγS (Fig. 2B, e).

Data analysis. Data were analysed using GraphPad Prism v5.04. The kinetic data were fit to a single exponential function ($Y = Y_0 + (\text{Plateau} - Y_0)(1 - e^{-kx})$). In Fig. 4b, c, the time courses do not pass through the origin. We are not certain whether this is an intrinsic characteristic of the homology search or a limitation of the experimental procedure: for example, the time for the DNA to relax from flow-induced stretching after movement into the reservoir. We note that the half-time for the relaxation of extended λ DNA is ~ 6 s (ref. 22); during this time the dsDNA is not in its equilibrium coiled configuration and initial interaction with the RecA nucleoprotein filament would be limited by the DNA polymer dynamics.

21. Bianco, P. R. *et al.* Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature* **409,** 374–378 (2001).
22. Perkins, T. T., Quake, S. R., Smith, D. E. & Chu, S. Relaxation of a single DNA molecule observed by optical microscopy. *Science* **264,** 822–826 (1994).

LETTER

Brassinosteroid regulates stomatal development by GSK3–mediated inhibition of a MAPK pathway

Tae-Wuk Kim[1,2], Marta Michniewicz[3], Dominique C. Bergmann[3] & Zhi-Yong Wang[1]

Plants must coordinate the regulation of biochemistry and anatomy to optimize photosynthesis and water-use efficiency. The formation of stomata, epidermal pores that facilitate gas exchange, is highly coordinated with other aspects of photosynthetic development. The signalling pathways controlling stomata development are not fully understood[1,2], although mitogen-activated protein kinase (MAPK) signalling is known to have key roles. Here we demonstrate in *Arabidopsis* that brassinosteroid regulates stomatal development by activating the MAPK kinase kinase (MAPKKK) YDA (also known as YODA). Genetic analyses indicate that receptor kinase-mediated brassinosteroid signalling inhibits stomatal development through the glycogen synthase kinase 3 (GSK3)-like kinase BIN2, and BIN2 acts upstream of YDA but downstream of the ERECTA family of receptor kinases. Complementary *in vitro* and *in vivo* assays show that BIN2 phosphorylates YDA to inhibit YDA phosphorylation of its substrate MKK4, and that activities of downstream MAPKs are reduced in brassinosteroid-deficient mutants but increased by treatment with either brassinosteroid or GSK3-kinase inhibitor. Our results indicate that brassinosteroid inhibits stomatal development by alleviating GSK3-mediated inhibition of this MAPK module, providing two key links; that of a plant MAPKKK to its upstream regulators and of brassinosteroid to a specific developmental output.

In animals and plants, steroid hormones have important roles in coordinating development and metabolism[3]. In contrast to animal steroid hormones, which act through nuclear receptor transcription factors[3], the plant steroid hormone brassinosteroid binds to the extracellular domain of the membrane-bound receptor kinase brassinosteroid insensitive 1 (BRI1). This activates intracellular signal transduction mediated by the serine/threonine protein kinase BSK1, the protein phosphatase BSU1, the GSK3-like BIN2 kinase, PP2A phosphatase and BRASSINAZOLE RESISTANT 1 (BZR1) family transcription factors[4–10]. When brassinosteroid levels are low, BZR1 is inactivated owing to phosphorylation by BIN2 (refs 11, 12). Brassinosteroid signalling leads to inactivation of BIN2, and PP2A-mediated dephosphorylation and activation of BZR1 (refs 4, 9, 10) (Supplementary Fig. 1a). Although the brassinosteroid signalling pathway has been characterized, its connections to other signalling and developmental pathways are not fully understood.

Stomata are epidermal pores that control gas exchange between the plant and the atmosphere and are critical for maintaining photosynthetic and water-use efficiency in the plant. The density and distribution of stomata in the epidermis of aerial organs is modulated by intrinsic developmental programs, by hormones and by environmental factors such as light, humidity and carbon dioxide[1,2,13,14]. The genetically defined signalling pathway that regulates stomatal development includes peptide ligands, a receptor protein (TMM), the ERECTA family of receptor-like kinases (ER, ERL1 and ERL2) and a MAPK module comprised of the MAPK kinase kinase (MAPKKK) YDA, the MAPK kinases (MAPKKs) MKK4, MKK5, MKK7 and MKK9, and MAPKs MPK3 and MPK6 (ref 15). Potential downstream targets include basic helix–loop–helix (bHLH) transcription factors SPEECHLESS (SPCH), MUTE, FAMA, ICE1 (also known as SCRM) and SCRM2, with SPCH being negatively regulated by direct MPK3- and MPK6-mediated phosphorylation[16,17] (Supplementary Fig. 1b). It is possible that the MAPK pathway integrates environmental and hormonal inputs to optimize stomatal production, but nothing is known about the nature of these signals and their biochemical mechanisms of MAPK pathway regulation.

Excess stomata have been observed in some brassinosteroid-deficient mutants[18]. To elucidate the function of brassinosteroid in regulating stomatal development, we examined the distribution of stomata on leaves of brassinosteroid-deficient and brassinosteroid-signalling mutants. In wild-type *Arabidopsis*, stomata are always distributed with at least one pavement cell between them (Fig. 1a). Brassinosteroid deficiency causes stomatal clusters (Fig. 1b, c), whereas treatment with brassinolide (the most active form of brassinosteroid) reduces stomatal density (Fig. 1d), indicating that brassinosteroid represses stomatal development. The brassinosteroid-insensitive mutants *bri1-116*, quadruple *amiRNA-BSL2,3 bsu1 bsl1* (*bsu-q*)[9], dominant *bin2-1* and plants that overexpress *BIN2* also exhibit stomatal clustering (Fig. 1e–h), and overproduce stomatal precursors (meristemoids and guard mother cells) (Fig. 1u and Supplementary Fig. 2). In contrast to the weak stomatal clustering phenotype of the *det2-1* and *bri1-116* mutants, *bsu-q* showed large stomatal clusters on hypocotyls (Supplementary Fig. 4) and cotyledon surfaces consisting almost entirely of stomata (Fig. 1f, u, and Supplementary Figs 2 and 3). Surprisingly, the hyperactive *bzr1-1D* mutation[10,19] did not affect stomatal development or suppress the stomatal phenotypes of *bri1-116*, *bsu-q* and *bin2-1*, although it suppressed their dwarf phenotypes (Fig. 1i–n and Supplementary Fig. 5). These results indicate that brassinosteroid regulation of stomatal development is mediated by upstream signalling components that include BRI1, BSU1 and BIN2, but that it is independent of the BIN2 substrate BZR1.

Consistent with increased stomatal development in brassinosteroid-insensitive mutants, fewer stomata were observed in cotyledons of plants overexpressing some of the positive brassinosteroid-signalling components of the BSU1 family (Fig. 1q, u and Supplementary Fig. 6) and in *bin2-3 bil1 bil2* loss-of-function mutants lacking 3/7 brassinosteroid-signalling GSK3-like kinases (Fig. 1o, p, u and Supplementary Fig. 2). We used bikinin (4-[[(5-bromopyridin-2-yl)amino]-4-oxobutanoic acid, ChemBridge Corporation), a highly specific inhibitor for the 7 *Arabidopsis* GSK3-like kinases that appear to be involved in brassinosteroid signalling[9,20,21], to investigate further the function of brassinosteroid-related GSK3-like kinases in stomatal development. When added to the growth medium, bikinin decreased stomatal production in wild-type plants, fully suppressed the stomatal clustering phenotypes of *bin2-1* and partially suppressed the severe stomatal phenotypes of *bsu-q* (Fig. 1r–u). These results confirm that increased activity of the GSK3-like kinases is responsible for enhanced stomatal production in brassinosteroid-deficient and brassinosteroid-insensitive mutants.

[1]Department of Plant Biology, Carnegie Institution for Science, Stanford, California 94305-4150, USA. [2]Department of Life Science, Hanyang University, Seoul 133-791, South Korea. [3]Department of Biology, Stanford University, Stanford, California 94305-5020, USA.
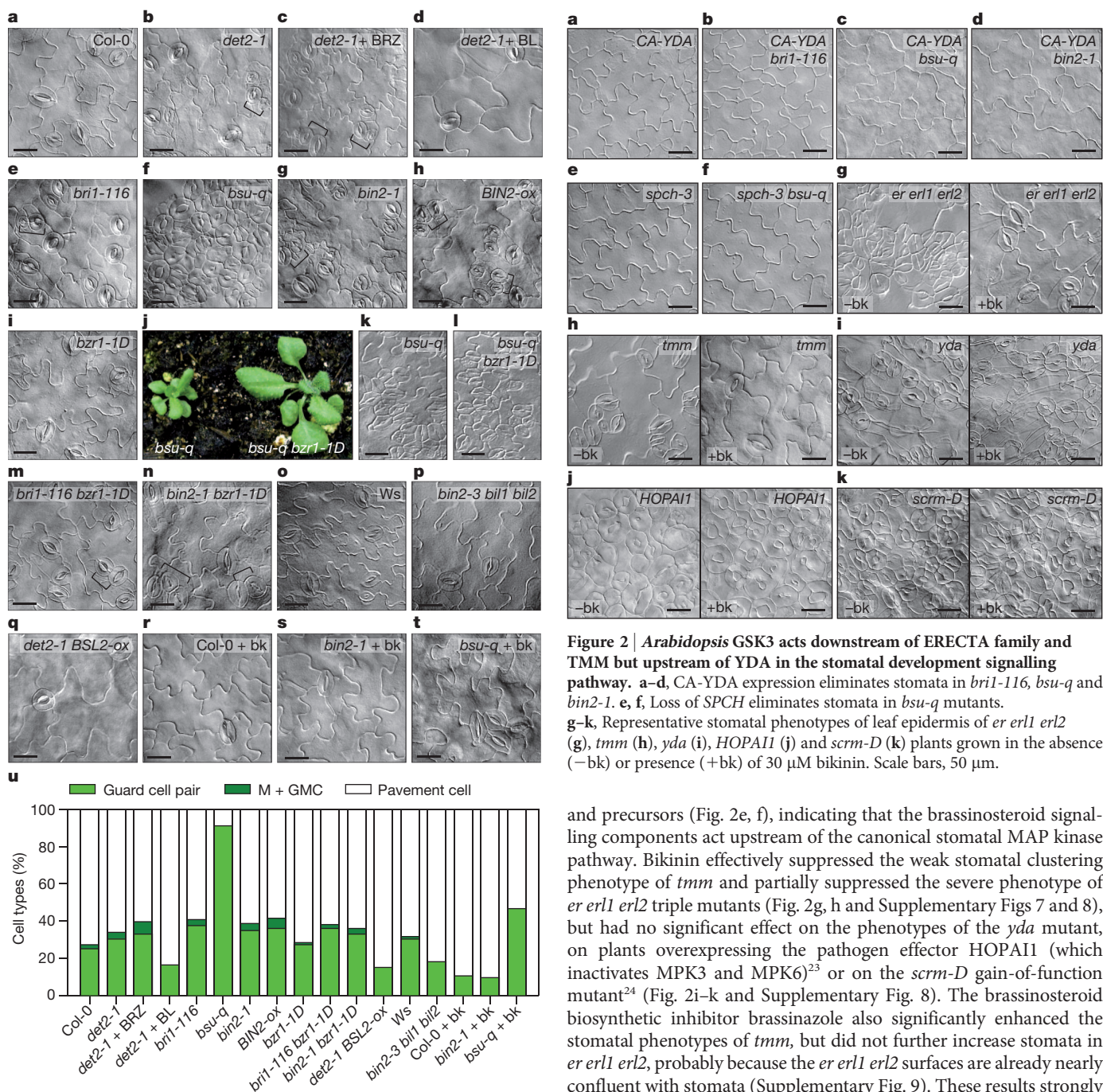
Figure 1 | Brassinosteroid negatively regulates stomatal development.
a–i, k–t, Differential interference contrast (DIC) microscopy images of abaxial cotyledon epidermis of 8-day-old seedlings or leaf epidermis of 4-week-old plants (**k, l**) with indicated genotypes (Col-0 and Ws are wild-type controls), grown on medium ± BRZ (2 μM), brassinolide (BL, 50 nM), or bikinin (bk, 30 μM). **j,** Growth phenotype of 4-week-old *bsu-q* and *bsu-q bzr1-1D* mutants. **u,** Quantification of epidermal cell types of the indicated 8-day-old mutants, expressed as percentage of total cells. GMC, guard mother cell; M, meristemoid. Brackets in **b, c, e, g, h, m, n** indicate clustered stomata. Scale bars, 50 μm.

Figure 2 | *Arabidopsis* GSK3 acts downstream of ERECTA family and TMM but upstream of YDA in the stomatal development signalling pathway. a–d, CA-YDA expression eliminates stomata in *bri1-116, bsu-q* and *bin2-1*. **e, f,** Loss of *SPCH* eliminates stomata in *bsu-q* mutants. **g–k,** Representative stomatal phenotypes of leaf epidermis of *er erl1 erl2* (**g**), *tmm* (**h**), *yda* (**i**), *HOPAI1* (**j**) and *scrm-D* (**k**) plants grown in the absence (−bk) or presence (+bk) of 30 μM bikinin. Scale bars, 50 μm.

We examined genetic interactions between brassinosteroid mutants and known stomatal mutants. Expression of constitutively active YDA (CA-YDA) can completely eliminate stomatal development[22] (Fig. 2a), probably through activation of a MAP kinase pathway that phosphorylates and inactivates SPCH[15,16]. Expression of CA-YDA completely suppressed stomatal development of the *bri1-116, bsu-q* and *bin2-1* mutants (Fig. 2b–d). Loss of SPCH was also completely epistatic to *bsu-q* in that a *bsu-q spch-3* (null) mutant lacked stomata
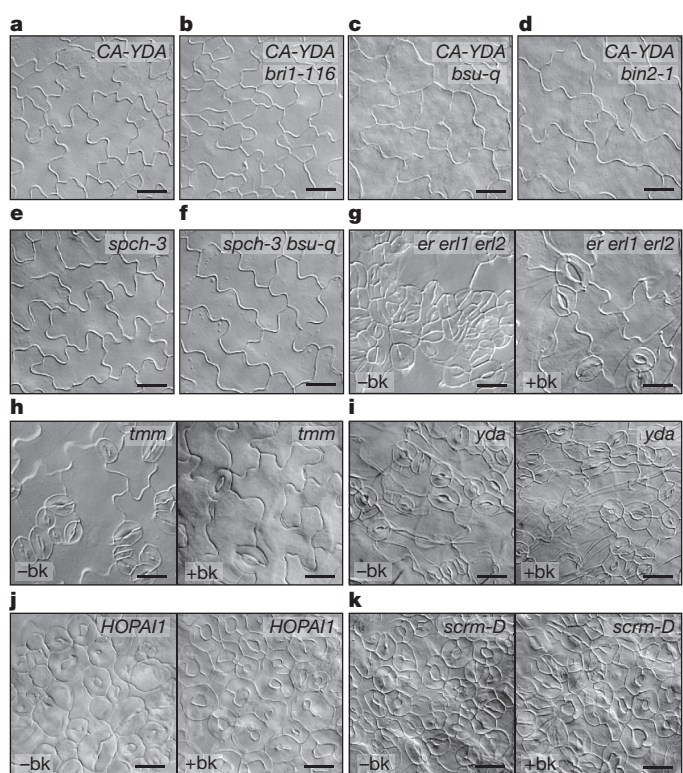
and precursors (Fig. 2e, f), indicating that the brassinosteroid signalling components act upstream of the canonical stomatal MAP kinase pathway. Bikinin effectively suppressed the weak stomatal clustering phenotype of *tmm* and partially suppressed the severe phenotype of *er erl1 erl2* triple mutants (Fig. 2g, h and Supplementary Figs 7 and 8), but had no significant effect on the phenotypes of the *yda* mutant, on plants overexpressing the pathogen effector HOPAI1 (which inactivates MPK3 and MPK6)[23] or on the *scrm-D* gain-of-function mutant[24] (Fig. 2i–k and Supplementary Fig. 8). The brassinosteroid biosynthetic inhibitor brassinazole also significantly enhanced the stomatal phenotypes of *tmm,* but did not further increase stomata in *er erl1 erl2,* probably because the *er erl1 erl2* surfaces are already nearly confluent with stomata (Supplementary Fig. 9). These results strongly indicate that GSK3-like kinases act downstream of the ER and TMM receptors, but upstream of the YDA MAPKKK.

YDA contains 84 putative GSK3 phosphorylation sites (Ser/Thr-X-X-X-Ser/Thr). Many of these sites are conserved in the two rice homologues of YDA, Os02g0666300 and Os04g0559800, and these homologues also share a highly conserved sequence just amino-terminal of the kinase domain. Importantly, YDA can be made constitutively active when part of this region (amino acids 185–322; Fig. 3a) is deleted[22]. The region that is deleted in CA-YDA contains 23 putative GSK3 phosphorylation sites, including successive phosphorylation sites that are similar to sites found in the known BIN2 target BZR1 (Fig. 3a and Supplementary Fig. 10).

We tested whether BIN2 directly interacts with and phosphorylates YDA. Maltose binding protein (MBP)-YDA was detected in an overlay assay by using GST–BIN2 and anti-GST antibody (Fig. 3b), demonstrating direct YDA binding to BIN2 *in vitro*. BIN2 also interacted with YDA and CA-YDA in yeast two-hybrid assays (Fig. 3c). *In vitro* kinase
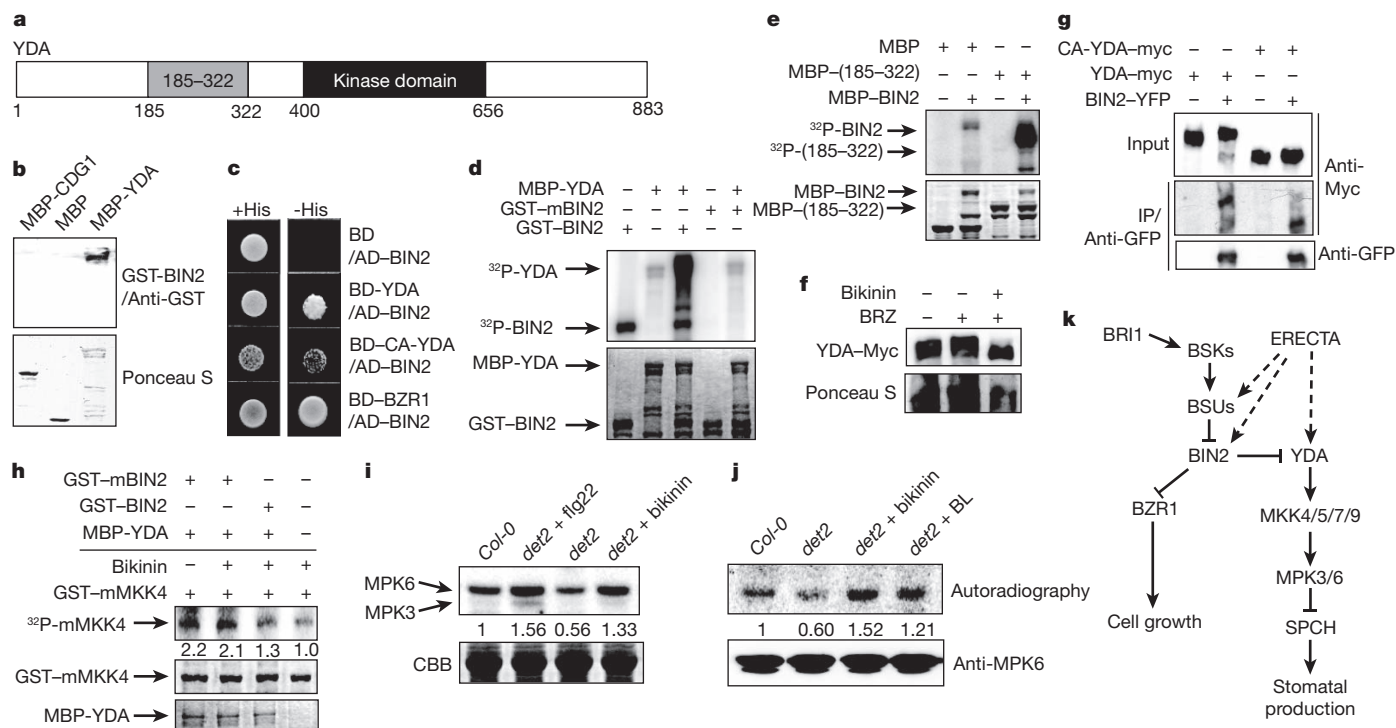
Figure 3 | BIN2 inhibits YDA kinase activity through phosphorylation.
a, Domain structure of YDA. b, Gel blot of indicated proteins (MBP–CDG1 is a negative control) sequentially probed with GST–BIN2 and anti-GST–HRP antibody. c, Yeast two-hybrid assays of indicated proteins. d, e, In vitro kinase assays of BIN2 phosphorylation of YDA or YDA fragment containing amino acids 185-322 (185-322). Upper panel shows autoradiography and bottom panel shows protein staining. Mutant BIN2 (mBIN2) is kinase inactive. f, YDA–Myc plants grown for 5 days on medium containing 2 μM BRZ ± 30 μM bikinin and analysed by anti-Myc immunoblot. g, Proteins transiently expressed in N. benthamiana leaves, immunoprecipitated (IP) with anti-YFP antibody, and immunoblotted with anti-Myc or anti-YFP antibody. h, YDA pre-incubated with BIN2 or mBIN2 (kinase-inactive mutant) and ATP was purified then incubated with mutant MKK4 (mMMK4) and [^{32}P]γATP,

± bikinin. Numbers indicate relative signal levels normalized to loading control. i–j, MPK6 and MPK3 activities in seedlings treated with flg22 (10 nM, positive control), bikinin (30 μM) or BL (100 nM) for 30 min (i) or 2 h (j), analysed by in-gel kinase assays. Numbers indicate relative signal levels (upper panel) normalized to the loading control (CBB or MPK6 immunoblot). k, A model for regulation of stomatal development by two receptor kinase-mediated signal transduction pathways. When brassinosteroid levels are low, BIN2 phosphorylates and inactivates YDA, increasing stomatal production. Brassinosteroid signalling through BRI1 inactivates BIN2, leading to activation of YDA and downstream MAPK proteins, and suppression of stomatal development. ERECTA is genetically upstream of YDA; a biochemical link is not known, but BSU1 and BIN2 or their homologues are strong candidates for intermediates (dashed line).

assays showed that BIN2 phosphorylated YDA, but YDA did not phosphorylate a kinase-inactive BIN2 mutant or other brassinosteroid signalling components (Fig. 3d and Supplementary Fig. 11). BIN2 strongly phosphorylated the region deleted in CA-YDA (Fig. 3e), indicating that BIN2 might inhibit YDA by phosphorylating its autoregulatory domain.

BIN2 phosphorylation of BZR1 causes mobility shifts of the phosphorylated BZR1 band in SDS–polyacrylamide gel electrophoresis (SDS–PAGE) gels[12]. Like BZR1, YDA that was phosphorylated by BIN2 in vitro also exhibited slower mobility (Fig. 3d and Supplementary Fig. 11). Consistent with the in vitro data, bikinin treatment of Arabidopsis seedlings increased the mobility of YDA–Myc in SDS–PAGE (Fig. 3f). When transiently expressed in Nicotiana benthamiana leaf cells, both YDA–Myc and CA-YDA–Myc were co-immunoprecipitated by anti-yellow fluorescent protein (YFP) antibody when co-expressed with BIN2–YFP but not when expressed alone (Fig. 3g), demonstrating that there is an interaction between BIN2 and YDA in vivo. Furthermore, co-expression of BIN2 retarded mobility of YDA, but not of CA–YDA bands in immunoblots (Fig. 3g). These results confirm that BIN2 mainly phosphorylates the YDA N-terminal regulatory domain.

Finally, we tested whether BIN2 phosphorylation of YDA affects YDA kinase activity and whether brassinosteroid and bikinin affect MAPK activity in plants. YDA was pre-incubated with BIN2 and ATP, or with a kinase-inactive mutant BIN2 as a control, and then purified and further incubated with MKK4 (its known substrate), bikinin and [^{32}P]γATP. Pre-incubation with BIN2, but not with mutant BIN2,

decreased YDA phosphorylation of MKK4 (Fig. 3h and Supplementary Fig. 12), indicating that BIN2 phosphorylation inhibits YDA activity. Consistent with BIN2 inactivation of YDA, the kinase activities of MPK3 and MPK6 were reduced in the det2 mutant but increased by treatment with bikinin or brassinolide (Fig. 3i and 3j).

Taken together, our genetic and biochemical analyses demonstrate that brassinosteroid negatively regulates stomatal development by inhibiting the BIN2-mediated phosphorylation and inactivation of YDA (Fig. 3k). When brassinosteroid levels are low, active BIN2 directly phosphorylates and inactivates YDA; reduced MAP kinase pathway activity can de-repress SPCH, allowing SPCH to initiate stomatal development. Brassinosteroid signalling through BRI1, BSK1 and BSU1 inactivates GSK3, resulting in activation of the MAP kinase pathway and inhibition of stomatal production (Fig. 3k).

This study supports a role of brassinosteroid as a master regulator that coordinates both physiological and developmental aspects of plant growth. Previous studies have demonstrated key functions of brassinosteroid in inhibiting photomorphogenesis and photosynthetic gene expression[25–27]. Here we find a role for brassinosteroid in stomatal production, which must be coordinated with other developmental processes to optimize photosynthetic and water-use efficiency. Notably, brassinosteroid represses light-responsive gene expression and chloroplast development mainly through the BZR1-mediated transcriptional network[26,27], but represses stomatal development through a BZR1-independent GSK3–MAPK crosstalk mechanism. Both GSK3 and MAPK are highly conserved in all eukaryotes, but it remains to be seen whether GSK3 directly inactivates MAPKKK

proteins in animals. This GSK3–MAPK connection has the potential to act in multiple receptor kinase-mediated signalling pathways, mediating crosstalk between these pathways in plants. The stronger stomata-clustering phenotype of *bsu-q* and suppression of *er erl1 erl2* stomata phenotypes by bikinin raise a possibility that members of the BSU1 and GSK3 families mediate signalling by the ERECTA family receptor kinases. However, the signals from BRI1 and ERECTA family must be partitioned differently downstream so that BRI1 controls GSK3 regulation of both BZR1 and YDA but ERECTA family mainly controls the GSK3 inactivation of YDA (Fig. 3k), because *er erl1 erl2* had no obvious effect on brassinosteroid-regulated BZR1 phosphorylation (Supplementary Fig. 13). Similar mechanisms and components might also be used by additional signalling pathways, such as the innate immunity pathway downstream of the FLS2 receptor kinase, which shares the BAK1 co-receptor[28] and downstream components MPK3 and MPK6 with BRI1 (ref. 23). In support of such an idea, overexpression of a GSK3-like kinase reduced the pathogen-induced activation of MPK3 and MPK6 (ref. 29). How signalling specificity is maintained when multiple pathways share the same components is a question for future study, and studies of the brassinosteroid model system will probably shed light on the hundreds of plant receptor kinases and their crosstalk during plant responses to complex endogenous and environmental cues.

METHODS SUMMARY

Stomatal quantification. Cotyledons of 8-day-old seedlings were cleared in ethanol with acetic acid (ratio of 19:1, v/v) and mounted on slides in Hoyer's solution (see ref. 22). Two to four images at ×400 magnification (180 µm^2) were captured per cotyledon from central regions of abaxial leaves. Guard cells, meristemoids, GMCs and pavement cells were counted. Statistical analysis was performed by Sigmaplot software (Systat Software). For treatment with bikinin[20], seedlings were grown on half-strength Murashige and Skoog (MS) medium containing dimethylsulphoxide (DMSO) or 30 µM bikinin (+10 µM oestradiol for HOPAI1-inducible lines) for 8 days before stomata were analysed.

Biochemical assays. To test the bikinin effect on YDA–Myc phosphorylation, homozygous *YDA-4Myc* plants were grown on 1/2 MS medium containing 2 µM BRZ (*BRASSINAZOLE*, an inhibitor of BR synthesis) for 5 days and treated with 30 µM bikinin or 2 µM BRZ solution for 30 min with gentle agitation. Yeast two-hybrid, *in vitro* interaction and kinase assays[9,12], and in-gel kinase assays[30] were carried out as described previously. Details of methods are available in the Supplementary Methods.

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

1. Bergmann, D. C. & Sack, F. D. Stomatal development. *Annu. Rev. Plant Biol.* **58**, 163–181 (2007).
2. Dong, J. & Bergmann, D. C. Stomatal patterning and development. *Curr. Top. Dev. Biol.* **91**, 267–297 (2010).
3. Thummel, C. S. & Chory, J. Steroid signaling in plants and insects—common themes, different pathways. *Genes Dev.* **16**, 3113–3129 (2002).
4. Kim, T. W. & Wang, Z. Y. Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant Biol.* **61**, 681–704 (2010).
5. Kinoshita, T. *et al.* Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433**, 167–171 (2005).
6. Hothorn, M. *et al.* Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* **474**, 467–471 (2011).
7. She, J. *et al.* Structural insight into brassinosteroid perception by BRI1. *Nature* **474**, 472–476 (2011).
8. Tang, W. *et al.* BSKs mediate signal transduction from the receptor kinase BRI1 in *Arabidopsis. Science* **321**, 557–560 (2008).
9. Kim, T.-W. *et al.* Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. *Nature Cell Biol.* **11**, 1254–1260 (2009).
10. Tang, W. *et al.* PP2A activates brassinosteroid-responsive gene expression and plant growth by dephosphorylating BZR1. *Nature Cell Biol.* **13**, 124–131 (2011).
11. Li, J. & Nam, K. H. Regulation of brassinosteroid signaling by a GSK3/SHAGGY-like kinase. *Science* **295**, 1299–1301 (2002).
12. He, J. X., Gendron, J. M., Yang, Y., Li, J. & Wang, Z. Y. The GSK3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in *Arabidopsis. Proc. Natl Acad. Sci. USA* **99**, 10185–10190 (2002).
13. Saibo, N. J., Vriezen, W. H., Beemster, G. T. & Van Der Straeten, D. Growth and stomata development of *Arabidopsis* hypocotyls are controlled by gibberellins and modulated by ethylene and auxins. *Plant J.* **33**, 989–1000 (2003).
14. Kang, C. Y., Lian, H. L., Wang, F. F., Huang, J. R. & Yang, H. Q. Cryptochromes, phytochromes, and COP1 regulate light-controlled stomatal development in *Arabidopsis. Plant Cell* **21**, 2624–2641 (2009).
15. Wang, H., Ngwenyama, N., Liu, Y., Walker, J. C. & Zhang, S. Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in *Arabidopsis. Plant Cell* **19**, 63–73 (2007).
16. Lampard, G. R., Macalister, C. A. & Bergmann, D. C. *Arabidopsis* stomatal initiation is controlled by MAPK-mediated regulation of the bHLH SPEECHLESS. *Science* **322**, 1113–1116 (2008).
17. Rowe, M. H. & Bergmann, D. C. Complex signals for simple cells: the expanding ranks of signals and receptors guiding stomatal development. *Curr. Opin. Plant Biol.* **13**, 548–555 (2010).
18. Szekeres, M. *et al.* Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in *Arabidopsis. Cell* **85**, 171–182 (1996).
19. Wang, Z. Y. *et al.* Nuclear-localized BZR1 mediates brassinosteroid-induced growth and feedback suppression of brassinosteroid biosynthesis. *Dev. Cell* **2**, 505–513 (2002).
20. De Rybel, B. *et al.* Chemical inhibition of a subset of *Arabidopsis thaliana* GSK3-like kinases activates brassinosteroid signaling. *Chem. Biol.* **16**, 594–604 (2009).
21. Rozhon, W., Mayerhofer, J., Petutschnig, E., Fujioka, S. & Jonak, C. ASKθ, a group-III *Arabidopsis* GSK3, functions in the brassinosteroid signalling pathway. *Plant J.* **62**, 215–223 (2010).
22. Bergmann, D. C., Lukowitz, W. & Somerville, C. R. Stomatal development and pattern controlled by a MAPKK kinase. *Science* **304**, 1494–1497 (2004).
23. Zhang, J. *et al.* A Pseudomonas syringae effector inactivates MAPKs to suppress PAMP-induced immunity in plants. *Cell Host Microbe* **1**, 175–185 (2007).
24. Kanaoka, M. M. *et al.* SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to *Arabidopsis* stomatal differentiation. *Plant Cell* **20**, 1775–1785 (2008).
25. Li, J., Nagpal, P., Vitart, V., McMorris, T. C. & Chory, J. A role for brassinosteroids in light-dependent development of *Arabidopsis. Science* **272**, 398–401 (1996).
26. Sun, Y. *et al.* Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in *Arabidopsis. Dev. Cell* **19**, 765–777 (2010).
27. Luo, X.-M. *et al.* Integration of light and brassinosteroid signaling pathways by a GATA transcription factor in *Arabidopsis. Dev. Cell* **19**, 872–883 (2010).
28. Chinchilla, D. *et al.* A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448**, 497–500 (2007).
29. Wrzaczek, M., Rozhon, W. & Jonak, C. A proteasome-regulated glycogen synthase kinase-3 modulates disease response in plants. *J. Biol. Chem.* **282**, 5249–5255 (2007).
30. Zhang, S. & Klessig, D. F. Salicylic acid activates a 48-kD MAP kinase in tobacco. *Plant Cell* **9**, 809–824 (1997).

METHODS

Materials and growth conditions. All mutants are in the Columbia ecotype except *yda* Y295 (C24 ecotype)[22], CA-YDA (Ler ecotype)[22] and *bin2-3 bil1 bil2* triple mutant obtained from J. Li (Ws ecotype)[31]. The *erecta* triple mutant *er105 erl1-2 erl2-1* (ref. 32) and *scrm-D* (ref. 24) were obtained from K. Torii. J.-M. Zhou provided seeds of oestradiol-inducible *HOPAI1* transgenic plants[23]. For all analyses, *Arabidopsis* seedlings were grown on MS agar medium for 8 days under continuous light in Percival growth chamber at 22 °C.

Stomatal quantification. Cotyledons of 8-day-old seedlings were cleared in ethanol with acetic acid and mounted on slides in Hoyer's solution (see ref. 22). Two to four images at ×400 magnification (180 μm^2) were captured per cotyledon from central regions of abaxial leaves. Guard cells, meristemoids, GMCs and pavement cells were counted. Statistical analysis was performed by Sigmaplot software (Systat Software). For treatment with bikinin[20], seedlings were grown on half-strength MS medium containing DMSO or 30 μM bikinin (+ 10 μM estradiol for HOPAI1-inducible lines) for 8 days before stomata were analysed.

Plasmids. For cloning MBP-185/322, a partial cDNA was amplified from a YDA cDNA clone using primers (forward; 5′-caccAGTAACAAAAACTCAGCTG AGATGTTT-3′, reverse; 5′-AGAGCTAG GACCAGGGCTTGTCATTCT-3′), cloned into pENTR-SD-D-TOPO vector (Invitrogen) and then subcloned into the gateway-compatible pMALc2 vector (New England Biolab). For expression in plants, cDNA entry clones of YDA and CA-YDA were subcloned into a gateway-compatible 35S::4myc-6His vector constructed in the pCAMBIA 1390 vector. *BSL2* cDNA in the pENTR vector was subcloned into Gateway-compatible pEarley-101 vector[33] to generate *35S::BSL2-YFP*.

Overlay assay. To test the interaction of YDA and BIN2 *in vitro*, a gel blot separating MBP, MBP–CDG1 (a protein kinase used as a negative control) and MBP–YDA was incubated with 20 μg GST–BIN2 in 5% non-fat dry milk/PBS buffer and washed four times. The blot was then probed with HRP-conjugated anti-GST antibody (Santa Cruz Biotechnology).

In vitro kinase assay. Induction and purification of proteins expressed from *Escherichia coli* was performed as described previously[12]. For Fig. 3d, e, 1 μg of GST–BIN2 or 0.5 μg of MBP–BIN2 was incubated with 1 μg of MBP–YDA or MBP-185/322 in the kinase buffer (20 mM Tris, pH 7.5, 1 mM MgCl$_2$, 100 mM NaCl and 1 mM DTT) containing 100 μM ATP and 10 μCi [^{32}P]γATP at 30 °C for 3 h. To examine whether BIN2 inhibits YDA activity, equal amounts of MBP–YDA were pre-incubated with GST–BIN2 or GST–mBIN2 (M115A) for 2 h. Pre-incubated MBP–YDA was subsequently purified using glutathione beads and amylose beads to remove GST–BIN2 or GST–mBIN2. Purified YDA was then incubated with GST–mMKK4 (K108R), 10 μCi [^{32}P]γATP and 10 μM bikinin (to inhibit any residual BIN2) at 30 °C for 3 h. YDA kinase activity towards mMKK4 was analysed by SDS–PAGE followed by autoradiography.

In-gel kinase assay. The in-gel kinase assay was performed as described previously[30], with some modifications. Total proteins were extracted with buffer containing 50 mM Tris, pH 7.5, 150 mM NaCl, 5% Glycerol, 1% Triton X-100, 1 mM phenylmethylsulphonyl fluoride (PMSF), 1 μM E-64, 1 μM bestatin, 1 μM pepstatin and 2 μM leupeptin. Supernatant obtained from 12,000 r.p.m. centrifugation was quantified by Bradford protein assay. Equal amounts of protein (40 μg) were loaded on 10% SDS–PAGE gel embedded with 0.2 mg ml^{-1} of myelin basic protein. After electrophoresis, SDS was removed by incubation with washing buffer (25 mM Tris, pH 7.5, 0.5 mM DTT, 5 mM NaF, 0.1 mM Na$_3$VO$_4$, 0.5 mg ml^{-1} bovine serum albumin and 0.1% Triton X-100) with three buffer exchanges at 22 °C for 1.5 h. The gel was incubated with renaturation buffer (25 mM Tris, pH 7.5, 0.5 mM DTT, 5 mM NaF and 0.1 mM Na$_3$VO$_4$) at 4 °C overnight with four buffer exchanges. After pre-incubation with 100 ml of kinase reaction buffer without ATP for 30 min, the gel was incubated with 30 ml of kinase reaction buffer (25 mM Tris, pH 7.5, 2 mM EGTA, 12 mM MgCl$_2$, 1 mM DTT, 0.1 mM Na$_3$VO$_4$, 200 nM ATP and 50 μCi [^{32}P]γ-ATP) for 1.5 h. The gel was washed with solution containing 5% trichloroacetic acid (w/v) and 1% potassium pyrophosphate (w/v) four times for 2–3 h. Dried gel was exposed with phosphor screen followed by phospho-imager analysis.

Transient interaction assays and analysis of bikinin effects on YDA in transgenic plants. *Agrobacterium* GV3101 strains transformed with *35S::CA-YDA-4Myc-6His* or *35S::YDA-4Myc-6His* constructs were alone or co-infiltrated with *35S-BIN2-YFP* expressing *Agrobacterium* into *N. benthamiana* leaves as described previously[9]. After 36 h, protein extracts were prepared from *N. benthamiana* leaves in immunoprecipitation buffer containing 50 mM Tris, pH 7.5, 150 mM NaCl, 5% Glycerol, 1% Triton X-100, 1 mM PMSF, 1 μM E-64, 1 μM bestatin, 1 μM pepstatin and 2 μM leupeptin. Supernatant obtained from 20,000g centrifugation was incubated with anti-YFP-antibody-bound protein A beads for 1 h. Beads were washed 5 times with immunoprecipitation buffer containing 0.2% Triton X-100. Immunoprecipitated proteins were eluted with 2× SDS Laemmli buffer, separated on SDS–PAGE and subjected to immunoblotting using anti-Myc antibody (Abcam) and anti-YFP antibody.

For transgenic *Arabidopsis* plants, wild-type *Arabidopsis* was transformed with *Agrobacterium* containing *35S::YDA-4Myc-6His* or *35S::BSL2-YFP* construct by floral dip. Hygromycin or Basta-resistant T1 plants were screened by immunoblot using anti-Myc or anti-YFP antibody, respectively.

To test the bikinin effect on YDA–Myc phosphorylation, homozygous *YDA-4Myc* plants were grown on half-strength MS medium containing 2 μM BRZ for 5 days and treated with 30 μM bikinin or 2 μM BRZ solution for 30 min with gentle agitation. YDA-Myc was analysed by immunoblot.

31. Yan, Z., Zhao, J., Peng, P., Chihara, R. K. & Li, J. BIN2 functions redundantly with other *Arabidopsis* GSK3-like kinases to regulate brassinosteroid signaling. *Plant Physiol.* **150,** 710–721 (2009).
32. Shpak, E. D., McAbee, J. M., Pillitteri, L. J. & Torii, K. U. Stomatal patterning and differentiation by synergistic interactions of receptor kinases. *Science* **309,** 290–293 (2005).
33. Earley, K. W. et al. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45,** 616–629 (2006).

LETTER

Expression of tumour–specific antigens underlies cancer immunoediting

Michel DuPage[1], Claire Mazumdar[1], Leah M. Schmidt[1], Ann F. Cheung[1] & Tyler Jacks[1,2]

Cancer immunoediting is a process by which immune cells, particularly lymphocytes of the adaptive immune system, protect the host from the development of cancer and alter tumour progression by driving the outgrowth of tumour cells with decreased sensitivity to immune attack[1,2]. Carcinogen-induced mouse models of cancer have shown that primary tumour susceptibility is thereby enhanced in immune-compromised mice, whereas the capacity for such tumours to grow after transplantation into wild-type mice is reduced[2,3]. However, many questions about the process of cancer immunoediting remain unanswered, in part because of the known antigenic complexity and heterogeneity of carcinogen-induced tumours[4]. Here we adapted a genetically engineered, autochthonous mouse model of sarcomagenesis to investigate the process of cancer immunoediting. This system allows us to monitor the onset and growth of immunogenic and non-immunogenic tumours induced *in situ* that harbour identical genetic and histopathological characteristics. By comparing the development of such tumours in immune-competent mice with their development in mice with broad immunodeficiency or specific antigenic tolerance, we show that recognition of tumour-specific antigens by lymphocytes is critical for immunoediting against sarcomas. Furthermore, primary sarcomas were edited to become less immunogenic through the selective outgrowth of cells that were able to escape T lymphocyte attack. Loss of tumour antigen expression or presentation on major histocompatibility complex I was necessary and sufficient for this immunoediting process to occur. These results highlight the importance of tumour-specific-antigen expression in immune surveillance, and potentially, immunotherapy.

To determine whether T lymphocytes influence tumour development, we adapted a mouse model of human soft tissue sarcomagenesis driven by Cre/LoxP-regulated expression of oncogenic K-ras^{G12D} and deletion of p53 to allow for the control of tumour immunogenicity[5]. Sarcomas were induced in either immune-competent $Kras^{LSL-G12D/+};p53^{fl/fl};Rag2^{+/-}$ (KP) or lymphocyte-deficient $Kras^{LSL-G12D/+};p53^{fl/fl};Rag2^{-/-}$ (KPR) mice by intramuscular injection of lentiviral vectors that expressed Cre recombinase alone (Lenti-x). To induce sarcomas with potentially immunogenic antigens, we used vectors that also expressed the T-cell antigens SIYRYYGL (SIY) and two antigens from ovalbumin (SIINFEKL (SIN, OVA$_{257-264}$) and OVA$_{323-339}$) fused to the carboxy terminus of luciferase (Lenti-LucOS). Intramuscular injection of Lenti-LucOS led to tumour formation in 100% of KPR mice but only 27% of KP mice by 140 days (Fig. 1a, $P < 0.0001$). Additional sarcomas ultimately developed in KP mice but with dramatically delayed kinetics (latency of 194.8 ± 43.4 days) compared with KPR mice (73.6 ± 4.3 days) (Fig. 1c, $P < 0.02$). We also observed a difference in the penetrance of sarcoma development in KPR versus KP mice by 140 days with Lenti-x (89% versus 43%, respectively), although the difference was less dramatic than observed with Lenti-LucOS (Fig. 1b, $P < 0.0005$). This suggests that in this model, tumour immunosurveillance may not necessitate the introduction of highly immunogenic tumour-specific

antigens (TSAs). The observed immunosurveillance against Lenti-x tumours could result from the lentiviral infection required to induce tumours, the acquisition of TSAs during tumour development, or the immunogenicity of Cre itself. However, in a previous study, we found that Cre was not highly immunogenic when expressed in developing lung adenocarcinomas[6]. Although Lenti-x-induced sarcoma development was slightly delayed in immune-competent (KP) mice (114.9 days in KP versus 79.5 days in KPR mice), it was not significant (Fig. 1c, $P = 0.11$). The increased latency that is specific to Lenti-LucOS tumours may be the result of an equilibrium between replicating tumour cells and T cells that recognize antigens expressed from the LucOS vector and restrain tumour progression[1,7].

Rag2 (recombination activating gene 2) deficiency prevents both T and B lymphocyte development and, therefore, could have pleiotropic effects on the immune response to tumour antigens. To specifically test
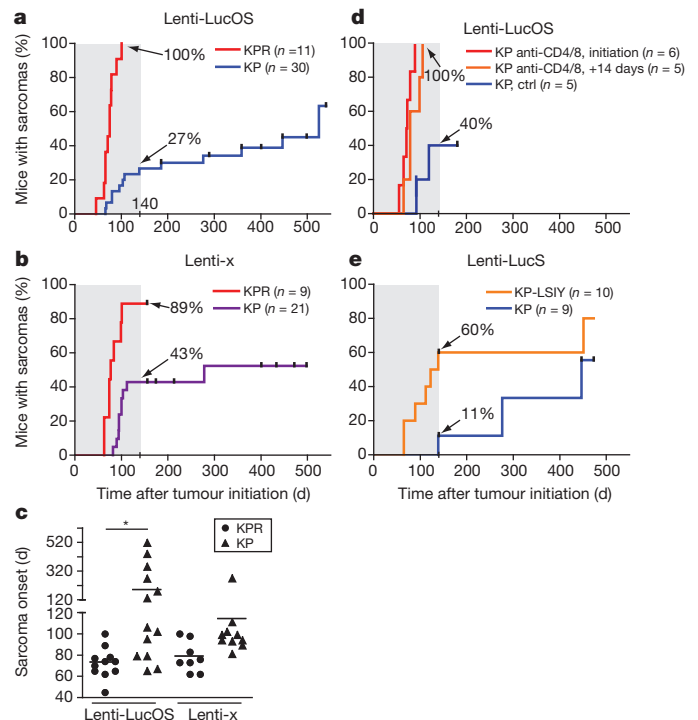

Figure 1 | Sarcoma formation in immunodeficient compared to wild-type mice occurs with increased penetrance and reduced latency. a, b, KPR or KP mice were injected intramuscularly with Lenti-LucOS (**a**) or Lenti-x (**b**) and the onset of palpable tumours was monitored. **c**, Time for palpable tumour formation with Lenti-LucOS or Lenti-x in KPR (circles) or KP (triangles) mice. **d**, Sarcoma formation in KP mice either untreated or treated with anti-CD4 and anti-CD8 antibodies beginning coincident with or 14 days after Lenti-LucOS injection. **e**, Sarcoma onset after injection of KP-LSIY or KP littermates with Lenti-LucS. The percentage of total mice (*n*) with sarcomas by 140 days (grey boxes) is indicated.

[1]Koch Institute for Integrative Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

the significance of T-cell responses, we treated mice with antibodies against CD4 and CD8 to deplete T cells concurrent with, or subsequent to, intramuscular injection of Lenti-LucOS. T-cell depletion at tumour initiation, or even 14 days after tumour initiation, led to sarcoma development with complete penetrance and early onset similar to KPR mice (Fig. 1d, $P = 0.001$ and $P = 0.013$ compared to untreated, respectively). To specifically test the importance of CD8[+] T cells that recognize the model TSAs, we made use of a regulatable luciferase–SIY fusion gene engineered into the murine Rosa26 locus ($R26^{LSL-LSIY}$)[8]. These mice develop specific tolerance to luciferase and SIY due to weak thymic expression and deletion of reactive T cells (Supplementary Fig. 1)[8]. $Kras^{LSL-G12D/+};p53^{fl/fl};R26^{LSL-LSIY/+}$ (KP-LSIY) mice injected with Lenti-LucS, a lenti-vector that expresses Cre and SIY fused to luciferase, were more susceptible to sarcoma formation and developed tumours earlier than KP littermates (Fig. 1e, $P = 0.058$). Thus, lymphocyte-mediated protection from sarcoma formation requires CD8[+] T cells that respond to non-self antigens expressed in tumours.

A key advantage of this conditional, genetically engineered cancer model over carcinogen-induced models is the capacity to track endogenous T cells specific for tumour antigens during primary tumour development. We used SIY and SIN loaded MHCI/K[b] reagents to track tumour-reactive CD8[+] T cells by flow cytometry. Only mice with Lenti-LucOS sarcomas harboured CD8[+] T cells specific to SIY and SIN in the lymph nodes nearest the tumour site as well as in the spleen (Fig. 2a, b). These CD8[+] T cells appeared to be completely functional because they produced both IFN-γ and TNF-α upon stimulation

(Fig. 2a–d). Interestingly, this contrasts sharply with results from an analogous model of lung adenocarcinoma in which the activity of T cells responding to the same tumour antigens was very weak, suggesting that different tumour types may use different mechanisms to escape immune attack[6]. We also investigated whether KP mice that did not develop sarcomas after injection with Lenti-LucOS harboured antigen-specific T cells, because such T cells could have protected these mice from sarcoma development. Indeed, we detected fully functional antigen-specific T cells in these mice (Fig. 2c, d and Supplementary Fig. 1), demonstrating that T cells specific to these model TSAs are functional and probably provide significant protection against the development of Lenti-LucOS sarcomas.

Pivotal experiments using methylcholanthrene (MCA)-induced sarcomas revealed that tumours generated in immune-compromised mice, and thus not immunoedited, are more susceptible to rejection upon transplantation into immune-competent mice[2]. To assay whether autochthonous sarcomas driven by targeted genetic mutations would also display an unedited phenotype, we transplanted independently derived sarcomas from KPR or KP mice into either wild-type or $Rag2^{-/-}$ mice. Whereas freshly isolated Lenti-LucOS-induced tumours (or cell lines) generated in KP mice grew similarly upon transplantation into either wild-type or $Rag2^{-/-}$ mice, most Lenti-LucOS tumours generated in KPR mice were rejected (1/7) or had significantly delayed growth (4/7) (Fig. 3a, b and Supplementary Fig. 2). These results recapitulate the original findings from carcinogen-induced sarcomas in a genetically engineered mouse model of sarcomagenesis.
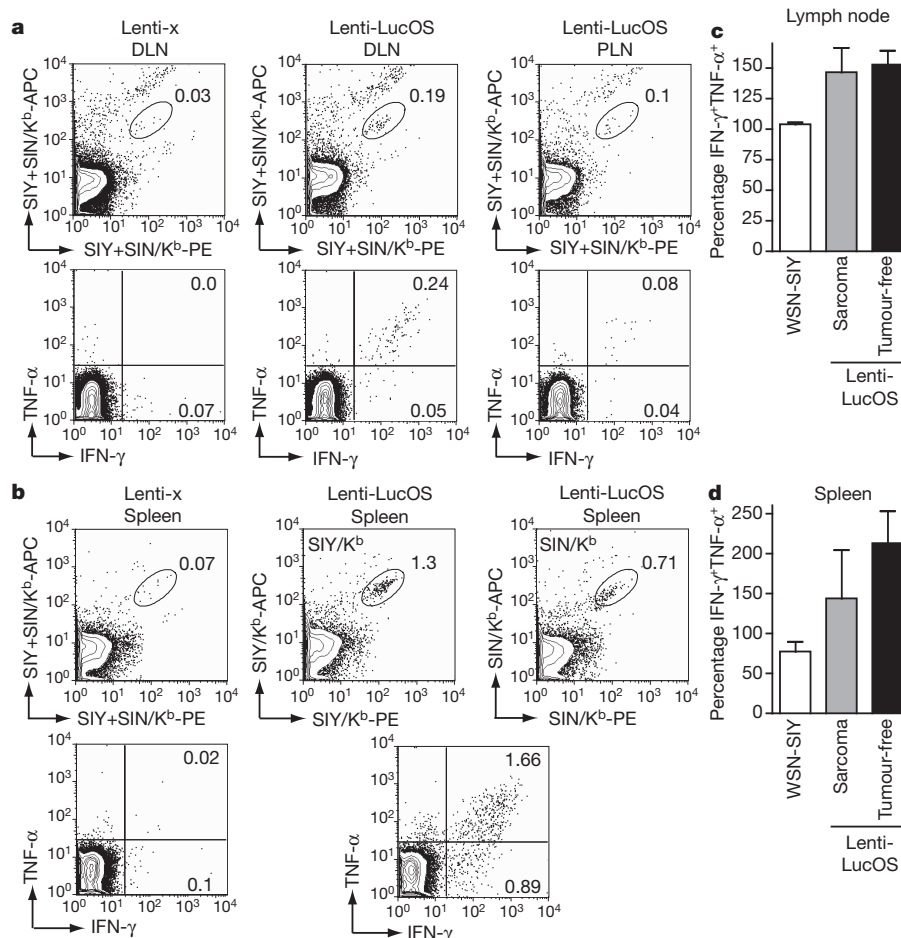

Figure 2 | Functional T-cell responses are generated against antigens expressed in sarcomas. a, Top: percentage of gated CD8[+] cells specific for SIY and SIN in the inguinal lymph nodes either draining (DLN) or peripheral to (PLN) Lenti-x or Lenti-LucOS tumours. Bottom: IFN-γ and TNF-α cytokine production in SIY+SIN-stimulated CD8[+] T cells from mice analysed above. **b**, Analysis of splenocytes as in **a**. **c, d**, Cumulative data depicting the percentage of SIY- and SIN-specific T cells that were IFN-γ^+TNF-α^+ from lymph nodes (**c**) or spleens (**d**) of KP mice infected with Lenti-LucOS that developed a 'sarcoma' or were 'tumour-free' at 170 days. T cells reactive to SIY were analysed four months after challenge with WSN-SIY (influenza strain expressing SIY). Data represent analysis of 3–4 mice per group, mean $\pm$ s.e.m.

Next we wanted to determine whether Lenti-x-induced sarcomas, which lack the strong T-cell antigens from LucOS, would yield similar results. Interestingly, Lenti-x tumours generated in KPR or KP mice grew equally well when transplanted into wild-type or $Rag2^{-/-}$ mice (Fig. 3c, d). It is noteworthy that while autochthonous tumours initiated by Lenti-x appeared partially inhibited by an adaptive immune response (Fig. 1b), in the context of transplantation, we found no evidence of immunoediting (Fig. 3c). This difference may be due to Rag-dependent innate immune cells (NKT and γδ T cells) that recognize stress or inflammatory ligands. These cells may be sufficient to eliminate a limited number of nascent tumour cells in the context of transformation by lentiviral infection, but not in response to the transplantation of

fully developed tumours[1,9]. Nevertheless, we suggest that Lenti-x sarcomas from KPR mice grew unabated after transplantation into KP mice because immunoediting by T lymphocytes requires potent TSAs, which these tumours lack. The observed immunogenicity of carcinogen-induced sarcomas generated in immune-compromised mice may be due to the *de novo* generation of potent tumour neoantigens during transformation with mutagens[9–12]. Importantly, in another study reported in this issue[13], somatically mutated spectrin-β2 in a MCA-induced sarcoma was found to act as a potent neoantigen that drove the immunoediting process. In an attempt to introduce immunogenic mutations in Lenti-x tumours, we treated cell lines from these tumours with MCA *in vitro*. Interestingly, such treatment rarely yielded clones with increased immunogenicity (Supplementary Fig. 3). This may indicate that although carcinogens can produce mutations that are immunogenic, it may be a rare event.

If cancer immunoediting by lymphocytes requires potent TSAs, then Lenti-LucOS-induced tumours that appear edited after forming in KP mice may have evaded an immune response by the selective outgrowth of cells lacking these potent antigens[14–16]. To assess antigen expression, we measured luciferase activity in tumours. Whereas tumours from KPR mice were universally luciferase positive, tumours from KP mice had drastically reduced luciferase activity in all but one of six sarcomas (Fig. 4a, b). Interestingly, this sarcoma had significantly reduced expression of H-2K^b, the MHC class I allele responsible for presenting the SIY and SIN antigens (Fig. 4c). Sarcomas from KP mice treated with anti-CD4 and anti-CD8 antibodies at tumour initiation also retained luciferase activity (5/6 sarcomas luc$^+$, Fig. 4a). However, fewer sarcomas retained luciferase expression when mice were treated with anti-CD4 and anti-CD8 antibodies beginning 14 days after tumour initiation (1/5 sarcomas luc$^+$), suggesting that immunoediting can occur very early during sarcoma development. Thus, by selectively eliminating cells that express potent TSAs, T lymphocytes drive the escape of tumour cells that either do not express potent antigens or cannot present the antigens to reactive T cells.

In a similar fashion to the antigen loss observed in autochthonous sarcomas, Lenti-LucOS-induced sarcomas from KPR mice lost antigen expression when transplanted into wild-type mice (Supplementary Fig. 4). Importantly, tumours that lost antigen expression after being passaged through wild-type mice grew comparably upon secondary transplantation into wild-type and $Rag2^{-/-}$ mice, whereas tumours passaged through $Rag2^{-/-}$ mice did not (Supplementary Fig. 4). To test whether antigen loss was sufficient to provide a means of escape for Lenti-LucOS sarcomas generated in KP mice, we reintroduced the LucOS antigens into sarcomas that had lost expression of the antigens after passage through wild-type mice (referred to as antigen$^-$ tumours). Indeed, re-expression of LucOS led to severely reduced tumour growth (Fig. 4d), indicating that loss of antigen expression was the primary means of tumour escape in this setting.

Epigenetic silencing of tumour antigen expression via DNA methylation could be responsible for antigen loss and tumour escape[17,18]. To test this hypothesis, we treated cell lines that had lost luciferase expression after transplantation into immune-competent mice with 5-aza-2′-deoxycytidine (Aza), which reverses epigenetic silencing by inhibiting DNA methylation. In several lines tested, luciferase activity was restored with Aza treatment (Fig. 4e). Therefore, epigenetic silencing of tumour antigens may represent an important mechanism by which tumours can be edited in response to immune surveillance.

Here we have overcome many of the obstacles of carcinogen-induced models of cancer by using an autochthonous, genetically engineered model of sarcomagenesis to show that T lymphocyte-driven tumour antigen loss is a critical means by which cancer immunoediting occurs in a primary tumour setting. Although this study was limited to investigating the role of anti-tumour immunity by T cells, this model could be adapted to investigate the role of other critical immune cells in cancer immunoediting, such as B cells or NK cells, by either introducing surface-expressed or stress-related antigens into
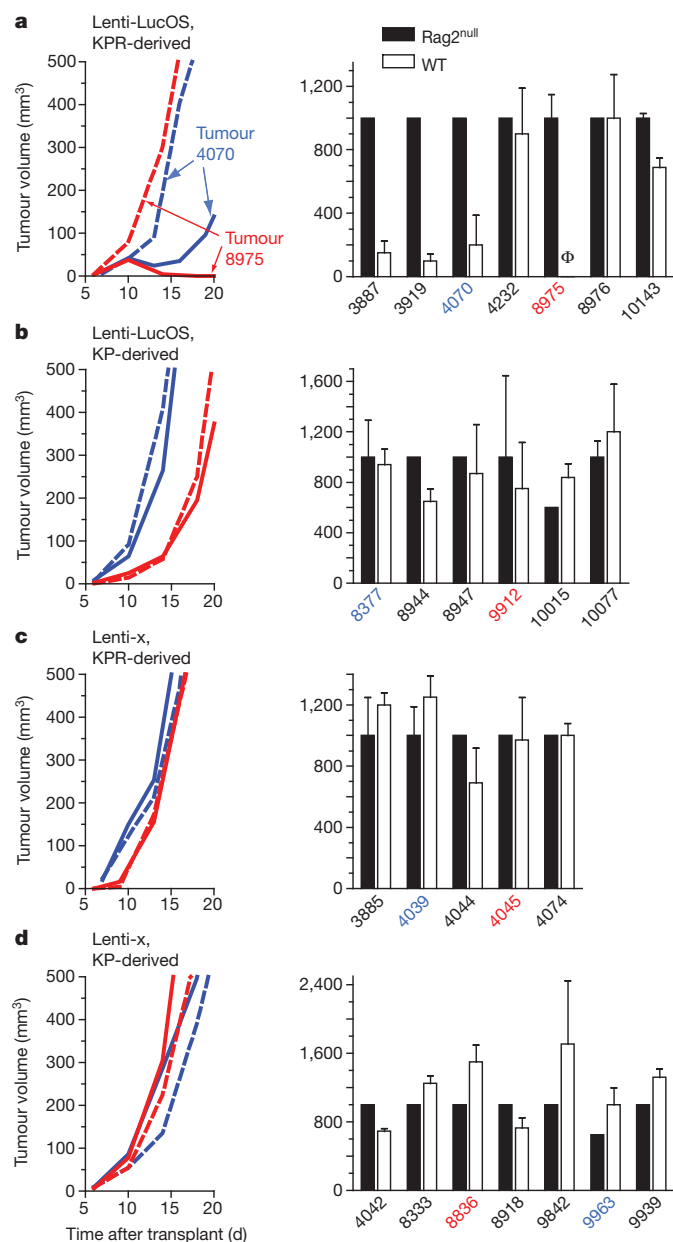

Figure 3 | Cancer immunoediting phenotypes require the presence of potent T-cell antigens. Transplanted tumour growth of Lenti-LucOS-induced sarcomas generated in KPR (**a**) or KP (**b**) mice and Lenti-x-induced sarcomas generated in KPR (**c**) or KP (**d**) mice. Left column, representative tumour growth curves from two different primary tumours (coloured red or blue) after transplantation into Rag2^{null} (dashed lines) or wild-type (WT, solid lines) mice. Right column, comparison of the mean tumour volume ± s.e.m. for all tumours transplanted. Φ indicates no detectable mass. See Supplementary Fig. 2 for growth curves of tumour lines.
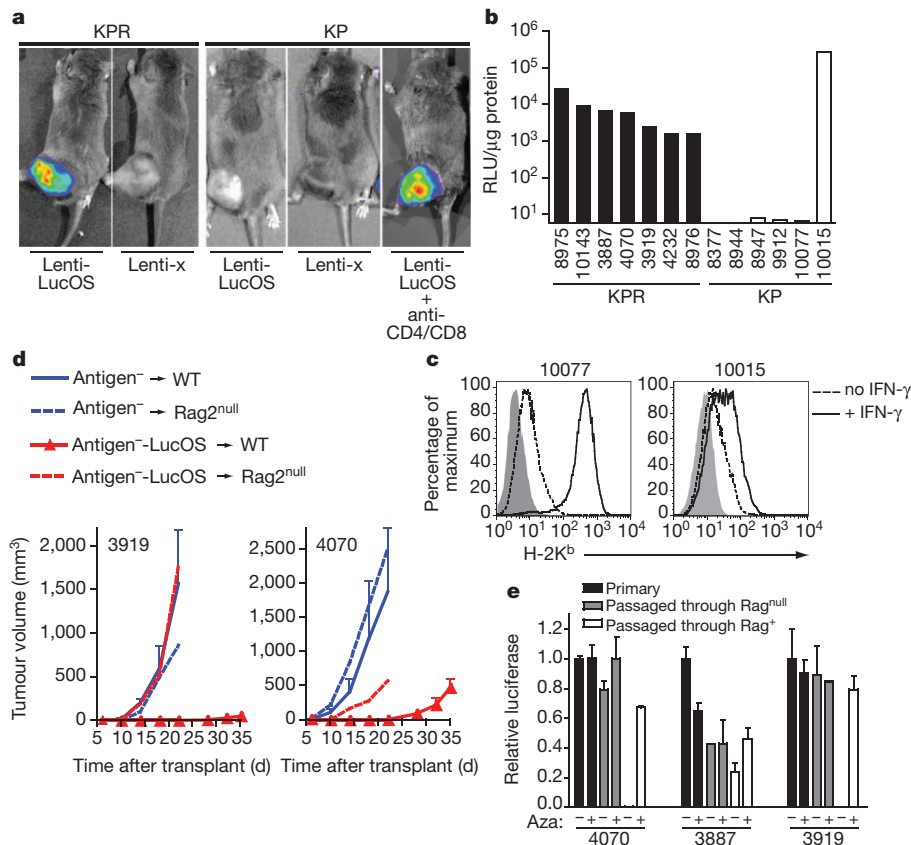
Figure 4 | Immunoediting occurs by selecting for tumour cells that do not express targeted antigens. a, Representative luciferase activity of Lenti-LucOS and Lenti-x-induced sarcomas in KPR, KP or anti-CD4/CD8 treated KP mice. **b**, Luciferase expression in Lenti-LucOS-induced sarcoma cell lines generated in KPR or KP mice. **c**, Freshly harvested sarcomas cultured with IFN-γ (solid line) or untreated (dashed line) were analysed for H-2K^b surface expression (shaded, control antibody). **d**, Growth of two independent tumours (3919 and 4070) that had lost antigen expression (antigen$^-$, blue lines) or the same

tumours, respectively[19–21]. This study resulted in two key discoveries. First, oncogene-driven, endogenous tumours can undergo immunoediting in a manner similar to carcinogen-driven tumours if engineered to express model TSAs. The immunogenicity of MCA-induced sarcomas is well-documented, and may be a direct consequence of TSAs that arise from carcinogen-induced mutations of normal genes during tumour development[9,11–13]. In contrast, cancers that arise spontaneously or by targeted genetic mutations in mice have been reported to be weakly immunogenic[22–25]. However, the mutational requirements for tumorigenesis in humans may be greater than in mice[26], and thus it is possible that spontaneous or genetically engineered mouse models of cancer might underestimate the mutational and antigenic load of most human cancers. This idea is supported by the second critical finding of this study—that tumour immunogenicity is not a universal characteristic of cancer development. By obviating the need for carcinogens, we could induce sarcomas that potentially lacked potent TSAs. These tumours had significantly reduced immunogenicity despite no previous engagement with the adaptive immune system and hence no opportunity for immunoediting. These results provide the first (to our knowledge) experimental system to unify the apparently conflicting results obtained using either carcinogen-induced or genetically targeted mouse models of cancer by identifying TSAs as the critical determinants that invoke adaptive immunosurveillance and immunoediting[2,22]. We propose that identifying and characterizing TSAs in human cancers may be critical for the generation of more effective anti-cancer immunotherapies in patients suffering from this disease.

tumour lines after reintroduction of LucOS (antigen$^-$-LucOS, red lines). Mean tumour volume $\pm$ s.e.m. after transplantation into three wild-type mice (solid lines) or one Rag2^{null} mouse (dashed lines). **e**, Relative luciferase activity (compared to the primary sarcoma) $\pm$ 5-aza-2'-deoxycytidine (Aza) of Lenti-LucOS sarcomas from KPR mice (Primary, black columns) that were passaged through Rag2$^{-/-}$ (Passaged through Rag^{null}, grey columns) or wild-type mice (Passaged through Rag$^+$, white columns). Mean $\pm$ s.e.m. from two experiments.

METHODS SUMMARY

Experiments used mice of the 129S$_4$/SvJae strain. All animal studies and procedures were approved by the Massachusetts Institute of Technology's Committee for Animal Care. Sarcomas were induced in KP and KPR mice by intramuscular injection of the hind limb with replication-incompetent lentiviruses expressing Cre recombinase as reported previously[5,6]. To deplete T cells, anti-CD4 (GK1.5) and anti-CD8 (YTS169.4) antibodies were administered at a dose of 250 µg per mouse by i.p. injection once weekly for the duration of the experiment. Flow cytometry was performed as described[6]. For transplantation experiments, 2×10^5 freshly isolated tumour cells or cultured tumour cells were transplanted subcutaneously into immune-competent or Rag2$^{-/-}$ mice of the 129S$_4$/SvJae background. Tumour volumes were calculated by multiplying the length $\times$ width $\times$ height of each tumour. To detect luciferase activity, freshly explanted tumours or cell lines were lysed, mixed with Luciferin reagent (Promega), and relative light units (RLU) were detected with a luminometer (MGM Instruments). Aza treatment used 1 µM 5-aza-2'-deoxycytidine for three days. In vivo bioluminescence images were acquired with the NightOWLII LB983 (Berthold Technologies) or the IVIS Spectrum (Xenogen) after intraperitoneal injection of 1.5 mg beetle luciferin (Promega). Statistical analyses used unpaired two-tailed Fisher exact probability tests or Student's t-tests.

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

1. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. *Nature Immunol.* **3**, 991–998 (2002).

2. Shankaran, V. *et al.* IFNγ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410,** 1107–1111 (2001).
3. Swann, J. B. *et al.* Demonstration of inflammation-induced cancer and cancer immunoediting during primary tumorigenesis. *Proc. Natl Acad. Sci. USA* **105,** 652–656 (2008).
4. Qin, Z. & Blankenstein, T. A cancer immunosurveillance controversy. *Nature Immunol.* 5, 3–4 (2004); author reply **5,** 4–5 (2004).
5. Kirsch, D. G. *et al.* A spatially and temporally restricted mouse model of soft tissue sarcoma. *Nature Med.* **13,** 992–997 (2007).
6. DuPage, M. *et al.* Endogenous T cell responses to antigens expressed in lung adenocarcinomas delay malignant tumor progression. *Cancer Cell* **19,** 72–85 (2011).
7. Koebel, C. M. *et al.* Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* **450,** 903–907 (2007).
8. Cheung, A. F., Dupage, M. J., Dong, H. K., Chen, J. & Jacks, T. Regulated expression of a tumor-associated antigen reveals multiple levels of T-cell tolerance in a mouse model of lung cancer. *Cancer Res.* **68,** 9459–9468 (2008).
9. Khong, H. T. & Restifo, N. P. Natural selection of tumor variants in the generation of "tumor escape" phenotypes. *Nature Immunol.* **3,** 999–1005 (2002).
10. Prehn, R. T. & Main, J. M. Immunity to methylcholanthrene-induced sarcomas. *J. Natl. Cancer Inst.* **18,** 769–778 (1957).
11. Dubey, P. *et al.* The immunodominant antigen of an ultraviolet-induced regressor tumor is generated by a somatic point mutation in the DEAD box helicase p68. *J. Exp. Med.* **185,** 695–705 (1997).
12. Monach, P. A., Meredith, S. C., Siegel, C. T. & Schreiber, H. A unique tumor antigen produced by a single amino acid substitution. *Immunity* **2,** 45–59 (1995).
13. Matsushita, H. *et al.* Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* http://dx.doi.org/10.1038/nature10755 (this issue).
14. Uyttenhove, C., Maryanski, J. & Boon, T. Escape of mouse mastocytoma P815 after nearly complete rejection is due to antigen-loss variants rather than immunosuppression. *J. Exp. Med.* **157,** 1040–1052 (1983).
15. Zhou, G., Lu, Z., McCadden, J. D., Levitsky, H. I. & Marson, A. L. Reciprocal changes in tumor antigenicity and antigen-specific T cell function during tumor progression. *J. Exp. Med.* **200,** 1581–1592 (2004).
16. Stauss, H. J., Van Waes, C., Fink, M. A., Starr, B. & Schreiber, H. Identification of a unique tumor antigen as rejection antigen by molecular cloning and gene transfer. *J. Exp. Med.* **164,** 1516–1530 (1986).
17. Guo, Z. S. *et al.* De novo induction of a cancer/testis antigen by 5-aza-2'-deoxycytidine augments adoptive immunotherapy in a murine tumor model. *Cancer Res.* **66,** 1105–1113 (2006).
18. Güre, A. O., Wei, I. J., Old, L. J. & Chen, Y. T. The SSX gene family: characterization of 9 complete genes. *Int. J. Cancer* **101,** 448–453 (2002).
19. Guerra, N. *et al.* NKG2D-deficient mice are defective in tumor surveillance in models of spontaneous malignancy. *Immunity* 28, 571–580 (2008); correction **28,** 723 (2008).
20. Schietinger, A. *et al.* A mutant chaperone converts a wild-type protein into a tumor-specific antigen. *Science* **314,** 304–308 (2006).
21. Zitvogel, L., Tesniere, A. & Kroemer, G. Cancer despite immunosurveillance: immunoselection and immunosubversion. *Nature Rev. Immunol.* **6,** 715–727 (2006).
22. Willimsky, G. & Blankenstein, T. Sporadic immunogenic tumours avoid destruction by inducing T-cell tolerance. *Nature* **437,** 141–146 (2005).
23. Embleton, M. J. & Heidelberger, C. Antigenicity of clones of mouse prostate cells transformed *in vitro*. *Int. J. Cancer* **9,** 8–18 (1972).
24. Hewitt, H. B., Blake, E. R. & Walder, A. S. A critique of the evidence for active host defence against cancer, based on personal studies of 27 murine tumours of spontaneous origin. *Br. J. Cancer* **33,** 241–259 (1976).
25. Scott, O. C. Tumor transplantation and tumor immunity: a personal view. *Cancer Res.* **51,** 757–763 (1991).
26. Rangarajan, A. & Weinberg, R. A. Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nature Rev. Cancer* **3,** 952–959 (2003).

METHODS

Mice and tumour induction. 129S$_4$/SvJae strains backcrossed 8 generations were used for all experiments. *Trp53*^{fl} mice were provided by A. Berns, *Kras*$^{LSL-G12D}$ were generated in our laboratory, and *Rag2*$^{-/-}$ mice were purchased from The Jackson Laboratory. Sarcomas were induced in KP and KPR mice by intramuscular injection of the left hind limb with replication-incompetent lentiviruses expressing Cre recombinase as reported previously[5,6]. Mice were monitored twice weekly for palpable sarcoma formation beginning 50 days after intramuscular injection. All animal studies and procedures were approved by the Massachusetts Institute of Technology's Committee for Animal Care.

Lentiviral production. Lentivirus was produced by transfection of 293T cells with Δ8.2 (gag/pol), CMV-VSV-G, and the various transfer vectors expressing Cre as described[27].

Antibody depletion. Anti-CD4 (GK1.5) and anti-CD8 (YTS169.4) antibodies were administered at a dose of 250 μg per mouse by i.p. injection once weekly for the duration of the experiment.

Preparation, culture and transplantation of primary sarcomas. Primary sarcomas were explanted and single cell suspensions were generated by mincing and digesting the tissues for ~1 h at 37 °C in 125 U ml^{-1} collagenase type I (Gibco), 60 U ml^{-1} hyaluronidase (Sigma), and 2 mg ml^{-1} collagenase/dispase (Roche), followed by passage through a 70 μm filter. Subcutaneous transplantation used 2×10^5 cells from freshly isolated tumour cells or cell lines from primary autochthonous tumours that were trypsinized and washed three times in plain DME medium. Transplant recipients were immune-competent or *Rag2*$^{-/-}$ mice on the 129S$_4$/SvJae background from the same mouse colony used to generate the autochthonous tumours. Subcutaneously transplanted tumour volumes were calculated by multiplying the length × width × height of each tumour. In Fig. 3, the mean volume ± s.e.m. of each tumour line is depicted after transplantation into wild-type mice (WT, open columns) at the time point when the same tumour line reached a volume of 1,000 mm^3 in the Rag2^{null} transplanted mice (Rag2^{null}, filled columns).

Flow cytometry. Cell suspensions from lymphoid organs were prepared by mechanical disruption between frosted slides. Cells were then stained with antibodies for 20–30 min after treatment with FcBlock (BD Pharmingen). Anti-CD8α (53-6.7), anti-IFNγ (XMG1.2), anti-TNFα (MP6-XT22), and DimerX I (Dimeric Mouse H-2K^b:Ig) were from BD Pharmingen. All antibodies were used at 1:200 dilution. Peptide-loaded DimerX reagents were prepared as directed and used at 1:75 dilution. To improve the sensitivity of the DimerX reagent, we used both PE and APC labelled dimers to co-stain CD8$^+$ T cells. Propidium iodide was used to exclude dead cells. Cells were read on a FACSCalibur and analysed using Flowjo software (Tree Star). In Fig. 2c, d, data were determined by comparing the fraction of CD8$^+$ cells in duplicate samples stained with K^b dimers or for cytokine production and exceeds 100% due to the incomplete sensitivity of the K^b dimers to detect antigen specific cells. In Fig. 4c, freshly harvested sarcomas were cultured for 24 h in the presence of 10 U IFN-γ (solid line) or untreated (dashed line) and analysed for H-2K^b surface expression (shaded, control antibody).

Cytokine production. Cells were resuspended in the presence or absence of SIYRYYGL and SIINFEKL peptides in OPTI-MEM I (Gibco) supplemented with GolgiPlug (BD Pharmingen) for ~4 h at 37 °C, 5% CO$_2$. Cells were then fixed and stained for intracellular cytokines using the Cytofix/Cytoperm kit (BD Biosciences).

Luciferase detection. Freshly explanted tumours or cell lines were lysed in Cell Culture Lysis Reagent, mixed with Luciferase Assay Reagent according to the manufacturer's instructions (Promega), and relative light units (RLU) were detected using the Optocomp I luminometer (MGM Instruments). RLUs were standardized by the total amount of protein (Bio-Rad Protein Asssay) in each sample. *In vivo* bioluminescence images were acquired with the NightOWLII LB983 (Berthold Technologies) or the IVIS Spectrum (Xenogen Corp.) after intraperitoneal injection of 1.5 mg beetle luciferin (Promega).

5-aza-2′-deoxycytidine treatment. Tumour cell lines were plated at low confluency (2×10^5 cells per well of 6-well plate), and treated with 1 μM 5-aza-2′-deoxycytidine replaced daily for three consecutive days and then analysed for luciferase activity.

Influenza. WSN-SIY (20 p.f.u. per mouse) provided by J. Chen. FACs analysis performed four months after intratracheal infection.

Statistical analyses. P-values were generated using unpaired two-tailed Fisher exact probability tests or Student's t-tests.

27. DuPage, M., Dooley, A. L. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nature Protocols* **4,** 1064–1072 (2009).

LETTER

DNase I sensitivity QTLs are a major determinant of human expression variation

Jacob F. Degner[1,2]*, Athma A. Pai[1]*, Roger Pique-Regi[1]*, Jean-Baptiste Veyrieras[1,3], Daniel J. Gaffney[1,4], Joseph K. Pickrell[1], Sherryl De Leon[4], Katelyn Michelini[4], Noah Lewellen[4], Gregory E. Crawford[5,6], Matthew Stephens[1,7], Yoav Gilad[1] & Jonathan K. Pritchard[1,4]

The mapping of expression quantitative trait loci (eQTLs) has emerged as an important tool for linking genetic variation to changes in gene regulation[1–5]. However, it remains difficult to identify the causal variants underlying eQTLs, and little is known about the regulatory mechanisms by which they act. Here we show that genetic variants that modify chromatin accessibility and transcription factor binding are a major mechanism through which genetic variation leads to gene expression differences among humans. We used DNase I sequencing to measure chromatin accessibility in 70 Yoruba lymphoblastoid cell lines, for which genome-wide genotypes and estimates of gene expression levels are also available[6–8]. We obtained a total of 2.7 billion uniquely mapped DNase I-sequencing (DNase-seq) reads, which allowed us to produce genome-wide maps of chromatin accessibility for each individual. We identified 8,902 locations at which the DNase-seq read depth correlated significantly with genotype at a nearby single nucleotide polymorphism or insertion/deletion (false discovery rate = 10%). We call such variants 'DNase I sensitivity quantitative trait loci' (dsQTLs). We found that dsQTLs are strongly enriched within inferred transcription factor binding sites and are frequently associated with allele-specific changes in transcription factor binding. A substantial fraction (16%) of dsQTLs are also associated with variation in the expression levels of nearby genes (that is, these loci are also classified as eQTLs). Conversely, we estimate that as many as 55% of eQTL single nucleotide polymorphisms are also dsQTLs. Our observations indicate that dsQTLs are highly abundant in the human genome and are likely to be important contributors to phenotypic variation.

It is now well established that eQTLs are abundant in a wide range of cell types and in diverse organisms, and recent studies have implicated human eQTLs as being important contributors to phenotypic variation[1–5]. However, the underlying regulatory mechanisms by which eQTLs affect gene expression remain poorly understood. One mechanism that may be important is when the alternative alleles at a particular single nucleotide polymorphism (SNP) lead to different levels of transcription factor binding or nucleosome occupancy at regulatory sites; this in turn may lead to allele-specific differences in transcription rates[9–12]. In this study we used DNase-seq in a panel of 70 individuals and found that a large fraction of eQTLs are indeed probably caused by this type of mechanism.

DNase-seq is a genome-wide extension of the classical DNase I footprinting method[13–15]. This assay identifies regions of chromatin that are accessible (or 'sensitive') to cleavage by the DNase I enzyme. Such regions are referred to as DNase I-hypersensitive sites (DHSs). DNase I sensitivity provides a precise, quantitative marker of regions of open chromatin and is well correlated with a variety of other markers of active regulatory regions including promoter-associated and enhancer-associated histone marks. Furthermore, bound transcription factors protect the DNA sequence within a binding site from DNase I cleavage, often producing recognizable 'footprints' of decreased DNase I sensitivity[13,15–17].

We collected DNase-seq data for 70 HapMap Yoruba lymphoblastoid cell lines for which gene expression data and genome-wide genotypes were already available[6–8]. We obtained an average of 39 million uniquely mapped DNase-seq reads per sample, providing individual maps of chromatin accessibility for each cell line (see Supplementary Information for all analysis details). Our data allowed us to characterize the distribution of DNase I cuts within individual hypersensitive sites at extremely high resolution. As expected, the DHSs coincided to a great extent with previously annotated regulatory regions, and DNase I sensitivity was positively correlated with the expression levels of nearby genes (Supplementary Figs 6 and 7). Overall, the locations of hypersensitive sites were highly correlated across individuals (Supplementary Information)[11].

We tested for genetic variants that affect local chromatin accessibility. To do this, we divided the genome into non-overlapping 100-base-pair (bp) windows, and then focused our analysis on the 5% of windows with the highest DNase I sensitivity (see Supplementary Information). For each individual we treated the number of DNase-seq reads in a given window, divided by the total number of mapped reads, as a quantitative trait that estimated the level of chromatin accessibility. We then tested for association between individual-specific DNase I sensitivity in each window and genotypes of all SNPs and insertions/deletions (indels) in a cis-candidate region of 40 kilobases (kb) centred on the target window.

Using this procedure, we identified associations between genotypes and inter-individual variation in DNase-seq read depth in 9,595 windows at a false discovery rate (FDR) of 10% (corresponding to 8,902 distinct DHSs, once we combined adjacent windows whose hypersensitivity data were associated with the same SNP or indel; Fig. 1a). We refer to these 8,902 loci as 'DNase I sensitivity QTLs', or dsQTLs, and show an example in Fig. 1c–f. We additionally considered a much smaller cis-candidate region of only 2 kb around each target window and found that most of the dsQTLs were detected within this smaller region (7,088 associated windows in 6,070 DHSs), suggesting that most dsQTLs lie close to the target DHS. In contrast, we found only weak evidence of trans-acting dsQTLs, probably because our experiment was underpowered for detecting these (Supplementary Information). For dsQTLs with enough DNase-seq reads overlapping the most significant SNP ($n = 892$), we confirmed that the fraction of reads carrying each allele in heterozygotes was well correlated with the dsQTL effect sizes (correlation coefficient $r = 0.72$, $P \ll 10^{-16}$; Fig. 1b).

We observed that dsQTLs typically affected chromatin accessibility for about 200–300 bp (Fig. 2a). Of the DHSs affected by dsQTLs, 77%

[1]Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. [2]Committee on Genetics, Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. [3]BioMiningLabs, 69001 Lyon, France. [4]Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637, USA. [5]Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina 27708, USA. [6]Department of Pediatrics, Division of Medical Genetics, Duke University School of Medicine, Durham, North Carolina 27708, USA. [7]Department of Statistics, University of Chicago, Chicago, Illinois 60637, USA.
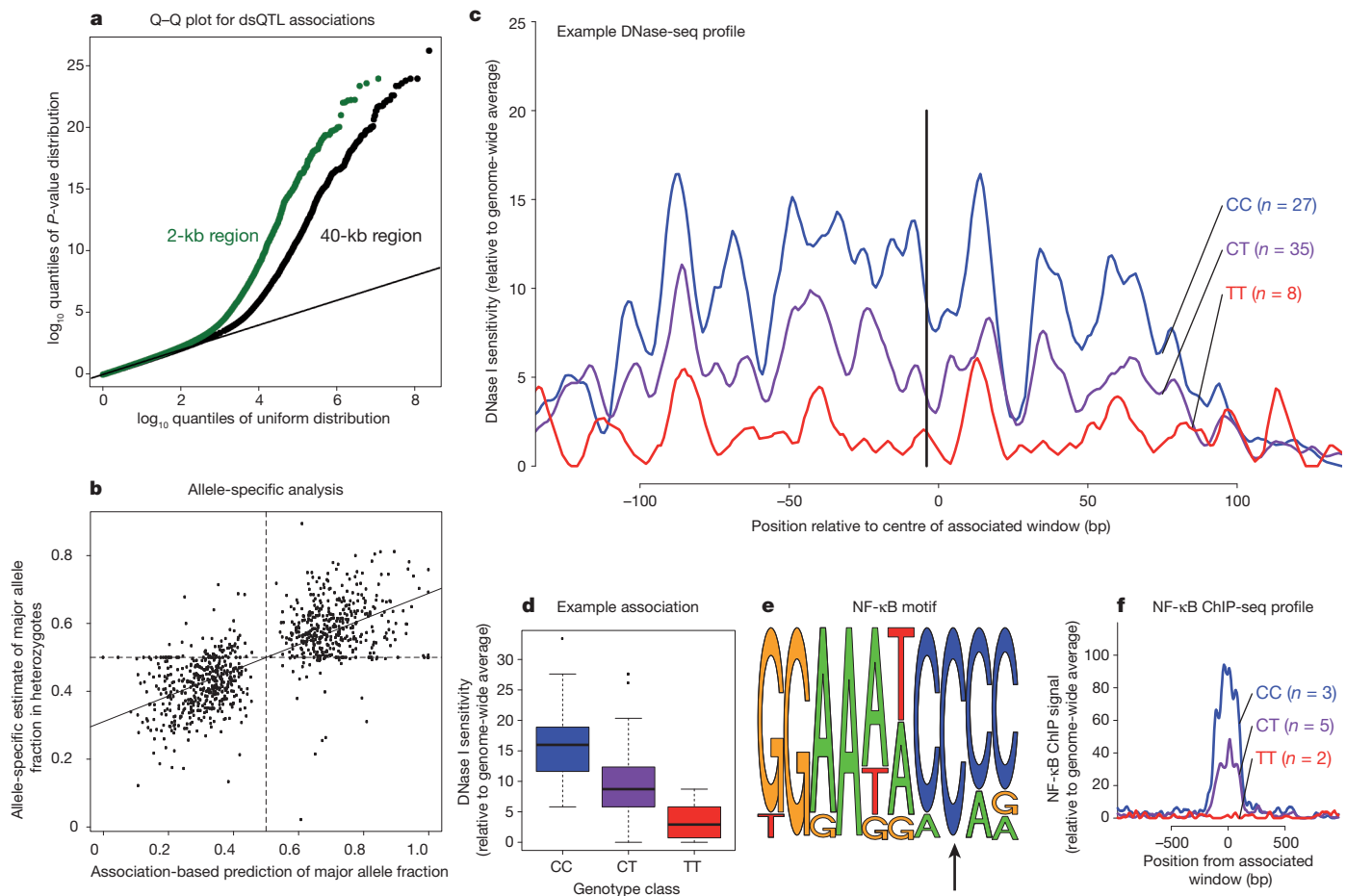*These authors contributed equally to this work.

Figure 1 | Genome-wide identification of dsQTLs and a typical example.
a, Q–Q plots for all tests of association between DNase I cut rates in 100-bp windows, and variants within 2-kb (green) and 40-kb (black) regions centred on the target DHS windows. **b**, Allele-specific analysis of dsQTLs in heterozygotes. Plotted are the predicted (*x* axis) and observed (*y* axis) fractions of reads carrying the major allele based on the genotype means. **c**, Example of a dsQTL (rs4953223). The black line indicates the position of the associated SNP. **d**, Box plot showing that rs4953223 is strongly associated with local chromatin accessibility ($P = 3 \times 10^{-13}$). **e**, The T allele, which is associated with low DNase I sensitivity, disrupts the binding motif of a previously identified NF-κB-binding site at this location[14]. **f**, NF-κB ChIP-seq data from ten individuals[7] indicates a strong effect of this SNP on NF-κB binding.

lie in chromatin regions previously predicted[18] to be functional in lymphoblastoid cell lines: 41% in predicted enhancers, 26% in promoters, and 10% in insulators, even though those chromatin states together cover only 6.7% of the genome overall (and 38% of our hypersensitive sites).

We next studied the properties of *cis*-acting variants that generated dsQTLs, with the use of a Bayesian hierarchical model that accounted for the uncertainty about which sites are causal[19] (Supplementary Information). This model obtained unbiased estimates of the average properties of causal sites even though, because of linkage disequilibrium, it was typically uncertain which site was causal for any individual dsQTL (Supplementary Information). As shown in Fig. 2b, c, most dsQTLs were generated by variants close to the target window. We estimate that 56% of the dsQTLs were due to variants that lay within the same DHSs and that 67% lay within 1 kb of the target window. dsQTLs that lay more than 1 kb from the target window were themselves significantly enriched in non-adjacent DHS windows (2.4-fold compared with matched random SNPs) and were often associated with changes in sensitivity in multiple non-adjacent DHS windows (Supplementary Fig. 15).

One intuitive mechanism for dsQTLs is that these may be caused by variants that strengthen or weaken individual transcription factor binding sites, thereby changing transcription factor affinity and local nucleosome occupancy[20–22] and hence DNase I cut rates. Consistent with this model, an aggregated plot of DNase I sensitivity at dsQTLs showed a distinct drop in chromatin accessibility around putatively

causal SNPs that was reminiscent of transcription factor binding footprints, especially in the genotypes associated with high sensitivity[15–17].

To test the importance of disruption of transcription factor binding sites as a mechanism underlying dsQTLs, we again turned to the Bayesian hierarchical model. We used the union of all published footprint locations in lymphoblastoid cell lines[16,17] and a set of footprints that we identified from the DNase-seq data reported in this study (Supplementary Methods). Analysis using the hierarchical model indicated a 3.6-fold enrichment of dsQTLs within transcription factor binding footprints ($P \ll 10^{-16}$), controlling for the overall enrichment within DHSs. In addition, the allele associated with a higher score of the position weight matrix is typically associated with higher chromatin accessibility ($P \ll 10^{-16}$), which is consistent with the expectation that higher transcription factor binding affinity leads to more open chromatin (Fig. 2d). Of the dsQTLs that fell within DNase-seq footprints tied to specific transcription factor motifs (using CENTIPEDE[17]), CCCTC binding factor (CTCF), cAMP-response element (CRE) and interferon-stimulated response element (ISRE) were the most enriched, whereas MADS box transcription enhancer factor 2 (MEF2) was significantly depleted.

To further understand the functional consequences of dsQTLs, we examined ChIP-seq data for nine transcription factors collected by the ENCODE Project in one or more lymphoblastoid cell lines[10,23]. Overall, the alleles that were associated with increased DNase I sensitivity were highly associated with increased transcription factor
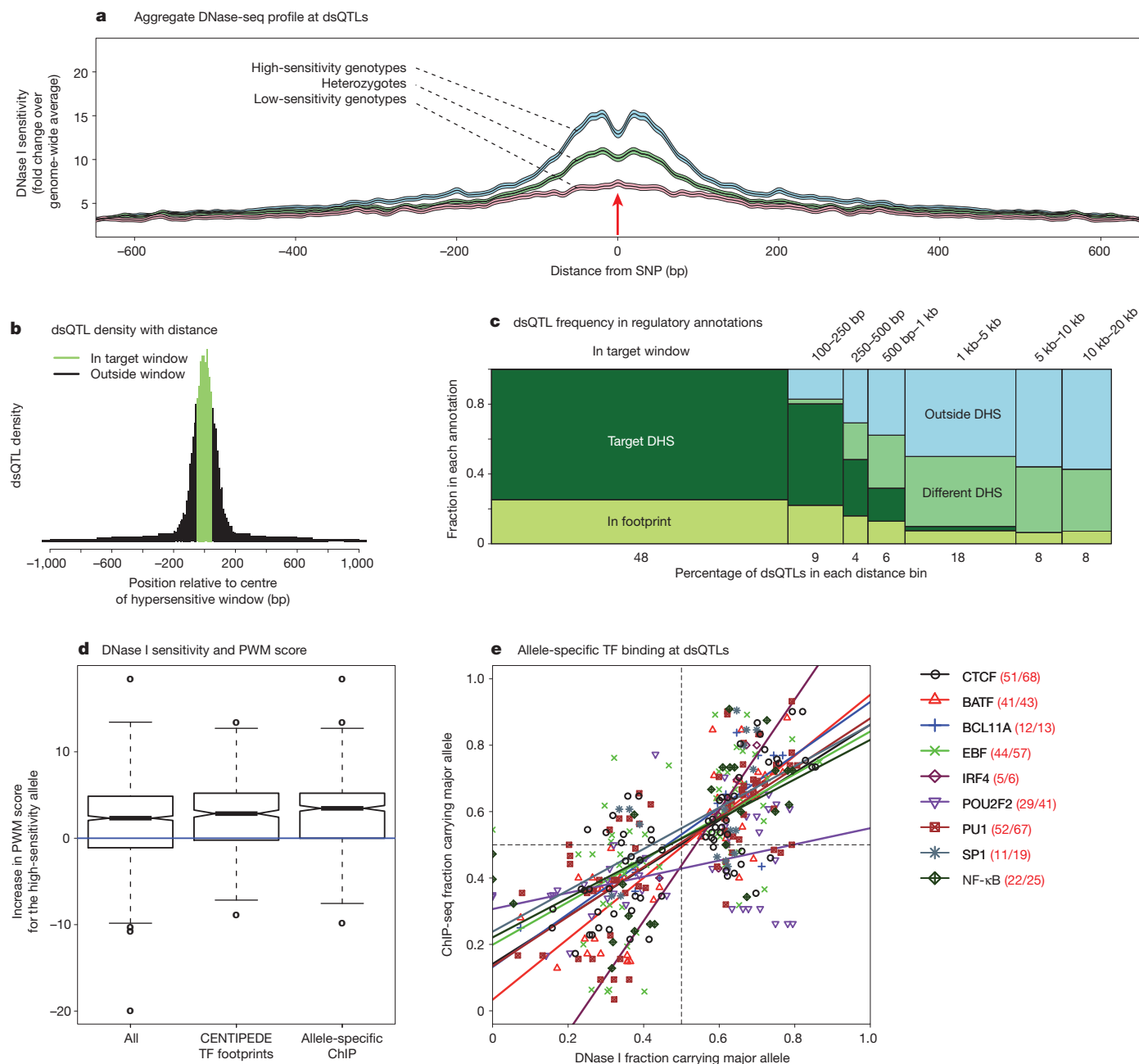
Figure 2 | Properties of dsQTLs. a, Aggregated plot of DNase I sensitivity for high-confidence dsQTLs that lie within the target DHS. Individuals were separated into the high-sensitivity (blue), heterozygote (green), and low-sensitivity (red) classes. The shading indicates the bootstrap 95% confidence intervals. **b**, The peak density of dsQTLs is very tightly focused around the target DHS window. **c**, Total fraction of *cis*-dsQTLs that fall into different categories of distance from the target window (*x* axis) and different annotations (*y* axis). The total area of each rectangle is proportional to the estimated number of dsQTLs in that category. **d**, Box plot showing distribution of position weight matrix (PWM) score differences between high-sensitivity and low-sensitivity dsQTL alleles, respectively. Notches indicate 95% confidence intervals for median. **e**, The *x* axis shows the fraction of sequence reads predicted to carry the major allele based on the DNase I genotype means; the *y* axis shows the observed fraction in ChIP-seq data. The lines show the regression fits for each factor separately; the numbers in the key show the fraction of sites that are in a concordant direction for each factor. CTCF, CCCTC binding factor; BATF, basic leucine zipper transcription factor; BCL11A, B-cell CLL/lymphoma 11A zinc-finger protein; EBF, early B-cell factor 1; IRF4, interferon regulatory factor 4; POU2F2, POU class 2 homeobox 2; PU1, proviral integration oncogene spi1; SP1, Sp1 transcription factor; NF-κB, nuclear factor of κ light polypeptide gene enhancer in B-cells 1.

binding ($P < 10^{-16}$; Fig. 2e), indicating that dsQTLs are strong predictors of changes in occupancy by a range of DNA-binding proteins.

Given that dsQTLs produce sequence-specific changes in chromatin accessibility and, frequently, changes in transcription factor binding, we speculated that a fraction of the dsQTL variants might also affect expression levels of nearby genes. We examined this by testing for associations between the most significant variant at each of the dsQTLs detected by using the 2 kb window size and expression

levels of nearby genes (that is, genes with transcription start sites (TSSs) within 100 kb) estimated by sequencing RNA from the same cell lines[8]. Using this approach, we found that 16% of dsQTL SNPs were also significantly associated with variation in expression levels of at least one nearby gene (FDR = 10%). This represents a huge enrichment over random expectation (450-fold, $P \ll 10^{-16}$; Fig. 3). One example of a joint dsQTL–eQTL is illustrated in Fig. 3a, in which a SNP disrupts an ISRE located in the first intron of the *SLFN5* gene,
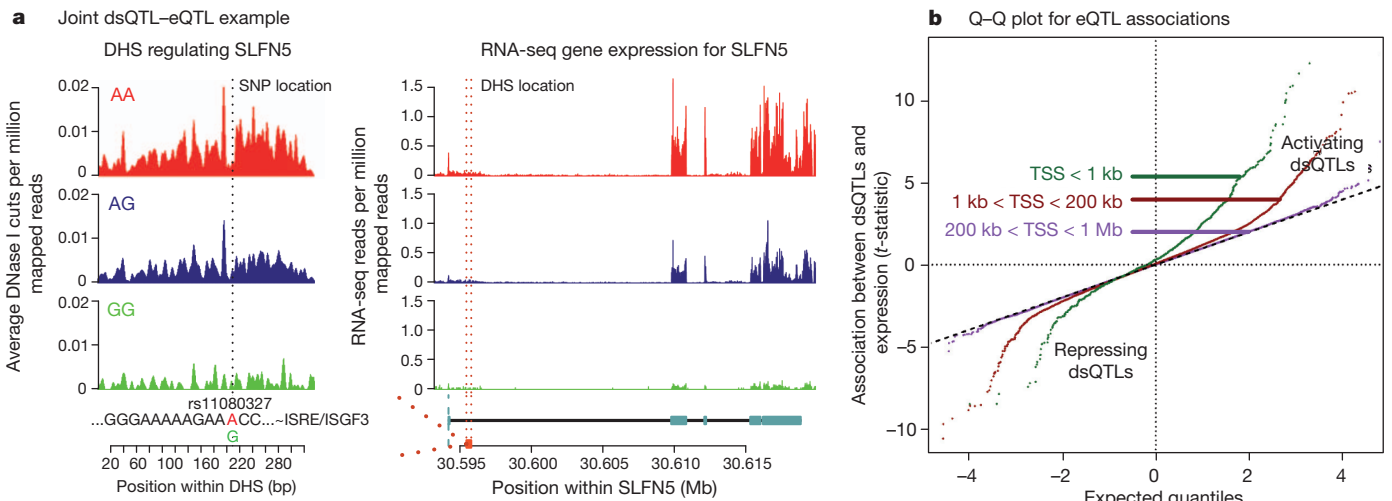
Figure 3 | Relationship between dsQTLs and eQTLs. a, Example of a dsQTL SNP that is also an eQTL for the gene *SLFN5*. The SNP disrupts an interferon-sensitive response element, thereby changing local chromatin accessibility within the first intron of *SLFN5*. Expression of *SLFN5* has been shown to be inducible by interferon α in melanoma cell lines. DNase-seq (left) and RNA-seq (right) measurements from DNase-seq and RNA-seq are plotted, stratified by genotype at the putative causal SNP. **b,** Q–Q plot of the *t*-statistic for association with gene expression changes (eQTL) of dsQTL SNPs. The sign of the eQTL *t*-statistic is with respect to the genotype that increases DNase sensitivity.

leading to both a strong dsQTL and an eQTL for *SLFN5*. Conversely, out of 1,271 eQTLs detected by using RNA-seq data from these cell lines[8], 23% of the most significant SNPs were also dsQTLs (FDR = 10%). Using the method in ref. 24 for estimating the proportion of tests in which the null hypothesis is false (while accounting for incomplete power), we estimate that 55% of the most significant eQTL SNPs are also dsQTLs and that 39% of the dsQTLs are also eQTLs. dsQTLs are therefore a major mechanism by which genetic variation may affect gene expression levels.

We observed that for most (70%) of the joint dsQTL–eQTLs, the allele that was associated with increased chromatin accessibility was also associated with increased gene expression levels (Fig. 3b). Because higher DNase I sensitivity generally correlates with higher transcription factor occupancy, this suggests that transcription factors that are bound to DHSs usually act as enhancers. CRE-box and ETS-box were the most enriched motifs among repressors and enhancers, respectively. The dsQTLs that were also eQTLs (FDR = 10%) were highly enriched around the TSSs of the target genes: for 23% of the joint dsQTL–eQTLs, the associated DHS was within 1 kb of the TSS, and for 39% it was within 10 kb (Fig. 4a). This is consistent with previous work showing strong clustering of eQTLs around TSSs[19,25,26]. Nonetheless, there was a significant signal of long-range regulation as far as 100 kb. In addition, 14% of the joint dsQTL–eQTLs were significant eQTLs for two or more genes, suggesting that some regulatory regions affect more than one gene.

We sought to identify additional factors that might influence whether a dsQTL regulates gene expression of nearby genes, while controlling for the very strong effect of distance from TSS (Fig. 4b). We observed that a dsQTL was more likely to be an eQTL for the gene with the nearest TSS (1.6-fold, $P = 3 \times 10^{-4}$) and was more likely to be an eQTL if it was located within the transcribed region of the gene (2.7-fold, $P = 2 \times 10^{-9}$). Further, a dsQTL was 2.6-fold more likely to be an eQTL if it was associated with a DHS that overlapped a DNA methylation QTL[27] ($P = 4 \times 10^{-4}$), and showed a 2.4-fold increase if the associated DHS overlapped a RNA polymerase II ChIP-seq peak[10] ($P = 4 \times 10^{-4}$). Conversely, a dsQTL was significantly less likely to be an eQTL for a gene if an active binding site for the insulator protein CTCF[17] lay between the dsQTL and the gene's TSS (2.4-fold decrease, $P = 10^{-12}$). Finally, the presence of the enhancer mark P300 (from ENCODE ChIP-seq data[28]) in the dsQTL window increased the probability that a distal dsQTL (TSS > 1.5 kb) was an eQTL (1.7-fold, $P = 10^{-5}$).

We have shown here that common genetic variants affect chromatin accessibility at thousands of hypersensitive regions across the human genome. The putative causal variants most often lie within or very near the hypersensitive regions, and frequently act by changing the binding affinity of transcription factors. Mapping of dsQTLs provides

Figure 4 | Relationship between dsQTLs and eQTLs. a, Most joint dsQTL–eQTLs lie close to the gene TSS. **b,** Effect of various factors on the log odds that a given dsQTL is also an eQTL, while controlling for the strong distance relationship observed in **a**. In annotations (1) and (2) we do not consider the direction of transcription. In annotations (6–8) ChIP-seq is measured on the dsQTL window. In annotations (4) and (6), 'meQTL' refers to a dsQTL that is also associated with methylation levels of a nearby CpG site[27] and 'Pol II' refers to the presence of an RNA polymerase II ChIP-seq peak overlapping the DHS associated with the dsQTL[23]. One of the most significant annotations in delineating the regulatory regions is defined by the presence of the CTCF insulator element, which decreases 2.4-fold the probability that a dsQTL is an eQTL. Error bars represent 95% confidence intervals.

a powerful tool for detecting potentially functional changes in a variety of different types of regulatory element, and roughly 50% of eQTLs are also dsQTLs. Furthermore, analysis of significantly associated SNPs from genome-wide association studies additionally implicates some of these dsQTLs as potentially underlying a variety of genome-wide association study hits (Supplementary Information). Changes in chromatin accessibility may be a major mechanism linking genetic variation to changes in gene regulation and, ultimately, organismal phenotypes.

METHODS SUMMARY

DNase-seq libraries were created as described previously[29], with small modifications. Each library was sequenced on at least two lanes of an Illumina GAIIx. Resulting 20-bp sequencing reads were mapped to the human genome sequence (hg18) using an algorithm that we designed specifically to eliminate mappability biases between sequence variants. We divided the genome into 100-bp windows and selected the top 5% in terms of total DNase I sensitivity. DNase I sensitivity for each individual in each window was normalized by the total number of mapped reads for that individual. For QTL mapping, the data were further rescaled within and across individuals, and we adjusted the data for an observed individual $\times$ GC interaction, as well as for the top four principal components of the DNase I sensitivity matrix. Genotypes for all available SNPs and indels were obtained from HapMap and 1,000 Genomes data and imputed where necessary[6,7,30]. We performed DNase-seq association mapping by regressing the adjusted sensitivity in each window against the genotypes at variants in a 40-kb region centred on each DHS. As validation, we used our DNase-seq reads as well as ChIP-seq reads and DNase-seq reads from ENCODE to confirm that allele-specific reads spanning heterozygous sites at dsQTLs were consistent with the association analysis. We also used RNA-seq data from the same cell lines[8] to study the links between dsQTLs and eQTLs. Finally, we explored the properties of dsQTLs that made them more or less likely to influence gene expression by fitting a logistic model on all dsQTLs, where the eQTL status of each dsQTL–eQTL test was modelled as a function of distance from the TSS and a variety of other annotations. For full details of all methods see Supplementary Information.

1. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296,** 752–755 (2002).
2. Cheung, V. G. *et al.* Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437,** 1365–1369 (2005).
3. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6,** e1000888 (2010).
4. Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6,** e1000895 (2010).
5. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).
6. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449,** 851–861 (2007).
7. Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
8. Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464,** 768–772 (2010).
9. Gaulton, K. J. *et al.* A map of open chromatin in human pancreatic islets. *Nature Genet.* **42,** 255–259 (2010).
10. Kasowski, M. *et al.* Variation in transcription factor binding among humans. *Science* **328,** 232–235 (2010).
11. McDaniell, R. *et al.* Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328,** 235–239 (2010).
12. Zheng, W., Zhao, H., Mancera, E., Steinmetz, L. M. & Snyder, M. Genetic analysis of variation in transcription factor binding in yeast. *Nature* **464,** 1187–1191 (2010).
13. Galas, D. & Schmitz, A. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5,** 3157–3170 (1978).
14. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132,** 311–322 (2008).
15. Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nature Methods* **6,** 283–289 (2009).
16. Boyle, A. P. *et al.* High-resolution genome-wide *in vivo* footprinting of diverse transcription factors in human cells. *Genome Res.* **21,** 456–464 (2011).
17. Pique-Regi, R. *et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **21,** 447–455 (2011).
18. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
19. Veyrieras, J. B. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4,** e1000214 (2008).
20. Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl Acad. Sci. USA* **107,** 22534–22539 (2010).
21. Wasson, T. & Hartemink, A. J. An ensemble model of competitive multi-factor binding of the genome. *Genome Res.* **19,** 2101–2112 (2009).
22. Raveh-Sadka, T., Levo, M. & Segal, E. Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res.* **19,** 1480–1496 (2009).
23. Myers, R. M. *et al.* A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9,** e1001046 (2011).
24. Storey, J. D., Taylor, J. E. & Siegmund, D. Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc., B* **66,** 187–205 (2004).
25. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nature Genet.* **39,** 1202–1207 (2007).
26. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nature Genet.* **39,** 1217–1224 (2007).
27. Bell, J. T. *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12,** R10 (2011).
28. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457,** 854–858 (2009).
29. Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protocols.* doi:10.1101/pdb.prot5384 (2010).
30. Guan, Y. & Stephens, M. Practical issues in imputation-based association mapping. *PLoS Genet.* **4,** e1000279 (2008).

Author Contributions A.A.P. led the data collection with assistance from S.D.L., K.M. and N.L. The data analysis was performed jointly by J.F.D. and R.P.R., with contributions from A.A.P., J.B.V., D.J.G. and J.K.Pi. G.E.C. and M.S. provided technical assistance and discussion of methods and results. The manuscript was written by J.F.D., A.A.P., R.P.R., Y.G. and J.K.Pr. The project was jointly supervised by Y.G. and J.K.Pr.

LETTER

Outgrowth of single oncogene–expressing cells from suppressive epithelial environments

Cheuk T. Leung[1] & Joan S. Brugge[1]

Tumorigenesis is a clonal evolution process that is initiated from single cells within otherwise histologically normal tissue[1]. It is unclear how single, sporadic mutant cells that have sustained oncogenic alterations evolve within a tightly regulated tissue environment. Here we investigated the effects of inducing oncogene expression in single cells in organotypic mammary acini as a model to elucidate the processes by which oncogenic alterations initiate clonal progression from organized epithelial environments. Sporadic cells induced to overexpress oncogenes that specifically perturb cell-cycle checkpoints (for example, E7 from human papilloma virus 16, and cyclin D1), deregulate Myc transcription or activate AKT signalling remained quiescent within growth-arrested acini. By contrast, single cells that overexpress ERBB2 initiated a cellular cascade involving cell translocation from the epithelial layer, as well as luminal outgrowth that is characteristic of neoplastic progression in early-stage epithelial tumours. In addition, ERBB2-mediated cell translocation to the lumen was found to depend on extracellular-regulated kinase and matrix metalloproteinase activities, and genetic alterations that perturb local cell–matrix adhesion drove cell translocation. We also provide evidence that luminal cell translocation may drive clonal selection by promoting either the death or the expansion of quiescent oncogene-expressing cells, depending on whether the pre-existing alterations allow anchorage-independent survival and growth. Our data show that the initial outgrowth of single oncogene-expressing cells from organized epithelial structures is a highly regulated process, and we propose that a cell translocation mechanism allows sporadic mutant cells to evade suppressive micro-environments and elicits clonal selection for survival and proliferative expansion outside the native niches of these cells.

The outgrowth of sporadic mutant cells within tightly regulated cellular environments is fundamental to tumour evolution. However, oncogenic alterations are usually not sufficient to predict the behaviours and fates of sporadic cell variants[2–7], particularly within complex cellular contexts such as tissues. Limitations in examining single-cell evolution within native tissues have precluded a systematic analysis. Three-dimensional (3D) organotypic cultures recapitulate many of the characteristics of cell dynamics and organization that are found in tissues, while allowing complex manipulations and long-term monitoring at single-cell resolution. MCF10A cells, a non-transformed human mammary epithelial cell line, develop into polarized, growth-arrested acinar structures containing a hollow lumen when cultured on reconstituted basement membrane (Matrigel) (Fig. 1a, b). By modelling the induction of oncogenes in single cells within 3D acini, we explored how sporadic mutant cells evolve within organized epithelial environments.

To induce oncogenes in single cells, growth-arrested (day 16) MCF10A acini that stably expressed the reverse tetracycline transactivator (rtTA) were infected with low-dose lentiviral vectors (pLT-iG) driving the tetracycline (Tet)-inducible bicistronic expression of oncogenes and fluorescent reporters, transducing <0.5% of cells (1 cell per ~10 acini) (Fig. 1b and Supplementary Fig. 2). Overexpression of Myc (also known as c-Myc), a master transcription factor that is deregulated in many cancers, or myrAKT1, which constitutively activates AKT

signalling, or perturbation of cell-cycle checkpoints by overexpressing E7 from human papilloma virus 16 (HPV16-E7) or cyclin D1[T286A] (degradation-resistant cyclin D1) was not sufficient to drive clonal outgrowth. Transduced cells remained quiescent as single cells in the acinar epithelial layer, similar to green fluorescent protein (GFP)-expressing controls (Fig. 1c, d). MCF10A cells that were induced to constitutively express these oncogenes from day 1 of 3D culture developed aberrant hyperproliferative structures (Supplementary Fig. 3), indicating that the lack of clonal expansion from the single-cell context is not due to subthreshold expression. By contrast, overexpression of ERBB2, a receptor tyrosine kinase encoded by a gene that is amplified in 30% of breast tumours[8], in sporadic cells within 3D acini effectively drove clonal outgrowth (90 ± 2% of GFP-labelled cell clusters contained multiple nuclei; mean ± s.d.) (Fig. 1c, d and Supplementary Fig. 4). Interestingly, these ERBB2-overexpressing

Figure 1 | Single-cell induction of oncogenic alteration in mammary acinar culture. a, Representative images of day 16 MCF10A acini. Shown are a low-magnification phase-contrast image (left) and acini immunostained with laminin-γ2 (LAMC2, red; centre) and GM130 (green; right). Nuclei were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) (blue). Scale bars, 50 μm (left) and 10 μm (centre and right). **b**, Schematic of single-cell lentiviral infection and doxycycline induction of oncogenes and reporters (green) in growth-arrested, polarized acini, with the basement membrane (red) depicted. d, days in culture. **c, d**, Representative images (**c**) and quantification (**d**) of clonal expansion 8 days after induction of GFP only (green) or the indicated genes with GFP (green). The nuclei of GFP-expressing clones are outlined (dashed white lines) in some cases to aid visualization. Scale bar, 10 μm (**c**). Data are presented as the mean ± s.d. from four experiments; *, statistically significant difference (two-tailed t-test) (**d**). **e**, Representative images of 3D confocal reconstructions of acini infected with vectors carrying GFP only (green) or ERBB2 and GFP (green). Scale bar, 15 μm. **f**, Acini with expanded ERBB2-overexpressing clones in the lumen were immunostained for GFP (indicating ERBB2-overexpressing cells) and LAMC2 or ITGA6, or for E-cadherin (Ecad) and ERBB2. Scale bar, 10 μm. **a–f**, The raw data are shown in Supplementary Table 1.

[1]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

clones were confined to the lumen (Fig. 1e), resembling the histological feature of early-stage carcinoma-*in-situ* breast tumours[9]. Immuno-staining for laminin-γ2, α6-integrin and E-cadherin indicated that gross acinar structures remained intact and that luminal translocation is not associated with epithelial–mesenchymal transition (Fig. 1f). No invasive structures were observed (data not shown). Overexpression of ERBB2 in single cells within 3D acini derived from primary mouse mammary epithelial cells or a highly polarized ovarian cell line, MCAS, also led to luminal localization of the transduced cells (Supplementary Fig. 5). The unique ability of ERBB2 to initiate clonal expansion and the striking pattern of luminal filling suggest that the outgrowth of sporadic mutant cells from organized epithelial structures is tightly regulated.

Long-term (56–85 h) time-lapse confocal microscopy indicated that single GFP-expressing control cells remained quiescent within growth-arrested acini (in 5 of 5 acini). By contrast, single ERBB2-overexpressing cells dissociated from the epithelial layer, showed increased protrusive activity and translocated to the lumen (in 6 of 7 acini) (Fig. 2a, Supplementary Movies 1 and 2 and Supplementary Fig. 6). Blocking cell proliferation with aphidicolin did not block translocation (Fig. 2b–d), further indicating that translocation is independent of proliferation. Taken together, these data reveal an initial luminal cell-translocation step in ERBB2-mediated clonal outgrowth.
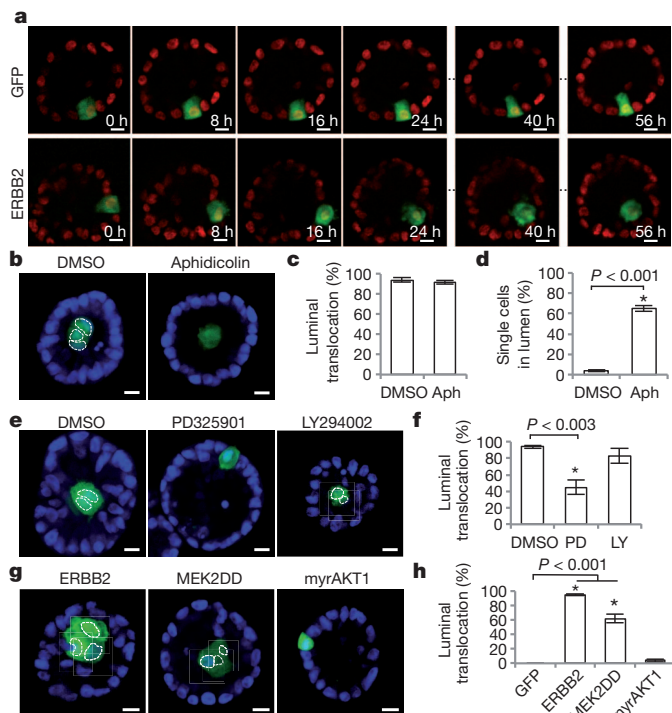
We next examined the involvement of two major pathways downstream of ERBB2, the mitogen-activated protein kinase (MAPK, or ERK) pathway and the phosphatidylinositol-3-OH kinase (PI(3)K) pathway, in luminal translocation. MAPK kinase (MEK) inhibition, but not PI(3)K–mammalian target of rapamycin (mTOR) inhibition, greatly reduced ERBB2-mediated translocation (Fig. 2e, f). Moreover, constitutively active MEK (MEK2DD), but not constitutively active AKT (myrAKT1), was sufficient to drive cell translocation to the lumen (Fig. 2g, h). We cannot rule out possible ERK-independent ERBB2 downstream mechanisms, because perturbing ERK activity only partially inhibited or promoted ERBB2-mediated luminal translocation. Recent studies reported that Ras^{V12}- and v-*Src*-transformed Madin–Darby canine kidney (MDCK) cells adjacent to normal neighbours are extruded from monolayer cultures by an ERK-dependent mechanism, although Raf-driven ERK activation is not sufficient to drive extrusion[10,11]. Another study showed that universal expression of activated Raf induces overall cell motility in MCF10A acini[12]. Our results suggest that this conserved role of ERK in cell motility is also crucial for ERBB2-mediated single-cell luminal translocation in 3D cultures.

We found that the ERK-regulated transcription factor ETS1 can drive cell translocation to the lumen (Fig. 3a, b). ETS1 transactivates proteinases, including matrix metalloproteinases (MMPs)[13], which have been implicated in tissue remodelling. Interestingly, MMP inhibition significantly blocked ERBB2-, MEK2DD- and ETS1-induced luminal

Figure 2 | Mechanisms of cell translocation to the lumen. a, Confocal sections from a time series of acini (nuclei, red) containing single ERBB2-overexpressing cells (green) or GFP-only expressing cells (green) captured starting from approximately 24 h after oncogene induction (indicated as 0 h). **b–d**, Representative images and quantification of the translocation of ERBB2-overexpressing cells (**b**, **c**) and of single ERBB2-overexpressing cells in the lumen (**d**) of day 16 acini treated with 10 μM aphidicolin (Aph) or dimethylsulphoxide (DMSO) for 8 days. Nuclei were counterstained with DAPI (blue). The nuclei of GFP-expressing clones are outlined (dashed white lines) in some cases to aid visualization. **e**, **f**, Representatives images (**e**) and quantification (**f**) of ERBB2-overexpressing cell translocation in 3D acini treated with a MEK inhibitor (1 μM PD325901 (PD)), a PI(3)K–mTOR inhibitor (20 μM LY294002 (LY)) or DMSO for 8 days. **g**, **h**, Representative images (**g**) and quantification (**h**) of the luminal translocation of MEK2DD- or myrAKT1-expressing cells from growth-arrested acini. **a**, **b**, **e**, **g**, Scale bars, 10 μm. **c**, **d**, **f**, **h**, Data are presented as the mean ± s.d. from four experiments; *, statistically significant difference (two-tailed *t*-test). **a–h**, The raw data are shown in Supplementary Table 2.

Figure 3 | Luminal translocation and clonal outgrowth from a suppressive epithelial environment. a–c, Representative images (**a**), quantification of luminal translocation (**b**) and single cells in the lumen (**c**) 8 days after induction of ETS1 or MT1-MMP expression (green) in 3D acini. Nuclei were counterstained with DAPI (blue). **d–f**, Representative images of ERBB2 cells/clones (green) (**d**) and quantification of translocation (**e**) and single cells in the epithelial layer (**f**) of acini treated with the broad-spectrum MMP inhibitors GM6001 or Batimastat, MMP2/MMP9 inhibitor III (MMP inh. III) or DMSO for 8 days. The nuclei of GFP-expressing clones are outlined (dashed white lines) in some cases to aid visualization. **g**, **h**, Representative images (**g**) and quantification (**h**) of translocation of cells in which talin 1 expression was knocked down (talin-1 KD) (red) in 3D acini. Immunoblotting (**h**, right) indicates efficient knockdown of talin-1 expression. Similar results were obtained with a different talin-1 knockdown construct (data not shown). NS, non-silencing construct. **a**, **d**, **g**, Scale bars, 10 μm. **b**, **c**, **e**, **f**, **h**, Data are presented as the mean ± s.d. from three to four experiments; *, statistically significant difference (two-tailed *t*-test). **a–h**, The raw data are shown in Supplementary Table 3.

translocation (Fig. 3d, e and Supplementary Fig. 7). Although the identity of the MMPs involved is unclear, these data indicate that MMP activity is important for cell translocation. We overexpressed MT1-MMP (also known as MMP14) to examine whether MMP activity can promote translocation. MT1-MMP was chosen because its membrane localization and broad substrate specificity make it an attractive tool for modulating local MMP activity. Single-cell MT1-MMP overexpression in MCF10A acini was sufficient to drive cell translocation to the lumen. Notably, neither ETS1-induced translocation nor MT1-MMP-induced translocation drove clonal expansion (Fig. 3a–c), and both were independent of ERK signalling (Supplementary Fig. 8). Taken together, these results identify specific proliferation-independent pathways that promote cell translocation to the lumen (Fig. 3a, b).

We proposed that perturbation of local cell–matrix adhesion by MMPs could trigger cell translocation. Consistent with this idea, MCF10A cells overexpressing ERBB2, but not Myc, myrAKT1 or GFP, showed impaired adhesion to Matrigel-coated plates (Supplementary Fig. 9). In addition, we observed compromised basement membrane adjacent to single ERBB2-overexpressing cells, but not adjacent to Myc-, myrAKT1- or GFP-overexpressing cells, in 3D acini (Supplementary Fig. 9). Moreover, weakening the cell–matrix adhesion strength by knocking down expression of the gene encoding talin 1 (ref. 14), a protein that links integrins to actin filaments and localizes to the basal cell membrane in 3D acini (Supplementary Fig. 10), was sufficient to induce translocation (Fig. 3g, h).

Intriguingly, the ERBB2-overexpressing cells that stayed in the epithelial layer as a result of MMP inhibition were unable to proliferate (Fig. 3f and Supplementary Fig. 11). MMP inhibition did not affect the proliferation of ERBB2-overexpressing cells in monolayer cultures, and the treatment of acini with an MMP inhibitor 4 days after single-cell ERBB2 induction, when most induced cells had already translocated, did not affect luminal outgrowth (Supplementary Fig. 12), suggesting that MMP activities are required specifically for the initial translocation step but not for proliferation.

Cell displacement has been proposed as a mechanism for removing aberrant cells from epithelia[11,15–17]. Our data predict that cell displacement by translocation may also facilitate the outgrowth of sporadic mutant cells. Using MT1-MMP as a tool to drive cell translocation, and myrAKT1- and Myc-overexpressing cells as models, we examined the outcome of forced translocation of quiescent oncogene-expressing cells. Single cells within MCF10A acini that stably carried Tet-inducible MT1-MMP-IRES-GFP (pLT-MT1-MMP-iGSP) or IRES-GFP (pLT-iGSP) vectors were infected with another lentiviral vector encoding Tet-inducible myrAKT1-IRES-mCherry or Myc-IRES-mCherry as well as constitutive rtTA expression (pLT-myrAKT1-iCSA or pLT-Myc-iCSA). Only these transduced single cells contain all of the Tet-inducible components required to drive doxycycline-dependent expression of myrAKT1 (or Myc), MT1-MMP and the two fluorescent reporters (Supplementary Fig. 13). This combinatorial inducible approach overcomes the size limitations on efficient virus packaging and allows flexible, multiplex genetic manipulations in single cells.

Forced translocation of cells co-expressing myrAKT1-mCherry and MT1-MMP-GFP, but not cells co-expressing Myc-mCherry and MT1-MMP-GFP (or mCherry and MT1-MMP-GFP), led to luminal expansion (Fig. 4a, b). Cells co-expressing myrAKT1-mCherry and MT1-MMP-GFP that failed to translocate remained as single cells in the epithelial layer (99 ± 2% of GFP-labelled cells; mean ± s.d.), indicating that translocation, but not simply MT1-MMP and myrAKT1 co-expression, facilitates clonal outgrowth. Indeed, MT1-MMP did not increase the colony formation of myrAKT1-expressing MCF10A cells in soft agar (Supplementary Fig. 14).

Both translocated Myc-overexpressing cells and myrAKT1-overexpressing cells re-entered the cell cycle (Fig. 4c), but only Myc-overexpressing cells showed increased apoptosis (Fig. 4d, e). These results suggest that the anti-apoptotic activity of myrAKT1 contributes to support clonal expansion in the matrix-deprived

lumen. Consistent with these observations, 42 ± 8% and 29 ± 12% (mean ± s.d.) of translocated ETS1-overexpressing and MT1-MMP-overexpressing single cells, respectively, also underwent apoptosis (Fig. 4e). Although we could not directly trace these cells, the apoptosis of single translocated ETS1-, MT1-MMP- or Myc-overexpressing cells suggests that the clonal lineage would probably be eliminated.

This dichotomous fate of Myc-overexpressing cells and myrAKT1-overexpressing cells suggests that translocation is not sufficient for outgrowth but instead might unleash cells from their suppressive epithelial environment. We tested whether perturbing the epithelial organization of 3D acini allows the expansion of quiescent mutant cells. Knockdown of *CTNND1* (which encodes p120-catenin), to destabilize the epithelial cell–cell junctions[18] in preformed acini, greatly reduced staining for β-catenin and E-cadherin at cell junctions, without disrupting gross acinar structures (Supplementary Fig. 15). In contrast to the distinct outcomes of forced translocation, both Myc-overexpressing cells and myrAKT1-overexpressing cells, but not GFP-expressing cells, underwent clonal expansion on p120-catenin
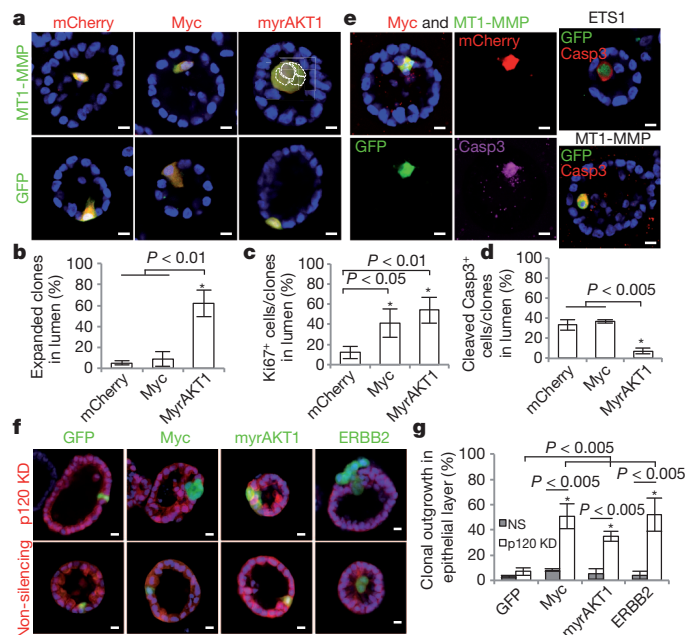

Figure 4 | Cell translocation elicits clonal selection or outgrowth of quiescent mutant cells. a–d, Single cells in day 16 MCF10A pLT-MT1-MMP-iGSP- (top) or pLT-iGSP (IRES-GFP)-infected acini (bottom) were infected with pLT-Myc-iCSA, pLT-myrAKT1-iCSA or pLT-iCSA (IRES-mCherry) to inducibly drive oncogene and reporter co-expression. Nuclei were counterstained with DAPI (blue). Representative images (**a**) and quantification (**b**) of Myc and mCherry, myrAKT1 and mCherry, or mCherry cells/clones co-expressing either MT1-MMP and GFP or GFP only (yellow) 12 days after doxycycline induction. Nuclei of clones co-expressing MT1-MMP and myrAKT1 are outlined (white dashed line). Quantification of proliferation (**c**) and apoptosis (**d**) in translocated cells/clones. **e**, Acini with cells overexpressing ETS1 or MT1-MMP or with cells co-expressing Myc and MT1-MMP were immunostained with antibody specific for cleaved caspase 3 (Casp3) 8 days after induction. **f, g**, Single cells within preformed, day 16 MCF10A pTRIPZ-p120KD- or pTRIPZ-NS (non-silencing)-expressing acini were infected with pLT-Myc-iG, pLT-myrAKT1-iG, pLT-ERBB2-iG or pLT-iG.. Acini were induced with doxycycline to drive expression of a p120-catenin knockdown (p120 KD; red) or non-silencing (NS; red) short hairpin RNA in all cells with co-expression of Myc, myrAKT1, ERBB2 or GFP (all in green) in single cells. Acini were induced with doxycycline for 48 h to express the p120-catenin KD or non-silencing short hairpin RNA before infection with pLT-ERBB2-iG. Representative images (**f**) and quantification (**g**) of expanded clones in the epithelial layer 8 days after induction. **a, e, f**, Scale bars, 10 μm. **b, c, d, g**, Data are presented as the mean ± s.d. from four experiments; *, statistically significant difference (two-tailed *t*-test). **a–g**, The raw data are shown in Supplementary Table 4.

downregulation (Fig. 4f, g). These data suggest that the epithelial organization mediated by p120-catenin and cadherin junctions is crucial for suppressing oncogene-induced proliferation in mature acini. Notably, single ERBB2-overexpressing cells in preformed acini that were subjected to p120-catenin knockdown also proliferated but did not translocate (Fig. 4f, g), suggesting that intact epithelial organization has a role in supporting cell translocation to the lumen.

These findings highlight the suppressive influence of a mature epithelial environment on sporadic mutant cells and raise the question of whether the expression of oncogenes in neighbouring cells would abrogate this proliferative suppression. Interestingly, induction of Myc or myrAKT1 in all cells of growth-arrested acini did not drive cell proliferation or disrupt acinar structure (Supplementary Fig. 16). This observation is consistent with a previous finding that tamoxifen-induced Myc activation in growth-arrested MCF10A acini does not drive proliferation[19]. Taken together, these results further demonstrate the strong suppressive control of organized epithelium, because an oncogene such as *Myc* or *myrAKT1* is not sufficient to abrogate this suppressive effect.

We used 3D organotypic cultures to model the genetic and tissue architectural context in which sporadic oncogene-expressing cells arise in the early stages of human breast tumorigenesis, and we demonstrated that the initial outgrowth of these sporadic mutant cells within organized epithelial environments is highly regulated. We showed that although perturbation of a suppressive epithelial environment allows a general expansion of quiescent mutant cells, a different process, involving cell translocation to the lumen, allows mutant cells to evade the suppressive epithelial environment and drives selection for survival and expansion in the matrix-deprived lumen (a model is shown in Supplementary Fig. 1). Our data also highlight the suppressive influence of organized epithelial structures on pre-neoplastic cell expansion.

The displacement of cell variants from epithelia[10,11,15,17,20–22] and stem cell compartments[16] has been widely observed in diverse organ systems and species. Previous studies have shown that the extrusion of apoptotic cells from epithelial monolayers involves a force-dependent expulsion process from neighbouring cells[17]. We show that cell translocation to the lumen in 3D cultures is induced by ERK and MMP activities that are intrinsic to the translocating cells and involves local perturbations of cell–matrix adhesion.

The migration of cells from specialized niches and micro-environments underlies cell differentiation and tissue morphogenesis during development and regeneration[23,24]. Our data suggest that a similar spatial cell translocation within tissue compartments may also have a role in tumorigenesis by eliciting clonal selection. Oncogenic alterations have been observed in cells in otherwise histologically normal tissue of healthy individuals[3,5], and the disruption of tissue organization has been associated with tumour progression. Our findings raise the possibility that mechanisms such as cell translocation or compromised tissue integrity may initiate neoplastic progression from these dormant mutant cells.

METHODS SUMMARY

3D Matrigel culture and virus infection. MCF10A cells were set up in 3D cultures on basement membrane in 8-well chamber slides (BD Biosciences) or coverglass-bottom 8-chamber slides (MatTek) as previously described[25], with 4,500–5,000 cells in assay medium (DMEM/F12 supplemented with 2% donor horse serum, 5 ng ml^{-1} epidermal growth factor (EGF), 10 µg ml^{-1} insulin, 1 ng ml^{-1} cholera toxin, 100 µg ml^{-1} hydrocortisone, 50 U ml^{-1} penicillin, 50 µg ml^{-1} streptomycin and 2% Matrigel). The medium was replaced at 4-day intervals. On day 16, 3D cultures were infected with the indicated lentiviruses diluted in assay medium without EGF or Matrigel and incubated for 6–8 h at 37 °C. Virus dosages were adjusted to infect less than 1 cell per 10 acini, to achieve sporadic single-cell infection. The virus was removed, and the chamber wells were rinsed with 500 µl PBS, which was replaced with normal 3D assay medium without Matrigel. Doxycycline (1 µg ml^{-1}) was added on the following day, together with drug treatment or vehicle control as indicated. Drugs and vehicles were replenished at 2-day intervals, and the complete medium was changed at 4-day intervals. Acinar

structures were analysed 8 days after induction with doxycycline or at longer time intervals as indicated.

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

1. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194,** 23–28 (1976).
2. Dolberg, D. S. & Bissell, M. J. Inability of Rous sarcoma virus to cause sarcomas in the avian embryo. *Nature* **309,** 552–556 (1984).
3. Holst, C. R. et al. Methylation of p16^{INK4a} promoters occurs *in vivo* in histologically normal human mammary epithelia. *Cancer Res.* **63,** 1596–1601 (2003).
4. Illmensee, K. & Mintz, B. Totipotency and normal differentiation of single teratocarcinoma cells cloned by injection into blastocysts. *Proc. Natl Acad. Sci. USA* **73,** 549–553 (1976).
5. Jonason, A. S. et al. Frequent clones of p53-mutated keratinocytes in normal human skin. *Proc. Natl Acad. Sci. USA* **93,** 14025–14029 (1996).
6. Michaloglou, C. et al. BRAF^{E600}-associated senescence-like cell cycle arrest of human naevi. *Nature* **436,** 720–724 (2005).
7. Weaver, V. M. et al. Reversion of the malignant phenotype of human breast cells in three-dimensional culture and in vivo by integrin blocking antibodies. *J. Cell Biol.* **137,** 231–245 (1997).
8. Slamon, D. J. et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235,** 177–182 (1987).
9. Hebner, C., Weaver, V. M. & Debnath, J. Modeling morphogenesis and oncogenesis in three-dimensional breast epithelial cultures. *Annu. Rev. Pathol.* **3,** 313–339 (2008).
10. Kajita, M. et al. Interaction with surrounding normal epithelial cells influences signalling pathways and behaviour of Src-transformed cells. *J. Cell Sci.* **123,** 171–180 (2010).
11. Hogan, C. et al. Characterization of the interface between normal and transformed epithelial cells. *Nature Cell Biol.* **11,** 460–467 (2009).
12. Pearson, G. W. & Hunter, T. Real-time imaging reveals that noninvasive mammary epithelial acini can contain motile cells. *J. Cell Biol.* **179,** 1555–1567 (2007).
13. Lincoln, D. W. II & Bove, K. The transcription factor Ets-1 in breast cancer. *Front. Biosci.* **10,** 506–511 (2005).
14. Bershadsky, A. D., Balaban, N. Q. & Geiger, B. Adhesion-dependent cell mechanosensitivity. *Annu. Rev. Cell Dev. Biol.* **19,** 677–695 (2003).
15. Gibson, M. C. & Perrimon, N. Extrusion and death of DPP/BMP-compromised epithelial cells in the developing *Drosophila* wing. *Science* **307,** 1785–1789 (2005).
16. Li, X., Han, Y. & Xi, R. Polycomb group genes *Psc* and *Su(z)2* restrict follicle stem cell self-renewal and extrusion by controlling canonical and noncanonical Wnt signaling. *Genes Dev.* **24,** 933–946 (2010).
17. Rosenblatt, J., Raff, M. C. & Cramer, L. P. An epithelial cell destined for apoptosis signals its neighbors to extrude it by an actin- and myosin-dependent mechanism. *Curr. Biol.* **11,** 1847–1857 (2001).
18. Davis, M. A., Ireton, R. C. & Reynolds, A. B. A core function for p120-catenin in cadherin turnover. *J. Cell Biol.* **163,** 525–534 (2003).
19. Partanen, J. I., Nieminen, A. I., Makela, T. P. & Klefstrom, J. Suppression of oncogenic properties of c-Myc by LKB1-controlled epithelial organization. *Proc. Natl Acad. Sci. USA* **104,** 14694–14699 (2007).
20. Shen, J. & Dahmann, C. Extrusion of cells with inappropriate Dpp signaling from *Drosophila* wing disc epithelia. *Science* **307,** 1789–1790 (2005).
21. Bullen, T. F. et al. Characterization of epithelial cell shedding from human small intestine. *Lab. Invest.* **86,** 1052–1063 (2006).
22. Marshall, T. W., Lloyd, I. E., Delalande, J. M., Nathke, I. & Rosenblatt, J. The tumor suppressor adenomatous polyposis coli controls the direction a cell extrudes from an epithelium. *Mol. Biol. Cell* **22,** 3962–3970 (2011).
23. Brawley, C. & Matunis, E. Regeneration of male germline stem cells by spermatogonial dedifferentiation *in vivo*. *Science* **304,** 1331–1334 (2004).
24. Aman, A. & Piotrowski, T. Cell migration during morphogenesis. *Dev. Biol.* **341,** 20–33 (2010).
25. Debnath, J. et al. The role of apoptosis in creating and maintaining luminal space within normal and oncogene-expressing mammary acini. *Cell* **111,** 29–40 (2002).

Author Contributions C.T.L. conceived the study, performed the experiments, analysed the data and drafted the manuscript. J.S.B. supervised the study and edited the manuscript.

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

METHODS

Cell culture. MCF10A cells were cultured as described previously[25]. MCAS cells were cultured in 1:1 Medium 199:MCDB 105 medium with 2 mM L-glutamine, 10% heat-inactivated FCS and 100 U ml^{-1} penicillin and streptomycin. Primary mouse mammary epithelial cells (MECs) were cultured in DMEM/F12 with 5% FBS, 5 μg ml^{-1} insulin, 1 mg ml^{-1} hydrocortisone, 3 ng ml^{-1} EGF and 100 U ml^{-1} penicillin and streptomycin.

Chemicals. Doxycycline was used at 1 μg ml^{-1}. The following doses of inhibitors were used: 20 μM LY294002, 1 μg ml^{-1} PD325901, 10 μM GM6001, 5 μM Batimastat, 5 μg ml^{-1} MMP2/MMP9 inhibitor III and10 μM aphidicolin.

Retroviral vectors. The vectors pBABE-hygro-rtTA and pBABE-puro-rtTA were constructed by subcloning the coding sequence of rtTA from pTet-On-Advanced (Clontech) into the pBABE-hygro or pBABE-puro retroviral vectors, respectively. The vector pBABE-neo-MT1-MMP was constructed by amplifying the coding sequence of MT1-MMP (Invitrogen) using PCR and subcloning it into pBABE-neo. The vector pBABE-puro-H2B-mCherry was constructed by amplifying the coding sequence of amino acids 1–125 of the human H2B gene using PCR, fusing it to the amino terminus of mCherry (Clontech) and subcloning it into pBABE-puro.

Lentiviral vectors. The pLT-iG lentiviral inducible vector was constructed by replacing the CAG promoter of the pCSC-SP-PW lentiviral vector (Addgene, plasmid 12335) with the tetracycline response element (TRE) from pTre-Tight (Clontech) and inserting a downstream IRES-GFP cassette from pIRES2-GFP (Clontech). The coding sequences from the following sources were either directly subcloned or PCR-amplified and then subcloned between the TRE and IRES-GFP cassette of pLT-iG. ERBB2 (pBABE-puro-ERBB2), myrAKT1 (pBABE-puro-myrAKT1), MEK2DD (pBABE-puro-MEK2DD) and HPV16-E7 (pLCNX-E7) were obtained as previously described[25,26]. CyclinD^{T286A} (pcDNA-CyclinD1 HA T286A, plasmid 11182) and Myc (pcDNA3-Myc, plasmid 16011) were obtained from Addgene, and ETS1 (pCR4/Ets1) and MT1-MMP (pCMV-Sport6/MT1-MMP) were purchased from Invitrogen. The pLT-iGSP vector sets were constructed by subcloning an SV40-puro cassette downstream of IRES-GFP. pLT-iGSA vector sets were subcloned by replacing SV40-puro of pLT-iGSP with SV40-rtTA. pLT-iCSA vector sets were subcloned by replacing GFP from pLT-iGSA with mCherry. Inducible p120 (pTRIPZ-p120), talin-1 (pTRIPZ-TLN1) and non-silencing (pTRIPZ-NS) short hairpin RNA knockdown constructs were purchased from Open Biosystems. The hairpin sequences are underlined: pTRIPZ-p120KD, 5′-TGCTGTTGACAGTGAGCGACCTGTGGAGCTCTCAAGAATATAGTGAAGCCACAGATGTATATTCTTGAGAGCTCCACAGGCTGCCTACTGCCTCGGA-3′; and pTRIPZ-TLN1 KD, 5′-TGCTGTTGACAGTGAGCGCGCGCAGAATGCCATCAAGAAATAGTGAAGCCACAGATGTATTTCTTGATGGCATTCTGCGCATGCCTACTGCCTCGGA-3′.

Virus production. Retroviruses and lentiviruses were produced by co-transfecting the corresponding viral vectors with the packaging vectors pCL-Ampho (retroviruses, IMGENEX) or psPAX2 and pMD2.G (lentiviruses, Addgene 12260 and 12259) into 293T cells with FuGENE HD (Roche). Virus-containing supernatants were collected on days 2 and 3 following transfection and were stored at −80 °C.

Primary MEC isolation. Primary MECs were isolated from 10–12-week-old virgin FVB females. The number four glands from two to four animals were dissected, and the lymph nodes were removed, minced into small pieces and digested with 2 mg ml^{-1} collagenase and 100 μg ml^{-1} hyaluronidase in DMEM/F12 medium for 1 h at 37 °C with shaking. At the end of the digestion, the samples were treated with DNase I for 5 min and centrifuged for 10 min at 600g. Organoid pellets were washed five times with 30 ml DMEM/F12 medium in a 50 ml conical tube by pulse spinning at 600× RCF. Organoids were then rinsed once with PBS and digested in 0.05% trypsin for 5 min at 37 °C. The trypsin was neutralized with growth medium, and the dissociated cells were pelleted and the supernatant removed. The cells were then seeded directly for 3D cultures.

Statistical analysis. Acini were fixed in 4% paraformaldehyde for 20 min at room temperature and counterstained for nuclei with 4′,6-diamidino-2-phenylindole (DAPI). Isolated infected acinar structures were scored for expansion and for spatial location of the infected cells/clones with respect to the structures. Reporter-labelled cell clusters containing more than one nucleus based on DAPI counterstaining were scored as expanded clones. Cells with cytoplasm directly adjacent to the surrounding matrix were scored as cells in the epithelial layer. Cells that were separated from the surrounding matrix by at least one cell layer were scored as cells in the lumen. At least three individual experiments with 50–200 acini counted in each were analysed per assay. Standard deviations were calculated, and two-tailed t-tests were performed.

Immunofluorescence. Immunofluorescent analyses were performed as previously described[25]. Primary antibodies were incubated overnight. The following antibody dilutions were used: α$_6$-integrin (1:100, Millipore), laminin-γ2 (1:100, BD Biosciences), GM130 (1:100, Cell Signaling Technology), E-cadherin (1:100, Cell Signaling Technology), β-catenin (1:100, Cell Signaling Technology), GFP (1:100, Invitrogen), activated caspase 3 (1:100, Cell Signaling Technology), Ki67 (1:100, Invitrogen), ERBB2 (1:100, Cell Signaling Technology), ZO1 (1:100, Invitrogen) and talin 1 (1:100, Abcam). Alexa-Fluor-conjugated secondary antibodies (Invitrogen) were used at 1:200–1:100 for 2 h at room temperature. Nuclei were counterstained with DAPI. Images were acquired using Nikon C1 or Nikon A1 confocal microscopes.

Immunoblotting. Inducible overexpression or knockdown of the target genes was confirmed by immunoblotting after induction for 2 days in 3D assay medium with 1 μg ml^{-1} doxycycline on monolayer culture cells. The following antibody dilutions were used: HPV16-E7 (1:100, Zymed), ERBB2 (1:1,000, Cell Signaling Technology), pan-AKT (1:500, Cell Signaling Technology), cyclin D1 (1:500, Cell Signaling Technology), Myc (1:1,000, Cell Signaling Technology), p120-catenin (1:1,000, BD Biosciences) and talin 1 (1:1,000, Cell Signaling Technology). β-Tubulin was used as a loading control.

Live cell imaging. MCF10A cells stably expressing H2B-mCherry and rtTA were set up for 3D culture in 8-well Lab-Tek II chambered coverglasses (MatTek) as indicated above. Live imaging was performed with a custom-built spinning disk confocal system based on a Ti inverted microscope (Nikon) with a CSU-X1 confocal head (Yokogawa), an ORCA-AG cooled charge-coupled-device camera (Hamamatsu) and a 37 °C, 5% CO$_2$ environmental control chamber. Images were acquired starting approximately 24–36 h after doxycycline induction. Image stacks of complete acini were captured with a 20× multi-immersion objective every 15–30 min for a duration of 56–84 h using MetaMorph software and were analysed with MetaMorph or Imaris software.

Soft-agar colony-formation assay. Cells (50,000) were seeded in 0.4% soft agar in normal MCF10A growth medium on a layer of 0.5% soft agar in normal growth medium. Cultures were fed every 6 days with 1 ml of 0.4% soft agar in growth medium.

Growth curve. ERBB2-overexpressing MCF10A cells or empty vector control cells were seeded at 100,000 cells per well in 6-well plates for 48 h in 3D assay medium with 25 μM GM6001 or dimethylsulphoxide (DMSO). The cell doubling rate was averaged from three individual experiments.

Adhesion assay. The expression of ERBB2, Myc, myrAKT1 or GFP was induced in MCF10A monolayer cultures for 72 h, and the cells were then trypsinized, plated at 50,000 cells per well in 24-well Matrigel-coated plates and incubated at 37 °C for 1 h. The plates were washed three times with PBS, and the cells were fixed with paraformaldehyde. Attached cells were counterstained with DAPI. The average number of attached cells from four arbitrary 20× fields in the centre of the wells was calculated from three individual experiments.

26. Gunawardane, R. N. et al. Novel role for PDEF in epithelial cell migration and invasion. Cancer Res. 65, 11572–11580 (2005).

LETTER

Recent contributions of glaciers and ice caps to sea level rise

Thomas Jacob[1]†, John Wahr[1], W. Tad Pfeffer[2,3] & Sean Swenson[4]

Glaciers and ice caps (GICs) are important contributors to present-day global mean sea level rise[1–4]. Most previous global mass balance estimates for GICs rely on extrapolation of sparse mass balance measurements[1,2,4] representing only a small fraction of the GIC area, leaving their overall contribution to sea level rise unclear. Here we show that GICs, excluding the Greenland and Antarctic peripheral GICs, lost mass at a rate of 148 ± 30 Gt yr^{-1} from January 2003 to December 2010, contributing 0.41 ± 0.08 mm yr^{-1} to sea level rise. Our results are based on a global, simultaneous inversion of monthly GRACE-derived satellite gravity fields, from which we calculate the mass change over all ice-covered regions greater in area than 100 km^2. The GIC rate for 2003–2010 is about 30 per cent smaller than the previous mass balance estimate that most closely matches our study period[2]. The high mountains of Asia, in particular, show a mass loss of only 4 ± 20 Gt yr^{-1} for 2003–2010, compared with 47–55 Gt yr^{-1} in previously published estimates[2,5]. For completeness, we also estimate that the Greenland and Antarctic ice sheets, including their peripheral GICs, contributed 1.06 ± 0.19 mm yr^{-1} to sea level rise over the same time period. The total contribution to sea level rise from all ice-covered regions is thus 1.48 ± 0.26 mm yr^{-1}, which agrees well with independent estimates of sea level rise originating from land ice loss and other terrestrial sources[6].

Interpolation of sparse mass balance measurements on selected glaciers is usually used to estimate global GIC mass balance[1,2,4]. Models are also used[3,7], but these depend on the quality of input climate data and include simplified glacial processes. Excluding Greenland and Antarctic peripheral GICs (PGICs), GICs have variously been reported to have contributed 0.43–0.51 mm yr^{-1} to sea level rise (SLR) during 1961–2004[3,7,8], 0.77 mm yr^{-1} during 2001–2004[8], 1.12 mm yr^{-1} during 2001–2005[1] and 0.95 mm yr^{-1} during 2002–2006[2].

The Gravity Recovery and Climate Experiment (GRACE) satellite mission[9] has provided monthly, global gravity field solutions since 2002, allowing users to calculate mass variations at the Earth's surface[10]. GRACE has been used to monitor the mass balance of selected GIC regions[11–14] that show large ice mass loss, as well as of Antarctica and Greenland[15].

Here we present a GRACE solution that details individual mass balance results for every region of Earth with large ice-covered areas. The main focus of this paper is on GICs, excluding Antarctic and Greenland PGICs. For completeness, however, we also include results for the Antarctic and Greenland ice sheets with their PGICs. GRACE does not have the resolution to separate the Greenland and Antarctic ice sheets from their PGICs. All results are computed for the same 8-yr time period (2003–2010).

To determine losses of individual GIC regions, we cover each region with one or more 'mascons' (small, arbitrarily defined regions of Earth) and fit mass values for each mascon (ref. 16 and Supplementary Information) to the GRACE gravity fields, after correcting for

hydrology and for glacial isostatic adjustment (GIA) computed using the ICE-5G deglaciation model. We use 94 monthly GRACE solutions from the University of Texas Center for Space Research, spanning January 2003 to December 2010. The GIA corrections do not include the effects of post-Little Ice Age (LIA) isostatic rebound, which we separately evaluate and remove. All above contributions and their effects on the GRACE solutions are discussed in Supplementary Information.

Figure 1 shows mascons for all ice-covered regions, constructed from the Digital Chart of the World[17] and the Circum-Arctic Map of Permafrost and Ground-Ice Conditions[18]. Each ice-covered region is chosen as a single mascon, or as the union of several non-overlapping mascons. We group 175 mascons into 20 regions. Geographically isolated regions with glacierized areas less than 100 km^2 in area are excluded. Because GRACE detects total mass change, its results for an ice-covered region are independent of the glacierized surface area (Supplementary Information).

Mass balance rates for each region are shown in Table 1 (see Supplementary Information for details on the computation of the rates and uncertainties). We note that Table 1 includes a few positive rates, but none are significantly different from zero. We also performed an inversion with GRACE fields from the GFZ German Research Centre for Geosciences and obtained results that agreed with those from the Center for Space Research (Table 1) to within 5% for each region.

The results in Table 1 are in general agreement with previous GRACE studies for the large mass loss regions of the Canadian Arctic[12] and Patagonia[11], as well as for the Greenland and Antarctic ice sheets with

Figure 1 | Mascons for the ice-covered regions considered here. Each coloured region represents a single mascon. Numbers correspond to regions shown in Table 1. Regions containing more than one mascon are outlined with a dashed line.

[1]Department of Physics and Cooperative Institute for Environmental Studies, University of Colorado at Boulder, Boulder, Colorado 80309, USA. [2]Institute of Arctic and Alpine Research, University of Colorado at Boulder, Boulder, Colorado 80309, USA. [3]Department of Civil, Environmental, and Architectural Engineering, University of Colorado at Boulder, Boulder, Colorado 80309, USA. [4]National Center for Atmospheric Research, Boulder, Colorado 80305, USA. †Present address: Bureau de Recherches Géologiques et Minières, Orléans 45060, France.

Table 1 | Inverted 2003–2010 mass balance rates

Region	Rate (Gt yr^{-1})
1. Iceland	−11 ± 2
2. Svalbard	−3 ± 2
3. Franz Josef Land	0 ± 2
4. Novaya Zemlya	−4 ± 2
5. Severnaya Zemlya	−1 ± 2
6. Siberia and Kamchatka	2 ± 10
7. Altai	3 ± 6
8. High Mountain Asia	−4 ± 20
8a. Tianshan	−5 ± 6
8b. Pamirs and Kunlun Shan	−1 ± 5
8c. Himalaya and Karakoram	−5 ± 6
8d. Tibet and Qilian Shan	7 ± 7
9. Caucasus	1 ± 3
10. Alps	−2 ± 3
11. Scandinavia	3 ± 5
12. Alaska	−46 ± 7
13. Northwest America excl. Alaska	5 ± 8
14. Baffin Island	−33 ± 5
15. Ellesmere, Axel Heiberg and Devon Islands	−34 ± 6
16. South America excl. Patagonia	−6 ± 12
17. Patagonia	−23 ± 9
18. New Zealand	2 ± 3
19. Greenland ice sheet + PGICs	−222 ± 9
20. Antarctica ice sheet + PGICs	−165 ± 72
Total	−536 ± 93
GICs excl. Greenland and Antarctica PGICs	−148 ± 30
Antarctica + Greenland ice sheet and PGICs	−384 ± 71
Total contribution to SLR	1.48 ± 0.26 mm yr^{-1}
SLR due to GICs excl. Greenland and Antarctica PGICs	0.41 ± 0.08 mm yr^{-1}
SLR due to Antarctica + Greenland ice sheet and PGICs	1.06 ± 0.19 mm yr^{-1}

Uncertainties are given at the 95% (2σ) confidence level.

their PGICs[19]. Our results for Alaska also show considerable mass loss, although our mass loss rate is smaller than some previously published GRACE-derived rates that used shorter and earlier GRACE data spans (Supplementary Information). The global GIC mass balance, excluding Greenland and Antarctic PGICs, is −148 ± 30 Gt yr^{-1}, contributing 0.41 ± 0.08 mm yr^{-1} to SLR.

Mass balance time series for all GIC regions are shown in Fig. 2. The seasonal and interannual variabilities evident in these time series have contributions from ice and snow variability on the glaciers, as well as from imperfectly modelled hydrological signals in adjacent regions and from random GRACE observational errors. Interannual variability can affect rates determined over short time intervals. Figure 2 and Supplementary Table 2 show that there was considerable interannual variability during 2003–2010 for some of the regions, especially High Mountain Asia (HMA). The HMA results in Supplementary Table 2 show that this variability induces large swings in the trend solutions when it is fitted to subsets of the entire time period. These results suggest that care should be taken in extending the 2003–2010 results presented in this paper to longer time periods.

For comparison with studies in which PGICs are included with GICs, we upscale our GIC-alone rate to obtain a GIC rate that includes PGIC, based on ref. 3 (Supplementary Information). The result is that GICs including PGICs lost mass at a rate of 229 ± 82 Gt yr^{-1} (0.63 ± 0.23 mm yr^{-1} SLR), and that the combined ice sheets without their PGICs lost mass at 303 ± 100 Gt yr^{-1} (0.84 ± 0.28 mm yr^{-1} SLR). Although no other study encompasses the same time span, published non-GRACE estimates for GICs plus PGICs are larger: 0.98 ± 0.19 mm yr^{-1} over 2001–2004[8], 1.41 ± 0.20 mm yr^{-1} over 2001–2005[1] and 0.765 mm yr^{-1} (no uncertainty given) over 2006–2010[20]. These differences could be due to the small number of mass balance measurements those estimates must rely on, combined with uncertain regional glacier extents. In addition, there are indications from more recent non-GRACE measurements that the GIC mass loss rate decreased markedly beginning in 2005[20].

Our results for HMA disagree significantly with previous studies. A recent GRACE-based study[5] over 2002–2009 yields significantly
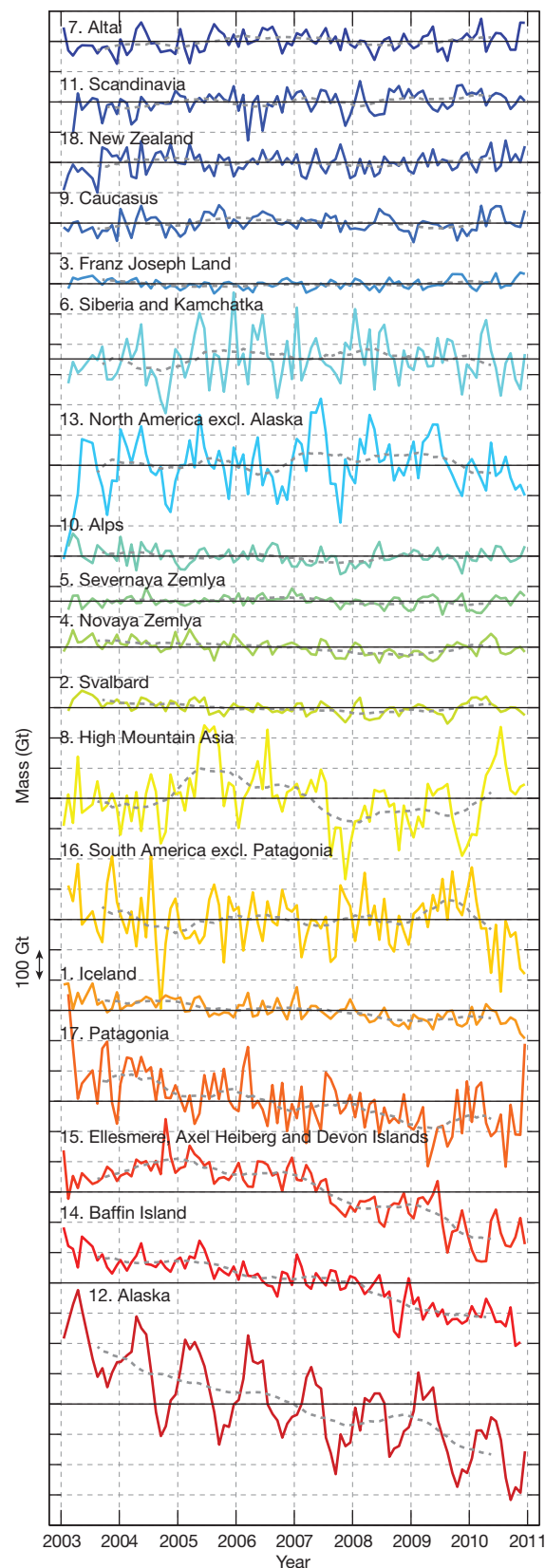

Figure 2 | Mass change during 2003–2010 for all GIC regions shown in Fig. 1 and Table 1. The black horizontal lines run through the averages of the time series. The grey lines represent 13-month-window, low-pass-filtered versions of the data. Time series are shifted for legibility. Modelled contributions from GIA, LIA and hydrology have been removed.

larger mass loss for HMA than does ours; we explain why the result of ref. 5 may be flawed in Supplementary Information. Conventional mass balance methods have been used to estimate a 2002–2006 rate of $-55 \, \text{Gt yr}^{-1}$ for this entire region[2], with $-29 \, \text{Gt yr}^{-1}$ over the eastern Himalayas alone, by contrast with our HMA estimate, of $-4 \pm 20 \, \text{Gt yr}^{-1}$ (Table 1). We show results for the four subregions of HMA (Fig. 3) in Table 1.

This difference prompts us to examine this region in more detail. GRACE mass trends show considerable mass loss across the plains of northern India, Pakistan and Bangladesh, centred south of the glaciers and at low elevations (Fig. 3a, b). Some of the edges of this mass loss region seem to extend over adjacent mountainous areas to the north, but much of that, particularly above north-central India, is leakage of the plains signal caused by the 350-km Gaussian smoothing function used to generate the figure. The plains signal has previously been identified as groundwater loss[16,21]. To minimize leakage in the HMA GIC estimates, additional mascons are chosen to cover the plains (Fig. 3a), the sum of which gives an average 2003–2010 water loss rate of $35 \, \text{Gt yr}^{-1}$. Our plains results are consistent with the results of refs 16 and 21, which span shorter time periods.

The lack of notable mass loss over glacierized regions is consistent with our HMA mascon solutions that indicate relatively modest losses (Table 1). We simulate what the ice loss rates predicted by ref. 2 would

look like in the GRACE results. We use those rates to construct synthetic gravity fields and process them using the same methods applied to the GRACE data, to generate the trend map shown in Fig. 3c. It is apparent that an ice loss of this order would appear in the GRACE map as a large mass loss signal centred over the eastern Himalayas, far larger in amplitude and extent than the GRACE results in that region (compare Fig. 3b with Fig. 3c).

It is reasonable to wonder whether a tectonic process could be causing a positive signal in the glacierized region that offsets a large negative glacier signal in HMA. To see what this positive rate would have to look like, we remove the simulated gravity field (based on ref. 2) from the GRACE data and show the resulting difference map in Fig. 3d. If the ice loss estimate were correct, the tectonic process would be causing an anomalous mass increase over the Himalayas of $\sim 3 \, \text{cm yr}^{-1}$ equivalent water thickness, equivalent to $\sim 1 \, \text{cm yr}^{-1}$ of uncompensated crustal uplift. Although we cannot categorically rule out such a possibility, it seems unlikely. Global Positioning System and levelling observations in this region indicate long-term uplift rates as large as 0.5–$0.7 \, \text{cm yr}^{-1}$ in some places[22,23]. But it is highly probable that any broad-scale tectonic uplift would be isostatically compensated by an increasing mass deficiency at depth, with little net effect on gravity[24] and, consequently, no significant contribution to the GRACE results. The effects of compensation are evident in the static gravity field. Supplementary Fig. 4

Figure 3 | HMA mass balance determination. a, Topographic map overlaid with the HMA mascons (crosses) and India plain mascons (dots); the dashed lines delimit the four HMA subregions (labelled as in Table 1). **b,** GRACE mass rate corrected for hydrology and GIA and smoothed with a 350-km Gaussian smoothing function, overlaid with the HMA mascons. w.e., water equivalent. **c,** Synthetic GRACE rates that would be caused by a total mass loss of $55 \, \text{Gt yr}^{-1}$ over HMA mascons, with $29 \, \text{Gt yr}^{-1}$ over the eastern Himalayas, after ref. 2. **d,** The difference between **b** and **c**.

shows the free-air gravity field, computed using a 350-km Gaussian smoothing function (used to generate Fig. 3) applied to the EGM96 mean global gravity field[25]. The topography leaves no apparent signature on the static gravity field at these scales, indicating near-perfect compensation.

For a solid-Earth process to affect GRACE significantly, it must be largely isostatically uncompensated, which for these broad spatial scales would require characteristic timescales of the same order or less than the mantle's viscoelastic relaxation times (several hundred to a few thousand years). One possible such process might be the ongoing viscoelastic response of the Earth to past glacial unloading. We have investigated this effect, as well as possible contributions from erosion, and find that neither is likely to be important (Supplementary Information).

Another possible explanation for the lack of a large GRACE HMA signal is that most of the glacier melt water might be sinking into the ground before it has a chance to leave the glaciated region, thus causing GRACE to show little net mass change. Some groundwater recharge undoubtedly does occur, but it seems unlikely that such cancellation would be this complete. Much of HMA, for example, is permafrost, so local storage capacity is small (see the Circum-Arctic Map of Permafrost and Ground-Ice Conditions; http://nsidc.org/fgdc/maps/ipa_browse.html). Therefore, although there would be surface melt, the frozen ground would inhibit local recharge and there would be little ability to store the melt water locally. How far the water might have to travel before finding recharge pathways, we do not know. It is true that some rivers originating in portions of HMA do not reach the sea. Most notable are the Amu Darya and Syr Darya, which historically feed the Aral Sea but have been diverted for irrigation. Any fraction of that diverted water that ends up recharging aquifers will not directly contribute to SLR. However, the irrigation areas lie well outside our HMA mascons, and so even if there is notable recharge it is unlikely to affect the HMA mascon solutions significantly.

Our emphasis here is on GICs; the Greenland and Antarctic ice sheets have previously been well studied with GRACE[15]. But for comparison with non-GRACE global estimates, we combine our GIC results with our estimates for Greenland plus Antarctica to obtain a total SLR contribution from all ice-covered regions of $1.48 \pm 0.26 \, \text{mm yr}^{-1}$ during 2003–2010. Within the uncertainties, this value compares favourably with the estimate of $1.8 \pm 0.5 \, \text{mm yr}^{-1}$ for 2006 from ref 4. However, there are regional differences between these and prior results, which need further study and reconciliation.

SLR from the addition of new water can be determined from GRACE alone as well as by subtracting Argo steric heights from altimetric SLR measurements[6]. The most recent new-water SLR estimate, comparing the two methods, is $1.3 \pm 0.6 \, \text{mm yr}^{-1}$ for 2005–2010[6], which agrees with our total ice-covered SLR value to within the uncertainties. The difference, $0.2 \pm 0.6 \, \text{mm yr}^{-1}$, could represent an increase in land water storage outside ice-covered regions, but we note that it is not significantly different from zero.

METHODS SUMMARY

GRACE solutions consist of spherical harmonic (Stokes) coefficients and are used to determine month-to-month variations in Earth's mass distribution[9,10]. We use monthly values of C_{20} (the zonal, degree-2 spherical harmonic coefficient of the geopotential) from satellite laser ranging[26], and include degree-one terms[27].

To determine mass variability for each mascon, we find the set of Stokes coefficients produced by a unit mass distributed uniformly across that mascon. We fit these sets of Stokes coefficients, simultaneously, to the GRACE Stokes coefficients, to obtain monthly mass values for each mascon. This method is similar to previously published mascon methods[28], though here we fit to Stokes coefficients rather than to raw satellite measurements and we do not impose smoothness constraints. To determine the optimal shape and number of mascons in a region, we construct a sensitivity kernel for several possible configurations, and choose the configuration that optimizes that kernel and minimizes the GRACE trend residuals (Supplementary Fig. 1c).

The average of two land surface models is used to correct for hydrology, and the model differences are used to estimate uncertainties (Supplementary Information).

LIA loading corrections have been previously derived for Alaska[13] and Patagonia[29], and equal 7 and $9 \, \text{Gt yr}^{-1}$, respectively. These numbers are subtracted from our Alaska and Patagonia inversions. For other GIC regions, where LIA characteristics are not well known, we estimate an upper bound for the correction by constructing a GIA model that tends to maximize the positive LIA gravity trend. Of all the additional GIC regions, only HMA has a predicted LIA correction that reaches $1 \, \text{Gt yr}^{-1}$. There, the model suggests we remove $5 \, \text{Gt yr}^{-1}$ from our inverted result. But because the LIA correction in this region is likely to be an overestimate (Supplementary Information), our preferred result splits the difference (Supplementary Table 1), and we use that difference to augment the total HMA uncertainty.

1. Cogley, J. G. Geodetic and direct mass-balance measurements: comparison and joint analysis. *Ann. Glaciol.* **50,** 96–100 (2009).
2. Dyurgerov, M. B. Reanalysis of glacier changes: from the IGY to the IPY, 1960–2008. *Data Glaciol. Stud.* **108,** 1–116 (2010).
3. Hock, R., de Woul, M., Radic, V. & Dyurgerov, M. Mountain glaciers and ice caps around Antarctica make a large sea-level rise contribution. *Geophys. Res. Lett.* **36,** L07501 (2009).
4. Meier, M. F. *et al.* Glaciers dominate eustatic sea-level rise in the 21st century. *Science* **317,** 1064–1067 (2007).
5. Matsuo, K. & Heki, K. Time-variable ice loss in Asian high mountains from satellite gravimetry. *Earth Planet. Sci. Lett.* **290,** 30–36 (2010).
6. Willis, J. K., Chambers, D. P., Kuo, C. Y. & Shum, C. K. Global sea level rise, recent progress and challenges for the decade to come. *Oceanography (Wash. DC)* **23,** 26–35 (2010).
7. Hirabayashi, Y., Doll, P. & Kanae, S. Global-scale modeling of glacier mass balances for water resources assessments: glacier mass changes between 1948 and 2006. *J. Hydrol. (Amst.)* **390,** 245–256 (2010).
8. Kaser, G., Cogley, J. G., Dyurgerov, M. B., Meier, M. F. & Ohmura, A. Mass balance of glaciers and ice caps: consensus estimates for 1961–2004. *Geophys. Res. Lett.* **33,** L19501 (2006).
9. Tapley, B. D., Bettadpur, S., Watkins, M. & Reigber, C. The gravity recovery and climate experiment: mission overview and early results. *Geophys. Res. Lett.* **31,** L09607 (2004).
10. Wahr, J., Swenson, S., Zlotnicki, V. & Velicogna, I. Time-variable gravity from GRACE: first results. *Geophys. Res. Lett.* **31,** L11501 (2004).
11. Chen, J. L., Wilson, C. R., Tapley, B. D., Blankenship, D. D. & Ivins, E. R. Patagonia icefield melting observed by gravity recovery and climate experiment (GRACE). *Geophys. Res. Lett.* **34,** L22501 (2007).
12. Gardner, A. S. *et al.* Sharply increased mass loss from glaciers and ice caps in the Canadian Arctic Archipelago. *Nature* **473,** 357–360 (2011).
13. Luthcke, S. B., Arendt, A. A., Rowlands, D. D., McCarthy, J. J. & Larsen, C. F. Recent glacier mass changes in the Gulf of Alaska region from GRACE mascon solutions. *J. Glaciol.* **54,** 767–777 (2008).
14. Riva, R. E. M., Bamber, J. L., Lavallee, D. A. & Wouters, B. Sea-level fingerprint of continental water and ice mass change from GRACE. *Geophys. Res. Lett.* **37,** L19605 (2010).
15. Rignot, E., Velicogna, I., van den Broeke, M. R., Monaghan, A. & Lenaerts, J. Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise. *Geophys. Res. Lett.* **38,** L05503 (2011).
16. Tiwari, V. M., Wahr, J. & Swenson, S. Dwindling groundwater resources in northern India, from satellite gravity observations. *Geophys. Res. Lett.* **36,** L18401 (2009).
17. Raup, B. H., Kieffer, H. H., Hare, J. M. & Kargel, J. S. Generation of data acquisition requests for the ASTER satellite instrument for monitoring a globally distributed target: glaciers. *IEEE Trans. Geosci. Remote Sens.* **38,** 1105–1112 (2000).
18. Brown, J., Ferrians, O. J., Heginbottom, J. A. & Melnikov, E. S. *Circum-Arctic Map of Permafrost and Ground-Ice Conditions.* National Snow and Ice Data Center/World Data Center for Glaciology (1998, revised, February 2001).
19. Velicogna, I. Increasing rates of ice mass loss from the Greenland and Antarctic ice sheets revealed by GRACE. *Geophys. Res. Lett.* **36,** L19503 (2009).
20. Cogley, J. G. in *Future Climates of the World* (eds Henderson-Sellers, A. & McGuffie, K.) 189–214 (Elsevier, 2012).
21. Rodell, M., Velicogna, I. & Famiglietti, J. S. Satellite-based estimates of groundwater depletion in India. *Nature* **460,** 999–1002 (2009).
22. Bettinelli, P. *et al.* Plate motion of India and interseismic strain in the Nepal Himalaya from GPS and DORIS measurements. *J. Geod.* **80,** 567–589 (2006).
23. Jackson, M. & Bilham, R. Constraints on Himalayan deformation inferred from vertical velocity-fields in Nepal and Tibet. *J. Geophys. Res. Solid Earth* **99,** 13897–13912 (1994).
24. Zhong, S. J. & Zuber, M. T. Crustal compensation during mountain-building. *Geophys. Res. Lett.* **27,** 3009–3012 (2000).
25. Lemoine, F. *et al. The Development of the Joint NASA GSFC and NIMA Geopotential Model EGM96.* NASA Goddard Space Flight Center (1998).
26. Cheng, M. K. & Tapley, B. D. Variations in the Earth's oblateness during the past 28 years. *J. Geophys. Res.* **109,** B09402 (2004).
27. Swenson, S., Chambers, D. & Wahr, J. Estimating geocenter variations from a combination of GRACE and ocean model output. *J. Geophys. Res. Solid Earth* **113,** B08410 (2008).

28. Rowlands, D. D. *et al.* Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements. *Geophys. Res. Lett.* **32,** L04310 (2005).
29. Ivins, E. R. & James, T. S. Bedrock response to Llanquihue Holocene and present-day glaciation in southernmost South America. *Geophys. Res. Lett.* **31,** L24613 (2004).

Author Contributions T.J. and J.W. developed the study and wrote the paper. W.T.P. and S.S. discussed, commented on and improved the manuscript. S.S. provided the CLM4 hydrology model output.

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

LETTER

Structural basis for recognition of H3K56–acetylated histone H3–H4 by the chaperone Rtt106

Dan Su[1]*, Qi Hu[1]*, Qing Li[1]*, James R. Thompson[2], Gaofeng Cui[1], Ahmed Fazly[1], Brian A. Davies[1], Maria Victoria Botuyan[1], Zhiguo Zhang[1] & Georges Mer[1]

Dynamic variations in the structure of chromatin influence virtually all DNA-related processes in eukaryotes and are controlled in part by post-translational modifications of histones[1–3]. One such modification, the acetylation of lysine 56 (H3K56ac) in the amino-terminal α-helix (αN) of histone H3, has been implicated in the regulation of nucleosome assembly during DNA replication and repair, and nucleosome disassembly during gene transcription[4–10]. In *Saccharomyces cerevisiae*, the histone chaperone Rtt106 contributes to the deposition of newly synthesized H3K56ac-carrying H3–H4 complex on replicating DNA[5], but it is unclear how Rtt106 binds H3–H4 and specifically recognizes H3K56ac as there is no apparent acetylated lysine reader domain in Rtt106. Here, we show that two domains of Rtt106 are involved in a combinatorial recognition of H3–H4. An N-terminal domain homodimerizes and interacts with H3–H4 independently of acetylation while a double pleckstrin-homology (PH) domain binds the K56-containing region of H3. Affinity is markedly enhanced upon acetylation of K56, an effect that is probably due to increased conformational entropy of the αN helix of H3. Our data support a mode of interaction where the N-terminal homodimeric domain of Rtt106 intercalates between the two H3–H4 components of the $(H3–H4)_2$ tetramer while two double PH domains in the Rtt106 dimer interact with each of the two H3K56ac sites in $(H3–H4)_2$. We show that the Rtt106–$(H3–H4)_2$ interaction is important for gene silencing and the DNA damage response.

To understand the mode of action of Rtt106, we characterized its three-dimensional (3D) structure and probed its association with histones. Rtt106 is modular (Supplementary Fig. 1), consisting of a homodimeric N-terminal domain (Rtt106DD; residues 1–42) and a double PH domain (Rtt106PH; residues 68–301) linked via a disordered region (residues 43–67) (Fig. 1a, b and Supplementary Figs 2 and 3). The 3D structure of Rtt106DD, determined using NMR spectroscopy, shows a previously undiscovered fold with each protomer adopting a V-shaped conformation consisting of two α-helices separated by a trans-proline residue (Fig. 1a, Supplementary Fig. 3 and Supplementary Table 1). The first and second α-helices of each protomer interact with the second and first α-helices of the other protomer, respectively, through extensive hydrophobic contacts (Fig. 1a). The 3D structure of Rtt106PH, determined by X-ray crystallography to a resolution of 1.4 Å (Fig. 1b and Supplementary Table 2), reveals similarity to the structure of Pob3, a protein thought to have a role in histone deposition[11] (Supplementary Fig. 4).

We performed isothermal titration calorimetry (ITC) measurements to probe the interaction of Rtt106DD–Rtt106PH (residues 1–301) with the histone H3–H4 complex reconstituted using non-acetylated H3 and H3 with an acetyl-lysine analogue chemically installed at position 56 (ref. 12). Rtt106 binds both non-acetylated and K56-acetylated H3–H4. However, acetylation results in enhanced affinity (Fig. 1c). In first approximation, the biphasic curve for the interaction of Rtt106 with non-acetylated H3–H4 in Fig. 1c can be

interpreted as two concurrent binding isotherms in a two-site binding model with dissociation constants $K_{d1} = 0.4 \pm 0.2\,\mu M$ and $K_{d2} = 1.5 \pm 0.2\,\mu M$. Agreeing well with two binding sites, one from Rtt106DD and the other from Rtt106PH, Rtt106DD (residues 1–42) alone binds H3–H4 in an acetylation-independent manner. ITC data are consistent with a one-site binding model with a K_d of $0.6 \pm 0.1\,\mu M$ (Fig. 1d), close to K_{d1} above. The reaction stoichiometry indicates that Rtt106DD, a dimer, binds two H3–H4 molecules, most likely in the form of an $(H3–H4)_2$ tetramer (Fig. 1d). Binding of Rtt106DD with H3–H4 was also demonstrated by NMR spectroscopy (Supplementary Fig. 5).

To interpret the biphasic ITC thermograms for the interaction of Rtt106 with K56-acetylated H3–H4, we considered a two-site binding model with an activation term that accounts for the effect of acetylation. The first dissociation constant $K_{d1ac} = 0.8 \pm 0.4\,\mu M$ is similar to K_{d1} (Fig. 1c and Supplementary Fig. 6). The second dissociation constant, K_{d2ac}, is $0.08 \pm 0.06\,\mu M$. In comparison to K_{d2}, the apparent gain in affinity for Rtt106PH is approximately 15-20-fold. These results indicate that the acetylated region of H3 is recognized by Rtt106PH but not by Rtt106DD. This was verified by monitoring NMR chemical shift perturbations in ^1H–^{15}N heteronuclear single-quantum coherence (HSQC) correlation spectra of ^{15}N-labelled Rtt106 (residues 1–67) and Rtt106PH upon titration with a non-labelled H3K56ac peptide (residues 51–61) (Fig. 1b). No binding to Rtt106 (residues 1–67) was observed upon addition of up to 15 molar excess of peptide (data not shown). In contrast, the H3K56ac peptide does specifically bind Rtt106PH as demonstrated by marked chemical shift changes ($\Delta\delta \geq 0.2$ p.p.m.) for 38 backbone amide signals of Rtt106PH (Fig. 1e and Supplementary Fig. 7). The affected residues are mapped to the second PH domain, specifically at the interface of the carboxy-terminal α-helix (α5), two underlying β-sheets and the flexible tether connecting the two PH domains (Fig. 1b). Noticeably, this region differs from the binding sites of previously characterized PH domains[13] (Supplementary Fig. 8). The K_d for the Rtt106PH–H3K56ac peptide interaction is 0.9 ± 0.1 mM (Supplementary Fig. 9). The markedly reduced affinity compared to that obtained for Rtt106DD–Rtt106PH (residues 1–301) and full-length H3–H4 is expected because the complete interaction is combinatorial with Rtt106DD and Rtt106PH both contacting H3–H4. Rtt106PH by itself has limited selectivity towards acetylation with only a twofold decrease in affinity for the non-acetylated H3K56 peptide ($K_d = 1.9 \pm 0.4$ mM).

How then does K56 acetylation specifically enhance the affinity of H3–H4 for Rtt106? Recent biophysical studies have highlighted the structural heterogeneity of H3 αN helix in the context of $(H3–H4)_2$, suggesting that the structure and dynamics of αN could be readily altered by post-translational modifications[14]. K56, being the C-terminal residue of αN, contributes favourably to helical stability via charge interaction with the α-helix dipole. Therefore, in neutralizing the charge of K56, acetylation may increase the conformational entropy

[1]Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, Minnesota 55905, USA. [2]Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, Minnesota 55905, USA.
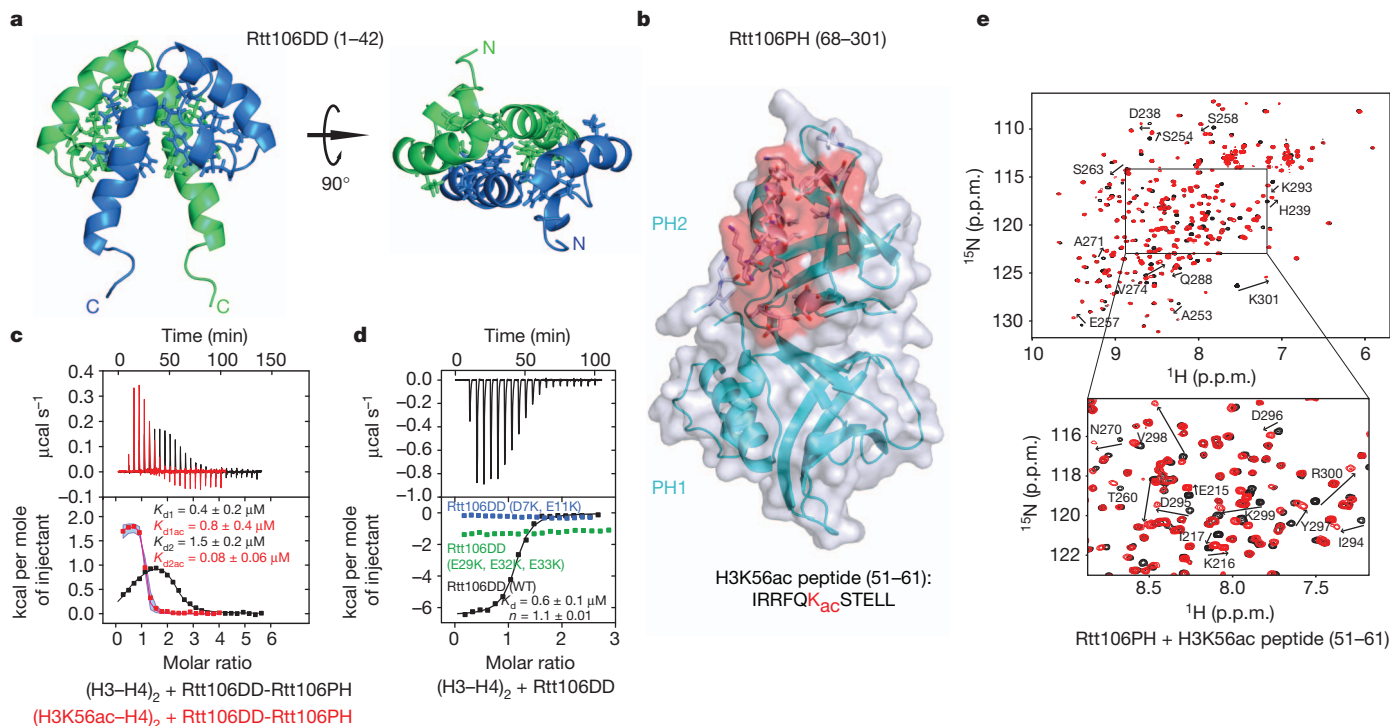*These authors contributed equally to this work.

Figure 1 | 3D structures of Rtt106 dimeric and double PH domains and their interaction with histones. **a**, NMR structure of Rtt106 dimeric region (residues 1–67) with the hydrophobic residues constituting the dimerization interface in stick representation. Two protomers are in blue and green. Residues 43–67 are disordered and omitted for clarity. **b**, Crystal structure of Rtt106 double PH domain (Rtt106PH, residues 68–301) with H3K56ac peptide-binding surface in flesh. **c**, ITC results (top, raw titration data; bottom, integrated heat measurements) for the interaction of Rtt106DD–Rtt106PH (residues 1–301) with non-acetylated $(H3–H4)_2$ (black) and K56-acetylated $(H3–H4)_2$ (red). For the former interaction, a two-site binding model

(dissociation constants K_{d1} and K_{d2}) was used. For the latter, an activation step accounting for the effect of acetylation was incorporated in the two-site binding model (K_{d1ac} and K_{d2ac}). K_ds are reported with s.d. determined by nonlinear least-squares analysis. The light blue envelop represents simulated data for K_{d2ac} varying from 0.01 to 0.1 μM and $K_{d1ac} = 0.4$ μM. **d**, ITC data for the interaction of Rtt106DD (residues 1–42) with $(H3–H4)_2$. Stoichiometry n and K_d are indicated. Data for two mutant forms of Rtt106DD, D7K, E11K and E29K, E32K, E33K, are also shown. **e**, $^1H–^{15}N$ HSQC spectra of H3K56ac peptide-bound (red) versus free (black) Rtt106PH. Perturbed signals are labelled on the spectra.

Figure 2 | Identification of a K56ac-binding cleft in Rtt106 and model of Rtt106 in complex with K56-acetylated $(H3–H4)_2$. **a**, Binding cleft and the side chain of K301 (red) in Rtt106PH and Rtt106PHL. **b**, Chemical shift changes in Rtt106PH $^1H–^{15}N$ HSQC spectra upon titration with the H3K56ac peptide (from red to purple signals) are compared to the chemical shifts of free

Rtt106PHL (black signals) with respect to seven residues in the vicinity of the binding cleft. A214, E215, K216 and I217 belong to the disordered loop adjacent to the binding cleft. **c**, Structural model of Rtt106 (residues 1–301) in complex with the $(H3–H4)_2$ tetramer. Atomic coordinates of $(H3–H4)_2$ are from the structure of budding yeast nucleosome core particle (PDB access code 1ID3)[16].

of αN, favouring interaction with Rtt106. This may be correlated to the observed entropy-driven increase in affinity upon acetylation (Supplementary Fig. 6).

In comparison to Rtt106PH (residues 68–301), the crystal structure of the longer Rtt106PHL (residues 68–315), with an α-helical extension from residues 302 to 311 (Supplementary Fig. 4), reveals a change in the conformation of K301 that leads to the identification of a K56/K56ac binding cleft in Rtt106 (Fig. 2a and Supplementary Table 2). Via an approximately 180° hinged motion, the side chain of K301 changes from being solvent-exposed in the structure of Rtt106PH to being partially buried in a cleft that is part of the H3K56ac peptide-binding region identified using NMR spectroscopy (Figs 1b, e and 2a). Noticeably, the ^{1}H–^{15}N chemical shifts of several Rtt106PHL residues in the vicinity of the K301-occupied cleft match well the end-point chemical shifts of corresponding Rtt106PH residues upon titration with the H3K56ac peptide (Fig. 2b and Supplementary Fig. 10). With only one lysine (K56) in the H3K56ac and H3K56 peptides (Fig. 1b), the correspondence in NMR chemical shifts between the Rtt106PH–H3K56ac complex and Rtt106PHL strongly suggests that the K301-binding cleft is also the binding pocket for acetylated and non-acetylated K56.

The structural difference between Rtt106PH and Rtt106PHL likely reflects a dynamic exchange between an open and a closed state of the binding site. Consistent with dual states, H3K56ac peptide binding to Rtt106PHL occurs, but with lower affinity than for Rtt106PH (Supplementary Fig. 11). The relatively large crystallographic B-factor values for the C-terminal residues of Rtt106PHL (residue 299 and onwards) are also consistent with conformational flexibility (Supplementary Fig. 12). Moreover, in another crystal structure of Rtt106 (residues 65–320)[15], there is no detectable electron density for residues 303–320 encompassing the helical extension of Rtt106PHL. Also supporting a two-state binding site with the open conformation favouring histone binding, four mutations at the C-terminal end of Rtt106PH markedly increased Rtt106PH affinity for the H3K56ac peptide (for example, $K_d = 0.4 \pm 0.1$ mM for K299A, Supplementary Table 3).

To illustrate how Rtt106 may associate with (H3–H4)$_2$, we derived a structural model of the complex (Fig. 2c). In the model, in which the dyad symmetry axes of the Rtt106DD and (H3–H4)$_2$ structures coincide, Rtt106DD is placed in a positively charged cavity formed between the two H3–H4 subcomplexes[16] and without any contact with the region of H3 encompassing K56, in accordance with experimental data. Furthermore, we verified by ITC that two different sets of mutations (D7K and E11K; and E29K, E32K and E33K) that reverse negatively charged surface areas of Rtt106DD without affecting the 3D structure disrupt binding to H3–H4 (Fig. 1d). We also note that removal of the dimeric domain renders Rtt106 non-functional in vivo (data not shown). In addition, in the model the flexible Rtt106 interdomain linker (residues 43–67) has an appropriate length to position Rtt106PH in contact with the K56-containing region of H3.

The histone-binding surface of Rtt106PH was further validated by measuring the affinity of twenty Rtt106PH mutants for the H3K56ac peptide (Fig. 3a, Supplementary Table 3, Supplementary Fig. 9 and Supplementary Discussion). Several of the mutations that affect binding were then incorporated into full-length Rtt106 to assess interactions with intact histones in vivo. Wild-type and mutant Rtt106 proteins were produced from budding yeast, isolated by tandem affinity purification and probed for association with histones by western blot. Whereas histone H3 co-purified with wild-type Rtt106, several surface mutations introduced in Rtt106 blocked (Y261A, F269A, Y291A and I294A) or diminished (I259A and Q288A) histone binding in vivo (Fig. 3b). Also consistent with the in vitro binding data (Fig. 3a and Supplementary Table 3), reduced amounts of H3 were detected with Rtt106 harbouring the Y297A mutation in the putative K56ac binding cleft (Supplementary Fig. 13a). These results indicate that the interaction interface identified in vitro from titration of Rtt106PH with
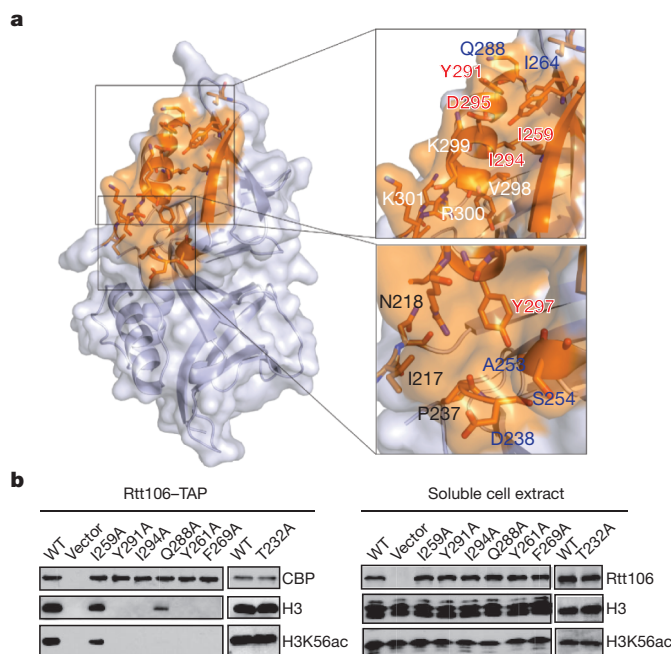

Figure 3 | Effects of Rtt106PH mutations on H3K56ac interaction.
a, Surface representation of Rtt106PH with the H3K56ac peptide-binding region in orange. The upper box represents a mainly hydrophobic area of the binding site whereas the lower box highlights the K56/K56ac binding cleft. Affinities of Rtt106PH mutants for the H3K56ac peptide were measured and K_ds are reported in Supplementary Table 3. Mutated residues that totally abolish, decrease, have no effect or enhance binding are labelled in red, blue, black and white, respectively. **b**, Wild-type (WT) and mutant tandem affinity purification (TAP)-tagged full-length Rtt106 were purified from yeast cells and analysed by western blot using indicated antibodies. CBP, calmodulin-binding peptide tag. Rtt106 mutated outside the binding site (T232A) was used as control.

an H3K56ac peptide is important for the proper interaction of Rtt106 and H3 in vivo.

Rtt106 is crucial for heterochromatin silencing in yeast in the absence of the Cac1 (also known as Rfl2) subunit of CAF-1, another histone chaperone implicated in K56ac-dependent replication-coupled chromatin assembly[5,17]. Using the green fluorescent protein (GFP) as a reporter in a gene silencing assay[18], we confirmed that there was significant loss of silencing of the *GFP* gene in *cac1Δrtt106Δ* cells (Fig. 4a and Supplementary Fig. 13b). *GFP* silencing was restored to almost the level in control W303-1A cells by expressing wild-type Rtt106 but not by expressing Rtt106 mutants (Y261A, F269A, Y291A and I294A) that are highly defective in H3 binding in vivo. Expression of Rtt106 mutants (I259A, Q288A and Y297A) that showed reduced H3K56ac binding in vivo slightly reduced *GFP* silencing in *cac1Δrtt106Δ* cells compared to wild-type Rtt106 expression (Fig. 4a and Supplementary Fig. 13b). These results show that the ability of Rtt106 to contact the K56-containing surface of H3 via the double PH domain is important for transcriptional silencing.

Rtt106 is also critical for maintenance of genomic stability in the absence of Cac1 (ref. 5). To test if the Rtt106 mutants that showed defects in H3 binding in vivo (Fig. 3b and Supplementary Fig. 13a) would have increased DNA damage sensitivity, yeast cells harbouring wild-type or mutant Rtt106 but lacking Cac1 were grown in media containing the genotoxic agents methyl methanesulphonate (MMS) or camptothecin (CPT). Rtt106 mutants severely defective for H3 interaction were more susceptible to MMS and CPT treatment than wild-type or Rtt106 mutants having little or no defect in H3 binding, indicating that the ability of Rtt106 to bind H3K56ac is connected to its role in preserving genomic integrity (Fig. 4b and Supplementary Fig. 13c).
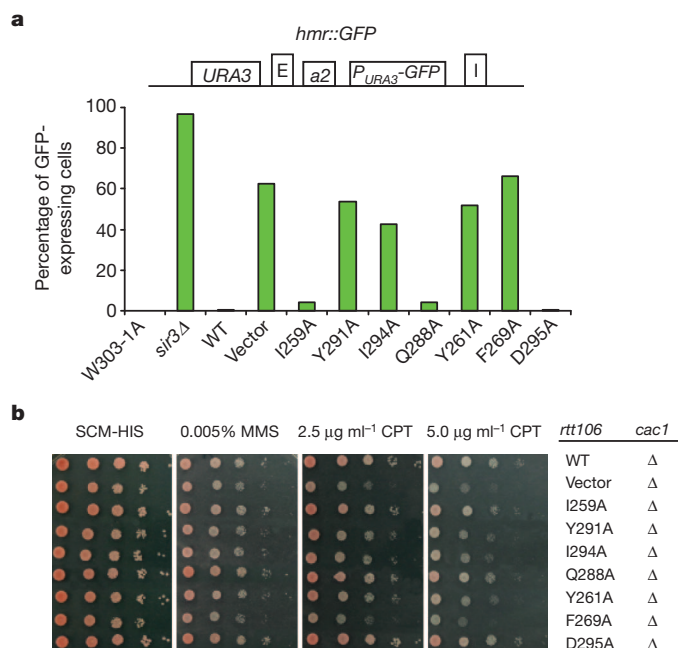
a

b

Figure 4 | Effects of Rtt106 mutations on *HMR* silencing and genome stability. a, Schematic of the GFP-based gene silencing reporter assay. The *GFP* gene (P_{URA3}-*GFP*) at the silent mating type locus *HMR* (*hmr::GFP*) is controlled by the *URA3* gene promoter, silencers E and I, and *a2* gene. Gene silencing is reported as percentage of yeast cells expressing GFP. One representative of three independent experiments is shown. Silencing is observed in W303-1A control cells, but not in cells lacking the silent information regulator gene *SIR3*. Expression of wild-type (WT) Rtt106 in *cac1Δrtt106Δ* cells restores silencing. Expression of Rtt106 mutated in the H3K56ac-binding surface does not or only partially restores silencing. **b**, Mutations in the H3 binding sites of Rtt106 enhance the DNA-damage sensitivity of *cac1Δ* mutant cells. Cells of the indicated genotypes were spotted onto media lacking histidine (SCM-HIS) for plasmid selection, without or with MMS or CPT for DNA damage assessment.

In conclusion, our study supports a working model where direct binding of Rtt106 to H3K56-acetylated (H3–H4)$_2$ tetramers contributes to nucleosome assembly with implications for DNA replication, gene silencing and maintenance of genomic stability. Our findings strongly suggest that the preferential association of Rtt106 with acetylated (H3–H4)$_2$ originates primarily from a K56 acetylation-triggered increase in conformational entropy of H3 αN. This mode of specific association with a modified histone is fundamentally different from that of so-called histone mark reader domains[19].

METHODS SUMMARY

Wild-type and mutant Rtt106 proteins were expressed in *Escherichia coli* as His$_6$-fusions and purified by immobilized metal affinity and gel filtration chromatography. Purification and reconstitution of H3–H4 followed established procedures[20]. The homogeneous site-specific installation of an acetylated lysine analogue (methylthiocarbonyl-thiaLys) in place of H3K56 in H3–H4 was done as reported[12] (Supplementary Fig. 14). Protein labelling with selenomethionine (SeMet) for X-ray crystallography studies and with ^{15}N, $^{15}N/^{13}C$ and $^{15}N/^{13}C/^{2}H$ for NMR spectroscopy studies was achieved by growing *E. coli* cells in SeMet- and isotope-enriched media following standard procedures. All proteins were crystallized at 15 °C. X-ray diffraction data were collected on-site and at the Advance Photon Source (APS) synchrotron facility, Argonne National Laboratory. NMR experiments were collected at 25 °C using a Bruker Avance 700 MHz spectrometer. The solution NMR structure of Rtt106 (residues 1–67) was determined using a simulated annealing-based protocol. Tandem affinity purification, gene silencing assays and DNA damage assays were done as reported[5,17,18].

1. Luger, K. Dynamic nucleosomes. *Chromosome Res.* **14,** 5–16 (2006).
2. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).
3. Corpet, A. & Almouzni, G. Making copies of chromatin: the challenge of nucleosomal organization and epigenetic information. *Trends Cell Biol.* **19,** 29–41 (2009).
4. Chen, C. C. *et al.* Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair. *Cell* **134,** 231–243 (2008).
5. Li, Q. *et al.* Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. *Cell* **134,** 244–255 (2008).
6. Williams, S. K., Truong, D. & Tyler, J. K. Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc. Natl Acad. Sci. USA* **105,** 9000–9005 (2008).
7. Xie, W. *et al.* Histone H3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. *Mol. Cell* **33,** 417–427 (2009).
8. Das, C., Lucia, M. S., Hansen, K. C. & Tyler, J. K. CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature* **459,** 113–117 (2009).
9. Tjeertes, J. V., Miller, K. M. & Jackson, S. P. Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *EMBO J.* **28,** 1878–1889 (2009).
10. Yuan, J., Pu, M., Zhang, Z. & Lou, Z. Histone H3–K56 acetylation is important for genomic stability in mammals. *Cell Cycle* **8,** 1747–1753 (2009).
11. VanDemark, A. P. *et al.* The structure of the yFACT Pob3-M domain, its interaction with the DNA replication factor RPA, and a potential role in nucleosome deposition. *Mol. Cell* **22,** 363–374 (2006).
12. Huang, R. *et al.* Site-specific introduction of an acetyl-lysine mimic into peptides and proteins by cysteine alkylation. *J. Am. Chem. Soc.* **132,** 9986–9987 (2010).
13. Blomberg, N., Baraldi, E., Nilges, M. & Saraste, M. The PH superfold: a structural scaffold for multiple functions. *Trends Biochem. Sci.* **24,** 441–445 (1999).
14. Bowman, A., Ward, R., El-Mkami, H., Owen-Hughes, T. & Norman, D. G. Probing the (H3–H4)$_2$ histone tetramer structure using pulsed EPR spectroscopy combined with site-directed spin labeling. *Nucleic Acids Res.* **38,** 695–707 (2010).
15. Liu, Y. *et al.* Structural analysis of Rtt106p reveals a DNA binding role required for heterochromatin silencing. *J. Biol. Chem.* **285,** 4251–4262 (2010).
16. White, C. L., Suto, R. K. & Luger, K. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J.* **20,** 5207–5218 (2001).
17. Huang, S. *et al.* Rtt106p is a histone chaperone involved in heterochromatin-mediated silencing. *Proc. Natl Acad. Sci. USA* **102,** 13410–13415 (2005).
18. Huang, S., Zhou, H., Tarara, J. & Zhang, Z. A novel role for histone chaperones CAF-1 and Rtt106p in heterochromatin silencing. *EMBO J.* **26,** 2274–2283 (2007).
19. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399,** 491–496 (1999).
20. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304,** 3–19 (1999).

Author Contributions Q.H. carried out the NMR spectroscopy experiments, prepared methylthiocarbonyl-aziridine and acetylated H3–H4, and performed the ITC measurements and analysis with assistance from G.M.; D.S. obtained crystals of the Rtt106 constructs, performed the X-ray diffraction measurements and solved the structures. Q.L. did the *in vivo* experiments. G.C. helped with the NMR experiments and ITC data analysis, A.F. and B.A.D. purified Rtt106 for initial crystal screening. J.R.T. worked on the refinement of crystal structures, M.V.B. contributed extensively to plasmid design, mutagenesis and protein preparations. G.M. and Z.Z. supervised the research. G.M. wrote the manuscript. All authors contributed in editing the manuscript.

Author Information The atomic coordinates and structure factors or NMR restraints of Rtt106DD, Rtt106PH, Rtt106PHL and Rtt106PH–acetyl-histamine have been deposited with the Protein Data Bank under accession codes 2LH0, 3FSS, 3TVV and 3TW1, respectively. The NMR chemical shifts of Rtt106 (residues 1–67) have been deposited in the Biological Magnetic Resonance Bank with the accession code 17832. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.M. (mer.georges@mayo.edu) or Z.Z. (zhang.zhiguo@mayo.edu).

METHODS

Protein preparation. Constructs of Rtt106 encompassing amino acids 1–315, 1–301 (Rtt106DD–Rtt106PH), 1–67, 1–42 (Rtt106DD), 68–299, 68–301 (Rtt106PH) and 68–315 (Rtt106PHL) were cloned in a modified pET vector (Novagen) encoding an N-terminal His6-tag and a tobacco etch virus (TEV) protease cleavage site. The wild-type proteins were overexpressed in *Escherichia coli* BL21(DE3) initially grown in LB broth at 37 °C to $D_{600\,nm}$ of ~0.8, then transferred to 15 °C, and after 45 min, induced by 1 mM isopropyl-β-D-thiogalactoside (IPTG) for 14 to 18 h. The cells were collected by centrifugation and resuspended in 50 mM sodium phosphate buffer, pH 7.5, 300 mM NaCl and 1 mM phenylmethanesulphonylfluoride, and lysed using an Emulsiflex C5 high-pressure homogenizer (Avestin). After centrifugation, the proteins were purified by affinity chromatography with Ni^{2+}-loaded NTA resin (Qiagen) according to the manufacturer's recommended protocol. The His6-tag was cleaved with TEV protease at 4 °C overnight and further purification was performed by size exclusion chromatography using preparative Superdex 75 or 200 columns (GE Healthcare). Cleavage of the His6-tag leaves three residues (GHM) at the N terminus of each protein.

Point mutants of Rtt106 were made following the QuikChange site-directed mutagenesis protocol (Stratagene) and were verified by DNA sequencing. All mutants were purified as described for the wild-type protein.

The preparation of isotopically labelled Rtt106 (residues 1–67), Rtt106PH and Rtt106PHL followed similar steps as above except that instead of LB broth, M9 media containing $1\,g\,l^{-1}$ ^{15}N NH4Cl, $4\,g\,l^{-1}$ [$^{12}C_6$]-D-glucose and $1\,g\,l^{-1}$ ^{15}N Isogro (Isotec) (for ^{15}N-labelled samples); and $1\,g\,l^{-1}$ $^{15}NH_4Cl$, $2\,g\,l^{-1}$ [$^{13}C_6$]-D-glucose and $1\,g\,l^{-1}$ $^{15}N/^{13}C$ Isogro (Isotec) (for $^{15}N/^{13}C$-labelled samples) were used[21]. The procedure was similar for producing $^2H/^{15}N/^{13}C$-labelled samples but with the *E. coli* cells grown in culture media prepared with 99.9% D_2O.

For producing selenomethionine (SeMet)-enriched Rtt106PH and Rtt106PHL, a similar protocol as above was used but with protein overexpression in the methionine-auxotroph *E. coli* strain B834(DE3) (Novagen) grown in M9 media with SeMet and all amino acids except methionine[22].

The preparation of histones H3 and H4 is based on a published protocol[20]. Histones H3 and H4 from *Xenopus laevis* were overexpressed in *E. coli* BL21 (DE3) Rosetta pLysS, purified separately under denaturing conditions and then combined to reconstitute the (H3–H4)2 tetramer. For each histone, cells were grown at 37 °C to $D_{600\,nm}$ of 0.6–0.8, induced with 1 mM IPTG, collected after 3 h and lysed using an Emulsiflex C5 high-pressure homogenizer (Avestin). After centrifugation, the pellet was washed with 1 M L-arginine monohydrochloride, 5 mM 2-mercaptoethanol, 10 mM Tris-HCl, pH 7.5 three times and with 1 M L-arginine monohydrochloride, 5 mM 2-mercaptoethanol, 1 M guanidine hydrochloride, 10 mM Tris-HCl, pH 7.5 once. Next, the pellet was dissolved in 10 mM dithiothreitol, 7 M guanidine hydrochloride, 20 mM Tris-HCl, pH 7.5 and centrifuged. The supernatant was dialysed several times in water containing 5 mM 2-mercaptoethanol for a period of 3 days and then lyophilized. The lyophilized solid was dissolved in 10 mM dithiothreitol, 6 M urea, 20 mM Tris-HCl, pH 7.5 (buffer A) and residual solids were spun down. The supernatant was passed through a Resource S cation exchange column (GE Healthcare) using buffer A as running buffer and eluted with a 0 to 1 M NaCl gradient. Resulting solutions of H3 and H4 were mixed at equimolar ratio and dialysed in 2 M NaCl, 1 mM EDTA, 5 mM 2-mercaptoethanol, 10 mM Tris-HCl, pH 7.5 (refolding buffer). Refolded H3–H4 was purified by size-exclusion chromatography using a Superdex 200 column (GE Healthcare) and refolding buffer as the running buffer.

Incorporation of an acetyl-lysine analogue in H3–H4. For incorporation of an acetyl-lysine analogue at position 56 of histone H3, we closely followed a published protocol[12]. The single cysteine (C110) in histone H3 was replaced by an alanine and K56 was replaced by a cysteine. C56 of H3 (K56C, C110A) or H3 (K56C, C110A)–H4 was alkylated with methylthiocarbonyl-aziridine (MTCA) to generate the acetylated lysine analogue methylthiocarbonyl-thiaLys (Supplementary Fig. 14).

MTCA was prepared as reported previously[12]. Briefly, to pre-cooled diethyl ether (30 ml) in a round-bottom flask (−80 °C, maintained with dry ice and acetone), aziridine (100 μl, 1.93 mmol), triethylamine (0.27 ml, 1.93 mmol) and methyl chlorothiolformate (0.16 ml, 1.89 mmol) were added with stirring. The reaction was allowed to proceed for 3 h before diluting the reaction mixture with diethyl ether:H_2O (15 ml:10 ml). The mixture was next shaken in a separation funnel and the organic layer isolated and washed as follow: 0.01 M HCl (5 ml, twice), H_2O (5 ml, twice) and brine (10 ml, twice). The organic layer was dried over anhydrous $MgSO_4$ and then concentrated using a rotary evaporator. The product, MTCA (100–200 mg yield), was verified by NMR spectroscopy (Supplementary Fig. 14). Aziridine was purchased from ChemService. All other chemicals were purchased from Sigma-Aldrich. Commercial reagents were used as received without further purification.

Resource S-purified H3 mutant (K56C, C110A) was dialysed extensively in water with 5 mM 2-mercaptoethanol and then in water alone. Next, it was lyophilized

and then dissolved in 100 mM ammonium bicarbonate, pH 8.0, to a final concentration of ~2 mg ml^{-1}. MTCA was added to a final concentration of 50–200 mM. The reaction proceeded for ~200 min at room temperature. The modified product was verified by mass spectrometry, lyophilized and subsequently used for reconstitution with H4 as explained above (Supplementary Fig. 14). Alternatively, the acetylation reaction was carried out on refolded H3 (K56C, C100A)–H4. After purification with Superdex 200, H3 (K56C, C100A)–H4 was dialysed in 100 mM ammonium bicarbonate, pH 8.0, 300 mM NaCl and concentrated to 2–10 mg ml^{-1}. 100 mM of MTCA was added and the reaction was left at room temperature for ~200 min. Acetylated H3–H4 obtained from either method was then extensively dialysed in the final buffer of 20 mM Tris-HCl, pH 7.5, 100 mM NaCl for ITC experiments.

Isothermal titration calorimetry. Measurements were carried out at 10 °C using a VP-ITC titration calorimeter (MicroCal). All proteins were prepared in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl. Rtt106DD, Rtt106 (residues 1–67) or Rtt106DD–Rtt106PH (residues 1–301) in the calorimeter injection syringe at concentrations ranging from 0.5 mM to 0.74 mM were delivered as a series of 5- to 8-μl injections every 5 min to the reaction cell containing non-acetylated or K56-acetylated H3–H4 at concentrations of 20 μM or 30 μM. Measurements were paired with control experiments for heat of mixing and dilution. Data were analysed with Levenberg−Marquardt nonlinear regression using different models programmed in Origin 7.0 and in-house software. Data were also simulated using Mathematica (Wolfram Research).

Crystallization, data collection and structure determination. Crystals of SeMet-labelled Rtt106PH were grown at 15 °C by vapour diffusion of hanging drops by mixing 1 μl of Rtt106PH at 30 mg ml^{-1} in 20 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM dithiothreitol, 10% glycerol with 1 μl of reservoir solution containing 4% (v/v) Tacsimate, pH 5.0, 12% PEG 3350. The crystals were cryoprotected by transfer to reservoir solution supplemented with 20% glycerol for 10–15 min, and were quick-frozen in a cryoloop (Hampton Research) with liquid nitrogen.

Single anomalous diffraction data were collected for Rtt106PH at APS 19BM beamline, Argonne National Laboratory. Processing of diffraction images and scaling of the integrated intensities were performed using HKL3000 (ref. 23). Crystals were of the space group C2 with one molecule of Rtt106PH in the asymmetric unit and a Matthews coefficient of 2.32 Å^3 Da^{-1}. The four Se atom positions were determined using SHELXD[24], followed by density modification with RESOLVE[25] and initial automatic model building using ARP/wARP[26]. Model correction and refinement were undertaken using the programs COOT[27] and REFMAC5 (ref. 28). Resolution is 1.4 Å. For the Ramachandran geometry, 91.5% of all dihedral angles are located in most favoured regions and 8.5% in additionally allowed regions.

A complex of Rtt106PH and acetyl-histamine (AHN) was prepared by soaking the centred monoclinic crystals of SeMet-labelled Rtt106PH for 5 min in a 1 M solution of AHN (Sigma) prepared in the mother liquor and by flash-freezing in liquid nitrogen. Diffraction data were collected at 100 K using a Rigaku Microfocus 007 generator and Rigaku R-AXIS IV^{++} area detector. Data processing and scaling of the integrated intensities were performed using HKL2000 (ref. 29). The structure was solved by molecular replacement using PHASER[30] with the crystal structure of Rtt106PH as a search model. Model correction and refinement were undertaken using the programs COOT[27] and PHENIX[31]. Resolution is 1.8 Å. For the Ramachandran geometry, 92% of all dihedral angles are located in most favoured regions and 8% in additionally allowed regions.

Crystals of SeMet-labelled Rtt106PHL (residues 68–315) were grown at 15 °C by hanging drop vapour diffusion after mixing 1 μl of Rtt106PHL at 30 mg ml^{-1} in 20 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM dithiothreitol, 10% glycerol with 1 μl of reservoir solution containing 8% (v/v) Tacsimate, pH 5.0, 20% PEG 3350. The crystals were cryoprotected as for Rtt106PH above.

Diffraction data were collected at 100 K at the APS 19ID beamline, Argonne National Laboratory. Processing of diffraction images and scaling of the integrated intensities were performed using HKL3000 (ref. 23). Crystals were of the space group P2₁ with two molecules of Rtt106PHL in the asymmetric unit and a Matthews coefficient of 2.86 Å^3 Da^{-1}.

The structure was solved by molecular replacement using PHASER[30] with the crystal structure of Rtt106PH as a search model. Model correction and refinement were undertaken using the programs COOT[27] and PHENIX[31]. Resolution is 2.6 Å. For the Ramachandran geometry, 86.8% of all dihedral angles are located in most favoured regions and 12.9% in additionally allowed regions.

NMR spectroscopy. NMR experiments were conducted at 25 °C using a Bruker Avance 700 MHz spectrometer equipped with a cryogenic probe. The Rtt106DD, Rtt106 (residues 1–67), Rtt106PH and Rtt106PHL protein samples (wild-type and mutants) were at concentrations of ~0.6 mM in 20 mM sodium phosphate buffer, pH 6.9, 30 mM NaCl and 5 mM dithiothreitol. 95% of backbone carbon and

nitrogen resonances of Rtt106PH and 98% of all resonances of Rtt106 (residues 1–67) were assigned from regular and transverse relaxation-optimized spectroscopy (TROSY)-based experiments[32] using $^2H/^{13}C/^{15}N$-labelled or $^{13}C/^{15}N$-labelled Rtt106 samples. NMR data were processed with NMRPipe/NMRDraw[33] and analysed with SPARKY (Goddard, T. D. & Kneller, D.G., http://www.cgl.ucsf.edu/home/sparky/). The solution structure of Rtt106 (1–67) homodimer was calculated and refined using the program CNS version 1.2, using a simulated annealing protocol for torsion angle dynamics[34]. A total of 1,150 distance constraints derived from 3D ^{15}N-resolved nuclear Overhauser enhancement spectroscopy (NOESY), 3D ^{13}C-resolved NOESY, 2D NOESY and 2D ^{13}C filtered–edited NOESY spectra were used in the structure calculations. Also included were 48 hydrogen bond distance constraints derived from hydrogen–deuterium exchange measurements and NOESY data, and 112 dihedral angles derived from chemical shift index analysis of $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO and $^1H\alpha$ atoms. The 20 lowest energy conformers from 200 refined structures were selected to represent the NMR ensemble. For the well-folded part of the molecule (residues 6–42), 94.3% of all dihedral angles are located in most favoured regions and 5.4% in additionally allowed regions of the Ramachandran plot.

The interactions of non-acetylated histone H3 (H3K56) and H3K56ac (residues 51–61 for both) peptides with wild-type Rtt106PH were quantified by recording a series of $^1H–^{15}N$ HSQC spectra of ^{15}N-labelled Rtt106PH at increasing concentrations of the peptides. The mutated Rtt106PH proteins were similarly titrated with the H3K56ac peptide. The dissociation constants (K_ds) were estimated from the NMR chemical shift perturbations of five to eight non-overlapping $^1H–^{15}N$ resonances by nonlinear least-squares fitting of the following equation:

$$\frac{\Delta\delta}{\Delta\delta_{max}} = 0.5\left[\left(1+\frac{K_d}{C_p}+M\right) - \sqrt{\left(\left(1+\frac{K_d}{C_p}+M\right)^2 - 4M\right)}\right]$$

where M is the molar ratio of H3K56 or H3K56ac peptide to Rtt106PH, C_p, the concentration of Rtt106PH and $\Delta\delta$, the normalized chemical shift change calculated as:

$$\Delta\delta = \sqrt{(\delta_{HN})^2 + (\delta_N)^2}$$

where δ_{HN} and δ_N denote the amide hydrogen and nitrogen atoms chemical shift differences, respectively, between the free and peptide-bound states for Rtt106PH. $\Delta\delta_{max}$ is the normalized difference in chemical shifts of the free and peptide-saturated Rtt106PH.

Molecular illustrations. Molecular illustrations were prepared with PyMOL (http://www.pymol.org/) and MOLMOL[35].

Yeast strains, plasmids and antibodies. All budding yeast strains used in this study were derived from the parental W303 background strain (*leu2-3, ura3-1, his3-11, trp1-1, ade2-1, can1-100*) and are listed in Supplementary Table 4. Rtt106 constructs were tagged at their C termini with the tandem affinity purification (TAP) tag according to published procedures[36]. Full-length Rtt106 containing its endogenous promoter and its C-terminal TAP tag was cloned into pRS313 vector and the resulting plasmid was used as a template to make Rtt106 mutants using the QuikChange site-directed mutagenesis kit (Stratagene). Mutant strains were constructed in the W303 background strain by standard yeast cloning methods[37]. Antibodies used in this study were produced as described previously[38].

Binding of Rtt106 with H3 in yeast cells using tandem affinity purification. To test the effect of Rtt106 mutations on H3 binding, wild-type and mutant Rtt106 proteins were purified from yeast cells using the TAP tag procedure and co-purified proteins were detected by western blotting with antibodies against calmodulin-binding peptide (CBP) and histones H3 and H3K56ac as described previously[5].

Assay for silencing at the *HMR* locus using the *GFP* reporter. The silencing assay was performed as described previously[17]. Briefly, exponentially growing wild-type or mutant cells were collected, washed with PBS, resuspended in SCM-TRP media, and analysed by flow cytometry.

Assay for the sensitivity towards DNA-damaging agents. To analyse the sensitivity of *cac1Δ* yeast cells harbouring wild-type or mutant Rtt106 to different DNA-damaging agents, tenfold serial dilutions of freshly grown yeast cells were spotted onto selective media SCM-HIS containing different concentrations of methyl methanesulphonate (MMS: 0.001, 0.005 and 0.01% (v/v)) or camptothecin (CPT: 0, 1, 2.5 and 5 μg ml^{-1}). Plates were incubated at 30 °C for 3 days and then photographed.

21. Botuyan, M. V. *et al.* Structural basis of BACH1 phosphopeptide recognition by BRCA1 tandem BRCT domains. *Structure* **12,** 1137–1146 (2004).
22. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9,** 1665–1672 (1990).
23. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62,** 859–866 (2006).
24. Sheldrick, G. M. A short history of *SHELX*. *Acta Crystallogr. A* **64,** 112–122 (2008).
25. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374,** 22–37 (2003).
26. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6,** 458–463 (1999).
27. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
29. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
30. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **60,** 432–438 (2004).
31. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
32. Ferentz, A. E. & Wagner, G. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **33,** 29–65 (2000).
33. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
34. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
35. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14,** 51–55 (1996).
36. Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24,** 218–229 (2001).
37. Thomas, B. J. & Rothstein, R. Elevated recombination rates in transcriptionally active DNA. *Cell* **56,** 619–630 (1989).
38. Zhou, H., Madden, B. J., Muddiman, D. C. & Zhang, Z. Chromatin assembly factor 1 interacts with histone H3 methylated at lysine 79 in the processes of epigenetic silencing and DNA repair. *Biochemistry* **45,** 2852–2861 (2006).

ARTICLE

DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes

Ofir Hakim[1]*, Wolfgang Resch[2]*, Arito Yamane[2]*, Isaac Klein[3], Kyong-Rim Kieffer-Kwon[2], Mila Jankovic[3], Thiago Oliveira[3,4], Anne Bothmer[3], Ty C. Voss[1], Camilo Ansarah-Sobrinho[2], Ewy Mathe[5], Genqing Liang[2], Jesse Cobell[2], Hirotaka Nakahashi[2], Davide F. Robbiani[3], Andre Nussenzweig[6], Gordon L. Hager[1], Michel C. Nussenzweig[3,7]* & Rafael Casellas[2,8]*

Recurrent chromosomal translocations underlie both haematopoietic and solid tumours. Their origin has been ascribed to selection of random rearrangements, targeted DNA damage, or frequent nuclear interactions between translocation partners; however, the relative contribution of each of these elements has not been measured directly or on a large scale. Here we examine the role of nuclear architecture and frequency of DNA damage in the genesis of chromosomal translocations by measuring these parameters simultaneously in cultured mouse B lymphocytes. In the absence of recurrent DNA damage, translocations between *Igh* or *Myc* and all other genes are directly related to their contact frequency. Conversely, translocations associated with recurrent site-directed DNA damage are proportional to the rate of DNA break formation, as measured by replication protein A accumulation at the site of damage. Thus, non-targeted rearrangements reflect nuclear organization whereas DNA break formation governs the location and frequency of recurrent translocations, including those driving B-cell malignancies.

Most cancers bear cytogenetic abnormalities including chromosomal translocations and rearrangements[1]. Although translocations and rearrangements are central to the development of cancer, their origins are poorly understood. One possibility is that they arise from rare and random events that are selected in tumour precursors because they provide a growth advantage. However, increasing evidence indicates that mechanistic factors other than simple selection may have a role in their genesis. In B lymphocytes, V(D)J recombination, class switch recombination (CSR) and somatic hypermutation (SHM) produce obligate single- and double-strand DNA break intermediates that can become substrates for translocations[2,3]. Consistent with this idea, genetic ablation of the enzymes that create DNA lesions during V(D)J recombination (RAGs) or CSR and SHM (AID; also called AICDA) has a profound protective effect on B-cell transformation[2,4].

A second mechanism that may also influence the incidence of chromosomal translocations is nuclear architecture. Two decades of imaging and recent molecular approaches have established that the spatial organization of the genome is not random, but compartmentalized into chromosome territories as well as transcriptionally active and silent subnuclear environments[5–8]. These compartments are believed to influence the frequency with which genes from different chromosomes can interact and recombine. Furthermore, there is a strong association between transcriptional activity and translocation[9].

Using new methods that capture rearrangements genome-wide, thousands of translocations were recently isolated in primary B cells in the absence of growth selection[9,10]. The studies confirmed the notion that the formation of chromosomal translocations is influenced by spatial conformation, targeted DNA damage and open chromatin.

Consistent with the distribution of mammalian chromosomes in discrete nuclear territories, most rearrangements occurred intrachromosomally[9,10]. Moreover, rearrangements in *trans* were biased towards transcriptionally active genes, and particularly those targeted by AID[9,10]. What the studies did not resolve, however, was to what extent recurrent DNA damage, chromatin accessibility, or spatial genome organization influence the location and frequency of cancer-inducing translocations. Here we make use of deep-sequencing techniques to establish the relationship between genome-wide spatial interactions, DNA damage and translocations in activated B cells.

A map of *Igh* and *Myc* long-range nuclear associations

To identify genomic regions that are in close spatial proximity to *Igh*, *Myc* and *Mycn* (also called *N-myc*) loci, we performed chromosome conformation capture experiments[11] followed by deep-sequencing (4C-seq). We used *Igh* and *Myc* as baits because they are actively transcribed and targeted by AID[12]. As controls, we analysed *Mycn*, which is transcriptionally silent in peripheral B cells and does not recruit AID[13], and *Igh* in mouse embryonic fibroblasts (MEFs), where immunoglobulin genes are not expressed. Because of the large size of *Igh*, we used two 4C-seq baits specific for 5′Eμ and 3′Eα enhancers (Supplementary Fig. 1a). Two independent 4C libraries (HindIII and BglII) were constructed for each condition (see Methods). In all experiments, most of the 4C sequence reads (76% on average) originated from the *cis* chromosome (Fig. 1a, Supplementary Table 1 and Supplementary Fig. 2a), an observation consistent with the finding that loci on the same chromosome preferentially interact in *cis* within a chromosome territory[5,6,9].

[1]Laboratory of Receptor Biology and Gene Expression, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA. [2]Genomics & Immunity, NIAMS, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA. [3]Laboratory of Molecular Immunology, The Rockefeller University, New York, New York 10065, USA. [4]Medical School of Ribeirao Preto/USP, Departmento of Genetics, 8 National Institute of Science and Technology for Stem Cells and Cell Therapy and Center for Cell-based Therapy, Ribeirao Preto, SP 14051-140, Brazil. [5]Biodata Mining and Discovery, NIAMS, National Institutes of Health, Bethesda, Maryland 20892, USA. [6]Laboratory of Genome Integrity, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA. [7]Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10065, USA. [8]Center of Cancer Research, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA.
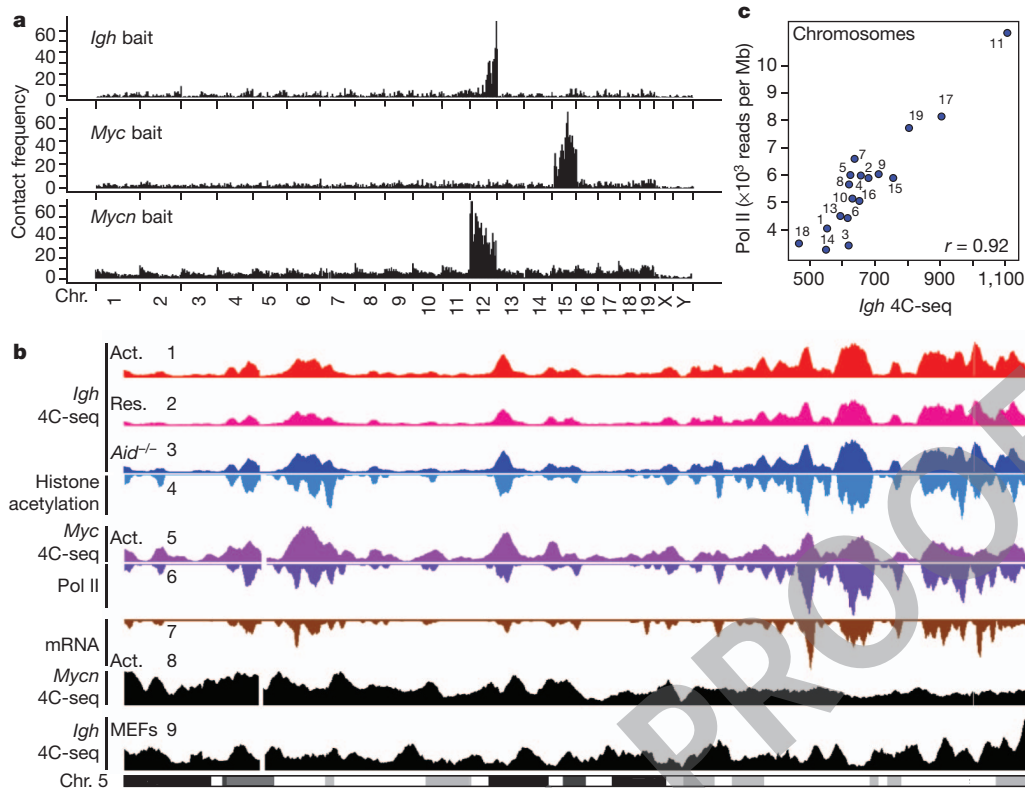*These authors contributed equally to this work.

Figure 1 | Characterization of the *Igh*, *Myc* and *Mycn* interactomes in B lymphocytes. **a**, Genome-wide interaction profile of *Igh* 3′Eα, *Myc* and *Mycn* in activated B cells. Plots show the percentage of HindIII fragments carrying 4C-seq reads. **b**, Contact frequency of *Igh* with chromosome 5 in activated B cells (lane 1), anti-HEL homozygous activated B cells (lane 2), or resting B cells (lane 3). Histone acetylation (lane 4), RNA Pol II (lane 6) and mRNA (lane 7) density is also shown. Lanes 5 and 8 show *Myc* and *Mycn* contacts, respectively. Lane 9 represents *Igh* contacts in MEFs. **c**, Comparison of Pol II and *Igh* 4C-seq data per chromosome normalized as reads per mappable megabase.

To explore contact frequencies in *trans*, the mouse genome was partitioned into 200-kilobase (kb) non-overlapping windows and the number of 4C-seq-positive fragments was calculated for each window (see Methods). In activated B cells the contact profile of *Igh* was nonrandom, following a peaks-and-valley pattern similar to that reported for transcriptionally active loci in other cell types[14] (Fig. 1b, lane 1). This pattern was comparable for Eμ and Eα baits (Spearman's $\rho = 0.70$, Supplementary Fig. 1b), and was further reproduced in resting wild-type and activated AID-deficient B cells ($Aid^{-/-}$) (Spearman's $\rho = 0.93$ (resting) and 0.94 ($Aid^{-/-}$); Fig. 1b, lanes 1–3, and Supplementary Fig. 2b). Nearly identical profiles were observed in B cells homozygous for an anti-hen egg lysozyme VDJ knock-in (anti-HEL, Spearman's $\rho = 0.89$, Supplementary Table 2), where most of the *Igh* variable domain is in germline configuration. Thus, globally, *Igh* nuclear interactions in peripheral B cells are largely independent of cell activation, AID expression, or *Ig* variable and constant region gene recombination.

4C-seq was validated by three-dimensional DNA fluorescence *in situ* hybridization (3D DNA FISH) using Perkin Elmer's ultra-high-throughput imaging system. This new approach allowed the automated and unbiased screening of 48,162 activated B cells. Analysis of *Igh* interactions with 14 genomic sites showed 3D FISH measurements to be in good agreement with 4C-seq ($R^2 = 0.99$, Supplementary Fig. 3).

Genome features enriched in contacting loci

Even though the *Igh* and *Myc* loci are on different chromosomes their interactome was significantly correlated (Spearman's $\rho = 0.58$ ($P < 1 \times 10^{-8}$); Fig. 1b, lanes 1–3 and 5). This finding is consistent with the notion that these genes frequently associate and thus may share a common subnuclear environment in B cells[15–17]. To characterize

the genomic properties of loci interacting with *Igh* and *Myc*, we compared their 4C-seq profiles to genome-wide epigenetic and transcription maps[13,18]. The analyses revealed a good concordance between *Igh*- and *Myc*-interacting loci and activating chromatin acetylation, RNA polymerase II (Pol II) and messenger RNA transcripts (Fig. 1b, lanes 1–7, and Supplementary Fig. 2c, d). This correlation was particularly evident when entire chromosomes were considered. For example, *Igh* contact probability and Pol II were highest for chromosomes 11, 17 and 19, and lowest for chromosomes 3, 14 and 18 (Pearson's $r = 0.92$ ($P = 1.7 \times 10^{-8}$); Fig. 1c). This hierarchical ordering closely followed gene density estimates, which are highest (2.1%) for mouse chromosome 11, and lowest (1%) for chromosomes 3 and 14 (Supplementary Table 3). Similar correlations were obtained for *Myc*, although the correlations were lower than for *Igh* (Pearson's $r = 0.61$ ($P = 0.0013$); Supplementary Fig. 2d, e and Supplementary Table 4). Altogether, the data recapitulate the spatial compartmentalization of transcriptionally active, gene-dense domains[6,11,14]. In marked contrast, the interactome of transcriptionally silent *Mycn* in B cells or *Igh* in MEFs seemed to be random and did not correlate well with any of the genomic features surveyed, with the exception of centromeric regions for *Myc* (Fig. 1b, lanes 8 and 9, and Supplementary Fig. 4). This latter feature might reflect the tendency of some silent loci to co-localize with peri-centromeric, repressive heterochromatin[19]. We conclude that in peripheral B cells *Igh* and *Myc* are more closely associated with transcribed, epigenetically accessible genomic sites, whereas interactions of transcriptionally inactive *Mycn* (or *Igh* in MEFs) are more randomly distributed.

Translocations in the absence of AID

To examine the role of nuclear contacts on the genesis of chromosomal translocations in the absence of programmed DNA damage, we

compared the 4C-seq genomic profiles to translocation-capture sequencing (TC-seq) data sets obtained from $Aid^{-/-}$ activated B cells[9]. The TC-seq assay was recently developed to map genomic rearrangements comprehensively in B cells where a specific DNA break at $Ig\mu$ (also called $Ighm$) or Myc is created via expression of the I-SceI mega-nuclease[9,12]. In the absence of AID, Igh^{I-SceI} or Myc^{I-SceI} translocate to loci that occasionally suffer DNA damage as a result of normal metabolic processes such as transcription or DNA replication.

A total of 68,403 and 28,548 rearrangements were captured between Igh^{I-SceI} and Myc^{I-SceI}, respectively, and the rest of the genome (Fig. 2a; see also ref. 9). Visual comparison of the aligned 4C- and TC-seq reads revealed a nonrandom distribution of AID-independent rearrangements across chromosomes (Fig. 2b). Notably, the translocation profiles resembled the Igh and Myc interactome as well as accessible chromatin as measured by histone acetylation (Fig. 2b and Supplementary Fig. 5). Conversely, there was no obvious concordance between $Mycn$ nuclear contacts in the same cells and Igh^{I-SceI} or Myc^{I-SceI} translocations (Supplementary Fig. 5). To validate these observations genome-wide, Igh, Myc and $Mycn$ nuclear contacts were subdivided into quartiles (Q) and the data plotted as a function of total Igh or Myc translocations per 200-kb non-overlapping windows. The results showed that the greater the interaction frequency between Igh (or Myc) and a given genomic site, the more likely that the two loci were translocated (Q1 versus Q4, $P = 0.0005$ (permutation test); Fig. 2c and Supplementary Fig. 6). In the case of Igh, where the number of captured rearrangements was substantial, translocations per chromosome were directly proportional to the contact frequency between Igh and a given chromosome (Pearson's $r = 0.77$ ($P = 0.0002$); Fig. 2d). Conversely, we observed little or no correspondence between Igh or Myc translocations and the interactome of $Mycn$ in B cells (Q1 versus Q4, $P = 0.35$; Supplementary Fig. 6). The data are thus consistent with the notion that AID-independent translocations occur preferentially between interacting genomic loci that are epigenetically accessible.

AID-targeted translocations

AID produces lesions in a large number of defined hotspots, many of which are recurrent translocation partners for Igh in lymphoma[2,9,10]

(Fig. 3a). To determine whether the location of these hotspots could be explained by B-cell nuclear architecture, we ranked RefSeq genes on the basis of 4C values. The analysis showed that a large fraction of loci carrying Igh translocation hotspots engaged in recurrent long-range interactions with this locus in activated B cells (Fig. 3b, c). At the same time, we identified thousands of genes that interacted repeatedly with Igh but that were not associated with translocation hotspots (Fig. 3c). For instance, up to 2,361 genes (11% of all RefSeq genes) outranked Myc in Igh contact frequency (Supplementary Table 5 and Supplementary Fig. 7), even though only 58 of them were recurrently rearranged to Igh (Fig. 3c). In a similar manner, whereas translocation hotspots in Myc^{I-SceI} B cells were biased for domains co-localizing with Myc, physical proximity per se could not predict the presence of translocation hotspots in these cells (Supplementary Fig. 8). Furthermore, a subset of hotspot genes (for example, $Socs1$ or $Dusp4$) associated infrequently with Igh (Fig. 3b, c), and we found no direct correlation between the number of translocations per hotspot and contact frequency with Igh (Spearman's $\rho = 0.1$ ($P = 0.4$); Fig. 3d). As an example, $Tmed8$ and $Dusp4$ genes were rearranged to Igh at roughly equal proportions (38 versus 36 translocations respectively), in spite of the fact that $Tmed8$ was physically associated with Igh ~10 times more frequently than $Dusp4$ in the B cell nucleus (Fig. 3c).

To exclude formally the possibility that Igh and Myc share translocation targets primarily because of shared contacts, we generated two additional 4C-seq libraries using baits specific for $Rac2$ (chromosome 15) and $Rplp2$ (chromosome 7). These genes are highly transcribed in B cells[13], and both interact more frequently with Igh than Myc does, but neither is associated with AID-mediated translocation hotspots (Fig. 3c; see also ref. 9). The interactome of these two genes was then compared to that of Igh in activated B cells. As controls for this analysis, we included the interactomes of Igh from resting B cells, anti-HEL knock-in B cells and MEFs, as well as the $Mycn$ interactome from stimulated B cells. As expected, the Igh interactome was similar in all B-cell types, but not in MEFs where it is not transcribed (Fig. 3e). Similarly, $Mycn$, which is transcriptionally silent in activated B cells, shows 4C-seq profiles with little correlation to Igh (Fig. 3e). Notably, the interactomes of $Rac2$ and $Rplp2$ were significantly more correlated to Igh than was Myc ($P < 1 \times 10^{-5}$ (bootstrapping test);
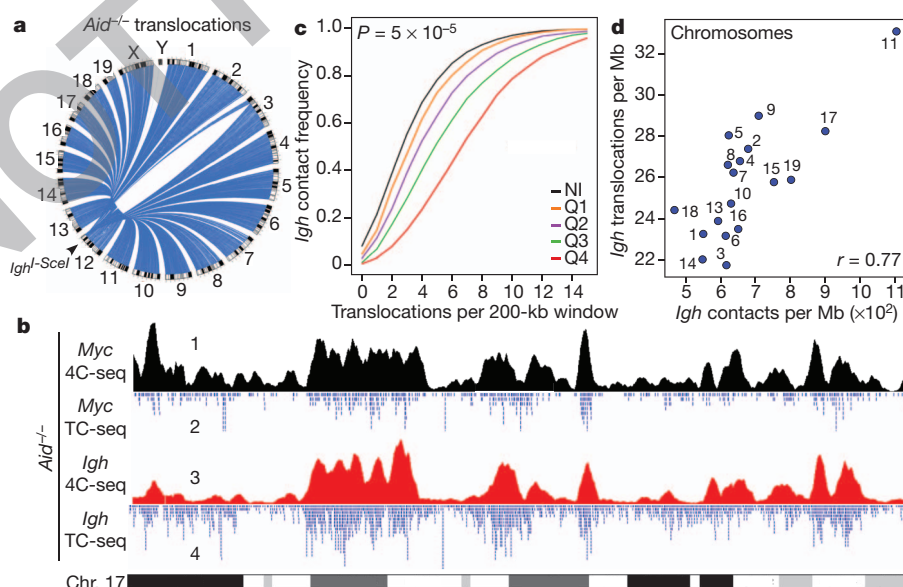

Figure 2 | Genomic distribution of AID-independent translocations correlates with nuclear contact profiles. a, Genome-wide view of rearrangements to Igh^{I-SceI} in $Aid^{-/-}$ B cells. **b**, Cross-comparison of contacts and translocations between Myc or Igh and mouse chromosome 17 in activated $Aid^{-/-}$ B cells. **c**, Empirical cumulative distribution showing Igh^{I-SceI} $Aid^{-/-}$ translocations per 200-kb non-overlapping windows as a function of Igh 4C-seq data subdivided as quartiles (Q1–4). NI represents windows with no aligned reads. **d**, Comparison of AID-independent translocations versus Igh 4C-seq per chromosome per mappable megabase. The degree of correlation is represented by Pearson's r.
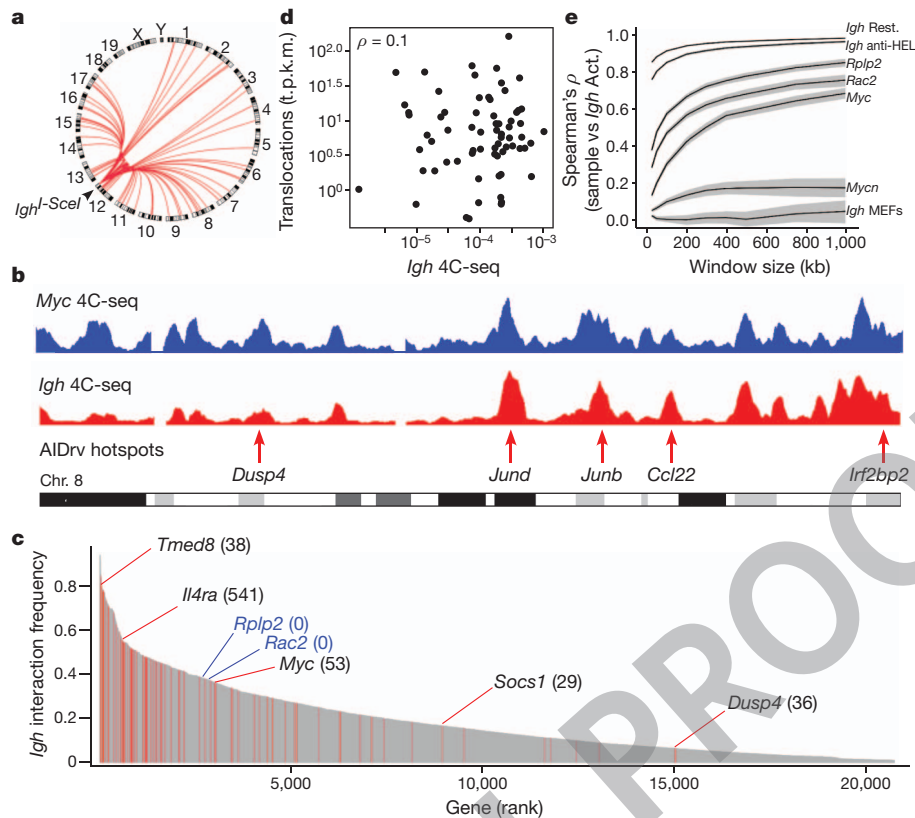
Figure 3 | Lack of correlation between translocation hotspots and nuclear architecture. a, Genome-wide hotspots in activated B cells transduced with I-SceI and AID retroviruses (rv). **b**, Translocation hotspots (bottom lane) and *Myc* and *Igh* contacts in chromosome 8. **c**, *Igh* contacts with RefSeq genes. Hotspot[+] genes are highlighted in red, and for a subset of them the number of translocations is provided in parentheses. The two hotspot[−] genes (*Rplp2* and *Rac2*) highlighted in blue are discussed in more detailed in panel **e**. **d**, Scatter plot showing *Igh* translocations per hotspot versus contacts. Data are plotted as sequence tags per kb per million sequences (t.p.k.m.). **e**, Line graph showing the 4C-seq correlation (Spearman's ρ) between *Igh* (in various cell types), *Rpbp2*, *Rac2*, *Myc* and *Mycn* (in activated B cells) versus *Igh* in activated B cells. The 99% bootstrapping confidence intervals are shown in grey.

Fig. 3e). Thus, nuclear interactions alone do not predispose transcriptionally active genes to high levels of translocation with *Igh*.

A genome-wide map of AID–mediated dsDNA breaks

Our observations challenge the current view that preferential chromosome and/or gene locus interactions govern tumour-inducing translocations in AID-expressing B cells[15,20], and indicate that the amount of AID-mediated DNA damage could account for the frequency of these events. To explore this idea we created a genome-wide map of AID-mediated DNA damage in activated B cells by measuring recruitment of replication protein A[13,21] (RPA) (Fig. 4a, lane 2). Because RPA accumulation is partially blocked by 53BP1 (also called Trp53bp1; refs 22, 23), we reasoned that genetic deletion of 53BP1 in B cells might increase RPA recruitment, thus providing a more sensitive means to map sites of AID-induced lesions by ChIP-seq. Consistent with this idea, RPA accumulation at *Igh* was markedly increased (7.8-fold relative to control) in the absence of 53BP1 (Fig. 4a, lane 3), and an even higher RPA signal (11-fold) was observed in *53BP1*[−/−] mice overexpressing AID (*Igκ*AID[12]; Fig. 4a, lane 4). Conversely, there was no detectable accumulation of RPA at *Igh* in activated *Aid*[−/−]*53BP1*[−/−] B cells (Fig. 4a, lane 1). Thus, RPA recruitment to *Igh* is AID dependent and is enhanced in the absence of 53BP1.

In agreement with our previous findings[13], we did not detect RPA recruitment at AID targets outside the *Igh* locus, such as *Cd83* (Fig. 4b, lane 2). However, we found prominent RPA ChIP signal at the same locus upon *53BP1* deletion (Fig. 4b, lanes 3 and 4). Analogous to *Igμ* and *Igγ1* (also called *Ighg1*), the *Cd83* RPA island extended nearly 50 kb upstream and downstream of the transcription start site (TSS) (Fig. 4a, b, lane 4). In total, 153 non-*Ig* genes accumulated RPA in an AID-dependent fashion (Fig. 4c and Supplementary Table 6) including known *Igh* translocation partners such as *Pax5*, *Pim1* and *Mir155* (Fig. 4d and Supplementary Table 6).

To ascertain the precise nature of RPA islands, we measured somatic hypermutation at a subset of RPA[+] and RPA[−] genes[13]. We found a strong positive correlation between the rate of hypermutation and the extent of RPA recruitment (Spearman's $\rho = 0.71$; Fig. 4e and Supplementary Table 7). We conclude that RPA-seq can be used as a surrogate to measure AID-mediated DNA damage across the B-cell genome.

Recurrent translocations are proportional to DNA damage

To evaluate the relative contribution of DNA damage to targeted translocations we compared the results of RPA-seq and TC-seq obtained from AID-expressing cells. We found a substantial overlap between the two data sets: out of a total of 97 genes with translocation hotspots with an average of 60 translocations per gene, 78 showed RPA accumulation (Supplementary Fig. 9 and Supplementary Table 8). A second group of genes (75) was also associated with RPA islands but displayed fewer translocations (mean = 7, Supplementary Fig. 9), and thus fell below our hotspot criteria cutoff. This result indicates that TC-seq underestimates the number of AID-mediated translocations, possibly due to lack of saturation[9]. Only 19 genes associated with translocation hotspots did not recruit RPA above background levels (Supplementary Fig. 9), suggesting that the RPA-seq data set is also not fully saturated. Thus, RPA demarcates sites of recurrent translocations in B lymphocytes.

In addition to the qualitative correlation above, we found that the absolute number of *Igh* translocations per hotspot was directly proportional to RPA recruitment (Spearman's $\rho = 0.6$ ($P = 2.9 \times 10^{-6}$);
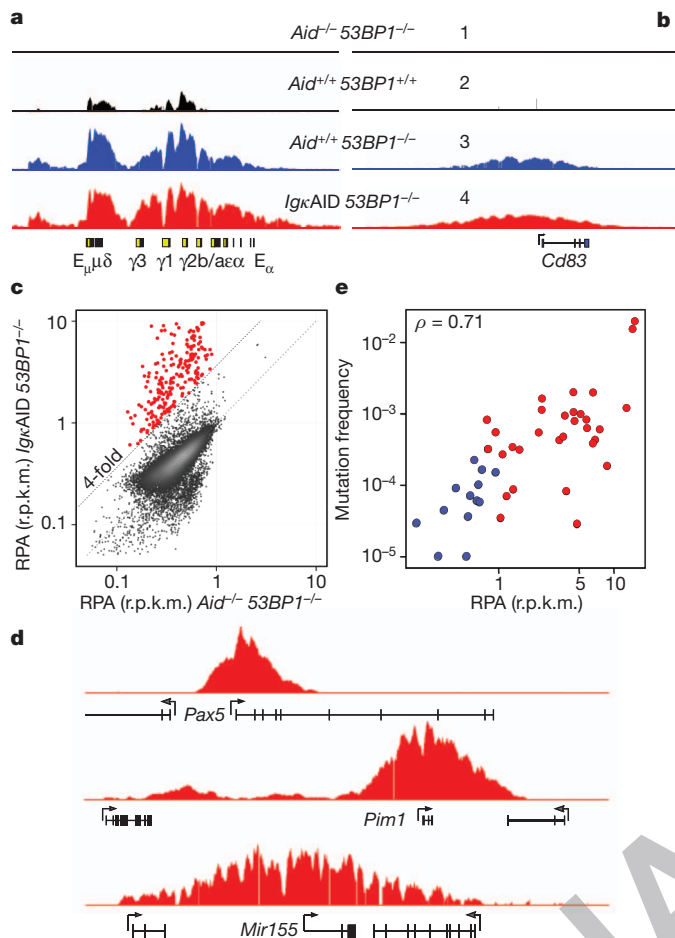
Figure 4 | Genome-wide map of AID-mediated DNA damage. a, RPA occupancy at *Igh* in activated B cells. Genotype and RPA sequence reads per million values are shown. **b**, Same analysis as in panel **a** for *Cd83*. For all tracks, background sequencing was filtered out via a threshold. **c**, One-hundred and fifty-three RPA islands (red dots) detected in *IgκAID 53BP1$^{-/-}$* B cells fourfold above background (measured in *Aid$^{-/-}$ 53BP1$^{-/-}$* cells). Data are plotted as reads per kb per million sequences (r.p.k.m.). **d**, RPA islands associated with TSSs from *Pax5*, *Pim1* and *Mir155*. **e**, Hypermutation frequency relative to RPA recruitment at TSSs (±2 kb) in a subset of RPA$^+$ (red dots) and RPA$^-$ (blue dots) genes. Spearman's ρ is provided.

Figure 5 | AID activity predicts the location and frequency of targeted chromosomal translocations. a, Scatter plot showing the correlation between *Igh* translocations per hotspot and RPA recruitment. **b**, Same as panel **a** but *Igh* contact frequency is used instead of translocations. **c**, Upper schematic: distribution of RPA islands (red dots) and translocation hotspots (blue dots) in chromosome 7. Middle: *Igh* 4C-seq profile demarcating the *Nsmce1-Il4ra-Il21r* loci. Bottom: RPA islands, translocations and contact frequency for each gene.

Fig. 5a). This result contrasts with the lack of correlation observed between nuclear contacts and total rearrangements per hotspot (Fig. 3d) or RPA accumulation (Spearman's ρ = 0.03 (P = 0.8); Fig. 5b). Similar results were observed for *Myc* (Spearman's ρ < 0.07; data not shown). These findings demonstrate that with regard to AID-targeted translocations, DNA damage is the primary determinant of rearrangement location and frequency. This was particularly evident for AID targets that are clustered within ~200-kb genomic domains, such as *Nsmce1*, *Il4ra* and *Il21r* in chromosome 7 (Fig. 5c), or the *Hist1h1c* gene family in chromosome 13 (Supplementary Fig. 10 and Supplementary Table 9). Whereas we found little variation in *Igh* (or *Myc*) proximity for different genes within these clusters, translocations varied substantially and in a manner that was proportional to AID-mediated damage (Fig. 5c, Supplementary Fig. 10 and Supplementary Table 9). Taken together, these results clearly demonstrate that DNA break formation, but not nuclear interactions, governs the rate of recurrent chromosomal translocations.

Discussion

We have shown that in the absence of programmed DNA damage, translocation partner selection is largely dictated by physical proximity, following principles of nuclear organization, chromatin accessibility
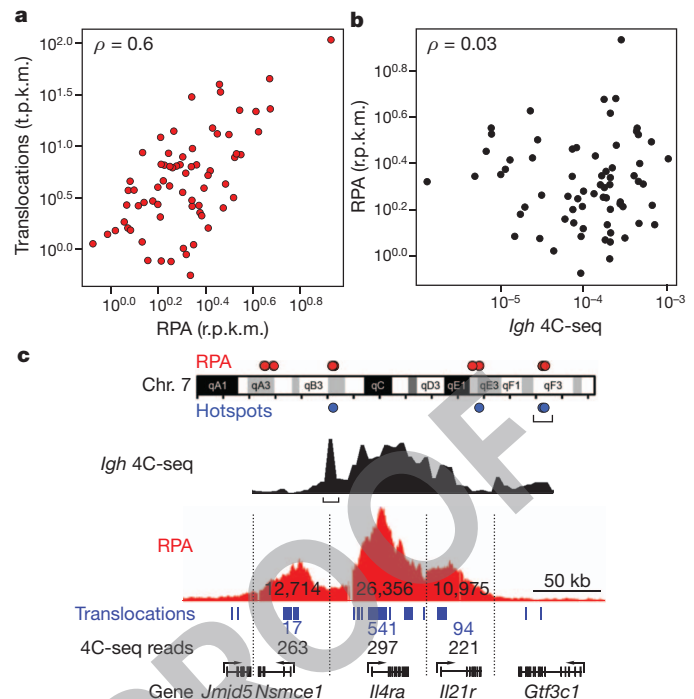
and gene expression. Physical proximity has also been suggested to have an impact on the formation of recurrent translocations[20]. For instance, rearrangements between *BCR–ABL1* in chronic myeloid leukaemia, *RET–CDC6* in thyroid malignancies, *TMPRSS2–ERG/ETV1* in prostate cancer, and *PML–RARA* in acute promyelocytic leukaemia have all been ascribed to preferential interactions between translocating partners in tumour cell precursors[1]. Similarly, *Igh* and *Myc* chromosomes have been shown to associate in mouse and human B cells[16,17], and RNA FISH has shown that the *Myc* and *Igh* alleles are frequently found in the same RNA Pol-II-enriched transcription factories[15]. Bystander translocations between *Igh* and *Igλ* have also been proposed to result from frequent contacts, as determined by 3D DNA FISH[24]. One limitation of FISH technology however is that it can only monitor a limited number of loci simultaneously. Consequently, it has been difficult to ascertain whether the documented contacts are truly unique relative to the broad array of genomic interactions. Our 4C measurements now clarify this issue in that they show that 29% of all genes interact with *Igh* at equal or higher frequency than *Myc*, *Igλ1*, or many of the oncogenes frequently rearranged in B-cell tumours. Thus, the rate of interaction between *Igh* and its recurrent translocation partners is not a specific feature of these loci and cannot account for their high rate of translocation.

Translocation requires joining of two double-stranded breaks (DSBs). Therefore, when breaks are limiting, increasing their frequency increases the rate of translocation[24,25]. Similarly, repair deficiencies augment the rate of translocation by increasing the half-life of dsDNA breaks and thereby the availability of substrates for aberrant repair[2]. However, it has not been possible to relate directly the frequencies of DNA damage and translocations because neither the extent nor the location of DSBs in the B-cell genome was known. We have overcome these limitations by measuring RPA deposition at sites of DNA damage in 53BP1 mutant B cells. On the basis of this new strategy we uncovered ~150 non-*Ig* genes that suffer AID-mediated DSBs. These genes

coincide with translocation hotspots, and are sites of ongoing hypermutation. Most importantly, by relating RPA-seq with TC-seq and 4C-seq data sets we found that the frequency of DNA damage directly accounts for the rate of translocation, as shown by the marked concordance between the amount of RPA deposition and the absolute number of rearrangements at any given genomic site. This view is also supported by the relative lack of correlation between proximity and the absolute number of rearrangements per hotspot.

The genomic distribution of sporadic translocations is best explained by nuclear architecture, whereas the location and incidence of recurrent translocations, including those involved in B-cell malignancies, directly reflect site-specific DNA damage.

METHODS SUMMARY

Full details of B-cell culture, hypermutation analysis, chromatin immunoprecipitation, chromosome conformation capture on Chip (4C), translocation capture sequencing analysis, deep sequencing and bioinformatics techniques are provided in Methods. The NIAMS-NIH Animal Care and Use Committee approved all animal protocols and experiments.

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

1. Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nature Rev. Cancer* **7,** 233–245 (2007).
2. Nussenzweig, A. & Nussenzweig, M. C. Origin of chromosomal translocations in lymphoid cancer. *Cell* **141,** 27–38 (2010).
3. Tsai, A. G. *et al.* Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell* **135,** 1130–1142 (2008).
4. Zhang, Y. *et al.* The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. *Adv. Immunol.* **106,** 93–133 (2010).
5. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2,** a003889 (2010).
6. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
7. Hakim, O., Sung, M. H. & Hager, G. L. 3D shortcuts to gene regulation. *Curr. Opin. Cell Biol.* **22,** 305–313 (2010).
8. Chakalova, L. & Fraser, P. Organization of transcription. *Cold Spring Harb. Perspect. Biol.* **2,** a000729 (2010).
9. Klein, I. A. *et al.* Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* **147,** 95–106 (2011).
10. Chiarle, R. *et al.* Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147,** 107–119 (2011).
11. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genet.* **38,** 1348–1354 (2006).
12. Robbiani, D. F. *et al.* AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol. Cell* **36,** 631–641 (2009).
13. Yamane, A. *et al.* Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nature Immunol.* **12,** 62–69 (2011).
14. Hakim, O. *et al.* Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. *Genome Res.* **21,** 697–706 (2011).
15. Osborne, C. S. *et al.* Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol.* **5,** e192 (2007).
16. Roix, J. J., McQueen, P. G., Munson, P. J., Parada, L. A. & Misteli, T. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nature Genet.* **34,** 287–291 (2003).
17. Parada, L. A., McQueen, P. G. & Misteli, T. Tissue-specific spatial organization of genomes. *Genome Biol.* **5,** R44 (2004).
18. Kuchen, S. *et al.* Regulation of microRNA expression and abundance during lymphopoiesis. *Immunity* **32,** 828–839 (2010).
19. Brown, K. E., Baxter, J., Graf, D., Merkenschlager, M. & Fisher, A. G. Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol. Cell* **3,** 207–217 (1999).
20. Meaburn, K. J. & Misteli, T. Cell biology: chromosome territories. *Nature* **445,** 379–381 (2007).
21. Vuong, B. Q. *et al.* Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nature Immunol.* **10,** 420–426 (2009).
22. Bunting, S. F. *et al.* 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* **141,** 243–254 (2010).
23. Bothmer, A. *et al.* 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J. Exp. Med.* **207,** 855–865 (2010).
24. Wang, J. H. *et al.* Mechanisms promoting translocations in editing and switching peripheral B cells. *Nature* **460,** 231–236 (2009).
25. Robbiani, D. F. *et al.* AID is required for the chromosomal translocations in c-myc that lead to c-myc/IgH translocations. *Cell* **135,** 1028–1038 (2008).

Author Contributions R.C., O.H., G.L.H. and M.C.N. planned studies and interpreted data. Experiments were performed as follows: O.H. and C.A.-S., 4C-seq and FISH; A.Y., RPA-seq; I.K., A.B., D.F.R. and M.J., TC-seq; W.R., E.M. and T.O., bioinformatics; K.-R.K.-K., T.C.V., H.N. and J.C., FISH; G.L. and H.N., hypermutation; A.N., 53BP1 expertise; M.C.N. and R.C. wrote the manuscript.

Author Information All sequence data are available at the NCBI SRA database under accession number SRP010565. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.C. (casellar@mail.nih.gov) or M.C.N. (nussen@rockefeller.edu).

METHODS

B-cell activation and hypermutation analysis. Miltenyi microbead-isolated CD43⁻ splenic B cells from wild-type, *Igk*AID-*Ung*$^{-/-}$, or *Aicda*$^{-/-}$ mice were cultured at 0.1×10^6 cells per ml with $50\,\mu g\,ml^{-1}$ lipopolysaccharide (LPS) (Sigma), $2.5\,ng\,ml^{-1}$ mouse recombinant IL-4 (Sigma) and $0.5\,\mu g\,ml^{-1}$ of aCD180 (RP105) antibody (RP/14, BD Pharmingen). For 4C-seq and TC-seq procedures, cells were collected at 72 h. For hypermutation analysis, cells were diluted 1:4 at 72 h and cultured for another 48 h under the same conditions. Fifty nanograms of genomic DNA was then amplified for 30 cycles with Phusion DNA polymerase and gene-specific primers. When nested PCR was applied, 40 (20 + 20) cycle amplifications were performed in the presence of DMSO. The amplicon was cloned using PCR Zero blunt (Invitrogen) and sequenced.

Chromosome conformation capture on chip (4C) followed by deep-sequencing. The 4C assay was performed as previously described[14] with minor modifications. Ten million mouse B cells were crosslinked in 2% formaldehyde at 37 °C for 10 min. The reaction was quenched by the addition of glycine (final concentration of 0.125 M). Cells were then washed with cold PBS and lysed (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.2% NP-40, 1× complete protease inhibitors (Roche)) at 4 °C for 1 h. Nuclei were incubated at 37 °C for 1 h in 500 μl of restriction buffer (New England Biolabs buffer 2 for HindIII or buffer 3 for BglIII digestion) containing 0.3% SDS. To sequester SDS, Triton X-100 was then added to a final concentration of 1.8%. DNA digestion was performed with 400 U of HindIII or BglII (New England Biolabs) at 37 °C overnight. After heat inactivation (65 °C for 30 min), the reaction was diluted to a final volume of 7 ml with ligation buffer containing 100 U T4 DNA Ligase (Roche) and incubated at 16 °C overnight. Samples were then treated with 500 μg Proteinase K (Ambion) and incubated overnight at 65 °C to reverse formaldehyde crosslinking. DNA was then purified by phenol extraction and ethanol precipitation. For circularization, the ligation junctions were digested with Csp6I (Fermentas) or DpnII (New England Biolabs) at 37 °C overnight. After enzyme inactivation and phenol extraction, the DNA was religated in a 7-ml volume (1,000 U T4 DNA Ligase, Roche). Three micrograms of 4C library DNA was amplified with Expand Long template PCR System (Roche). Thermal cycle conditions were DNA denaturing for 2 min at 94 °C, followed by 30 cycles of 15 s at 94 °C, 1 min at 60 °C, 3 min at 68 °C, and a final step of 7 min at 68 °C. Baits were amplified with inverse PCR primers as follows: *Igh* with HindIII: IgH_R_4C 5′-CCAGACATGTGG GCTGAGAT-3′, Igh_Hind_Read 5′-CTACCCACCTAACTCCAAGC-3′; *Mycn* with HindIII: Mycn_R_4C 5′-CTCCCATTTTGCACTCCTGT-3′, Mycn_Hind_Read 5′-GATTTATCCTTAAACCCTTAAGC-3′; *Igh* with BglII: IgH_Bgl_R_4C 5′-CATGGACATTTGCGTGTGTA-3′, IgH_Bgl_Read 5′-GTG CCCCCAGGAGCAGATCT-3′; *Mycn* with BglII: Mycn_Bgl_R_4C 5′-AG TCTGCGGGAGGTAAGAAG-3′, Mycn_Bgl_Read 5′-CCCTTTTAGACAGCC AGATCT-3′; *Myc* with BglII: Myc_Bgl_R_4C 5′-AAGAATGTGCCCAGTC AACA-3′, Myc_Bgl_Read 5′-AGTGAATTGCCAACCCAGAT-3′; *Rplp2* with HindIII: 5′-GCCATCTCTCCAGTCAAAAAGC-3′, CTTCTCACTTCCATT CCCTGAG-3′; Rac2 with HindIII: 5′-GCCATGGAGACCGGAAGCTT-3′, 5′-GGGACTGTCCACTCCACCT-3′; Eμ with HindIII: 5′-TGTGGCTGCTGC TCTTAAAGC-3′, 5′-TGTGAAGCCGTTTTGACCAGAATGT-3′. 4C amplified DNA was microsequenced with the Illumina platform. For multiplexing purposes, extra nucleotides were added to the 5′ end of read primers: a T for LPS + IL-4 activated B cells, TT for resting B cells, A for IgH$^{\alpha HEL/\alpha HEL}$ B cells, and AA for MEFs.

Translocation capture sequencing. All experimental procedures involving translocation capture sequencing (TC-seq) library preparation and computational analysis is provided in ref. 9.

Bioinformatics. For 4C-seq, standard Illumina pipeline software (version ≥ 1.8) was used to process raw data and obtain Fastq files of paired short reads. Each read pair was then tested for the presence of a perfect match to the respective bait primer as well as the bait spacer between the end of the primer and the restriction sites used in the corresponding experiment. In some experiments up to three mismatches in the non-index portion of the PCR primer/flank sequence was allowed without any relevant changes in the resulting data. These flanking sequences other than the restriction site were trimmed and the remainder was aligned against the mouse genome (build mm9/NCBI37) with Bowtie[26] with the following command line options: '-X 500 -p 3 -v2 -k2 -m1 –phred64-quals –sam', which reported all unique alignments with at most two mismatches. In the case of Eμ, only the HindIII spanning read in each pair was sufficiently long to reach the ligated interaction partner. In that case, single end alignments of the flank were carried out with the command line options '–best–all–strata–chunkmbs 256 -m1 –sam'. Alignments were then matched up with restriction sites and assigned to a HindIII or BglII fragment. Fragments were combined into 200-kb non-overlapping windows to determine (1) the

total number of 4C reads per fragment, and (2) the fraction of restriction fragments for which 4C reads were found. The latter part of the analysis was carried out with a combination of custom software written in Bash, Python, R, and BedTools[27]. Processing of fragment- or window-level data was carried out in R using standard methods.

For RPA-seq, short reads obtained from Illumina pipeline were aligned against the mouse genome (build mm9/NCBI37) with Bowtie[26] using options '–threads = 8–phred64-quals–best–all–strata -m1 -n2 -l36 –sam'. Raw tag densities of *Igk*AID *53BP1*$^{-/-}$ and *Aid*$^{-/-}$ *53BP1*$^{-/-}$ samples flanking transcription start sites of RefSeq genes were compared and genes with >4-fold enrichment were selected.

For histone acetylation, short reads obtained for ChIP against a set of histone acetylations (see below) were aligned above. Areas of local enrichment over a random background model (islands) were identified with SICER 1.03[28] and the density of reads overlapping these islands was averaged in the same 200-kb non-overlapping windows used for 4C analysis.

High-throughput 3D DNA FISH. For 3D FISH, cultured B cells were set in a 384-well, poly-D-lysine-coated microplate (Perkin Elmer) by centrifugation at 1,000 r.p.m. for 5 min. After fixation in 4% PFA for 10 min, and permeabilization in 0.5% saponin (Sigma Aldrich)/0.5% Triton X-100/PBS for 20 min, cells were incubated in 0.1 N HCl for 10 min. Two PBS washes were applied between each step. After a 2× SSC wash, cells were kept in 50% formamide/2× SSC buffer for at least 30 min. Bacterial artificial chromosomes (BACs) were used as probes as follows: *Mlh3*, RP24-139J8; *Klhdc1*, RP24-109D18; *Clec2d*, RP24-149B3; *Mir142*, RP24-376D9, *Il4ra*, RP23-60A3; *Cyth1*, RP23-267B12; *Pim1*, RP24-331E7; *Furin*, RP24-377F13; *Pax5*, RP23-258E20; *Myc*, RP24-297E9; *Cxcr5*, RP24-308P6; *Rasa3*, RP24-247P3; *Gata3*, RP24-402N11. For each BAC, single colonies were grown and the presence of BAC DNA was verified by PCR. DNA was isolated and labelled with biotin (Roche; Biotin-Nick translation mix) or digoxigenin (Roche; DIG-Nick translation mix). A probe mix containing 250 ng of digoxigenin- and biotin-labelled probes, 3 μg mouse COT1 DNA (Invitrogen), and 20 μg tRNA (Ambion) was ethanol precipitated, and re-suspended in 15 μl of hybridization buffer (10% dextran sulphate, 50% formamide, 2× SSC, and 1% Tween 20). Probe was then added to each well, denatured together with nuclei at 85 °C for 7 min and left to hybridize at 37 °C overnight in a humidified chamber. To discard non-hybridizing probe, cells were then washed in 50% formamide with 2× SSC at 45 °C, followed by washes with 1× SSC at 60 °C. Each wash was repeated three times for 5 min. Cells were blocked with 3% BSA/0.05% Tween 20/4× SSC for 20 min at room temperature and then incubated for 1 h with Fluorescein Avidin (Vector) and anti-Dig-Rhodamin (Roche) diluted 1:200 in blocking solution. Next, cells were washed three times with 0.05% Tween 20/4× SSC and mounted in DAPI-containing Vectashield (Vector) for imaging.

Microscopy. Cells were imaged in 384-well plates with the Opera (Perkin Elmer) confocal high-throughput imaging system using a ×40 water objective lens with 9 optical steps of 1.0 μm.

Automated image analysis algorithms. To quantify distances between FISH signals, we customized the automated image analysis computer algorithm from ref. 14 to allow analysis of Opera images. This algorithm determines centre-to-centre distances of FISH signals in 3D. Nuclei and DNA FISH loci were automatically identified by a combination of the Acapella image analysis software (Perkin Elmer) together with a series of custom algorithms developed using Matlab technical computing software and the Matlab Image Processing toolbox (The Mathworks). The resulting morphometric information from each nucleus was automatically stored in a Matlab database file, which could be accessed by the custom algorithms. First, nuclei region of interest (nucROI) were segmented using the DAPI signal channel by the Acapella software. Then DNA FISH loci (fishROI) were automatically identified by a separate algorithm in the nucROI positional information. FishROIs were identified by intensity-based thresholding of the FISH fluorescent channel images. A third algorithm assigned the nuclear position of each FISH locus, based on the x–y–z location of the centre of the brightest 3×3 FISH fluorescent channel pixel that was located within the fishROI. For each nucleus, three-dimensional (x–y–z) distances were calculated between all possible pairs of FISH loci. The closest distance between the two probes in each nucleus was used to calculate the frequency of cells with distances <1 μm as previously described[14]. Interaction frequencies between each of the genes and *Igh* was calculated by computing the percentage of cells carrying at least one pair of FISH signals separated by distances smaller than 1 μm (Supplementary Fig. 12a). This distance threshold was recently shown to correlate well with 4C contact frequencies[14]. To determine the minimum number of cells necessary to reach statistical significance in our experimental setting, we first examined 17,462 LPS plus IL-4 activated lymphocytes with *Igh*- and *Myc*-specific probes. We found that at least 2,000

cells were necessary to reach a standard error of 0.01 (Supplementary Fig. 12b). For different probes, the optimal number of cells was as follows: 2,200 for *Myc*, 4,300 for *Mlh3*, 2,100 for *Clec2d*, 3,800 for *Klhdc1*, 1,500 for RP24-137A18, 1,300 for *Gata3*, 2,200 for *Il4ra*, 1,900 for *Mir142*, 2,200 for *Cyth1*, 2,100 for *Pim1*, 2,100 for *Furin*, 1,900 for *Pax5*, 1,800 for *Cxcr5* and 1,300 for *Rasa3*.

26. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
27. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
28. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25,** 1952–1958 (2009).

NEWS & VIEWS

CLIMATE CHANGE

Shrinking glaciers under scrutiny

Melting glaciers contribute to sea-level rise, but measuring their mass loss over time is difficult. An analysis of satellite data on Earth's changing gravity field does just that, and delivers some unexpected results.
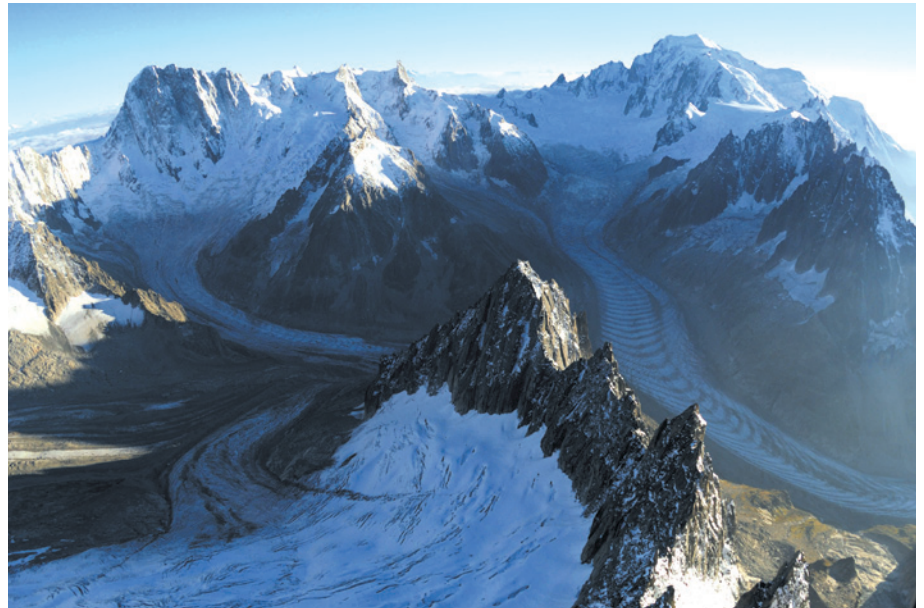
JONATHAN BAMBER

Glaciers and ice caps are pivotal features of both water resources and tourism. They are also a significant contributor to sea-level rise. About 1.4 billion people are dependent on the rivers that flow from the Tibetan plateau and Himalayas[1]. Yet significant controversy[2] and uncertainty surround the recent past and future behaviour of glaciers in this region. This is not so surprising when one considers the problem in hand. There are more than 160,000 glaciers and ice caps worldwide. Fewer than 120 (0.075%) have had their mass balance (the sum of the annual mass gains and losses of the glacier or ice cap) directly measured, and for only 37 of these are there records extending beyond 30 years. Extrapolating this tiny sample of observations to all glaciers and ice caps is a challenging task that inevitably leads to large uncertainties.

In an article published on *Nature*'s website today, Jacob and colleagues[3] describe a study based on satellite data for Earth's changing gravity field that tackles this problem. Their results have surprising implications for both the global contribution of glaciers to sea level and the changes occurring in the mountain regions of Asia.

Melting glaciers are an iconic symbol of climate change. On the basis of the limited data mentioned above, they seem to have been receding, largely uninterrupted, almost everywhere around the world for several decades[4]. Scaling up the small sample of ground-based observations to produce global estimates is, however, fraught with difficulty. Size, local topography, altitude range, aspect and microclimate all affect the response of individual glaciers in complex ways. Even the seasonality of changes in temperature and precipitation strongly influence the glaciers' response, and those that terminate in a lake or ocean behave differently again.

Nonetheless, until recently there was little alternative to some form of extrapolation of the terrestrial observations to large regions and numbers of glaciers. One such high-profile assessment[5] concluded that, during the period 1996–2006, the mass loss from glaciers and ice caps (GICs) increased steadily, contributing a sea-level rise of 1.1 ± 0.24 millimetres

Figure 1 | The Leschaux and Talèfre glaciers in the French Alps. The photograph highlights the complex and intricate topographic setting of these mountain glaciers and the difficulty in extrapolating observations from one glacier to others. Jacob and colleagues[3] avoided these difficulties by using the area-integrated signal from satellite gravity data.

per year by 2006. In this study[5], the authors concluded that GICs had been the dominant mass contributor to sea-level rise over the study period, and they extrapolated their results forward to argue that this would also be the case in the future.

Then along came the Gravity Recovery and Climate Experiment (GRACE), which consists of a pair of satellites that have been making global observations of changes in Earth's gravity field since their launch in 2002. They have been used in various studies to examine the changing mass of the great ice sheets of Antarctica and Greenland[6] and several other large glaciated regions[7]. But, so far, the data have not been analysed simultaneously and consistently for all areas.

The difficulty with doing this is that GRACE measures the gravity field of the complete Earth system. This includes mass exchange and/or mass redistribution in the oceans, atmosphere, solid Earth and land hydrology, in addition to any changes in GIC volume. To determine the latter, it is clearly essential to be able to separate it from the other sources of

mass movement that affect the gravity field. A second, related issue is the effective resolution of the observations. The GRACE satellites are sensitive to changes in the gravity field over distances of a few hundred kilometres. They cannot 'see' the difference between the signal from one glacier or small ice cap and another.

To isolate the GIC signal from others at the surface, Jacob and colleagues defined units of mass change — called mass concentrations, or mascons — within each of their 18 GIC regions (including the European Alps; Fig. 1). Each region might have many tens of mascons defining the geographic extent of significant ice volume within the sector[3]. Combined with global models of land hydrology and atmospheric-moisture content, the authors were able to isolate the GIC mass trends over the eight-year (2003–10) period of the observations. What they found was unexpected.

First, the contribution of GICs (excluding the Antarctica and Greenland peripheral GICs) to sea-level rise was less than half the value of the most recent, comprehensive estimate[8] obtained from extrapolation of *in situ*

measurements for 2001–05 (0.41 ± 0.08 compared with 1.1 mm yr^{-1}). Second, losses for the High Mountain Asia region — comprising the Himalayas, Karakoram, Tianshan, Pamirs and Tibet — were insignificant. Here, the mass-loss rate was just 4 ± 20 gigatonnes per year (corresponding to 0.01 mm yr^{-1} of sea-level rise), compared with previous estimates that were well over ten times larger. By a careful analysis, the authors discounted a possible tectonic origin for the huge discrepancy, and it seems that this region is more stable than previously believed.

What is the significance of these results[3]? Understanding, and closing, the sea-level budget (the relative contributions of mass and thermal expansion to ocean-volume change) is crucial for testing predictions of future sea-level rise. Estimates of the future response of GICs to climate change are, in general, based on what we know about how they have responded in the past. A better estimate of past behaviour, such as that obtained by Jacob and colleagues, will therefore result in better estimates of future behaviour. Discussion of the demise of the Himalayan glaciers has been mired in controversy, partly because of basic errors[2], but also because of the dearth of reliable data on past trends. Given their role as a water supply for so many people[1], this has been a cause for concern and an outstanding issue.

Of course, eight years is a relatively short observation period. Some of the regions, such as the Gulf of Alaska, experience large inter-annual variations in mass balance that are mainly due to variability in precipitation[7]. This is also true for the High Mountain Asia region[3], and, as a consequence, a different measurement period could significantly alter the estimated trend for this sector. Furthermore, some areas, such as the European Alps and Scandinavia, have been relatively well monitored, and thus constrained, using other approaches. Nonetheless, Jacob and colleagues have dramatically altered our understanding of recent global GIC volume changes and their contribution to sea-level rise. Now we need to work out what this means for estimating their future response. ∎

Jonathan Bamber *is at the Bristol Glaciology Centre, School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK.*
e-mail: j.bamber@bristol.ac.uk

1. Immerzeel, W. W., van Beek, L. P. H. & Bierkens, M. F. P. *Science* **328,** 1382–1385 (2010).
2. Cogley, J. G., Kargel, J. S., Kaser, G. & van der Veen, C. J. *Science* **327,** 522 (2010).
3. Jacob, T., Wahr, J., Pfeffer, W. T. & Swenson, S. *Nature* http://dx.doi.org/10.1038/nature10847 (2012).
4. Kaser, G., Cogley, J. G., Dyurgerov, M. B., Meier, M. F. & Ohmura, A. *Geophys. Res. Lett.* **33,** L19501 (2006).
5. Meier, M. F. *et al. Science* **317,** 1064–1067 (2007).
6. Velicogna, I. *Geophys. Res. Lett.* **36,** L19503 (2009).
7. Luthcke, S. B., Arendt, A. A., Rowlands, D. D., McCarthy, J. J. & Larsen, C. F. *J. Glaciol.* **54,** 767–777 (2008).
8. Cogley, J. G. *Ann. Glaciol.* **50,** 96–100 (2009).