Time to open up

If scientists want the public to continue to volunteer for research projects, they must learn to be a lot more forthcoming about the ways in which the information they garner will be used.

Few biomedical researchers who work with human volunteers are entirely happy with the rules that govern and preserve research ethics. Scientists working on an experiment or a clinical trial can find that the complex web of regulations that they must follow is a confusing burden. These rules, they feel, simply add time, cost and complication to their studies.

Institutions are wary of falling foul of regulations that even highly trained staff are hard-pressed to interpret. And research participants — the volunteers without whom research such as clinical trials would not be possible — often don't understand all the implications contained in the voluminous documents that they must sign before they take part in the research. Although these forms are ostensibly designed to protect participants, they are more often directed at protecting institutions from liability.

The notion of informed consent has become especially problematic in this modern age of 'big data'. The need to seek and obtain permission from volunteers can be traced back to research-ethics principles introduced in response to revelations about the way Nazi doctors tortured people during the Second World War. Decades ago, researchers could not glean much information from a piece of stored tissue. But now, it is no longer possible to explain to a research subject all the possible ways in which their offered data might be used in the future, and it is becoming less feasible to guarantee their privacy under (correct) rules that mandate the deposition of research data in public databanks. Some large data-collection projects have therefore adopted very broad informed-consent provisions, in which research participants simply agree that their data may be used for a wide and unspecified variety of purposes. But as potential volunteers become savvier to the value and vulnerability of their personal data, they may become less receptive to this approach. Indeed, as our feature on page 312 shows, this may already be happening.

Technology and some creative thinking should be able to provide a solution to this problem. In an era when people can control who sees what types of personal information on their Facebook page, programmers should be able to design similar tools to give research volunteers a degree of influence over who uses their data and for which types of research. Experiments along these lines are now being tried by the company Private Access, based in Irvine, California.

Scholars of ethics and law are also trying to think of new models for informed consent that could accommodate the needs of both researchers and research participants. For instance, a group based at Duke University in Durham, North Carolina, and the University of North Carolina at Chapel Hill has proposed remodelling consent provisions based on laws that protect trade secrets. In its favour, this model would provide the option of giving donors something in return for handing over their samples to a large data collection, such as a biobank. It would ask the donors what they want in return for the information embedded in the samples, such as financial compensation (which many already

receive) or a say in the types of research their information is used for.

Other researchers have proposed reforms that head in the opposite direction — for instance, by relaxing the rules on consent for research involving 'de-identified' specimens, which are stripped of information that could link them back to their source.

This approach may seem attractive because it would seem to lessen cost and burden on the researchers, and could make the most of currently under-exploited medical records and data samples. Extreme caution is needed, however. It would take only one high-profile case of an unwanted data disclosure to undermine support for research — and not just for the field in question. Researchers need only look to the backlash against newborn screening in the United States and tissue research in the United Kingdom to realize how failing to obtain consent can set back their cause.

"Scholars of ethics and law are trying to think of new models for informed consent."

As this journal has argued previously, a more appropriate solution to the conundrum of informed consent is to introduce greater openness to the process.

Scientists could agree to give results containing medically relevant findings back to participants in genetic studies, such as information about their health. Although some have argued that this approach, as well as patient-controlled approaches, will add time, cost and complication to studies, no one really knows if this is the case. Such concerns must not be allowed to derail the idea.

Most large studies are funded by taxpayers or patient-advocacy groups. Researchers are therefore obliged to listen when patients and members of the public argue that they want to have more information — not less — to ensure that they agree to offer their continued participation in research. It might not make everybody happy, but it should keep everybody involved. ∎

Renewed vigour

Stem-cell researchers must engage with politicians to keep their work alive in Europe.

Research involving human embryonic stem (ES) cells is once more under scrutiny in Europe. In a situation that will stir memories of the acrimonious debates of 2006, legislators must again assess whether this kind of work should still be funded under the forthcoming €80-billion (US$100-billion) Horizon 2020 research programme.

Now, as then, opinion is split. Some countries and some members of

the European Parliament are in favour of the research, recognizing the long-term potential to treat debilitating disease. Others maintain that it is immoral to exploit a technique that uses human embryos — even spare embryos from *in vitro* fertilization clinics that would otherwise be destroyed, and from which nearly all experimental human ES cell lines are derived. The 2006 debate was resolved with both the European Parliament and Council agreeing to fund such research, provided that it didn't involve the creation of new human ES cells, and provided that it was not carried out in those countries — such as Germany — whose national law banned it.

Stem-cell researchers around the world breathed a sigh of relief, knowing how a decision in Europe could also influence funding decisions elsewhere. The outcome of the present debate offers similar influence. But there are already signs that some members of the European Parliament will once again try to outlaw funding of research involving human ES cells.

Rather than wait for these views to gain unchecked momentum, a group including UK research-funding bodies the Wellcome Trust, the Medical Research Council, the British Heart Foundation and Parkinson's UK last week issued a joint statement in support of the research, explaining the benefits of the work and the rationale to include it in Horizon 2020.

It is a wise move that should help to anchor the coming debate to reality. Biology is complicated, which makes it easy for politicians to mislead the public and colleagues — intentionally or unintentionally — in emotive areas. The general public, and politicians, have every right to question whether the ends justify the means used by medical researchers. But they also have the right to reliable information.

The statement outlines the remarkable progress that stem-cell researchers have made since 2006, which has led in the past 12 months to the first approvals for clinical trials of potential therapies involving human ES cells — for a type of blindness called macular degeneration and for spinal-cord injury. In addition, scientists have discovered how to force adult cells back to an embryonic-like state. The resulting induced pluripotent stem (iPS) cells can then be grown into particular cell types and used to understand mechanisms of disease at the cellular level. In the long term, they may also be useful for therapy.

The most insidious claim of those who oppose human ES cell research holds that iPS cells, which can be derived from a particular patient's own cells and are ethically unburdened, eliminate the need for human ES cells in research and therapy. That concept sounds appealing, but it is simply not true. Scientists understand little of the differences between the two sorts of stem cells and it will take years of comparative work to do so.

> "The general public, and politicians, have every right to question whether the ends justify the means used by medical researchers."

One particular event that makes biomedical researchers worry is a decision taken last year by the European Court of Justice. In October it ruled that patenting of inventions involving human ES cells was illegal because it was immoral — and as a consequence, human ES cell research must also be immoral (see *Nature* **480,** 310–312; 2011). *Nature* condemned this ruling as being beyond the court's juridical and technical competencies (see *Nature* **480,** 291–292; 2011). But members of the European Parliament who oppose human ES cell research stealthily inserted a reference to it during an unrelated resolution on broad patenting of essential biological processes in animal and plant breeding, which was adopted on 10 May.

Horizon 2020 has to be approved by the European Parliament and Council by mid-2013 so that first calls for proposals can be launched at the start of 2014. The UK research funding agencies' statement is a good start to a continuous campaign of education and transparency that stem-cell researchers from all interested European countries must maintain for the next year. Just as cell-culture medium needs to be renewed to keep its cargo alive, a political message has to be constantly renewed if it is to stay alive in political minds. ∎

Serious questions

Nature Publishing Group's reader survey on lab–safety practices needs your input.

Scientific laboratories are dangerous places. Noxious chemicals, naked flames and nasty microbes abound. The white laboratory coat, a long-standing symbol of science to many outsiders, offers some protection against these implicit threats. White coats are ubiquitous in fictional labs in films and on television, but how many lab scientists actually wear one? And, perhaps more importantly, how many should do, but don't? Are you wearing one right now? Are your colleagues? Does it matter? Would you tell anybody if it did? And, if not a lab coat, what about those protective goggles? They get so hot in summer, don't they? Is it really that big a deal if you leave them on the hook just this once?

It is easy for scientists, especially those who have been around for a while and so tend to be in charge, to take a cavalier attitude to safety, purely because science is so much safer now than it was when they began. And although it is true that laboratories and lab culture have improved since the reckless days of the 1950s and 1960s, accidents still happen. Sometimes, these accidents are fatal. Laboratories do still kill people.

In an Editorial on the subject last April (*Nature* **472,** 259; 2011), prompted by the death of physics and astronomy undergraduate student Michele Dufault in a workshop at Yale University in New Haven, Connecticut, this publication warned against complacency when it comes to lab safety. A common complaint among environmental health and safety officers in universities and elsewhere, we noted, is that "there is no good source of consistent data on laboratory accidents, which could be studied to determine effective safety interventions". That the working environment for scientists is safer now than in times past is less important than whether it is as safe now as it could, or should, be — and there is at present no way to say for sure that it is.

> "It is easy for scientists to take a cavalier attitude to safety."

Nature Publishing Group (NPG) has now joined with the University of California, Los Angeles, and the software firm BioRAFT to try to fill in some of the blanks. (BioRAFT, based in Cambridge, Massachusetts, has investment from Digital Science, owned by NPG's parent company, Macmillan.) Together, the three have launched an online survey of international laboratory safety and working culture. Some readers will already have received invitations to participate, but everyone else is welcome, too: the survey can be found at go.nature.com/7ldjli. It should take about 15 minutes to complete and is anonymous — although there is an option to leave an e-mail address for follow-up questions. The organizers hope that tens of thousands of working scientists will respond to questions about the environments they work in and the attitudes that they and their colleagues have to health and safety regulations. The survey also addresses research practice, including how many people regularly work alone in a lab, and how often; training provision; and whether scientists feel able to raise concerns about safety. Please take the survey. Someone, some day, will benefit. ∎

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

Do not censor science in the name of biosecurity

Security officials should not be concerned about the publication of mutant-flu research, says bio-weapons expert **Tim Trevan**.

The recent controversy over research into mutated versions of the H5N1 flu virus has focused on biosecurity concerns. It is easy to get the impression that this debate has created a clear split between a scientific community that wants the research to proceed and the results to be published and a biosecurity community that doesn't.

As a member of this biosecurity community for more than 30 years — I was special adviser to the chairman of the United Nations weapons inspectors in Iraq and covered chemical and biological disarmament with the UK Foreign Office in both London and Geneva, Switzerland — I believe this to be a false dichotomy. The research should be published in full, as it will be this week.

In fact, I will go further and say that the whole concept of dual-use biological research that is 'of concern' is flawed. It is a dangerous distraction, an inappropriate hangover from nuclear-threat analysis. Almost all biological knowledge can be either misused or applied for good.

Those concerned about publishing full details of the mutant-flu work say that they fear the research will be misused to develop more-effective biological weapons. But who would want to use a live, highly transmissible, virulent organism as a weapon, and to what purpose? And would censorship stop them?

Although such a weapon would strike terror and harm economies, its impacts would be uncontrollable, indiscriminate and unpredictable. And compared with conventional weapons, it would be slow to take effect and relatively easy to combat, through prompt vaccination and treatment.

That severely restricts the number of potential users. An uncontrollable weapon is unsuited to targeted attacks and its use would heap opprobrium on the user. And if insurgents or terrorists unleashed a catastrophic and indiscriminate attack on civilians, it would devastate sympathy for their cause.

The only groups who might logically consider using such a weapon are those for whom humans are the problem, such as environmental extremists and animal-rights activists, or apocalyptic sects, such as the Japanese terrorist cult Aum Shinrikyo, which released sarin gas in the Tokyo underground in 1995. Then there are those who do not care about casualties, such as a state or a regime that believes it faces imminent existential threat, or suicide fighters.

Censorship of the H5N1 papers would not have kept the genie in the bottle. Suppressing such papers or limiting access to their findings might even encourage proliferation by drawing attention to the risks and by provoking those researchers denied access to the results to seek to replicate them.

Can we prevent proliferation by controlling research? Certainly, researchers, institutional review boards and funders must consider the implications of proposed research from the

CONTINUED RESEARCH INTO MUTANT FLU WILL ULTIMATELY DETER THE USE OF BIOLOGICAL WEAPONS.

outset and implement a full biosafety and biosecurity plan. Major efforts have been made in this area. But to deny funding to projects with clear scientific or public-health value, even if they have some biosecurity risks, will drive research to undesirable sources of funding and prevent valuable research from being done.

If the knowledge and the science cannot be contained, then what about access to the materials and equipment required to turn research results into weapons? The direction in which technology and scientific services are heading does not bode well for controlling proliferation in this way. Companies already make genes for mail order. Free gene-design software exists. DNA printers will probably be on lab bench tops within the decade. But it cannot be morally or politically defensible to prevent wide distribution of tools that are indispensable to public health and basic research.

Warfare and terrorism are not the only biological risks that confront humanity. There is an entire spectrum of risks, from natural and accidental to deliberate. We are mostly helpless to prevent the periodic creation of new deadly diseases. We know that we face regular flu pandemics and that some will be particularly deadly.

An analysis of the effect of carrying out and publishing such research must compare two factors. The first — the cost — is the risk that publication will lead to deliberate release, multiplied by the impact of the release, multiplied by the frequency of release. The second — the benefit — is the possible reduction in the 250,000–500,000 annual deaths worldwide due to seasonal flu and the more than 12 million lives lost annually to other infectious diseases, among other public-health benefits.

Precise calculation is not possible, but the evidence strongly suggests that the increase in risk is quite small. The known benefits of addressing public-health challenges from nature will almost always far outweigh the potential and unknowable increased risk of misuse.

The bigger argument in favour of continued research into viral transmissibility and pathogenicity (the focus of the mutant-flu work) is that it will ultimately deter the use of biological weapons.

The best strategy to stop biological attacks is to make biological weapons unattractive by making preparedness and responses so effective that the consequences are no worse than those of a train wreck. Increased understanding of transmissibility and pathogenicity will enable countries to identify threats earlier, develop better vaccines, produce them more quickly and develop broad-spectrum defences to diseases. This will protect against both nature and warfare. ∎

Tim Trevan *is executive director of the International Council for the Life Sciences in McLean, Virginia.*
e-mail: trevan@iclscharter.org

RESEARCH HIGHLIGHTS
Selections from the scientific literature

Sequencing tracks outbreak

Whole-genome sequencing has enabled researchers to confirm which patients in a hospital were affected by a particular outbreak of an antibiotic-resistant pathogen — at a cost and on a timescale that are clinically relevant.

Sharon Peacock at the University of Cambridge, UK, and her team sequenced isolates of methicillin-resistant *Staphylococcus aureus* (MRSA) from 14 hospitalized patients, half of whom had become carriers of a specific strain of MRSA during an outbreak in the neonatal intensive care unit. In about 1.5 days and at a cost of some US$150 per isolate, the authors generated sequences that showed a clear genetic distinction between outbreak and non-outbreak isolates.

The researchers predict that, as whole-genome sequencing costs and turnaround times fall, this will become a standard tool for controlling the spread of dangerous pathogens.
N. Engl. J. Med. 366, 2267–2275 (2012)

Castration boosts spider stamina

The male orb-web nephilid spider often castrates himself during sex, reducing his body weight by up to 9%. This could increase his endurance when defending his mate from competitors.

Daiqin Li at the National University of Singapore and his colleagues partly or fully castrated male spiders (*Nephilengys malabarensis*; **fully castrated pictured right**) and then assessed their stamina by repeatedly poking them with a paintbrush to goad them into moving.

Spiders that retained their genitals, or palps, quickly became exhausted, whereas the much lighter eunuchs showed 80% greater locomotor endurance. Spiders with one remaining palp kept going for 32% longer than intact males (**left**).
Biol. Lett. http://dx.doi.org/10.1098/rsbl.2012.0285 (2012)

Bat culls do not stop spread of rabies

Culling adult vampire bats might not be an effective means of reducing outbreaks of rabies in humans and livestock.

Daniel Streicker at the University of Georgia in Athens and his colleagues tested common vampire bats (*Desmodus rotundus*; pictured) sampled from 20 colonies across Peru between 2007 and 2010 for exposure to the rabies virus. Exposure prevalence ranged from 3% to 28% and was highest in immature bats. Culling during the test period did not reduce the probability of exposure to rabies.

Adult bats might be developing immunity to rabies after repeated exposure to the virus, the authors suggest, so culling could increase virus transmission in part because it targets immune adults and leaves behind young bats that are more likely to carry and to transmit the disease.
Proc. R. Soc. B http://dx.doi.org/10.1098/rspb.2012.0538 (2012)

Good microbes fight bad

Microbes living in the guts and airways of mammals help their hosts to fend off pathogens.

John Wherry and David Artis at the University of Pennsylvania in Philadelphia and their team treated mice with antibiotics to kill off their gut microbes and then infected them with an influenza virus. The mice lost more weight and were more likely to die than those that did not receive antibiotics. The antibiotic-fed mice also mounted a reduced immune response to the virus.

The authors suggest that the bacteria living in mammals prime the immune system to respond to pathogens, and say that harnessing this ability could aid in the treatment of viral infections in humans.
Immunity http://dx.doi.org/10.1016/j.immuni.2012.04.011 (2012)

Graphene can desalinate water

Salt could be separated from water using an ultrathin porous membrane made up of a single sheet of carbon atoms, a computational study suggests.

David Cohen-Tanugi

---

and Jeffrey Grossman at the Massachusetts Institute of Technology in Cambridge simulated the molecular interactions between graphene, salt and water, and looked at the effects of different pore sizes on salt filtration. The researchers found that graphene with a pore size of 0.7–0.9 nanometres should stop the passage of salt while letting water through. Attaching hydroxyl groups to the edges of the graphene pores would boost water flow-rate but impair salt rejection — because the chemical groups can swap with water molecules surrounding the salt ions. By contrast, attaching hydrogen atoms to the pores would improve filtration.

Graphene promises to be many times more permeable to water than conventional membranes used in desalination, the authors say.
Nano Lett. http://dx.doi.org/10.1021/nl3012853 (2012)

MICROBIOLOGY

Bacterial border defence

Insects combat pathogenic bacteria by producing a polymer called melanin and depositing it onto the microbe's surface. But one bacterial species has a weapon of its own — a cell-surface molecule that inhibits insects' melanin-producing enzymes.

Jon Clardy at Harvard Medical School in Boston, Massachusetts, and his team pinpointed the molecule, rhabduscin, on the surface of the bacterium *Xenorhabdus nematophila*. Nanomolar levels of the chemical blocked the activity of a melanin-producing enzyme from waxmoth larvae. *X. nematophila* that lack a rhabduscin-producing enzyme were less effective at killing the larvae than were normal bacteria.

Genes that encode enzymes involved in rhabduscin production are also found in the pathogen that causes cholera, *Vibrio cholerae*. A similar defence mechanism

might exist in this bacterium, the authors speculate.
Proc. Natl Acad. Sci. USA http://dx.doi.org/10.1073/pnas.1201160109 (2012)

APPLIED PHYSICS

Terahertz-wave detector

Devices that emit and detect radiation in the terahertz part of the spectrum — between the infrared and microwave regions — have potential applications in imaging, including in medical diagnostics. Researchers have developed a compact and efficient terahertz detector that works at room temperature.

Miriam Vitiello of the National Enterprise for nanoScience and nanoTechnology in Pisa, Italy, and her team built their detector out of indium arsenide nanowires 1.5 micrometres long and 30 nanometres in diameter. Radiation from a 1.5-terahertz emitter was funnelled to the detector from a bow-tie-shaped antenna.

The researchers suggest that the detector could be tuned to respond to even higher frequencies, and could be built into multi-pixel arrays, which are ideal for detectors.
Appl. Phys. Lett. 100, 241101 (2012)

STEM CELLS

Human eye parts in a dish

Retinal cells made from human embryonic stem cells could one day be used to help restore sight in people with certain forms of blindness.

Yoshiki Sasai at the RIKEN Center for Developmental Biology in Kobe, Japan, and his colleagues used the stem cells to generate retinal epithelial cells, which are precursors for the retina. After a few weeks, a single layer of these cells spontaneously formed a curved structure called an optic cup. After several more weeks, the cup developed into a multilayered structure with

multiple retinal cell types, including light-sensitive photoreceptors.

The researchers also devised a method for cryopreserving the retinal tissue. They foresee that stored material could ultimately be triggered to develop into specific cell types that can be grafted onto a patient's retina.
Cell Stem Cell 10, 771–785 (2012)
For a longer story on this research, see go.nature.com/ibgaqa

ARCHAEOLOGY

One mummy but three people

A mummy found at a site off the coast of Scotland consists of remains from at least three individuals.

Human skeletons dating from 1400–1100 BC were previously unearthed at the Cladh Hallan settlement. A male skeleton was identified as a composite of multiple individuals on the basis of isotope data, but the status of a female mummy (**pictured**) was less certain. Terry Brown at the University of Manchester, UK, and his colleagues extracted DNA

from bones of the jaw, right arm and right leg of the female remains. After excluding possible contamination, the researchers found that all three body parts came from different individuals. DNA analysis of the skull was inconclusive.

The team suggests that the remains at Cladh Hallan were deliberately merged, perhaps to symbolically combine different ancestries into one lineage.
J. Archaeol. Sci. http://dx.doi.org/10.1016/j.jas.2012.04.030 (2012)

SEVEN DAYS
The news in brief

China celebrates space-station success

In a milestone for China's space programme, three astronauts boarded the country's orbiting Tiangong 1 space module on 18 June. Their flight on the *Shenzhou 9* craft was the country's fourth manned space launch, but is the first of a series of missions in efforts to build a manned space station, the Tiangong ('Heavenly Palace'), by 2020 (see *Nature* **473**, 14–15; 2011). The mission carried China's first female astronaut, Liu Yang. See go.nature.com/f5qkka for more.

POLICY

Marine reserves
Australia's government has unveiled its final plans to create what will be the world's largest network of marine reserves, covering 3.1 million square kilometres of ocean along the nation's coasts. Researchers were worried by draft proposals last year (see *Nature* **480,** 14–15; 2011), but the plans released on 14 June addressed their concerns by including small but significant extensions to the reserve boundaries, to increase protection to regions such as the Coral Sea. The government expects the reserves to become law at the end of this year, after a public consultation. See go.nature. com/vnq8fw for more.

US research thrift
Research universities in the United States need to become more efficient and more productive, a 14 June report from the US National Academies urges. The report, requested by Congress in 2009, recommends ten fixes to maintain quality, including improving cost-effectiveness by sharing expensive research equipment and facilities, and encouraging greater collaboration. Federal funding has flattened or declined, and state funding has dropped by 25% on average and up to 50% in some cases, the report notes. See go.nature.com/6ppdtj for more.

Fisheries reform
Europe's attempts to overhaul its much-criticized fisheries policy took a step forward at a meeting in Luxembourg on 12 June. Ministers on the council of the European Union agreed to a partial ban on the practice of throwing away unwanted by-catch, and pledged to return fish stocks to levels that produce the maximum sustainable yield by 2020. But critics said that these agreements were insufficient, effectively allowing Europe's fisheries to continue on their current unsustainable path. Any reforms still need to be negotiated with the European Parliament. See go.nature. com/tn8oea for more.

Ocean acidity
The International Atomic Energy Agency (IAEA), based in Vienna, is to create a centre to facilitate and communicate research into ocean acidification, it announced on 18 June. The centre, to be launched this summer at the IAEA's Environmental Laboratories in Monaco, is part of the agency's remit to support peaceful uses of nuclear technology. Tracing of the radioisotope calcium-45, for example, allows sensitive measurements of calcium carbonate uptake in the skeletons of marine organisms, affected by acidifying oceans.

German excellence
Thirty-nine universities have each won a share of €2.4 billion (US$3 billion) in the second round of Germany's Excellence Initiative — a competition among the country's institutions to win elite status and funding for specialized research clusters and graduate schools for 2012–17. The decisions, announced on 15 June, spread cash across 99 separate projects.

Biomedical careers
The US National Institutes of Health on 14 June received two key reports on how to improve prospects for young biomedical scientists. One suggests raising minimum salaries for postdocs, among other recommendations, while the other addresses the plight of under-represented minorities in biomedical science. See page 304 for more.

Open access
A report commissioned by the UK government has laid out how the country should accelerate a shift to open-access publishing. The 19 June report, chaired by sociologist Janet Finch at the University of Manchester, UK, recommended that researchers should prepare to pay up front for the cost of publishing their research, a route known as gold open access. See page 302 for more.

Unethical research
A US court has dismissed a lawsuit by Guatemalan citizens against US officials over American researchers

who intentionally infected Guatemalans with sexually transmitted diseases in the 1940s (see *Nature* **482,** 148–152; 2012). A presidential bioethics commission has already issued a series of reports condemning the experiments, but on 13 June a judge ruled that the government was immune to prosecution in the case. Lawyers said they would appeal. See go.nature. com/1ffh6l for more.
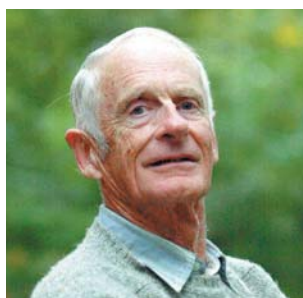
PEOPLE

Mikovits theft case

Chronic-fatigue-syndrome researcher Judy Mikovits is no longer facing criminal charges for stealing lab notebooks, computers and other material from her former employer, the Whittemore Peterson Institute for Neuro-Immune Disease in Reno, Nevada. A Nevada district attorney dropped the charges last week, although Mikovits (known for her now-retracted work linking chronic fatigue syndrome to a virus) still faces a civil suit from the institute. See go.nature.com/ ukgegy for more.

Nobel chemist dies

Organic chemist William Knowles (**pictured**), who shared the 2001 Nobel Prize in Chemistry, died on 13 June aged 95. Knowles worked for four decades at the agricultural

giant Monsanto in St Louis. He won the Nobel with Ryoji Noyori and Barry Sharpless for his work on chemical syntheses that selectively create one of two mirror-image forms (enantiomers) of a molecule.

BUSINESS

GM soya levy

The biotechnology giant Monsanto is one step closer to losing billions of dollars in revenues from its genetically modified (GM) Roundup Ready soya beans in Brazil. Monsanto, headquartered in St Louis, Missouri, levies a charge on Brazilian farmers who grow soya beans that turn out to be GM. Farmers say it is impossible to avoid growing GM soya because of contamination, and in April they won a challenge in the state of Rio Grande do Sul, where a judge ruled that the company's levy was illegal. The ruling is currently suspended, pending consideration by a higher court. But on 12 June,

the Brazilian Supreme Court of Justice said it should apply nationwide. See go.nature. com/mowmyh for more.

Fossil smuggling

A nearly complete tyrannosaurid fossil that sold for US$1 million was illegally smuggled out of Mongolia, according to a civil complaint seeking the fossil's return, which was filed on 18 June by the US Department of Justice. Despite a restraining order obtained by the Mongolian government from a Texas court to prevent the sale or transfer of the fossil, Heritage Auctions based in Dallas, Texas, auctioned off the *Tarbosaurus bataar* fossil in New York last month.

RESEARCH

Diesel cancer links

Diesel exhaust is carcinogenic to humans, the International Agency for Research on Cancer (IARC) declared on 12 June after a meeting in Lyons, France. Diesel emissions were previously classed as 'probably' carcinogenic; the latest conclusion followed the publication in March of a long-delayed US government study showing how exposure to diesel exhaust increases the risk of lung cancer in miners. The IARC's pronouncement is purely scientific; it will be up to national regulatory agencies

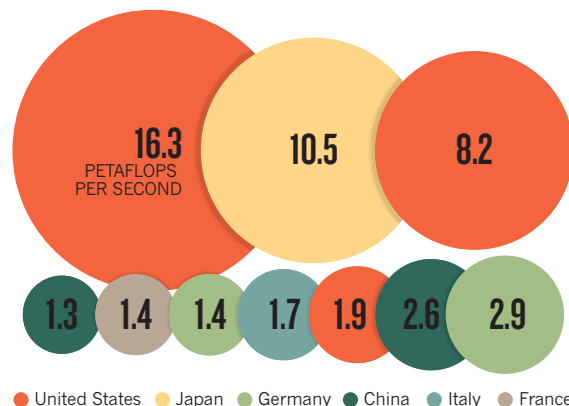to decide how to proceed. See go.nature.com/dddpaz for more.

Space X-rays

NASA's NuSTAR telescope, which will examine high-energy X-rays produced at the thresholds of black holes (see *Nature* **483,** 255; 2012), was launched into low-Earth orbit on 13 June. The low-cost mission is one of only a few available to X-ray astronomers. See go.nature. com/dcye8k for more.

COMING UP

24 JUNE–6 JULY
In St Petersburg, Russia, the United Nations Educational, Scientific and Cultural Organization's World Heritage Committee meets to discuss the state of conservation sites including Australia's Great Barrier Reef.
whc36-russia2012.ru

26–28 JUNE
In Washington, Seattle, marine scientists plan out an international network to monitor the acidification of the oceans.
go.nature.com/lopgt6

CORRECTIONS
The story 'Nobel laureate dies' (*Nature* **486,** 11; 2012) should have said that Andrew Huxley did his award-winning work at the Laboratory of the Marine Biological Association in Plymouth, not the Plymouth Marine Laboratory. And 'Piezonuclear row' (*Nature* **486,** 162; 2012) gave the wrong name for INRiM: it is the National Institute for Metrological Research.

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

TREND WATCH

The United States, which two years ago yielded the title of host to the world's top supercomputer — bested first by China and then by Japan — has roared back into the lead in this month's list of the world's top 500 supercomputers, with the Sequoia machine at Lawrence Livermore National Laboratory in California. It also claimed third place with a supercomputer at Argonne National Laboratory in Illinois. Italy made its top-ten debut with a system at the CINECA computing centre near Bologna.

WORLD'S FASTEST COMPUTERS
The United States reclaims the top spot for first time in two years with a supercomputer at Lawrence Livermore National Laboratory in California.

16.3 PETAFLOPS PER SECOND · 10.5 · 8.2

1.3 · 1.4 · 1.4 · 1.7 · 1.9 · 2.6 · 2.9

● United States ● Japan ● Germany ● China ● Italy ● France

who intentionally infected Guatemalans with sexually transmitted diseases in the 1940s (see *Nature* **482**, 148–152; 2012). A presidential bioethics commission has already issued a series of reports condemning the experiments, but on 13 June a judge ruled that the government was immune to prosecution in the case. Lawyers said they would appeal. See go.nature. com/1ffh6l for more.
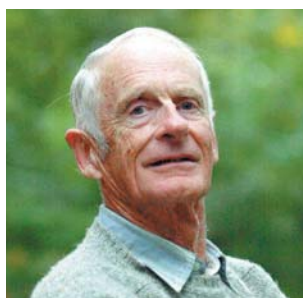
PEOPLE

Mikovits theft case

Chronic-fatigue-syndrome researcher Judy Mikovits is no longer facing criminal charges for stealing lab notebooks, computers and other material from her former employer, the Whittemore Peterson Institute for Neuro-Immune Disease in Reno, Nevada. A Nevada district attorney dropped the charges last week, although Mikovits (known for her now-retracted work linking chronic fatigue syndrome to a virus) still faces a civil suit from the institute. See go.nature.com/ukgegy for more.

Nobel chemist dies

Organic chemist William Knowles (**pictured**), who shared the 2001 Nobel Prize in Chemistry, died on 13 June aged 95. Knowles worked for four decades at the agricultural

giant Monsanto in St Louis. He won the Nobel with Ryoji Noyori and Barry Sharpless for his work on chemical syntheses that selectively create one of two mirror-image forms (enantiomers) of a molecule.

BUSINESS

GM soya levy

The biotechnology giant Monsanto is one step closer to losing billions of dollars in revenues from its genetically modified (GM) Roundup Ready soya beans in Brazil. Monsanto, headquartered in St Louis, Missouri, levies a charge on Brazilian farmers who grow soya beans that turn out to be GM. Farmers say it is impossible to avoid growing GM soya because of contamination, and in April they won a challenge in the state of Rio Grande do Sul, where a judge ruled that the company's levy was illegal. The ruling is currently suspended, pending consideration by a higher court. But on 12 June,

the Brazilian Supreme Court of Justice said it should apply nationwide. See go.nature. com/mowmyh for more.

Fossil smuggling

A nearly complete tyrannosaurid fossil that sold for US$1 million was illegally smuggled out of Mongolia, according to a civil complaint seeking the fossil's return, which was filed on 18 June by the US Department of Justice. Despite a restraining order obtained by the Mongolian government from a Texas court to prevent the sale or transfer of the fossil, Heritage Auctions based in Dallas, Texas, auctioned off the *Tarbosaurus bataar* fossil in New York last month.

RESEARCH

Diesel cancer links

Diesel exhaust is carcinogenic to humans, the International Agency for Research on Cancer (IARC) declared on 12 June after a meeting in Lyons, France. Diesel emissions were previously classed as 'probably' carcinogenic; the latest conclusion followed the publication in March of a long-delayed US government study showing how exposure to diesel exhaust increases the risk of lung cancer in miners. The IARC's pronouncement is purely scientific; it will be up to national regulatory agencies

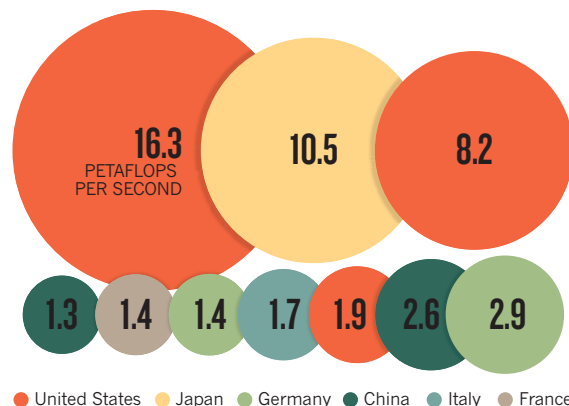to decide how to proceed. See go.nature.com/dddpaz for more.

Space X-rays

NASA's NuSTAR telescope, which will examine high-energy X-rays produced at the thresholds of black holes (see *Nature* **483,** 255; 2012), was launched into low-Earth orbit on 13 June. The low-cost mission is one of only a few available to X-ray astronomers. See go.nature. com/dcye8k for more.

CORRECTIONS

The story 'Nobel laureate dies' (*Nature* **486,** 11; 2012) should have said that Andrew Huxley did his award-winning work at the Laboratory of the Marine Biological Association in Plymouth, not the Plymouth Marine Laboratory. And 'Piezonuclear row' (*Nature* **486,** 162; 2012) gave the wrong name for INRiM: it is the National Institute for Metrological Research.

TREND WATCH

The United States, which two years ago yielded the title of host to the world's top supercomputer — bested first by China and then by Japan — has roared back into the lead in this month's list of the world's top 500 supercomputers, with the Sequoia machine at Lawrence Livermore National Laboratory in California. It also claimed third place with a supercomputer at Argonne National Laboratory in Illinois. Italy made its top-ten debut with a system at the CINECA computing centre near Bologna.

WORLD'S FASTEST COMPUTERS
The United States reclaims the top spot for first time in two years with a supercomputer at Lawrence Livermore National Laboratory in California.

16.3 PETAFLOPS PER SECOND
10.5
8.2

1.3 1.4 1.4 1.7 1.9 2.6 2.9

● United States ● Japan ● Germany ● China ● Italy ● France

Luc Montagnier has controversial scientific views that may harm Cameroon's AIDS institute, allege critics.

AIDS

Nobel fight over African HIV centre

Laureates question choice of interim scientific director.

BY DECLAN BUTLER

A fledgling AIDS research centre in Cameroon, already struggling to find a scientific leader, is now facing insurrection from an unlikely quarter: a group of 35 Nobel prizewinners.

The laureates are calling for the centre's interim scientific director, fellow prizewinner Luc Montagnier, to be removed from the part-time post. Observers say that unless the leadership crisis is resolved quickly and decisively, it could harm the prospects of the Chantal Biya International Reference Centre (CIRCB) in Yaoundé.

The centre has a comprehensive AIDS research and health-care programme, in particular testing and treating newborn babies to reduce maternal transmission of HIV. It is the only research institution in central Africa with the technology and expertise to monitor people with HIV thoroughly, and one of the few African sources of hard data about the spread of the disease. It has an annual budget of about US$1 million, an array of international collaborations and around 20 local staff members, most of whom trained abroad.

Nature has learned that the Nobel laureates wrote on 9 June to Paul Biya, president of Cameroon, asking him to reconsider Montagnier's appointment. Montagnier, head of the World Foundation for AIDS Research and Prevention in Paris, shared the 2008 Nobel Prize in Physiology or Medicine for discovering HIV.

The laureates argue that his embrace of theories that are far from the scientific mainstream, as well as what they claim are anti-vaccination views, risk hurting the CIRCB's research, health-care programme and reputation. Montagnier has suggested, for example, that water can retain a 'memory' of pathogens that are no longer present[1]; that the DNA sequences of pathogens emit electromagnetic waves that could be used to diagnose disease[2,3]; and that stimulating the immune system with antioxidants and nutritional supplements may help people to fight off AIDS[4].

HIGH-PROFILE OPPOSITION

The letter was coordinated by Richard Roberts, a Nobel-prizewinning molecular biologist and chief scientific officer of New England Biolabs in Ipswich, Massachusetts, who also wrote personally to Biya on 4 June, to resign from the CIRCB's scientific board. Roberts says he is concerned that Montagnier plans to pursue his unorthodox research at the centre. Several other board members have also resigned.

Robert Gallo, head of the Institute of Human Virology at the University of Maryland, Baltimore, who had battled with Montagnier over which of them had discovered HIV, has also entered the fray. On 4 June, Gallo wrote to Biya expressing concerns similar to those of the Nobel laureates and informing Biya that his institute, a founding sponsor of the CIRCB, was immediately severing its links with the centre.

Montagnier deplores what he describes as "ad hominem attacks" and "plain lies", and says that there is an "ignominious campaign" against him and his group. He says that history is full of pioneers whose ideas were at first given a chilly reception by a conservative research community. "I believe this is happening again to me, and it is very sad that it involves Nobel Prize laureates attacking a fellow laureate," he says.

The last straw for Montagnier's critics seems to have been his appearance in May alongside vaccine sceptics at a conference in Chicago, Illinois, organized by US patient-advocacy groups AutismOne and Generation Rescue. Montagnier's talk, on his hypothesis that bacterial infections may be one of many causes of autism spectrum disorder, states: "There is in ▶

▶ the blood of most autistic children — but not in healthy children — DNA sequences that emit, in certain conditions, electromagnetic waves."

Montagnier defends his research, pointing out that some clinicians have observed improvements in symptoms of autism after long-term treatment with antibiotics. He says that he has never argued that vaccination could cause autism. "Many parents have observed a temporal association, which does not mean causation, between a vaccination and the appearance of autism symptoms," he says. "Presumably vaccination, especially against multiple antigens, could be a trigger of a pre-existing pathological situation in some children."

LEADERSHIP CRISIS
The CIRCB, founded in 2006, is named after President Biya's wife, who has championed efforts to fight AIDS in Africa. Montagnier's AIDS foundation was a founding partner; Montagnier is also president of the now-defunct scientific advisory board, and vice-president of the management board.

The current crisis compounds problems caused by the centre's lack of stable full-time leadership. In March, its management committee appointed Montagnier to replace former interim scientific director Vittorio Colizzi, an AIDS researcher on secondment from the Tor Vergata University in Rome, who had held the post since 2009. Colizzi was standing in until a full-time scientific director could be hired, but a recruitment process last year failed to settle on an agreed candidate. Some candidates had also expressed misgivings about the job, because at the time the scientific director and administrative director had to share power, a situation that caused tensions, says Colizzi. To address this issue, a presidential decree issued on 31 May merged the positions to create the post of permanent director, with full control of the centre. The move should make it much easier to attract a leading scientist to the post, says Jacques Theze, an immunologist at the Pasteur Institute in Paris, a former member of the CIRCB's scientific board.

The decree also required that many of the centre's posts and committees be disbanded or renewed, creating an uncertain transitional period. On the day that Roberts resigned, for example, the scientific board was officially dissolved, and no clear timetable has been set to reestablish it. Colizzi is concerned that this deprives the centre of its main mechanism for enforcing rigorous peer-review and ethical oversight of research proposals. Montagnier says that he intends to continue all research previously approved by the board, and that he will ask the next board to review the programme. He also plans to embark on new research, including a "key project" using his electromagnetic-wave theory to detect reservoirs of HIV in the body that persist after antiretroviral treatment. Any new projects, including his own, will need to be approved by the centre's science board and ethics committee, he says.

Jean Stéphane Biatcha, head of the centre's management board and a presidential adviser, recognizes the "very serious disagreement" but says that the president and the Ministry of Health will quickly enact the 31 May decree, and so will renew the scientific advisory board and begin the search for a permanent director.

Theze says that he would have preferred Montagnier's detractors to have taken a more diplomatic approach, and warns that the high-level criticism, and the resulting controversy, risks tarnishing the credibility and reputation of the centre, which he says is unfair, because the CIRCB has enormous potential. He worries that the episode might also discourage scientists from applying for the position of director. ■

1. Montagnier, L., Aïssa, J., Ferris, S., Montagnier, J.-L. & Lavalléee, C. *Interdiscipl. Sci.* **1,** 81–90 (2009).
2. Montagnier, L., *et al.* Preprint at http://arxiv.org/abs/1012.5166 (2010).
3. Montagnier, L. *et al. Interdiscipl. Sci.* **1,** 245–253 (2009).
4. Butler, D. *Nature* **468,** 743 (2010).

Britain aims for broad open access

But critics claim plan seeks to protect publishers' interests.

BY RICHARD VAN NOORDEN

For years, countries have been edging towards open access for research, with some funding agencies requiring that researchers make their papers publicly available within a set period after publication. A report commissioned by the UK government recommends a more radical step: making all papers open access from the start, with authors paying publishers up-front to make their work free to read.

The shift towards this 'gold' form of open access will create short-term financial burdens for research funders, the report acknowledges, but the economic and cultural benefits far outweigh the risks. Not everyone is convinced, however: research-intensive universities say they are concerned that the report plays down potentially cheaper ways to move to open access, in favour of sustaining publishers' profits.

"Momentum for open access is already under way, and it's important for the United Kingdom to embrace that change, to accelerate it, and to manage it," says Janet Finch, a sociologist at the University of Manchester, UK, who chaired the panel behind the report, which was released on 19 June. It is expected to set the national agenda for open access, and influence other countries to follow Britain's lead.

"The ultimate goal is to have a system where the full costs of research publication are met in advance," says Martin Hall, another member of the panel and vice-chancellor of the University of Salford in Manchester. Globally, the number of gold articles is growing by about 30% each year, aided by the rise of journals such as *PLoS ONE*. But they still make up a minority of the world's output — comprising about 12% of research articles indexed in Elsevier's Scopus database in 2011, according to preliminary estimates by Mikael Laakso and Bo-Christer Björk at the Hanken School of Economics in Helsinki (see 'Rise of gold'). UK researchers tend to publish in higher-impact selective journals, so only 5% of their articles are gold open access, according to data collected by Yassine Gargouri, a informatician at the University of Quebec in Montreal, Canada (see 'Open access in the UK').

As that proportion rises, the report notes, authors' open-access costs will grow — but university libraries will still have to subscribe to most of the journals that currently line their shelves. Subscription costs will fall substantially only when most research articles are freely available. During the transition period, gold and subscription models will exist side by side, potentially increasing the overall costs of access. The report also recommends subsidising subscription licences for health and business users to give them better access. Overall, the panel estimates that these transitional costs will amount to roughly £50 million–60 million (US$78 million–94 million) per year, on top of the country's existing annual spending of about £175 million to publish and access research. If the costs were to be met by research funders, they would total about 1% of Britain's annual science budget.

The report does not recommend a figure for

⊃ NATURE.COM
For *Nature*'s policy on open access, see:
go.nature.com/pnxzup

the cost of a gold article, but notes that the UK Wellcome Trust, a major biomedical research funder, last year paid an average of £1,422 per paper on behalf of the scientists it supports. Costs could be greater in more selective journals — *Nature*'s editor-in-chief Philip Campbell says that the journal would have to charge more than £6,500 for gold open-access articles.

Universities and funders will have to work out how to transform their payment systems under a gold regime, with each institution likely to set up a central publishing fund supported by a percentage of every research grant. Whatever the solution, academics will be much more aware of the costs of publishing. This could, in turn, modify their behaviour, with researchers submitting papers to the journals they can afford to publish in, or trying to publish fewer, broader articles.

GOING GREEN

An alternative open-access model is already thriving around the world, and particularly in the United Kingdom. Under green open access, research funders can require that peer-reviewed papers be made openly accessible in online repositories, without the author paying a fee. This usually happens some months after publication, a time period that allows publishers to sell access to the paper for long enough to turn a profit. Researchers can also post pre-publication versions of their papers in institutional repositories.

Paul Ayris, director of library services at University College London, says that scaling up green publishing would be a cheaper short-term route to expanding open access, together with a nationwide scheme to pay for researchers' access to subscription journals en masse. "The gold route does nothing about publisher profits, which many commentators feel are already too high," he says. Open-access advocate Stevan Harnad, a cognitive scientist at the University of Montreal, is even more critical of the report's overt support for gold access. "Some publishers seem to be successfully persuading some politicians that what is at issue is protecting their current revenue streams and modus operandi from the threat of green open access," he says.

But the Finch group says that it was expressly asked to find sustainable ways to grow open

OPEN ACCESS IN THE UK

Of the 85,215 research papers published by UK academics in 2010 (as indexed by Web of Science), around 5% were gold open access, whereby authors pay for open publication. Another 35% were green open access — published behind a pay wall and then put in a free repository. However, the proportion varied between disciplines.

SOURCE: THOMSON REUTERS/Y. GARGOURI

- ■ Not open access
- ■ Green
- ■ Gold
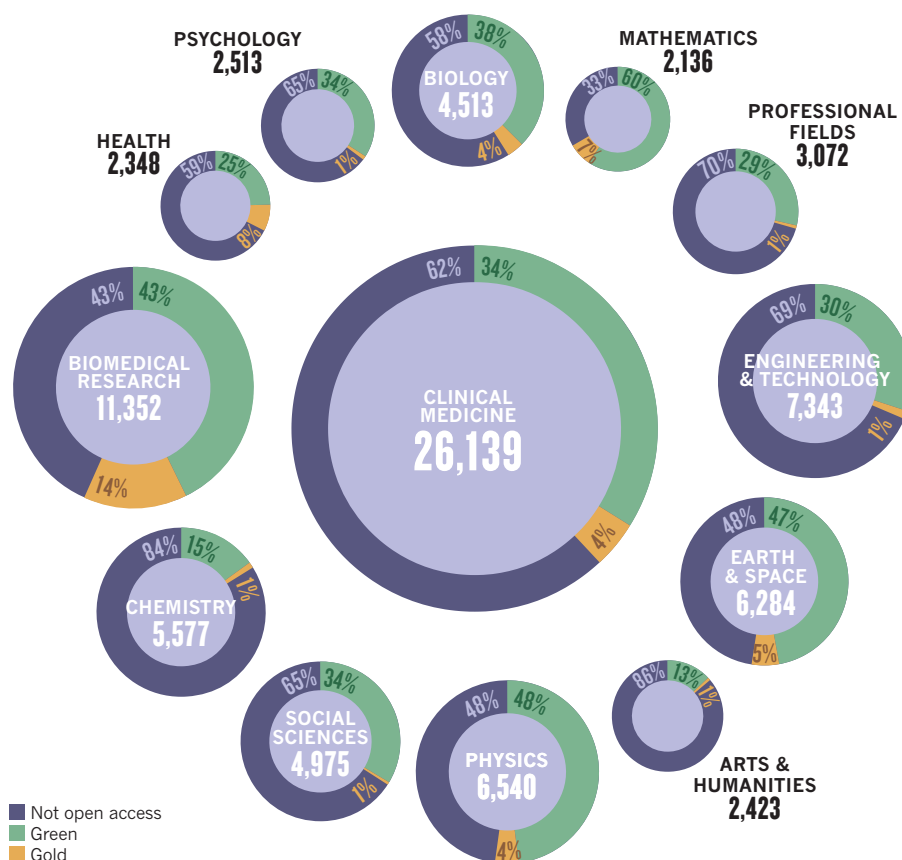
access, which it says only a gold route can provide. "It's not in the interests of UK scholarship to make recommendations which undermine the sustainability of the publishing industry," says Philip Sykes, another Finch group member and a librarian at the University of Liverpool. Universities can use their collective lobbying power to drive down both subscription and gold costs, he adds. Gold open access will eventually result in lower incomes for publishers anyway, Finch members note, by making the research-publishing market more transparent and competitive.

That's particularly worrying for learned societies, because they rely on subscription publishing for much of their income. The London-based Institute of Physics, for example, earns some £10 million each year — more than 60% of its total income — from publishing, which it spends on activities such as science education and outreach, says its president Peter Knight. "The mood of the community is to get costs down — but if scientific publishing only just covered its costs, an awful lot of our programmes would be in jeopardy," he says.

What matters now is how the agencies that support UK scientists require them to make their research freely available. Existing open-access mandates have been spottily enforced. The Wellcome Trust has only 55% compliance, although it will soon make grant funding

conditional on open-access publishing. A similar condition from the US National Institutes of Health currently has 75% compliance.

In March, Research Councils UK (the umbrella body for the United Kingdom's seven government-funded grant agencies) released a draft policy that suggested it, too, would toughen up on open access. The Higher Education Funding Council for England, another major research funder, could go the same way. But the devil will be in the detail, says Hall. "If research funders go soft on open access, the Finch report will be of only academic interest."

Most uncertain of all is how rapidly the United Kingdom's efforts might drive other countries towards open access. British scientists produce 6% of the research papers published worldwide each year, and the country could find itself paying to make its research free for others' benefit. But there is growing momentum internationally. The European Commission hopes to push for an open-access mandate in its 2014–20 research-funding programme Horizon 2020, and the newly formed Global Research Council — a forum for funding-agency heads worldwide (see *Nature* **485**, 427; 2012) — has open access on its agenda for its second meeting next year in Berlin. As the report concludes, "measures to promote open access need to be ... international in scope if they are to achieve their full potential". ■

RISE OF GOLD

The world's gold open-access articles are rising as a share of the total.
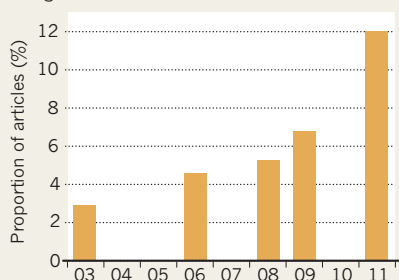
SOURCE: SCOPUS/ISI/M. LAAKSO/B.-C. BJÖRK

COMMUNITY

A workforce out of balance

Too many biomedical PhDs and too few minorities are a demographic dilemma for the NIH.

BY MEREDITH WADMAN

The US biomedical workforce has a glut of young researchers but a dearth of some minority groups, members of which are struggling to establish themselves in the field. The double-barrelled problem is laid out in detail in a pair of reports presented to the US National Institutes of Health (NIH) in Bethesda, Maryland, on 14 June.

NIH leaders have long worried about a steep increase in the number of biomedical PhDs, a consequence of the doubling of the NIH budget from 1998 to 2003. Now that boom is making it increasingly difficult for young scientists to launch academic careers (see 'Swelling ranks').

"This is dysfunctional and it's not sustainable in the long term," says Shirley Tilghman, president of Princeton University in New Jersey and a co-chair of the working group that authored a report on structural problems in the workforce. It calls for several measures to address the over-supply of PhDs, including a six-year cap on the number of years that a graduate student can be supported by NIH funds and an increase in the proportion of students on career-oriented training grants rather than on research grants.

A second report, focusing on diversity, was spurred by a study published in *Science* last year, which found that after factors such as education and publication record are controlled for, black applicants are 10% less likely than white applicants to win NIH research funding (D. K. Ginther *et al. Science* **333,** 1015–1019; 2011).

The diversity report confirms that minority applicants have significantly reduced success rates for grant applications (see 'Uneven playing field'). Confronted with data such as these, "there are a number of scientists of colour who feel, at the end of the day, 'What's the point?'" says Reed Tuckson, vice-president and chief of medical affairs at the health-insurance firm UnitedHealth Group in Minnetonka, Minnesota, and one of three co-chairs of the diversity working group. That sentiment "really, really bothers me", he adds. "We have got to turn [that] around."

The report recommends, among other things, that the NIH launch a "bold", well-funded competitive grant process to build infrastructure at institutions with a record of producing minority scientists, and that it launch an experiment to make applicants' identities and institutions anonymous in the review process.

Francis Collins, director of the NIH, has promised to respond to the recommendations by December. ∎

SWELLING RANKS

The number of US biomedical PhDs has ballooned in the past decade, driven by NIH funding of research assistantships. These typically offer less career development compared with traineeships and fellowships.

SOURCE: NSF
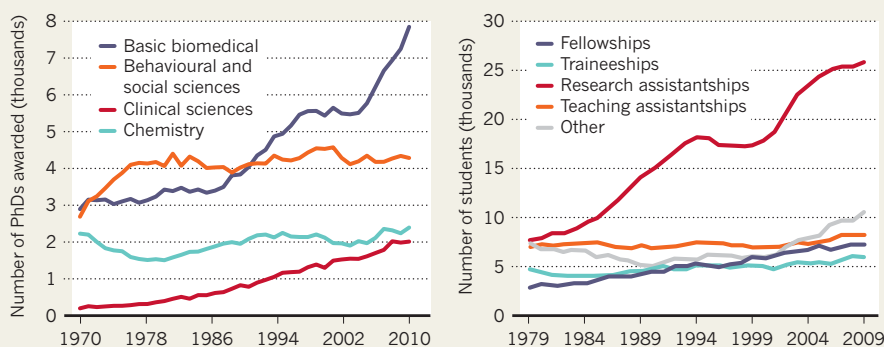
Researchers become faculty members and win a first NIH grant significantly later in life now than in 1980.

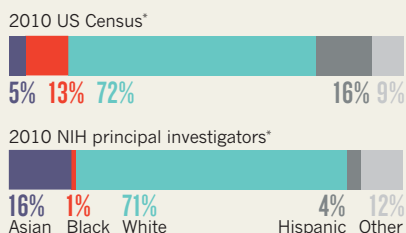SOURCE: NIH; ASSOC. AM. MED. COLL.
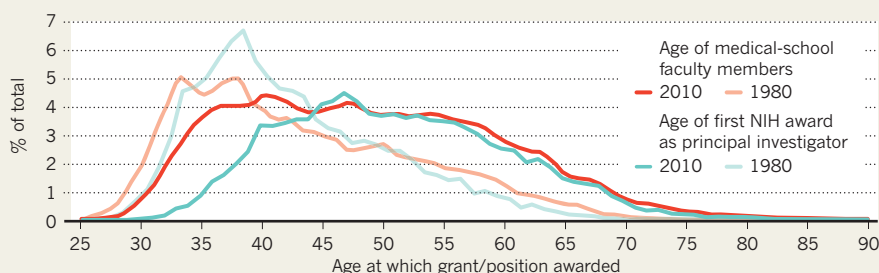
A growing number of US biomedical scientists are foreign workers. Early salaries are lower than in other fields.
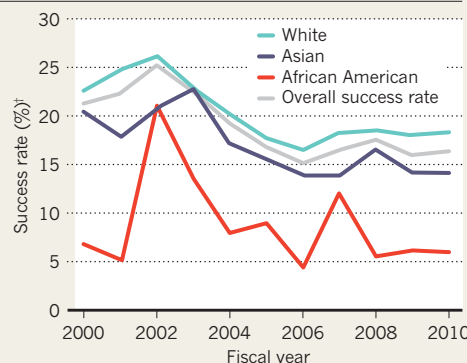

SOURCE: NSF
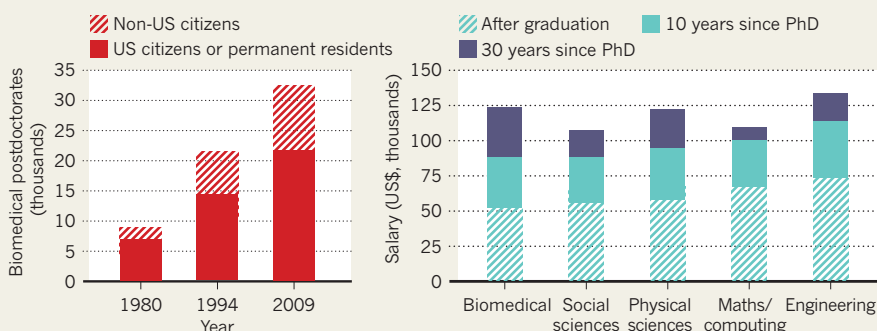
UNEVEN PLAYING FIELD

Relative to the US population, black and Hispanic people are under-represented in biomedical fields. Minority applicants have lower success rates than white applicants at winning new NIH grants (right).

2010 US Census*

5% | 13% | 72% | 16% | 9%

2010 NIH principal investigators*

16% | 1% | 71% | 4% | 12%
Asian | Black | White | Hispanic | Other

*Adds to more than 100% because some respondents declared two races.

†Success rates of NIH new grant applications in basic sciences category.

SOURCE: 2010 US CENSUS BUREAU REPORT; NIH

ACADEMIC CONDUCT

Romanian prime minister accused of plagiarism

Allegations prompt questions about government's ability to tackle misconduct in academia.

BY QUIRIN SCHIERMEIER

Romania's new government, still reeling from a misconduct scandal that forced its research minister to resign last month, has been hit by fresh allegations of plagiarism that strike at the very top.

Prime Minister Victor Ponta has been accused of copying large sections of his 2003 PhD thesis in law from previous publications, without proper reference. If the charges are substantiated, they could spark public pressure for Ponta to resign, say political insiders. The allegations are also raising fresh doubts about the government's ability to tackle corruption in the higher-education system.

Nature has seen documents compiled by an anonymous whistle-blower indicating that more than half of Ponta's 432-page, Romanian-language thesis[1] on the functioning of the International Criminal Court consists of duplicated text. Moreover, the thesis was republished with very minor amendments as a Romanian-language book in 2004 (ref. 2), and also forms the basis of a 2010 book on liability in international humanitarian law[3]. A former PhD student of Ponta's, Daniela Coman, is named as co-author of the books.

Substantial sections of text in all three publications seem to be identical, or almost so, to material in monographs written in Romanian by law scholars Dumitru Diaconu[4] and Vasile Creţu[5]. They also feature direct Romanian translations of parts of an English-language publication by law scholar Ion Diaconu[6].

"The evidence of plagiarism is overwhelming," says Marius Andruh, a chemist at the University of Bucharest and president of the Romanian council for the recognition of university diplomas. If the allegations are borne out, "a serious discussion is needed in Romania and abroad to prevent this in the future," says Andruh.

"I understand that in law studies it can be necessary to copy extensive legal articles and definitions," says Paul Dragos Aligica, a Romanian political scientist at George Mason University in Arlington, Virginia. But Ponta's alleged plagiarism "goes way beyond that. It's astonishing", he adds. Ponta did not respond to *Nature*'s request for comment on the allegations, and Coman could not be contacted.

Ponta, leader of the Romanian Social Democratic Party, took office as prime minister only last month, replacing Emil Boc, who stepped down in February following protests against austerity measures that he had introduced. Ponta obtained his PhD from the University of Bucharest while acting as Secretary of State in the government of an earlier prime minister, Adrian Năstase — who was also his PhD supervisor.

> *"There is very clear evidence in these excerpts that the matter should be investigated further."*

"There is very clear evidence in these excerpts that the matter should be investigated further," says Vlad Perju, a Romanian political scientist and director of the Clough Center for the Study of Constitutional Democracy at Boston College in Chestnut Hill, Massachusetts.

The episode follows the resignation last month of the education and research minister, computer scientist Ioan Mang, following accusations of plagiarism in at least eight papers[7]. An investigation of that case by the Romanian Research Ethics Council is ongoing.

The latest allegations add to complaints about declining academic standards in Romania. The previous government had introduced measures to make the country's struggling science and education system more competitive and transparent, but the plans met ferocious opposition from large parts of the academic establishment, and have been substantially relaxed by the current government.

"It's more than unlikely that this government is fit to create institutional structures in science that Romania urgently needs," says Aligica. "How can it be, when some of its leaders don't seem to even remotely understand, or care about, the standards of good science?"

Members of Romania's post-communist elite — including many politicians — have been eager to acquire academic credentials. In the view of some critics, a number of private and public universities in the country are consequently degenerating into 'degree mills' that care little about the quality or novelty of the knowledge that they produce, and which are a breeding ground for academic plagiarism.

"One could almost feel pity for all these guys who have power and money, and who are now craving intellectual recognition," says Aligica. "Unfortunately these incidents just add to the disrepute of Romanian academic standards and create extra pressure that real Romanian scholars and scientists will now have to fight against." ∎

1. Ponta, V. *Curtea Penală Internaţionala*. PhD thesis, Univ. Bucharest (2003).
2. Ponta, V. & Coman, D. *Curtea Penală Internaţionala* (Lumina Lex, 2004).
3. Coman, D. & Ponta, V. *Răspunderea în Dreptul Internaţional Umanitar* (Universul Juridic, 2010).
4. Diaconu, D. *Curtea Penală Internaţionlă, Istorie şi Realitate* (Editura All Beck, 1999).
5. Creţu, V. *Drept Internaţional Penal* (Editura Tempus, 1996).
6. Diaconu, I. *The International Criminal Court: A New Stage* (Nicolae Titulescu Romanian Inst. International Studies, 2002).
7. Abbott, A. *Nature* **485,** 289 (2012).

MORE ONLINE

TOP STORY

Biologists grow human-eye precursor from stem cells
go.nature.com/ibgaqa

MORE NEWS

● Budget cuts threaten Japanese scientists' pay go.nature.com/ahwioz
● Risk assessment of US agro-biosafety lab found wanting go.nature.com/ozze9w
● Boron finally gets a triple bond go.nature.com/3motmh

NEWS SPECIAL

Rio+20 Earth summit kicks off in Brazil, with low expectations of progress
go.nature.com/d69weu

Many unregulated gold mines in Madre de Dios, Peru, leave a trail of environmental devastation in their wake.

ENVIRONMENT

Peru battles the golden curse of Madre de Dios

Attempts to reduce the environmental and health impacts of mining cause unrest.

BY ELIE GARDNER IN LIMA

"You go out to these communities that are so incredibly poor, and there is money buried in the dirt," says Jason Scullion. "It is not surprising that they want to go out there and dig it up."

Scullion, a graduate student in environmental and forest sciences at the University of Washington, Seattle, arrived in the Peruvian region of Madre de Dios last September, just after the price of gold hit a record high of US$68 per gram. Madre de Dios is at ground zero of Peru's gold rush: an estimated 30,000 artisanal and small-scale miners work in this lush Amazonian area, one of the most biologically diverse places on Earth. Scientists and conservationists are alarmed by the damage that mining is causing to the land and its people.

In the 1930s, settlers armed with shovels, picks and pans came to search the riverbanks for deposits of gold washed down from the Andes mountains. Now, mechanical diggers and dump trucks are much more common. Peru is the sixth-largest producer of gold in the world, and the metal was its main export in the first quarter of this year. Yet about 20% of Peru's bullion is mined illegally, using techniques that destroy forests and pollute local rivers — as well as depriving the government of an estimated US$305 million in taxes each year. The informal operations rarely assess their effects on the environment, or develop plans for what to do with the mines once they are exhausted, and they leave behind mountains of sand and rock, dead trees and deep pits filled with murky water.

Now the government is tightening the screws on illegal mining, and scientists monitoring its impacts are on the front lines of a battle between miners, environmental campaigners and the authorities.

In February, the Peruvian government banned mining in Madre de Dios outside a designated 500,000-hectare corridor (see map) and ordered that all miners must formally register — a year-long process that requires the mine operators to produce a work plan, an environmental-impact assessment and a clean-up strategy, among other requirements. By setting aside a specific area for mining, the government hopes to regulate the industry more effectively and to protect parks and the territories of indigenous people. But miners who have worked outside this newly designated corridor for decades with the government's tacit approval fear for their livelihoods, and the Madre de Dios

NATURE.COM
For *Nature*'s special on the changing Amazon, see:
go.nature.com/pvnkj5

Mining Federation in regional capital Puerto Maldonado has campaigned strongly against the legislation. In March, three people died as police clashed with federation-organized protests involving about 15,000 people. The government began raiding mining camps outside the corridor in late March, but in many of those areas miners have since returned to work.

Only around 4,000 miners in the region have met the government's deadline of 13 June to register their mining activities, suggesting that many more intend to disregard the order. Miners who have not registered, or who break environmental laws, will face up to 10 years in prison. But locals say that corruption and lack of government resources in Madre de Dios will make enforcing the law difficult.

SPOILT WILDERNESS

At stake are broad swathes of fertile rainforest, including that in the Tambopata National Reserve and Manu National Park, which is the largest national park in Peru and has been designated a World Heritage Site by the United Nations Educational, Scientific and Cultural Organization. "This is the epitome of a healthy ecosystem," says Enrique Ortiz, vice-president of the Amazon Conservation Association, based in Washington DC. "It's the capital of biodiversity." But mining is already beginning to encroach on these areas, and was threatening to become more widespread. A study published last year (J. J. Swenson *et al. PLoS ONE* **6**, e18875; 2011) showed that mining is deforesting Madre de Dios faster than any other activity. Using satellite imagery, the study's authors found that deforestation in two prominent mining zones increased sixfold between 2003 and 2009, destroying 6,600 hectares of wetlands and primary tropical forest. And they predicted that the trend will only get worse.

Scullion came to Madre de Dios to find out whether that prediction is coming true. Funded by a 10-month grant from the US government's Fulbright Program, he is mapping 2 million hectares of the region, including many of the mining hotspots. Most previous mapping studies have assigned land to only two categories — forested and deforested — lumping agriculture, towns and mining areas together, and making it difficult to track the impact of the gold rush. Scullion's study will use ten categories, including mining, agriculture and five different classes of forestry.

Scullion says that there is a misconception among locals that researchers are against mining. Not so, he says — he just wants it to be done in a more sustainable way, staying out of parks and reserves, and ensuring that miners reforest areas after operations have finished. Scullion hopes that his maps will identify the most vulnerable areas in Madre de Dios, as well as the most biologically diverse, and says that they could provide the high-quality information that the government needs to decide whether land should be used for logging, mining or conservation. The data could also help to guide where — and how much — mining takes place in the designated corridor.

Mining is also taking its toll on local people. An estimated 45–50 tonnes of mercury are used each year in Madre de Dios to extract the prized gold, and a large proportion of that ends up in rivers or is released into the atmosphere.

Miners combine mercury with sediments that contain gold — typically using their feet to mix them in a bucket or drum — to form a solid amalgam of the two metals. That amalgam is then heated, often in frying pans over open flames in non-ventilated spaces, to boil off the mercury and leave gold behind.

In March, Katy Ashe, a graduate student in environmental engineering at Stanford University in California, published the first study (K. Ashe *PLoS ONE* **7**, e33305; 2012) to show the scale of the health threat from mercury in Madre de Dios. She found that in mining zones, the proportion of people burdened with unhealthy levels of the metal — 6 micrograms or more per gram of dry hair tested — was more than twice that in Puerto Maldonado. Mercury poisoning can cause vomiting and

THE GOLD RUSH
The Peruvian government hopes to prevent gold mining from encroaching into national parks in the Madre de Dios region.

BRAZIL

BOLIVIA

Madre de Dios

Manu National Park

Government mining corridor: 500,000 hectares

Puerto Maldonado

Tambopata National Reserve

PERU

Mining allowed by government

Mining allowed with permission from native communities

100 km

diarrhoea and, in more extreme cases, brain or kidney damage. Because the metal accumulates in rivers, elevated mercury levels were much more common in those who ate a lot of fish: 18% of people who ate 12 or more fish meals each month had unhealthy mercury levels, in contrast to just 6% and 7% of low and moderate fish consumers, respectively.

That finding tallies with as-yet unpublished research by Luis Fernandez, a tropical ecologist at the Carnegie Institution for Science in Stanford, with whom Ashe is about to begin working. In 2009, Fernandez discovered that the most-consumed fish species in Madre de Dios, such as the mota (*Calophysus macropterus*) and doncella (*Pseudoplatystoma fasciatum*), had the highest levels of mercury. Fernandez is now leading a project to conduct a more extensive survey of the levels of mercury in fish and humans.

Despite the overwhelming evidence of harm, Peruvians are still divided over the findings. Some march through Puerto Maldonado's main plaza shouting through megaphones that mercury is killing everyone, whereas others are willing to drink the toxic liquid metal to prove it is safe.

> *"Mining is deforesting Madre de Dios faster than any other activity."*

Variations of the mercury-amalgamation technique have been used in gold mining for centuries, and it is difficult to dislodge such deep-rooted practices among artisanal miners. The global gold-mining industry, including leading mines in Peru, has mostly switched to an extraction process that uses cyanide, and recovers about twice as much gold as does mercury amalgamation. But cyanide requires more careful handling than mercury and few artisanal or small-scale miners have the necessary knowledge and skills. The more modern process of thiosulphate leaching might offer a non-toxic alternative, but it is most effective with very fine particles of gold — and Madre de Dios tends to yield larger, coarser grains.

To reduce miners' exposure to mercury, non-governmental organizations have distributed retorts that can capture the toxic vapour. Two years ago, Peruvian engineer Carlos Villachica unveiled the ECO-100V, a US$4,500 machine that uses water and jets of air to separate gold from sediments. And local development organizations such as Caritas Peru, based in Callao, and the Association for Integral Research and Development in Lima have developed other mercury-free technologies for extracting gold.

But Cesar Ascorra, director of Caritas Peru's office in Madre de Dios, says that miners will not switch methods unless the alternative works just as quickly, recovers at least as much gold, and is no more expensive than mercury amalgamation. For now, he adds, miners are more worried about the government's demands for them to formalize their work, and until the price of mercury goes up, or its use is banned, there is little incentive for them to change their practices.

Ortiz says that this underscores the value of research in the region — and the importance of disseminating the results through public-awareness campaigns. Studies such as Fernandez's are not done "for the sake of knowledge", says Ortiz. "This has a direction." ∎

Unlike the Acropolis, many classical sites are under threat from looting.

ARCHAEOLOGY

Cuts leave Greek heritage in ruins

Austerity measures damaging archaeological research.

BY LEIGH PHILLIPS

The economic and political turmoil in Greece is not just jeopardizing the country's economic future, it is also having a devastating effect on the country's rich cultural past, according to archaeologists in Athens.

Last month, the Association of Greek Archaeologists warned that the economic policies dictated by the European Union and the International Monetary Fund would cause "the destruction of both our country and our cultural heritage". The austerity measures intended to cut government debt have forced the state archaeological service to slash staff numbers by more than 10%, with a further 35–50% reduction possible. Research and excavations are being abandoned. Museums that can no longer afford to pay for security are being plagued by armed robbers. And organized criminals are exploiting the chaos in an explosion of illegal digs and the trafficking of illicitly procured antiquities.

Despina Koutsoumba, president of the archaeologists' association, says that the government no longer funds new research projects — other than those involving foreign partners to whom they are contractually obliged.

"Last year, a very important excavation underneath metro rail lines uncovered long walls connecting Athens to the port of Piraeus, built in the age of Pericles," she says. "The walls to the port were vital and played a major role in the Peloponnesian War as they linked Athens to the sea, and their strength was in their fleet." But the excavation was stopped uncompleted, she says, and there will be no opportunity to return to it once the rail lines are finished.

Koutsoumba says that the ministry also wanted to halt work at a site in Thessaloniki, after major buildings from the time of Roman emperor Gaius Galerius around AD 305–11 were discovered on land where a shopping centre was to be built. The archaeologists' association won a small victory when it persuaded the ministry to continue the work, she says, "but it is a continuous struggle". Net salaries for young

archaeologists in the service have been cut by 35%, to €670 (US$842) a month, and many have taken part-time jobs to supplement their income. Meanwhile, senior employees with the highest salaries and most experience are being forced into early retirement. Those who remain are reduced to bureaucratic 'box-ticking' — visiting construction sites to assess whether excavation is needed. They no longer have time for research or analysis, says Koutsoumba.

"Excavation is of course part of our job, but we also need time and funds to do research, to publish our findings. We are not giving anything to the scientific community and letting people know about our new discoveries — we are just digging up pretty objects," she says.

While legitimate archaeology is being hampered, looting is on the rise. The country is pockmarked with holes dug by the poor and desperate hoping for 'buried treasure', and organized criminals perform more professional excavations. "There is no doubt that there has been an increase in the past 3–4 years in both organized and amateur illegal digs and this is definitely related to the cuts," says Christos Tsirogiannis, a forensic archaeologist and researcher at the University of Cambridge, UK, who specializes in investigating the criminal networks behind trafficking in antiquities. He escorts the Greek police art squad on raids to identify looted antiquities.

"There is always a rise in this form of crime in regions rich with antiquities during times of crisis — in recent years in Egypt, Iraq and Afghanistan," says Tsirogiannis.

He says that auction houses and museums are not doing enough to avoid dealing in artefacts of "unprovenanced origin". "I wouldn't be surprised if in the coming years there is a large increase in the number of unprovenanced Greek antiquities sold openly," he says.

Christie's, the London auction house, has strict internal policies to ensure they only offer objects for sale legally, says Matthew Patton, head of communications.

The European Commission denies that spending cuts are to blame. "Greece has received very large sums that go towards cultural heritage. Many cultural institutions have been saved in Greece because of the work of the European Union," says Dennis Abbott, spokesman for Androulla Vassiliou, European Commissioner for culture and education. "But there are limits to what we can do. The main responsibility lies with the states," he says. ■

CORRECTION

The News story 'Journal offers flat fee for "all you can publish"' (*Nature* **486,** 166; 2012) stated that all co-authors on a paper must be members of *PeerJ*. In fact, only 12 co-authors need to be paying members. It also wrongly noted that *mBio* does not assess for impact or importance: it does.

Y. KARAHALIS/REUTERS

AN ILL WIND

WITH TURBINES THREATENING SOME BIRD AND BAT POPULATIONS, RESEARCHERS ARE SEEKING WAYS TO KEEP THE SKIES SAFE FOR WILDLIFE.

BY MEERA SUBRAMANIAN

Marc Bechard turned a worried eye skywards as he walked among the limestone hills at the southern tip of Spain. It was October 2008, and thousands of griffon vultures — along with other vulnerable raptors — were winging towards the Strait of Gibraltar and beyond to Africa. But first they had to navigate some treacherous airspace. The landscape on either side of the strait bristles with wind turbines up to 170 metres high, armed with blades that slice the air at 270 kilometres per hour.

Bechard, a biologist at Boise State University in Idaho, and colleagues from the Doñana Biological Station in Seville, Spain, had been hired to help the birds make it safely past 13 wind farms in Cádiz province. Each time the researchers spotted a raptor heading towards a turbine, they called the wind farm's control tower. Within minutes the blades slowed to a stop, and one more migrating bird soared past unharmed. Then the turbine swung back into action.

When the biologists weren't looking up at the sky, they were scouring the ground for carcasses of griffon vultures (*Gyps fulvus*), Spanish imperial eagles (*Aquila adalberti*) and other species. The Spanish Ornithological Society in Madrid estimates that Spain's 18,000 wind turbines may be killing 6 million to 18 million birds and bats annually. "A blade will cut a griffon vulture in half," says Bechard. "I've seen them just decapitated."

Wind turbines kill far fewer birds in general each year than do many other causes linked to humans, including domestic cats and collisions with glass windows. But wind power has a disproportionate effect on certain species that are already struggling for survival, such as the precarious US population of golden eagles (*Aquila chrysaetos canadensis*).

"The troubling issue with wind development is that we're seeing a growing number of birds of conservation concern being killed by wind turbines," says Albert Manville, a biologist with the US Fish and Wildlife Service in Arlington, Virginia.

The deaths caused by turbines have the potential to harm not only wildlife, but also the wind-energy industry, which is the fastest-growing source of power worldwide, according to the World Bank. With critics vilifying wind turbines as 'bird blenders', wind companies, governments and researchers are teaming up to mitigate the problem before it reaches a crisis point. Cádiz province, for example, requires all wind-energy projects to consider environmental issues, and helps to fund research on reducing any damage.

The early signs are that with targeted efforts, wind power and wildlife can cautiously coexist. Bechard and his colleagues, for example, lowered mortality at the Cádiz wind farms by 50%, with only a 0.07% loss in energy production[1]. Others are finding that minor changes in the design or operation of wind farms can bring major reductions in animal deaths.

Bechard and others believe research is crucial to making wind energy viable: "In the long run it'll save a lot of money and a lot of headaches."

THE GATHERING STORM

Wind power is poised to take off as the world seeks new sources of renewable energy. The industry is growing most quickly in China, which plans a 60% increase in wind power in the next three years. The US Department of Energy is aiming for a sixfold jump by 2030, and the European Union is working towards supplying 20% of its energy demand using renewable sources by 2020, much of that from wind.

But the rapid expansion of wind power can harm wildlife in multiple ways. Beyond direct collisions with turbines, wind farms threaten species by displacing habitat. And bats can develop fatal internal haemorrhaging

Raptors favour sites ideal for turbines, like these in Cadiz, Spain.

as a result of air-pressure changes when they fly through the wake of a spinning blade.

The industry maintains that the effects on wildlife are minor. Although there are only a few, limited estimates of bird fatalities at a national level, the available data for the United States suggest that wind farms account for a tiny fraction of avian deaths (see 'Bird killers').

But the concern is that turbines threaten species that are already struggling, such as bats, which in North America have been hit hard by white-nose fungus. Another vulnerable group is raptors, which are slow to reproduce and favour the wind corridors that energy companies covet. "There are species of birds that are getting killed by wind turbines that do not get killed by autos, windows or buildings," says Shawn Smallwood, an ecologist who has worked extensively in Altamont Pass, California, notorious for its expansive wind farms and raptor deaths. Smallwood has found that Altamont blades slay an average of 65 golden eagles a year[2]. "We could lose eagles in this country if we keep on doing this," he says.

Other species at risk include the critically endangered California condors (*Gymnogyps californicus*) — which number only 226 in the wild — and the few hundred remaining whooping cranes (*Grus americanus*), concentrated in the central United States. Biologists can't say whether the increase in wind farms will cause the collapse of these or other bird species, which already face many threats. But waiting for an answer is not an option, says Smallwood. "By the time we do understand the population-level impacts, we might be in a place we don't want to be."

DAMAGE CONTROL

In Cádiz, temporarily shutting down turbines has worked because the biggest threat is to migratory birds, which pass through only occasionally. Similar methods could reduce mortalities along the migratory bottlenecks in Central America, Europe and Asia, says Miguel Ferrer, a conservation biologist at Doñana and a co-author of the Cádiz study.

But that tactic will not work in Altamont Pass, which has both migratory and permanent avian populations. Instead, companies there are making headway by replacing small, ageing turbines with fewer large ones. Choosing sites carefully can help, too. "Raptors do not use the landscape randomly," explains Doug Bell, wildlife programme manager with the East Bay Regional Park District, which manages parklands and monitors wind farms around Altamont.

When the Buena Vista Wind Energy Project at Altamont replaced 179 turbines with 38 taller ones in 2006, Smallwood advised the company to avoid ridge saddles between hills and other hotspots for raptor traffic. Since then, golden eagle fatalities at Buena Vista have dropped by 50% and other raptor deaths by 75%, says Smallwood.

In the eastern United States, Todd Katzner, a biologist at West Virginia University in Morgantown, finds troublesome locations using a tracker designed to fit on golden eagles. "We can identify places where there are wins for both sides," he says. Moving a turbine site by a few hundred metres can substantially reduce the risk of collisions, says Katzner.

Companies say that they have learned from past mistakes. "That there hasn't been another [location like] Altamont Pass for years should be an assurance that the wind industry has gotten better at siting," says Stu Webster, director of permitting and environmental affairs

at Iberdrola Renewables, a renewable-energy firm in Portland, Oregon.

Sometimes a slight change in procedures can make a big difference. For example, most turbines are set to turn on when wind speeds reach 4.0 metres per second. But when the Iberdrola Renewables Casselman Wind Project in Pennsylvania increased the threshold to 5.5 metres per second, it slashed deaths of bats — which don't fly as much in high winds — by 93% while shaving just 1% off of power production, says Ed Arnett, who conducted a study there[3] while working at Bat Conservation International in Austin, Texas.

Some wind farms are betting on technology to make a difference. The MERLIN radar system, made by DeTect in Panama City, Florida, scans the skies for up to 6.5 kilometres around and uses algorithms to detect incoming flocks and even individual birds and bats, says Gary Andrews, chief executive of DeTect. Iberdrola Renewables and Pattern Energy, a wind-power company based in San Francisco, California, are using the system at wind farms, as is Torsa Renewables, based in Malaga, Spain.

Some researchers question the effectiveness of the radar. In 2010, a US Fish and Wildlife Service biologist saw a blade hit an American white pelican (*Pelecanus erythrorhynchos*) at Iberdrola's Peñascal wind farm in Texas, which was using MERLIN. But Iberdrola says that the system there is set up to prevent mass mortalities, not to detect individual birds.

According to Iberdrola, birds are most at risk in fog or other periods of low visibility, so Peñascal sometimes shuts down turbines during bad spells, unless the radar reveals that there are no birds around. Generally, however, the radar indicates that birds avoid the turbines by themselves. "We have terabytes' worth of data that are showing us how birds actually react to wind farms," says Webster.
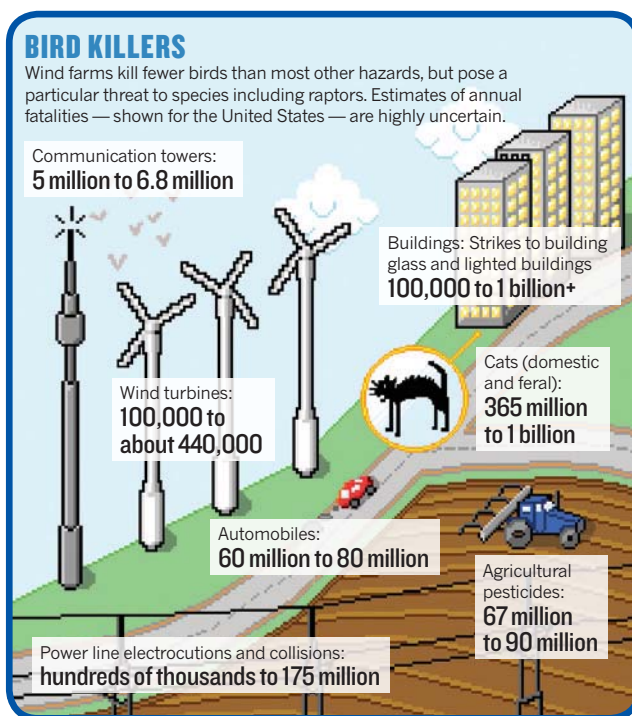
Researchers would love to see such data, but so far they have not been given access. "I haven't seen any results in the published literature," says Arnett, "so in my mind, [the radar] remains untested."

Companies worry that any monitoring data that they release could be used against them in lawsuits by environmental organizations or in political attacks from groups that support the fossil-fuel industries and want to scuttle wind power. The American Wind Wildlife Institute, a coalition of industry and conservation organizations based in Washington DC, is attempting to remedy the situation by creating a limited-access data repository that protects company privacy. The institute expects to complete the pilot phase of the project by mid-summer.

Even as they start to open up, some sectors of the wind industry acknowledge that they should do more to head off problems. "We as an industry need to do a better job of incorporating mitigation strategies into our economics," says John Calaway, director of wind development at Pattern Energy.

The strategies seem to be working in Cádiz, but Bechard still frets when he sees vultures on the horizon. "It's always nerve-wracking" wondering whether the turbines will stop in time, he says. When the birds pass safely, he breathes a sigh of relief. But with wind farms popping up around the globe, Bechard worries about what the vultures will encounter as they disappear into the distance. ∎

Meera Subramanian is a freelance writer in Cape Cod, Massachusetts.

1. de Lucas, M., Ferrer, M., Bechard, M. J. & Muñoz, A. R. *Biol. Conserv.* **147**, 184–189 (2012).
2. Smallwood, K. S. & Karas, B. *J. Wildlife Manage.* **73**, 1062–1071 (2009).
3. Arnett, E. B., Huso, M. M. P. Shirmacher, M. R. & Hayes, S. P. *Front. Ecol. Environ.* **9**, 209–214 (2011).

SOURCE: A. MANVILLE, US FISH AND WILDLIFE SERVICE

BIRD KILLERS

Wind farms kill fewer birds than most other hazards, but pose a particular threat to species including raptors. Estimates of annual fatalities — shown for the United States — are highly uncertain.

Communication towers: 5 million to 6.8 million

Buildings: Strikes to building glass and lighted buildings 100,000 to 1 billion+

Cats (domestic and feral): 365 million to 1 billion

Wind turbines: 100,000 to about 440,000

Automobiles: 60 million to 80 million

Agricultural pesticides: 67 million to 90 million

Power line electrocutions and collisions: hundreds of thousands to 175 million

A BROKEN CONTRACT

AS RESEARCHERS FIND MORE USES FOR DATA, INFORMED CONSENT HAS BECOME A SOURCE OF CONFUSION. SOMETHING HAS TO CHANGE.

BY ERIKA CHECK HAYDEN

Late in May, the direct-to-consumer gene-testing company 23andMe proudly announced the impending award of its first patent. The firm's research on Parkinson's disease, which used data from several thousand customers, had led to a patent on gene sequences that contribute to risk for the disease and might be used to predict its course. Anne Wojcicki, co-founder of the company, which is based in Mountain View, California, wrote in a blog post that the patent would help to move the work "from the realm of academic publishing to the world of impacting lives by preventing, treating or curing disease".

Some customers were less than enthusiastic. Holly Dunsworth, for example, posted a comment two days later, asking: "When we agreed to the terms of service and then when some of us consented to participate in research, were we consenting to that research being used to patent genes? What's the language that covers that use of our data? I can't find it."

The language is there, in both places. To be fair, the terms of service is a bear of a document — the kind one might quickly click past while installing software. But the consent form is compact and carefully worded, and approved by an independent review board to lay out clearly the risks and benefits of participating in research. "If 23andMe develops intellectual property and/or commercializes products or services, directly or indirectly, based on the results of this study, you will not receive any compensation," the document reads.

The example points to a broad problem in research on humans — that informed consent is often not very well informed (see 'Reading between the lines'). Protections for participants have been cobbled together in the wake of past controversies and have always been difficult to uphold. But they are proving even more problematic in the 'big data' era, in which biomedical scientists are gathering more information about more individuals than ever before. Many studies now include the collection of genetic data, and researchers can interrogate those data in a growing number of ways. Several US states, including California, are considering laws that would curtail the way in which researchers, law-enforcement officials and private companies can use a person's DNA.

The research coordinators who develop consent forms cannot predict how such data might be used in the future, nor can they guarantee that the data will remain protected. Many people argue that participants should have more control over how their data are used, and efforts are afoot to give them that control. Researchers, meanwhile, often bristle at the added layers of bureaucracy wrought by the protections, which sometimes provide no real benefits to the participants. The result is a mess of opinions and procedures that sow confusion and risk deterring people from participating in research.

"A lot of times researchers will say, 'Why can't we just go back to the way it was?', which was basically that we take these samples and people do it for altruistic reasons and everything's lovely," says Sharon Terry, president of the patient-advocacy group Genetic Alliance in Washington DC. "That worked in a prior age. I don't think it works today."

The concept of informed consent was first set out in the Nuremberg Code, a set of research-ethics principles adopted in the wake of revelations of torture by Nazi doctors during the Second World War. But in recent years, a series of mishaps over consent have undermined support for research. In 2004, for example, scandal erupted in the United Kingdom after parents found out that from the late 1980s to the mid-1990s doctors and researchers had removed and stored organs and tissues from patients — including infants and children — without parental consent. New laws were passed that required explicit consent for such collections.

Then, in 2010, the Havasupai tribe of Arizona won a US$700,000 settlement against Arizona State University in Phoenix. Individuals believed that they had provided blood for a study on the tribe's high rate of diabetes, but the samples had also been used in mental-illness research and population-genetics studies that called into question the tribe's beliefs about its origins. In the settlement, the university's board of regents said that it wanted to "remedy the wrong that was done".

The cases illustrate the divide between researchers and the public over what people need to know before agreeing to participate in research.

Many of the recent concerns over consent are driven by the rapid growth of genome analysis. Decades ago, researchers weren't able to glean much information from stored tissue; now, they can identify the donor, as well as his or her susceptibilities to many diseases. Researchers try to protect the genetic data through technological and legal mechanisms, but both approaches have weaknesses.

It is not enough to stripping out any information that would identify the donor, such as names and full health records, before the data are stored. In 2008, geneticists showed that they could easily identify individuals within pooled, anonymized data sets if they had a small amount of identified genetic information for reference (N. Homer *et al. PLoS Genet* **4,** e1000167; 2008). And it may become possible to identify a person in a public database from other information collected during a study, such as data on ethnic background, location and medical factors unique to the study participants, or to predict a person's appearance from his or her DNA.

Even legal mechanisms have vulnerabilities. In 2004, Jane Costello, a social psychologist at Duke University in Durham, North Carolina, was forced to go to court to defend the confidentiality of patient records from the Great Smoky Mountains Study. The study, which is just going into its third decade, examines emotional and behavioural problems in a cohort of people who enrolled as adolescents. A participant in the study was testifying against her grandfather, John Trosper 'JT' Bradley, who had been accused of sexual abuse. JT's lawyers subpoenaed the granddaughter's records from the study in hope that the information would undermine her credibility as a witness.

It meant a major crisis of confidence for Costello. "I was telling 1,400-plus people every time we saw them that 'your data are absolutely safe', and now I was in a position where I was told, 'No, that's not true'," she says. After Costello's day in court, in August 2004, the records remained sealed, but mostly because the judge did not believe that they would exonerate JT. The result provided no clarity about patient protections.

BETTER MODELS FOR CONSENT

One solution is to keep genetic information separate from demographic data. The BioVU databank at Vanderbilt University Medical Center in Nashville, Tennessee, for instance, contains DNA samples from patients treated at the hospital — 143,939 people as of 11 June. The DNA is linked to health records in a second database, called a 'synthetic derivative', in which the data are anonymized and scrambled in ways that, its creators say, make it difficult for anyone to work back from the database to verify a patient's identity. Sample-collection dates are altered, for example, and some records are discarded at random, so that it is not possible to know that someone is in the database just because he or she was treated at the hospital. Even researchers who work with the data cannot determine whose data they are using. The databank expects to include as many as 200,000 individuals by 2014, making it one of the largest collections of linked genetic and health records in the world.

But when it comes to consent, BioVU takes a different approach from many other programmes. Patients don't choose to participate; rather they are given the chance to opt out. Patients are asked to sign a 'consent to treat' form every year. It includes a box that they can tick to keep their DNA out of the database. That model helps BioVU to collect many more samples, and much more cheaply, than other projects can.

The opt-out model — which is used in only a few other places — troubles Misha Angrist, a genome policy analyst at Duke University, who says that it risks taking advantage of people when they are ill. "Even a routine visit to the clinic can be a vulnerable moment, and they're saying, 'Would you mind doing this for future generations, to help people just like you?'."

And legal challenges have shown the weaknesses of opt-out policies. Health officials are now destroying millions of blood samples taken from newborn babies in Texas and Minnesota because the families were not adequately informed that the samples, collected to screen for specific inherited disorders, would also be used in research.

Vanderbilt officials and researchers counter that they have run extensive public campaigns to ensure that people in Nashville are aware of BioVU and are comfortable with the way it works. They regularly consult a community advisory board about the project. And Vanderbilt's approach actually goes above and beyond what is required by federal law; because the synthetic derivative includes de-identified

READING BETWEEN THE LINES

Despite the work that goes into making consent forms clear and detailed, some participants say they are confused by the wording.

---

5. What are the benefits and risks of participating?

......If 23andMe develops intellectual property and/or commercializes products or services, directly or indirectly, based on the results of this study, you will not receive any compensation.

23andMe received a patent in May for its work on Parkinson's disease.
Some participants did not expect it to seek intellectual-property rights.
From 23andMe's research consent form **www.23andme.com/about/consent/**

---

2. I have been informed that the purpose of the research is to study the causes of behavioral/medical disorders.

Many participants in the Medical Genetics at Havasupai study in Arizona in the 1990s were unaware of how broad the research goals were and what kind of studies would be performed.
Courtesy of Pilar Ossorio

Re-identification
We are quickly learning that with powerful computers and good mathematicians, it is increasingly possible to uniquely identify people inside large data sets ... Researchers will sign a contract in which they agree that, even if they were able to identify you, they won't do it.

In an attempt to be more transparent about privacy risks, the Consent to Research project's 'Portable Legal Consent' essentially states that although anonymity cannot be guaranteed, researchers must pledge to uphold it.
Courtesy of John Wilbanks Consent to Research

---

data, it doesn't legally require informed consent at all. Last July, the US Department of Health and Human Services signalled that it might be rethinking the rules that exempt de-identified data from the consent requirement, as part of a broad overhaul of research ethics regulations.

Irrespective of the outcome, obliterating patient identities has drawbacks. Researchers can't perform some types of research on the scrambled data. Because dates are changed, studies on the timing of influenza infections, for example, are impossible. And patients can't be told if the research has revealed that they carry individual genetic risks linked to disease.

FULL DISCLOSURE

Returning study results to research participants has been another thorny issue for consent. Doctors might learn about genetic predispositions to disease that are separate from the ailments that led a patient to participate in the research in the first place, but it is not clear what they should do with this information.

UK researchers, for example, are forbidden from sharing genetic results with participants. But US research societies, such as the American College of Medical Genetics and Genomics in Bethesda, Maryland, are moving towards adopting standards that would encourage the practice for some types of findings, such as those that are medically relevant.

Some countries, such as Germany, Austria, Switzerland and Spain, are already feeding back such information. And some clinical sequencing programmes are considering offering patients 'tiered' consent, in which people can decide whether to be told about their data and how much they want to learn.

This is what Han Brunner, a geneticist at the Radboud University Nijmegen Medical Centre in the Netherlands, had hoped to do.

Last year, he began a project to sequence the exomes — the protein-coding regions of the genome — of 500 children and adults, looking for the genetic causes of intellectual disabilities, blindness, deafness and other disorders. Brunner proposed allowing participants to choose from three options: they could learn everything that researchers had divined about disease susceptibility; just information relevant to the disease for which their genomes were examined; or no information at all. Ethics reviewers shot down his proposal. "They said that in practice, it would be impossible for people to draw those lines, because people giving consent cannot foresee all the possible outcomes of the study," Brunner says. Instead, everyone participating in the studies must agree to learn all medically relevant information arising from the analysis of their genomes. As a consequence, Brunner recently had to tell the family of a child with a developmental disability that the child also has a genetic predisposition to colon cancer. Not all researchers endorse the idea of informing children about diseases that might affect them as adults. In this case, doctors recommended early screening, and Brunner says, "the family handled it very well; they said, 'This is not what we anticipated, but it's useful information'".

Many of the studies done now ask patients to give consent for research linked to particular investigators or diseases. But that means that researchers cannot pool data from separate studies to tackle different research questions. Many researchers say that the obvious solution is a broad consent document that gives researchers free rein with the data. But many non-scientists think participants should be able to control how their data are used, says lawyer Tim Caulfield of the University of Alberta in Calgary, Canada, who has surveyed patients about this idea. "There's an emerging consensus within the research community about the need to adopt things like broad consent, but that hasn't translated out to the legal community or to the public," he says.

Another solution might be called 'radical honesty'. A US project called Consent to Research, which aims to provide a large pool of user-contributed genomic and health data, has devised what it calls a 'Portable Legal Consent', which allows anyone to upload information about him or herself, such as direct-to-consumer genetic results and lab tests ordered through medical providers, to an interface that strips the data of identifiers. It makes the data widely available to researchers under broad guidelines, but also requires data donors to go through a much more rigorous consent process than most studies do. The Portable Legal Consent specifically informs participants that researchers might be able to determine their identities, but that they are forbidden from doing so under the project's terms of use.

Such approaches could help scientists by giving them access to a trove of data with no restrictions on use. But the participant protections system that is in place might not be ready for such frank dialogues, says Angrist, who serves on one of Duke's institutional review boards.

While reviewing a research proposal for a large biobank, for example, Angrist suggested that the researchers send the participants an annual e-mail explaining how their samples were being used, and thanking them for donating their time and tissue. The review board voted this suggestion down after its chair argued that e-mailing the patients would create a problem in light of the Health Insurance Portability and Accountability Act (HIPAA) — the US law that guarantees the privacy of health records. "The irony is that the HIPAA is supposed to protect people, and what I was hearing was, 'We can't talk to people because we're too busy protecting them'," Angrist says. "Institutions use informed consent to mitigate their own liability and to tell research participants about all the things they cannot have, and all the ways they can't be involved. It borders on farcical."

But as patient data become more precious to researchers, and as advocacy organizations become more involved in driving research agendas and in funding the work, such paternalistic attitudes will probably not survive, says Terry. She adds that technologies that allow research participants to control and track how researchers use their data will soon catch on. These approaches could benefit patients, who gain transparency and control over their data, and researchers, who gain access to richer data sets. "I think we're going to have to get to a place where consenting people becomes customizable easily through technology, and we're not there yet," Terry says. ■ **SEE EDITORIAL P.293**

Erika Check Hayden *writes for* Nature *from San Francisco, California.*

COMMENT

S. MAXWELL/ASPEN CENTER FOR PHYSICS

Summer workshops at the Aspen Center for Physics give researchers respite from their academic duties.

Aspen physics turns 50

Michael S. Turner reflects on how mountain serenity has bred big breakthroughs at the Aspen Center for Physics in Colorado.

Theoretical physicists are an odd lot: bad communicators (Niels Bohr and Werner Heisenberg); brilliant showmen (Richard Feynman and George Gamow); the 'strangest man' (Paul Dirac); lots of Hungarians (Leó Szilárd, Edward Teller and Eugene Wigner); bad hair (Albert Einstein); and too few women. They don't need fancy equipment — a pencil and paper will do. But they do like a serene environment, with blackboards and other people of their ilk, in which to come up with big ideas: among them relativity, the Big Bang, quantum mechanics and the atomic bomb.

Over the past 50 years, the Aspen Center for Physics (ACP), nestled in a beautiful valley at 2,400 metres above sea level in the Colorado Rocky Mountains, has provided a 'circle of serenity' during the summer months for

10,000 theoretical physicists, including 53 Nobel laureates, from 65 countries. The centre can lay claim to the string-theory revolution, the birth of the arXiv preprint archive and to setting the agenda for condensed-matter physics. Its history is tied to the revival of a silver-mining town and the American entrepreneurial spirit, and features a fascinating cast of characters, from philosopher Mortimer Adler to journalist Hunter S. Thompson.

The centre's story cannot be separated from that of the town. The 1893 repeal of the Sherman Silver Purchase Act demonetized silver and almost overnight turned Aspen, with a population of about 15,000, into a ghost town. Elizabeth Paepcke, wife of Chicago industrialist Walter Paepcke, visited in 1939, describing it as a place that "had slept since 1893". She found 700 residents, decaying

Victorian buildings and wonderful skiing. She persuaded her husband, a devotee of German writer Johann Wolfgang von Goethe, to visit in 1945. Seeing it as the ideal place to bring together the three aspects of life — economic, cultural and physical — he invested millions of dollars in rebuilding it. In 1946, he formed the Aspen Skiing Corporation, which remains the financial engine of the valley.

Aspen's cultural transformation came with the 1949 Goethe bicentennial. Organized by Walter Paepcke (with guidance from Adler and Robert Maynard Hutchins, then chancellor of the University of Chicago in Illinois), the bicentennial aimed to rehabilitate German culture and to revive humanism in the wake of the Second World War and the dawn of the atomic age. Around 2,000 people gathered in a tent designed by architect ▶

Eero Saarinen for the 20-day celebration. They included German–French theologian Albert Schweitzer, pianist Artur Rubinstein, philosopher José Ortega y Gasset and poet Stephen Spender. The event led to the formation of the Aspen Music Festival (now the Aspen Music Festival and School) and, in 1950, of the Aspen Institute for Humanistic Studies (now the Aspen Institute). Just as Paepcke had imagined, today the town brings together culture, wealth and athleticism — and a touch of glitz.

BEGINNINGS

The ACP's origins lie with physicist George Stranahan, heir to the fortunes of the Champion spark-plug company in California and a graduate student at the Carnegie Institute of Technology in Pittsburgh, Pennsylvania. In the late 1950s, he decided that he would rather do his physics during the summer months in the mountains of Colorado, where fishing and hiking provided a more enjoyable backdrop than did an office in steamy Pittsburgh. After a few years, he realized that theoretical physics was best done with others, and set out to draw physicists to Aspen. When he later moved to Colorado, Stranahan became the landlord and close friend of Thompson.

Stranahan got things going with help from Michael Cohen, a condensed-matter theorist at the University of Pennsylvania in Philadelphia, who was one of Feynman's few PhD students, and Robert Craig, executive director of the Aspen Institute. The Stranahan family's Needmor Fund paid for the first building, Stranahan Hall, designed by Bauhaus architect Herbert Bayer, who also planned the Aspen Institute campus. Cohen found the physics talent, and Craig convinced the Aspen Institute to create a physics division.

In spring 1962, a letter was sent out to the physics community tentatively announcing "the possibility of a summer physics institute". The purpose was "to provide a place for physicists to work on their own problems during the summer, in a stimulating physics atmosphere, and in a location with pleasant surroundings and natural beauty". That year, 42 brave souls came to Aspen to "pursue their work with minimal distractions".

The Aspen formula was — and still is — to bring the best theorists together in an informal setting for weeks or months, free from their usual responsibilities of students and teaching, and isolated from distractions. There, they could talk with one another, think big thoughts and come up with game-changing ideas. Physicists were housed two to an office and held discussions on a patio with a small blackboard, often accompanied by beautiful music from the town's music tent. Graduate students were excluded, differentiating Aspen from teaching summer schools. For many years, the buildings had only a handful of public phones. This put a limit on interruptions, but sometimes provided entertainment. I once overheard particle physicist Murray Gell-Mann of the California Institute of Technology quipping, "I don't know the English for it, but the Japanese is …".

An early attempt to bring together physicists and philosophers failed because of Adler's insistence that they agree on "the pyramid of knowledge", which had physics at the bottom and philosophy on top. Because of the clash of cultures and egos, the centre did not stay tied to the Aspen Institute for long and became an independent entity in 1968. Since then, the ACP has been run by physicists who volunteer their time, helped by just two full-time staff. More than 200 top theorists have shaped and guided the centre — including five Nobel laureates and Stephen Hawking.

An early grant from the Sloan Foundation was crucial, and Hans Bethe of Cornell University in New York donated part of his 1967 Nobel prize money. Bethe Hall, built in 1978, was named in his honour. Robert Rathbun Wilson, the first director and builder of Fermilab in Batavia, Illinois, visited Aspen in 1967 and convinced the US Department of Energy to build a large, temporary office building there, where Fermilab's facilities for particle-physics experiments were designed. The construction of Hilbert Hall, named after mathematician David Hilbert, almost tripled the number of physicists that the centre could accommodate.

SOLID FOUNDATIONS

In 1972, the US National Science Foundation became the ACP's main funder, with support from other US science agencies including the Department of Energy and NASA. In the mid-1990s, a $3-million fund-raising campaign led by astrophysicist David Schramm of the University of Chicago financed the final and largest building, Smart Hall.

George Stranahan, Michael Cohen and Robert Craig (left to right): the centre's three founders.

Contributions came from physicists, friends in the Aspen community and the Smart Family Foundation in Connecticut.

Three figures played a major part in establishing the ACP in the theoretical community: Philip Anderson of Princeton University, Bethe and Gell-Mann. Coincidentally, they all began coming to Aspen two years before receiving a Nobel prize. They set the agenda, served as scientific magnets and gave early legitimacy. Any high-energy theorist would kill to spend three weeks discussing physics with Gell-Mann; Bethe helped to get astrophysics going at the ACP; and Anderson shaped condensed-matter physics there for three decades.

Anderson set the tone for the condensed-matter field with his influential paper 'More is Different' (P. W. Anderson *Science* **177,** 393–396; 1972). Contrary to particle physics, in which scientists pursue a reductionist quest for simplicity at smaller and smaller scales, condensed-matter physics applies the basic rules to discover and study the often unexpected, emergent phenomena that arise in large systems with complicated interactions, such as superconductivity or biological systems. Today, biological physics has emerged as a major activity at the ACP.

Two other condensed-matter theorists played a crucial part: David Pines of the University of Illinois at Urbana–Champaign and Elihu Abrahams of Rutgers University in Piscataway, New Jersey. They pioneered workshops on the latest topics to attract a balance of researchers from universities and from industry (mostly Bell Labs). These workshops brought in young hotshots, keeping the talent pool fresh. One area of scientific focus, strongly correlated electrons in metals, laid the foundations for the current understanding of high-temperature and other unconventional superconductors.

In addition to Bethe's presence, astrophysics at the ACP was jump-started by the discovery in 1967 of pulsars and their identification as neutron stars. The exotic properties of pulsars — rapid rotation, superfluidity and superconductivity — intrigued Pines and other condensed-matter theorists. They brought in astrophysicists with expertise in relativity and nuclear physics, and work done at the centre linked pulsar glitches to superfluidity within neutron stars, advancing both fields. In 1972, NASA started funding an annual workshop, and astrophysics had a foothold in Aspen.

But it was cosmology that caused astrophysics to rise to the same level as particle physics and condensed matter. Around 1980, Schramm and others began to realize that theories of unification in particle physics might revolutionize the sleepy field of cosmology, which had been the province of astronomers since the time of Edwin Hubble. Aspen was the ideal incubator for this
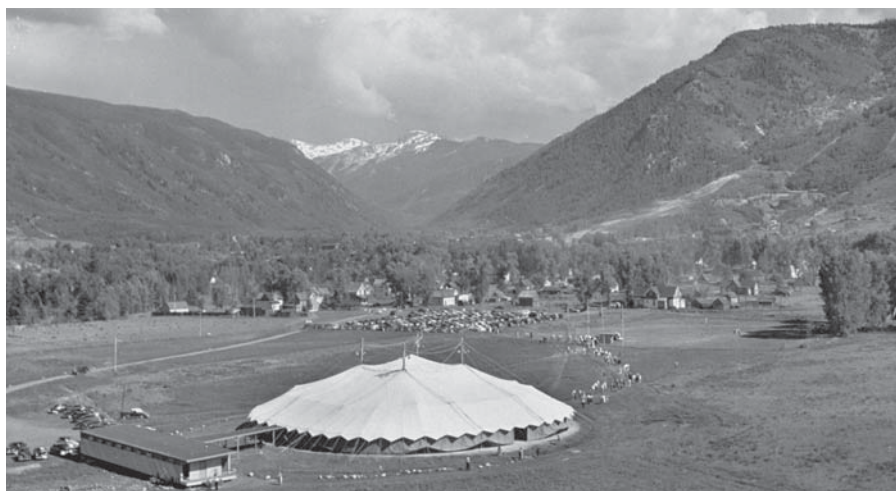
Aspen's cultural shift started with big-top celebrations for Johann von Goethe's bicentennial in 1949.

young, interdisciplinary field. Workshops brought together astronomers and physicists to discuss the hot topics — Big Bang nucleosynthesis, dark matter, inflation, large-scale structure, the cosmic microwave background and cosmic strings.

HIGH IMPACT

A staggering 10,000 or more papers are attributed to visits to the ACP. But its real impact is the big ideas that originated there. Much of today's consensus cosmology, with its particle dark matter, inflationary origins and dark energy, can trace its roots to the ACP. In his Nobel prize acceptance speech last December, Adam Riess of Johns Hopkins University described how his team, which co-discovered that the expansion of the Universe is speeding up, regularly met at the centre to chart its activities.

Whether or not string theory is the theory of everything, it has changed the course of physics. It began as a way to describe the strong interactions between neutrons, protons and related particles. Supersymmetry, the symmetry between bosons and fermions and a hallmark of today's string theory, traces its origins to theorist Pierre Ramond's first summer in Aspen in 1970, where, as he put it, he "stopped calculating and started thinking". When he got back to Fermilab, he prepared the paper that added supersymmetry to string theory. This has become the pattern: think in Aspen, calculate and write at home.

String theory was declared dead at a 1974 Aspen workshop, having been beaten by quantum chromodynamics as the best description of the strong (colour) interactions between the quark constituents of the hadrons. But John Schwarz of the California Institute of Technology decided to be bold and think bigger, touting strings as the path to unifying the forces of the subatomic world with gravity. For the next ten years, Schwarz, his collaborator Michael Green and a handful of others tried to make

good on this promise at the centre. In the summer of 1984, their breakthrough came with a historic paper that showed the mathematical consistency of string theory (technically the cancellation of anomalies), triggering the first string-theory revolution.

The Green–Schwarz discovery was announced immediately in grand fashion during a 'physics cabaret' at Aspen's historic Hotel Jerome. In a skit, Schwarz, playing the role of Gell-Mann, rushed onto the stage to announce that he had discovered the theory of everything — and was eventually carried off the stage by a man in a white coat.

Another revolution, in physics publishing, traces its origin to a chance encounter on an ACP bench in June 1991. Joanne Cohn, a young theorist at the Institute for Advanced Study in Princeton, had been running an informal preprint distribution service, e-mailing papers to hundreds of string theorists who wanted to get the latest results as quickly as possible. Paul Ginsparg, then at the Los Alamos National Laboratory, asked her why she hadn't automated the system. By the next day, Ginsparg had written some scripts, and two months later the Los Alamos arXiv (now residing at Cornell University) was opened for business. Today, more than 1 million articles are downloaded every week.

Physicists have always been attracted to mountains, to hike and to think. ACP co-founder Robert Craig was a world-class mountaineer who scaled some of the toughest and tallest mountains, including K2 in Asia, and many physicists who came to Aspen were serious climbers. A topographical map of the surrounding mountains is displayed prominently in Stranahan Hall. Aspen's combination of challenging hikes, unpredictable mountain weather and crumbling rock means that tragedies are not uncommon.

In 1988, ACP trustee Heinz Pagels, executive director of the New York Academy of Sciences, slipped on a loose stone while climbing on Pyramid Peak in Colorado and

fell to his death. A respected popularizer of physics as well as a brilliant theoretical physicist, Pagels' name was given to the centre's summer public lecture series.

Schramm, my mentor, was also an expert climber. One summer, bad weather trapped him and his climbing partner on the face of Colorado's Capitol Peak in freezing rain; they made it back to Aspen two days later, after many had given up hope. Sadly, in December 1997, Schramm died in a plane crash while flying his twin-engine turboprop from Denver to Aspen.

THE FUTURE

Today, the Aspen Center for Physics is thriving on a 1.6-hectare campus, surrounded by a large 'circle of serenity' of open space on the Aspen Meadows. The Aspen Institute, the Aspen Music Festival and School and the ACP are recognized as the town's three major cultural institutions.

Each summer, the ACP's 16-week programme and 10–15 workshops attract more than 500 leading theorists to work on the most important problems in physics. With delight, I note that about 20% of the attendees and members of the governing board are now women, and that the centre now has its first female president — progress for theoretical physics. Discussions fill the offices, halls, alcoves and patios; spontaneous volleyball games occur regularly. New ideas and collaborations made in this informal environment have launched the careers of hundreds of young theorists (myself included) and moved physics forwards.

Will the Aspen formula continue to be as successful as it has been for the past 50 years? Several theoretical-physics institutes now exist — the Isaac Newton Institute in Cambridge, UK, and the Kavli Institutes in Santa Barbara and Beijing. These have longer, more formal programmes, and lack the serenity of Aspen. The bigger challenge comes with the change in the way science is done today. It is more collaborative, more connected and 'more faster'. When the ACP was founded, collaboration required face-to-face interaction. Now, with e-mail and the Internet, many collaborators have never met in person.

Aspen continues to be a place to think, free from the constraints of everyday existence — as Paepcke said, a place for "lifting us out of our usual lives". In today's fast-paced world of science, the need for a circle of serenity is only more acute. ∎

Michael S. Turner *is director of the Kavli Institute for Cosmological Physics at the University of Chicago, Illinois 60637, USA. He is chair of the trustees of the Aspen Center for Physics, which he has attended since 1979. e-mail: mturner@kicp.uchicago.edu*
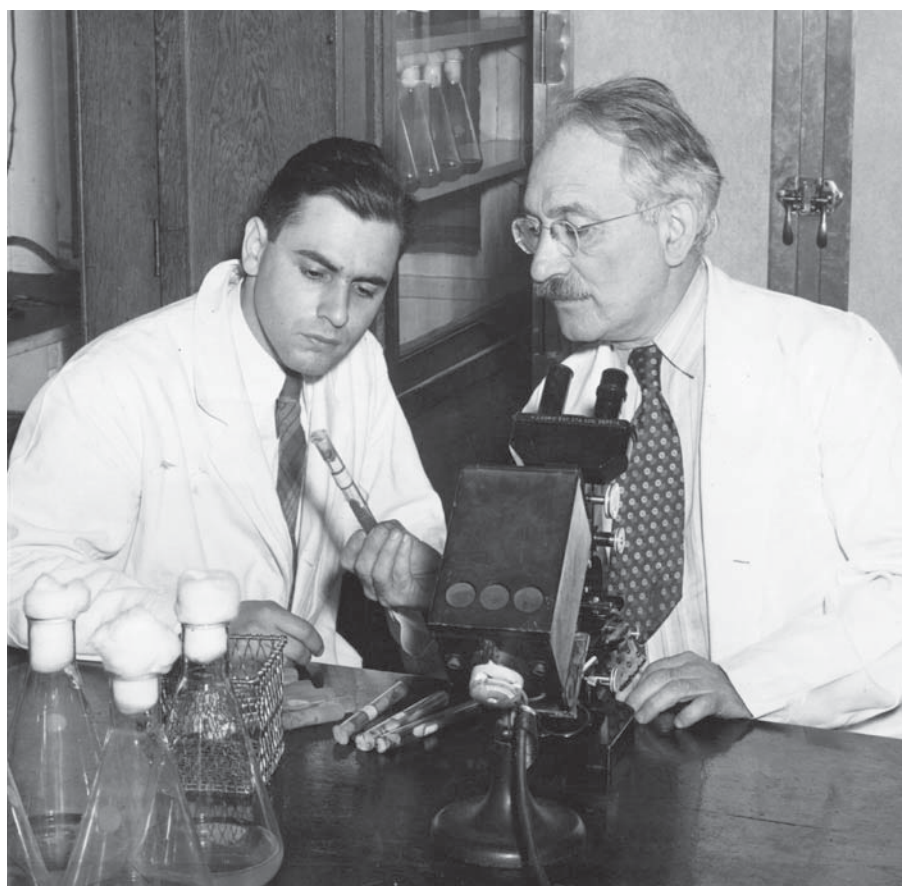
Albert Schatz (left) felt that he deserved a share of the Nobel prize awarded to Selman Waksman (right).

NOBEL PRIZE

A dark edge to the glory

An examination of the battles behind the prestige of top awards grips **Hidde Ploegh**.

Those who enjoy making predictions ahead of each year's Nobel announcements, only to criticize the winners when their choices lose out, may find satisfaction in *Prize Fight*.

Two stories of contested Nobel prizes drive the narrative in this readable book by the distinguished radiologist Morton Meyers. Meyers succeeds in chronicling these events without being overtly partisan, and includes a few vignettes of well-known and not-so-ennobling kerfuffles, with elliptical references to current events. Finally, Meyers discusses the general and fascinating ethical, professional and philosophical issues that arise from the prize fights.

Meyers first tackles biochemist Selman Waksman's 1943 discovery of streptomycin, and the ensuing altercation with Albert Schatz about attribution, intellectual property and commercial exploitation of this early antibiotic. Waksman — who developed soil microbiology and recognized that actinomycetes bacteria are nature's drug store for antibiotics — assigned to Schatz, his graduate student, the search for compounds active against Gram-negative bacteria. Schatz soon came up with *Streptomyces griseus* as a source of streptomycin.

This drug, the lucrative patent for which lists both Schatz and Waksman as its inventors, turned out to be highly active against *Mycobacterium tuberculosis*. It was responsible for remarkable therapeutic successes until the emergence of streptomycin-resistant strains and isoniazid, an antibiotic used in combination therapy, in the early 1950s.

Drawing on historical records and interviews with Schatz's widow, the author lays out a plausible and fascinating story. The failure to communicate and manage expectations led to a painful clash between Schatz and Waksman. A culture that viewed the appropriation of subordinates' discoveries by laboratory chiefs as legitimate deepened the chasm.

Schatz's attempts to redress the perceived injustice — including a lawsuit against Waksman and Rutgers University in New Brunswick, New Jersey, later settled out of court — probably excluded him from academic positions to which he was entitled by talent and previous accomplishment. Reclamations of this type inevitably invite the opprobrium of the larger scientific community, not unlike the fate of many a whistle-blower similarly castigated for attempting to right a wrong. On the fiftieth anniversary of streptomycin's discovery, however, Rutgers finally awarded Schatz its highest decoration: the Rutgers University Medal.

Meyers's second story deals with the development of nuclear magnetic resonance (NMR) imaging as a tool in medical diagnostics, featuring physician Raymond Damadian and his battle with the late Paul Lauterbur, a chemist. This is the livelier portion of the book. Meyers — as a radiologist, effectively gardening in his own backyard — had the added benefit of interviewing both Damadian and Lauterbur.

With physicist Peter Mansfield, Lauterbur in 2003 bested Damadian to Nobel recognition of magnetic resonance imaging (MRI). This outcome prompted the appearance of full-page ads financed by the Friends of Raymond Damadian in a number of newspapers, including *The New York Times*.

As Meyers shows, Damadian was the first to suggest using NMR in imaging to distinguish healthy from cancerous tissues. But he could not claim to have made the essential steps leading to wider practicability of the method. He built a massive machine, 'Indomitable', a prototype of which included a permanent magnet, as well as cardboard and copper foil components. Damadian used this monster to produce a low-resolution image of postdoc Larry Minkoff's chest — which was, indeed, a first. However, many
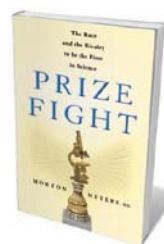
Prize Fight: The Race and the Rivalry to be the First in Science.
MORTON MEYERS
Palgrave Macmillan: 2012. 272 pp. £16.99/$27

of the financial dividends from Damadian's inventions accrued from patent infringement suits, not from building commercially successful instruments. Lauterbur's and Mansfield's contributions were what ultimately brought the technique to diagnostic practice.

More generally, Meyers uses *Prize Fight* to muse on the obsession with awards and publication in top-flight periodicals, which can ultimately devalue the passion and ingenuity of so many who will never share that limelight. External validation by one's peers is an important, but not the sole, driver of ambition. Taken to pathological extremes, the drive to satisfy this need fuels unethical behaviour and scientific misconduct. Those victimized by it may bear permanent scars or decide to leave science entirely. Milder cases abound. Disagreements and fights over authorship, priority and recognition occur wherever science is practised.

Meyers emphasizes the human nature of scientific pursuit. He reminds us that the individual scientist's contribution is ephemeral: the field moves on. Moreover, every scientist who assembles and leads a team of graduate students, postdocs and technicians has to contend with issues of priority, authorship and, less frequently, assignment of intellectual-property rights. The book has yet to be written that lays out such points of friction for the kinds of research that will never be recognized by the Nobel committee, yet drive entire disciplines relentlessly forward — a category into which most of the science that has a positive impact on society probably falls.

> *"Obsession with awards can devalue the passion and ingenuity of so many who will never share that limelight."*

In the airline industry and medicine, checklists have become an essential step in preventing human error, whether in tightening bolts or measuring drug dosages. In the coda to *Prize Fight*, Meyers provides a checklist of sorts on how to avoid landing oneself in the intellectual and emotional morass that permanently colours the outlook of deserving but unrecognized scientists.

This starts with being aware of the problem, making an effort to be consistent in attribution of authorship, and setting criteria for establishing credit. Easier said than done: these issues are unlikely to disappear any time soon. But we ignore them at the risk of creating toxic working environments. ∎

Hidde Ploegh *is professor of biology at the Massachusetts Institute of Technology and the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts. e-mail: ploegh@wi.mit.edu*

Books in brief
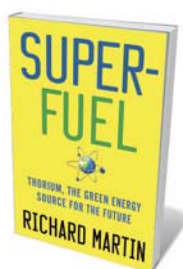
SuperFuel: Thorium, the Green Energy Source for the Future
Richard Martin PALGRAVE MACMILLAN *272 pp. £18.99 (2012)*
Post-Fukushima, uranium-powered plants face being phased out in many countries. But there is a nuclear alternative, argues clean-energy-research analyst Richard Martin: thorium. Less volatile than uranium, four times as abundant, energy-dense and efficient, thorium has major potential, not least because liquid fluoride thorium reactors create no nuclear waste. Martin's investigation reveals how the technology, developed at Oak Ridge National Laboratory in Tennessee, was dropped by President Richard Nixon in 1972 — and how interest is now picking up in China, India and elsewhere.

Born Together — Reared Apart
Nancy L. Segal HARVARD UNIVERSITY PRESS *416 pp. £36.95 (2012)*
The 'Jim twins' constituted a watershed in the nature–nurture debate. When Jim Lewis and Jim Springer — twins separated at four months — were reunited at 39, both were found to have loved maths, worked as sheriffs and practised carpentry, among other startling parallels. The case underlined the importance of genetics and led to the Minnesota Study of Twins Reared Apart. In this inclusive overview, Nancy Segal, director of the Twin Studies Center at California State University, Fullerton, examines the study that turned ideas on parenting, teaching, health and sexual orientation upside down.
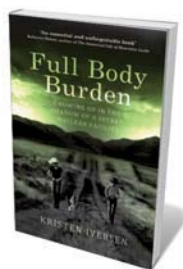
Full Body Burden: Growing Up in the Shadow of a Secret Nuclear Facility
Kristen Iversen HARVILL SECKER *416 pp. £14.99 (2012)*
For years, Kristen Iversen's mother thought that the industrial complex in their small Colorado town manufactured cleaning agents. But this was Rocky Flats — the US government facility where the plutonium 'pits' of nuclear weapons were manufactured. And, as Iversen reveals, it was plagued by safety issues. Among the appalling twists in this tale are high levels of testicular cancer among teenage boys in the area. After an inter-agency raid in 1989, pit production ceased; but Rocky Flats makes for a story with a long half-life.

Picturing the Book of Nature: Image, Text, and Argument in Sixteenth-Century Human Anatomy and Medical Botany
Sachiko Kusukawa UNIVERSITY OF CHICAGO PRESS *304 pp. £29 (2012)*
Science historian Sachiko Kusukawa probes the role of illustration in sixteenth-century medical treatises, before the advent of the microscope. Looking at Leonhart Fuch's *De historia stirpium*, Vesalius's *De humani corporis fabrica* and the unpublished *Historia plantarum* of Conrad Gessner, Kusukawa argues that such anatomical and botanic images were not simply records of natural phenomena, but varied visual experiments. His book is studded with illustrative gems, not least John Dee's 'pop-up' pyramids in *Of Euclid's Elements*.

Psychology in the Bathroom
Nick Haslam PALGRAVE MACMILLAN *184 pp. £50 (2012)*
Arcane sexual behaviours are the stuff of cocktail-party chat, whereas the "psychology of flatulence" and incontinence remain taboo. Psychologist Nick Haslam eases open the bathroom door on the many human behaviours associated with excretion. Drawing on clinical research, psychoanalytical theory, language, gender and more, he conducts a fascinating neurogastroenterological journey, from scatological slang and toilet graffiti to the psychological aspects of constipation and diarrhoea. 'Toilet reading' of a high order.

Birds such as golden eagles eat carrion, recycling the nutrients captured during an animal's lifetime.

ANIMAL BEHAVIOUR

Dissecting decay

Clive D. L. Wynne celebrates a lively exploration of the life-and-death cycle in the wild.

Despite focusing on death and decay, *Life Everlasting* is far from morbid; instead, it is life-affirming. Bernd Heinrich, emeritus professor of biology at the University of Vermont in Burlington, does a tremendous job of convincing the reader that physical demise is not an end to life, but an opportunity for renewal.

He was prompted to ponder mortality by an odd request from a friend, who asked if, when he dies, his corpse might be left out for the ravens on Heinrich's land in Maine. That set Heinrich off on a quest to understand the role of death in life. His journey in this book starts and ends in, and often returns to, his beautiful Maine woodland, where nature can be observed at close range, and occasionally friends show up with elderberry wine and guitars.

Heinrich opens with a succession of animal corpses that he watches being buried by beetles, colonized by maggots, hauled off by ravens and vultures, and putrefied by bacteria. He starts small, with a deceased mouse, and proceeds through ever larger bodies: a freshly killed squirrel, a rooster, deer and pig, and finally a massive bull moose. Most of these he places so that he can watch their decay comfortably from a chair in his cabin. Each corpse provides opportunities to describe, in loving detail, the life that death provides.

Beetles with beautiful wings use the dead mouse as a romantic meeting place. Having found each other, they carry the carcass to a safe place for burial. The female lays her eggs in soil close by. When the larvae hatch, the parents feed them regurgitated food from the corpse, staving off decay with antibiotic anal secretions.

Maggots prefer the deer carcass. Being larger, it stays warm for longer and encourages bacteria, whose "soupy by-products" provide sustenance for maggots. The maggots quickly colonize the corpse, and, because of their high metabolic rates, actually raise the temperature inside it and accelerate their own growth rate, creating a frenzied positive-feedback loop.

Looking at marine life, Heinrich contemplates the death of salmon and the dismantling of whales. When a whale carcass comes to rest on the ocean floor, it may be kilometres deep, in total darkness and at temperatures very close to freezing. Nonetheless, there is no shortage of specialist scavengers even in these extreme conditions, including sleeper sharks, hagfish and a wealth of tiny crustaceans.

Heinrich also reminisces about his time in Africa during the 1970s. In these time travels

Life Everlasting: The Animal Way of Death
BERND HEINRICH
Houghton Mifflin Harcourt: 2012. 256 pp. £15.99/$25

he considers the biggest of land beasts — the elephant. This animal's death brings humans into the picture: Heinrich argues that our ancestors were to elephants "what the sleeper sharks and hagfish are to the whales of the ocean depths: the ultimate recyclers".

Discussion of starving people in Zimbabwe today hacking at an elephant carcass segues into consideration of how early humans would have had to learn to cooperate to bring down mammoths, armed only with sharpened sticks and stones. Heinrich suggests that learning to process elephant meat taught us so much that tackling any other animal left us unfazed. He even details the decomposition of elephant dung, recounting with relish the time he watched half a litre of it be colonized by some 3,800 beetles in just 15 minutes.

Heinrich also considers the death of plants. When a tree falls in the forest and no one is around to take the wood away, what becomes of all that timber? In a gripping section on "plant undertakers", he considers how dead plants are broken down and the nutrients captured during their lives recycled. This process is led by sawyer beetles, jewel beetles and bark beetles, followed by horntail wasps, countless other beetles, fungi, birds such as woodpeckers, centipedes, millipedes and, if the tree falls into water, fish. Even other trees may use a rotting log as a base from which to grow.

Heinrich argues passionately that we cannot and should not fight the return of life to life through death. Death is the ultimate recycler. In the United States, most people choose to have their corpse either pumped with toxic formaldehyde and sealed in a steel box, or incinerated. The incineration of bodies in the United States burns enough fuel each year to power 80 trips to the Moon. In this way, we perpetuate in death the exclusion of ourselves from the natural world that many of us also proclaim in life.

Extinction, not death, is the real problem. Heinrich notes that unloved "undertaker" species such as ravens, vultures and condors are especially vulnerable to humans' effects on ecosystems. The largest cause of loss for such species is the complete replacement of large wild animals with farming activities that displace habitat and take away carcasses that nature would have left to return to the earth.

Ultimately, Heinrich is unable to indulge his friend's desire for a "green burial". In the United States, the movement and disposal of dead bodies is tightly bound by legal restrictions. He succeeds, however, in a larger aim. He replaces the inanimate biblical bookends to our lives, 'dust to dust', with what we really are at death: conduits for life. If nature has its way, the real progression is 'life to life'. ∎

Clive Wynne *is a professor of psychology at the University of Florida, Gainesville, Florida 32611, USA.*
e-mail: wynne@ufl.edu

M. GREY/NHPA/PHOTOSHOT

MATHEMATICS

A life computed

James Poskett navigates a sophisticated account of Alan Turing's extraordinarily varied intellectual world.

Alan Turing did not invent the computer. During the 1930s, well before the Manchester 'Baby', Pilot ACE or EDVAC machines, thousands were in operation all across Britain. These 'computers' were women, working in teams and each performing a discrete step of a complex mathematical operation.

Employed by the Scientific Computing Service (SCS) in the United Kingdom, these doughty women solved problems in everything from X-ray crystallography to jet-engine design. That world of 'computers before computers' is featured in the opening gallery of *Codebreaker*, the celebration of mathematician Alan Turing's life and legacy at London's Science Museum.

Although the SCS may seem far removed from Turing's world, knowledge of its work can help us to make sense of his seminal 1936 article 'On computable numbers, with an application to the *Entscheidungsproblem*'. This is viewed by some as the origin of the concepts behind the modern computer. But the real story is more complex.

Turing didn't pluck the idea of the modern computer out of thin air. He took the idea of a team of human computers working together

and abstracted it, imagining a universal computing machine that could take on all of the individual tasks allocated to the women in the SCS. This was merely a thought experiment for him at first — an aid for approaching David Hilbert's notorious 'decision problem' on the question of whether an algorithm exists for deciding if a given mathematical statement has a proof or not.

*Codebreaker —
Alan Turing's life
and legacy*
*The Science Museum,
London
(21 June 2012–
31 July 2013)*

Interestingly, although sometimes armed with little more than a pencil and slide rule, the women working for the SCS also used basic calculating machines — one of which sits in the first room of *Codebreaker*. These typewriter-like contraptions, covered in strips of red and white buttons, helped to speed up ballistics calculations crucial to the war effort.

The parade of objects that follows reveals Turing's impact, as well as his influences, leading the visitor through a vast intellectual landscape, from aeronautical design and biochemistry to cryptography and artificial intelligence.

In one gallery, a German Enigma machine, tucked into its neat wooden case, sits among archival photos that evoke the atmosphere of Second World War intelligence. Turing is well known to have worked in wartime cryptanalysis at the Government Code & Cipher School at Bletchley Park, UK. But he did not operate in isolation, collaborating with Gordon Welchman on the design of the electromechanical machines used to crack the code. These devices, each of which could mimic the action of several Enigma machines, in turn originated in the earlier work of Polish cryptanalysts such as Marian Rejewski.

In a nearby gallery, a twisted white metal fuselage serves as a poignant reminder of the part Turing's ideas played in developing safer air travel after the war. In 1954, the world's first commercial jet airliner, the de Havilland Comet registered 'Yoke Peter', exploded in mid-air, killing everyone on board — and prompted the Royal Aircraft Establishment at Farnborough, UK, to find the cause.

By this time, Turing's abstract idea of a universal computing machine had become a reality. The National Physical Laboratory in Teddington, UK, had in 1950 completed the Pilot ACE, an electronic computer designed to one of Turing's first practical specifications. For a short time, this computer was the fastest in the world. The huge rack of wires, relays and coloured transistors is displayed alongside the Yoke Peter wreckage. It helped to process the enormous amounts of data needed to complete detailed analysis of the debris, eventually revealing the point of structural weakness and prompting improvements in the design and manufacture of de Havilland jets.

The exhibition also explores links between the Pilot ACE and Dorothy Hodgkin's work on the structure of vitamin B_{12}. One of her original models of the vitamin, an intricate web of more than 100 red and blue balls, is on show next to the story of her use of computers. Cryptography and X-ray crystallography had much in common at this time: each involved recovering information from a scrambled signal.

Hodgkin's problem was that the images she produced indicated only the amplitudes of the diffracted waves. To establish the correct structure of a molecule, she needed to churn through an enormous number of possible permutations of different wave phases. So she enlisted the help of both the SCS and the Pilot ACE, going on to win the Nobel Prize in Chemistry for her work in 1964.

Turing's relationships with artificial intelligence and developmental biology are also on show through video interviews with contemporary mathematical biologists run next to programmable robotic tortoises. Although Turing didn't work on these machines himself, he was fascinated by the possibility of an artificial mind, coming to London especially to see robots such as these scurry across the floor at the Festival of Britain in 1951.

Codebreaker does an impressive job of bringing these diverse histories together. Turing is rightly celebrated, not as a lone genius, but as an impressive intellect and brilliant collaborator. ∎

Alan Turing worked on devices to crack the German Enigma machine's code during the Second World War.

James Poskett *is a science writer based in Cambridge, UK, specializing in the history and philosophy of science.*
e-mail: james.poskett@cantab.net

SCIENCE & SOCIETY PICTURE LIBRARY

Correspondence

More ways to govern geoengineering

You call for stronger governance of climate-mitigation strategies that reflect the Sun's energy away from Earth (*Nature* **485,** 415; 2012). We see the scientists' cancellation of a controversial field trial for the UK Stratospheric Particle Injection for Climate Engineering (SPICE) project (*Nature* **485,** 429; 2012) as responsible self-governance in the absence of the governmental oversight that is needed for solar geoengineering research.

The decision to cancel the SPICE balloon experiment can advance norms for research priorities and conditions of research. The underlying governance principles have been articulated by the Bipartisan Policy Center's Task Force on Climate Remediation Research and the Solar Radiation Management Governance Initiative (SRMGI), sponsored by the UK Royal Society, the Environmental Defense Fund and the Academy of Sciences for the Developing World (TWAS).

On the basis of the SPICE example, scientists can now decide — through projects, workshops and professional societies — that there should be no immediate research into deployment methods for geoengineering technologies, and that they will not engage in research that has intellectual-property implications. They can also learn from SPICE about public engagement and ways to make research transparent.

Eventually, legitimate governance must grow out of consultations with diverse constituencies (such as those sponsored by the SRMGI) and needs to come from governmental institutions that are fully accountable to society.
Jane C. S. Long *The Bipartisan Policy Center, Washington DC, USA. janecslong@gmail.com*
Steve Hamburg *Environmental Defense Fund, Washington DC, USA.*
John Shepherd *University of Southampton, UK.*

Use fast reactors to burn plutonium

Frank von Hippel and colleagues review some disposal options for radioactive plutonium waste (*Nature* **485,** 167–168; 2012). Another option is the profitable consumption of plutonium from thermal nuclear plants in a fast-spectrum breeder reactor with fuel recycling.

A prototype Integral Fast Reactor was operated at the Argonne West National Laboratory in Idaho for 30 years until 1994. 'Burning' spent nuclear fuel produces a fraction of the waste of current reactors, and it has low radiotoxicity (W. H. Hannum (ed.) *Prog. Nucl. Energy* **31,** 1–217; 1997).

The reactor's metal fuel (mainly uranium, plutonium and zirconium) and liquid-sodium coolant provide passive safety. An unpressurized pool vessel disperses decay heat by natural convection, even when cooling pumps are inoperable and the heat sink is lost.

The fuel-recycling system generates vast amounts of clean electricity, extending uranium supplies 150-fold — unlike today's once-through-and-throw-away cycle. Its proliferation risk is low because the products are unsuitable for use in fissile weapons.

The company GE Hitachi has designed an integral fast reactor, the 311-megawatt electric Power Reactor Innovative Small Module (PRISM), that is intended for commercial use. A prototype plant is already being considered in the United States, and the company has recommended these plants to the UK government for plutonium disposal (see go.nature.com/dwiqvg).
Barry W. Brook *University of Adelaide, Australia.*
barry.brook@adelaide.edu.au
Tom Blees, William H. Hannum *Science Council for Global Initiatives, Woodland, California, USA.*

Spread the risk of antibiotic research

There are commercial as well as scientific barriers to seeking out new antibiotics (*Nature* **485,** 439–440; 2012). These discourage the pharmaceutical industry from investing in further research and clinical development, particularly as current antibiotics are cheap, usually work satisfactorily and generate profits for their manufacturers and distributors.

Producing new antibiotics is costly, particularly at the clinical-trial stage because of concerns over safety and resistance. Intellectual property is an issue in developing new drugs based on established antibiotic classes.

We could wait until the clinical situation becomes severe enough for the private sector to step in with substantial investment, but that would be risky. A mix of private and public investment might work, particularly if the public sector were to take on some of the commercial risk in clinical trials (see *Nature* **472,** 32; 2011). A pragmatic strategy would be to define the clinical profiles of desirable new antibiotics and then to devise commercially viable routes for delivering them to the clinic.
Chris Schofield *University of Oxford, UK.*
christopher.schofield@chem.ox.ac.uk

Monitor sea pollution to stop strandings

Hundreds of small cetaceans were stranded along Peru's northern coast earlier this year. While the event is under investigation, Peru's government should be setting up programmes to monitor marine pollution and taking precautions to protect the coastal ecosystem.

The dead animals comprised mainly long-beaked common dolphins (*Delphinus capensis*) and Burmeister's porpoises (*Phocoena spinipinnis*). They had internal trauma and lesions that could have been caused by underwater noise effects (see go.nature.com/tbfi7n). Although military sonar is known to induce cetacean strandings, no naval exercises had been reported in the area. Neither had there been any seismic testing associated with gas and oil exploration, which can also be a contributor.

Persistent pollutants that accumulate in cetaceans could be a factor. These weaken cetacean immune systems, making them more susceptible to infection (P. Ross *Hum. Ecol. Risk Assess.* **8,** 277–292; 2002), exacerbated by food shortages during El Niño episodes and harmful algal blooms.
Juan Jose Alava *Simon Fraser University, Burnaby, Canada.*
jalavasa@sfu.ca

China must provide education on HIV

Cultural factors spanning 5,000 years force homosexual men in China to endure huge psychological and social pressures (H. Shang *et al. Nature* **485,** 576–577; 2012). Government authorities need to face up to reality and promote HIV research and public education about transmission of the virus to curb its spread.

Alongside the promotion campaigns of World AIDS Day, China should provide more ways to access up-to-date information on prevention and treatment of HIV infection. Infected people also need proper medical assistance, legal safeguards and humane care.
Jian Zhang, Nan Jiang *South China University of Technology, Guangzhou, China.*
zhangjian3954@126.com

Akira Tonomura

(1942–2012)

Physicist who pioneered electron holography.

Akira Tonomura changed the field of fundamental physics through microscopy. Like botanist Robert Brown before him, he opened up a new world to observation. In the nineteenth century, Brown's microscope revealed Brownian motion and the cell nucleus. In the twentieth and twenty-first, Tonomura's has shown us basic principles of the quantum landscape and its applications.

Over decades, Tonomura developed the extremely stable, phase-coherent electron beams needed for good holographic imaging, a technique that enables measurement of both the intensity and the phase of transmitted electrons. This allowed many 'thought experiments' of quantum mechanics to be done in practice, and revealed details of magnetic and electric fields at the nanoscale. Tonomura used electron holography to illuminate the wave–particle duality of electrons and to measure the magnetic fields of superconductors and other quantum effects under challenging conditions, earning him heroic status among electron microscopists. His achievements meant he was tipped several times for a Nobel prize. He died of pancreatic cancer in early May, at the age of 70.

Tonomura spent some of his early childhood in Hiroshima, Japan. Fortunately, his family moved away from the city two months before the fateful morning in early August 1945 when the nuclear bomb was dropped. Soon after graduating in physics from the University of Tokyo in 1965, Tonomura joined the central research laboratory of the Hitachi Corporation in Tokyo and, with crucial encouragement from the distinguished electron microscopist Hiroshi Watanabe, began his long career.

At that time, physicist Dennis Gabor, working at the engineering company British Thomson-Houston, based in London, had already proposed the idea of using holography to increase the resolution of electron microscopes, and the technique had passed some feasibility tests. The most immediate and spectacular applications of holography were in the field of light optics, in which it was used to create stunning three-dimensional images. This led to a Nobel prize for Gabor. Electron holography could advance no further until the invention of the electron biprism (a positively charged filament that causes the electron paths on either side of it to cross) by Gottfried Möllenstedt at the University of Tübingen in Germany, where Tonomura worked briefly in 1973–4. Using this technology, Tonomura created the first practical electron holography microscope for Hitachi in 1978.
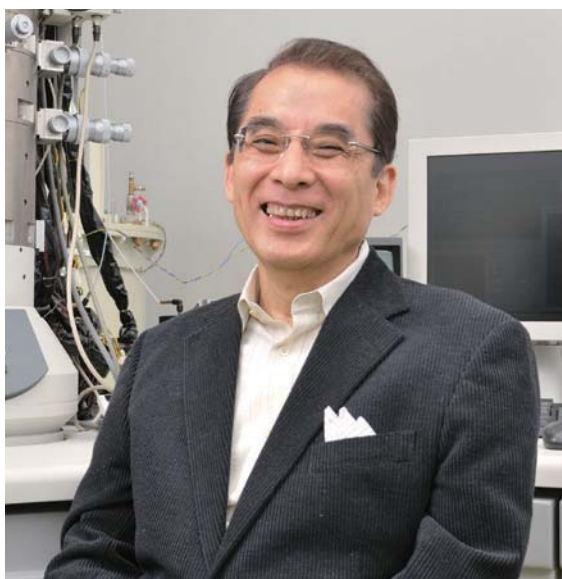
Tonomura's microscope proved crucial in settling a controversy over a bizarre quantum phenomenon. The Aharonov–Bohm (A–B) effect states that the phase of an electron's wavefunction can be shifted by a nearby magnetic field, even if the electron doesn't pass through that field. This idea sits uneasily with the classical theories that were used to develop practical electron optics. Early experiments hinted at confirmation of the A–B effect, but critics argued that stray fields might have caused the observed phase shift. So Tonomura placed a ring magnet inside a superconducting sheath to eliminate any stray magnetic fields, and encased that within a copper layer to stop a passing electron beam from entering the magnetic field region. He then used electron holography to confirm the predicted phase shift between electron paths inside and outside the ring. This conclusive and elegant experiment of 1986 finally silenced the critics, and was immediately recognized beyond the world of electron microscopy as a remarkable tour de force.

Few electron holography microscopes outpaced the resolution of ordinary electron microscopes as Gabor had envisaged, but Tonomura's use of holography to detect electron phases allowed him to pioneer and dominate the technique's practical application. In 1989, he scored a major success by imaging a magnetic vortex emerging from a superconducting film, and built on that with observations of vortices in various metallic and ceramic superconductors. Tonomura and others mapped magnetic fields in small particles, in magnetic tape and, most recently, in the skyrmion lattice — a periodic arrangement of magnetic vortices generated by a complex structure of electron spins. In quantum computing, observations of superconducting vortices are key to investigating the behaviour of these potential 'qubits'.

Tonomura combined stubborn persistence and experimental skill with imagination and excellent communication skills. He presented material with scrupulous care in publications and seminars. In a memorable Royal Institution lecture in 1994, he filled the strictly allotted one-hour time slot almost to the second. His images of magnetic phenomena were so striking that they were often on the cover of major journals. An Internet video of his version of the classic 'double-slit experiment' continues to demonstrate for many the central mystery of quantum mechanics (see go.nature.com/722hph). It shows how electrons travelling through a biprism arrive at a detector one by one, as particles, but over time build up a wave interference pattern.

Tonomura's stellar reputation and powers of persuasion helped to secure financial support from the Japanese government for his ambitious ideas. In 2010, he was awarded the largest grant for an individual research project in the country's history (see *Nature* **464,** 966–967; 2010). He became seriously ill a year later. In the spirit of Gabor's original idea for boosting resolution, the project aims to use high-voltage electron holography to create 3D images of electron wavefunctions. Its future now depends on securing a leader as inspiring as Akira Tonomura. ∎

Archie Howie is a physicist and electron microscopist in the Cavendish Laboratory, University of Cambridge, UK.
e-mail: ah30@cam.ac.uk

21ST-CENTURY GIRL

A child out of time.

BY ADRIAN TCHAIKOVSKY

"Do you dream of mammoths?" a talk-show host once asked me. I knew not to give him my first choice of answer; that would have enlightened his audience over the course of three and a half hours — so long as they had the basic grounding in biosciences required to understand it. I also knew enough to avoid my second answer, which was that his question was unintelligent, and that the entire interview had taken time I could have spent better in the laboratory.

What I actually said was: "No more than you do," and the sound that the audience made told me that they liked that. I had gained their sympathy somehow. I couldn't see precisely the mechanism by which this had happened, but I filed the memory away with all the others, my empirical evidence of the human condition by which I attempt to govern my social interaction with my fellow hominids.

Another one I get is: "You must have had a difficult childhood." That throws me because it isn't a question, and so you can't really answer it. It's a statement, to which the only response, if response is even required, is: "Yes." Of course I must. Why say the obvious, except that interviews are all about them saying the obvious, and me replying with lies and simplifications because that is, I have learned, what they want to hear.

When I was 15, my foster-parents took me aside for The Talk. The bulk of my difficult childhood was behind me, although I had yet to learn most of the coping strategies I now rely on. I was essentially friendless, more comfortable interacting online than off, an academic overachiever and unable to understand why that didn't come with the positive social pay-off that I had been led to expect. I didn't like crowds or strangers much. My world was comfortable with only a few other people in it.

No different, really, from hundreds of other children across the world.

I had thought that I knew what The Talk was going to be. They had never said, but I knew I wasn't their natural offspring, by deduction from first principles. I didn't look like them. I was built differently, and I'd spent ages looking at my face in the mirror, tracing the contours of nose and chin and forehead.

I was a striking girl. People who know about me now say I'm ugly, but that's a judgement influenced by their foreknowledge — they think I *should* be ugly, and so they recast my features in that unflattering light. Striking, is the word I prefer. Not even unique, if

you take each feature on its own. Not resembling my foster-parents, though.

I know, I told them. *I'm adopted*. They were unsure how to proceed. I could see that this was not, in fact, what The Talk was to be about. Perhaps I shouldn't have said anything, but that is something I still have difficulties with: knowing when to withhold knowledge. It seems so counter-intuitive to do so.

We had The Talk, at last. I wonder how many other disaffected, unsociable children wait for just that revelation: *you are something special; there is a reason why you are not like them*. The Talk was about adoption, in a way, about telling me who my parents were. My mother was stem-cell research and my father was gene sequencing.

When I was 20, and had been accepted for my doctorate, I made the decision to go public. You will recall the media storm. Nobody knew quite what to do with me. The geneticists behind my genesis had done something unethical, and yet at the same time their detractors wanted to study

me. There were legal battles, in which I was a determined participant. If I was to be a test subject, stripped of human rights, then at the same time my reviled creators were guilty of nothing more than making a *thing*. Alternatively, if they had broken the boundaries of professional ethics, then it could only be because I was a human being.

In having me raised among their own kind, in showing that I was intellectually, and at least borderline socially, functional, they sealed their own professional fate and secured mine. They must have known. They have been forgiven, since, because genius is too valuable a quality to waste. As for me …

I did not go into my current discipline purely because of my unique past. I became a geneticist because it is an area in which my cognitive strengths shine. My ability to find patterns in complex data, and to focus without distraction on the small details of my work, is as apposite for the minutiae of modelling gene-sequencing outcomes as it would have been for the painstaking production of exacting stone tools. In fact, the further I progress in my profession, the more I meet people who are just like me, despite our different heritages — and the less our different heritages matter in any meaningful way.

I will have sisters, soon, and brothers, as close to me as blood-kin. That project is proceeding specifically because I have been more than a success: I and my people have valuable intellectual traits that the world can use. I am not working on that team, though. I have other genomes to sequence, other verdicts of history to reverse.

And still people want to know, "What's it like, being you? What is it like to be brought back, to be taken from your proper time?" And I answer that this is my time, that I am a child of the twenty-first century. And if I, *Homo sapiens neoneanderthalensis*, did not evolve to live in cities and use the Internet and make advances in the field of genetics, then neither did *Homo sapiens sapiens*, and we will both have to make do.

And why would I need to dream of mammoths when these days I can step out of my office and just watch them? ∎

Adrian Tchaikovsky *was born in Lincolnshire, studied psychology and zoology at Reading and now practises law in Leeds. More of his work can be found at www.shadowsoftheapt.com.*

BRIEF COMMUNICATIONS ARISING

The Great Eruption of η Carinae

During the years 1838–1858, the very massive star η Carinae became the prototype supernova impostor: it released nearly as much light as a supernova explosion and shed an impressive amount of mass, but survived as a star[1]. In the standard interpretation, mass was driven outward by excess radiation pressure, persisting for several years. From a light-echo spectrum of that event, Rest *et al.*[2] conclude that "other physical mechanisms" are required to explain it, because the gas outflow appears cooler than theoretical expectations. Here we note that (1) theory predicted a substantially lower temperature than they quoted, and (2) their inferred observational value is quite uncertain. Therefore, analyses so far do not reveal any significant contradiction between the observed spectrum and most previous discussions of the Great Eruption and its physics.

Rest *et al.* state that a temperature of 7,000 K was expected, and that 5,000 K is observed. These refer to outflow zones that produced most of the emergent radiation. For the 7,000 K value those authors cite a 1987 analysis by one of us[3], but they quote only a remark in the text, not the actual calculated values. According to figure 1 in ref. 3, η Carinae's Great Eruption should have had a characteristic radiation temperature in the range 5,400–6,500 K, not 7,000 K. (Here we assume mass loss exceeding one solar mass per year and luminosity exceeding 10^7 solar luminosities[1].) The mention of 7,000 K in ref. 3 concerned less extravagant outbursts, and η Carinae was explicitly stated to differ from them. Moreover, to establish a conflict between observations and expectations, new calculations with modernized opacities would be needed.

The approximately 5,000 K temperature 'observed' by Rest *et al.* is based on a derived classification for the light-echo spectrum, using automated cross-correlations with a set of normal supergiant stars. This technique may be suitable for mass-production normal spectra, but any non-routine object requires specific feature-by-feature comparisons instead. One of the first principles of stellar classification is to separate luminosity from temperature criteria, but all the reference stars in this case were far less luminous than η Carinae's eruption. (Luminosity correlates with surface gravity, which affects gas density and thereby the spectrum.) Furthermore, emission lines appear to be present and may contaminate an automated analysis; but without access to the spectrum we cannot verify this. Rest *et al.* used a temperature calibration from a 1984 reference[4] taken from an even older publication in 1977[5]. Considerable work has been done since then, and for the highest luminosities, each spectral type has a substantial range of possible temperatures—for example, 5,100–6,200 K for the G2–G5

spectral types favoured in their paper[6–8]. The temperature range indicated by stellar classification thus overlaps the theoretical expectations.

Moreover, η Carinae's eruption was a large-scale mass outflow, not a static atmosphere with definable surface gravity. This distinction quantitatively alters the relation between absorption lines and the underlying continuum. The characteristic radiation temperature T_0 in the 1987 theoretical description[3] is therefore defined differently from a normal star's 'effective temperature'. If spectral types are assigned to outflows, there is no reason to expect their temperatures to coincide with the stellar-atmosphere calibration adopted by Rest *et al.*[1]. This is not a question of stellar wind versus explosion; dense winds, stellar eruptions, and opaque explosions are basically alike in their emergent radiation physics[1,3], and their density dependences $\rho(r)$ differ in character from normal stellar atmospheres. In conclusion, as far as existing models allow anyone to say, the observed spectrum appears to be consistent with what one expects for a giant eruption with η Carinae's parameters.

Kris Davidson[1] & Roberta M. Humphreys[1]
[1]Minnesota Institute for Astrophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA.
email: kd@umn.edu

1. Davidson, K. & Humphreys, R. M. (eds) *Eta Carinae and the Supernova Impostors* Vol. 384 (Astrophysics and Space Science Library, Springer Media, 2012).
2. Rest, A. *et al.* Light echoes reveal an unexpectedly cool η Carinae during its nineteenth-century Great Eruption. *Nature* **482,** 375–378 (2012).
3. Davidson, K. The relation between apparent temperature and mass-loss rate in hypergiant eruptions. *Astrophys. J.* **317,** 760–764 (1987).
4. Humphreys, R. M. & McElroy, D. B. The initial mass function for massive stars in the Galaxy and the Magellanic Clouds. *Astrophys. J.* **284,** 565–577 (1984).
5. Flower, P. J. Transformations from theoretical H–R diagrams to C–M diagrams: effective temperatures, colors and bolometric corrections. *Astron. Astrophys.* **54,** 31–39 (1977).
6. Soubiran, C. *et al.* The PASTEL catalogue of stellar parameters. *Astron. Astrophys. Suppl.* **515,** 111–116 (2010).
7. Cenarro, A. J. *et al.* Empirical calibration of the near-infrared Ca II triplet—II. The stellar atmosphere parameters. *Mon. Not. R. Astron. Soc.* **326,** 981–994 (2001).
8. Kovtyukh, V. V. *et al.* Accurate luminosities for F–G supergiants from FeII/FeI line depth ratios. *Mon. Not. R. Astron. Soc.* **408,** 1568–1575 (2010).

Rest *et al.* reply

In our Letter[1] reporting the light echoes of η Carinae we analysed the spectral characteristics of η Carinae during the Great Eruption of the mid-1800s, and found the line content to be similar to that of G supergiant stars. This we interpret as evidence that η Carinae's Great Eruption was not a typical luminous blue variable (LBV) outburst, because spectra similar to those of F and A supergiant stars, earlier and hotter than G-type stars, are observed in LBV eruptions of all kinds, in agreement with theoretical predictions. Davidson & Humphreys[2] object that our spectral type and temperature estimate

are not sufficiently robust, and that the spectral features are in agreement with their theoretical predictions.

The issues raised by Davidson and Humphreys[2] involve elementary considerations of stellar atmospheres and LBV spectra, as well as deeper ones of epistemology. We are well aware of these issues, but they are outside the scope of a *Nature* Letter. As they say, other things (such as chemical abundances) being equal, a spectrum is determined by the temperature and pressure. In stellar atmospheres, the latter is in turn determined by the gravity, or more fundamentally the mass and

BRIEF COMMUNICATIONS ARISING

radius. In a LBV outburst, the physical situation is entirely different and at present unknown; in the accompanying News & Views[3] Soker and Kashi describe their favoured hypothesis of a role for episodic binary mass transfer in the Great Eruption, but they also say that an instability in the primary itself is an equally viable hypothesis. Basically, we do not know whether Great Eruptions are late evolutionary events in all hypermassive stars, or whether they occur only in binary systems. There is no definitive model for these events, whether primitive or modern, and hence any derived physical parameters are highly uncertain. The comparison with stellar spectral types provides only a description of the line content of the LBV spectrum. By the same token, the comment that the Great Eruption of η Carinae was more luminous than the comparison supergiant stars is irrelevant.

The absolute temperature derived for an LBV outburst spectrum, whether observationally or theoretically, is virtually meaningless, because there are no reliable models for the physical structure that produced it. However, the relative spectral types and temperatures at different stages of these events, or among different LBV stars at similar stages, may be more meaningful and indeed are traditionally used by all LBV specialists for descriptive purposes. For example, during LBV outbursts the spectral type becomes later and the apparent temperature lower towards the visual-light maximum. The G spectral type at the Great Eruption peak, which we derived from detailed comparison of several spectral features with those of the supergiants, both visually and by cross-correlation, is unusually late and unprecedented for an LBV outburst. It may be related to the huge amplitude of this event or additional physical mechanisms in η Carinae's Great Eruption. Our suggestion of an explosion and blast wave is motivated by the large ratio of kinetic to radiative energy in the event[4], and by the direct observation of velocities up to several thousand kilometres per second in some of the ejecta[5].

A. Rest[1], **J. L. Prieto**[2,3], **N. R. Walborn**[1], **N. Smith**[4], **F. B. Bianco**[5,6], **R. Chornock**[7], **D. L. Welch**[8], **D. A. Howell**[5,6], **M. E. Huber**[9], **R. J. Foley**[7], **W. Fong**[7], **B. Sinnott**[8], **H. E. Bond**[1], **R. C. Smith**[10], **I. Toledo**[11], **D. Minniti**[12] & **K. Mandel**[7,13]

[1]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA.
email: arest@stsci.edu
[2]Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101, USA.
[3]Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, New Jersey 08544, USA.
[4]Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA.
[5]Las Cumbres Observatory Global Telescope Network, Goleta, California 93117, USA.
[6]Department of Physics, University of California, Santa Barbara, California 93106, USA.
[7]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.
[8]Department of Physics and Astronomy, McMaster University, Hamilton, Ontario, L8S 4M1, Canada.
[9]Department of Physics and Astronomy, Johns Hopkins University, Baltimore, 3400 North Charles Street, Maryland 21218, USA.
[10]Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Colina el Pino S/N, La Serena, Chile.
[11]ALMA, KM 121 CH 23, San Pedro de Atacama, II Region, Chile.
[12]Department of Astronomy and Astrophysics, Pontifica Universidad Catolica, Avenida Vicuna Mackenna 4860, Santiago 22, Chile.
[13]Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK.

1. Rest, A. et al. Light echoes reveal an unexpectedly cool η Carinae during its nineteenth-century Great Eruption. Nature **482,** 375–378 (2012).
2. Davidson, K. & Humphreys, R. M. Nature **486,** http://dx.doi.org/10.1038/nature11166 (2012).
3. Soker, N. & Kashi, A. Echoes from an old outburst. Nature **482,** 317–318 (2012).
4. Smith, N. et al. Mass and kinetic energy of the Homunculus Nebula around η Carinae. Astron. J. **125,** 1458–1466 (2003).
5. Smith, N. A blast wave from the 1843 eruption of η Carinae. Nature **455,** 201–203 (2008).

DRUG DISCOVERY

Computer model predicts side effects

Drug candidates are usually found to be unsafe only late in the drug discovery process. A method for predicting the many biological targets of a given molecule might allow drug safety to be considered much earlier. SEE ARTICLE P.361

KYLE KOLAJA

Creating safe, effective drugs is not for the faint hearted. Pharmaceutical companies screen up to millions of compounds to find the handful that are suitable for clinical development. The process typically begins with analogues of biologically active compounds being made to find those that interact potently and selectively with the drug target of interest. Attention then shifts to other properties of the compounds that could affect their use as drugs, including their safety profile. Unfortunately, predicting compound toxicity — let alone adverse side effects in patients — in preclinical studies is an onerous task. But on page 361 of this issue, Lounkine et al.[1] report a computational approach for predicting side effects that might help to streamline drug discovery*.

Identifying the mechanisms behind adverse drug effects remains inherently difficult, despite major investment in predictive tools by pharmaceutical and biotechnology companies (Fig. 1). The challenge is often magnified by the limited availability of many compounds, the need to identify toxicity risks in a short space of time and the difficulty in identifying the biological targets that mediate any observed toxicity[2]. Side effects can result from the intended action of a drug (such as the bleeding associated with medications that prevent blood clots) or because of unintended 'off-target' activity (such as the increase in blood pressure caused by torcetrapib, a compound that was designed to raise levels of 'good' cholesterol in the blood[3]). Any method that can rapidly and reliably identify the off-target activities of a compound, and the associated commercial liabilities, early in the drug-discovery process would be a powerful, must-have tool for a toxicologist.

Lounkine and colleagues use two steps to determine the biological targets for which a molecule may have affinity. The structural similarity of known ligand molecules for a given biological target is first calculated and the structural relationship of the group of ligands with a test compound is then assessed. This approach helps to reveal associations between

*This article and the paper[1] under discussion were published online on 10 June 2012.

Figure 1 | Assessing the risks. The adverse side effects of drugs are difficult to foresee. Lounkine and colleagues' computational method[1] for predicting such effects opens up a fresh approach for safety assessment in drug discovery.

compounds that would not be obvious from more conventional analyses of chemical structures or the amino-acid sequences of receptors. This mathematical method for predicting 'guilt by association' doesn't require any compound material to generate testable hypotheses, making it a neat way to rapidly interpret observations of toxicity.

To test their computational approach, Lounkine et al. used it to estimate the binding affinities of a comprehensive set of 656 approved drugs for 73 biological targets. They identified 1,644 possible drug–target interactions, of which 403 were already recorded in ChEMBL, a publicly available database of biologically active compounds. However, because the authors had used this database as a training set for their model, these predictions were not really indicative of the model's effectiveness, and so were not considered further.

A further 348 of the remaining 1,241 predictions were found in other databases (which the authors hadn't used as training sets), leaving 893 predictions, 694 of which were then tested experimentally. The authors found that 151 of these predicted drug–target interactions were genuine. So, of the 1,241 predictions not in ChEMBL, 499 were true; 543 were false; and

199 remain to be tested. Many of the newly discovered drug–target interactions would not have been predicted using conventional computational methods that calculate the strength of drug–target binding interactions based on the structures of the ligand and of the target's binding site.

The authors built on their results by identifying adverse side effects associated with a particular chemical structural fingerprint, again using a guilt-by-association mathematical approach. Their model predicted an impressive 247 new drug–target–side effect associations, shedding light onto previously unappreciated possible mechanisms of side effects. For example, it suggested that prenylamine (a drug that widens blood vessels) causes sedation by acting at the histamine H_1 receptor. And it predicted that the anti-allergy drug diphenhydramine causes tremors by acting not only at a known off-target receptor (the sodium channel SCN10A), but also at a dopamine transporter protein, SLC6A3, for which the drug was not previously known to be a ligand.

Lounkine et al. went on to provide strong experimental evidence of a previously unrecognized mechanism for the upper abdominal

pain caused by chlorotrianisene, a drug used to treat symptoms of menopause and deficiencies in ovary function. Their computational method predicted that chlorotrianisene has potent affinity towards the COX-1 enzyme; COX-1 is a target of non-steroidal anti-inflammatory drugs, a class of molecules that has long been associated with adverse gastrointestinal events[4]. The authors found that chlorotrianisene potently inhibits platelet aggregation in samples of human blood, which is a widely accepted biomarker of the *in vivo* effects of COX-1 inhibition and connects chlorotrianisene's adverse effect to a plausible mechanism. This example demonstrates how rapidly Lounkine and colleagues' computational predictions can glean information about the off-target toxicology of drugs.

Although the authors' computational approach is highly informative, it does have some deficiencies. Roughly half of its predictions of adverse drug activity are false, which

could prompt unnecessary biological experimentation in drug-discovery programmes, or even stop a safe compound from being developed further as a drug. And, as with any test, Loukine and colleagues' approach will overlook certain drug activities — although such false negatives are less disconcerting at the early stages of drug discovery because thorough testing of on- and off-target effects will subsequently be conducted as a matter of course. It is also worth noting that the authors' assessment of their technique's performance is based on a retrospective training set; prospective use will be required to fine-tune the model and to discover how well it truly performs.

Nonetheless, a computational method for predicting which biological sites are likely to be targeted by compounds provides a new way to improve safety assessment in drug discovery. Because Lounkine and colleagues' approach doesn't require compounds to be synthesized

or wet-lab experiments to be conducted, it will be useful very early in the discovery process to predict the potential for off-target binding of drug candidates, and so to prioritize which compounds should be made by medicinal chemists. The ability to associate side effects with predicted off-target activity means that chemists can be guided not only by the efficacy of compounds, but also by safety concerns, as soon as a drug-discovery programme is initiated. ∎

Kyle Kolaja *is in the Department of Nonclinical Safety, Hoffman-La Roche, Nutley, New Jersey 07110, USA. e-mail: kyle.kolaja@roche.com*

1. Lounkine, E. *et al. Nature* **486,** 361–367 (2012).
2. Sasseville, V. G. *et al. Chem. Biol. Interact.* **150,** 9–25 (2004).
3. Schwartz, G. G. *Curr. Atheroscler. Rep.* **14,** 41–48 (2012).
4. Park, S. C., Chun, H. J., Kang, C. D. & Sul, D. *World J. Gastroenterol.* **17,** 4647–4653 (2011).

MATERIALS CHEMISTRY

Carbon origami

A reaction that folds up large aromatic molecules and fixes them into bowl shapes expands opportunities for making nanometre-scale objects from single sheets of carbon. Such objects have potential applications in electronics.

JAY S. SIEGEL

Single layers of carbon atoms, known as graphene sheets, have been fêted as the material of choice for future generations of electronic devices. But carbon can form a broad range of networks besides the hexagonal 'chicken-wire' arrangement of atoms found in graphene. For example, networks formed from combinations of rings of five, six or seven carbon atoms can result in bowl, tube and saddle morphologies[1] that have diverse physical and electronic properties of use in applications ranging from structural engineering to light–energy conversion. Although they may form spuriously in nature, their selective preparation is difficult, requiring a high level of expertise in the crafts of molecular design and chemical synthesis. Reporting in *Angewandte Chemie*, Amsharov *et al.*[2] describe a practical method for preparing bowl-shaped carbon structures (buckybowls) in high yields, offering the tantalizing prospect that other extended, non-planar networks of carbon atoms will be accessible in the future.

The authors start from polynuclear aromatic compounds (Fig. 1), which can be thought of as small graphene fragments. Transformation of these fragments into buckybowls can be achieved by forming carbon–carbon (C–C)

bonds exclusively across 'cove' regions in the molecules — regions in which the carbon skeleton of the molecule curves inwards like the coastline around a cove. These reactions result in the formation of five-membered rings that induce the curvature found in carbon bowls and tubes; however, selectively forming C–C bonds across cove regions is easier said than done.

Amsharov *et al.* offer a solution to this problem by heating a polynuclear aromatic compound that bears a fluorine atom in the cove region with an alumina catalyst (Al_2O_3). This is a remarkable approach because carbon–fluorine bonds are generally regarded as being very strong, and therefore unreactive. The authors observed that only fluorine atoms in cove regions take part in the transformation — when they performed their reactions on starting materials containing fluorine atoms at non-cove regions, these atoms remained unaltered in the product.

The transformation occurs at temperatures substantially lower than those required by alternative methods for making carbon networks, such as flash vacuum pyrolysis[3] (FVP), in which the reactant is rapidly heated in a vacuum, or chemical vapour deposition[4], in which starting materials react on the surface of a substrate to make a thin film. What's more, the method works well with solid substrates

in the absence of a solvent, affording well-characterized products in good yields. This is handy, because many of the substrates that are likely to be used in the reaction are insoluble in most solvents. The authors also report that, in compounds that have several cove regions bearing a fluorine atom, C–C bonds can form at multiple regions.

Amsharov and colleagues' reaction is one of several transformations[5] that create new C–C bonds between aromatic rings, starting from aryl halides — compounds in which an aromatic ring is attached through a carbon atom to a halogen, typically chlorine, bromine or iodine. For example, FVP is known to effect intramolecular reactions of this sort by eliminating hydrogen halide molecules (HX, where X is a halogen) from aryl halides. These FVP processes involve highly reactive intermediates, such as free radicals, carbenes or benzyne. Aryl halides can also react on certain metal surfaces to form ribbons or star-shaped junctions of graphene[6]; catalysis by the metal makes the reaction possible.

Most such reactions of aryl halides work best with starting materials that bear weak carbon–halogen bonds: carbon–iodine bonds are generally the weakest, then carbon–bromine bonds, followed by carbon–chlorine bonds. Carbon–fluorine bonds are typically even stronger than carbon–chlorine bonds, and so Amsharov and colleagues' use of these in C–C bond-formation reactions is unusual, with few other examples known[7,8].

The development of strategies to manufacture graphene and related carbon-based materials that have designed shapes and symmetries[9] is one of the foremost challenges in materials chemistry. Similarly, the race to find an industrial method for producing graphene sheets of decimetre-to-metre dimensions is hotly contested. Recent developments, such
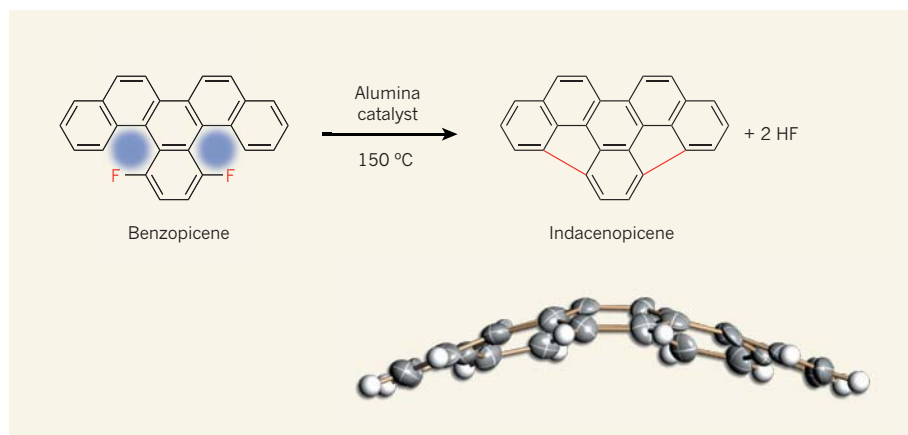
Figure 1 | Bent into shape. Amsharov *et al.*[2] report a method for making buckybowls — nanometre-scale objects that can be thought of as a single layer of carbon atoms bent into a bowl shape. They used a reaction in which a carbon–carbon bond forms across the 'cove' regions (blue areas) of polynuclear aromatic compounds, such as the benzopicene molecule in this example. The reaction requires an alumina (Al_2O_3) catalyst and occurs only where fluorine atoms are present at cove regions. The fluorine atoms in the starting material, and the bonds formed during the reaction, are shown in red. The structure shown beneath the reaction is the X-ray crystal structure of the buckybowl product, indacenopicene.

as kilogram-scale processes for preparing the archetypal buckybowl, corannulene[10], will help by increasing the commercial availability of building blocks for carbon nanotubes and graphene fragments.

To understand one way in which such fragments could be used, consider the example of a fullerene — a 'buckyball' of carbon atoms, in which the atoms form the vertices of a network of polygons that makes up a spherical surface. Imagine cutting the network open. Depending on where you cut, various two-dimensional projections can be obtained[11,12] that can be folded back into the ball. If polynuclear aromatic compounds are available that correspond to such projections, then synthetic methods such as that of Amsharov *et al.* could be used to fold up the molecules and 'glue' the edges together.

The synthesis of specific projections is therefore a fashionable strategy for preparing designer fullerenes[13]. Compounds such as corannulene are good starting points for many such projections, so having a readily available supply opens up opportunities to make these materials[14]. In the same way, the synthesis of molecules that can be thought of as cross-sections of carbon nanotubes has become something of an art form[15,16], with one idea being that these could be stacked up and joined together to make nanotubes.

It is worth noting that, although many aspects of chemical synthesis may be thought of as having the character of engineering, chemical methods are not generally transferable to every synthetic problem — unlike universal engineering tools, which can be used transferably to make almost anything. The optimal construction of molecular targets often requires finding exactly the right tool for each job. Nevertheless, seemingly specialized synthetic methods do evolve into

everyday tools used by chemists around the world and, ultimately, into scalable processes for industrial chemical production. In this way, Amsharov and colleagues' findings, and those of others aiming to make similar materials, are vital because they reveal new types of efficient chemical transformation that might

some day form the basis of the manufacture of graphenic materials tailored to a variety of applications. ∎

Jay S. Siegel *is at the Organic Chemistry Institute, University of Zurich, Zurich CH-8057, Switzerland.*
e-mail: jss@oci.uzh.ch

1. Ebbesen, T. W. *Phys. Today* **49** (June), 26–32 (1996).
2. Amsharov, K. Y., Kabdulov, M. A. & Jansen, M. *Angew. Chem. Int. Edn* **51**, 4594–4597 (2012).
3. Brown, R. F. C. *Pyrolytic Methods in Organic Chemistry: Application of Flow and Flash Vacuum Pyrolytic Techniques* (Academic, 1980).
4. Bae, S. *et al. Nature Nanotechnol.* **5**, 574–578 (2010).
5. Hassan, J., Sévignon, M., Gozzi, C., Schulz, E. & Lemaire, M. *Chem. Rev.* **102**, 1359–1469 (2002).
6. Cai, J. *et al. Nature* **466**, 470–473 (2010).
7. Allemann, O., Duttwyler, S., Romanato, P., Baldridge, K. K. & Siegel, J. S. *Science* **332**, 574–577 (2011).
8. Douvris, C. & Ozerov, O. V. *Science* **321**, 1188–1190 (2008).
9. Simpson, C. D. *et al. J. Am. Chem. Soc.* **126**, 3139–3147 (2004).
10. Butterfield, A. M, Gilomen, B. & Siegel, J. S. *Org. Proc. Res. Dev.* **16**, 664–676 (2012).
11. Scott, L. T. *et al. Science* **295**, 1500–1503 (2002).
12. Otero, G. *et al. Nature* **454**, 865–868 (2008).
13. Scott, L. T. *Angew. Chem. Int. Edn* **48**, 436–437 (2009).
14. Scott, L. T. *et al. J. Am. Chem. Soc.* **134**, 107–110 (2012).
15. Omachi, H., Segawa, Y. & Itami, K. *Acc. Chem. Res.* http://dx.doi.org/10.1021/ar300055x (2012).
16. Jasti, R., Bhattacharjee, J., Neaton, J. B. & Bertozzi, C. R. *J. Am. Chem. Soc.* **130**, 17646–17647 (2008).

GENOMICS

The breast cancer landscape

Whole-genome sequencing of breast cancers is exposing the scope of tumour diversity and helping to pinpoint avenues for precise diagnostics and targeted therapy. See Articles p.346 & p.353, Letters p.395, p.400 & p.405

JOE GRAY & BRIAN DRUKER

The information-generating power of genome-analysis technologies is increasing at a rate that surpasses even the doubling in computer performance that is achieved every 18 months by the semiconductor industry[1]. Genome-analysis methods are now sufficiently powerful, fast and reliable that they are underpinning efforts to elucidate the molecular architecture of human cancers[2,3] and, in some cases, can be used in routine clinical practice[4]. Five papers[5–9] published in this issue present whole-genome analyses of a range of different breast cancers, providing a comprehensive picture of breast cancer's genetic diversity and suggesting refined tumour-classification strategies and new lines of therapeutic attack.

Genome-analysis techniques allow detection of various genetic abnormalities, including DNA sequence changes, structural changes (such as translocations) that alter the order in which DNA segments occur in the genome, and copy-number variants, in which whole segments of DNA are deleted, duplicated or amplified. These techniques can also provide information about the transcriptome — the complement of RNA molecules that are transcribed from DNA — at a particular time or in a particular tissue. Furthermore, they can give an indication of epigenetic modifications, which are chemical modifications to DNA and associated proteins that regulate gene transcription rates without changing the nucleotide sequence.

Early transcriptome-wide studies of patients with breast cancer demonstrated that the

tumours could be organized into subtypes on the basis of patterns of gene expression that differed in clinical outcome[10]. This advance led to the development of clinical assays that are now used to identify patients at sufficiently low risk of cancer recurrence to warrant surveillance in lieu of chemotherapy[11,12]. Genome-wide measurements of DNA sequence, copy number, structure and gene-expression levels during the past decade have revealed remarkably diverse derangement in individual breast tumours, among different tumours and during various stages of tumour development. These aberrations involve many genes, including several implicated in cancer[13–15].

Expanding on these analyses are genome-wide functional studies[16], which have begun to identify aberrant genes that contribute to the changes in physiological processes that occur in breast cancer, in the hope that these might yield to therapeutic attack. These studies were inspired by the development of therapies that effectively treat tumours in which the gene *ERBB2* is amplified[17]. In parallel, *in vitro* systems that model aspects of breast-cancer diversity are being used to identify molecular features that predict therapeutic responses[18,19]. These efforts are now being stimulated by the development of high-resolution and high-throughput techniques, including microarray analysis and massively parallel DNA and RNA sequencing, that provide detailed information about the nature of cancer-causing aberrations[20].

The five papers in this issue describe the application of these techniques to a range of breast cancers. Curtis *et al.*[5] (page 346) analysed copy number, sequence changes known as single nucleotide polymorphisms, and gene-transcription rates in approximately 2,000 breast cancers encompassing all known types. They defined 45 regions of sequence amplification or deletion that deregulate genes that are likely to be involved in the pathophysiology of breast cancer. These regions are now sufficiently well defined for researchers to commence studies of the driver genes expected to lie therein. Curtis and colleagues also integrated their measurements of copy number and gene expression to define new breast-cancer subtypes that are associated with different patient outcomes, although these associations are not yet strong enough to be applied in the clinic.

Triple-negative breast cancers are those in which the tumours express neither *ERBB2* nor the genes that encode the receptors for oestrogen or progesterone. Shah *et al.*[6] (page 395) assessed mutations, copy number and gene expression in triple-negative breast cancers and found that the frequencies of copy-number abnormalities and mutations vary markedly between and within the tumours, which indicates that mutations can arise at multiple stages of tumour progression. Their

study also suggests that three genes, *TP53*, *PIK3CA* and *PTEN*, are involved in the early stages of breast-cancer development. Interestingly, only one-third of the low-prevalence mutated genes that the authors identified were transcribed into RNA, which suggests that they may be chance mutations unrelated to the cancer and/or that the mutations involved genes with tumour-suppressive activity.

Stephens *et al.*[7] (page 400) analysed somatic mutations and copy-number variants in 100 breast cancers and discovered numerous aberrations, including nine new cancer genes that were mutated rarely but more frequently than would be expected by chance, indicating that they have functional roles in the cancers in which they arise. Many of these mutations are predicted to lead to the expression of truncated proteins, which implies that the proteins, in their normal forms, are likely to have tumour-suppressor roles. The authors also report that certain DNA-base substitutions are strongly associated with the age of the patient in tumours that do not overexpress the ostrogen receptor (oestrogen-receptor-negative tumours), but not in oestrogen-receptor-positive (overexpressing) ones, suggesting an important difference in the dynamics of mutation accumulation in these two tumour types.

> "The data in these five manuscripts are a remarkable testament to the power of genomic technologies."

Ellis and colleagues[8] (page 353) focused on oestrogen-receptor-positive breast cancers. They assessed the genomes of tumours from patients participating in pre-operative clinical evaluation of their response to a class of drugs called aromatase inhibitors. The researchers showed that tumours that have a high frequency of cells expressing a protein, Ki67, which is associated with resistance to aromatase inhibitors, contained an elevated frequency of somatic mutations (those that arise during tumour progression, rather than being inherited) and genome-structure changes compared with tumours with a low frequency of Ki67-positive cells. This finding implicates genetic changes that lead to deregulation of DNA replication and repair processes in this drug resistance. However, a single 'smoking gun' predictor of resistance did not emerge from this study.

Finally, Banerji and colleagues[9] (page 405) examined somatic mutations in diverse breast cancers and report a rate of non-silent mutations (those that change the amino-acid sequence of a protein) of around one mutation per million DNA base pairs. Their study also added three aberrations to the list of those implicated in breast cancer — mutations in the genes *CBFB* and *RUNX1*, and the *MAGI3–AKT3* gene fusion.

The data in these five papers are a remarkable

testament to the power of genomic technologies to define the landscape of a complex disease state, and they give us the most thorough view yet of the molecular underpinnings of breast cancers. However, much remains to be accomplished to move this information into routine clinical practice. These studies, and those that preceded them, show that individual breast cancers typically carry a few consistent and functionally characterized abnormalities, along with tens to thousands of other changes that are rare or unique to the individual tumour and about which little is known. We need to understand which of these genes contribute to tumorigenesis or tumour progression and how this ensemble of aberrations collaborates, changes during tumour evolution and interacts with the tumour microenvironment — and thereby how it regulates each tumour's behaviour and response to therapy. Understanding such deregulated biology should facilitate the development of therapeutic approaches targeting specific cellular pathways, including combination therapies, which are likely to be needed to achieve more durable patient responses.

However, therapeutic-strategy development will require target validation, tumour-genome analysis (especially of tumours that respond to treatment), and the organization and analysis of clinical metadata on a scale that has not yet been attempted. Although this is a daunting prospect, the promise still remains that the precision with which breast cancers are diagnosed and managed will be improved by the identification of new drug targets and by the ability to assign cancers to molecular subtypes that are associated with effective treatments. This is being made possible only by the remarkable advances in measurement science and computational capability. ∎

Joe Gray *and* **Brian Druker** *are at the Knight Cancer Institute, Oregon Health and Science University, Portland, Oregon 97239, USA.*
e-mails: grayjo@ohsu.edu; drukerb@ohsu.edu

1. www.genome.gov/sequencingcosts
2. Collins, F. S. & Barker, A. D. *Sci. Am.* **296,** 50–57 (2007).
3. Hudson, T. J. *et al. Nature* **464,** 993–998 (2010).
4. Corless, C. L. *Science* **334,** 1217–1218 (2011).
5. Curtis, C. *et al. Nature* **486,** 346–352 (2012).
6. Shah, S. P. *et al. Nature* **486,** 395–399 (2012).
7. Stephens, P. J. *et al. Nature* **486,** 400–404 (2012).
8. Ellis, M. J. *et al. Nature* **486,** 353–360 (2012).
9. Banerji, S. *et al. Nature* **486,** 405–409 (2012).
10. Perou, C. M. *et al. Nature* **406,** 747–752 (2000).
11. Paik, S. *et al. N. Eng. J. Med.* **351,** 2817–2826 (2004).
12. van 't Veer, L. J. *et al. Nature* **415,** 530–536 (2002).
13. Gray, J. W. & Chin, L. *Nature* **452,** 553–563 (2008).
14. Meyerson, M. *et al. Nature Rev. Genet.* **11,** 685–696 (2010).
15. Kristensen, V. N. *et al. Proc. Natl Acad. Sci. USA* **109,** 2802–2807 (2012).
16. Sims, D. *et al. Genome Biol.* **12,** R104 (2011).
17. Slamon, D. *et al. N. Engl. J. Med.* **365,** 1273–1283 (2011).
18. Heiser, L. M. *et al. Proc. Natl Acad. Sci. USA* **109,** 2724–2729 (2012).
19. Barretina, J. *et al. Nature* **483,** 603–607 (2012).
20. Mardis, E. R. *Nature* **470,** 198–203 (2011).

NUCLEAR PHYSICS

Symmetrical tin

The tin isotope ^{100}Sn is the heaviest 'doubly magic nucleus' that has an equal number of protons and neutrons. It is now finally starting to give up its secrets, thanks to the persistent efforts of nuclear physicists. SEE ARTICLE P.341

DANIEL BAZIN

On page 341 of this issue, Hinke *et al.*[1] report how, after years of endeavour, they have achieved a significant leap forward in the study of the heaviest 'symmetrical doubly magic' nucleus — the tin isotope ^{100}Sn. Composed of 50 protons and 50 neutrons, this nucleus is drawing the attention of nuclear physicists around the globe because of its unique location in the nuclear landscape.

Nuclei are complex and challenging quantum objects. In contrast to the structure of atoms, for which the fundamental interaction between the electrons and the nucleus — the electromagnetic force — is known with great precision, the interaction between the constituents of a nucleus, the strong nuclear force, is not so well known. This is due in part to the composite nature of the nuclear constituents, or nucleons, and the nature of the fundamental forces that bind nucleons

together (quantum chromodynamics and the electroweak force).

As far as we know, nuclei are the smallest objects that can be split up into their constituents. They are therefore the smallest entities in which emergent properties — patterns that arise from complexity — can be studied. Nuclear scientists study these emergent phenomena and are using them to decipher the nature of the nuclear force. Magic numbers are numbers of protons or neutrons that form full shells in an atomic nucleus, and are perhaps the foremost emergent property of nuclei. Thought to have first been coined by the physicist Eugene Wigner, the term reflects the unexpected shell structure of nuclei, as opposed to the liquid-like behaviour expected for such densely packed and strongly interacting objects. In fact, the independent particle model used to describe atoms, in which electrons (the particles) are assumed to move independently of each other, also works

remarkably well for nuclei. The model has been able to explain — at least for stable nuclei[2] — the observed sequence of magic numbers: 2, 8, 20, 28, 50, 82 and 126, all of which, by virtue of their 'magic' nature, correspond to increased stability.

In recent years, however, as physicists have expanded their reach across the nuclear landscape, a different picture has emerged. The magic numbers observed in stable nuclei seem to be either vanishing or evolving, especially on the neutron-rich side of the nuclide chart[3,4], which plots proton number against neutron number (Fig. 1).

Nuclei that have a magic number of neutrons or protons are more tightly bound than their non-magic counterparts, and their intrinsic simplicity makes them prime candidates for testing proposed models of nuclear structure. Particularly interesting are nuclei in which the number of both protons and neutrons reaches one of the magic numbers. These doubly magic nuclei have even greater binding energy than singly magic nuclei.

One would expect that symmetrical doubly magic nuclei, which have an equal magic number of protons and neutrons, would follow the magic number sequence, and this is indeed the case for light nuclei: those of helium (^{4}He), oxygen (^{16}O) and calcium (^{40}Ca). However, because of the repulsion between protons, the line of stable nuclei in the nuclide chart veers away from the symmetry line, as ever more neutrons are required to bind heavier nuclei (Fig. 1). As a result, the only two other nuclei that follow the magic sequence are the nickel nucleus ^{56}Ni and ^{100}Sn. These nuclei are bound but unstable: they undergo β-decay, in which a positron (the antiparticle of an electron) is emitted to produce a daughter nucleus.

Although ^{56}Ni is not too far from being a stable nucleus (^{58}Ni is stable), ^{100}Sn is very close to the edge of nuclear stability, where the nuclear force between nucleons can no longer bind them into a nucleus. The nucleus ^{100}Sn has 12 neutrons fewer than the lightest stable isotope of tin, ^{112}Sn. Therein lies the particular attraction of ^{100}Sn: it is at the same time doubly magic and at the edge of the nuclear landscape. Many long-standing questions about this oddball are now beginning to be answered. For example, is it really doubly magic and simple in structure? How is the strength of its β-decay distributed across the energy levels of its daughter nucleus, indium-100? Does it have isomeric (metastable) states? The study of its β-decay is particularly interesting because of the large energy gap between the ground state of ^{100}Sn and that of its daughter, a characteristic of nuclei that are close to the limits of nuclear binding.

Unfortunately, what makes this nucleus attractive is also what makes it difficult to study. It is so far away from stable isotopes that it is extremely difficult to produce. Two types of nuclear reaction have typically been used to attempt this feat. One,
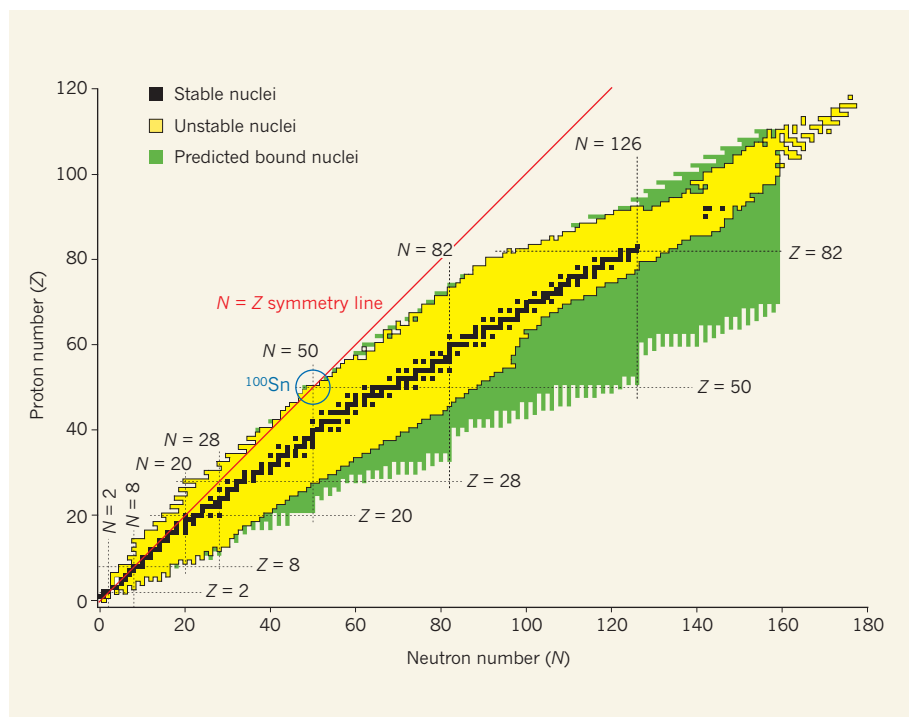
Figure 1 | Nuclear landscape. The chart shows the location of all nuclei as a function of their neutron number (*N*) and proton number (*Z*). Dashed lines represent magic numbers, which correspond to full shells of protons or neutrons. Doubly magic nuclei lie at the intersections of magic-number lines. Symmetrical nuclei have an equal number of protons and neutrons. Hinke *et al.*[1] have produced and studied ^{100}Sn, the heaviest symmetrical doubly magic nucleus. Note that the calculation used to predict bound nuclei in this chart is limited to those with *N* less than 160.

called fusion–evaporation[5], is a bottom-up approach, in which two nuclei are fused together with minimum excitation energy, so as to minimize the subsequent loss of protons or α-particles ([4]He nuclei). The other reaction, called projectile fragmentation[6], is more brutal but at present more effective. In this approach, a high-energy projectile that is heavier than [100]Sn — the xenon nucleus [124]Xe, at 1 gigaelectronvolt per atomic mass unit in Hinke and colleagues' experiment[1] — is sheared by making it collide with a target, leaving a residue that is composed of 50 protons and 50 neutrons. The chance of ending up with the desired nucleus is greater in the former approach than in the latter, but because of the underlying high energy involved, the latter method is more efficient at finding the 'needle in the haystack', and the experiment is actually feasible. To give a sense of the filtering (isolating the desired nuclei from the multitude of other species produced by the nuclear reaction) needed to pull this off, consider that, out of the 1.2×10^{15} [124]Xe projectiles accelerated during Hinke and colleagues' experiment at the GSI Helmholtz Centre for Heavy Ion Research in Darmstadt, Germany, only 259 [100]Sn nuclei were identified.

The results of the authors' experiment represent a giant step compared with previous attempts to study [100]Sn. Not only have the experimenters greatly improved the precision of the half-life measurement of this isotope, but they have, for the first time, also determined the end point of the energy spectrum of β-decay (the maximum energy of the emitted positrons) and have observed γ-ray transitions, which correspond to decays between states of the daughter nucleus. Their deductions are stunning: [100]Sn seems to have the highest-known β-decay strength of all nuclei, and has been classified as a 'superallowed Gamow–Teller decay' (Gamow–Teller transitions allow the spin to change by 0 or ±1 between the initial state of the parent and the final states of the daughter). Usually, this label is reserved for Fermi decays (transitions that occur between states of the same spin), because they typically have the largest strengths.

As always happens with scientists, once they have been given a taste of a new delicacy, they crave more. Other laboratories have joined the race and are working to improve on the GSI [100]Sn production rates. They include: the Radioactive Isotope Beam Factory in Wako, part of Japan's RIKEN national network of labs, which has recently synthesized [100]Sn nuclei; SPIRAL2 at the heavy-ion accelerator GANIL in France; and the Facility for Rare Isotope Beams at Michigan State University. These facilities will produce this remarkable nucleus, as well as many others, in even larger quantities. Deciphering the emergent properties of [100]Sn, and of other nuclei located far from the stability line on the nuclide chart, should lead

scientists towards a fuller understanding of the nuclear force. ∎

Daniel Bazin *is at the National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, Michigan 48824–1321, USA.*
e-mail: bazin@nscl.msu.edu

1. Hinke, C. B. *et al. Nature* **486,** 341–345 (2012).
2. Mayer, M. G. *Phys. Rev.* **74,** 235–239 (1948).
3. Mayer, M. G. & Jensen, J. H. D. *Elementary Theory of Nuclear Shell Structure* (Wiley, 1955).
4. Sorlin, O. & Porquet, M.-G. *Prog. Part. Nucl. Phys.* **61,** 602–673 (2008).
5. Seweryniak, D. *et al. Phys. Rev. Lett.* **99,** 022504 (2007).
6. Morrissey, D. J. & Sherrill, B. M. *Phil. Trans. R. Soc. Lond. A* **356,** 1985–2006 (1998).

PLANETARY SCIENCE

Early start for rocky planets

The chemical composition of stars that host small planets seems to be more varied than that of large planets. This finding may push back the clock for the start of rocky planets and of life around stars other than the Sun. SEE LETTER P.375

DEBRA FISCHER

On page 375 of this issue, Buchhave *et al.*[1] describe a spectroscopic analysis of the chemical composition of stars hosting planet candidates that have been detected by NASA's Kepler mission. The authors have focused on 152 stars harbouring planet candidates that are a similar size to Neptune or smaller. The key result of this work is that small planets — those with a radius up to four times that of Earth — form under a broader range of environmental conditions than do gas-giant planets. This is an important finding, as it shows that the formation of small planets is a less constrained process than is the formation of large planets. It also implies that small, rocky planets formed at an earlier epoch in the Universe's history than gas giants*.

To appreciate this result, flash back to only a few hundred million years after the Big Bang, when the first stars in the Universe had just formed; the Universe today is approximately 14 billion years old. Those stars were composed of only hydrogen and helium, because heavier elements did not yet exist. Metals — defined by astronomers as all elements that are heavier than helium — are manufactured by the fusion of lighter elements in the 'pressure-cooker' cores of stars. When stars explode at the end of their lifetimes, their repositories of metals are cast into space, where they seed molecular clouds to produce new generations of stars that are enriched in heavy elements. Thus powered by the life cycle of stars, the chemical composition of the Universe has evolved from a simple mixture of hydrogen and helium gases to its current state replete with all of the elements in the periodic table.

The chemical evolution of the Universe has

*This article and the paper[1] under discussion were published online on 13 June 2012.

profound implications for the formation of planets. Planets are born in 'protoplanetary' disks around nascent stars; both the star and the protoplanetary disk inherit the chemical composition of the parent molecular cloud. A star that consists of only hydrogen and helium will have a disk with that same composition. This makes it unlikely that the first stars in the Universe formed with planetary companions or life; the elements from which planets and our bodies are made did not exist. Even if exotic gas-giant planets composed of hydrogen and helium existed, they could not have harboured pools of liquid water or a veneer of prebiotic chemistry. Therefore, planets that could have acted as platforms for biology did not populate the early Universe.

Fast-forward 13.6 billion years to the present time and the situation is markedly different. Ground-based surveys, a hallmark of the past two decades, have detected hundreds of planets orbiting nearby stars. As the first of these planets were discovered, an interesting correlation was immediately noted: a large fraction of gas-giant planets was being detected around metal-rich stars[2–4]. Conversely, gas-giant planets were rarely detected around stars that had metallicities (metal abundances) lower than that of the Sun[5]. The general interpretation of this planet–metallicity correlation was that the presence of heavy elements is an environmental condition that increases the density of the protoplanetary disk and therefore the efficiency of planet formation. In the past few years, however,

"This result is exciting because it illuminates the history of the environmental requirements for planet formation."

ground-based surveys have also detected super-Earths — planets that have masses from a few to ten times the mass of Earth. Interestingly, the host stars of these low-mass planets do not seem to exhibit the same high metal content as do hosts of gas-giant planets.

NASA's Kepler mission has revolutionized exoplanet science by virtue of its vast number of planet detections — more than 2,000 planet candidates to date. Buchhave and colleagues' analysis is the first comprehensive assessment of host-star metallicity for the super-Earth- and Neptune-sized planets detected by Kepler, and it elevates the trend inferred by ground-based surveys[6] to a solid statistical result: small planets can form around stars that have a lower metal content than do the host stars of gas-giant planets. The authors obtained the element composition of Kepler stars by matching observed spectra to reference spectra.

This result is exciting because it illuminates the history of the environmental requirements for planet formation. In the beginning, stars formed in isolation, without planets or life. As the chemical evolution of the Universe proceeded, protoplanetary disks developed with a sufficient inventory of heavy elements to build the cores of planets, whether rocky or gaseous. However, for a gas-giant planet to form, there is a race against time: the core needs to reach a critical mass before the gas-rich material dissipates from the disk, which occurs within about 5 million years[7,8]. If the core reaches the critical mass before the gas is lost, runaway gas accretion occurs. High metallicity helps the core to win this race. Small, rocky planets can accrete dust and rocky material long after the volatile gas has dissipated from the disk. They then acquire tenuous atmospheres by releasing light, volatile elements as gas or by accreting trace amounts of gas from the disk. This process can continue for 50 million to 100 million years[9–11].

Unfortunately, it is difficult to use knowledge of metallicity to pinpoint the beginning of the 'age of planet formation', because the chemical enrichment of the Universe is not a monotonic process — some regions inside galaxies may have developed high metallicity rapidly whereas other regions are still metal-poor today. However, knowing that the formation of rocky planets can occur in environments of lower metallicity than those of gas giants implies that there could be some places in the Universe where rocky planets and life got an earlier start than did Earthlings. ∎

Debra Fischer *is in the Department of Astronomy, Yale University, New Haven, Connecticut 06511, USA.*
e-mail: debra.fischer@yale.edu

1. Buchhave, L. A. *et al. Nature* **486,** 375–377 (2012).
2. Gonzalez, G. *Mon. Not. R. Astron. Soc.* **285,** 403–412 (1997).
3. Santos, N. C., Israelian, G. & Mayor, M. *Astron. Astrophys.* **415,** 1153–1166 (2004).
4. Fischer, D. A. & Valenti, J. *Astrophys. J.* **622,** 1102–1117 (2005).
5. Mortier, A. *et al. Astron. Astrophys.* (in the press); preprint at http://arxiv.org/abs/1205.3723 (2012).
6. Mayor, M. *et al.* preprint at http://arxiv.org/abs/1109.2497 (2011).
7. Haisch, K. E. Jr, Lada, E. A. & Lada, C. J. *Astrophys. J.* **553,** L153–L156 (2001).
8. Pascucci, I. & Tachibana, S. *Protoplanetary Dust: Astrophysical and Cosmochemical Perspectives* (eds Apai, D. & Lauretta, D. S.) 263–298 (Cambridge Univ. Press, 2010).
9. Lissauer, J. J. *Icarus* **69,** 249–265 (1987).
10. Raymond, S. N., Quinn, T. & Lunine, J. I. *Icarus* **168,** 1–17 (2004).
11. Raymond, S. N., O'Brien, D. P., Morbidelli, A. & Kaib, N. A. *Icarus* **203,** 644–662 (2009).

VIROLOGY

Bird flu in mammals

An engineered influenza virus based on a haemagglutinin protein from H5N1 avian influenza, with just four mutations, can be transmitted between ferrets, emphasizing the potential for a human pandemic to emerge from birds. SEE LETTER P.420

HUI-LING YEN &
JOSEPH SRIYAL MALIK PEIRIS

Influenza pandemics in humans arise from animal influenza viruses, yet the molecular changes required for an animal virus to be transmitted efficiently between humans are poorly understood. Highly pathogenic H5N1 avian flu viruses have circulated in poultry for more than 16 years, only rarely resulting in human infections. But when people do catch H5N1 bird flu, their disease is of unusual severity, raising concerns that a human H5N1 pandemic might have a catastrophic impact on public health. However, an H5N1 virus that can be efficiently transmitted from human to human has not yet emerged, leading some researchers to question whether these viruses are inherently incapable of acquiring this capacity. On page 420 of this issue, Imai *et al.*[1] demonstrate that H5N1 viruses do have the potential to cause a human pandemic. The authors identify mutations in the avian virus that permit viral transmission between ferrets by means of respiratory droplets — the best available model for influenza transmission in humans*.

Imai *et al.* focused on the haemagglutinin (HA) protein of a highly pathogenic H5N1 influenza virus, which is involved in binding and fusion of the virus to the cells that it infects; the HA type is used in the viral nomenclature, together with the other influenza surface glycoprotein involved, a neuraminidase (NA). The HA of H5N1 viruses preferentially binds sialic acids in receptors on the surface of avian cells (Siaα2,3)**,** whereas cells of the human upper airway predominantly have another type of sialic acid (Siaα2,6), which is recognized by human influenza viruses. The researchers introduced random mutations into the globular head of the HA molecule, where the receptor-binding domain is located, and searched for mutated viruses that exhibited enhanced binding to Siaα2,6. Using reverse genetics, a technique that allows genetic manipulation of the virus genome, they then made a 'hybrid' H5N1 virus in which the gene encoding one of these mutated H5 HA proteins replaced the HA gene in the H1N1 virus that caused a human pandemic in 2009.

The researchers infected ferrets with this hybrid H5N1 virus and, following multiple rounds of infection, isolation of the virus from the upper respiratory tract of the infected animals, and reinfection in animals not previously exposed to the virus, they obtained a virus that can be transmitted efficiently between the animals. Four key amino-acid changes in the HA are associated with ferret respiratory-droplet transmissibility in the authors' virus (Fig. 1a). Three of the mutations (N158D, N224K and Q226L) contribute to Siaα2,6 specificity. The fourth (T318I) lowers the pH at which the protein undergoes a structural change that allows it to release its genetic material into the cytoplasm of an infected cell by fusion of the viral envelope with intracellular membranes.

There have been many previous attempts to determine whether H5N1 can acquire transmission capacity in ferrets. Two studies[2,3] assessed H5N1 and H3N2 hybrid viruses, and another study[4] introduced mutations into the H5N1 HA that are known to increase Siaα2,6 binding in H2 and H3 haemagglutinin — but neither approach conferred transmissibility by respiratory droplets. One study achieved partial transmissibility by introducing three HA mutations (Q196R, Q226L and G228S) into an H5N1 virus and combining this with the NA protein of a human seasonal H3N2 virus (Fig. 1b). It is interesting that the two H5 HAs that demonstrate transmissibility among ferrets[1,5] (Fig. 1a,b) contain similar mutations: one at amino-acid residues 158–160, which removes an *N*-linked glycosylation site of the globular head[6], and the other at residues 221–228, which alters the structure of the loop of the receptor-binding domain[6].

Another research group previously produced a transmissible H9N2 hybrid virus[7] using the HA and NA of a low-pathogenic avian virus

*This article and the paper[1] under discussion were published online on 2 May 2012.

and other genes from a human seasonal H3N2 virus. The HA of this hybrid contains a Q226L mutation, which confers human-like Siaα2,6 binding[7] (Fig. 1c). During the 10 rounds of infection that were needed for this H9N2 virus to acquire transmissibility, it accumulated two additional HA mutations, one located at residue 189 of HA1 (close to the receptor-binding domain) and the other at residue 192 of HA2 (close to the membrane-fusion domain), as well as one NA mutation located in the transmembrane domain. Together, these studies demonstrate that mutations that increase Siaα2,6 binding[1,5,7], as well as those that stabilize HA structure[1,7], seem to be functionalities required for the HA of viruses that are transmissible among mammals.

Thus, Imai and colleagues' generation of a respiratory-droplet-transmissible H5N1 virus represents the culmination of a sustained effort by numerous groups to better understand the mechanisms that can confer mammalian transmissibility in avian influenza viruses. However, it is likely that different combinations of HA mutations can achieve the same effects, and further studies are needed to explore this possibility. Furthermore, although transmission of influenza A virus among humans is already known[8,9] to involve the virus HA, NA and basic polymerase protein 2, it is possible that other viral proteins, not explored by Imai and colleagues, also contribute to mammalian transmissibility.

It is intriguing that, although the parent H5N1 virus strain (A/Vietnam/1203/2004) used by Imai et al. causes lethal disease in ferrets when they are directly infected, infection with the ferret-transmissible H5N1 virus does not kill the animals. It is possible that the change in receptor binding from Siaα2,3, which is present on alveolar epithelial cells in ferret and human lungs, to Siaα2,6, found on cells of the upper respiratory tract, may change the virus from one that causes alveolar infection, likely to result in more severe disease, to one targeting the upper airways, which is associated with milder symptoms. However, Imai and colleagues did not find that the two viruses differ markedly in their targeting of the upper or lower respiratory regions.

Imai and colleagues' H1N1 virus with an H5 HA is a laboratory creation, but it should not be considered an experimental artefact. Natural emergence of an H5N1–H1N1 hybrid virus is plausible. Some H1N1 and H5N1 viruses readily swap genes with one another in vitro[10,11], generating hybrid viruses. Furthermore, pandemic H1N1 viruses are established in pigs in many parts of the world, and H5N1 viruses have been isolated from pigs[12], suggesting that opportunities exist for the viruses to combine in these animals.

However, these findings do not only provide further indication that such a virus may arise naturally; they also pave the way for improved influenza surveillance and pandemic
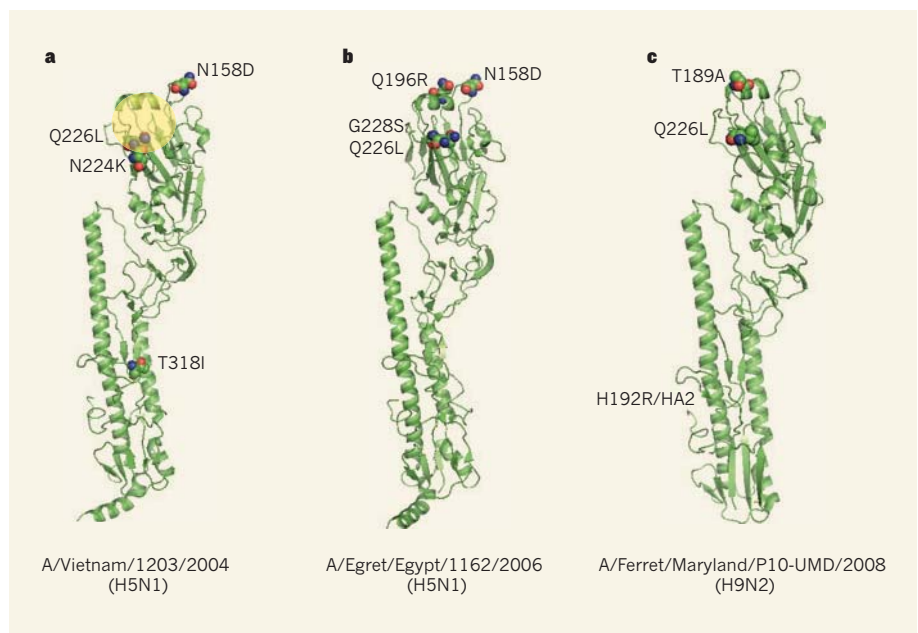
Figure 1 | Avian haemagglutinins transmissible in mammals. The haemagglutinin (HA) protein of influenza determines the type of target cell that the virus can infect. By mutating the site of the HA that binds to sialic acids in receptors on target cells, which differ between avian and mammalian cells, and other protein regions that determine the pH at which virus–cell fusion can occur, researchers have generated viruses that have avian HA proteins that can be transmitted from mammal to mammal. **a,** Imai et al.[1] identify four mutations (N158D, N224K, Q226L and T318I) in the HA of a highly pathogenic avian influenza virus, H5N1, that allow a virus with this HA to be transmitted by respiratory droplets between ferrets. The receptor-binding site of the HA is shaded in yellow. **b,** Chen et al.[5] also created a virus with an H5N1 HA that can be partially transmitted between ferrets, by introducing three mutations (Q196R, Q226L and G228S) into the H5 HA. The virus used as the basis for this hybrid already contained the N158D mutation that Imai and colleagues also identified in their mutated HA. **c,** Sorrell et al.[7] used the HA protein from a low-pathogenic H9N2 avian virus to achieve similar transmissibility. The Q226L mutation was already present in this virus, and two additional mutations (T189A and H192R/HA2) were acquired during ferret infection studies. The H192R/HA2 mutation cannot be shown because the protein is cleaved at a site that lies in front of this amino-acid residue. Full virus names and their subtypes are given below each structure. Mutated amino-acid residues are indicated by red and blue spheres.

preparedness. For example, one of the four mutations reported by Imai et al., N158D, results in a loss of N-linked glycosylation, and loss of glycosylation at this residue is increasingly seen among naturally occurring H5N1 isolates. Although the other three mutations have only very rarely been seen in H5N1 isolates from the field, mutations in HA that change the virus from binding Siaα2,3 to binding both Siaα2,3 and Siaα2,6 have been reported in both birds and humans[13]. These findings reinforce the need for focusing even greater attention on H5N1 infections in humans and other mammals (including pigs). Influenza viruses exist as genetic variants termed quasi-species even within a single clinical specimen, but such genetic diversity may not be fully assessed by conventional sequencing methods. Investigation of mammalian clinical specimens using new deep sequencing methods for these mutations, and for other mutations that confer similar functionality, will allow us to evaluate the extent of H5N1 adaptation that is occurring in mammalian hosts. More broadly, understanding the mutations that confer mammalian transmission of avian influenza viruses will allow better risk assessment of the animal viruses

that represent a pandemic threat, and help to select virus strains against which pre-pandemic vaccines should be generated. ∎

Hui-Ling Yen and **Joseph Sriyal Malik Peiris** are at the School of Public Health and the Centre of Influenza Research, LKS Faculty of Medicine, The University of Hong Kong, and the HKU-Pasteur Research Centre, Pokfulam, Hong Kong.
e-mails: malik@hku.hk; hyen@hku.hk

1. Imai, M. et al. Nature **486,** 420–428 (2012).
2. Jackson, S. et al. J. Virol. **83,** 8131–8140 (2009).
3. Maines, T. R. et al. Proc. Natl Acad. Sci. USA **103,** 12121–12126 (2006).
4. Maines, T. R. et al. Virology **413,** 139–147 (2011).
5. Chen, L.-M. et al. Virology **422,** 105–113 (2012).
6. Stevens, J. et al. J. Mol. Biol. **381,** 1382–1934 (2008).
7. Sorrell, E. M., Wan, H., Araya, Y., Song, H. & Perez, D. R. Proc. Natl Acad. Sci. USA **106,** 7565–7570 (2009).
8. Belser, J. A., Maines, T. R., Tumpey, T. M. & Katz, J. M. Expert Rev. Mol. Med. **12,** e39 (2010).
9. Sorrell, E. et al. Curr. Opin. Virol. **1,** 635–642 (2011).
10. Cline, T. D. et al. J. Virol. **85,** 12262–12270 (2011).
11. Octaviani, C. P., Ozawa, M., Yamada, S., Goto, H. & Kawaoka, Y. J. Virol. **84,** 10918–10922 (2010).
12. Nidom, C. A. et al. Emerg. Infect. Dis. **16,** 1515–1523 (2010).
13. Yamada, S. et al. Nature **444,** 378–382 (2006).

PERSPECTIVE

Engineering H5N1 avian influenza viruses to study human adaptation

David M. Morens[1], Kanta Subbarao[1] & Jeffery K. Taubenberger[1]

Two studies of H5N1 avian influenza viruses that had been genetically engineered to render them transmissible between ferrets have proved highly controversial. Divergent opinions exist about the importance of these studies of influenza transmission and about potential 'dual use' research implications. No consensus has developed yet about how to balance these concerns. After not recommending immediate full publication of earlier, less complete versions of the studies, the United States National Science Advisory Board for Biosecurity subsequently recommended full publication of more complete manuscripts; however, controversy about this and similar research remains.

Knowledgeable observers operating within a legitimate framework for the public good have expressed divergent opinions about the importance and public safety implications of two papers, one recently published[1] and one soon-to-be published[2], describing the production of ferret-transmissible H5N1 influenza viruses, and about related influenza transmission and pathogenesis research[3–7]. Some have emphasized that understanding the underlying principles of influenza virus host adaptation and transmission can lead to better prevention and control of viruses that arise naturally, whereas others have drawn attention to 'dual use' implications—that is, bioterrorism—or to accidental release of potentially deadly viruses. The most commonly mentioned public safety concerns relate to three assumptions: (1) H5N1 viruses are currently highly lethal to humans but are poorly transmissible; (2) genetic manipulation of H5N1 viruses to increase transmissibility in mammals such as ferrets will create variant viruses that remain highly pathogenic and that become transmissible in humans; and (3) if accidentally or intentionally released, such a virus could precipitate a historically severe influenza pandemic. How do these assumptions hold up against scientific data? In this perspective, we address research evidence related to the epidemic/pandemic potential of genetically engineered H5N1 viruses, and discuss limitations in understanding how influenza viruses become pathogenic, transmissible and potentially pandemic in humans.

Background

Influenza is among the leading global infectious causes of death, periodically causing pandemics that can kill millions of people. Countless influenza A viruses circulate globally in a reservoir that consists of hundreds of avian species. Rarely, one of these viruses undergoes changes that enable it to switch hosts to infect mammals, including humans, although it is not clear whether human transmission can result directly from adaptation of an avian influenza virus (this has not been documented to occur), or only indirectly via further adaptation of pre-existing human or mammalian-adapted viruses, the mechanism that has been associated with all known pandemic and seasonal viruses after 1918. The factors underlying all such emergences are poorly understood[8]. In the past 80 years of influenza virology, three pandemics have resulted from reassortments of pre-existing human-adapted or mammalian-adapted viruses with one or more avian-influenza-derived genes, but no purely avian influenza virus has emerged to cause a pandemic or human outbreak, or has even become stably adapted to humans.

However, because avian influenza viruses have adapted to other mammals, it is considered plausible that such an emergence could occur in humans.

Among many other important research areas related to influenza, it is therefore critical to study the mechanisms by which influenza viruses emerge from birds to become adapted to mammals and ultimately humans, and to learn how the phenotypic properties of such evolving viruses may be associated with human transmission and disease. Among the many subtypes and strains of avian influenza A viruses that exist in nature, those that have at least occasionally infected mammals (for example, H5N1, H7N7 and H9N2) are of interest because they might theoretically be more likely than other influenza A viruses to adapt directly or indirectly to humans. Highly pathogenic avian influenza (HPAI) H5N1 viruses have been of particular interest with respect to theoretical pandemic potential because they have been unusually pathogenic in domestic poultry and have infected and killed several hundred people over a 15-year period.

In seeking to understand such influenza viruses, a research approach used widely in virology is to engineer specific genetic mutations into naturally occurring viruses, and then study the resulting viral phenotypic properties in animals, including infectivity, cell tropism, viral replication, pathogenicity and transmissibility. These types of experiments can potentially provide clues about whether and how a virus might adapt to humans, and what prevention and control options might be useful if that virus did emerge. Much H5N1 research of this type has already been published, including viral genetic engineering to evaluate properties such as pathogenicity and transmissibility in ferrets and other animal models. In the context of this published research literature, we comment on questions relevant to the two papers under discussion[1,2].

H5N1 infectivity for humans

The ongoing HPAI H5N1 enzootic continues to cause 'spill-over' human infections. World Health Organization (WHO) data indicate that since 2003, HPAI H5N1 viruses have infected 603 people and killed 356 (ref. 9). Technically, the term 'highly pathogenic' refers only to the effects of certain H5 and H7 influenza viruses in poultry, not in humans or other mammals; most such viruses either cannot infect, or are relatively harmless in, humans. HPAI H5 and H7 phenotypes are both associated with mutations in the haemagglutinin (HA) gene that usually result from insertion of a sequence of codons encoding multiple basic amino acids at the location where the two linked protein domains comprising the

[1]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

mature HA are cleaved during infection[10]. This cleavage site change leads to disseminated viral replication in multiple organs of avian species, resulting in high mortality. However, in humans, who cannot be easily infected with most low pathogenicity or HPAI viruses, if they can be infected at all, efficient replication outside the respiratory tract generally does not occur. Therefore, despite the current unusual situation with H5N1 viruses in humans since 2003 (see below), neither H5N1 nor other HPAI viruses would necessarily replicate systemically or cause extreme pathogenicity should human adaptation occur. Although the basis of HPAI H5N1 viral pathogenicity in severe and fatal human cases remains unknown, there is no evidence suggesting that it results from changes known to be associated with viral adaptation to gallinaceous poultry; in fact, no human-adapted or pandemic influenza virus contains genetic changes indicative of prior poultry adaptation.

Solely on the basis of publicly available information about pathogenicity of intranasally inoculated H5N1 virus (the model for natural human and animal infection), the laboratory-derived H5N1 viruses produced in the two papers under discussion[1,11–15] do not have enhanced pathogenicity in ferrets compared to the 2009 pandemic H1N1 virus, which is considered to be mildly pathogenic for humans[14,15]. An apparent misconception has nevertheless arisen in recent public discussions of these studies, namely that the engineered, ferret-transmissible H5N1 viruses were extremely pathogenic in ferrets after intranasal inoculation or aerosol transmission. This notion seems to have resulted in part from misunderstandings about a technique—intratracheal inoculation—used in a separate sub-study reported in the manuscript by the Fouchier group[14,15], a method that is not directly relevant to viral transmissibility or natural pathogenesis. As documented since the 1940s[16], intratracheal inoculation of influenza viruses is not a 'model' for natural viral pathogenicity; influenza viruses that are otherwise considered to be of low pathogenicity often induce severe and even fatal disease in animals when administered by this route, including the 2009 pandemic H1N1 virus[15]. The presentation of transmissibility studies alongside high-dose intratracheal inoculation pathogenesis studies in the Fouchier manuscript seems to have suggested (incorrectly) to some that the engineered transmissible H5N1 virus is deadly after intranasal inoculation or aerosol transmission between ferrets and, by extension, might be both transmissible and deadly for humans, that is, a virus of deadly pandemic potential. No evidence has yet been presented to support this, although the possibility that additional unspecified genetic changes might do so cannot be excluded.

Potential for human adaptation of H5N1 viruses

It is questionable to what extent HPAI H5N1 is adapted to, or capable of adapting to, humans. It is not clear why one evolving lineage of avian HPAI H5N1 viruses, out of a large and genetically divergent pool of H5 and other avian influenza viruses that rarely infect humans (much less cause severe human disease), has recently infected hundreds or perhaps thousands of people. It may be that the human cases are a result of unusual high-dose exposures or rare individual genetic susceptibilities. Alternatively, H5N1 viruses may be beginning to do something no other HPAI virus has ever been documented to do—adapt directly to humans. And if H5N1 did adapt, could it cause a pandemic?

No HPAI virus in the historic record has ever been efficiently transmitted between humans, let alone caused a pandemic. Even when avian influenza virus genes have been imported by reassortment into existing human influenza viruses, as happened for example in 1957 with H2N2 influenza and in 1968 with H3N2 influenza, the sources seem to have been circulating low-pathogenicity avian viruses, not poultry-adapted viruses such as HPAI viruses[10]. Conceivably, the considerable host-switching mutations associated with adaptation of wild bird viruses to gallinaceous poultry, or at least of wild viruses to HPAI poultry viruses, represent an evolutionary pathway divergent from those pathways associated with mammalian adaptation, seemingly presenting an additional challenge for poultry-adapted influenza viruses to achieve efficient mammalian adaptation[17]. After 15 years of high-density enzootic circulation in domestic poultry around the world, no human-adapted

H5N1 virus has emerged from a natural reservoir, suggesting the existence of unknown biological barriers.

Despite circulation of influenza A viruses of 16 HA subtypes in billions of birds over a very long time span, the four pandemics in the last century have been restricted to influenza viruses bearing HA subtypes H1, H2 or H3. Decades ago, many experts predicted that influenza pandemics could be explained by 'recycling' of a small number of HAs in new human generations; more recently, this belief has been expanded to posit that the other HA subtypes (including H5) are fundamentally incapable of adapting to humans, being selected against by biological constraints or unappreciated selection pressures[18–20]. Despite widespread influenza virus circulation and dynamic evolution at the human–animal interface, with many billions of quasispecies, mutations and gene constellations circulating, only four influenza pandemics have occurred in the last century, and in the three of those with a known viral origin the viruses resulted from reassortment of pre-existing human or swine viruses[21], not by mutation or adaptation of existing avian viruses.

This suggests that *de novo* emergence of a human pandemic influenza virus is an extremely rare event that is not easily achieved in nature[10], and presumably would not be easily achieved by engineering a small number of laboratory mutations. As some of the key engineered H5N1 mutations in the two studies occur spontaneously during normal laboratory passage[22], or have been found singly or in combination in natural H5N1 and in other influenza viruses[23–26], including strains from wild birds, it remains unclear whether or how the engineered viruses in question create or increase the risk of a pandemic.

Engineering H5N1 phenotypic changes

Serial passage of a virus in intact animals or in tissues derived from a particular species often results in enhanced species-specific virulence, which can be applied to establish an animal model with measurable morbidity and/or mortality outcomes useful for evaluating antiviral therapeutics, passive immunization and vaccines. Influenza viruses, SARS coronavirus and Ebola virus have all been passaged in mice to enhance virulence; the resulting host-adapted viruses have been studied biologically and used to evaluate strategies for control and prevention. However, adaptational mutations resulting from serial passage tend to be host-specific and may not produce the same outcomes in other species. For example, the classical swine influenza virus, A/swine/Iowa/1930 (H1N1), is very pathogenic in ferrets and mice but poses no threat to humans[27]. Another example is mouse-adapted Ebola virus, which is lethal for mice and guinea pigs but attenuated for nonhuman primates[28,29]. Ferrets are susceptible to a wide range of viruses including influenza viruses, SARS coronavirus, canine distemper virus and some parvoviruses, many of which do not infect humans or other mammals. A number of influenza viruses that replicate efficiently in ferrets[30–33] seem poorly able, or unable, to infect humans, even after experimental challenge[34]. Thus, pathogenicity and transmissibility of any influenza virus in ferrets cannot be used directly to predict what type and severity of disease the same virus might produce in humans and human populations.

Predicting human transmissibility

It is unclear whether genetic manipulation of an H5N1 virus to achieve transmissibility in a particular mammal such as a ferret can predict human transmissibility. Because natural history and viral challenge studies cannot always be performed in humans, they have been conducted in experimental animals including mice, guinea pigs, ferrets, non-human primates and various other mammals. Unfortunately, there is no perfect animal model capable of reproducing all of the important variables involved in human influenza infection, although each animal model may be useful in understanding some aspect of influenza biology. Unlike most other mammals, ferrets generally can be infected with many or most avian, mammalian and human influenza viruses without prior viral adaptation, and often transmit efficiently between them[35], providing useful general information about the viral genetic basis of phenotypic properties such as infectivity, pathogenicity, transmissibility and

immune responses[35], even though the findings cannot necessarily be directly applied to human infections[36–38]. Furthermore, in decades of research, using a large number of different avian and mammalian influenza viruses, severe or fatal disease has not often been observed in ferrets following intranasal inoculation or aerosol exposure.

These useful traits of easy infectability and mildly symptomatic infection have rendered the ferret a 'permissive' influenza model. Specifically, many naturally occurring influenza viruses that infect, and often transmit between, ferrets are not known to infect people or cause human disease[27,30–33,39,40]. Ferrets are thus an imperfect model for predicting human infectivity or transmissibility, let alone the high level of transmissibility characteristic of pandemic spread. On the basis of public presentations by the senior authors of the two studies in question, neither of the engineered H5N1 viruses was as efficiently transmissible in ferrets as the human-adapted 2009 pandemic H1N1 virus[14,15]. Phenotypic properties such as replication, pathogenicity and transmissibility are likely to be polygenic traits driven by mutations that are independent and possibly competing[10,41]. Transmissibility is a complex phenotype that probably requires cooperative changes in more than one gene segment, and these may differ greatly between different viruses that become transmissible. Mutations that confer transmissibility in a ferret may be species-specific and irrelevant to other hosts[42]. There are probably multiple unique virus-specific pathways to transmissibility for particular viruses infecting particular hosts[43]. For example, transmissibility of the 1918 pandemic H1N1 virus has been linked to changes in the genes encoding HA and PB2 proteins[36,37], whereas transmissibility of the 2009 pandemic H1N1 virus, which has a closely related HA, has been linked to gene segments encoding the neuraminidase and matrix proteins[44].

Moreover, because determinants of viral pathogenicity may be lost upon adaptation to a new host, H5N1 viruses, whether or not transmissible, do not always cause severe disease in ferrets or non-human primates[45–47]. For these reasons viral phenotypes observed in animal models like the ferret may not predict what would happen in humans. Indeed, given that many influenza viruses that are non-pathogenic for humans easily transmit and may cause illness in ferrets, the 'ferret model' does not reliably predict human transmissibility or pathogenicity, although the model remains valuable for understanding the principles and dynamics of infection.

In addition to data from experiments in mammals, it is noteworthy that of the many mammalian-adapted influenza viruses that infect and transmit explosively among pigs, among horses and among dogs, few infect humans and none are transmitted between them[48]. Although swine influenza viruses caused sporadic human infections before 2009[49,50], and caused the 2009 H1N1 pandemic[21], outbreaks associated with human influenza viruses are rare in pigs. It is even conceivable that H5N1 viruses have already become adapted to mammals without causing severe disease or onward transmission to humans. Evidence from China's Qinghai Lake, for example, shows 13.4% H5N1 seropositivity and 3.4% active infection in apparently healthy, live-caught rodents called pikas[51]. Clearly, adaptation of an influenza virus to a specific mammalian host does not necessarily predict its infectiousness, pathogenicity or transmissibility in other mammals, even though valuable insights into mechanisms of disease, host responses, and prevention and treatment may be obtained by studying these particular animals. Such insights can provide valuable clues in designing countermeasures against deadly epidemics and pandemics.

H5N1 case–fatality rate

Belief that an H5N1 virus could produce a 59% pandemic case-fatality rate is the most frightening aspect of the current controversy surrounding aerosol transmission of H5N1 virus in ferrets. In 500 years of observation, no influenza pandemic is believed to have caused a case-fatality rate over about 2% (ref. 52); pandemic and seasonal circulation of H1, H2 and H3 influenza viruses over the past century have all produced lower overall mortality rates[53]. The widely cited 59% figure is not a mortality rate but a case-fatality rate computed from cumulative cases

reported to WHO since 2003[9]. (Case fatality refers to fatal cases divided by all fatal plus non-fatal cases combined.) By general consensus, the WHO figure probably overestimates actual mortality. Among other concerns common to epidemiological data, the WHO case definition[54] features diagnostic severity criteria (evidence of an acute pneumonia on chest radiograph plus evidence of respiratory failure) that constitute a self-fulfilling prophecy for fatality; as with many illnesses studied epidemiologically, severe diseases are more likely to receive optimal diagnostic work-up ('detection bias'); and most H5N1 cases have been reported from countries with limited resources for identifying milder cases, if they occurred. These factors together could combine to erroneously inflate case-fatality calculations by over-counting severe cases and under-counting mild cases[55].

However, potentially more important clues to actual H5N1 pathogenicity and human case-fatality rates come from epidemiological studies, which taken together suggest to us that H5N1 may not be highly lethal except in people with rare susceptibilities. Forty-six published H5N1 seroprevalence studies of various exposure categories (household contacts, healthcare workers, poultry workers, and so on) show generally low H5N1 seroprevalence (mean, 1.7% of 21,435 persons examined in all 46 studies combined (a bibliography of these studies is available from the authors on request)). Given intense poultry and other exposures in many study areas, these low rates at first seem perplexing, especially when compared to the much higher seroprevalence rates in humans for other avian influenza viruses such as H9N2 (ref. 56). When such information is considered in light of statistically significant clustering of non-human-transmitted (that is, presumably avian-acquired) household cases in genetically related versus unrelated persons[57,58], a reasonable explanation seems to us to be that H5N1 is so poorly adapted to humans that exposure does not normally lead to infection or even the development of a detectable immune response[57,59], except in persons with specific but undefined genetic susceptibilities, many of whom become cases[60–62]. There are few data on what the basis of such genetic susceptibilities may be, although recent evidence has linked severe human influenza to a minor *IFITM3* allele[63], supporting the suspicion that genetic determinants of influenza infection and replication in humans do exist.

A published meta-analysis of selected seroprevalence studies implies that the actual H5N1 case-fatality rate may be far below 1% (ref. 56), and thus probably far below the case-fatality rates for seasonal influenza. This has been disputed because it has been difficult to find mild cases, and because of the possibility that some low-level antibody titres (<1:80) might be false positives[5]. On the other hand, rapid disappearance of human H5N1 vaccine-induced antibody[64] suggests that the opposite problem of false negatives could be occurring and, if so, might be especially problematic in cross-sectional studies in which the time since infection is not known, and which could in some cases be long enough for antibody titres to wane to sub-threshold or undetectable levels.

Given such confusing information, there has been little agreement so far on the important question of asymptomatic and undetected H5N1 infections. But whatever the case, unless healthy seropositive people detected in seroprevalence studies temporally and geographically associated with H5N1 cases are all falsely seropositive, their addition to exposure denominators greatly decreases case-fatality determinations. For example, the 1997 Hong Kong H5N1 outbreak case-fatality rate of 33.3% (ref. 65) drops to around 3% with the addition of exposed seropositive persons detected in the related seroprevalence studies. Similar recalculations of other data would yield far lower rates, and wider seroprevalence studies would undoubtedly lower case-fatality rates even further.

Thus, an explanation for the apparent case-fatality rate/seroprevalence paradox may not be purely one of missing cases. Like other poorly adapted viruses that rarely infect humans[34], the H5N1 virus may simply be productively infecting too few of the people exposed to it, even in situations of widespread human contact, leaving minimal immunological evidence of exposure at the population level, while at the same time 'finding' and

infecting those occasional persons with unusual susceptibilities to it; that is, cases[59]. Even so, it should be remembered that limited spread of a deadly H5N1 virus, or pandemic H5N1 spread associated with a far lower case-fatality rate, would still be of public health concern.

The dangers of information release

Owing to global concern over a possible H5N1 influenza pandemic, the pathogenicity, immunogenicity and transmissibility of naturally occurring and laboratory-derived H5N1 viruses have been examined extensively and safely using high-containment facilities and appropriate regulatory and safety oversight (see later). The two H5N1 studies under discussion[1,2] build upon and are the logical extensions of dozens of similar published studies performed in the wake of the 1997 Hong Kong H5N1 outbreak. This research includes another published study in which genetic engineering of the H5N1 virus was able to newly create transmissibility in ferrets[66], a similar study in which increased ferret transmissibility was not documented[67], and a study in which transmissibility was restored and arguably increased in guinea pigs[68]. None of these publications, including the prior publication of engineered H5N1 transmissibility in ferrets, led to concern among scientists, federal agencies or the public.

Such studies feature numerous pathogenicity-associated, and sometimes transmissibility-associated, mutations affecting the HA-receptor-binding site, including changes that enhance receptor affinity for α2-6-linked sialic acid receptors, thought to be important for human adaptation[25,69–71]. Other studies have examined mutations associated with changing antigenicity[72,73], changes associated with fusion[25,74], changes associated with the polybasic HA cleavage site[75–77], and virulence factors in the polymerase proteins, crucial for viral replication[24,68], and in the non-structural protein (NS1), involved in antagonizing host type I interferon responses[78–80]. This widely available body of published research complicates determination of what to do with these two and with similar research manuscripts that seem likely to continue to appear. Withholding or redacting them does not prevent anyone from piecing together the basic information that they contain. Most of this information is generally known and relatively obvious, has already been published, and is now being widely publicized and discussed as a result of increased attention[11,13,14,81,82].

Some would argue that even this background research should not have been done, or should henceforth be classified and made available only to 'approved' scientists who would be vetted by yet-to-be-determined mechanisms[83,84]. But had these former studies not been made available in the open literature, the field of influenza research would have been considerably impeded and our current state of knowledge and readiness for responding to future outbreaks and/or pandemics would be lessened. Some proposed that 'censoring' this information actually increases the risk of bioterrorism[85]. The two studies under discussion[1,2] can help augment surveillance to detect naturally emerging viruses with pandemic potential and expand our knowledge of the principles underlying host adaptation. Although the dangers of 'information release' in the case of these two studies is probably small or nil—because all or most of the critical information is widely available anyway—it nevertheless remains important to rethink larger questions about balancing safety (accidental or deliberate release of an influenza virus or any dangerous pathogen) with the need to study such viruses to learn enough of their biology to prevent and control them. These are important issues that should be discussed broadly among scientists, policy makers and the public.

Biosafety and biosecurity concerns

As novel pathogens emerge, scientists must be able to continue to work with them safely and appropriately in teams using the talents of many highly trained researchers. Numerous layers of robust biosafety and biosecurity protection and oversight are in place to safeguard the scientists and the public alike, including rigorous safety training, biocontainment practices, regulations and oversight, select agent rules, background investigations and biosurety oversight[86]. The H5N1 studies under discussion[1,2]

were both performed in high containment laboratories with rigorous and appropriate oversight and biosecurity measures[15], as is the case for all such research in the US.

Few disagree that it is crucial to continue research with H5N1 and other emerging infections, including investigation of how emerging pathogens adapt to new hosts and cause disease. However, it is important to ask whether some types of infectious diseases research should not be done, or not published openly. If so, we need criteria to identify such research in advance, and processes to balance the importance of the research knowledge with the importance of preventing adverse consequences of the research[87]. Even with the eventual publication of the two H5N1 studies, questions about how such research should be approved, evaluated and made public remain unanswered[83,84]. The biomedical field is built on more than a century of openness and full publication/broad discussion of all findings; it is unclear how redacted publications of future scientific data can be accomplished, and what effect such a system would have on science and scientific progress.

These complex questions have been asked and answered in the past[87], and are being asked again in the context of these two papers. Continued discussion and decisions about how to deal with this research will be of importance to scientific progress and public health. We believe that it is important to consider the broader context of research aimed at understanding how influenza viruses adapt to humans. H5N1 is only one of many avian influenza viruses. If, as we believe existing data suggest, pathways to human adaptation are many, virus-specific, and with few common denominators, it will be important to study not just H5N1 but a wide range of avian and mammalian- and human-adapted viruses, including studies that feature backward genetic engineering to remove phenotypic determinants of adaptation, studies in nonhuman primates and, when safe and appropriate to do so, in human challenge studies[88].

Future directions

The H5N1 controversy underscores how little is known about determinants of human influenza pathogenicity and transmissibility, which are among the most fundamentally important questions in infectious disease research because of the huge burden of influenza.

In the past two decades the question of pursuing and publishing potential 'dual use' infectious disease research has always been decided in favour of conducting and publishing the research; for example, delineating the genomes of smallpox[89] and SARS viruses[90], defining the pathogenicity of neuraminidase-inhibitor-resistant influenza viruses[91,92], genetically altering and making ferret transmissible both a more pathogenic pandemic H2N2 influenza virus[93] and an H9N2 avian influenza virus[40], and resurrecting from RNA fragments, recreating and studying *in vivo* the 1918 pandemic influenza virus[94–96]. In the latter case, important findings already have markedly enhanced our understanding of the emergence, transmissibility and pathogenicity of that important virus, helping us to prepare for and respond to the emergence of other influenza viruses. Examples include using the 1918 HA crystal structure in vaccine design[97], investigating the role of the host immune response in disease[98,99], identification of mutations associated with pathogenicity and host adaptation[36,100,101], understanding influenza evolution[21], and helping guide and target the response to the 2009 H1N1 pandemic[53,102]. All of this work has been done safely with appropriate oversight, and without negative consequences.

In considering the threat of bioterrorism or accidental release of genetically engineered viruses, it is worth remembering that nature is the ultimate bioterrorist. Indeed, H5N1 mutations, including some of those made in the two studies under discussion[1,2], occur spontaneously in nature, probably at a high rate, although they have not yet led to a pandemic. Given the relative rarity of pandemics caused by newly emerging influenza viruses, their explosive transmissibility may result from unique and virus-specific mutational changes that arise at very low frequency. For past pandemics, we have had limited ability to detect such changes by surveillance or by animal model experimentation. Thus, our best hope in

preventing and controlling the microbial agents that continually challenge us is to increase fundamental knowledge about the mechanisms by which they emerge, spread and cause disease, so that we can develop countermeasures such as enhanced surveillance, better diagnostics, vaccines and drug therapies. In moving forward we need to be safety conscious and to have consensus safety measures and policies in place, while at the same time using all available tools to seek broad understanding about the complex relationships between viruses and hosts. It is only this knowledge that stands between us and the devastation of future influenza pandemics. In reconsidering the proper balance between progress and safety, the critical importance of advancing scientific knowledge needs to be kept front and centre.

1. Imai, M. *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* http://dx.doi.org/10.1038/nature10831 (this issue).
2. Herfst, S. *et al.* Aerosol transmission of avian influenza A/H5N1 virus. *Science* (22 June 2012).
3. Doherty, P. C. & Thomas, P. G. Dangerous for ferrets: lethal for humans? *BMC Biol.* **10,** 10 (2012).
4. Fouchier, R. *et al.* Preventing pandemics: the fight over flu. *Nature* **481,** 257–259 (2012).
5. Osterholm, M. T. & Kelley, N. S. Mammalian-transmissible H5N1 influenza: facts and perspective. *MBio* **3,** 2 (2012).
6. Palese, P. Don't censor life-saving science. *Nature* **481,** 115 (2012).
7. Webster, R. G. Mammalian-transmissible H5N1 influenza: the dilemma of dual-use research. *MBio* **3,** 1 (2012).
8. Parrish, C. R. *et al.* Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev.* **72,** 457–470 (2008).
   Authors review data on viral host-switch events.
9. WHO. Cumulative number of confirmed human cases of avian influenza A(H5N1) reported to WHO. http://www.who.int/influenza/human_animal_interface/H5N1_cumulative_table_archives/en/index.html (2012).
10. Taubenberger, J. K. & Kash, J. C. Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe* **7,** 440–451 (2010).
    A review of influenza host adaptation and pandemic formation.
11. Brown, T. H5N1 research discussed in special session at ASM. http://www.medscape.com/viewarticle/759465 (2012).
12. Fouchier, R. A. *et al.* Pause on avian flu transmission research. *Science* **335,** 400–401 (2012).
13. Kawaoka, Y. H5N1: Flu transmission work is urgent. *Nature* **482,** 155 (2012).
14. Osterholm, M. T., Fauci, A. S., Alberts, B. & Fouchier, R. A. M. Discussion of NSABB's publication recommendations for the NIH-funded research on the transmissibility of H5N1. http://www.asmbiodefense.org/index.php/program-information/nsabbs-recommendations-for-h5n1-research (2012).
15. Kawaoka, Y. & Fouchier, R. A. Royal Society meeting. H5N1 research: biosafety, biosecurity and bioethics. http://royalsociety.org/events/2012/viruses/ (2012).
16. Wilson, H. E., Saslaw, S., Doan, C. A., Woolpert, O. C. & Schwab, J. L. Reactions of monkeys to experimental mixed influenza and *Streptococcus* infections: an analysis of the relative roles of humoral and cellular immunity, with the description of an intercurrent nephritic syndrome. *J. Exp. Med.* **85,** 199–215 (1947).
17. Swayne, D. E. Understanding the complex pathobiology of high pathogenicity avian influenza viruses in birds. *Avian Dis.* **51,** 242–249 (2007).
    A review of highly pathogenic avian influenza viruses.
18. Francis, T. Jr. Influenza: the new acquayantance. *Ann. Intern. Med.* **39,** 203–221 (1953).
19. Masurel, N. & Marine, W. M. Recycling of Asian and Hong Kong influenza A virus hemagglutinins in man. *Am. J. Epidemiol.* **97,** 44–49 (1973).
20. Hilleman, M. R. Realities and enigmas of human viral influenza: pathogenesis, epidemiology and control. *Vaccine* **20,** 3068–3087 (2002).
21. Morens, D. M., Taubenberger, J. K. & Fauci, A. S. The persistent legacy of the 1918 influenza virus. *N. Engl. J. Med.* **361,** 225–229 (2009).
22. Bogs, J. *et al.* Reversion of PB2-627E to -627K during replication of an H5N1 Clade 2.2 virus in mammalian hosts depends on the origin of the nucleoprotein. *J. Virol.* **85,** 10691–10698 (2011).
23. Chen, H. *et al.* Properties and dissemination of H5N1 viruses isolated during an influenza outbreak in migratory waterfowl in western China. *J. Virol.* **80,** 5976–5983 (2006).
24. Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science* **293,** 1840–1842 (2001).
    A study showing the key role that changes in a viral polymerase gene has in H5N1 virulence in a mammalian model.
25. Stevens, J. *et al.* Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* **312,** 404–410 (2006).
    A study describing the haemagglutinin crystal structure of an H5N1 influenza virus and studies of mutations associated with receptor-binding specificity.
26. Yen, H. L. *et al.* Inefficient transmission of H5N1 influenza viruses in a ferret contact model. *J. Virol.* **81,** 6890–6898 (2007).
27. Memoli, M. J. *et al.* An early 'classical' swine H1N1 influenza virus shows similar pathogenicity to the 1918 pandemic virus in ferrets and mice. *Virology* **393,** 338–345 (2009).
28. Bray, M., Davis, K., Geisbert, T., Schmaljohn, C. & Huggins, J. A mouse model for evaluation of prophylaxis and therapy of Ebola hemorrhagic fever. *J. Infect. Dis.* **178,** 651–661 (1998).
29. Bray, M., Hatfill, S., Hensley, L. & Huggins, J. W. Haematological, biochemical and coagulation changes in mice, guinea-pigs and monkeys infected with a mouse-adapted variant of Ebola Zaire virus. *J. Comp. Pathol.* **125,** 243–253 (2001).
30. Chen, G. L. *et al.* Evaluation of replication and cross-reactive antibody responses of H2 subtype influenza viruses in mice and ferrets. *J. Virol.* **84,** 7695–7702 (2010).
31. Hinshaw, V. S., Webster, R. G., Easterday, B. C. & Bean, W. J. Jr. Replication of avian influenza A viruses in mammals. *Infect. Immun.* **34,** 354–361 (1981).
32. Joseph, T. *et al.* Evaluation of replication and pathogenicity of avian influenza a H7 subtype viruses in a mouse model. *J. Virol.* **81,** 10558–10566 (2007).
33. Song, H., Wan, H., Araya, Y. & Perez, D. R. Partial direct contact transmission in ferrets of a mallard H7N3 influenza virus with typical avian-like receptor specificity. *Virol. J.* **6,** 126 (2009).
34. Beare, A. S. & Webster, R. G. Replication of avian influenza viruses in humans. *Arch. Virol.* **119,** 37–42 (1991).
35. Belser, J. A., Katz, J. M. & Tumpey, T. M. The ferret as a model organism to study influenza A virus infection. *Dis. Model Mech.* **4,** 575–579 (2011).
    A review of the ferret as a model organism in influenza virus pathogenicity studies.
36. Tumpey, T. M. *et al.* A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science* **315,** 655–659 (2007).
    A study elucidating the importance of the haemagglutinin-receptor-binding region in conferring ferret transmissibility in the 1918 pandemic influenza virus.
37. Van Hoeven, N. *et al.* Human HA and polymerase subunit PB2 proteins confer transmission of an avian influenza virus through the air. *Proc. Natl Acad. Sci. USA* **106,** 3366–3371 (2009).
    A study describing the importance of both the 1918 haemagglutinin and polymerase PB2 genes in conferring ferret transmissibility in viruses expressing 1918 pandemic influenza virus genes.
38. Watanabe, T. *et al.* Viral RNA polymerase complex promotes optimal growth of 1918 virus in the lower respiratory tract of ferrets. *Proc. Natl Acad. Sci. USA* **106,** 588–592 (2009).
39. Kimble, J. B., Sorrell, E., Shao, H., Martin, P. L. & Perez, D. R. Compatibility of H9N2 avian influenza surface genes and 2009 pandemic H1N1 internal genes for transmission in the ferret model. *Proc. Natl Acad. Sci. USA* **108,** 12084–12088 (2011).
40. Sorrell, E. M., Wan, H., Araya, Y., Song, H. & Perez, D. R. Minimal molecular constraints for respiratory droplet transmission of an avian-human H9N2 influenza A virus. *Proc. Natl Acad. Sci. USA* **106,** 7565–7570 (2009).
41. Koster, F. *et al.* Exhaled aerosol transmission of pandemic and seasonal H1N1 influenza viruses in the ferret. *PLoS ONE* **7,** e33118 (2012).
42. Yamanaka, T. *et al.* No evidence of horizontal infection in horses kept in close contact with dogs experimentally infected with canine influenza A virus (H3N8). *Acta Vet. Scand.* **54,** 25 (2012).
43. Mänz, B., Brunotte, L., Reuther, P. & Schwemmle, M. Adaptive mutations in NEP compensate for defective H5N1 RNA replication in cultured human cells. *Nature Commun.* **3,** 802 (2012).
44. Lakdawala, S. S. *et al.* Eurasian-origin gene segments contribute to the transmissibility, aerosol release, and morphology of the 2009 pandemic H1N1 influenza virus. *PLoS Pathog.* **7,** e1002443 (2011).
45. Chen, Y. *et al.* Pathological lesions and viral localization of influenza A (H5N1) virus in experimentally infected Chinese rhesus macaques: implications for pathogenesis and viral transmission. *Arch. Virol.* **154,** 227–233 (2009).
46. Govorkova, E. A. *et al.* Lethality to ferrets of H5N1 influenza viruses isolated from humans and poultry in 2004. *J. Virol.* **79,** 2191–2198 (2005).
47. Maines, T. R. *et al.* Avian influenza (H5N1) viruses isolated from humans in Asia in 2004 exhibit increased virulence in mammals. *J. Virol.* **79,** 11788–11800 (2005).
48. Morens, D. M. & Taubenberger, J. K. Historical thoughts on influenza viral ecosystems, or behold a pale horse, dead dogs, failing fowl, and sick swine. *Influenza Other Respir. Viruses* **4,** 327–337 (2010).
49. Myers, K. P., Olsen, C. W. & Gray, G. C. Cases of swine influenza in humans: a review of the literature. *Clin. Infect. Dis.* **44,** 1084–1088 (2007).
50. Shinde, V. *et al.* Triple-reassortant swine influenza A (H1) in humans in the United States, 2005–2009. *N. Engl. J. Med.* **360,** 2616–2625 (2009).
51. Zhou, J. *et al.* Characterization of the H5N1 highly pathogenic avian influenza virus derived from wild pikas in China. *J. Virol.* **83,** 8957–8964 (2009).
52. Morens, D. M. & Taubenberger, J. K. Pandemic influenza: certain uncertainties. *Rev. Med. Virol.* **21,** 262–284 (2011).
    A historical review of influenza pandemics.
53. Morens, D. M., Taubenberger, J. K. & Fauci, A. S. The 2009 H1N1 pandemic influenza virus: what next? *MBio* **1,** 4 (2010).
54. WHO. WHO case definitions for human infections with influenza A(H5N1) virus. http://www.who.int/influenza/resources/documents/case_definition2006_08_29/en/index.html (2006).
55. Palese, P. & Wang, T. T. H5N1 influenza viruses: facts, not fear. *Proc. Natl Acad. Sci. USA* **109,** 2211–2213 (2012).
56. Wang, T. T., Parides, M. K. & Palese, P. Seroevidence for H5N1 influenza infections in humans: meta-analysis. *Science* **335,** 1463 (2012).

57. Aditama, T. Y. et al. Risk factors for cluster outbreaks of avian influenza A H5N1 infection, Indonesia. Clin. Infect. Dis. 53, 1237–1244 (2011).
A study describing human H5N1 case clusters in Indonesia.
58. Aditama, T. Y. et al. Avian influenza H5N1 transmission in households, Indonesia. PLoS ONE 7, e29971 (2012).
59. Roos, R. No sign of missed H5N1 cases in Bangladesh study. http://www.cidrap.umn.edu/cidrap/content/influenza/avianflu/news/mar1312seroprev.html (2012).
60. Horby, P. et al. What is the evidence of a role for host genetics in susceptibility to influenza A/H5N1? Epidemiol. Infect. 138, 1550–1558 (2010).
61. Trammell, R. A. & Toth, L. A. Genetic susceptibility and resistance to influenza infection and disease in humans and mice. Expert Rev. Mol. Diagn. 8, 515–529 (2008).
62. Zhang, L., Katz, J. M., Gwinn, M., Dowling, N. F. & Khoury, M. J. Systems-based candidate genes for human response to influenza infection. Infect. Genet. Evol. 9, 1148–1157 (2009).
63. Everitt, A. R. et al. IFITM3 restricts the morbidity and mortality associated with influenza. Nature 484, 519–523 (2012).
64. Brady, R. C. et al. Safety and immunogenicity of a subvirion inactivated influenza A/H5N1 vaccine with or without aluminum hydroxide among healthy elderly adults. Vaccine 27, 5091–5095 (2009).
65. Katz, J. M. et al. Antibody response in individuals infected with avian influenza A (H5N1) viruses and detection of anti-H5 antibody among household and social contacts. J. Infect. Dis. 180, 1763–1770 (1999).
66. Chen, L. M. et al. In vitro evolution of H5N1 avian influenza virus toward human-type receptor specificity. Virology 422, 105–113 (2012).
67. Maines, T. R. et al. Effect of receptor binding domain mutations on receptor binding and transmissibility of avian influenza H5N1 viruses. Virology 413, 139–147 (2011).
68. Steel, J., Lowen, A. C., Mubareka, S. & Palese, P. Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N. PLoS Pathog. 5, e1000252 (2009).
69. Ilyushina, N. A., Govorkova, E. A., Gray, T. E., Bovin, N. V. & Webster, R. G. Human-like receptor specificity does not affect the neuraminidase-inhibitor susceptibility of H5N1 influenza viruses. PLoS Pathog. 4, e1000043 (2008).
70. Yamada, S. et al. Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. Nature 444, 378–382 (2006).
71. Yen, H. L. et al. Changes in H5N1 influenza virus hemagglutinin receptor binding domain affect systemic spread. Proc. Natl Acad. Sci. USA 106, 286–291 (2009).
72. Rudneva, I. A. et al. Antigenic epitopes in the hemagglutinin of Qinghai-type influenza H5N1 virus. Viral Immunol. 23, 181–187 (2010).
73. Wang, W. et al. Glycosylation at 158N of the hemagglutinin protein and receptor binding specificity synergistically affect the antigenicity and immunogenicity of a live attenuated H5N1 A/Vietnam/1203/2004 vaccine virus in ferrets. J. Virol. 84, 6570–6577 (2010).
74. Sui, J. et al. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. Nature Struct. Mol. Biol. 16, 265–273 (2009).
75. Bogs, J. et al. Highly pathogenic H5N1 influenza viruses carry virulence determinants beyond the polybasic hemagglutinin cleavage site. PLoS ONE 5, e11826 (2010).
76. Gohrbandt, S. et al. Amino acids adjacent to the haemagglutinin cleavage site are relevant for virulence of avian influenza viruses of subtype H5. J. Gen. Virol. 92, 51–59 (2011).
77. Suguitan, A. L. Jr et al. The multibasic cleavage site of the hemagglutinin of highly pathogenic A/Vietnam/1203/2004 (H5N1) avian influenza virus acts as a virulence factor in a host-specific manner in mammals. J. Virol. 86, 2706–2714 (2012).
78. Seo, S. H., Hoffmann, E. & Webster, R. G. Lethal H5N1 influenza viruses escape host anti-viral cytokine responses. Nature Med. 8, 950–954 (2002).
79. Spesock, A. et al. The virulence of 1997 H5N1 influenza viruses in the mouse model is increased by correcting a defect in their NS1 proteins. J. Virol. 85, 7048–7058 (2011).
80. Zielecki, F. et al. Virulence determinants of avian H5N1 influenza A virus in mammalian and avian hosts: role of the C-terminal ESEV motif in the viral NS1 protein. J. Virol. 84, 10708–10718 (2010).
81. MacKenzie, D. Five easy mutations to make bird flu a lethal pandemic. New Sci. http://www.newscientist.com/article/mg21128314.600-five-easy-mutations-to-make-bird-flu-a-lethal-pandemic.html (2011).
82. WHO. Public health, influenza experts agree H5N1 research critical, but extend delay. http://www.who.int/mediacentre/news/releases/2012/h5n1_research_20120217/en/index.html (2012).
83. NSABB. National Science Advisory Board for Biosecurity Findings and Recommendations March 29–30, 2012. http://oba.od.nih.gov/biosecurity/biosecurity_documents.html (NSABB, 2012).
84. WHO. Report on technical consultation on H5N1 research issues. http://www.who.int/influenza/human_animal_interface/consensus_points/en/index.html (2012).
85. MacKenzie, D. Censoring flu data could raise bioterror threat. New Sci. http://www.newscientist.com/article/dn21675-censoring-flu-data-could-raise-bioterror-threat.html (2012).
86. CDC in Biosafety in Microbiological and Biomedical Laboratories (BMBL) (eds Chosewood, L. C. & Wilson, D. E.) Ch. 8 (Government Printing Office, 2009).
87. National Research Council in Science and Security in a Post 9/11 World: A Report Based on Regional Discussions Between the Science and Security Communities Ch. 4, 57–68 (NAS, 2007).
88. Killingley, B. et al. Potential role of human challenge studies for investigation of influenza transmission. Lancet Infect. Dis. 11, 879–886 (2011).
89. Massung, R. F. et al. Potential virulence determinants in terminal regions of variola smallpox virus genome. Nature 366, 748–751 (1993).
90. Marra, M. A. et al. The genome sequence of the SARS-associated coronavirus. Science 300, 1399–1404 (2003).
91. Herlocher, M. L. et al. Influenza viruses resistant to the antiviral drug oseltamivir: transmission studies in ferrets. J. Infect. Dis. 190, 1627–1630 (2004).
92. Memoli, M. J. et al. Multidrug-resistant 2009 pandemic influenza A(H1N1) viruses maintain fitness and transmissibility in ferrets. J. Infect. Dis. 203, 348–357 (2011).
93. Pappas, C. et al. Receptor specificity and transmission of H2N2 subtype viruses isolated from the pandemic of 1957. PLoS ONE 5, e11158 (2010).
94. Taubenberger, J. K. & Kash, J. C. Insights on influenza pathogenesis from the grave. Virus Res. 162, 2–7 (2011).
95. Taubenberger, J. K. et al. Characterization of the 1918 influenza virus polymerase genes. Nature 437, 889–893 (2005).
96. Tumpey, T. M. et al. Characterization of the reconstructed 1918 Spanish influenza pandemic virus. Science 310, 77–80 (2005).
97. Fleishman, S. J. et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science 332, 816–821 (2011).
98. Kash, J. C. et al. Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. Nature 443, 578–581 (2006).
99. Kobasa, D. et al. Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. Nature 445, 319–323 (2007).
100. Pappas, C. et al. Single gene reassortants identify a critical role for PB1, HA, and NA in the high virulence of the 1918 pandemic influenza virus. Proc. Natl Acad. Sci. USA 105, 3064–3069 (2008).
101. Qi, L. et al. The ability of pandemic influenza virus hemagglutinins to induce lower respiratory pathology is associated with decreased surfactant protein D binding. Virology 412, 426–434 (2011).
102. Morens, D. M., Taubenberger, J. K., Harvey, H. A. & Memoli, M. J. The 1918 influenza pandemic: lessons for 2009 and the future. Crit. Care Med. 38, e10–e20 (2010).

Author Contributions D.M.M., K.S. and J.K.T. jointly co-authored this manuscript.

ARTICLE

Superallowed Gamow–Teller decay of the doubly magic nucleus ^{100}Sn

C. B. Hinke[1], M. Böhmer[1], P. Boutachkov[2], T. Faestermann[1], H. Geissel[2], J. Gerl[2], R. Gernhäuser[1], M. Górska[2], A. Gottardo[3], H. Grawe[2], J. L. Grębosz[4], R. Krücken[1,5], N. Kurz[2], Z. Liu[6], L. Maier[1], F. Nowacki[7], S. Pietri[2], Zs. Podolyák[8], K. Sieja[7], K. Steiger[1], K. Straub[1], H. Weick[2], H.-J. Wollersheim[2], P. J. Woods[6], N. Al-Dahan[8], N. Alkhomashi[8], A. Ataç[9], A. Blazhev[10], N. F. Braun[10], I. T. Čeliković[11], T. Davinson[6], I. Dillmann[2], C. Domingo-Pardo[12], P. C. Doornenbal[13], G. de France[14], G. F. Farrelly[8], F. Farinon[2], N. Goel[2], T. C. Habermann[2], R. Hoischen[2], R. Janik[15], M. Karny[16], A. Kaşkaş[9], I. M. Kojouharov[2], Th. Kröll[17], Y. Litvinov[2], S. Myalski[4], F. Nebel[1], S. Nishimura[13], C. Nociforo[2], J. Nyberg[18], A. R. Parikh[19], A. Procházka[2], P. H. Regan[8], C. Rigollet[20], H. Schaffner[2], C. Scheidenberger[2], S. Schwertel[1], P.-A. Söderström[13], S. J. Steer[8], A. Stolz[21] & P. Strmeň[15]

The shell structure of atomic nuclei is associated with 'magic numbers' and originates in the nearly independent motion of neutrons and protons in a mean potential generated by all nucleons. During β⁺-decay, a proton transforms into a neutron in a previously not fully occupied orbital, emitting a positron–neutrino pair with either parallel or antiparallel spins, in a Gamow–Teller or Fermi transition, respectively. The transition probability, or strength, of a Gamow–Teller transition depends sensitively on the underlying shell structure and is usually distributed among many states in the neighbouring nucleus. Here we report measurements of the half-life and decay energy for the decay of ^{100}Sn, the heaviest doubly magic nucleus with equal numbers of protons and neutrons. In the β-decay of ^{100}Sn, a large fraction of the strength is observable because of the large decay energy. We determine the largest Gamow–Teller strength so far measured in allowed nuclear β-decay, establishing the 'superallowed' nature of this Gamow–Teller transition. The large strength and the low-energy states in the daughter nucleus, ^{100}In, are well reproduced by modern, large-scale shell model calculations.

Gamow–Teller transitions, in which a proton is transformed into a neutron or vice versa, while possibly flipping its spin, represent an important spin–isospin degree of freedom in atomic nuclei. They are important in many astrophysical processes: they govern, for example, electron capture during the core collapse of supernovae. Furthermore, a detailed understanding of Gamow–Teller transitions will provide an essential constraint on the neutrino mass, in the event that neutrinoless double β-decay is ever observed. Most of the Gamow–Teller strength is found in the collective Gamow–Teller giant resonance (GTGR) of the neighbouring nucleus, which is typically a broad structure composed of many states. Whereas in charge-exchange reactions in stable nuclei the full GTGR is accessible, the Gamow–Teller strength in unstable nuclei can, so far, only be studied through β-decay. However, β-decay studies can observe only the fraction of the total Gamow–Teller strength within the decay energy window. Towards more proton-rich nuclei, this window becomes larger. Nevertheless, it is still experimentally challenging to detect all small components of the Gamow–Teller strength[1,2]. Thus, in most nuclei, measuring the full Gamow–Teller strength is difficult because it is fragmented and only partly accessible in β-decays.

^{100}Sn has $N = 50$ neutrons and $Z = 50$ protons, and as a result has completely occupied shells. It is therefore called 'doubly magic' and is particularly suited both experimentally and theoretically to the study of Gamow–Teller transitions. The closed $N = Z = 50$ shells reduce the effect of long-range correlations, thus decreasing the amount of fragmentation of the GTGR. Theoretical predictions suggest that a single state is dominantly populated in this decay. At the same time, the energy window for β-decay is ~7.4 MeV (ref. 3), and most of the GTGR is therefore accessible. Such a situation for a doubly magic system is realized nowhere else in the Segrè chart (a two-dimensional lattice in which all known nuclei are arranged with N and Z on the x and y axes, respectively): ^{16}O and ^{40}Ca are stable nuclei; ^{56}Ni has too small a Q_{EC} value (the energy available for β⁺-decay or electron-capture decay) to make the Gamow–Teller resonance observable in β-decay; and doubly magic nuclei with $N = Z$ that are heavier than ^{100}Sn are unbound. Also, a recent experiment shows that ^{56}Ni has a much more fragmented Gamow–Teller strength[4] as a result of a less robust $N = Z = 28$ doubly magic shell closure as well as subtle differences in the shell structure (Methods Summary and Supplementary Information).

In an extreme, pure single-particle picture, the only possible Gamow–Teller transition of ^{100}Sn is the decay of a proton in the completely filled $g_{9/2}$ shell to a neutron in the empty $g_{7/2}$ shell because the $g_{9/2}$ neutron orbital is filled and no levels above $Z = 50$ are occupied. The large energy separation (shell gap) between these spin-parallel ($g_{9/2}$) and spin-antiparallel ($g_{7/2}$) orbitals, for which the orbital angular momentum is $L = 4$, is responsible for 50 being a magic number. The β-decay of ^{100}Sn is supposed to be enhanced as a result of the large

[1]Physik Department E12, Technische Universität München, D-85748 Garching, Germany. [2]GSI Helmholtzzentrum für Schwerionenforschung GmbH, D-64291 Darmstadt, Germany. [3]Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Legnaro, 35020 Legnaro, Italy. [4]The Henryk Niewodniczanski Institute of Nuclear Physics (IFJ PAN), 31-342 Krakow, Poland. [5]TRIUMF, Vancouver, British Columbia V6T 2A3, Canada. [6]School of Physics & Astronomy, The University of Edinburgh, Edinburgh EH9 3JZ, UK. [7]Université de Strasbourg, IPHC, 67037 Strasbourg Cedex, France. [8]Department of Physics, University of Surrey, Guildford GU2 7XH, UK. [9]Physics Department, Faculty of Science, Ankara University, 06100 Tandogan, Turkey. [10]Institute of Nuclear Physics, University of Cologne, D-50937 Köln, Germany. [11]Institute Vinca, University of Belgrade, 11000 Belgrade, Serbia. [12]IFIC, CSIC-University of Valencia, E-46071 Valencia, Spain. [13]RIKEN Nishina Center, Wako, Saitama 351-0198, Japan. [14]Grand Accélérateur National d'Ions Lourds, CEA/DSM-CNRS/IN2P3, 14076 Caen, France. [15]Comenius University, 818 06 Bratislava 16, Slovakia. [16]Institute of Experimental Physics, University of Warsaw, PL-00681 Warsaw, Poland. [17]Institut für Kernphysik, Technische Universität Darmstadt, D-64289 Darmstadt, Germany. [18]Department of Physics & Astronomy, Uppsala University, SE-75120 Uppsala, Sweden. [19]Departamento de Fisica i Enginyeria Nuclear, Universitat Politecnica de Catalunya (EUETIB), E-08036 Barcelona, Spain. [20]KVI, University of Groningen, 9747AA Groningen, The Netherlands. [21]National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, Michigan 48824-1321, USA.

number of protons occupying the $g_{9/2}$ shell, which can decay to the mostly empty neutron $g_{7/2}$ shell. This would lead to a GTGR consisting of only a single $I^\pi = 1^+$ level (I, spin; π, parity) with a large Gamow–Teller strength of about $B_{GT} = 10$, taking into account the standard renormalization factor (0.75) of the Gamow–Teller matrix element due to configurations outside the model space[5]. This unique situation has been termed 'superallowed' Gamow–Teller decay[6]. Even in more realistic models, including particle–hole correlations, the Gamow–Teller decay of the ground state of ^{100}Sn is predicted to populate with more than 95% probability a single 1^+ state in ^{100}In at an excitation energy of about 3 MeV. In these calculations, a Gamow–Teller strength of around 8–14 is obtained[7–10], leading to renormalized predictions of 5–7 (refs 8, 10). These theoretical results are summarized in Methods Summary and Supplementary Information.

The production of ^{100}Sn, and the study of its decay properties, has been the aim of several experiments[11–16], but in these only a few ^{100}Sn nuclei were uniquely identified. Here we report a new measurement of the half-life and Q_{EC} value from 259 identified ^{100}Sn nuclei, which yields the smallest $\log(ft)$ value of any known β-decay (here t is the half-life and f is a phase space factor that takes into account the trivial decay energy dependence of the half-life). The Gamow–Teller strength is inversely proportional to ft and thus is greatest for the ^{100}Sn decay establishing the robustness of $N = Z = 50$ shell closures. The experimentally observed Gamow–Teller strength is well described in modern, large-scale shell model (LSSM) calculations, which are able to handle an unprecedentedly large degree of configuration mixing in the case of ^{100}Sn. This ^{100}Sn doubly magic shell closure is the benchmark for various topics currently discussed in this mass region, such as spin-aligned pairing in $N = Z$ nuclei, alpha clustering and quadrupole collectivity in the Sn isotopic chain.

Experimental details

The experiment was performed at the GSI Helmholtzzentrum für Schwerionenforschung, Germany. A ^{124}Xe beam with a kinetic energy of $1.0A$ GeV (A, nucleon number) and 1-s-long spills of 3×10^9 ions every 3 s was directed onto a beryllium target placed in front of the fragment separator[17]. Neutron-deficient nuclei were produced through relativistic projectile fragmentation, transmitted to the final focal plane of the fragment separator, and identified event by event (Fig. 1). The correct identification was verified by observing the γ-radiation depopulating known isomers, for example the 8^+ isomer in ^{98}Cd. In total, 259 ^{100}Sn nuclei were unambiguously identified.

This corresponds to a production rate of 0.75 per hour and a cross-section of 5.8 ± 2.1 pb. All uncertainties correspond to one standard deviation.

The ions were implanted into a stack of highly segmented silicon strip detectors surrounded by the RISING array, which consists of 105 germanium detectors arranged in the stopped-beam configuration[18] to detect γ-rays with high efficiency. Of the 259 identified ^{100}Sn nuclei, 163 were stopped in the 2.1-mm-thick implantation layer.

Analysis and results

Following a ^{100}Sn implantation in a pixel of the implantation zone of the silicon detector, all decay events were recorded that occurred within 15 s in that pixel or in the directly neighbouring ones. During this correlation time, it was possible to assign 126 decay chains to the 163 ^{100}Sn implantations. A maximum-likelihood (MLH) analysis with a maximum of three decay events during the correlation time was used to analyse these decay chains. The half-life of ^{100}Sn was deduced to be 1.16 ± 0.20 s in the MLH analysis using established values for the lifetimes of the daughter nuclei. The measurement is much more precise than previous experiments yielding $0.94^{+0.54}_{-0.27}$ s (ref. 14) and $0.55^{+0.70}_{-0.31}$ s (ref. 16). In Fig. 2, we show the decay curve for ^{100}Sn.

Figure 3 shows the γ-ray spectrum observed in coincidence with decay events following ^{100}Sn implantations. Notably, discrete γ-transitions from the ^{100}Sn decay could be observed. The five transitions denoted in Fig. 3 are associated with the depopulation of excited states in the daughter nucleus ^{100}In.

The statistics were sufficient only to establish a coincidence between the 436-keV and 96-keV transitions, and it is thus impossible to deduce an unambiguous level scheme for ^{100}In. Within the uncertainties, the transitions could have the same intensity, which would allow for a single cascade of five transitions from the excited 1^+ state to the ground state. However, this would lead to an excitation energy of more than 4 MeV, which is higher than the value of about 2.5 MeV predicted with realistic shell model calculations (see, for example, refs 19, 20). The large uncertainties in the observed intensities also allow for two parallel cascades originating from this 1^+ state.

Figure 4 shows the relevant level scheme for ^{100}In obtained from LSSM calculations. In this approach, ^{100}Sn is not treated as an inert, doubly magic core but instead excitations across the $N = Z = 50$ shell closures were allowed within the fifth ($4\hbar\omega$) harmonic oscillator shell (Methods Summary and Supplementary Information).

The states of two multiplets that are relevant for the decay are shown. The states originate in the coupling of proton (π) holes in the $g_{9/2}$ orbital either to neutron (v) particles in the $g_{7/2}$ orbital ($\pi g_{9/2}^{-1} \otimes v g_{7/2}^{1}$) with total spin and parity $I^\pi = 1^+$–8^+ or to neutron

Figure 1 | Particle identification plot. Events are plotted with respect to Z and the mass-to-charge ratio, A/Z, for the full statistics of the ^{100}Sn fragment separator setting. In total, 259 ^{100}Sn nuclei (those indicated in the figure) were unambiguously identified. Resolutions (full-widths at half-maximum) in mass of $\Delta A = 0.32$ ($A = N + Z$) and in nuclear charge of $\Delta Z = 0.25$ were obtained. The colours indicate the number of events per bin in a logarithmic scale as indicated on the right-hand side.

Figure 2 | Time distribution of first decay events. The histogram shows the observed time distribution of all first decay events in the nearest-neighbouring pixels after implantation of ^{100}Sn nuclei. Decay curves resulting from the MLH analysis are shown individually for ^{100}Sn (dashed) and its daughter nucleus ^{100}In (dash–dot). The solid line shows the sum of these decay curves and takes into account a small amount of random background.
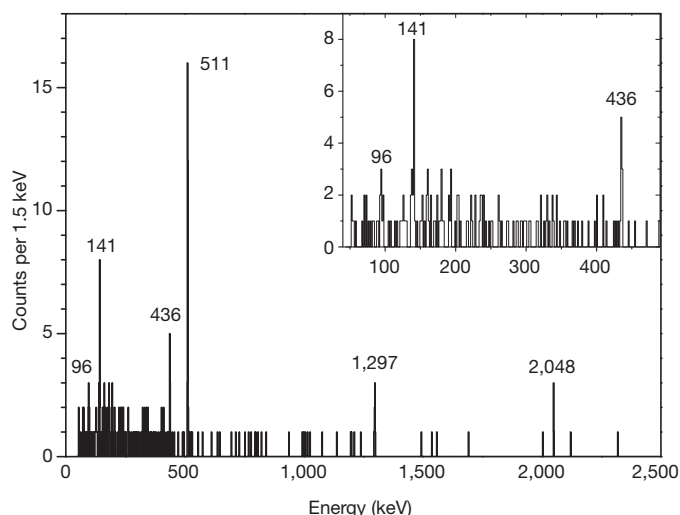
Figure 3 | Spectrum of γ-radiation. Energy distribution of the γ-radiation observed within 4 s after implantation of ^{100}Sn. With 65% probability these are directly following the ^{100}Sn decay. The other contributions are uncorrelated background decays and daughter decays of ^{100}In. None of the observed lines corresponds to known transitions from these minor contributions. The line at 511 keV is due to positron annihilation radiation. The measured absolute numbers of transitions of the five lines with the energies 96, 141, 436, 1,297 and 2,048 keV are respectively 79 ± 40, 100 ± 31, 59 ± 22, 72 ± 26 and 53 ± 26 corrected for electron conversion assuming M1 (magnetic dipole) transitions. Inset, enlarged view of the energy range up to 500 keV.

particles in the $d_{5/2}$ orbital ($\pi g_{9/2}^{-1} \otimes v d_{5/2}^{1}$) with $I^{\pi} = 2^{+}$–7^{+}. The predictions reflect the observed γ-ray transitions well if the high-energy, 2,048-keV, transition populates the lowest 2^{+} state, this then decays to lower-lying states via the 436-, 141- and 96-keV transitions, and the decay chain ends either in the 6^{+} ground state or a low-lying isomeric state with unobserved decay. In this picture, the second 2^{+} state is populated by the 1,297-keV transition and decays to the lower-lying 2^{+} and 3^{+} states. This may lead to a fragmentation of the intensities, making it impossible to observe these transitions in the present experiment. This picture is supported by three experimental facts: the measurement of the total γ-ray energy ($E^{*}_{1+} = 2.76 \pm 0.43$ MeV) in a previous experiment with a bismuth germanate detector[14]; the known mass difference between ^{100}Sn and ^{100}In[13], combined with our

measured β-decay end-point energy ($E^{*}_{1+} = 2.6 \pm 1.0$ MeV); and our observation of a single event of β-delayed proton emission ($E^{*}_{1+} = 2.93 \pm 0.34$ MeV). It is fully consistent with the expectation that a single 1^{+} state is dominantly populated in the decay. Further details are given in Methods Summary and Supplementary Information.

As a key feature of this experiment, we measured the kinetic energy of the decay positrons fully absorbed in the compact silicon detector array. The spectrum resulting from the summed energies deposited by a β-particle in the pixels of the calorimeter up to 3 s after a ^{100}Sn implantation is shown in Fig. 5. It was fitted using a MLH analysis based on a single-component β-decay phase space function to determine the end-point energy in the decay of ^{100}Sn. For the fit of the end-point energy, only data in the energy region between 400 and 2,600 keV were used. In the analysis, corrections were applied to account for the emission of conversion electrons instead of low-energy γ-rays during the de-excitation of the daughter nucleus ^{100}In, for bremsstrahlung emitted when the positrons are slowed down and for the annihilation of positrons in flight before the deposition of their total kinetic energy. The end-point energy of the β-decay, if populating a single final state in the daughter nucleus ^{100}In, was determined to be 3.29 ± 0.20 MeV. The corresponding fraction of electron-capture decays is 13% of all ^{100}Sn decays.

Discussion

Using the measured half-life and the end-point energy, we calculated a $\log(ft)$ value of $2.62^{+0.13}_{-0.11}$, which is the smallest such value found so far for any nuclear β-decay. Thus, the Gamow–Teller decay of ^{100}Sn has a much larger strength than the known $0^{+} \rightarrow 0^{+}$ superallowed Fermi decays of $N = Z$ nuclei and can indeed be considered a superallowed Gamow–Teller decay[21]. This finding is also illustrated in Fig. 6, which shows the distribution of $\log(ft)$ values for allowed Gamow–Teller and Fermi transitions.

The extracted Gamow–Teller strength of the ^{100}Sn ground-state decay to the single excited 1^{+} state in ^{100}In is $B_{GT} = 9.1^{+2.6}_{-3.0}$. The measured value is extraordinarily large but consistent with the value of $B_{GT} = 5.8^{+5.5}_{-3.2}$ deduced from previous results for Q_{EC} and half-life[22], within the large error bars of the earlier measurement. The uncertainty in the new B_{GT} value is dominated by the uncertainty in the β-decay end-point energy. The extraction of the strength was done under the assumption that the Gamow–Teller decay was into only one final 1^{+}
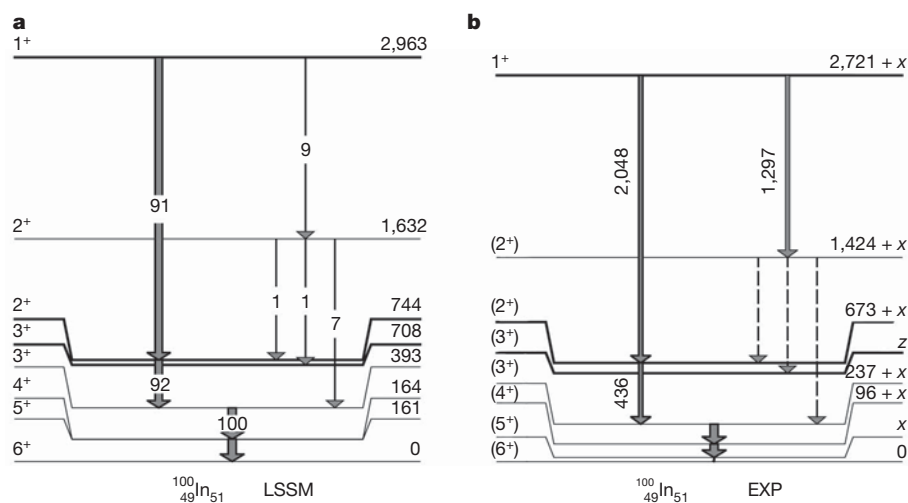
Figure 4 | Tentative level scheme of ^{100}In. a, LSSM calculation of the low-lying excited states in ^{100}In. Spin and parity (I^{π}) are shown on the left and energy (keV) is shown on the right. Populated levels with an almost pure $\pi g_{9/2}^{-1} \otimes v g_{7/2}^{1}$ configuration are indicated with bold lines, and the remaining levels are part of the $\pi g_{9/2}^{-1} \otimes v d_{5/2}^{1}$ multiplet. Gamma-transitions with their relative intensities (in per cent and indicated by arrow width) are shown for

selected transitions. **b,** Most likely level scheme for the five observed γ-transitions in ^{100}In (three with energies (keV) shown). Because one low-energy transition might have been missed, the energy of the levels might have a systematic shift of up to $x = 80$ keV. The dashed transitions and level z were not observed. The assignment of spin and parity is certain only for the 1^{+} state; the others are tentative assignments based on theory.
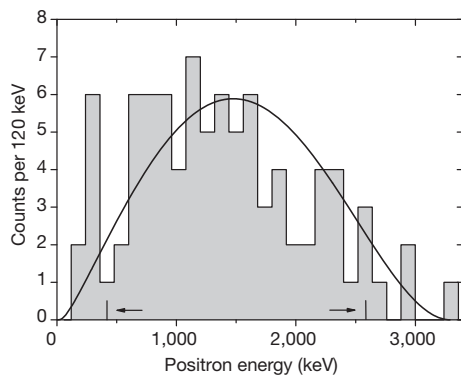
Figure 5 | Distribution of the positron energies emitted in the β-decay of ^{100}**Sn.** The spectrum contains only decay events that can be assigned to ^{100}Sn decays with a probability of at least 75%. The MLH fit was applied to the region between 400 and 2,600 keV, which is indicated with markers. The solid curve illustrates the shape of the best-fitting single-component β-decay phase space function determined by MLH analysis.

state in ^{100}In. However, if 1^+ states at excitation energies above the observed state were also populated, the summed Gamow–Teller strength would increase while the strength of the decay into the first excited 1^+ state would decrease. It would have been difficult to observe such higher-energy states because the reduced phase space for lower-energy $β^+$-particles would have led to a strongly reduced population.

The LSSM calculations, which within the *gds* harmonic oscillator shell take into account most of the long-range correlations across the $N = Z = 50$ doubly magic shell closure and include up to five particle–hole excitations (Methods Summary and Supplementary Information), yield a total summed Gamow–Teller strength of $B_{GT} = 8.19$ for all possible final states in the daughter nucleus up to 60 MeV. The standard renormalization due to correlations beyond the 0g, 1d, 2s model space has been applied[5]. The distribution of strength up to an excitation energy of 10 MeV is shown in Methods Summary and Supplementary Fig. 1.

A Gamow–Teller strength of $B_{GT} = 7.82$ is predicted in the experimental Q_{EC} window of 7.03(20) MeV. This corresponds to a reduction

in the total renormalized Gamow–Teller strength of the extreme single-particle estimate ($B_{GT,ESPM} \approx 10$) by 18% for excitation energies up to 60 MeV and by 22% in the Q_{EC} window. It is due to mixing in the *gds* harmonic oscillator shell, that is, emptying of the proton $g_{9/2}$ orbital, pre-filling of the neutron $g_{7/2}$ orbital and destructive interference of the four possible combinations of Gamow–Teller transitions within the g orbital ($L = 4$). The occupation numbers of the two orbitals that are linked by the Gamow–Teller operator, which acts on spin and isospin but does not change L, directly influence the strength of the transition matrix element.

The calculation predicts that the largest fraction of the strength remains located in the first excited 1^+ state, in agreement with earlier calculations[7,21]. Nevertheless, according to our LSSM calculations it is reasonable to consider that several 1^+ states in ^{100}In are populated in the decay of ^{100}Sn. If, as an exercise, we take from the LSSM calculation the four lowest 1^+ states in ^{100}In with their energy splittings and relative Gamow–Teller strengths (Methods Summary and Supplementary Information), the value of $B_{GT}(1_1^+) = 9.1^{+2.6}_{-3.0}$ (assuming a single 1^+ state) would be reduced to $B_{GT}(1_1^+) = 7.6^{+2.2}_{-2.5}$ for the first excited 1^+ state using the experimental half-life and β-spectrum. The corresponding summed Gamow–Teller strength would be $\sum_{i=1}^{4} B_{GT}(1_i^+) = 9.9^{+2.8}_{-3.2}$. Because this exercise served only to gauge the effect of branching on the experimental B_{GT} value, no error for the branching ratios is included.

The LSSM result of $B_{GT} = 5.7$ for this first excited 1^+ state agrees within the statistical uncertainty with the value, $B_{GT} = 7.6^{+2.2}_{-2.5}$, extracted from the experimental log(ft) value under the above assumptions. The experimental concentration of most of the Gamow–Teller strength in the first excited 1^+ state clearly classifies the ^{100}Sn Gamow–Teller decay as superallowed. This large experimental Gamow–Teller strength of the transition to the first excited 1^+ state proves that both the ^{100}Sn ground state and the first excited 1^+ state in ^{100}In have relatively pure wavefunctions. As expected, the LSSM calculation reveals that the respective wavefunctions consist predominantly of the $\pi g_{9/2}^{10} \otimes \nu g_{7/2}^0$ (82% probability) and $\pi g_{9/2}^9 \otimes \nu g_{7/2}^1$ (54% probability) components. The high purity of the wavefunctions within the *gds* model space establishes the simultaneous robustness of the $Z = 50$ and $N = 50$ shell closures in ^{100}Sn, which is only ~3 MeV from the proton drip line, corroborating for $N = 50$ the results of refs 23, 24, and excludes the need for explicitly treating the unbound proton orbits as continuum states.

The LSSM calculations allow enough configuration mixing in the *gds* shell that convergent results are obtained, leading to meaningful conclusions in this exotic region far from the valley of stability. This indicates that it should be possible to obtain reliable, more accurate results for nuclei in the neighbourhood of ^{100}Sn, especially close to the proton drip line.

The underlying shell structures of nuclei in the vicinity of ^{100}Sn have to be determined with the highest possible accuracy to address the important issues in nuclear structure, such as the possibility of a new coupling scheme developing in the $N = Z$ nuclei in the vicinity of ^{100}Sn (ref. 25). The present measurement is a stringent test for LSSM calculations, in which the realistic character of such a coupling scheme still needs to be probed.

A better understanding of the nuclear structure is of major importance for modelling weak interaction rates in nuclei, which depend on the underlying shell structure and are important in many astrophysical processes. For example, Gamow–Teller transitions govern electron capture during the core collapse of supernovae. Also, Gamow–Teller transitions are an essential constraint on the theoretical calculations of neutrino-less double-β-decay matrix elements, the knowledge of which is necessary to relate the neutrino mass to the rate of this yet undiscovered process. Further interest in the decay rates of nuclei around ^{100}Sn comes from the study of certain astrophysical processes, as this region has been considered the end of the rapid proton capture process due to the Sn-Sb-Te cycle.

Figure 6 | Log(ft) values of allowed nuclear β-decays. Number distribution of log(ft) values for allowed β-transitions (obeying the selection rules). The data are from ref. 26. The values are for generally allowed Gamow–Teller transitions between 0^+ and 1^+ states (black), mixed Fermi/Gamow–Teller transitions (blue) and the well-established pure, superallowed Fermi transitions from 0^+ to 0^+ states (green). The decay of ^{100}Sn is unique because it has the smallest known log(ft) value (red) of any nuclear β-decay.

METHODS SUMMARY

^{100}Sn and neighbouring nuclei were produced by fragmentation of a 1.0A-GeV ^{124}Xe beam from the GSI accelerators, separated in the fragment separator and identified by multiple energy-loss, magnetic rigidity and time-of-flight measurements. The nuclei were stopped in an implantation detector with high spatial resolution to correlate implantations with succeeding decays. The device was surrounded by the stopped-beam RISING array of 15 × 7 germanium detectors in close geometry. In this configuration, the set-up enabled us to do nearly 4π spectroscopy of the emitted γ-radiation and particle-decay radiation. With a photopeak efficiency of about 10% (1 MeV) for γ-ray detection and nearly 100% for full energy detection of decay particles up to 5 MeV, this high-resolution set-up allowed for a maximum use of the secondary beam.

The ^{100}Sn half-life and the β-decay end-point energy were calculated in the framework of a MLH analysis applied respectively to the time distribution of β-decays after implantation and the energy distribution of emitted positrons. This analysis also considered the daughter decays and the presence of uncorrelated random background decays from previous implantations.

To interpret the measured Gamow–Teller strength and the observed γ-rays emitted from ^{100}In, we carried out LSSM calculations. The valence space used in the LSSM consists of the fifth (4$\hbar\omega$) harmonic oscillator shell, that is, proton and neutron $\pi\nu(g,d,s)$ orbitals outside the ^{80}Zr core. The calculations included up to five particle–hole excitations from the $g_{9/2}$ proton and neutron orbitals to the rest of the shell, which made it possible for us to obtain convergent results for excitation spectra and the Gamow–Teller strength.

1. Ichimura, M., Sakai, H. & Wakasa, T. Spin-isospin responses via (p,n) and (n,p) reactions. *Prog. Part. Nucl. Phys.* **56,** 446–531 (2006).
2. Hardy, J. C., Carraz, L. C., Jonson, B. & Hansen, P. G. The essential decay of pandemonium: a demonstration of errors in complex beta-decay schemes. *Phys. Lett. B* **71,** 307–310 (1977).
3. Audi, G., Wapstra, A. H. & Thibault, C. The Ame2003 atomic mass evaluation: (ll). Tables, graphs and references. *Nucl. Phys. A* **729,** 337–676 (2003).
4. Sasano, M. *et al.* Gamow-Teller transition strengths from ^{56}Ni. *Phys. Rev. Lett.* **107,** 202501 (2011).
5. Caurier, E., Martínez-Pinedo, G., Nowacki, F., Poves, A. & Zuker, P. A. The shell model as a unified view of nuclear structure. *Rev. Mod. Phys.* **77,** 427–488 (2005).
6. Brown, B. A. The nuclear shell model towards the drip lines. *Prog. Part. Nucl. Phys.* **47,** 517–599 (2001).
7. Brown, B. A. & Rykaczewski, K. Gamow-Teller strength in the region of ^{100}Sn. *Phys. Rev. C* **50,** R2270–R2273 (1994).
8. Dean, D. J., Koonin, S. E., Kuo, T. T. S., Langanke, K. & Radha, P. B. Complete 0$\hbar\omega$ shell model Monte Carlo calculations of ^{94}Ru, ^{96}Pd, ^{96,98}Cd and ^{100}Sn. *Phys. Lett. B* **367,** 17–20 (1996).
9. Bobyk, A., Kaminski, W. & Borzov, I. N. Gamow-Teller beta-decay strengths of neutron-deficient tin isotopes: comparison of FFST and pnBCS+QRPA results. *Acta Phys. Pol. B* **31,** 953–963 (2000).
10. Batist, L. *et al.* Systematics of Gamow-Teller beta decay "Southeast" of ^{100}Sn. *Eur. Phys. J. A* **46,** 45–53 (2010).
11. Schneider, R. *et al.* Production and identification of ^{100}Sn. *Z. Phys. A* **348,** 241–242 (1994).
12. Lewitowicz, M. *et al.* Identification of the doubly-magic nucleus ^{100}Sn in the reaction ^{112}Sn+^{nat}Ni at 63 MeV/nucleon. *Phys. Lett. B* **332,** 20–24 (1994).
13. Chartier, M. *et al.* Mass measurement of ^{100}Sn. *Phys. Rev. Lett.* **77,** 2400–2403 (1996).
14. Sümmerer, K. *et al.* Identification and decay spectroscopy of ^{100}Sn at the GSI projectile fragment separator FRS. *Nucl. Phys. A* **616,** 341–345 (1997).
15. Stolz, A. *et al.* Projectile fragmentation of ^{112}Sn at E_{lab} = 1A GeV. *Phys. Rev. C* **65,** 064603 (2002).
16. Bazin, D. *et al.* Production and beta-decay of rp-process nuclei ^{96}Cd, ^{98}In, and ^{100}Sn. *Phys. Rev. Lett.* **101,** 252501 (2008).
17. Geissel, H. *et al.* The GSI projectile fragment separator (FRS): a versatile magnetic system for relativistic heavy ions. *Nucl. Instrum. Methods B* **70,** 286–297 (1992).
18. Pietri, S. *et al.* Recent results in fragmentation isomer spectroscopy with rising. *Nucl. Instrum. Methods B* **261,** 1079–1083 (2007).
19. Plettner, C. *et al.* Beta decay of ^{100}In. *Phys. Rev. C* **66,** 044319 (2002).
20. Coraggio, L., Covello, A., Gargano, A. & Itaco, N. Structure of particle-hole nuclei around ^{100}Sn. *Phys. Rev. C* **70,** 034310 (2004).
21. Hamamoto, I. & Sagawa, H. Gamow-Teller beta decay and isospin impurity in nuclei near the proton drip line. *Phys. Rev. C* **48,** R960–R963 (1993).
22. Faestermann, T. *et al.* Decay studies of N ≈ Z nuclei from ^{75}Sr to ^{102}Sn. *Eur. Phys. J. A* **15,** 185–188 (2002).
23. Blazhev, A. *et al.* Observation of a core-excited E4 isomer in ^{98}Cd. *Phys. Rev. C* **69,** 064304 (2004).
24. Boutachkov, P. *et al.* High-spin isomers in ^{96}Ag: excitations across the Z = 38 and Z = 50, N = 50 closed shells. *Phys. Rev. C* **84,** 044311 (2011).
25. Cederwall, B. *et al.* Evidence for a spin-aligned neutron–proton paired phase from the level structure of ^{92}Pd. *Nature* **469,** 68–71 (2011).
26. Singh, B., Rodriguez, J. L., Wong, S. S. M. & Tuli, J. K. Review of logft values in β-decay. *Nucl. Data Sheets (N.Y. N.Y.)* **84,** 487–563 (1998).

Author Contributions Fragment separator: H.W., P.B., H. Geissel, M.G., Zs.P. and C.N.; particle detectors: C.B.H., K. Straub, R.G., T.F., L.M. and F. Nebel; RISING γ-array: P.B., M.G., S.P., J.G., I.M.K. and H.-J.W.; data acquisition and analysis software: M.B., R.G., J.L.G., N.K. and L.M.; data analysis and interpretation: C.B.H., K. Straub, T.F., M.G., H. Grawe, R.K., K. Steiger, F. Nowacki and K. Sieja; shell model calculations: F. Nowacki and K. Sieja; writing of manuscript: C.B.H., T.F., M.G., H. Grawe, R.K., F. Nowacki and K. Sieja. All authors except H. Grawe, F. Nowacki and K. Sieja took part in the preparation and the experiments, and all authors commented on the final paper.

ARTICLE

The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups

Christina Curtis[1,2]†*, Sohrab P. Shah[3,4]*, Suet-Feung Chin[1,2]*, Gulisa Turashvili[3,4], Oscar M. Rueda[1,2], Mark J. Dunning[2], Doug Speed[2,5]†, Andy G. Lynch[1,2], Shamith Samarajiwa[1,2], Yinyin Yuan[1,2], Stefan Gräf[1,2], Gavin Ha[3], Gholamreza Haffari[3], Ali Bashashati[3], Roslin Russell[2], Steven McKinney[3,4], METABRIC Group‡, Anita Langerød[6], Andrew Green[7], Elena Provenzano[8], Gordon Wishart[8], Sarah Pinder[9], Peter Watson[3,4,10], Florian Markowetz[1,2], Leigh Murphy[10], Ian Ellis[7], Arnie Purushotham[9,11], Anne-Lise Børresen-Dale[6,12], James D. Brenton[2,13], Simon Tavaré[1,2,5,14], Carlos Caldas[1,2,8,13] & Samuel Aparicio[3,4]

The elucidation of breast cancer subgroups and their molecular drivers requires integrated views of the genome and transcriptome from representative numbers of patients. We present an integrated analysis of copy number and gene expression in a discovery and validation set of 997 and 995 primary breast tumours, respectively, with long-term clinical follow-up. Inherited variants (copy number variants and single nucleotide polymorphisms) and acquired somatic copy number aberrations (CNAs) were associated with expression in ~40% of genes, with the landscape dominated by *cis*- and *trans*-acting CNAs. By delineating expression outlier genes driven in *cis* by CNAs, we identified putative cancer genes, including deletions in *PPP2R2A*, *MTAP* and *MAP2K4*. Unsupervised analysis of paired DNA–RNA profiles revealed novel subgroups with distinct clinical outcomes, which reproduced in the validation cohort. These include a high-risk, oestrogen-receptor-positive 11q13/14 *cis*-acting subgroup and a favourable prognosis subgroup devoid of CNAs. *Trans*-acting aberration hotspots were found to modulate subgroup-specific gene networks, including a TCR deletion-mediated adaptive immune response in the 'CNA-devoid' subgroup and a basal-specific chromosome 5 deletion-associated mitotic network. Our results provide a novel molecular stratification of the breast cancer population, derived from the impact of somatic CNAs on the transcriptome.

Inherited genetic variation and acquired genomic aberrations contribute to breast cancer initiation and progression. Although somatically acquired CNAs are the dominant feature of sporadic breast cancers, the driver events that are selected for during tumorigenesis are difficult to elucidate as they co-occur alongside a much larger landscape of random non-pathogenic passenger alterations[1,2] and germline copy number variants (CNVs). Attempts to define subtypes of breast cancer and to discern possible somatic drivers are still in their relative infancy[3–6], in part because breast cancer represents multiple diseases, implying that large numbers (many hundreds or thousands) of patients must be studied. Here we describe an integrated genomic/transcriptomic analysis of breast cancers with long-term clinical outcomes composed of a discovery set of 997 primary tumours and a validation set of 995 tumours from METABRIC (Molecular Taxonomy of Breast Cancer International Consortium).

A breast cancer population genomic resource

We assembled a collection of over 2,000 clinically annotated primary fresh-frozen breast cancer specimens from tumour banks in the UK and Canada (Supplementary Tables 1–3). Nearly all oestrogen receptor (ER)-positive and/or lymph node (LN)-negative patients did not receive chemotherapy, whereas ER-negative and LN-positive patients did. Additionally, none of the HER2[+] patients received trastuzumab. As such, the treatments were homogeneous with respect to clinically relevant groupings. An initial set of 997 tumours was analysed as a discovery group and a further set of 995 tumours, for which complete data later became available, was used to test the reproducibility of the integrative clusters (described below). An overview of the main analytical approaches is provided in Supplementary Fig. 1. Details concerning expression and copy number profiling, including sample assignment to the PAM50 intrinsic subtypes[3,4,7] (Supplementary Fig. 2), copy number analysis (Supplementary Tables 4–8) and validation (Supplementary Figs 3 and 4 and Supplementary Tables 9–11), and *TP53* mutational profiling (Supplementary Fig. 5) are described in the Supplementary Information.

Genome variation affects tumour expression architecture

Genomic variants are considered to act in *cis* when a variant at a locus has an impact on its own expression, or in *trans* when it is associated

[1]Department of Oncology, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK. [2]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. [3]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. [4]Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia V5Z 1L3, Canada. [5]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Cambridge CB3 0WA, UK. [6]Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Montebello, 0310 Oslo, Norway. [7]Department of Histopathology, School of Molecular Medical Sciences, University of Nottingham, Nottingham NG5 1PB, UK. [8]Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. [9]King's College London, Breakthrough Breast Cancer Research Unit, London WC2R 2LS, UK. [10]Manitoba Institute of Cell Biology, University of Manitoba, Manitoba R3E 0V9, Canada. [11]NIHR Comprehensive Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London, London WC2R 2LS, UK. [12]Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, 0316 Oslo, Norway. [13]Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK. [14]Molecular and Computational Biology Program, University of Southern California, Los Angeles, California 90089, USA. †Present addresses: Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA (Ch.C.); University College London, Genetics Institute, WC1E 6BT, UK (D.S.).
*These authors contributed equally to this work.
‡Lists of participants and affiliations appear at the end of the paper.

with genes at other sites in the genome. We generated a map of CNAs, CNVs (Supplementary Fig. 6, Supplementary Tables 12–15) and single nucleotide polymorphisms (SNPs) in the breast cancer genome to distinguish germline from somatic variants (see Methods), and to examine the impact of each of these variants on the expression landscape. Previous studies[8] have shown that most heritable gene expression traits are governed by a combination of *cis* (proximal) loci, defined here as those within a 3-megabase (Mb) window surrounding the gene of interest, and *trans* (distal) loci, defined here as those outside that window. We assessed the relative influence of SNPs, CNVs and CNAs on tumour expression architecture, using each of these variants as a predictor (see Methods) to elucidate expression quantitative trait loci (eQTLs) among patients.

Both germline variants and somatic aberrations were found to influence tumour expression architecture, having an impact on >39% (11,198/28,609) of expression probes genome-wide based on analysis of variance (ANOVA; see Methods), with roughly equal numbers of genes associated in *cis* and *trans*. CNAs were associated with the greatest number of expression profiles (Fig. 1, Supplementary Figs 7–13 and Supplementary Tables 16–20), but were rivalled by SNPs to explain a greater proportion of expression variation on a per-gene basis genome-wide, whereas the contribution from CNVs was more moderate (Fig. 1b and Supplementary Table 21). The true ratio of putative *trans* versus *cis* eQTLs is hard to estimate[9]; however, the large sample size used here allowed the detection of small effects, with 5,401 and 5,462 CNAs significantly (Šidák adjusted *P* value <0.0001) associated in *cis* or in *trans*, respectively. Whereas *cis*-associations tended to be stronger, the *trans*-acting loci modulated a larger number of messenger RNAs, as described below.

Expression outliers refine the breast cancer landscape

As shown above, ~20% of loci exhibit CNA-expression associations in *cis* (Supplementary Fig. 14). To refine this landscape further and identify the putative driver genes, we used profiles of outlying expression (see Methods and ref. 10) and the high resolution and sensitivity of the

Affymetrix SNP 6.0 platform to delineate candidate regions. This approach markedly reduces the complexity of the landscape to 45 regions (frequency > 5, Fig. 2) and narrows the focus, highlighting novel regions that modulate expression. The full enumeration of regions delineated by this approach and their subtype-specific associations (Supplementary Figs 15 and 16 and Supplementary Tables 22–24) includes both known drivers (for example, *ZNF703* (ref. 11), *PTEN* (ref. 12), *MYC*, *CCND1*, *MDM2*, *ERBB2*, *CCNE1* (ref. 13)) and putative driver aberrations (for example, *MDM1*, *MDM4*, *CDK3*, *CDK4*, *CAMK1D*, *PI4KB*, *NCOR1*).

The deletion landscape of breast cancer has been poorly explored, with the exception of *PTEN*. We illustrate three additional regions of significance centred on *PPP2R2A* (8p21, Fig. 2, region 11), *MTAP* (9p21, Fig. 2, region 15) and *MAP2K4* (17p11, Fig. 2, region 33), which exhibit heterozygous and homozygous deletions (Supplementary Figs 15, 17–19 and Supplementary Table 24) that drive expression of these loci. We observe breast cancer subtype-specific (enriched in mitotic ER-positive cancers) loss of transcript expression in *PPP2R2A*, a B-regulatory subunit of the PP2A mitotic exit holoenzyme complex. Somatic mutations in *PPP2R1A* have recently been reported in clear cell ovarian cancers and endometrioid cancers[14,15], and methylation silencing of *PPP2R2B* has also been observed in colorectal cancers[16]. Thus, dysregulation of specific PPP2R2A functions in luminal B breast cancers adds a significant pathophysiology to this subtype.

MTAP (9p21, a component of methyladenosine salvage) is frequently co-deleted with the *CDKN2A* and *CDKN2B* tumour suppressor genes in a variety of cancers[17] as we observe here (Supplementary Figs 17c and 18). The third deletion encompasses *MAP2K4* (also called *MKK4*) (17p11), a p38/Jun dual specificity serine/threonine protein kinase. *MAP2K4* has been proposed as a recessive cancer gene[18], with mutations noted in cell lines[19]. We show, for the first time, the recurrent deletion of *MAP2K4* (Supplementary Figs 17d and 19) concomitant with outlying expression (Supplementary Fig. 15) in predominantly ER-positive cases, and verify homozygous deletions (Supplementary Table 9) in primary tumours, strengthening the evidence for *MAP2K4* as a tumour suppressor in breast cancer.
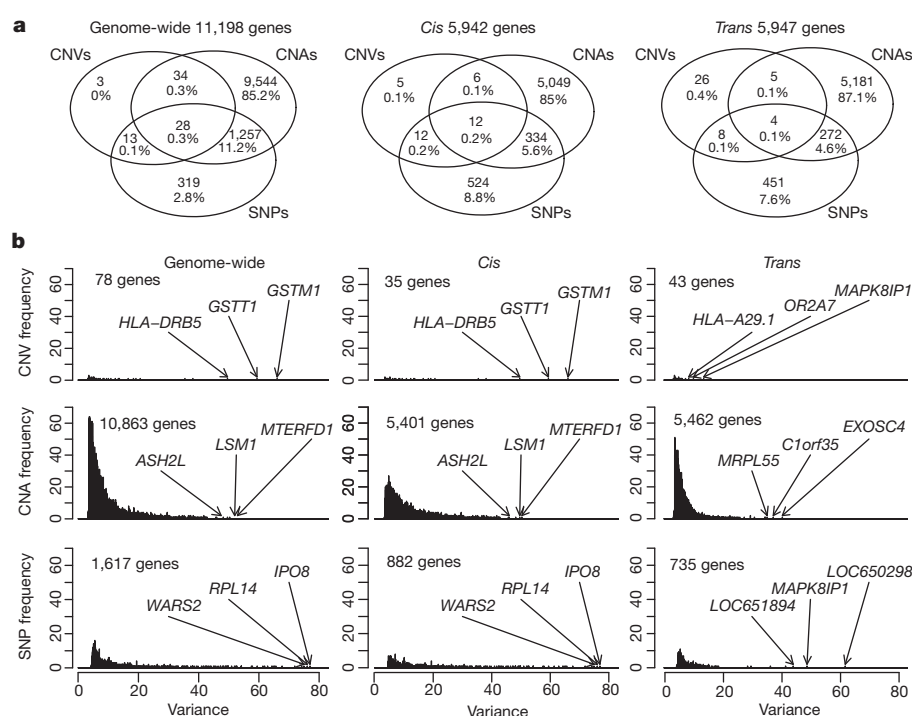
Figure 1 | Germline and somatic variants influence tumour expression architecture. a, Venn diagrams depict the relative contribution of SNPs, CNVs and CNAs to genome-wide, *cis* and *trans* tumour expression variation for significant expression associations (Šidák adjusted *P*-value ≤0.0001). **b**, Histograms illustrate the proportion of variance explained by the most significantly associated predictor for each predictor type, where several of the top associations are indicated.
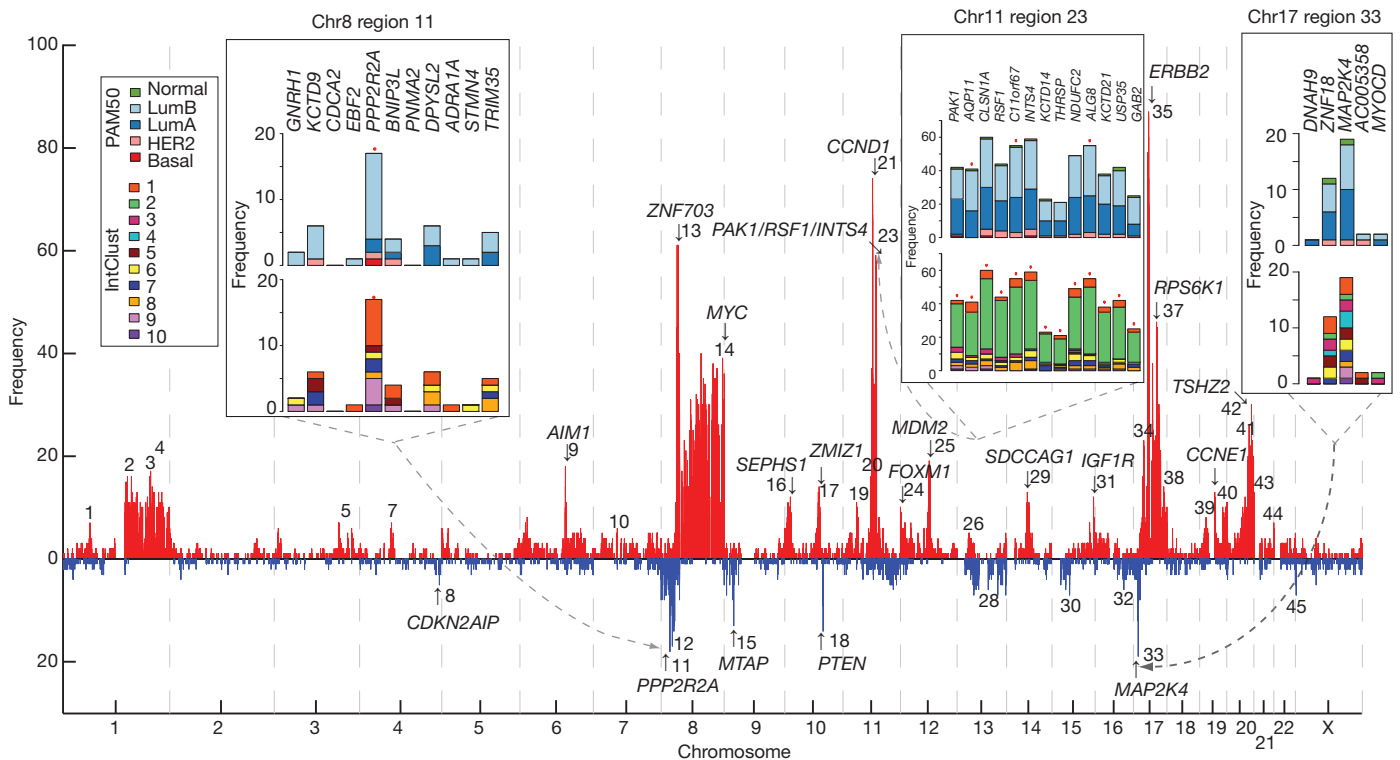
Figure 2 | Patterns of *cis* outlying expression refine putative breast cancer drivers. A genome-wide view of outlying expression coincident with extreme copy number events in the CNA landscape highlights putative driver genes, as indicated by the arrows and numbered regions. The frequency (absolute count) of cases exhibiting an outlying expression profile at regions across the genome is shown, as is the distribution across subgroups for several regions in the insets. High-level amplifications are indicated in red and homozygous deletions in blue. Red asterisks above the bar plots indicate significantly different observed distributions than expected based on the overall population frequency (χ^2 test, $P < 0.0001$).

Trans-acting associations reveal distinct modules

We next asked how *trans*-associated expression profiles are distributed across the genome. We mapped these in the expression landscape by examining the matrices of CNA–expression associations (see Methods). This revealed strong off-diagonal patterns at loci on chromosomes 1q, 7p, 8, 11q, 14q, 16, 17q and 20q (Fig. 3a), including both positive and negative associations, as well as numerous *trans*-acting aberration hotspots (defined as CNAs associated with >30 mRNAs). Importantly, these aberration hotspots can be grouped into pathway modules, which highlight known driver loci such as *ERBB2* and *MYC*, as well as novel loci associated with large *trans* expression modules (Supplementary Tables 25 and 26). The T-cell-receptor (TCR) loci on chromosomes 7 (*TRG*) and 14 (*TRA*) represent two such hotspots that modulated 381 and 153 unique mRNAs, respectively, as well as 19 dually regulated genes (Supplementary Fig. 20). These cognate mRNAs were highly enriched for T-cell activation and proliferation, dendritic cell presentation, and leukocyte activation, which indicate the induction of an adaptive immune response associated with tumour-infiltrating lymphocytes (Fig. 3b, Supplementary Fig. 20 and Supplementary Tables 27 and 28), as described later.

In a second approach, we examined the genome-wide patterns of linear correlation between copy number and expression features (see Methods), and noted the alignment of several off-diagonal signals, including those on chromosome 1q, 8q, 11q, 14q and 16 (Supplementary Fig. 21). Additionally, a broad signal on chromosome 5 localizing to a deletion event restricted to the basal-like tumours was observed (Supplementary Fig. 21), but was not detected with the eQTL framework, where discrete (as opposed to continuous) copy number values were used. This basal-specific *trans* module is enriched for transcriptional changes involving cell cycle, DNA damage repair and apoptosis (Supplementary Table 29), reflecting the high mitotic index typically associated with basal-like tumours, described in detail below.

Integrative clustering reveals novel subgroups

Using the discovery set of 997 breast cancers, we next asked whether novel biological subgroups could be found by joint clustering of copy number and gene expression data. On the basis of our finding that *cis*-acting CNAs dominated the expression landscape, the top 1,000 *cis*-associated genes across all subtypes (Supplementary Table 30) were used as features for input to a joint latent variable framework for integrative clustering[20] (see Methods). Cluster analysis suggested 10 groups (based on Dunn's index) (see Methods and Supplementary Figs 22 and 23), but for completeness, this result was compared with the results for alternative numbers of clusters and clustering schemes (see Methods, Supplementary Figs 23–27 and Supplementary Tables 31–33). The 10 integrative clusters (labelled IntClust 1–10) were typified by well-defined copy number aberrations (Fig. 4, Supplementary Figs 22, 28–30 and Supplementary Tables 34–39), and split many of the intrinsic subtypes (Supplementary Figs 31–33). Kaplan–Meier plots of disease-specific survival and Cox proportional hazards models indicate subgroups with distinct clinical outcomes (Fig. 5, Supplementary Figs 34, 35 and Supplementary Tables 40 and 41). To validate these results, we trained a classifier (754 features) for the integrative subtypes in the discovery set using the nearest shrunken centroids approach[21] (see Methods and Supplementary Tables 42 and 43), and then classified the independent validation set of 995 cases into the 10 groups (Supplementary Table 44). The reproducibility of the clusters in the validation set is shown in three ways. First, classification of the validation set resulted in the assignment of a similar proportion of cases to the 10 subgroups, each of which exhibited nearly identical copy number profiles (Fig. 4). Second, the groups have substantially similar hazard ratios (Fig. 5b, Supplementary Fig. 35 and Supplementary Table 40). Third, the quality of the clusters in the validation set is emphasized by the in-group proportions (IGP) measure[22] (Fig. 4).

Among the integrative clusters, we first note an ER-positive subgroup composed of 11q13/14 *cis*-acting luminal tumours (IntClust 2,
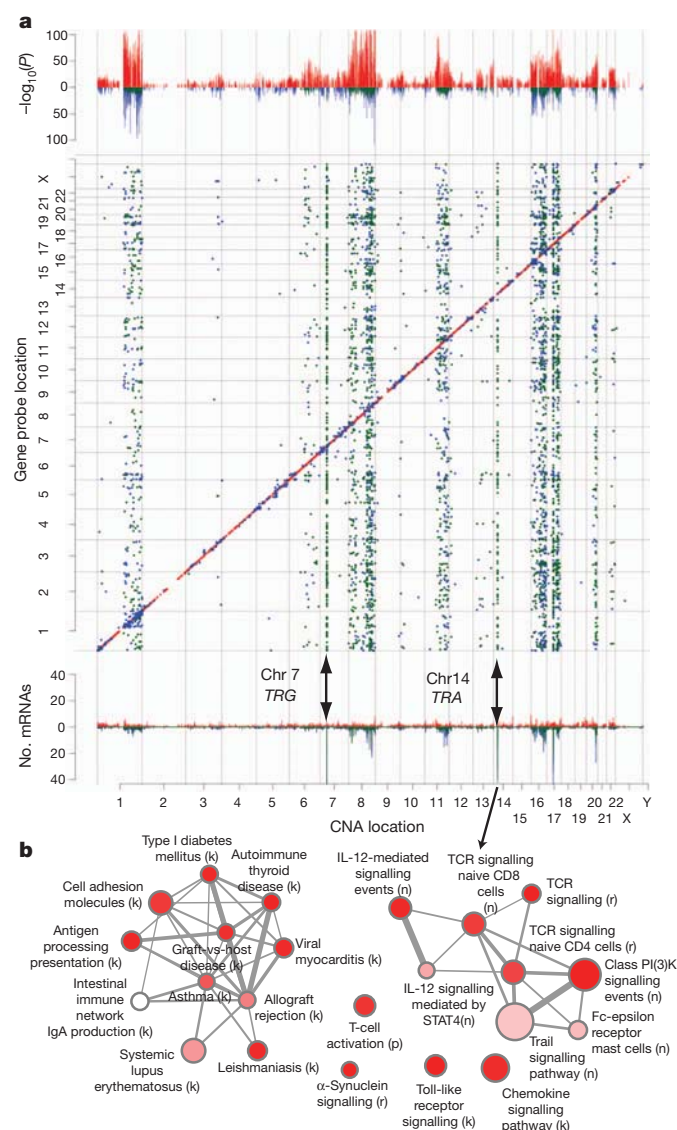
Figure 3 | *Trans*-acting aberration hotspots modulate concerted molecular pathways. a, Manhattan plot illustrating *cis* and *trans* expression-associated copy number aberrations from the eQTL analysis (top panel). The matrix of significant predictor–expression associations (adjusted *P*-value ≤0.0001) exhibits strong off-diagonal patterns (middle panel), and the frequency of mRNAs associated with a particular copy number aberration further illuminates these *trans*-acting aberration hotspots (bottom panel). The directionality of the associations is indicated as follows: *cis*: positive, red; negative, pink; *trans*: positive, blue; negative, green. **b**, Enrichment map of immune response modules in the *trans*-associated TRA network, where letters in parentheses represent the source database as follows: b, NCI-PID BioCarta; c, cancer cell map; k, KEGG; n, NCI-PID curated pathways; p, PANTHER; r, Reactome.

n = 45) that harbour other common alterations. This subgroup exhibited a steep mortality trajectory with elevated hazard ratios (discovery set: 3.620, 95% confidence interval (1.905–6.878); validation set: 3.353, 95% confidence interval (1.381–8.141)), indicating that it represents a particularly high-risk subgroup. Several known and putative driver genes reside in this region, namely *CCND1* (11q13.3), *EMSY* (11q13.5), *PAK1* (11q14.1) and *RSF1* (11q14.1), which have been previously linked to breast[13,23] or ovarian cancer[24]. Both the copy number (Fig. 4) and expression outlier landscapes (Fig. 2) suggest at least two separate amplicons at 11q13/14, one at *CCND1* (11q13.3) and a separate peak from 11q13.5-11q14.1 spanning *UVRAG–GAB2*, centred around *PAK1*, *RSF1*, *C11orf67* and *INTS4*, where it is more challenging to distinguish the driver[24]. Notably, the

expression outlier profiles for this region are enriched for samples belonging to IntClust 2 (Fig. 2, inset region 23) and all 45 members of this subgroup harboured amplifications of these genes, with high frequencies of amplification also observed for *CCND1* (*n* = 39) and *EMSY* (*n* = 34). In light of these observations, the 11q13/14 amplicon may be driven by a cassette of genes rather than a single oncogene.

Second, we note the existence of two subgroups marked by a paucity of copy number and *cis*-acting alterations. These subgroups cannot be explained by low cellularity tumours (see Methods). One subgroup (IntClust3, *n* = 156) with low genomic instability (Fig. 4 and Supplementary Fig. 22) was composed predominantly of luminal A cases, and was enriched for histotypes that typically have good prognosis, including invasive lobular and tubular carcinomas. The other subgroup (IntClust 4, *n* = 167) was also composed of favourable outcome cases, but included both ER-positive and ER-negative cases and varied intrinsic subtypes, and had an essentially flat copy number landscape, hence termed the 'CNA-devoid' subgroup. A significant proportion of cases within this subgroup exhibit extensive lymphocytic infiltration (Supplementary Table 45).

Third, several intermediate prognosis groups of predominantly ER-positive cancers were identified, including a 17q23/20q *cis*-acting luminal B subgroup (IntClust 1, *n* = 76), an 8p12 *cis*-acting luminal subgroup (IntClust 6, *n* = 44), as well as an 8q *cis*-acting/20q-amplified mixed subgroup (IntClust 9, *n* = 67). Two luminal A subgroups with similar CNA profiles and favourable outcome were noted. One subgroup is characterized by the classical 1q gain/16q loss (IntClust 8, *n* = 143), which corresponds to a common translocation event[25], and the other lacks the 1q alteration, while maintaining the 16p gain/16q loss with higher frequencies of 8q amplification (IntClust 7, *n* = 109). We also noted that the majority of basal-like tumours formed a stable, mostly high-genomic instability subgroup (IntClust 10, *n* = 96). This subgroup had relatively good long-term outcomes (after 5 years), consistent with ref. 26, and characteristic *cis*-acting alterations (5 loss/8q gain/10p gain/12p gain).

The *ERBB2*-amplified cancers composed of HER2-enriched (ER-negative) cases and luminal (ER-positive) cases appear as IntClust 5 (*n* = 94), thus refining the ERBB2 intrinsic subtype by grouping additional patients that might benefit from targeted therapy. Patients in this study were enrolled before the general availability of trastuzumab, and as expected this subgroup exhibits the worst disease-specific survival at both 5 and 15 years and elevated hazard ratios (discovery set: 3.899, 95% confidence interval (2.234–6.804); validation set: 4.447, 95% confidence interval (2.284–8.661)).

Pathway deregulation in the integrative subgroups

Finally, we projected the molecular profiles of the integrative subgroups onto pathways to examine possible biological themes among breast cancer subgroups (Supplementary Tables 46 and 47) and the relative impact of *cis* and *trans* expression modules on the pathways. The CNA-devoid (IntClust 4) group exhibits a strong immune and inflammation signature involving the antigen presentation pathway, OX40 signalling, and cytotoxic T-lymphocyte-mediated apoptosis (Supplementary Fig. 36). Given that *trans*-acting deletion hotspots were localized to the *TRG* and *TRA* loci and were associated with an adaptive immune response module, we asked whether these deletions contribute to alterations in this pathway. The CNA-devoid subgroup (IntClust 4) was found to exhibit nearly twice as many deletions (typically heterozygous loss) at the *TRG* and *TRA* loci (~20% of cases) as compared to the other subtypes (with the exception of IntClust 10), and deletions of both TCR loci were significantly associated with severe lymphocytic infiltration (χ^2 test, $P < 10^{-9}$ and $P < 10^{-8}$, respectively). Notably, these *trans*-associated mRNAs were significantly enriched in the immune response signature of the CNA-devoid subgroup (Supplementary Fig. 36) as well as among genes differentially expressed in CNA-devoid cases with severe lymphocytic infiltration (Supplementary Fig. 37). We conclude that genomic copy number loss
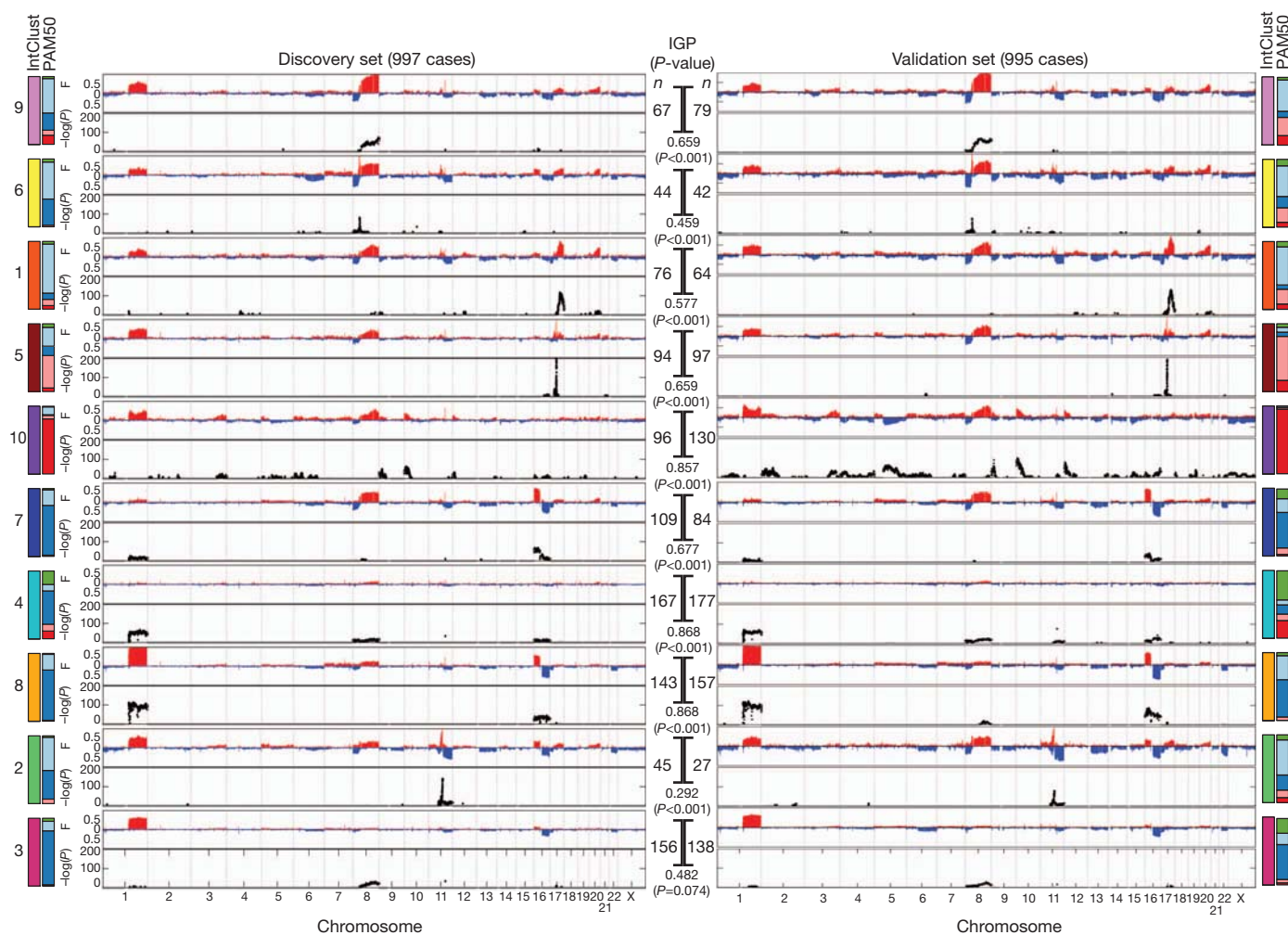
Figure 4 | The integrative subgroups have distinct copy number profiles.
Genome-wide frequencies (F, proportion of cases) of somatic CNAs (y-axis, upper plot) and the subtype-specific association ($-\log_{10}$ P-value) of aberrations (y-axis, bottom plot) based on a χ^2 test of independence are shown for each of the 10 integrative clusters. Regions of copy number gain are indicated in red and regions of loss in blue in the frequency plot (upper plot). Subgroups were

ordered by hierarchical clustering of their copy number profiles in the discovery cohort ($n = 997$). For the validation cohort ($n = 995$), samples were classified into each of the integrative clusters as described in the text. The number of cases in each subgroup (n) is indicated as is the in-group proportion (IGP) and associated P-value, as well as the distribution of PAM50 subtypes within each cluster.

at the TCR loci drives a *trans*-acting immune response module that associates with lymphocytic infiltration, and characterizes an otherwise genomically quiescent subgroup of ER-positive and ER-negative patients with good prognosis. These observations suggest the presence of mature T lymphocytes (with rearranged TCR loci), which may explain an immunological response to the cancer. In line with these findings, a recent study[27] demonstrated the association between CD8+ lymphocytes and favourable prognosis.

Also among the *trans*-influenced groups is IntClust 10 (basal-like cancer enriched subgroup), which harbours chromosome 5q deletions (Supplementary Fig. 21). Numerous signalling molecules, transcription factors and cell division genes were associated in *trans* with this deletion event in the basal cancers, including alterations in *AURKB*, *BCL2*, *BUB1*, *CDCA3*, *CDCA4*, *CDC20*, *CDC45*, *CHEK1*, *FOXM1*, *HDAC2*, *IGF1R*, *KIF2C*, *KIFC1*, *MTHFD1L*, *RAD51AP1*, *TTK* and *UBE2C* (Supplementary Fig. 38). Notably, *TTK* (*MPS1*), a dual specificity kinase that assists *AURKB* in chromosome alignment during mitosis, and recently reported to promote aneuploidy in breast cancer[28], was upregulated. These results indicate that 5q deletions modulate the coordinate transcriptional control of genomic and chromosomal instability and cell cycle regulation within this subgroup.

In contrast to these subtype-specific *trans*-associated signatures, the high-risk 11q13/14 subgroup was characterized by strong

cis-acting associations. Like the basal cancers, this subgroup also exhibited alterations in key cell-cycle-related genes (Supplementary Fig. 39), which probably have a role in its aggressive pathophysiology, but the nature of the signature differs. In particular, the regulation of the G1/S transition by BTG family proteins, which include CCND1, PPP2R1B and E2F2, was significantly enriched in the 11q13/14 *cis*-acting subgroup, but not the basal cancers, and this is consistent with *CCND1* and the *PPP2R* subunit representing subtype-specific drivers in these tumours.

Discussion

We have generated a robust, population-based molecular subgrouping of breast cancer based on multiple genomic views. The size and nature of this cohort made it amenable to eQTL analyses, which can aid the identification of loci that contribute to the disease phenotype[29]. CNAs and SNPs influenced expression variation, with CNAs dominating the landscape in *cis* and *trans*. The joint clustering of CNAs and gene expression profiles further resolves the considerable heterogeneity of the expression-only subgroups, and highlights a high-risk 11q13/14 *cis*-acting subgroup as well as several other strong *cis*-acting clusters and a genomically quiescent group. The reproducibility of subgroups with these molecular and clinical features in a validation cohort of 995 tumours suggests that by integrating multiple genomic
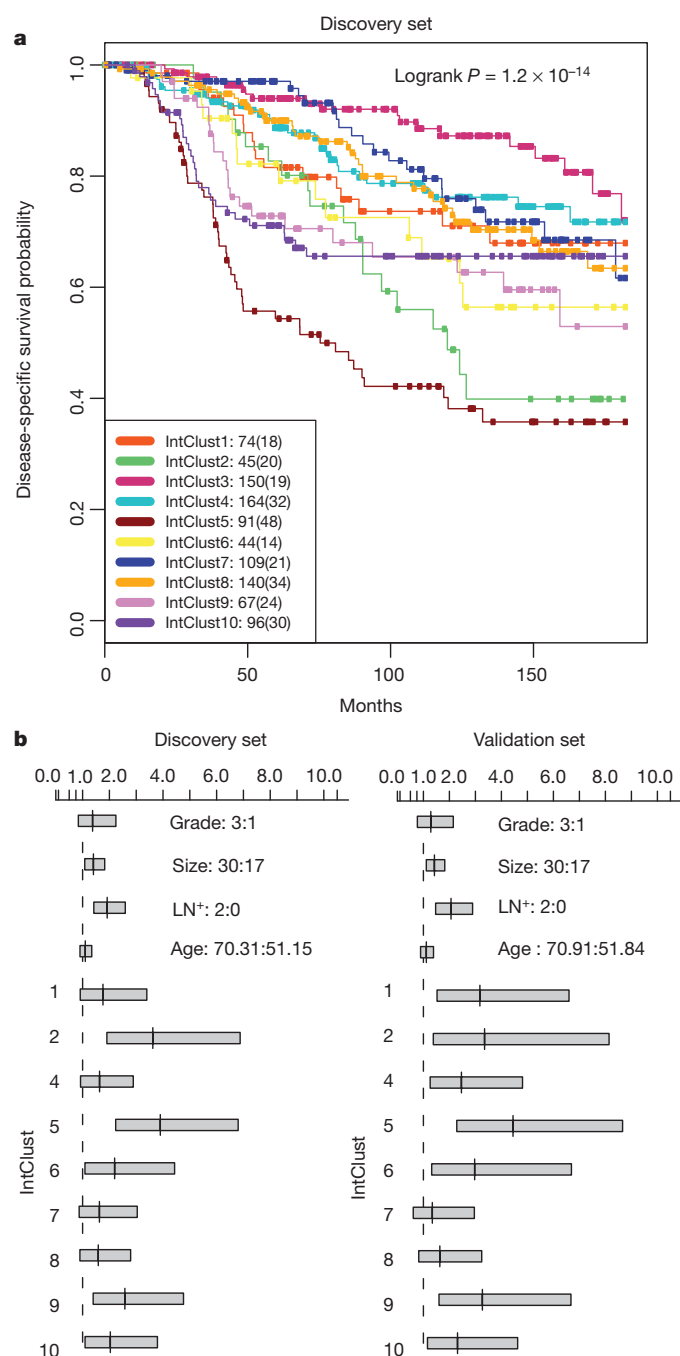
Figure 5 | The integrative subgroups have distinct clinical outcomes.
a, Kaplan–Meier plot of disease-specific survival (truncated at 15 years) for the integrative subgroups in the discovery cohort. For each cluster, the number of samples at risk is indicated as well as the total number of deaths (in parentheses). **b**, 95% confidence intervals for the Cox proportional hazard ratios are illustrated for the discovery and validation cohort for selected values of key covariates, where each subgroup was compared against IntClust 3.

features it may be possible to derive more robust patient classifiers. We show here, for the first time, that subtype-specific *trans*-acting aberrations modulate concerted transcriptional changes, such as the TCR deletion-mediated adaptive immune response that characterizes the CNA-devoid subgroup and the chromosome 5 deletion-associated cell cycle program in the basal cancers.

The integrated CNA-expression landscape highlights a limited number of genomic regions that probably contain driver genes, including *ZNF703*, which we recently described as a luminal B specific driver[11], as well as somatic deletion events affecting key subunits of the

PP2A holoenzyme complex and *MTAP*, which have previously been under-explored in breast cancer. The CNA-expression landscape also illuminates rare but potentially significant events, including *IGF1R*, *KRAS* and *EGFR* amplifications and *CDKN2B*, *BRCA2*, *RB1*, *ATM*, *SMAD4*, *NCOR1* and *UTX* homozygous deletions. Although some of these events have low overall frequencies (<1% patients) (Figs 2, Supplementary Fig. 15 and Supplementary Tables 22–24), they may have implications for understanding therapeutic responses to targeted agents, particularly those targeting tyrosine kinases or phosphatases.

Finally, because the integrative subgroups occur at different frequencies in the overall population, focusing sequencing efforts on representative numbers from these groups will help to establish a comprehensive breast cancer somatic landscape at sequence-level resolution. For example, a significant number (~17%, n = 167 in the discovery cohort) of breast cancers are devoid of somatic CNAs, and are ripe for mutational profiling. Our work provides a definitive framework for understanding how gene copy number aberrations affect gene expression in breast cancer and reveals novel subgroups that should be the target of future investigation.

METHODS SUMMARY

All patient specimens were obtained with appropriate consent from the relevant institutional review board. DNA and RNA were isolated from samples and hybridized to the Affymetrix SNP 6.0 and Illumina HT-12 v3 platforms for genomic and transcriptional profiling, respectively. A detailed description of the experimental assays and analytical methods used to analyse these data are available in the Supplementary Information.

1. Leary, R. J. *et al.* Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc. Natl Acad. Sci. USA* **105,** 16224–16229 (2008).
2. Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463,** 893–898 (2010).
3. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).
4. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98,** 10869–10874 (2001).
5. Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10,** 529–541 (2006).
6. Chin, S. F. *et al.* High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol.* **8,** R215 (2007).
7. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27,** 1160–1167 (2009).
8. Stranger, B. E. *et al.* Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1,** e78 (2005).
9. Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24,** 408–415 (2008).
10. Teschendorff, A. E., Naderi, A., Barbosa-Morais, N. L. & Caldas, C. PACK: Profile analysis using clustering and kurtosis to find molecular classifiers in cancer. *Bioinformatics* **22,** 2269–2275 (2006).
11. Holland, D. *et al.* ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. *EMBO Mol. Med.* **3,** 167–180 (2011).
12. Li, J. *et al.* PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275,** 1943–1947 (1997).
13. Santarius, T., Shiply, J., Brewer, D., Stratton, M. R. & Cooper, C. S. A census of amplified and overexpressed human cancer genes. *Nature Rev. Cancer* **10,** 59–64 (2010).
14. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330,** 228–231 (2010).
15. McConechy, M. K. *et al.* Subtype-specific mutation of PPP2R1A in endometrial and ovarian carcinomas. *J. Pathol.* **223,** 567–573 (2011).
16. Tan, J. *et al.* B55β-associated PP2A complex controls PDK1-directed MYC signaling and modulates rapamycin sensitivity in colorectal cancer. *Cancer Cell* **18,** 459–471 (2010).
17. Christopher, S. A., Diegelman, P., Porter, C. W. & Kruger, W. D. Methylthioadenosine phosphorylase, a gene frequently codeleted with p16 (CDKN2A/ARF), acts as a tumor suppressor in a breast cancer cell line. *Cancer Res.* **62,** 6639–6644 (2002).
18. Teng, D. H. *et al.* Human mitogen-activated protein kinase kinase 4 as a candidate tumor suppressor. *Cancer Res.* **57,** 4177–4182 (1997).
19. Hollestelle, A. *et al.* Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res. Treat.* **121,** 53–64 (2010).

20. Shen, R., Olshen, A. B. & Ladanyi, M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* **25**, 2906–2912 (2009).

21. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl Acad. Sci. USA* **99**, 6567–6572 (2002).

22. Kapp, A. V. & Tibshirani, R. Are clusters found in one dataset present in another dataset? *Biostatistics* **8**, 9–31 (2007).

23. Hughes-Davies, L. *et al.* EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell* **115**, 523–535 (2003).

24. Brown, L. A. *et al.* Amplification of 11q13 in ovarian carcinoma. *Genes Chromosom. Cancer* **47**, 481–489 (2008).

25. Russnes, H. G. *et al.* Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci. Transl. Med.* **2**, 38ra47 (2010).

26. Blows, F. M. *et al.* Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med.* **7**, e1000279 (2010).

27. Mahmoud, S. M. A. *et al.* Tumor-infiltrating CD8[+] lymphocytes predict clinical outcome in breast cancer. *J. Clin. Oncol.* **29**, 1949–1955 (2011).

28. Daniel, J., Coulter, J., Woo, J.-H., Wilsbach, K. & Gabrielson, E. High levels of the Mps1 checkpoint protein are protective of aneuploidy in breast cancer cells. *Proc. Natl Acad. Sci. USA* **108**, 5384–5389 (2011).

29. Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).

METABRIC Group

Co-chairs Carlos Caldas[1,2], Samuel Aparicio[3,4]

Writing committee Christina Curtis[1,2]†, Sohrab P. Shah[3,4], Carlos Caldas[1,2], Samuel Aparicio[3,4]

Steering committee James D. Brenton[1,2], Ian Ellis[5], David Huntsman[3,4], Sarah Pinder[6], Arnie Purushotham[6], Leigh Murphy[7], Carlos Caldas[1,2], Samuel Aparicio[3,4]

Tissue and clinical data source sites: University of Cambridge/Cancer Research UK Cambridge Research Institute Carlos Caldas (Principal Investigator)[1,2]; Helen Bardwell[2], Suet-Feung Chin[1,2], Christina Curtis[1,2]†, Zhihao Ding[2], Stefan Gräf[1,2], Linda Jones[8], Bin Liu[1,2], Andy G. Lynch[1,2], Irene Papatheodorou[1,2], Stephen J. Sammut[9], Gordon Wishart[9]; **British Columbia Cancer Agency** Samuel Aparicio (Principal Investigator)[3,4], Steven Chia[4], Karen Gelmon[4], David Huntsman[3,4], Steven McKinney[3,4], Caroline Speers[4], Gulisa Turashvili[3,4], Peter Watson[3,4,7]; **University of Nottingham:** Ian Ellis (Principal Investigator)[5], Roger Blamey[5], Andrew Green[5], Douglas Macmillan[5], Emad Rakha[5]; **King's College London** Arnie Purushotham (Principal Investigator)[6], Cheryl Gillett[6], Anita Grigoriadis[6], Sarah Pinder[6], Emanuele de Rinaldis[6], Andy Tutt[6]; **Manitoba Institute of Cell Biology** Leigh Murphy (Principal Investigator)[7], Michelle Parisien[7], Sandra Troup[7]

Cancer genome/transcriptome characterization centres: University of Cambridge/Cancer Research UK Cambridge Research Institute Carlos Caldas (Principal Investigator)[1,2], Suet-Feung Chin (Team Leader)[1,2], Derek Chan[1], Claire Fielding[2], Ana-Teresa Maia[1,2], Sarah McGuire[2], Michelle Osborne[2], Sara M. Sayalero[2], Inmaculada Spiteri[2], James Hadfield[2]; **British Columbia Cancer Agency** Samuel Aparicio (Principal Investigator)[3,4], Gulisa Turashvili (Team Leader)[3,4], Lynda Bell[4], Katie Chow[4], Nadia Gale[4], David Huntsman **[3,4]**, Maria Kovalik[4], Ying Ng**[4]**, Leah Prentice[4]

Data analysis subgroup: University of Cambridge/Cancer Research UK Cambridge Research Institute Carlos Caldas (Principal Investigator)[1,2], Simon Tavaré (Principal Investigator)[1,2,10,11], Christina Curtis (Team Leader)[1,2]†, Mark J. Dunning[2], Stefan Gräf[1,2], Andy G. Lynch[1,2], Oscar M. Rueda[1,2], Roslin Russell[2], Shamith Samarajiwa[1,2], Doug Speed[2,10]; Florian Markowetz (Principal Investigator)[1,2], Yinyin Yuan[1,2]; James D. Brenton (Principal Investigator)[1,2]; **British Columbia Cancer Agency** Samuel Aparicio (Principal Investigator)[3,4], Sohrab P. Shah (Team Leader)[3,4], Ali Bashashati[3], Gavin Ha[3], Gholamreza Haffari[3] & Steven McKinney[3,4]

[1]Department of Oncology, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK. [2]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. [3]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. [4]Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia V5Z 1L3, Canada. [5]Department of Histopathology, School of Molecular Medical Sciences, University of Nottingham, Nottingham NG5 1PB, UK. [6]King's College London, Breakthrough Breast Cancer Research Unit, London, WC2R 2LS, UK. [7]Manitoba Institute of Cell Biology, University of Manitoba, Manitoba R3E 0V9, Canada. [8]Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK. [9]Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. [10]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Cambridge CB3 0WA, UK. [11]Molecular and Computational Biology Program, University of Southern California, Los Angeles, California 90089, USA. †Present address: Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA.

ARTICLE

Whole-genome analysis informs breast cancer response to aromatase inhibition

Matthew J. Ellis[1,2,3]*, Li Ding[4,5]*, Dong Shen[4,5]*, Jingqin Luo[3,6], Vera J. Suman[7], John W. Wallis[4,5], Brian A. Van Tine[1], Jeremy Hoog[1], Reece J. Goiffon[8,9,10], Theodore C. Goldstein[11], Sam Ng[11], Li Lin[1], Robert Crowder[1], Jacqueline Snider[1], Karla Ballman[7], Jason Weber[1,8,12], Ken Chen[13], Daniel C. Koboldt[4,5], Cyriac Kandoth[4,5], William S. Schierding[4,5], Joshua F. McMichael[4,5], Christopher A. Miller[4,5], Charles Lu[4,5], Christopher C. Harris[4,5], Michael D. McLellan[4,5], Michael C. Wendl[4,5], Katherine DeSchryver[1], D. Craig Allred[3,14], Laura Esserman[15], Gary Unzeitig[16], Julie Margenthaler[2], G. V. Babiera[13], P. Kelly Marcom[17], J. M. Guenther[18], Marilyn Leitch[19], Kelly Hunt[13], John Olson[17], Yu Tao[6], Christopher A. Maher[1,4], Lucinda L. Fulton[4,5], Robert S. Fulton[4,5], Michelle Harrison[4,5], Ben Oberkfell[4,5], Feiyu Du[4,5], Ryan Demeter[4,5], Tammi L. Vickery[4,5], Adnan Elhammali[8,9,10], Helen Piwnica-Worms[8,12,20,21], Sandra McDonald[2,22], Mark Watson[6,14,22], David J. Dooling[4,5], David Ota[23], Li-Wei Chang[3,14], Ron Bose[2,3], Timothy J. Ley[1,2,4], David Piwnica-Worms[8,9,10,12,24], Joshua M. Stuart[11], Richard K. Wilson[2,4,5] & Elaine R. Mardis[2,4,5]

To correlate the variable clinical features of oestrogen-receptor-positive breast cancer with somatic alterations, we studied pretreatment tumour biopsies accrued from patients in two studies of neoadjuvant aromatase inhibitor therapy by massively parallel sequencing and analysis. Eighteen significantly mutated genes were identified, including five genes (*RUNX1*, *CBFB*, *MYH9*, *MLL3* and *SF3B1*) previously linked to haematopoietic disorders. Mutant MAP3K1 was associated with luminal A status, low-grade histology and low proliferation rates, whereas mutant TP53 was associated with the opposite pattern. Moreover, mutant *GATA3* correlated with suppression of proliferation upon aromatase inhibitor treatment. Pathway analysis demonstrated that mutations in *MAP2K4*, a MAP3K1 substrate, produced similar perturbations as MAP3K1 loss. Distinct phenotypes in oestrogen-receptor-positive breast cancer are associated with specific patterns of somatic mutations that map into cellular pathways linked to tumour biology, but most recurrent mutations are relatively infrequent. Prospective clinical trials based on these findings will require comprehensive genome sequencing.

Oestrogen-receptor-positive breast cancer exhibits highly variable prognosis, histological growth patterns and treatment outcomes. Neoadjuvant aromatase inhibitor treatment trials provide an opportunity to document oestrogen-receptor-positive breast cancer phenotypes in a setting where sample acquisition is easy, prospective consent for genomic analysis can be obtained and responsiveness to oestrogen deprivation therapy is documented[1]. We therefore conducted massively parallel sequencing (MPS) on 77 samples accrued from two neoadjuvant aromatase inhibitor clinical trials[2,3]. Forty-six cases underwent whole-genome sequencing (WGS) and 31 cases underwent exome sequencing, followed by extensive analysis for somatic alterations and their association with aromatase inhibitor response. Case selection for discovery was based on the levels of the tumour proliferation marker Ki67 in the surgical specimen, because high cellular proliferation despite aromatase inhibitor treatment identifies poor prognosis tumours exhibiting oestrogen-independent growth[4] (Supplementary Fig. 1). Twenty-nine samples had Ki67 levels above 10% ('aromatase-inhibitor-resistant tumours', median Ki67 21%, range 10.3–80%) and

48 were at or below 10% ('aromatase-inhibitor-sensitive tumours', median Ki67 1.2%, range 0–8%). Cases were also classified as luminal A or B by gene expression profiling[3]. We subsequently examined interactions between Ki67 biomarker change, histological categories, intrinsic subtype and mutation status in selected recurrently mutated genes in 310 cases overall. Pathway analysis was applied to contrast the signalling perturbations in aromatase-inhibitor-sensitive versus aromatase-inhibitor-resistant tumours.

Results

The mutation landscape of luminal-type breast cancer

Using paired-end MPS, 46 tumour and normal genomes were sequenced to at least 30-fold and 25-fold haploid coverage, respectively, with diploid coverage of at least 95% based on concordance with SNP array data (Supplementary Table 1). Candidate somatic events were identified using multiple algorithms[5,6], and were then verified by hybridization capture-based validation that targeted all putative somatic single-nucleotide variants (SNVs) and small insertions/deletions (indels) that

[1]Department of Internal Medicine, Division of Oncology, Washington University, St Louis, Missouri, USA. [2]Siteman Cancer Center, Washington University, St Louis, Missouri 63110, USA. [3]Breast Cancer Program, Washington University, St Louis, Missouri 63110, USA. [4]The Genome Institute, Washington University, St Louis, Missouri 63108, USA. [5]Department of Genetics, Washington University, St Louis, Missouri 63108, USA. [6]Division of Biostatistics, Washington University, St Louis, Missouri 63110, USA. [7]ACOSOG Statistical Center, Mayo Clinic, Rochester, Minnesota 55905, USA. [8]BRIGHT Institute, Washington University School of Medicine, St Louis, Missouri 63110, USA. [9]Molecular Imaging Center, Washington University, St Louis, Missouri 63110, USA. [10]Malinckrodt Institute of Radiology, Washington University, St Louis, Missouri 63110, USA. [11]Department of Biomolecular Engineering, University of California, Santa Cruz, California 95064, USA. [12]Department of Cell Biology and Physiology, Washington University, St Louis, Missouri 63110, USA. [13]M. D. Anderson Cancer Center, Houston, Texas 77030, USA. [14]Department of Pathology and Immunology, Washington University, St Louis, Missouri 63110, USA. [15]Helen Diller Cancer Center, University of California, San Francisco, California 94143, USA. [16]Doctors Hospital of Laredo, Laredo, Texas 78045, USA. [17]Duke University Cancer Center, Durham, North Carolina 27705, USA. [18]Good Samaritan Hospital, Cincinnati, Ohio 45406, USA. [19]Simmons Cancer Center, University of Texas Southwestern, Dallas, Texas 75390, USA. [20]Department of Internal Medicine, Washington University, St Louis, Missouri 63110, USA. [21]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. [22]ACOSOG Central Specimen Bank, Washington University, St Louis, Missouri 63110, USA. [23]ACOSOG Operations Center, Duke University, Durham, North Carolina 27705, USA. [24]Department of Developmental Biology, Washington University, St Louis, Missouri 63110, USA.
*These authors contributed equally to this work.

overlap coding exons, splice sites and RNA genes (tier 1), high-confidence SNVs and indels in non-coding conserved or regulatory regions (tier 2), as well as non-repetitive regions of the human genome (tier 3). In addition, somatic structural variants and germline structural variants that potentially affect coding sequences (Supplementary Information) were assessed. Digital sequencing data from captured target DNAs from the 46 tumour and normal pairs (Supplementary Table 2 and Supplementary Information) confirmed 81,858 mutations (point mutations and indels) and 773 somatic structural variants. The average numbers of somatic mutations and structural variants were 1,780 (range 44–11,619) and 16.8 (range 0–178) per case, respectively (Supplementary Table 3). Tier 1 point mutations and small indels predicted for all 46 cases also were validated using both 454 and Illumina sequencing (Supplementary Information). BRC25 was a clear outlier with only 44 validated tiers 1–3 mutations, all at low allele frequencies (ranging from 5% to 26.8%). This sample probably had low tumour content despite histopathology assessment, but the data are included to avoid bias.

The overall mutation rate was 1.18 validated mutations per megabase (Mb) (tier 1: 1.05; tier 2: 1.14; tier 3: 1.20). The mutation rate for tier 1 was higher than that observed for acute myeloid leukaemia (0.18–0.23)[6,7], but lower than that reported for hepatocellular carcinoma (1.85)[8], malignant melanoma (6.65)[9] and lung cancers (3.05–8.93)[10,11] (Supplementary Table 4). The background mutation rate (BMR) across the 21 aromatase-inhibitor-resistant tumours was 1.62 per Mb, nearly twice that of the 25 aromatase-inhibitor-sensitive tumours at 0.824 per Mb ($P = 0.02$, one-sided t-test). A trend for more somatic structural variations in the aromatase-inhibitor-resistant group was also observed, as the validated somatic structural variation frequency in the 21 aromatase-inhibitor-resistant tumour genomes was 21.69 versus an average of 12.76 in 25 aromatase-inhibitor-sensitive tumours ($P = 0.16$, one-sided t-test) (Fig. 1). If ten $TP53$ mutated cases were excluded, the background mutation rate still tended to be higher in the aromatase-inhibitor-resistant group ($P = 0.08$). To demonstrate that a single-tumour core biopsy produced representative genomic data, whole-genome sequencing of two pre-treatment biopsies was conducted for 5 of the 46 cases. The frequency of mutations in the paired

specimens showed high concordance in all cases (correlation co-efficiency ranged from 0.74 to 0.95) (Supplementary Fig. 2) and a somatic mutation was infrequently detected in only one of the two samples (4.65% overall).

Significantly mutated genes in luminal breast cancer

The discovery effort was extended by studying 31 additional cases by exome sequencing, producing an additional 1,371 tier 1 mutations. In total the 77 cases yielded 3,355 tier 1 somatic mutations, including 3,208 point mutations, 1 dinucleotide mutation and 146 indels, ranging from 1 to 28 nucleotides. The point mutations included 733 silent, 2,145 missense, 178 nonsense, 6 read-through, 69 splice-site mutations and 77 in RNA genes (Supplementary Table 5). Of 2,145 missense mutations, 1,551 were predicted to be deleterious by SIFT[12] and/or PolyPhen[13]. The MuSiC package[45] was applied to determine the significance of the difference between observed versus expected mutation events in each gene, on the basis of the background mutation rate. This identified 18 significantly mutated genes with a convolution false discovery rate (FDR) < 0.26 (Table 1 and Supplementary Table 6). The list contains genes previously identified as mutated in breast cancer ($PIK3CA$[14], $TP53$[15], $GATA3$[12], $CDH1$[13], $RB1$[16], $MLL3$[17], $MAP3K1$[18] and $CDKN1B$[19]) as well as genes not previously observed in clinical breast cancer samples, including $TBX3$, $RUNX1$, $LDLRAP1$, $STNM2$, $MYH9$, $AGTR2$, $STMN2$, $SF3B1$ and $CBFB$.

Thirteen mutations (3 nonsense, 6 frame-shift indels, 2 in-frame deletions and 2 missense) were identified in $MAP3K1$ (Table 1 and Fig. 2), a serine/threonine kinase that activates the ERK and JNK kinase pathways through phosphorylation of $MAP2K1$ and $MAP2K4$ (ref. 20). Of interest, a missense (S184L) and a splice-region mutation (e2+3 probably affecting splicing) in $MAP2K4$ were observed in two tumours with no $MAP3K1$ mutation (Fig. 2). Single nonsynonymous mutations in $MAP3K12$, $MAP3K4$, $MAP4K3$, $MAP4K4$, $MAPK15$ and $MAPK3$ were also detected (Supplementary Table 5). $TBX3$ harboured three small indels (one insertion and two deletions). $TBX3$ affects expansion of breast cancer stem-like cells through regulation of FGFR[21]. Two truncating mutations in the tumour suppressor $CDKN1B$ were
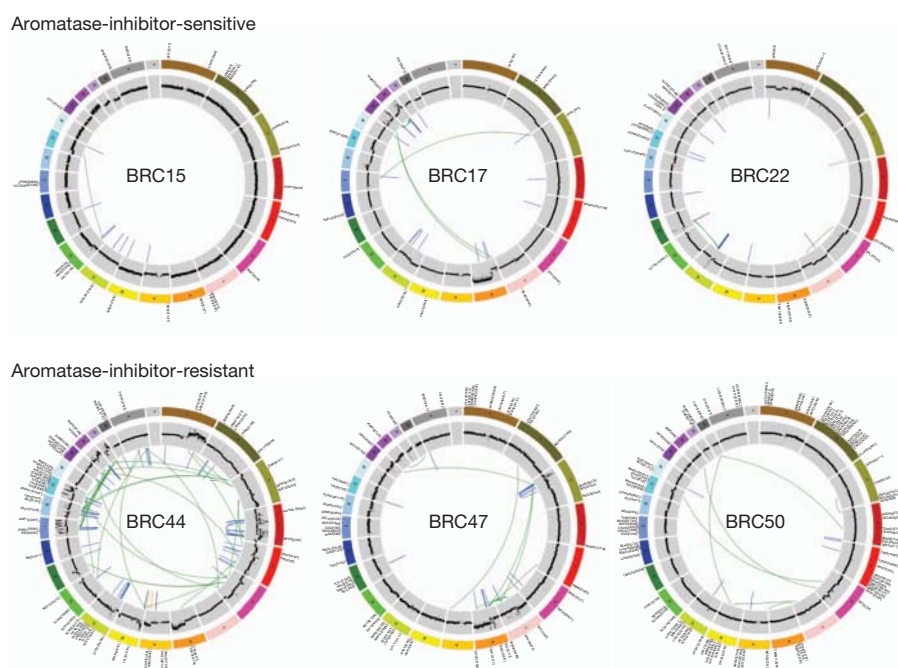
Figure 1 | Genome-wide somatic mutations. Circos plots[44] indicate validated somatic mutations comprising tier 1 point mutations and indels, genome-wide copy number alterations, and structural rearrangements in six representative genomes. Three on-treatment Ki67 less than or at 10% (top panel: BRC15,

BRC17 and BRC22) and three on-treatment Ki67 greater than 10% (bottom panel: BRC44, BRC47 and BRC50) cases are shown. Significantly mutated genes are highlighted in red. No purity-based copy number corrections were used for plotting copy number.

Table 1 | Significantly mutated genes identified in 46 whole genomes and 31 exomes sequenced in luminal breast cancer patients

Gene	Total	MS	NS	Indel	SS	P value	FDR
MAP3K1	13	2	3	8	0	0	0
PIK3CA	45	44	0	1	0	0	0
TP53	18	13	1	1	1	0	0
GATA3	8	1	0	4	3	1.15×10^{-19}	7.41×10^{-16}
CDH1	8	1	1	5	1	3.07×10^{-15}	1.59×10^{-11}
TBX3	3	0	0	3	0	2.58×10^{-6}	0.011
ATR	6	6	0	0	0	3.73×10^{-6}	0.014
RUNX1	4	4	0	0	0	6.59×10^{-6}	0.021
ENSG00000212670*	2	2	0	0	0	2.31×10^{-5}	0.066
RB1	4	2	1	0	1	2.76×10^{-5}	0.071
LDLRAP1	2	1	1	0	0	4.27×10^{-5}	0.092
STMN2	2	1	0	1	0	4.15×10^{-5}	0.092
MYH9	4	1	1	2	0	8.96×10^{-5}	0.178
MLL3	5	1	1	3	0	1.04×10^{-4}	0.191
CDKN1B	2	0	1	1	0	1.39×10^{-4}	0.240
AGTR2	2	2	0	0	0	1.71×10^{-4}	0.256
SF3B1	3	3	0	0	0	1.79×10^{-4}	0.256
CBFB	2	1	1	0	0	1.70×10^{-4}	0.256

* ENSG00000212670 is not in RefSeq release 50.
MS, Missense; NS, nonsense; SS, splice site.

identified[19]. Four missense *RUNX1* mutations were observed, with three in the RUNT domain clustered within the 8 amino acid putative ATP-binding site (R166Q, G168E and R169K). *RUNX1* is a transcription factor affected by mutation and translocation in the M2 subtype of acute myeloid leukaemia[22] and is implicated in tethering the oestrogen receptor to promoters independently of oestrogen response elements[23]. Two mutations (N104S and N140*) were also identified in *CBFB*, the binding partner of *RUNX1*. Additional mutations included 3 missense (2 K700E and 1 K666Q), in *SF3B1*, a splicing factor implicated in myelodysplasia[24] and chronic lymphocytic leukaemia[25]. One missense mutation, one nonsense mutation and two indels were found in the *MYH9* gene, involved in hereditary macrothrombocytopenia[26] as well as being observed in an ALK translocation in anaplastic large cell lymphoma[27].

We also identified three significantly mutated genes (*LDLRAP1*, *AGTR2* and *STMN2*) not previously implicated in cancer. A missense and a nonsense mutation were observed in *LDLRAP1*, a gene associated with familial hypercholesterolaemia[28]. *AGTR2*, angiotensin II receptor type 2, harboured two missense mutations (V184I and R251H). Angiotensin signalling and oestrogen receptor intersect in models of tissue fibrosis[29]. *STMN2*, a gene activated by JNK family kinases[30,31] and therefore regulated by *MAP3K1* and *MAP2K4*, harboured one frameshift deletion and one missense mutation. Three deletions and one point mutation (Supplementary Fig. 3) were identified in a large, infrequently spliced non-coding (lnc) RNA gene, *MALAT1* (metastasis associated lung adenocarcinoma transcript 1),

that regulates alternative splicing by modulating the phosphorylation of SR splicing factor[32]. Translocations and point mutations of *MALAT1* have been reported in sarcoma[33] and colorectal cancer cell lines[34]. Five additional MALAT1 mutations were found in the recurrent screening set (Supplementary Table 5d). The locations of these mutations clustered in a region of species homology (F1 and 2 domains) that could mediate interactions with SRSF1 (ref. 32, Supplementary Fig. 4). Non-coding mutation clusters were found in *ATR*, *GPR126* and *NRG3* (Supplementary Information and Supplementary Table 7).

Correlating mutations with clinical data

To study clinical correlations, mutation recurrence screening was conducted on an additional 240 cases (Supplementary Table 8 and Supplementary Fig. 1). By combining WGS, exome and recurrence screening data, we determined the mutation frequency in *PIK3CA* to be 41.3% (131 of 317 tumours) (Supplementary Table 5a–d and Supplementary Fig. 3). *TP53* was mutated in 51 of 317 tumours (16.1%) (Supplementary Table 5a–d and Supplementary Fig. 3). Additionally, 52 nonsynonymous *MAP3K1* mutations in 39 tumours and 10 mutations in its substrate *MAP2K4* were observed, representing a combined case frequency of 15.5% (Supplementary Table 5a–d and Fig. 3). Of note, 52 of the 62 non-silent mutations in *MAP3K1* and *MAP2K4* were scattered indels or other protein-truncating events strongly suggesting functional inactivation. In addition, 13 tumours harboured two non-silent *MAP3K1* mutations, indicative of bi-allelic loss and reinforcing the conclusion that this gene is a tumour suppressor. Twenty nine tumours harboured a total of 30 mutations in *GATA3*, consisting of 25 truncation events, one in-frame insertion, and 4 missense mutations including 3 recurrent mutations at M294K (Supplementary Table 5a–d and Supplementary Fig. 3). BRC8 harboured a chromosome 10 deletion that includes *GATA3*. *CDH1* mutation data were available for 169 samples and, as expected, its mutation status was strongly associated with lobular breast cancer[13] (Table 2a). We applied a permutation-based approach in MuSiC[45] to ascertain relationships between mutated genes. Negative correlations were found between mutations in gene pairs such as *GATA3* and *PIK3CA* ($P = 0.0026$), *CDH1* and *GATA3* ($P = 0.015$), and *CDH1* and *TP53* ($P = 0.022$). *MAP3K1* and *MAP2K4* mutations were mutually exclusive, albeit without reaching statistical significance ($P = 0.3$). In contrast, a positive correlation between *MAP3K1/MAP2K4* and *PIK3CA* mutations was highly significant ($P = 0.0002$) (Supplementary Table 9).

Two independent mutation data sets, designated 'Set 1' (discovery cohort) and 'Set 2' (validation cohort), from these clinical trial samples were analysed separately and then in combination, with a false discovery rate (FDR)-corrected *P* value to gauge the overall strength and
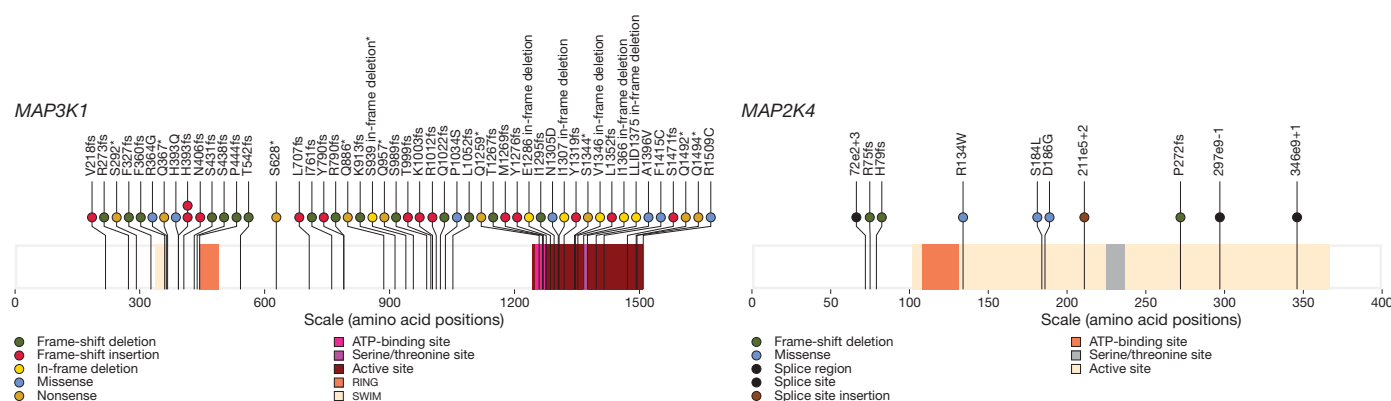
Figure 2 | *MAP3K1* and *MAP2K4* mutations observed in 317 samples. Somatic status of all mutations was obtained by Sanger sequencing of PCR products or Illumina sequencing of targeted capture products. The locations of conserved protein domains are highlighted. Each nonsynonymous substitution, splice site mutation or indel is designated with a circle at the representative protein position with colour to indicate translation effects of the mutation. Asterisk, nonsense mutations that cause truncation of the open reading frame.

Deletion involving *MAP3K1*

Deletion involving *MAP2K4*

Deletion involving *NRG1*
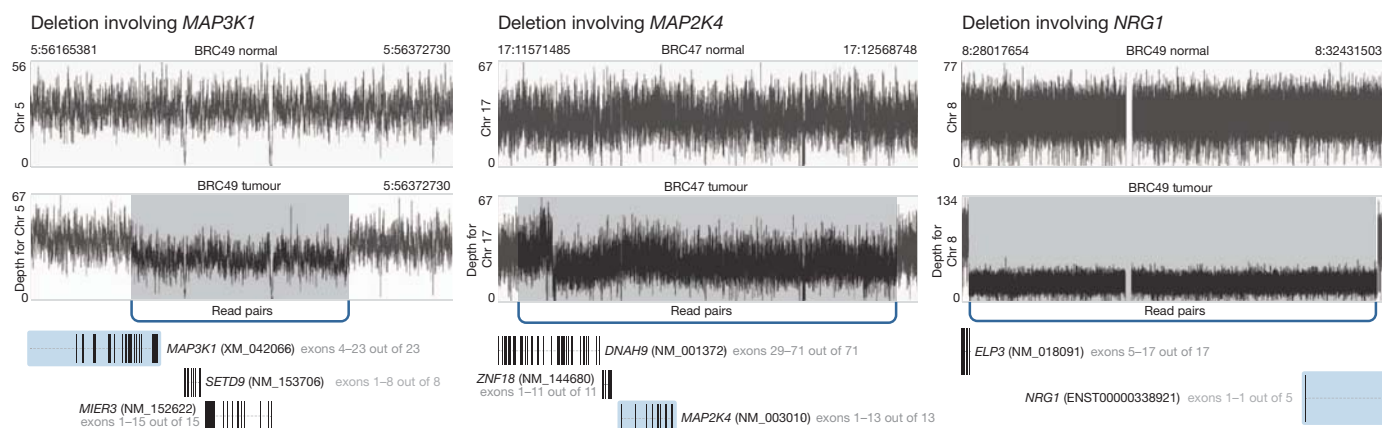
Figure 3 | Structural variants in significantly mutated or frequently deleted genes. One *MAP3K1* deletion in BRC49 and one *MAP2K4* deletion in BRC47, and one *ELP3-NRG1* fusion in BRC49 identified using Illumina paired-end reads from whole-genome sequence data. Arcs represent multiple breakpoint-spanning read pairs with sequence coverage depth plotted in black across the region. Chr, chromosome.

consistency of genotype–phenotype relationships (Table 2a, b and Supplementary Fig. 1). *TP53* mutations in both data sets correlated with significantly higher Ki67 levels, both at baseline ($P = 0.0003$) and at surgery ($P = 0.001$). Furthermore, *TP53* mutations were significantly enriched in luminal B tumours ($P = 0.04$) and in higher histological grade tumours ($P = 0.02$). In contrast, *MAP3K1* mutations were more frequent in luminal A tumours ($P = 0.02$), in grade 1 tumours ($P = 0.005$) and in tumours with lower Ki67 at baseline ($P = 0.001$) with consistent findings across both data sets. *GATA3* mutation did not influence baseline Ki67 levels but was enriched in samples exhibiting greater percentage Ki67 decline ($P = 0.01$). This finding requires further verification because it was significant in Set 1 (uncorrected *P* value 0.003) but was a marginal finding in Set 2 ($P = 0.08$). However, it suggests *GATA3* mutation may be a positive predictive marker for aromatase inhibitor response.

Structural variation and DNA repair mechanisms

Analysis of copy number alterations (CNAs) revealed arm-level gains for 1q, 5p, 8q, 16p, 17q, 20p and 20q and arm-level losses for 1p, 8p, 16q, and 17p in the 46 WGS tumour genomes (Supplementary Fig. 5). A total of 773 structural variants (579 deletions, 189 translocations and 5 inversions) identified by WGS were validated as somatic in 46 breast cancer genomes by capture validation. No recurrent translocations were detected but six in-frame fusion genes were validated by reverse transcription followed by PCR (Supplementary Information and Supplementary Tables 10–13). Seven tumours had multiple complex translocations with breakpoints suggestive of a catastrophic mitotic event ('chromothripsis'; Supplementary Table 11). Analysis of the structural variant genomic breakpoints shows the spectra of putative chromothripsis-related events are the same as seen for other somatic events, with the majority of structural variants arising from non-homologous end-joining. We classified somatic (mitotic) and germline (meiotic) structural variants into four groups: variable number tandem repeat (VNTR), non-allelic homologous recombination (NAHR), microhomology-mediated end joining (MMEJ), and non-homologous end joining (NHEJ), according to criteria described in Supplementary Information. The fraction of each classification is shown for germline and somatic (mitotic) events (Supplementary Table 14). There were significantly more somatic NHEJ events in tumour genomes than the other three types ($P < 2.2 \times 10^{-16}$).

Pathways relevant to aromatase inhibitor response

Pathscan[35] analysis (Supplementary Table 15 and Supplementary Information) indicated that somatic mutations detected in the 77 discovery cases affect a number of pathways, including caspase cascade/apoptosis, ErbB signalling, Akt/PI3K/mTOR signalling, TP53/RB signalling and MAPK/JNK pathways (Fig. 4a). To discern the pathways relevant to aromatase inhibitor sensitivity, we conducted separate pathway analyses for aromatase-inhibitor-sensitive versus aromatase-inhibitor-resistant tumours. Whereas the majority of top altered pathways (FDR ≤ 0.15) in each group are shared, several pathways were enriched in the aromatase-inhibitor-resistant group, including the TP53 signalling pathway, DNA replication, and mismatch repair. Specifically, 38% of the aromatase-inhibitor-resistant group (11 of 29 tumours) have mutations in the TP53 pathway with three having double or triple hits involving *TP53*, *ATR*, *APAF1* or *THBS1*. In contrast, only 16.6% (8 of 48 tumours) of the Ki67 low group had mutations in the TP53 signalling pathway, each with only a single hit in genes *TP53*, *ATR*, *CCNE2* or *IGF1*. (Supplementary Table 16).

GeneGo pathway analysis of MetaCore interacting network objects was used to identify genes in the 77 luminal breast cancers with low-frequency mutations that cluster into pathway maps. Eight networks assembled from significant maps encompassed mutations from 71 (92%) of the tumours (Fig. 4b). Many of the network objects shared pathways with significantly mutated genes such as *TP53*, *MAP3K1*, *PIK3CA* and *CDH1*. GeneGo analysis also revealed that several genes with low-frequency mutations were actually subunits of complexes, resulting in higher mutation rates for that object, for example, the condensin complex (4 mutations in 4 genes) and the MRN complex (4 mutations in 3 genes). Several pathways without multiple significantly mutated genes, such as the apoptotic cascade, calcium/phospholipase signalling and G-protein-coupled receptors, were significantly affected by low-frequency mutations. Grouping tumours by significantly mutated genes and pathway mutation status showed that whereas 55 (71%) of the tumours contained significantly mutated genes in significant pathways, an additional 16 (21%) contained only non-significantly mutated genes in these pathways. Thus, tumours without a given significantly mutated gene often had other mutations in the same relevant pathway (Fig. 4b, Supplementary Fig. 6, Supplementary Table 17 and Supplementary Information).

We also applied PARADIGM[36] to infer pathway-informed gene activities using gene expression and copy-number data to identify several 'hubs' of activity (Supplementary Fig. 7, Supplementary Fig. 8 and Supplementary Information). As expected, *ESR1* and *FOXA1* were among the hubs activated cohort-wide while other hubs exhibited high but differential changes in aromatase-inhibitor-resistant tumours including *MYC*, *FOXM1* and *MYB* (Supplementary Fig. 8). The concordance among the 104 MetaCore maps from GeneGo analysis described above is significant, with 75 (72%) matching one of the
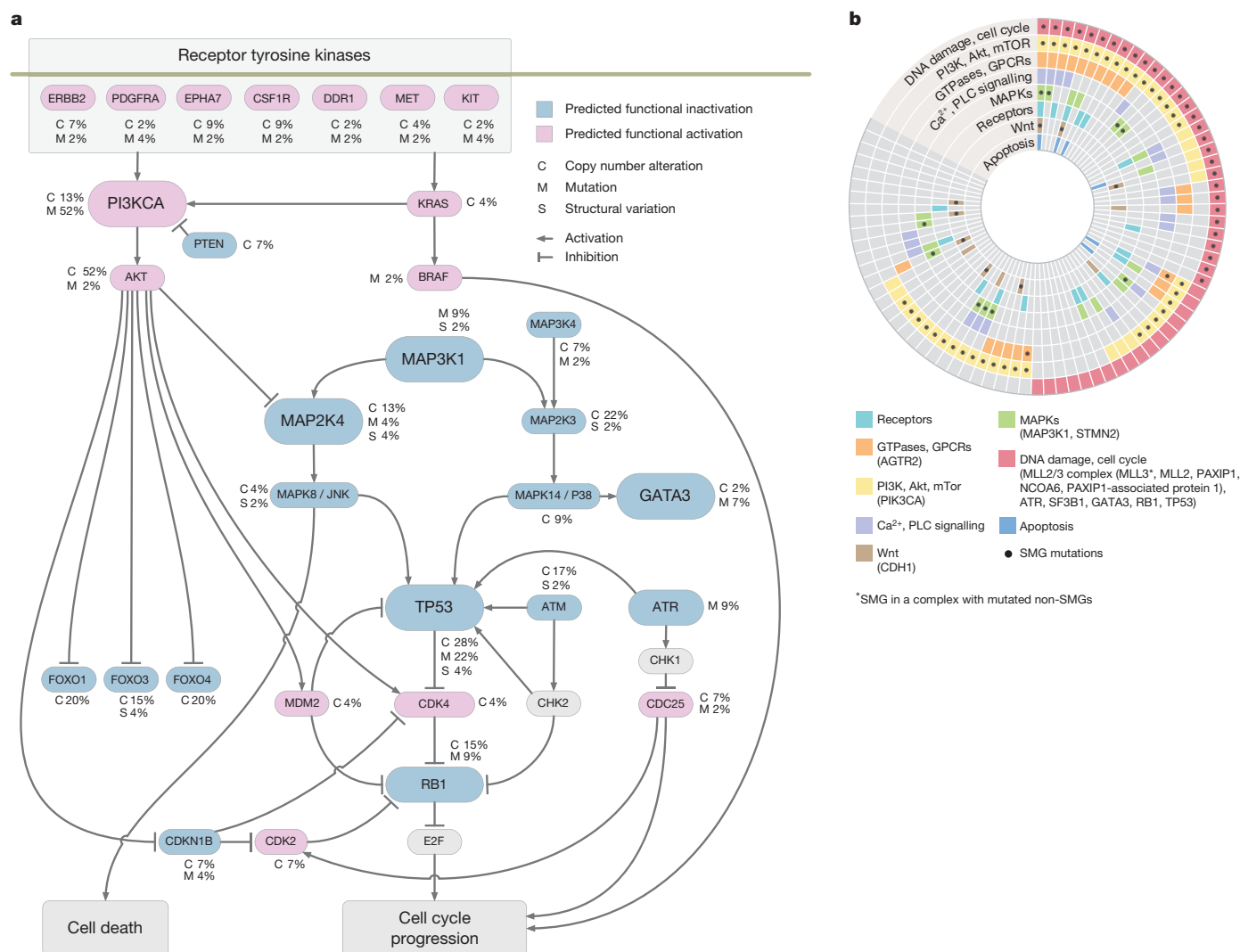
Figure 4 | Key cancer pathway components altered in luminal breast tumours. **a**, Only genetic alterations identified in 46 WGS cases are shown. Alterations were discovered in key genes in the *TP53/RB*, *MAPK*, *PI3K/AKT/mTOR* pathways. Genes coloured blue and red are predicted to be functionally inactivated and activated, respectively, through focused mutations including point mutations and small indels (M), copy number deletions (C), or other structural changes (S) that affect the gene. The inter-connectedness of this network (several pathways) shows that there are many different ways to perturb a pathway. **b**, Eight interaction networks from canonical maps are significantly over-represented by mutations in 77 luminal breast tumours (46 WGS and 31 exome cases). In the concentric circle diagram, tumours are arranged as radial spokes and categorized by their mutation status in each network (concentric ring colour) and significantly mutated gene mutation status (black dots). Tumour classification by pathway analysis shows many tumours unaffected by a given significantly mutated gene often harbour other mutations in the same network. For full annotation, see Supplementary Information and Supplementary Fig. 6. PLC, phospholipase C; SMG, significantly mutated gene.

PARADIGM subnetworks at the 0.05 significance level after multiple test correction ($P < 4.4 \times 10^{-6}$; Bonferroni-adjusted hypergeometric test) (Supplementary Fig. 9). We identified significant subnetworks associated with Ki67 biomarker status (Supplementary Fig. 10 and Supplementary Information) involving transcription factors controlling large regulons.

The PARADIGM-inferred pathway signatures were further used to derive a map of the genetic mechanisms that may underlie treatment response. A subnetwork was constructed in which interactions were retained only if they connected two features with higher than average absolute association with Ki67 biomarker status (Supplementary Figs 10 and 11 and Supplementary Information). Consistent with the PathScan results, among the largest of the hubs in the identified network were a central DNA damage hub with the second highest connectivity (55 regulatory interactions; 1% of the network) and *TP53* with the 14th highest connectivity (26 connections; 0.5% of the network). Additional highly connected hubs identified in order of connectivity were *MYC* with 79 connections (1.4%), *FYN* with 45 (0.8%), *MAPK3*

with 43, *JUN* with 40, *HDAC1* with 40, *SHC1* with 39, and *HIF1A/ARNT* complex with 39 (Supplementary Fig. 11).

To identify higher-level connections between mutations and clinical features, we compared the samples on the basis of pathway-derived signatures. For each clinical attribute and each significantly mutated gene, we dichotomized the discovery samples into a positive and a negative group to derive pathway signatures that discriminated between the groups (see details in Supplementary Information). We then computed all pair-wise Pearson correlations between pathway signatures and clustered the resulting correlations (Fig. 5). The entire process was repeated using validated mutations and signatures derived from the validation set (Supplementary Fig. 12). In line with expectation, *PIK3CA*, *MAP3K1*, *MAP2K4*, and low risk preoperative endocrine prognostic index (PEPI) scores (PEPI is an index of recurrence risk post neoadjuvant aromatase inhibitor therapy[4]) cluster with the luminal A subtypes and with each other, and are supported by the validation set analysis. The luminal B-like signatures included *TP53*, *RB1*, *RUNX1* and *MALAT1*, which also associated
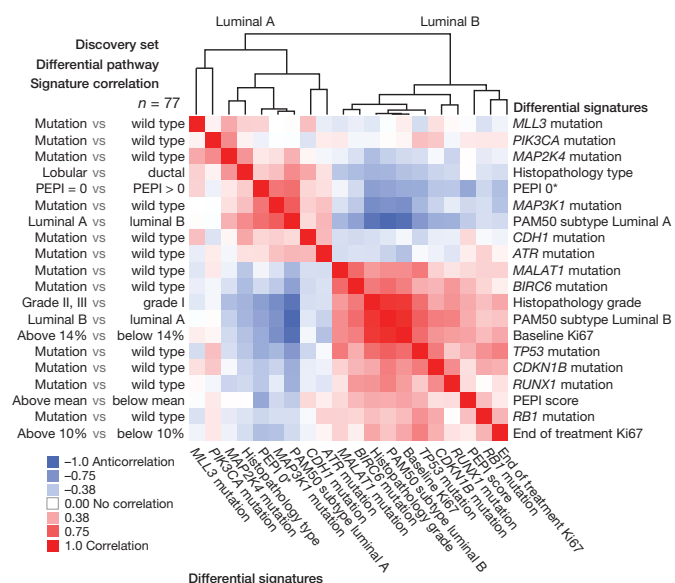
Figure 5 | Pathway signatures reveal connections between mutations and clinical outcomes. PARADIGM-based pathway signatures were derived for tumour feature dichotomies including mutation driven gene signatures (mutant versus non-mutant), histopathology type (lobular versus ductal), preoperative endocrine prognostic index (PEPI) score (PEPI = 0 favourable versus PEPI >0 unfavourable), PAM50 (50-gene intrinsic breast cancer subtype classifier) luminal A subtype (luminal A versus luminal B) and the reverse (luminal B versus luminal A), histopathology grade (grades II and III versus I), baseline Ki67 levels ($\geq$ 14% versus < 14%), and end-of-treatment Ki67 levels ($\geq$ 10% versus < 10%) and overall PEPI score (higher than mean unfavourable versus lower than mean favourable). Pearson correlations were computed between all pair-wise signatures; positive correlations, red; negative correlations, blue; column features ordered identically as rows. Correlation analysis on the 77 samples in the discovery set is shown. Asterisk: Ki67 < 2.7%, oestrogen-receptor-positive, node negative and tumour size $\leq$ 5 cm.

with other poor outcome features such as high baseline and surgical Ki67 levels, high grade histology and high PEPI scores. The *TP53* and *MALAT1* associations in the discovery set also were supported by the validation set analysis.

Druggable gene analysis

We defined mutations in druggable tyrosine kinase domains including in *ERBB2* (a V777L and a 755–759[LRENT] in-frame deletion homologous to gefitinib-sensitizing *EGFR* mutations in lung cancer[37]), as well as in *DDR1* (A829V, R611C), *DDR2* (E583D), *CSF1R* (D735H, M875L), and *PDGFRA* (E924K). In addition, pleckstrin homology domain mutations were observed in *AKT1* (C77F) and *AKT2* (S11F) and a kinase domain mutation was identified in *RPS6KB1* (S375F) (Supplementary Table 18).

Discussion

The low frequency of many significantly mutated genes presents an enormous challenge for correlative analysis, but several statistically significant patterns were identified, including the relationship between *MAP3K1* mutation, luminal A subtype, low tumour grade and low Ki67 proliferation index. On this basis, for patients with *MAP3K1* mutant luminal tumours, neoadjuvant aromatase inhibitor could provide a favourable option. In contrast, tumours with *TP53* mutations, which are mostly aromatase inhibitor resistant, would be more appropriately treated with other modalities. MAP3K1 activates the ERK family, thus, loss of ERK signalling could explain the indolent nature of MAP3K1-deficient tumours[20]. However, MAP3K1 also activates JNK through MAP2K4, which also can be mutated[38]. Loss of JNK signalling produces a defect in apoptosis in response to stress, which would hypothetically explain why these mutations accumulate[39,40]. *PIK3CA* harboured the most mutations (41.3%) but was neither associated with clinical nor Ki67 response, confirming our earlier report[41]. However, the positive association between *MAP3K1*/*MAP2K4* mutations and *PIK3CA* mutation at both the mutation and pathway levels suggests cooperativity (Fig. 4a).

The finding of multiple significantly mutated genes linked previously to benign and malignant haematopoietic disorders suggests that breast cancer, like leukaemia, can be viewed as a stem-cell disorder

Table 2 | Correlations between mutations and clinical features

a Luminal subtype and histology grade

Gene	Expression/histo-pathology variable	Mutation frequency*	Set1 P†	Set2 P†	Whole set FDR P‡
TP53	Luminal subtype A	9.3% (13/140)	0.001	0.46	0.041
	Luminal subtype B	21.5% (38/177)			
TP53	Histological grade I	4.5% (3/66)	0.05	0.067	0.02
	Histological grade II/III	19.2% (48/250)			
MAP3K1	Luminal subtype A	20.0% (28/140)	0.018	0.028	0.005
	Luminal subtype B	6.2% (11/177)			
MAP3K1	Histological grade I	25.8% (17/66)	0.061	0.011	0.005
	Histological grade II/III	8.8% (22/250)			
CDH1	Histological type ductal	5.9% (10/169)	0.41§	2.8×10^{-11}	3.9×10^{-10}
	Histological type lobular	50.0% (20/40)			

b Mutation and Ki67 index

Gene	Ki67 variable	Wild type mean‖	Mutant mean‖	Set1 P¶	Set2 P¶	Whole set FDR P‡
TP53	Baseline	13.1	25.1	3.7×10^{-5}	0.012	0.0003
	Surgery	1.4	4	0.0002	0.014	0.001
	% change	−89.2	−84.3	0.09	0.28	0.24
MAP3K1	Baseline	15.8	8.1	0.049	0.001	0.002
	Surgery	1.86	0.75	0.11	0.1	0.05
	% change	−88.3	−90.5	0.49	0.65	0.55
GATA3	Baseline	14.8	11.5	0.13	0.95	0.56
	Surgery	1.95	0.38	0.001	0.23	0.012
	% change	−86.8	−96.9	0.003	0.08	0.012

* Mutation percentage (mutant cases/total cases in a category), counts are based on all cases (Set 1 and Set 2 combined).
† Unadjusted P value from Fisher's exact test or Chi-square test as appropriate.
‡ Benjamini–Hochberg false discovery rate (FDR)-adjusted P value using all cases (Set1 and Set 2 combined).
§ Only 77 cases in Set1 had CDH1 sequencing results.
‖ Geometric means are based on all cases (Set 1 and Set 2 combined).
¶ Unadjusted P value from Wilcoxon rank sum test.

that produces indolent or aggressive tumours that display varying phenotypes depending on differentiation blocks generated by different mutation repertoires[42]. Whereas only *MLL3* showed statistical significance in the analysis of 46 WGS cases, multiple mutations in genes related to histone modification and chromatin remodelling are worth noting (Supplementary Table 19). An array of coding mutations and structural variations was discovered in methyltransferases (*MLL2*, *MLL3*, *MLL4* and *MLL5*), demethyltransferases (*KDM6A*, *KDM4A*, *KDM5B* and *KDM5C*), and acetyltransferases (*MYST1*, *MYST3* and *MYST4*). Furthermore, our analysis identified several adenine-thymine (AT)-rich interactive domain-containing protein genes (*ARID1A*, *ARID2*, *ARID3B* and *ARID4B*) that harboured mutations and large deletions, reinforcing the role of members from the SNF/SWI family in breast cancer.

Pathway analysis enables the evaluation of mutations with low recurrence frequency where statistical comparisons are conventionally underpowered. For example, the eight samples with *MAP2K4* mutations were sufficient to derive a reliable pathway-based gene signature in PARADIGM that aligns with *MAP3K1*. This approach also pointed to a putative connection between *MALAT1* and the *TP53* pathway. Finally, we provide evidence that transcriptional associations to Ki67 response reside in a connected network under the control of several key 'hub' genes including *MYC*, *FYN* and *MAP* kinases, among others. Targeting these hubs in resistant tumours could produce therapeutic advances. In conclusion, the genomic information derived from unbiased sequencing is a logical new starting point for clinical investigation, where the mutation status of an individual patient is determined in advance and treatment decisions are driven by therapeutic hypotheses that stem from knowledge of the genomic sequence and its possible consequences. However, the accrual of large numbers of patients and the use of comprehensive sequencing and gene expression approaches will be required because of the extreme genomic heterogeneity documented by this investigation.

METHODS SUMMARY

Clinical trial samples were accessed from the preoperative letrozole phase 2 study (NCT00084396)[2] that investigated the effect of letrozole for 16 to 24 weeks on surgical outcomes and from the American College of Surgeons Oncology Group (ACOSOG) Z1031 study (NCT00265759)[3] that compared anastrozole with exemestane or letrozole for 16 to 18 weeks before surgery (REMARK flow charts, Supplementary Fig. 1). Baseline snap-frozen biopsy samples with greater than 70% tumour content (by nuclei) underwent DNA extraction and were paired with a peripheral blood DNA sample. Two formalin-fixed biopsies were obtained at baseline and at surgery, and were used to conduct oestrogen receptor and Ki67 immunohistochemistry as previously published[4]. Paired end Illumina reads from tumours and normal samples were aligned to NCBI build36 using BWA. Somatic point mutations were identified using SomaticSniper[43], and indels were identified by combining results from a modified version of the Samtools indel caller (http://samtools.sourceforge.net/), GATK and Pindel. Structural variations were identified using BreakDancer[5] and SquareDancer (unpublished). All putative somatic events found in 46 cases were validated by targeted custom capture arrays (Nimblegen)/Illumina sequencing and all tier 1 mutations for 46 WGS cases also were validated using PCR/454 sequencing. All statistical analyses, including significantly mutated gene, mutation relation and clinical correlation were done using the MuSiC package[45] and/or by standard statistical tests (Supplementary Information). Pathway analysis was performed with PathScan, GeneGo Metacore (http://www.genego.com/metacore.php) and PARADIGM. A complete description of the materials and methods used to generate this data set and results is provided in the Supplementary Methods section.

1. Chia, Y. H., Ellis, M. J. & Ma, C. X. Neoadjuvant endocrine therapy in primary breast cancer: indications and use as a research tool. *Br. J. Cancer* **103**, 759–764 (2010).
2. Olson, J. A. Jr *et al.* Improved surgical outcomes for breast cancer patients receiving neoadjuvant aromatase inhibitor therapy: results from a multicenter phase II trial. *J. Am. Coll. Surg.* **208**, 906–914; discussion 915–906 (2009).
3. Ellis, M. J. *et al.* Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor-rich stage 2 to 3 breast cancer: clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype—ACOSOG Z1031. *J. Clin. Oncol.* **29**, 2342–2349 (2011).
4. Ellis, M. J. *et al.* Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics. *J. Natl. Cancer Inst.* **100**, 1380–1388 (2008).
5. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods* **6**, 677–681 (2009).
6. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
7. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
8. Totoki, Y. *et al.* High-resolution characterization of a hepatocellular carcinoma genome. *Nature Genet.* **43**, 464–469 (2011).
9. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
10. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
11. Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473–477 (2010).
12. Usary, J. *et al.* Mutation of *GATA3* in human breast tumors. *Oncogene* **23**, 7669–7678 (2004).
13. Berx, G. *et al.* E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J.* **14**, 6107–6115 (1995).
14. Samuels, Y. *et al.* High frequency of mutations of the *PIK3CA* gene in human cancers. *Science* **304**, 554 (2004).
15. Prosser, J., Thompson, A. M., Cranston, G. & Evans, H. J. Evidence that p53 behaves as a tumour suppressor gene in sporadic breast tumours. *Oncogene* **5**, 1573–1579 (1990).
16. T'Ang, A., Varley, J. M., Chakraborty, S., Murphree, A. L. & Fung, Y. K. Structural rearrangement of the retinoblastoma gene in human breast carcinoma. *Science* **242**, 263–266 (1988).
17. Wang, X. X. *et al.* Somatic mutations of the mixed-lineage leukemia 3 (*MLL3*) gene in primary breast cancers. *Pathol. Oncol. Res.* **17**, 429–433 (2011).
18. Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466**, 869–873 (2010).
19. Spirin, K. S. *et al.* p27/Kip1 mutation found in breast cancer. *Cancer Res.* **56**, 2400–2404 (1996).
20. Fanger, G. R., Johnson, N. L. & Johnson, G. L. MEK kinases are regulated by EGF and selectively interact with Rac/Cdc42. *EMBO J.* **16**, 4961–4972 (1997).
21. Fillmore, C. M. *et al.* Estrogen expands breast cancer stem-like cells through paracrine FGF/Tbx3 signaling. *Proc. Natl Acad. Sci. USA* **107**, 21737–21742 (2010).
22. Mao, S., Frank, R. C., Zhang, J., Miyazaki, Y. & Nimer, S. D. Functional and physical interactions between AML1 proteins and an ETS protein, MEF: implications for the pathogenesis of t(8;21)-positive leukemias. *Mol. Cell. Biol.* **19**, 3635–3644 (1999).
23. Stender, J. D. *et al.* Genome-wide analysis of estrogen receptor α DNA binding and tethering mechanisms identifies Runx1 as a novel tethering factor in receptor-mediated transcriptional activation. *Mol. Cell. Biol.* **30**, 3943–3955 (2010).
24. Papaemmanuil, E. *et al.* Somatic *SF3B1* mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* **365**, 1384–1395 (2011).
25. Wang, L. *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506 (2011).
26. Chen, Z. *et al.* The May-Hegglin anomaly gene *MYH9* is a negative regulator of platelet biogenesis modulated by the Rho-ROCK pathway. *Blood* **110**, 171–179 (2007).
27. Lamant, L. *et al.* Non-muscle myosin heavy chain (*MYH9*): a new partner fused to *ALK* in anaplastic large cell lymphoma. *Genes Chromosom. Cancer* **37**, 427–432 (2003).
28. Wilund, K. R. *et al.* Molecular mechanisms of autosomal recessive hypercholesterolemia. *Hum. Mol. Genet.* **11**, 3019–3030 (2002).
29. Dellê, H. *et al.* Antifibrotic effect of tamoxifen in a model of progressive renal disease. *J. Am. Soc. Nephrol.* **23**, 37–48 (2012).
30. Tararuk, T. *et al.* JNK1 phosphorylation of SCG10 determines microtubule dynamics and axodendritic length. *J. Cell Biol.* **173**, 265–277 (2006).
31. Westerlund, N. *et al.* Phosphorylation of SCG10/stathmin-2 determines multipolar stage exit and neuronal migration rate. *Nature Neurosci.* **14**, 305–313 (2011).
32. Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925–938 (2010).
33. Rajaram, V., Knezevich, S., Bove, K. E., Perry, A. & Pfeifer, J. D. DNA sequence of the translocation breakpoints in undifferentiated embryonal sarcoma arising in mesenchymal hamartoma of the liver harboring the t(11;19)(q11;q13.4) translocation. *Genes Chromosom. Cancer* **46**, 508–513 (2007).
34. Xu, C., Yang, M., Tian, J., Wang, X. & Li, Z. MALAT-1: a long non-coding RNA and its important 3′ end functional motif in colorectal cancer metastasis. *Int. J. Oncol.* **39**, 169–175 (2011).
35. Wendl, M. C. *et al.* PathScan: a tool for discerning mutational significance in groups of putative cancer genes. *Bioinformatics* **27**, 1595–1602 (2011).
36. Vaske, C. J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26**, i237–i245 (2010).
37. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).

38. Johnson, G. L. & Lapadat, R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* **298,** 1911–1912 (2002).
39. Widmann, C., Johnson, N. L., Gardner, A. M., Smith, R. J. & Johnson, G. L. Potentiation of apoptosis by low dose stress stimuli in cells expressing activated MEK kinase 1. *Oncogene* **15,** 2439–2447 (1997).
40. Wagner, E. F. & Nebreda, A. R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nature Rev. Cancer* **9,** 537–549 (2009).
41. Ellis, M. J. *et al.* Phosphatidyl-inositol-3-kinase alpha catalytic subunit mutation and response to neoadjuvant endocrine therapy for estrogen receptor positive breast cancer. *Breast Cancer Res. Treat.* **119,** 379–390 (2010).
42. Prat, A. & Perou, C. M. Mammary development meets cancer genomics. *Nature Med.* **15,** 842–844 (2009).
43. Larson, D. E. *et al.* SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* **28,** 311–317 (2011).
44. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).
45. Dees, N. *et al.* MuSiC: Identifying mutational significance in cancer genomes. *Genome Res.* (in the press).

Author Contributions M.J.E. led the clinical investigations, biomarker analysis and chip-based genomics. E.R.M., M.J.E., L.D., R.S.F., T.J.L. and R.K.W. designed the experiments. L.D. and M.J.E. led data analysis. D.S., J.W.W., D.C.K., C.C.H., M.D.M., K.C., C.A.Mi., F.D., W.S.S., M.C.W., R.C. and C.K. performed data analysis. D.S., C.A.Ma., J.W.W., J.F.M., C.L. and L.D. prepared figures and tables. R.S.F., L.L.F., R.D., M.H., T.L.V., J.H., L.L., R.C. and J.S. performed laboratory experiments. L.E., G.U., J.M., G.V.B., P.K.M., J.M.G., M.L., K.H. and J.O. provided samples and clinical data. V.J.S., K.B., J.L., Y.T. and C.K. provided statistical and clinical correlation analysis. D.O. oversees the ACOSOG Operations Center that provides oversight and tracking for ACOSOG clinical trials. K.D., S.McD., D.C.A. and M.W. provided pathology analysis. B.A.V.T., J.W., R.J.G., A.E., D.P.-W., H.P.-W., J.M.S., T.C.G., S.N., C.K. and M.C.W. performed pathway analysis. L.-W.C. and R.B. analysed the druggable target mutation data. D.J.D. and B.O. provided informatics support. L.D., M.J.E. and E.R.M. wrote the manuscript. T.J.L., M.C.W. and R.K.W. critically read and commented on the manuscript.

Author Information DNA sequence data are deposited in the restricted access portal at dbGaP, accession number phs000472.v1.p1. Gene expression array data used in the Paradigm training set is deposited in GEO, accession number GSE29442, and a Superseries that covers both the Agilent gene expression data and the Agilent array CGH data used for the Paradigm test set is deposited in GEO, accession number GSE35191   . Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. 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.J.E., L.D. and E.R.M.

ARTICLE

Large–scale prediction and testing of drug activity on side–effect targets

Eugen Lounkine[1]*, Michael J. Keiser[2,3]*, Steven Whitebread[1], Dmitri Mikhailov[1], Jacques Hamon[4], Jeremy L. Jenkins[1], Paul Lavan[4], Eckhard Weber[4], Allison K. Doak[3], Serge Côté[4], Brian K. Shoichet[3] & Laszlo Urban[1]

Discovering the unintended 'off-targets' that predict adverse drug reactions is daunting by empirical methods alone. Drugs can act on several protein targets, some of which can be unrelated by conventional molecular metrics, and hundreds of proteins have been implicated in side effects. Here we use a computational strategy to predict the activity of 656 marketed drugs on 73 unintended 'side-effect' targets. Approximately half of the predictions were confirmed, either from proprietary databases unknown to the method or by new experimental assays. Affinities for these new off-targets ranged from 1 nM to 30 μM. To explore relevance, we developed an association metric to prioritize those new off-targets that explained side effects better than any known target of a given drug, creating a drug–target–adverse drug reaction network. Among these new associations was the prediction that the abdominal pain side effect of the synthetic oestrogen chlorotrianisene was mediated through its newly discovered inhibition of the enzyme cyclooxygenase-1. The clinical relevance of this inhibition was borne out in whole human blood platelet aggregation assays. This approach may have wide application to de-risking toxicological liabilities in drug discovery.

Adverse drug reactions (ADRs) can limit the use of otherwise effective drugs. Next to lack of efficacy, they are the leading cause for attrition in clinical trials of new drugs[1–3] and are more prominent still in the failure of molecules to advance from pre-clinical research into human trials[4]. Some ADRs are caused by modulation of the primary target of a drug[5], others result from non-specific interactions of reactive metabolites[6]. In many cases, however, ADRs are caused by unintended activity at off-targets. Notorious examples of off-target toxicity include that of the appetite suppressant fenfluramine–phentermine (fen–phen), which was withdrawn from the market after numerous patient deaths. These owed to the activation of the 5-hydroxytryptamine-2B (5-HT$_{2B}$) receptor by one of its metabolites, norfenfluramine, leading to proliferative valvular heart disease[7]. Similarly, well-known drugs, such as the antihistamine terfenadine, have been withdrawn because they caused arrhythmias and death, which have been attributed to their off-target inhibition of the human *ether-à-go-go*-related gene potassium channel (hERG, also known as KCNH2)[8,9]. Prediction of unknown off-target drug interactions might prevent such disastrous drug toxicities, which are often detected only after fatalities in the clinic, and might allow safer molecules to be prioritized for pre-clinical development. Methods to systematically predict off-targets, and associate these with side effects, have thus attracted intense interest[10–16], frequently in the form of either chemical genomics[17,18] or informatics[19–26] approaches.

Whereas the informatics methods have never been tested systematically on a large scale, in principle they can be deployed against thousands of targets. Here we present a large-scale, prospective evaluation of safety target prediction using one such method, the similarity ensemble approach (SEA)[25–27]. SEA calculates whether a molecule will bind to a target based on the chemical features it shares with those of known ligands, using a statistical model to control for random similarity. Because SEA relies only on chemical similarity, it can be applied systematically and, for those targets that have known ligands, comprehensively. For 656 drugs approved for human use (Supplementary Table 1), targets were predicted from among 73 proteins (Supplementary Table 2 and Methods) with established association of ADRs[22,28], for which assays were available at Novartis. Encouragingly, many of the predictions were confirmed, often at pharmacologically relevant concentrations. This motivated us to develop a guilt-by-association metric that linked the new targets to the ADRs of those drugs for which they are the primary or well-known off-targets, creating a drug–target–ADR network. The applicability and the limitations of this approach will be considered.
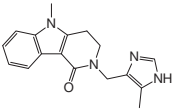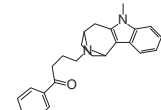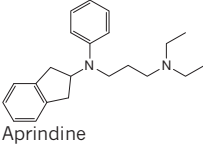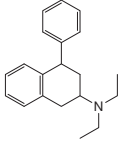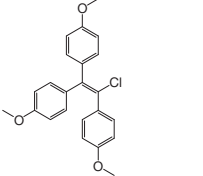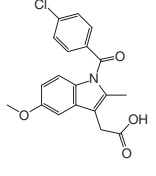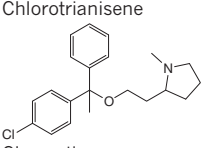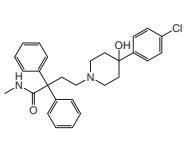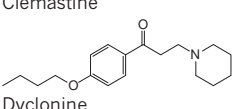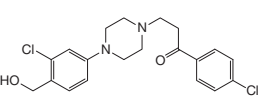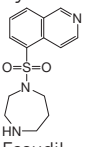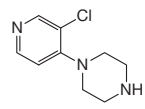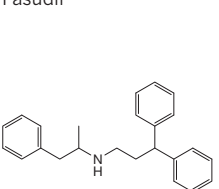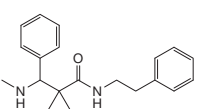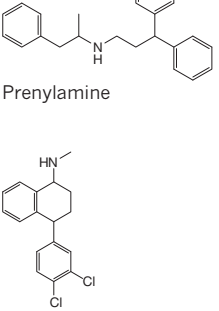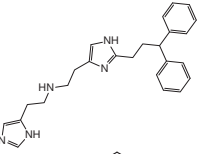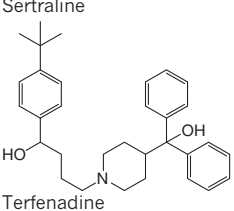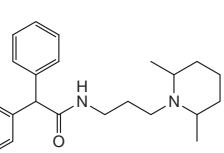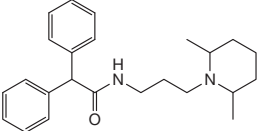
Testing the predictions

The 656 drugs were computationally screened for their likelihood to bind to 73 targets (Supplementary Table 2) using SEA[25–27]. The targets belong to the Novartis *in vitro* safety panels based on their association with ADRs[22,28]. Here we insisted that they also be described in the ChEMBL database[29], enabling correspondence with SEA predictions (Supplementary Table 2). ChEMBL annotates more than 285,000 ligands modulating more than 1,500 different human targets with affinities better than 30 μM. SEA calculated the similarity of each drug versus each set of ligands for the 73 targets, comparing the overall set similarity to a model of such expected at random. For instance, the sodium channel blocker aprindine loosely resembled the set of histamine H$_1$ ligands; although no single H$_1$ ligand was strongly similar to the drug (Table 1), the overall similarity of the set was much greater than expected at random, leading to a highly significant SEA expectation value (*E* value) of 5×10^{-26} between aprindine and H$_1$ receptor ligands. Only 1,644 of the more than 47,000 possible drug–target pairs had significant *E* values. Of these, 403 were already known in ChEMBL and so were trivially confirmed; we do not consider these further. Of the remaining 1,241 predictions, 348 (28%) were unknown to ChEMBL, but could be found in proprietary ligand–target databases that were unavailable to SEA (see Methods). The remaining

[1]Novartis Institutes for Biomedical Research, Cambridge, Massachusetts 02139, USA. [2]SeaChange Pharmaceuticals Inc, 409 Illinois Street, San Francisco, California 94158, USA. [3]Department of Pharmaceutical Chemistry, University of California, San Francisco, 1700 4th Street, Byers Hall Suite 508D, California 94158-2550, USA. [4]Novartis Institutes for Biomedical Research, 4056 Basel, Switzerland.
*These authors contributed equally to this work.

**Table 1 | New drug-off-target predictions confirmed by *in vitro* experiment**

Drug	Closest chEMBL molecule	Tc value	Target	SEA *E* value	IC$_{50}$ (μM)	Closest known target	BLAST *E* value
Alosetron		0.25	HTR2B	10.6×10^{-17}	0.02	KCNH7	3.6×10^2
Aprindine		0.38	HRH1	5.0×10^{-26}	0.78	SCN5A	3.3×10^{-1}
Chlorotrianisene		0.31	COX-1	1.9×10^{-17}	0.16	ESR1	9.0×10^2
Clemastine		0.31	SLC6A4	1.1×10^{-14}	0.42	KCNH2	6.1×10^1
Dyclonine		0.36	DRD4	1.5×10^{-17}	4.1	SLC6A3	2.3×10^2
Fasudil		0.37	ADRA2A	1.1×10^{-7}	4.0	CCR2	1.5×10^{-9}
Prenylamine		0.31	OPRM1	1.1×10^{-8}	1.8	CACNA1G	3.5×10^0
Prenylamine		0.30	HRH1	3.2×10^{-66}	7.9	SCN5A	3.3×10^{-1}
Sertraline		0.33	HRH2	5.1×10^{-45}	1.4	HTR4	6.8×10^{-51}
Terfenadine		0.24	SCN5A	8.9×10^{-8}	7.1	CYP2D6	5.9×10^1

Representative, confirmed predictions are shown. Tc values determined using ECFP_4-based molecular similarity to the closest ChEMBL reference molecule in the target set. The closest known target is a known target of the drug that has highest sequence similarity to the predicted target. The BLAST *E* value is based on the sequence identity of the predicted target to the closest known target; values greater than 10^{-5} represent unrelated proteins.

893 predictions represented previously unexplored drug–target associations.

Of these predictions, 694 were tested at Novartis. For 478, activity was less than 25% at 30 μM; these were considered disproved. For another 65 predictions, activity was between 25 and 50% at 30 μM; these were considered ambiguous. Finally, for 151 of the new drug–target predictions,

half-maximum inhibitory concentration (IC$_{50}$) values of less (better) than 30 μM were measured in concentration–response curves (Fig. 1a and Supplementary Fig. 1). In 125 cases, the drugs had an IC$_{50}$ value better than 10 μM, and in 48 cases activities were sub-micromolar (Table 1, Supplementary Table 3 and Supplementary Fig. 1). In summary, of the 1,042 predictions that were tested (694 by assay,
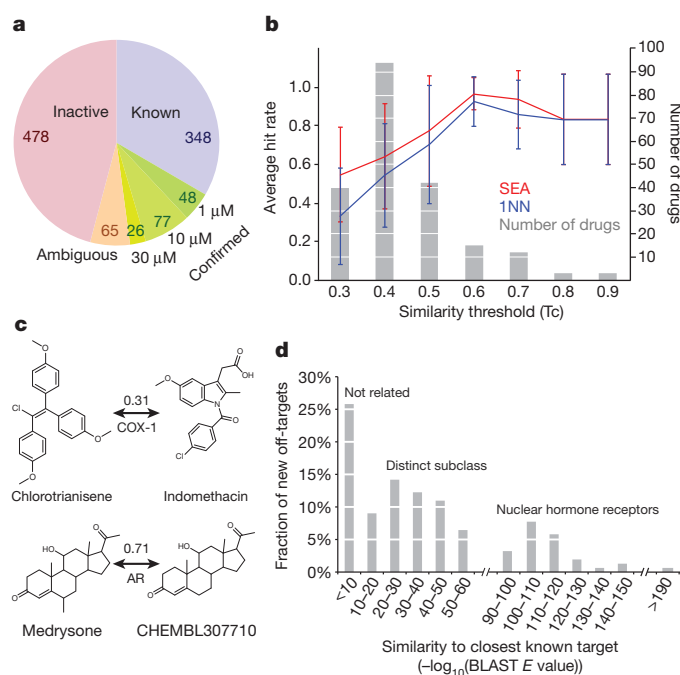
Figure 1 | Predicting off-targets, and their novelty. a, Success of SEA predictions. Known: predicted off-targets confirmed by proprietary databases. Confirmed: predictions tested *in vitro* achieving $IC_{50} < 30\,\mu M$. Ambiguous: predictions with 25–50% activity at $30\,\mu M$. Inactive: <25% activity. **b**, SEA enriched for non-trivial similarity. Drugs were binned (grey) by lowest Tc values yielding valid SEA predictions. Hit rates of SEA (red) and 1NN (blue) are shown. Error bars denote s.d. **c**, Non-intuitive (chlorotrianisene) and straightforward (medrysone) SEA predictions, with Tc values to closest references shown. Chlorotrianisene is only marginally similar to indomethacin, but is (correctly) predicted for COX-1. AR, androgen receptor. **d**, Sequence similarities of each confirmed drug off-target to the closest known target of the drug.

348 by databases), 48% were confirmed either in proprietary databases, unknown to the method and to those undertaking the SEA calculation, or in Novartis assays in full concentration responses, and just under 46% were disproved (Fig. 1a).

In assessing these results, one would like to compare the true predictions with the false-positive and the false-negative predictions. Whereas this work offers guidance on false-positive predictions, we can only address false negatives for a few compounds (Supplementary Results). Among these was astemizole, which had affinities ranging from 0.1 to $9\,\mu M$ on the 5-HT_{2A}, 5-HT_{2B} and 5-HT_{2C} serotonin receptors and the histamine H_2 and dopamine D_2 receptors, as measured in other projects at Novartis. These targets were missed owing to a charge post-filter, separate from SEA itself, which excluded compounds with net charge dissimilar from the reference ligands[30]. Astemizole was improperly assigned[31] a charge of +2, wrongly differentiating it from the known ligands; the SEA E values linking astemizole to these targets were themselves between 10^{-25} to 10^{-27}. Other failures could be attributed to SEA itself. For instance, promazine bound to the histamine H_1 and H_2 receptors with low to mid-nanomolar affinities, but the SEA E values at 10^{-4} to 10^{-3} were below our significance cut-off. This work was undertaken with ChEMBL_2 as a source of ligand–target association; had we used the more recent ChEMBL_10, the H_1 receptor would have been predicted with an E value of 10^{-9} (see http://sea.bkslab.org), and had we used ChEMBL_12 and a newer version of SEA, both targets would have been predicted. Clearly, with its reliance on topology and on inference from known ligand–target associations, SEA will have false negatives.

A key question is whether the new predictions were in any way surprising. One way to evaluate this is to compare the similarity of

drugs predicted for new targets with the closest previously known ligand for that target. We used Tanimoto coefficients (Tc), which compare the groups in common between two molecules, here represented by ECFP_4 fingerprints. Tc values between nearest molecules were small, often less than 0.4 (ref. 32); visual inspection of these pairs confirms the dissimilarity suggested by the low Tc values (Table 1). More systematically, SEA may be compared with a method that predicts targets based only on the one-nearest neighbour (1NN) model (Fig. 1b). For close analogues (Tc values > 0.7; Fig. 1c), the fraction of true positives was comparable between 1NN and SEA approaches (Fig. 1b). But across most similarity thresholds, SEA substantially outperformed 1NN, and by nearly twofold in the low similarity range. Thus, for the Rho kinase inhibitor fasudil, SEA predicted only the adrenergic α_{2A} receptor, with an E value of 1.1×10^{-7}, which was experimentally confirmed ($IC_{50} = 4\,\mu M$). This occurred despite the low similarity of the closest known α_2 ligand, which had a Tc value of 0.37 to fasudil. Conversely, at this similarity threshold the 1NN model predicted nine targets, only three of which were confirmed (Supplementary Table 4). For chlorotrianisene, two of the three targets predicted by SEA were confirmed; conversely, at its 0.31 Tc for cyclooxygenase-1 (COX-1, also known as PTGS1) the 1NN model predicted ten targets, only two of which were confirmed.

We also investigated how often the new off-target would have been obvious based on sequence similarity of the targets[25,26,33]. We calculated the BLAST sequence similarity of predicted targets to any known target of a drug (Table 1 and Supplementary Table 3). Of the 151 new off-target predictions, 39 (26%) had BLAST E values greater (worse) than 10^{-5}, suggesting the previously known targets shared no sequence similarity with the new off-targets (Table 1, Supplementary Table 3 and Fig. 1d). For example, the anaesthetic dyclonine was shown to bind the histamine H_2 receptor (HRH2), whereas the closest known target was the $Na_v1.8$ channel (SCN10A), which has no significant sequence similarity (BLAST E value > 1) and is functionally unrelated to the H_2 receptor. Similarly, the anti-nausea drug alosetron antagonized the 5-HT_{2B} receptor with an IC_{50} of 18 nM, although 5-HT_{2B} has no sequence similarity to the ion channel targets of this drug (Table 1). Chlorotrianisene potently inhibits the enzyme COX-1, which is unrelated by sequence to the primary nuclear hormone receptor of this drug, the oestrogen receptor (Table 1).

Associating *in vitro* targets with ADRs

To assess the potential clinical relevance of the discovered targets systematically, we developed a quantitative score that associated *in vitro* activity with patient ADRs. We enumerated all possible target–ADR pairs for 2,760 drugs with available adverse event annotations, expressing as an enrichment score the co-occurrence of pairs that were more common than expected by chance (Supplementary Table 5). For example, 'abdominal pain upper' has been reported for 45 drugs that interact with COX-1. The ADR abdominal pain upper was linked with 6,046 drug–target pairs, whereas COX-1 was linked with 2,188 drug–ADR pairs; there were a total of 681,797 target–ADR pairs overall. Thus the pair abdominal pain upper–COX-1 was enriched 2.3-fold above random (Methods), with a χ^2 P value of 9.9×10^{-9}. A total of 3,257 significant target–ADR associations were identified (Supplementary Table 5).

Having identified new off-targets for the drugs, and linked these with observed ADRs, we sought drug–target–ADR connections that illuminate the clinical relevance of the predictions. Of the 151 confirmed new drug–target associations tested at Novartis, 82 were significantly associated with one or more ADR, resulting in a total of 247 drug–target–ADR links. In 116 cases, the enrichment factor (EF) of the new drug–target–ADR link was stronger than that for any previously known target (Table 2 and Supplementary Table 6). For example, prenylamine was found to bind the histamine H_1 receptor (HRH1), which we associate with a sedation ADR (EF = 4.9). By contrast, none of the known targets of prenylamine was associated

Table 2 | Characteristic new, confirmed targets associated with ADRs of the drugs

Drug name	Target	Activity (μM) (median)	AUC (μM h)	C_{max} (μM)	Adverse event	EF ratio	Alternative target	Comparable drug
Chlorotrianisene	COX-1	0.16	NA	NA	Abdominal pain upper	2.32	None	None
					Rash	1.79	None	None
Clemastine	SLC6A4	0.42	NA	NA	Sleep disorder	2.15	None	None
Cyclobenzaprine	HRH1	0.02	0.16–4.10 (0.69)	0.01–0.13 (0.06)	Ataxia	1.73	None	**Desipramine**
					Somnolence	1.49	None	**Aripiprazole**
Diphenhydramine	SLC6A3	4.33	2.57–3.42 (3.00)	0.26–0.26 (0.26)	Tremor	2.02/1.90	SCN10A	**Citalopram**
Loxapine	CHRM2	1.12	0.03–0.43 (0.21)	0.02–0.41 (0.14)	Tachycardia	2.08/1.97	CHRM1	Sibutramine
Methylprednisolone	PGR	1.30	0.09–10.76 (1.28)	0.06–2.11 (0.31)	Depression	3.87/2.49	NR3C1	Flutamide
Prenylamine	HRH1	7.87	0.12–0.12 (0.12)	1.20–1.20 (1.20)	Sedation	4.94	None	None
Ranitidine	CHRM2	5.56	5.66–121.90 (9.67)	1.14–9.11 (2.12)	Constipation	1.63	None	Haloperidol
Ritodrine	OPRM1	9.18	0.03–0.32 (0.11)	0.01–0.15 (0.04)	Hyperhidrosis	3.21	None	**Oxycodone**

ADRs are listed that are more strongly associated with the predicted target than any known target, together with pharmacokinetic data (AUC and C_{max}), where available. Numbers in parentheses denote the median value, and the range denotes the minimal and maximal reported value. Where pharmacokinetics (PK) and pharmacodynamic activity (PD) were available, drugs have been identified that behave comparably to the predicted drug and also cause the adverse event. The EF ratio is the ADR–target enrichment for the predicted/best alternative known target ratio. Comparable drugs are those that are known to bind the predicted target (bold denotes predicted target is the primary target), share the ADR and behave similarly in terms of PK and PD (see Methods). NA, not available.

with this side effect. For other cases, known targets represented an alternative explanation for an ADR. For instance, we found that diphenhydramine binds to the dopamine transporter (SLC6A3; Table 2), which is associated with tremor[34]. Although tremor was also associated with one of the known targets of diphenhydramine, sodium channel SCN10A (EF = 1.9)[35], its association with the dopamine transporter was higher (EF = 2.02), indicating a possible mechanistic link with the new off-target. Conversely, the 'dry mouth' side effect of diphenhydramine was better explained by its known antagonism of the M_3 muscarinic receptor (CHRM3; EF = 2.45, Supplementary Table 5).

We asked whether the affinity of a drug for its predicted ADR–target was relevant given its pharmacology, comparing the predictions against other drugs with similar pharmacodynamics and pharmacokinetics (Table 2). This was possible for 36 drug–target–ADR links (Supplementary Table 6). For instance, cyclobenzaprine was shown to bind to the histamine H_1 receptor at 21 nM, whereas its median maximal plasma concentration (C_{max}) was 61 nM; nine other drugs binding the H_1 receptor in the nanomolar range with comparable C_{max} values were found (Table 2). Although some of the measured drug–target affinities were moderate, the pharmacokinetic data often confirmed that they were nevertheless relevant. For instance, the affinity of ranitidine (Zantac) for the M_2 muscarinic receptor, which we associate with its constipation ADR, is only 5.6 μM. Nevertheless, with an area under the curve (AUC) value of 5.7 to 122 μM h (minimal and maximal reported values) in plasma, this association seems plausible. Similarly, diphenhydramine has an IC_{50} value of only 4.3 μM against the dopamine transporter, a target that we associate with the tremor side effect of the drug. Nevertheless, the AUC value of diphenhydramine of 2.6 to 3.4 μM h supports the relevance of its modest IC_{50} value.

Drug–target–ADR networks

Network graphs help to visualize the new and known drug–target links, and the adverse events with which they are associated (Fig. 2a–c). For example, the oestrogen receptor (ESR1) modulator chlorotrianisene was found to inhibit COX-1, with an affinity substantially better than its affinity for ESR1. Drugs that modulate the two proteins can share two of the adverse reactions of chlorotrianisene ('erythema multiforme' and 'oedema'), but 'rash' and 'abdominal pain upper' link only to drugs inhibiting COX-1, and these are both associated with chlorotrianisene almost uniquely among the oestrogen receptor modulators (Fig. 2c and Supplementary Table 5). For prenylamine, a new G-protein-coupled receptor (GPCR) cluster (HRH1, OPRM1 and ADRB2) emerges that is unrelated to the primary ion channel activity of the drug, but uniquely link to its sedative and myocardial infarction ADRs (Fig. 2b). For domperidone, its known activity at dopamine receptors is associated with a Parkinsonism-like phenotype ('hyperprolactinaemia' and 'extrapyramidal disorder'),

whereas 'somnolence' associates only with the newly discovered opioid activity (Fig. 2a).

Several of the drug–target–ADR associations that emerged were surprising. Among them were the association of the muscle relaxant cyclobenzaprine with somnolence, the H_2 antagonist ranitidine with constipation, and chlorotrianisene with upper abdominal pain (Table 2). Cyclobenzaprine caught our attention because even its mechanism of action target for muscle relaxation has not been characterized, and its association with the off-target discovered here, the H_1 receptor ($IC_{50} = 20$ nM), precedes the identification of its primary target. The central nervous system H_1 receptor is strongly associated with somnolence, consistent with the ADR of the drug, and supported by its pharmacokinetics. Similarly, the constipation effect of ranitidine is consistent with its activity on the M_2 muscarinic receptor. Although its affinity for M_2 is moderate at 5.5 μM, the pharmacokinetics make this affinity relevant to this ADR.

Perhaps the most compelling demonstration of a drug–target–ADR association is one *in vivo*, or in an accepted *in vivo* biomarker. The observation that chlorotrianisene was a potent COX-1 inhibitor seemed a reasonable explanation for the upper abdominal pain (epigastralgia) side effect provoked by the drug, and one that lent itself to direct testing in an accepted biomarker. Epigastralgia is a well-known ADR of non-steroidal anti-inflammatory drugs (NSAIDs), which inhibit the cyclooxygenase enzymes COX-1 and COX-2. COX-1 has housekeeping effects in the gastric mucosa[36], and its inhibition can lead to mucosal thinning and gastroduodenal ulceration, leading to upper gastric pain and the thousands of annual hospitalizations that are associated with NSAID use[37]. NSAIDs also inhibit platelet aggregation by direct inhibition of their endogenous COX-1 enzyme[38]. Intriguingly, this effect is unreported for other synthetic oestrogens, which, to the contrary, are more likely to promote platelet aggregation[39,40]. A widely accepted model for platelet aggregation may be run *ex vivo*, in whole blood, allowing one to test for target engagement of COX-1 in this effect.

Accordingly, collagen-induced platelet aggregation was measured in freshly drawn human blood from six healthy volunteer donors. Acetylsalicylic acid, the active ingredient in aspirin, inhibited platelet aggregation by 42–48% at 250 μM. The more potent NSAID indomethacin inhibited platelet aggregation by 50% at 50 μM. Chlorotrianisene inhibited platelet aggregation in whole blood with a potency almost indistinguishable from that of indomethacin, and more potently than acetylsalicylic acid (Fig. 2d). These results are consistent with an *in vivo* inhibitory activity of chlorotrianisene on COX-1, and with the epigastralgia that is among its common side effects.

Drug and target promiscuity

To investigate overall patterns of drug and target promiscuity, we integrated the experimental results from this and other Novartis studies. The most promiscuous target was the voltage-gated sodium channel
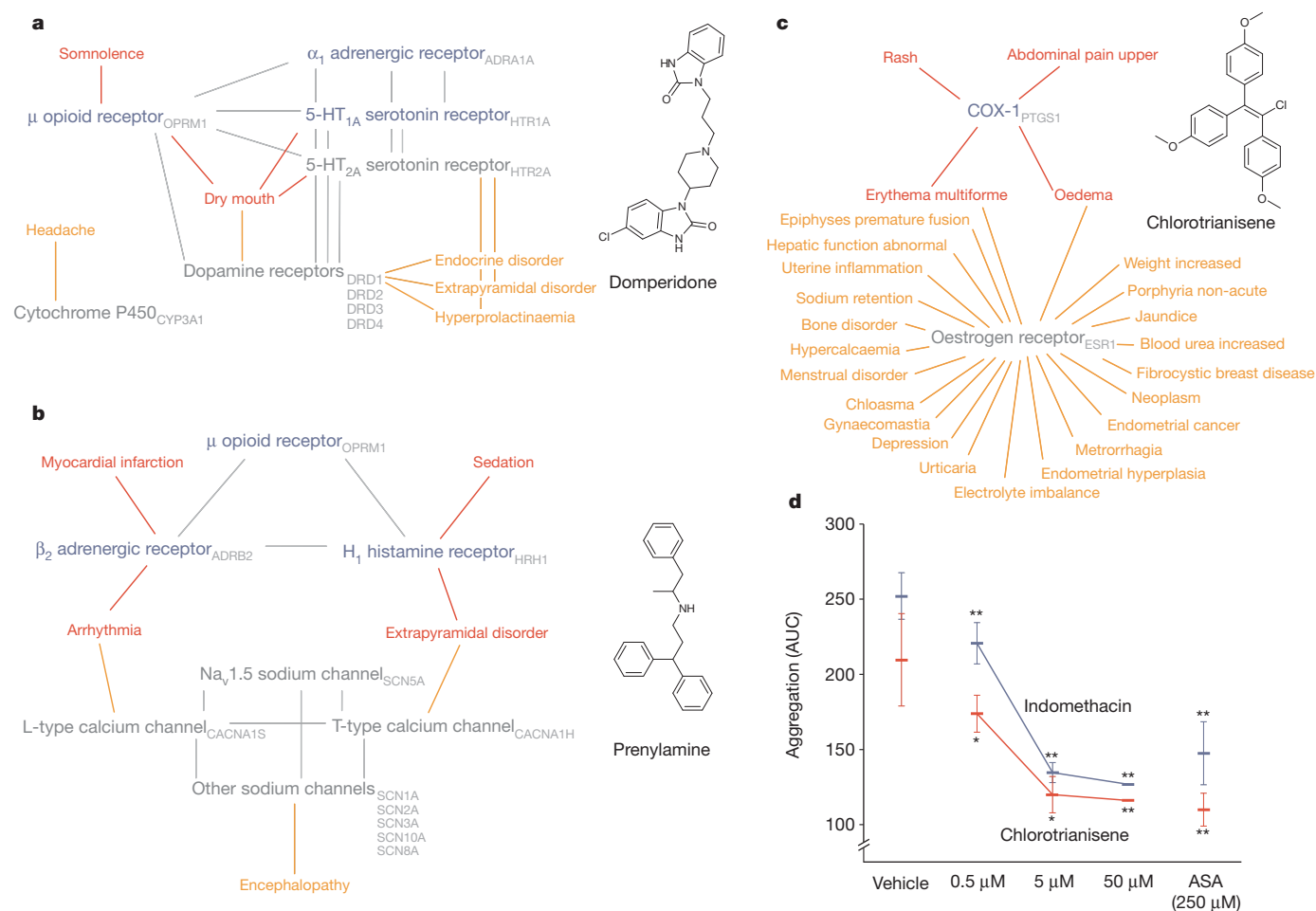
Figure 2 | Off-target networks. a–c, Off-target networks for three drugs. Known targets of the drugs are grey, whereas newly predicted targets are blue; the adverse events associated with each are orange and red, respectively. Red adverse events are significantly (EF > 1, q value < 0.05) associated with the new off-targets. Targets related by sequence are connected by grey edges. **d**, Chlorotrianisene inhibits platelet aggregation. Two independent experiments (red and blue) are shown for chlorotrianisene and indomethacin, respectively. ASA, acetylsalicylic acid (positive control); vehicle denotes negative control. *P < 0.05, **P < 0.01 (paired Student's t-test differences compared with vehicle control). Error bars denote s.d.

(SCN5A), to which 70 out of 126 (56%) tested drugs bound (Fig. 3a). From a target family standpoint, however, this was an exception, as most other promiscuous targets were small molecule-recognizing GPCRs; of non-GPCRs, only SCN5A and the ion channel hERG were targeted more than average (>13% of drugs tested). Transporters had mid-range promiscuity; enzymes, nuclear receptors and ligand-gated ion channels were less promiscuous, whereas peptide-recognizing receptors were hit least of all (Supplementary Tables 2 and 7).

Inverting this analysis, the most promiscuous drug, chlorhexidine, hit 34 out of 54 (64%) targets against which it was tested, and another nine drugs were active on more than 50% of their tested targets (Fig. 3b and Supplementary Table 8). Twenty-five drugs bound to proteins from among all major target classes. Highly promiscuous drugs were often lipophilic and cationic at physiological pH (Fig. 3b)[41,42].

Predicting off-targets and adverse events

This study begins to quantify drug polypharmacology at scale: the 656 drugs considered here each modulated an average of seven safety targets, sometimes across several classes, and more than 10% of the drugs acted on nearly half (45%) of the 73 targets (Fig. 3b). It is sobering that this promiscuity is observed for approved human drugs, which have typically already been optimized to minimize toxicity. For lead molecules that are progressing towards the clinic, this level of off-target promiscuity might be higher still[28]. Anticipating these off-targets is difficult, as they can be unrelated in sequence and structure

to the primary targets of a drug, and even known target–ADR associations are not always straightforward. Two results of this study begin to address these challenges. First, of the 1,042 predicted drug–target associations that were tested, 48% were confirmed (Fig. 1a). With 46% of the predictions disproved, the method remains imperfect, but this rate may nevertheless be high enough to prioritize compound classes and targets for testing. Second, a guilt-by-association metric can link off-targets with ADRs. A three-way association between drugs, molecular targets and ADRs may be systematically calculated and interpreted (Fig. 2a–c).

Surprisingly, drugs often modulated off-targets unrelated to their primary target. Of the 151 off-targets that were confirmed by new experiment, 39 were unrelated by sequence to any of the known drug targets (Fig. 1d). For example, the antitussive clemastine and the antihistamine diphenhydramine (an active ingredient in products such as Tylenol PM), both of which act on the histamine H_1 GPCR, also modulate the serotonin transporter (5-HTT, also known as SLC6A4), to which the primary target is unrelated by sequence or structure. Conversely, the serotonin transporter inhibitor sertraline acts on the histamine H_2 GPCR. The activity of drugs on targets that are unrelated by sequence or structure to their primary targets can seem capricious and certainly makes prioritization of likely targets more difficult. A ligand-based approach offers an orthogonal view of target relationships and so can illuminate similarities that are opaque from a molecular biology perspective. The converse is also true, and the two views will often be complementary.

Figure 3 | Target and drug promiscuity. a, Target promiscuity. Targets are sorted based on the percentage of drugs hitting the target below 30 μM (indicated by numbers next to target names). Colours code for seven distinct target classes. **b**, Promiscuous drugs are often hydrophobic and cationic. Each point represents one drug. Ionization at pH 7.4 and AlogP (calculated octanol–water partitioning coefficient) values were calculated from drug structures. Hit rate denotes the percentage of targets the drug binds below 30 μM.

The association between chlorotrianisene, COX-1 and epigastralgia illustrates the potential of the approach. The therapeutic target of chlorotrianisene, the oestrogen nuclear hormone receptor, bears no sequence or structural similarity to the COX-1 enzyme, but the likelihood of cross-activity between the two targets is articulated by ligand similarity (Table 1). Correspondingly, the linking of abdominal pain and COX-1 only emerges when one quantitatively compares the ADRs of known COX-1 inhibitors to what one would expect at random. The potent inhibition of platelet aggregation by chlorotrianisene in whole blood (Fig. 2d) is consistent with the systemic, *in vivo* activity of this drug at relevant concentrations on COX-1.

Certain shortcomings of the method should not escape the reader's attention. Almost 46% of the predicted drug–target associations were disproved, and the method, which is inference-based, undoubtedly has important false negatives. As such, SEA cannot replace compound-target testing. What it can do is identify compounds early in development for possible liabilities that would ordinarily be identified only much later in drug progression. Similarly, the guilt-by-association method is inference-based and mechanism-naive, and so will miss some target–ADR associations and make others that are invalid. Also, only some side effects fall into the remit of this approach, which assumes an off-target mechanism. Side effects, like other pharmacological events, have a strong exposure component,

and can result from complex interactions with regulatory networks[43]. Thus, topical drugs such as econazole or chlorhexidine, although promiscuous *in vitro*, have fewer ADRs than expected because they never achieve sufficient systemic exposure *in vivo*. Conversely, a drug such as tacrolimus might be relatively selective *in vitro*, but is associated with several side effects owing to its broad immunosuppressant effects.

These caveats should not obscure the potential of this approach to predict and understand drug side effects. The method was used automatically at scale, without human intervention. Whereas its predictions were sometimes disproved, they were just as often confirmed. If some of the targets it suggested required no imaginative leap—as when a steroid was predicted for a new nuclear hormone receptor—a quarter of the confirmed targets were unrelated by sequence or structure to any of the known targets of the drugs (Fig. 1c, d, Table 1 and Supplementary Table 2). Pragmatically, the ability to calculate drug–target–ADR networks provides a tool to anticipate liabilities among candidate drugs being advanced towards the clinic, or yet earlier, for prioritization of chemotypes in preclinical series. If such networks cannot replace direct experimentation, they can usefully prioritize off-targets for consideration. As we struggle to develop new therapeutics, this and related approaches[11,19,21–24,44,45] can identify molecules with liabilities early in their development, and so focus effort on those candidates that are least subject to them.

METHODS SUMMARY

A collection of on-hand 656 US Food and Drug Administration-approved drugs was computationally screened against a panel of 339 molecular targets representing the species-specific expansion of 73 target assays used in Novartis safety panels. Each of the 339 target proteins was represented by its set of known ligands, as extracted from the ChEMBL_2 database. The two-dimensional structural similarity of a drug to the ligand set of a target was quantified as an E value using the SEA, then subjected to a molecular charge filter. Predictions were tested retrospectively using proprietary databases including GeneGo Metabase, Thompson Reuters Integrity, Drugbank and GVKBio; new predictions were tested prospectively in Novartis *in vitro* assays. Binding assays, and, when available, functional assays were performed including scintillation proximity, fluorometric imaging, filtration, fluorescence polarization, patch clamp, time-resolved fluorescence resonance energy transfer, and homogenous time-resolved fluorescence assays. Concentration-response curves were calculated using XLfit (v.2 or v.4, IDBS) or corresponding in-house software. All curves were redrawn using GraphPad PRISM v.5. Adverse drug reaction data were extracted from the World Drug Index and encoded using the medicinal dictionary for regulatory affairs (MedDRA). Using target annotations from GeneGo Metabase, Integrity, Drugbank, ChEMBL and GVKBio, target–ADR pairs for all drugs were enumerated. Disproportionality analysis in conjunction with a chi-squared test for association was carried out for all drug–target pairs. The false discovery rate was controlled using the Benjamini–Hochberg correction for multiple hypothesis testing. Pharmacokinetic data were extracted from Integrity. For target and drug promiscuity analysis, combined external and internal target annotations were used. The computational workflow apart from SEA was implemented in Pipeline Pilot version 8 and statistical analyses were performed in R. Platelet aggregometry was performed in human blood for chlorotrianisene and indomethacin, with acetylsalicylic acid as a positive control.

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

1. Giacomini, K. M. *et al.* When good drugs go bad. *Nature* **446,** 975–977 (2007).
2. Arrowsmith, J. Trial watch: phase III and submission failures: 2007–2010. *Nature Rev. Drug Discov.* **10,** 87 (2011).
3. Arrowsmith, J. Trial watch: phase II failures: 2008–2010. *Nature Rev. Drug Discov.* **10,** 328–329 (2011).
4. Boyer, S. The use of computer models in pharmaceutical safety evaluation. *Altern. Lab. Anim.* **37,** 467–475 (2009).
5. Wong, D., Wang, M., Cheng, Y. & Fitzgerald, G. A. Cardiovascular hazard and non-steroidal anti-inflammatory drugs. *Curr. Opin. Pharmacol.* **5,** 204–210 (2005).
6. Antunes, A. M. M. *et al.* Protein adducts as prospective biomarkers of nevirapine toxicity. *Chem. Res. Toxicol.* **23,** 1714–1725 (2010).
7. Rothman, R. B. *et al.* Evidence for possible involvement of 5-HT$_{2B}$ receptors in the cardiac valvulopathy associated with fenfluramine and other serotonergic medications. *Circulation* **102,** 2836–2841 (2000).
8. Roy, M., Dumaine, R. & Brown, A. M. HERG, a primary human ventricular target of the nonsedating antihistamine terfenadine. *Circulation* **94,** 817–823 (1996).
9. Curran, M. E. *et al.* A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell* **80,** 795–803 (1995).
10. Ji, Z. L. *et al.* Drug Adverse Reaction Target Database (DART): proteins related to adverse drug reactions. *Drug Saf.* **26,** 685–690 (2003).
11. Kuhn, M., Campillos, M., Letunic, I., Jensen, L. J. & Bork, P. A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* **6,** 343 (2010).
12. Matthews, E. J. & Frid, A. A. Prediction of drug-related cardiac adverse effects in humans–A: Creation of a database of effects and identification of factors affecting their occurrence. *Regul. Toxicol. Pharmacol.* **56,** 247–275 (2010).
13. Yang, X. *et al.* Kinase inhibition-related adverse events predicted from *in vitro* kinome and clinical trial data. *J. Biomed. Inform.* **43,** 376–384 (2010).
14. Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nature Chem. Biol.* **4,** 682–690 (2008).
15. Zhang, J.-X. *et al.* DITOP: drug-induced toxicity related protein database. *Bioinformatics* **23,** 1710–1712 (2007).
16. Yang, L., Luo, H., Chen, J., Xing, Q. & He, L. SePreSA: a server for the prediction of populations susceptible to serious adverse drug reactions implementing the methodology of a chemical-protein interactome. *Nucleic Acids Res.* **37,** W406–W412 (2009).
17. Lee, S., Lee, K. H., Song, M. & Lee, D. Building the process-drug-side effect network to discover the relationship between biological processes and side effects. *BMC Bioinformatics* **12** (suppl. 2), S2 (2011).
18. Schreiber, G. & Keating, A. E. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **21,** 50–61 (2011).
19. Oprea, T. I. *et al.* Associating drugs, targets and clinical outcomes into an integrated network affords a new platform for computer-aided drug repurposing. *Mol. Inform.* **30,** 100–111 (2011).
20. Paolini, G. V., Shapland, R. H. B., van Hoorn, W. P., Mason, J. S. & Hopkins, A. L. Global mapping of pharmacological space. *Nature Biotechnol.* **24,** 805–815 (2006).
21. Scheiber, J. *et al.* Mapping adverse drug reactions in chemical space. *J. Med. Chem.* **52,** 3103–3107 (2009).
22. Bender, A. *et al.* Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem* **2,** 861–873 (2007).
23. Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L. J. & Bork, P. Drug target identification using side-effect similarity. *Science* **321,** 263–266 (2008).
24. Tatonetti, N. P. *et al.* Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. *Clin. Pharmacol. Ther.* **90,** 133–142 (2011).
25. Keiser, M. J. *et al.* Predicting new molecular targets for known drugs. *Nature* **462,** 175–181 (2009).
26. Keiser, M. J. *et al.* Relating protein pharmacology by ligand chemistry. *Nature Biotechnol.* **25,** 197–206 (2007).
27. Hert, J., Keiser, M. J., Irwin, J. J., Oprea, T. I. & Shoichet, B. K. Quantifying the relationships among drug classes. *J. Chem. Inf. Model.* **48,** 755–765 (2008).
28. Azzaoui, K. *et al.* Modeling promiscuity based on *in vitro* safety pharmacology profiling data. *ChemMedChem* **2,** 874–880 (2007).
29. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–D1107 10.1093/nar/gkr777 (2012).
30. Laggner, C. *et al.* Chemical informatics and target identification in a zebrafish phenotypic screen. *Nature Chem. Biol.* **8,** 144–146 (2012).
31. Shelley, J. C. *et al.* Epik: a software program for pK$_a$ prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **21,** 681–691 (2007).
32. Muchmore, S. W. *et al.* Application of belief theory to similarity data fusion for use in analog searching and lead hopping. *J. Chem. Inf. Model.* **48,** 941–948 (2008).
33. Yıldırım, M. A., Goh, K.-I., Cusick, M. E., Barabási, A.-L. & Vidal, M. Drug–target network. *Nature Biotechnol.* **25,** 1119–1126 (2007).
34. Marshall, V. & Grosset, D. G. Role of dopamine transporter imaging in the diagnosis of atypical tremor disorders. *Mov. Disord.* **18** (suppl. 7), S22–S27 (2003).
35. Kuo, C. C., Huang, R. C. & Lou, B. S. Inhibition of Na$^+$ current by diphenhydramine and other diphenyl compounds: molecular determinants of selective binding to the inactivated channels. *Mol. Pharmacol.* **57,** 135–143 (2000).
36. Schoen, R. T. & Vender, R. J. Mechanisms of nonsteroidal anti-inflammatory drug-induced gastric damage. *Am. J. Med.* **86,** 449–458 (1989).
37. Kong, S. X., Hatoum, H. T., Zhao, S. Z., Agrawal, N. M. & Geis, S. G. Prevalence and cost of hospitalization for gastrointestinal complications related to peptic ulcers with bleeding or perforation: comparison of two national databases. *Am. J. Manag. Care* **4,** 399–409 (1998).
38. G., Perrone, M., Scilimati, A., Simone, L. & Vitale, P. Selective COX-1 inhibition: A therapeutic target to be reconsidered. *Curr. Med. Chem.* **17,** 3769–3805 (2010).
39. Akaraseerenont, P., Tripatara, P., Chotewuttakorn, S., Palo, T. & Thaworn, A. The effects of estrone, estradiol and estriol on platelet aggregation induced by adrenaline and adenosine diphosphate. *Platelets* **17,** 441–447 (2006).
40. Norris, L. A. & Bonnar, J. Effect of oestrogen dose on whole blood platelet activation in women taking new low dose oral contraceptives. *Thromb. Haemost.* **72,** 926–930 (1994).
41. Leeson, P. D. & Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Rev. Drug Discov.* **6,** 881–890 (2007).
42. Peters, J.-U., Schnider, P., Mattei, P. & Kansy, M. Pharmacological promiscuity: dependence on compound properties and target specificity in a set of recent Roche compounds. *ChemMedChem* **4,** 680–686 (2009).
43. Cosgrove, B. D. *et al.* Cytokine-associated drug toxicity in human hepatocytes is associated with signaling network dysregulation. *Mol. Biosyst.* **6,** 1195–1206 (2010).
44. Mestres, J., Gregori-Puigjané, E., Valverde, S. & Solé, R. V. Data completeness–the Achilles heel of drug-target networks. *Nature Biotechnol.* **26,** 983–984 (2008).
45. Mestres, J., Gregori-Puigjané, E., Valverde, S. & Solé, R. V. The topology of drug-target interaction networks: implicit dependence on drug properties and target families. *Mol. Biosyst.* **5,** 1051–1057 (2009).

METHODS

Virtual target profiling of drugs. We assembled a set of 656 drugs (Supplementary Table 1) available for internal prospective testing together with 73 assay targets for which Novartis safety panel assays[28] were available. To compare activity annotations across databases, each target was mapped to human genes using Entrez gene and ChEMBL target identifiers (Supplementary Table 2). For target prediction, the 73 targets were represented by 339 orthologous proteins from human, rat, mouse, bovine and sheep, using the ChEMBL_2 database (released 25 March 2010); ligands for these targets with affinities $\leq 1\,\mu$M were grouped into sets for the SEA calculation.

We computationally screened the 656 drugs against the 339-target panel, using 1024-bit folded ECFP_4 (ref. 46) and 2048-bit Daylight[47] fingerprints independently, with the Tc value as the similarity metric. Tc values lie between 0 and 1, in which 1 corresponds to perfect overlap of two fingerprints. Where both fingerprints yielded the same SEA prediction, we took the prediction with the lower (that is, stronger) E value, unless otherwise noted. The maximum pair-wise Tc value was used in the 1NN model.

Predictions with $E < 10^{-4}$ were retained. As a final step, we subjected the SEA predictions to a pass/no-pass charge filter, to de-prioritize those predictions in which the total charge of the drug did not match the charges calculated for at least 5% of the known ligands of the predicted target[30,31]. This resulted in 4,195 drug–ChEMBL target pairs that were subsequently mapped to the 73 target panel, resulting in 1,644 unique predictions (the difference reflects the orthologous redundancies).

Testing predictions. Many SEA predictions could be confirmed by interrogation of proprietary databases, available at Novartis but unavailable in San Francisco where the calculations were performed. These included the Thompson Reuters Integrity (http://thomsonreuters.com/products_services/science/science_products/a-z/integrity, accessed January 2011), GeneGo Metabase (version 6.2, http://www.genego.com, accessed January 2011), and GVKBio (http://www.gvkbio.com/, accessed January 2011). In addition, we also compared predictions with the ChEMBL_11 (ref. 29) and Drugbank 3.0 (ref. 48) databases. For comparison across data sources, compounds were represented using the non-stereo-specific part of InChiKeys[49].

For prospective evaluation of the remaining predictions we used binding and functional assay data from internal Novartis profiling efforts, carried out in parallel to the SEA study. For some targets, functional assays were also available. Full concentration–responses curves were plotted for any compound with at least 50% inhibition or activity at the maximal tested concentration (30 μM; Supplementary Fig. 1). For detailed assay descriptions, see Supplementary Methods and Supplementary Table 2.

Comparison to a 1NN model. We evaluated two 1NN models, using either ECFP_4 or Daylight fingerprints. Each drug was compared with all reference ligands of a target. The highest Tc value resulting from that comparison was assigned to the drug–target pair. For each drug, we identified the lowest Tc value that yielded valid SEA predictions using the respective fingerprint, and collected all drug–target pairs with Tc scores above that threshold, irrespective of the SEA E value. We counted the predictions confirmed in the proprietary databases or by experiment at Novartis. We calculated an adjusted hit rate:

$$\text{Adjusted hit rate} = (\text{number of true positives} + 1)/$$
$$(\text{number of total predictions} + 1).$$

The additional count for both numerator and denominator distinguishes cases in which no predictions were confirmed, but one method or the other predicted fewer targets. For example, SEA predicted four targets for bezafibrate, none of which were confirmed (Supplementary Table 4). However, at the corresponding Tc threshold of 0.37, the ECFP_4 1NN model identified 12 potential targets, none of which was confirmed. The adjusted fraction for SEA is 0.2 ((0+1)/(4+1)), whereas the adjusted fraction for the 1NN model is 0.077 (1/13). We monitored the average adjusted hit rate for ten similarity threshold bins ranging from 0 to 1.

BLAST target comparison. To investigate how closely the predicted targets are related to already known primary or off-targets, we calculated a target similarity matrix for all known and predicted targets found in our study. Amino acid sequences of all targets were assembled from UniProt[50]. Sequences were compared in a pair-wise manner using BLASTp as implemented in Pipeline Pilot (version 8, http://www.accelrys.com)[51]. Target sequence similarity was quantified using BLAST E values. Target pairs with values smaller than 10^{-5} were considered related by sequence.

Target and drug promiscuity. Targets were classified using the ChEMBL target taxonomy, which consists of eight levels. The first three levels were used here to distinguish between small molecule and peptide GPCRs, as well as voltage- and ligand-gated ion channels (Supplementary Table 5). In-house and literature drug–target annotations were combined, and annotations with $IC_{50} < 30\,\mu$M were counted as hits. The lipophilicity of drugs was assessed by calculating AlogP values in Pipeline Pilot. Negative values correspond to hydrophilic compounds, and positive values to lipophilic compounds.

Associations between targets and ADRs. ADRs were extracted from the World Drug Index (WDI, http://thomsonreuters.com/products_services/science/science_products/a-z/world_drug_index/, accessed March 2011) and mapped to preferred terms from the medicinal dictionary for regulatory affairs (MedDRA)[52]. MedDRA organizes adverse reaction terms in a hierarchy reaching from low-level terms to system organ classes at the highest level. Original WDI terms were first mapped to low-level terms in the MedDRA hierarchy using text mining components in Pipeline Pilot (version 8). Low-level terms serve as synonyms for preferred terms in MedDRA. These preferred terms were used to identify each adverse event uniquely. For example, the low-level terms 'dry mouth' and 'xerostomia' both map to the preferred term dry mouth. This resulted in 1,685 unique ADR terms, 2,760 unique drug structures with ADR annotations, and a total of 51,101 drug–ADR pairs. Using drug–target associations from databases used for testing predictions, we enumerated all target–ADR pairs (681,797 total). The assessment was done separately for binding, antagonist and agonist annotations. Assuming that each ADR could potentially occur owing to any of the targets hit by the drug, we enumerated all possible target–ADR pairs for each drug. Target–ADR pairs occurring more than ten times were retained. The number of observations for each unique pair was then compared with the expected number of observations given the overall distribution of activity and adverse effect annotations. An enrichment score was calculated for each target–ADR pair:

$$\text{EF} = p/(A \times T/P)$$

in which p is the co-occurrence of target X and ADR Y, A is the number of times ADR Y was linked to any drug–target pair, T is the number of times target X was linked with any drug–ADR pair, and P is the total number of target–ADR pairs.

To assess the statistical significance of found associations, we applied the chi-squared test for association based on contingency tables calculated for each unique target–ADR pair with an EF score greater than one. The false discovery rate was controlled using Benjamini–Hochberg correction in R (version 2.12, http://www.r-project.org)[53]. P values and q values (that is, P values corrected for multiple hypothesis testing), as well as the χ^2 statistic were calculated using the R statistical package. A total of 3,257 associations with a q value of <0.05 were retained (Supplementary Table 5).

Adverse reactions associated with predicted targets. Enrichment factors of predicted target–ADR pairs were compared with the association of ADRs with any known targets of each drug. We prioritized adverse reactions that were stronger associated with the predicted than with any known target (that is, had a higher EF score). To prioritize further adverse reactions that were probably due to the newly predicted target we extracted pharmacokinetic data from Thompson Reuters Integrity. Maximal plasma concentration (C_{max}) and cumulative concentration (AUC) values measured in humans were assembled. Activity data were assembled from quantitative sources (ChEMBL_11 and GVKBio) for drugs that were not part of the predictions, but shared ADRs with the prediction drugs. Drugs were identified for each prediction and associated ADR that satisfied the following three criteria: (1) they shared the ADR with the prediction drug; (2) they were not more than ten times more active at the predicted target; and (3) their C_{max} value and/or AUC value was not more than ten times higher than for the prediction drug.

Platelet aggregation inhibition. Human blood samples from six healthy volunteer male donors were used to perform platelet aggregometry with a multiplate impedance aggregometer (Dynabyte Medical) as follows: chlorotrianisene or indomethacin was added to whole blood at final concentrations of 0.5, 5 and 50 μM, and incubated at room temperature for 10 min; platelet aggregation was induced with 1 μg ml^{-1} collagen and measured at 37 °C for 15 min; control aggregations were measured with vehicle only, and with 250 μM acetylsalicylic acid. Statistical analysis was performed using two-tailed Student's t-tests and $P \leq 0.05$ was considered significant. A detailed description can be found in the Supplementary Methods.

46. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50,** 742–754 (2010).
47. James, C., Weininger, D. & Delany, J. *Daylight Theory Manual — Daylight 4.91* (Daylight Chemical Information Systems, 2005).
48. Wishart, D. S. *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **36,** D901–D906 (2008).
49. Stein, S., Heller, S. & Tchekhovski, D. An open standard for chemical structure representation — the IUPAC chemical identifier. *Nimes Int. Chem. Inf. Conf. Proc.* **143,** 131–143 (2003).
50. The UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.* **39,** D214–D219 (2010).
51. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
52. Brown, E. G., Wood, L. & Wood, S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf.* **20,** 109–117 (1999).
53. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).

ARTICLE

Structure of yeast Argonaute with guide RNA

Kotaro Nakanishi[1]*, David E. Weinberg[2,3]*, David P. Bartel[2,3] & Dinshaw J. Patel[1]

The RNA–induced silencing complex, comprising Argonaute and guide RNA, mediates RNA interference. Here we report the 3.2 Å crystal structure of *Kluyveromyces polysporus* Argonaute (KpAGO) fortuitously complexed with guide RNA originating from small-RNA duplexes autonomously loaded and processed by recombinant KpAGO. Despite their diverse sequences, guide–RNA nucleotides 1–8 are positioned similarly, with sequence-independent contacts to bases, phosphates and 2′-hydroxyl groups pre-organizing the backbone of nucleotides 2–8 in a near–A-form conformation. Compared with prokaryotic Argonautes, KpAGO has numerous surface-exposed insertion segments, with a cluster of conserved insertions repositioning the N domain to enable full propagation of guide–target pairing. Compared with Argonautes in inactive conformations, KpAGO has a hydrogen-bond network that stabilizes an expanded and repositioned loop, which inserts an invariant glutamate into the catalytic pocket. Mutation analyses and analogies to ribonuclease H indicate that insertion of this glutamate finger completes a universally conserved catalytic tetrad, thereby activating Argonaute for RNA cleavage.

RNA interference (RNAi) is a eukaryote-specific gene-silencing pathway triggered by double-stranded RNA (dsRNA)[1–3]. In this pathway, the ribonuclease (RNase) III enzyme Dicer first cleaves the dsRNA trigger into small interfering RNAs (siRNAs), which have 5′-monophosphates and pair to each other with two-nucleotide 3′ overhangs[4–6]. The siRNA duplex is incorporated into the effector protein Argonaute (AGO), at which point one of the strands (designated the passenger strand) is cleaved[7–9]. After the cleaved passenger strand is discarded, the resulting ribonucleoprotein complex (the RNA-induced silencing complex, or RISC) uses the remaining siRNA strand (designated the guide strand) to specify interactions with target RNAs[10,11]. If sequence complementarity between guide and target is extensive, AGO again catalyses cleavage, resulting in 'slicing' of the target RNA[12].

The first structures of full-length AGOs were of prokaryotic proteins from *Pyrococcus furiosus*[12] and *Aquifex aeolicus*[13]. Early structures revealed that the PIWI domain adopts an RNase H-like fold, thereby implicating AGO as the 'slicer' enzyme that mediates RNAi[12,14,15]. Because these prokaryotic enzymes bind 5′-phosphorylated guide DNAs rather than RNAs[13,16], subsequent structures featured the binary complex of *Thermus thermophilus* Ago (TtAGO) with guide DNA[17] and ternary complexes with target RNAs of varying length[18,19]. These studies shed light on the nucleation, propagation and cleavage steps of the AGO catalytic cycle[19,20]. However, the physiological role of prokaryotic AGOs is enigmatic; the origin of the guide DNA is unknown, and bacteria lack recognizable components of the RNAi pathway[21]. Therefore, attention has turned to eukaryotic AGOs, which use RNA guides and have protein-binding partners absent in bacteria[22]. Eukaryotic AGOs are also larger than prokaryotic AGOs because of insertion elements of unknown structure and function. Structures of individual domains and the MID-PIWI lobe within eukaryotic AGO have been determined[23–28], but structural characterization of the entire protein has remained a challenge.

Although *Saccharomyces cerevisiae* lacks RNAi, some closely related budding-yeast species were recently shown to have retained RNAi, thereby offering fresh possibilities for the study of the eukaryotic pathway[29]. We previously determined the structure and mechanism of Dicer from the budding yeast *Kluyveromyces polysporus*[30] and thus turned our attention to the AGO of this species.

Cleavage activity of budding–yeast AGO

K. polysporus AGO (Ago1) has the four conserved domains (N, PAZ, MID, PIWI) and two linker regions (L1, L2) found in other AGOs (Fig. 1a). It also has an amino-terminal extension, predicted to be disordered, which we removed to facilitate crystallization. The resulting protein, KpAGO, can substitute for the full-length protein when reconstituting RNAi in *S. cerevisiae* (Fig. 1b).

KpAGO and other budding-yeast AGOs have acidic side chains at the three positions corresponding to active-site residues in slicing-competent AGOs[31] (Supplementary Fig. 1), which suggested that KpAGO might also cleave target RNAs. Indeed, after incubation with a single-stranded guide RNA, recombinant KpAGO cleaved a matched target RNA at the expected position (Fig. 1c). To examine whether slicing occurs *in vivo*, we performed degradome sequencing from a related RNAi-containing yeast, *Saccharomyces castellii*. This procedure identifies polyadenylated RNAs containing 5′-monophosphates, including products of AGO-catalysed slicing[32,33]. Many *AGO1*-dependent degradome tags mapped to Y′-element transcripts (major targets of *S. castellii* RNAi[29]), and these tended to pair to endogenous siRNAs in the register that implicated cleavage across from positions 10–11 of the guide RNA, which was diagnostic of slicing[34] (Supplementary Fig. 2). These results indicate that budding-yeast AGO functions as a slicer during endogenous RNAi, and with the *in vitro* results establish KpAGO as a eukaryotic slicer suitable for structure–function analyses.

Structural architecture of eukaryotic AGO

We crystallized recombinant KpAGO purified from *Escherichia coli*. Extensive screening identified several crystals that were free of twinning, one of which diffracted to 3.2 Å resolution. A crystal of selenomethionine-substituted KpAGO yielded reflections suitable

[1]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [3]Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.
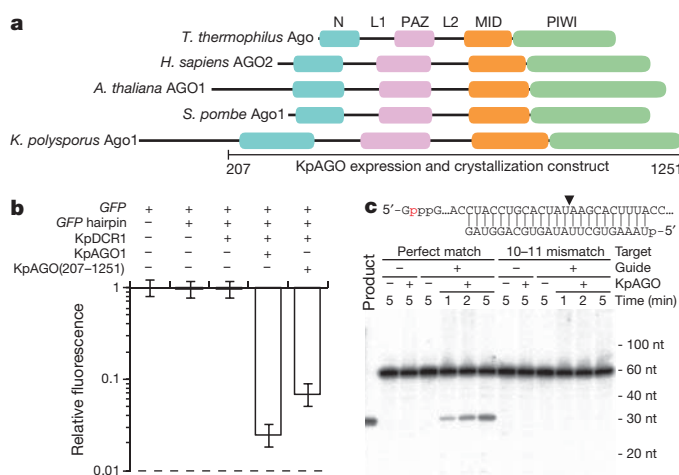
Figure 1 | Cleavage activity of budding-yeast AGO. a, Domain architectures of AGO proteins from *T. thermophilus*, *Homo sapiens*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe* and *K. polysporus*. **b,** RNAi reconstituted in *S. cerevisiae* using *K. polysporus* genes. Median green fluorescent protein (GFP) intensity is plotted as a fraction of GFP-only control. Error bars, quartiles; dashed line, background fluorescence. **c,** Cleavage activity of KpAGO. RNAs labelled at a cap phosphate (red) and matching the guide (either perfectly or with mismatches to positions 10–11) were incubated with/without (+/−) KpAGO that had been pre-incubated in the presence/absence (+/−) of synthetic guide RNA. Product was resolved on a denaturing polyacrylamide gel alongside cap-labelled synthetic product (left) and RNA standards (migration shown on right; nt, nucleotide).

for phasing by single-wavelength anomalous dispersion (Supplementary Table 1; representative electron density is shown in Supplementary Fig. 3).

The overall structure of KpAGO resembles the bilobal architecture of its prokaryotic counterparts, but with expansions throughout the protein (Fig. 2a and Supplementary Fig. 4). Of the 19 insertion segments not found in prokaryotic AGOs[12,13,17], 11 were conserved segments (cS) found in all eukaryotic AGOs, albeit with some differences in secondary structure and length, whereas the remaining eight were variable segments (vS) found in only some eukaryotic AGOs (Supplementary Fig. 1). All insertion segments are external, generating new surfaces for potential interactions with AGO-binding proteins.

Autonomously loaded guide RNA

After modelling the KpAGO protein, the $F_o - F_c$ map revealed continuous residual electron density lying along the nucleic-acid-binding channel (Fig. 2b). This unanticipated density resembled that of an oligonucleotide and could be fit well with an RNA octamer (Fig. 2c and Supplementary Fig. 5a). Analysis of end-labelled polynucleotides extracted from soluble and crystalline KpAGO confirmed the presence of small RNAs (Fig. 2d and Supplementary Fig. 5b), the high-throughput sequencing of which identified a diverse population with a bimodal length distribution centring at 12 and 17 nucleotides (Supplementary Table 2 and Supplementary Fig. 5c, d).

The location of the copurifying RNAs suggested that they might represent functional guide RNAs. Supporting this interpretation, they had two features of budding-yeast guide RNAs: 5′ uridine enrichment (Fig. 2e) and the presence of 5′ monophosphate, indicated by both electron density (Fig. 2c) and a phosphatase-sensitive block of 5′-end labelling (Supplementary Fig. 5e)[29]. Moreover, the KpAGO preparation sliced an RNA containing a site complementary to a copurifying 17-nucleotide RNA comprising approximately 0.1% of our sequencing reads (Fig. 2f and Supplementary Fig. 5c). Slicing was at the anticipated linkage and sensitive to mismatches to guide nucleotides 10–11. Reactions showed initial burst kinetics, as observed previously

Figure 2 | KpAGO architecture and copurifying RNA. a, KpAGO protein structure, with N (cyan), linker L1 (yellow), PAZ (violet), linker L2 (grey), MID (orange), and PIWI (green) domains in ribbon representation. Constant (cS) and variable (vS) insertion segments, blue and slate, respectively; disordered regions, dotted lines. **b,** $F_o - F_c$ map (blue) contoured at 2.8σ before modelling RNA. **c,** Simulated-annealing omit map (blue) contoured at 3.5σ around final RNA model (red). **d,** Nuclease sensitivity of copurifying nucleic acid. End-labelled polynucleotides extracted from the indicated KpAGO samples were either untreated (−) or incubated with RNase (R) or DNase (D) before analysis on a denaturing gel. **e,** Nucleotide composition and origin of copurifying RNA.

Sequences were analysed for enriched or depleted nucleotides (positive or negative bits, respectively) at each of the first eight positions (top). Numbers of sequencing reads mapping along each strand of the KpAGO expression plasmid are indicated (bottom, log scale). **f,** Cleavage activity guided by copurifying RNA. As in Fig. 1c, except labelled RNAs were designed to match the indicated copurifying RNA. **g,** Autonomous duplex loading and passenger-strand cleavage. Labelled single-stranded RNA (ssRNA) or siRNA duplex was incubated with or without KpAGO. **h,** Autonomous duplex loading and target cleavage. As in Fig. 1c, except KpAGO and RNA concentrations were reduced by 90% and 99.95%, respectively.
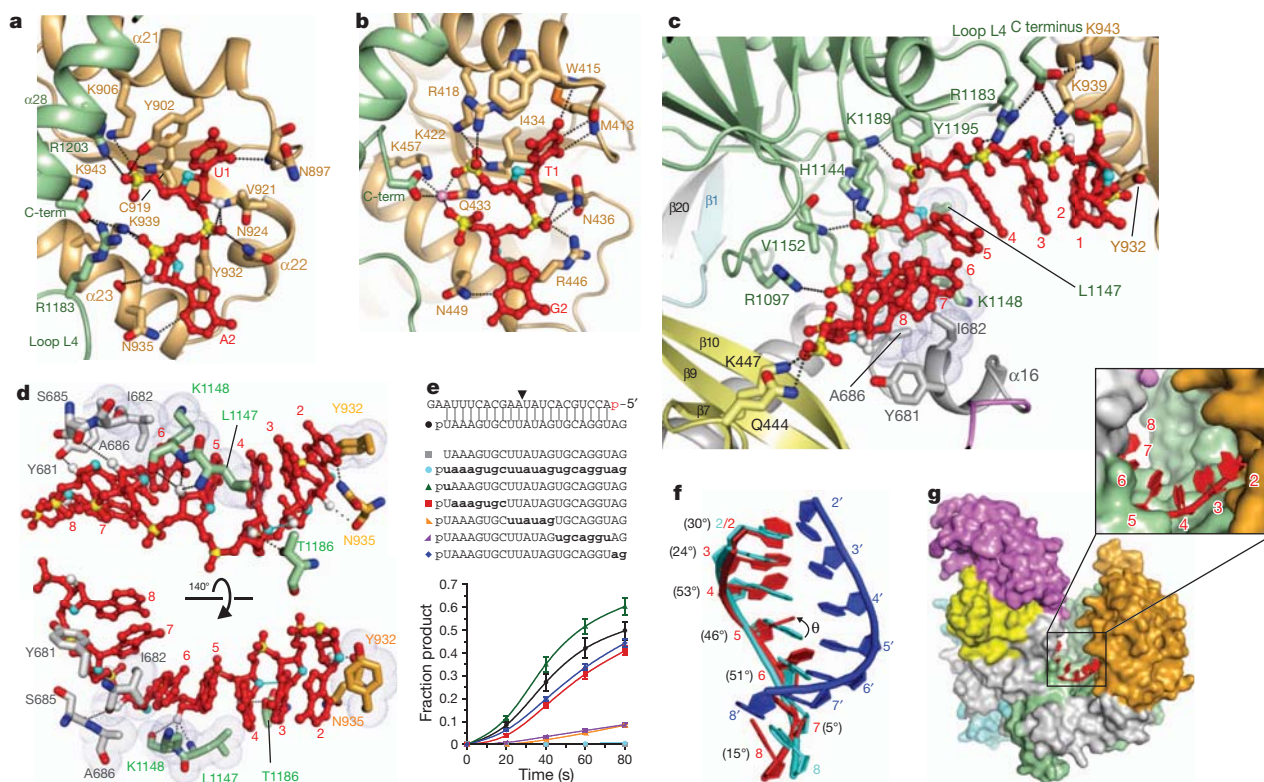
Figure 3 | Organization of the guide RNA. a, b, The 5′-nucleotide-binding pockets of KpAGO (**a**) and TtAGO[17] (**b**). Colours, as in Fig. 2a; protein, ribbon representation; highlighted residues and RNA, stick representation; O2′, O4′ and phosphate, white, cyan and yellow, respectively; hydrogen bonds, dotted lines. **c, d,** Interactions involving bases and either phosphates (**c**) or 2′-OH groups (**d**) of the seed region. Intermolecular (black) and intramolecular (blue) hydrogen bonds, dotted lines; hydrophobic interactions, van der Waals radii. **e,** Effects of guide-strand modifications on duplex loading and passenger-strand cleavage. KpAGO was incubated with siRNA duplexes with the indicated guide strands; p, 5′ monophosphate; upper case, 2′-OH; lower case bold, 2′-deoxy. The fraction of labelled passenger strand cleaved is plotted (average of three independent replicates; error bars, standard deviations; points connected by smooth curves). **f,** Superposition of guide-RNA nucleotides 2–8 (red) on A-form RNA (cyan and blue). Dihedral angles (θ) between guide-RNA bases and those of A-form RNA are in parentheses. **g,** Solvent-exposed seed nucleotides (red). KpAGO surface is rendered, domains coloured as in Fig. 2a.

for metazoan AGOs[31,35–38], although addition of Triton enabled sustained product formation (Supplementary Fig. 5f), perhaps by facilitating a conformational change that promotes product release.

Most copurifying RNAs mapped to the KpAGO expression plasmid (Fig. 2e and Supplementary Fig. 6) in a manner suggesting origins from siRNA-like duplexes loaded into KpAGO with subsequent passenger-strand cleavage (Supplementary Fig. 7). For this to occur, KpAGO must be able to load siRNA duplexes in the absence of RISC-loading factors. Indeed, purified KpAGO incubated with an siRNA duplex generated products diagnostic of passenger-strand cleavage (Fig. 2g) and formed active RISC able to slice cognate target RNA (Fig. 2h). Loading was more efficient with duplex than with single-stranded guide and occurred asymmetrically in a manner consistent with preference for 5′ uridine on the guide strand (Fig. 2h and Supplementary Fig. 8a–c).

We conclude that KpAGO can autonomously load an siRNA duplex, lose the passenger strand and then slice targets. This conclusion counters the prevailing view that loading of siRNA duplexes to form functional RISC requires RISC-loading factors[11]. We suspect that other AGOs might also autonomously load siRNA duplexes and that reports to the contrary resulted from assaying target-RNA slicing under conditions in which AGO retained inhibitory passenger-strand fragments. Autonomous loading explains how KpAGO RISC fortuitously formed in the absence of other RNAi proteins. In contrast to previous preparations of AGO complexes used for structural studies, the formation of KpAGO RISC through loading of a duplex resembles the physiological RISC-assembly pathway. From this perspective, the KpAGO structure reflects the natural state of eukaryotic RISC.

Organization of the guide RNA

Electron density corresponding to the base of nucleotide 1 was smaller than that corresponding to most other positions (Supplementary Fig. 9), which agreed with our sequencing results showing that KpAGO-bound RNAs were diverse but enriched for a 5′ uridine (Fig. 2e). Therefore, we modelled the first nucleotide as uridine and the next seven as adenine (the generic nucleotide used to minimize bias during refinement[39]) and refined the final structure as a KpAGO–pUAAAAAAAp binary complex (Supplementary Table 1 and Supplementary Fig. 10).

The guide-strand nucleotides 2–8 run along the nucleic-acid-binding channel, from the MID domain to the L2 domain. These nucleotides, including their bases, have electron-density quality resembling that of the KpAGO protein, even though this density represents a composite of thousands of different RNAs. Thus, for this segment of the guide RNA, known as the seed region, diverse RNA sequences are all presented in essentially the same orientation. The electron density disappeared after the ninth nucleotide (Fig. 2b, c), even though most copurifying RNAs were longer than nine nucleotides (Supplementary Fig. 5d). This density loss indicates that guide-RNA 3′ halves are either disordered or adopt diverse sequence-specific conformations. In addition, the PAZ domain is not well ordered, as observed in TtAGO complexes in which the PAZ domain has released the 3′ end of the guide[19], consistent with the idea that KpAGO holds the guide RNA without assistance from the PAZ domain.

Like prokaryotic AGOs[15,17–19,40], KpAGO recognizes the 5′ phosphate of the guide, the notable difference being that KpAGO uses the ammonium group of Lys 939 rather than a divalent cation (despite Mn[2+] in the crystallization buffer) to neutralize the negative charge

resulting from the close juxtaposition of the C-terminal carboxylate and phosphates 1 and 3 (Fig. 3a, b). The inserted carboxy terminus is anchored by Lys 939 and Lys 943 (Fig. 3a), with mutation of either residue impeding guide-RNA binding in *Drosophila* AGO1 (ref. 28). Another distinct facet involves Arg 1183, which hydrogen bonds with the C-terminal carboxylate and phosphate 4 (Fig. 3c). In the free *Neurospora crassa* QDE-2 (NcQDE-2) MID-PIWI structure[28] the analogous arginine is in a disordered loop, suggesting that guide RNA recruits Arg 1183 to the 5′-phosphate–binding pocket. Notably, conservation of Lys 939 and Arg 1183 is restricted to eukaryotic AGOs (Supplementary Fig. 1).

The Asn 897 main-chain amide interacts with the O2 carbonyl of the uridine at position 1 (Fig. 3a). Because analogous interactions with O2 of cytidine and N3 of purines would be isosteric, this hydrogen bond cannot explain the preference for a 5′ uridine. The preference might instead be attributed to the relatively weak stacking and pairing of uridine, which would facilitate the requisite flipping out of nucleotide 1 during siRNA loading[16,40].

KpAGO interacts with phosphates of the seed region primarily using contacts homologous to those observed in prokaryotic AGO complexes[16–19,40] (Fig. 3c). Structures of prokaryotic complexes, however, have not revealed intermolecular contacts to the guide-RNA 2′-OH groups. We find that KpAGO forms hydrogen bonds with most 2′-OH groups of the seed, using main-chain atoms at positions 2, 5 and 6 and hydroxyl groups of Thr 1186 and Tyr 681 at positions 4 and 7, respectively (Fig. 3d). We also observe an intra-RNA hydrogen bond between the 2′-OH group at position 3 and O4′ at position 4, a type of interaction proposed to facilitate base-pair fluctuations in A-form RNA helices[41]. A second intra-RNA hydrogen bond involves the 2′-OH group at position 1 and a non-bridging oxygen of phosphate 2, as previously observed in *Archaeoglobus fulgidus* PIWI–siRNA complex structures[16,40].

To examine the contributions of guide-strand 5′ phosphate and 2′-OH groups, we monitored autonomous loading and passenger-strand cleavage of modified siRNA duplexes. Removing the monophosphate or substituting all guide-strand 2′-OH groups with 2′-H (deoxy) greatly impaired activity (Fig. 3e and Supplementary Fig. 11), consistent with observations in transfected human cells[42,43]. To learn more about the 2′-OH groups contributing to this effect, we compared guide RNAs with deoxy substitutions at positions 1, 2–8, 9–14, 15–21 and 22–23. Substitution of the 2′-OH group at position 1 enhanced activity (perhaps by facilitating flipping out of nucleotide 1), whereas substitutions in all other regions impaired activity (Fig. 3e and Supplementary Fig. 11). Deoxy substitution at positions 2–8 impaired activity to a similar degree as at positions 22–23, which are presumably recognized by the PAZ domain[23–25]. Thus, the 2′-OH groups within the seed region contribute to duplex loading or passenger-strand cleavage. Nonetheless, greater effects were observed at positions 9–14 and 15–21, the understanding of which will require structural studies of additional states along the eukaryotic RISC-assembly pathway.

Together, contacts to the phosphate and 2′-OH groups maintain the sugar–phosphate backbone of the single-stranded guide-RNA seed in a near-A-form conformation resembling that of the siRNA duplex (Fig. 3f). Maintaining this conformation pre-organizes the seed backbone for pairing to the target, as anticipated from studies of microRNA targeting[1] and supported by structural and biophysical studies[16,40,44]. Also as anticipated, the bases of the seed nucleotides are stacked, with Watson–Crick faces (particularly those of nucleotides 2–4) displayed to solvent and accessible to nucleate pairing to target RNA (Fig. 3g).

The surprising feature of the guide-RNA conformation was the tilting of the bases away from the orientation required for helical pairing (Fig. 3f). KpAGO makes hydrophobic contacts with the bases at positions 2, 5 and 6 while anchoring the sugar–phosphate backbone (Fig. 3c, d). Base 2 packs against Tyr 932 (Fig. 3d), which is conserved as Tyr or Thr in eukaryotic AGOs (Supplementary Fig. 1) and thus might represent a conserved hydrophobic interaction that facilitates

the flipping out of nucleotide 1 by preventing its stacking on base 2. As observed in structures of prokaryotic AGO complexes[16–19,40], base 2 is recognized at N3 (purines) or O2 (pyrimidines) by the side chain of Asn 935 (Fig. 3d), which is conserved throughout all AGOs. Bases 5 and 6 are surrounded by a hydrophobic pocket comprising Ile 682, Ala 686, Leu 1147 and Lys 1148 (Fig. 3d). Bases 3 and 4 make no contact with KpAGO but are nonetheless tilted because of continuous stacking of the seed bases (Fig. 3c, d). Untilting of the seed stack, which would accompany nucleation of target pairing at positions 2–4, might disfavour contacts to Ile 682 and neighbouring residues, thereby facilitating repositioning of α16, a helix that would otherwise block full seed pairing. Such changes in base tilting and α16 might communicate the presence of target RNA.

Potentially unobstructed guide–target pairing

To compare the architectures of eukaryotic and prokaryotic AGOs, we structurally aligned each domain of KpAGO on its TtAGO counterpart. Except for the N domain, each of the domains superimposed well (Supplementary Fig. 12). The structural difference between the N domains is attributed to cS1, cS3 and vS2 (Fig. 4a, b). cS1 and cS3 cluster together with cS7 and cS10 (Supplementary Fig. 13a) such that they bury a space observed in prokaryotic AGO structures and concomitantly lengthen the nucleic-acid-binding channel[12,13,17–19] (Fig. 4c, d). These insertion segments interact with the L2-linker and PIWI domains through a hydrogen-bond network involving residues that are conserved throughout eukaryotic AGOs (Supplementary Figs 1 and 13a), which indicates that an extended nucleic-acid-binding channel is a general feature of eukaryotic AGOs.

In all crystallized conformations of TtAGO, the N domain blocks the channel and prevents propagation of guide–target pairing beyond position 16 (ref 19) (Fig. 4c and Supplementary Fig. 13c, d). In addition to lengthening the nucleic-acid-binding channel, the cS1/3/10 cluster positions the KpAGO N domain such that a slight widening of the channel would allow pairing to propagate to the 3′ end of the guide RNA (Fig. 4d and Supplementary Fig. 13b). The potential for unobstructed propagation of guide–target pairing is consistent with the prevalence of pairing throughout the 3′ region of plant small RNAs that guide target cleavage[45] and the contribution of such pairing to the stability of guide–target association *in vitro*[35,37].
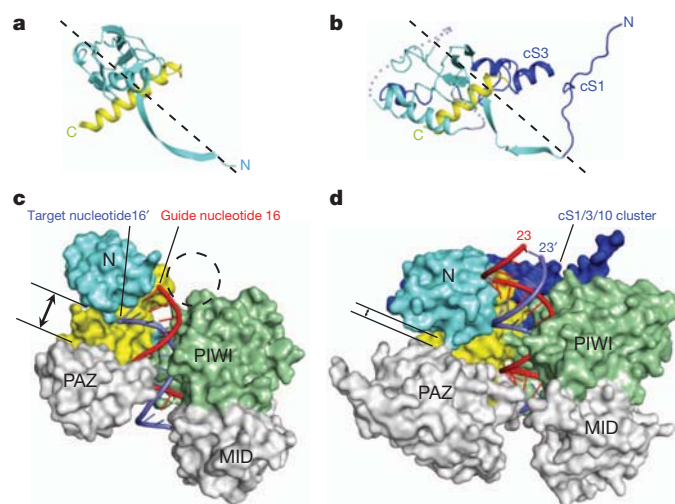
Figure 4 | An extended, potentially unobstructed nucleic-acid-binding channel in KpAGO. a, b, Position of N domain (cyan) relative to L1-linker domain (yellow) in TtAGO[19] (a) and KpAGO (b). Domains are oriented based on their N-terminal beta strands (dashed line connects strand termini). Colours, as in Fig. 2a. c, d, Channels of TtAGO ternary complex[19] (c) and KpAGO with modelled A-form duplex (d). Protein surfaces are rendered, highlighting distances between the N and PAZ domains (parallel lines) and the cS1/3/10 cluster (blue), which fills a cavity (dashed circle).

Glu 1013 completes a catalytic tetrad

When comparing the structures of KpAGO and the free NcQDE-2 MID-PIWI lobe[28], we observed notable differences in loops L1 and L2 (Supplementary Table 3). In KpAGO, loop L2 expands by partial unfolding of α25 (Fig. 5a) and packs into a cavity, such that the invariant Glu 1013 side chain inserts into the catalytic pocket, near the three Asp residues of the active site (Fig. 5b). This conformation is enabled by the movement of loop L1, which otherwise blocks access to the catalytic pocket (Fig. 5c). Opening of the loop L1 gate in KpAGO is accompanied by a conformational transition of cS11 and hydrophobic packing between aliphatic side chains on loop L1 and cS11. Notably, deletion of cS11 from *Drosophila* AGO1 inhibits guide-RNA binding and abolishes silencing activity[28].

The plugged-in conformation, in which the Glu 1013 finger is inserted into the catalytic pocket, is stabilized by an extensive hydrogen-bond network, with Glu 1013 bridging His 977 and Arg 1045, and loop L2 main-chain atoms interacting with Arg 1045 and Glu 1060 (Fig. 5b). These four residues are conserved throughout eukaryotic AGOs (and even most of the PIWI clade; Supplementary Fig. 1). Glu 1013, Arg 1045 and Glu 1060 are also conserved throughout prokaryotic AGOs, prompting a search for a similar plugged-in conformation in the available structures[12,13,17–19,28]. We found both plugged-in and unplugged conformations of TtAGO, with striking parallels to the eukaryotic hydrogen-bond network and

correlated loop movements (Fig. 5a–f). The plugged-in conformation was observed only in complexes in which the PAZ domain released the 3′ end of the guide and TtAGO assumed its catalytically active state (Supplementary Table 3). In contrast, the inactive states—either those of the apo proteins or complexes in which the PAZ domain engages the guide 3′ end—resembled the unplugged conformation. These observations indicate that the plugged-in conformation of loop L2 is correlated with release of the 3′ end of the guide and formation of active RISC. Supporting the functional importance of the plugged-in conformation, mutation of any of the four residues of the KpAGO hydrogen-bond network impaired RNAi (Fig. 5g).

The structures of ternary TtAGO complexes in the plugged-in conformation show the position of loop L2 in the context of guide and target strands. TtAGO loop L2 interacts with the guide DNA at positions 11–15 (Supplementary Table 3)[19]. Moreover, the carboxyl group of the glutamate finger approaches both the 2′-OH of the nucleotide adjacent to the scissile phosphate and one of the two active-site divalent metal ions (Fig. 5h), which indicates that the glutamate finger might act as a catalytic residue. Indeed, simultaneous coordination of the analogous 2′-OH and metal ion is the role of Glu 109 in the 'DEDD' catalytic tetrad at the active site of *Bacillus halodurans* RNase H1 (ref. 46). Although the PIWI domain of AGO has an RNase H fold, only a conserved 'DDX' catalytic triad (where 'X' is generally Asp or His) had been recognized in AGOs with slicer
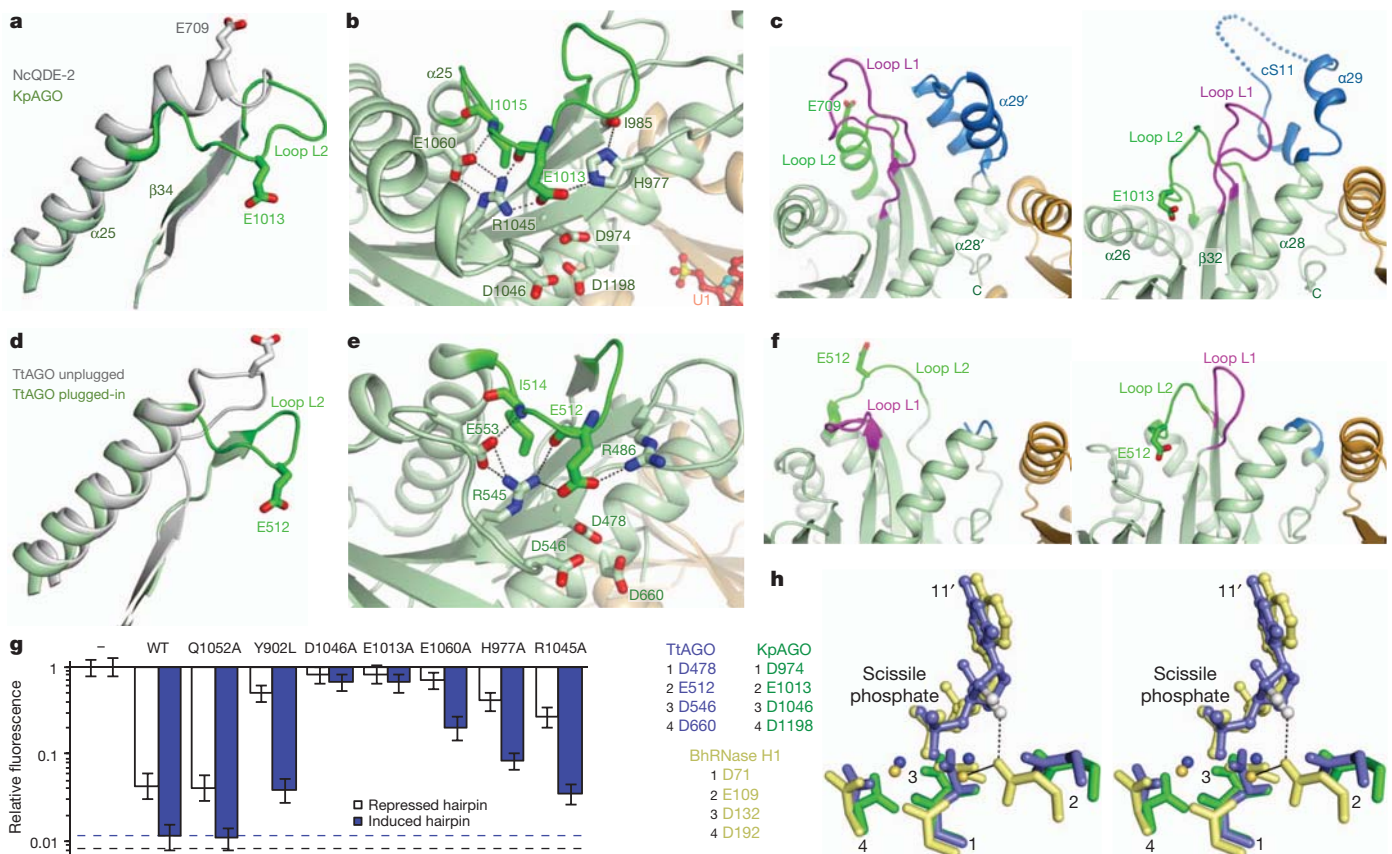
Figure 5 | A plugged-in glutamate finger at the active site. a, Superposition of α25 and β34 of KpAGO (green) on counterparts of NcQDE-2 MID-PIWI lobe (grey), highlighting the extended loop L2 (dark green). **b**, Hydrogen-bond network stabilizing the plugged-in loop L2. Loop L2, dark green; otherwise, as in Fig. 3a. **c**, Closed (left) and open (right) configurations of the loop L1 gate (purple) in NcQDE-2 MID-PIWI lobe and KpAGO, respectively. cS11, blue; otherwise, as in panel **b**. **d**, Superposition of the region flanking loop L2 in the unplugged (grey) and plugged-in (green) conformations of TtAGO, depicted as in panel **a**. **e**, Hydrogen-bond network stabilizing the plugged-in loop L2 in TtAGO, depicted as in panel **b**. **f**, Closed and open configurations of the loop L1 gate in the unplugged and plugged-in conformations of TtAGO, respectively,

depicted as in panel **c**. **g**, RNAi reconstituted in *S. cerevisiae* using wild-type (WT) *K. polysporus AGO1* or genes with the indicated substitutions. Silencing was monitored under either permissive (induced hairpin, blue bars) or stringent (repressed hairpin, open bars) conditions. Q1052 and Y902, conserved residues insensitive and sensitive to substitution, respectively[28], were included as controls. Dashed lines (blue and black), background fluorescence (permissive and stringent conditions, respectively); otherwise, as in Fig. 1b. **h**, Stereoview of KpAGO catalytic residues (green) superpositioned with catalytic residues, divalent cations, scissile phosphate and adjacent nucleoside in TtAGO (blue) and *B. halodurans* RNase H1 (BhRNase H1, yellow) ternary complexes.

activity[47,48]. On the basis of analogy to RNase H, a fourth catalytic residue had been suspected, but previous searches for this missing component had focused on the residues corresponding to Arg 1045 and Glu 1060 (refs 12, 13, 31, 46 and 47), whose conservation and proximity to the catalytic pocket are now explained instead by their roles in stabilizing the plugged-in conformation (Fig. 5b). In support of the glutamate finger as the missing catalytic residue that helps to coordinate an active-site metal ion (either directly or through outer-sphere contacts), the putative DEDD catalytic tetrads in the plugged-in conformations of both TtAGO and KpAGO are nearly isosteric with the RNase H DEDD tetrad (Fig. 5h). Moreover, when we assayed RNAi, only mutation of Glu 1013 abrogated RNAi to the extent observed for mutation of Asp 1046, a previously identified active-site residue (Fig. 5g). Thus, we propose that the glutamate finger constitutes the second residue of a universally conserved RNase-H-like DEDX catalytic tetrad at the active site of slicing AGOs.

Our new insights suggest the following model for AGO loading and catalysis. The apo protein in the unplugged conformation binds the siRNA duplex, in part using contacts between the two-nucleotide overhang of the guide strand and the PAZ domain. As the duplex loads and the 3′ end of the guide stand is released from the PAZ domain, the glutamate finger inserts into the active site, thereby completing the DEDX catalytic tetrad to enable cleavage of the passenger strand. After discarding the passenger-strand fragments, the resulting RISC remains in a plugged-in conformation resembling that of the current structure and is competent to bind and cleave suitably paired target RNAs.

While our manuscript was in review, a structure of human AGO2 (HsAGO2; also known as EIF2C2) with RNA of unknown biochemical origin and function was reported[49]. The authors noted many contacts to the RNA 5′ monophosphate and sugar–phosphate backbone analogous to those of KpAGO. Our inspection of the HsAGO2 structure further revealed that it also has an extended nucleic-acid-binding channel, an N domain positioned to allow unobstructed guide–target pairing, and a plugged-in glutamate finger that completes a DEDH catalytic tetrad (Supplementary Fig. 14). These similarities indicate that the HsAGO2 structure (for which conserved residues were mutated to improve diffraction) has features of active RISC and that studies of KpAGO will continue to provide insights relevant to metazoan AGOs.

In contrast to RNase H, which forms its active site during initial folding, AGO requires a conformational change to form its active site. What might explain this difference between these two related RNases? The constitutive active site of RNase H is well-suited to its role in nonspecifically cleaving RNA–DNA hybrids, whereas proper AGO function requires high specificity. Coupling siRNA duplex loading (in part through recognition by the PAZ domain) with active-site formation imparts specificity to AGO, thereby preventing it from cleaving any base-paired RNA. After passenger-strand cleavage and removal, activity of the licensed AGO is restricted by its guide RNA. In this way, AGO activity is tightly controlled and spurious endonucleolytic cleavage is prevented. The previous view was that among proteins adopting the RNase H fold, RNase H enzymes were unique in having a catalytic tetrad, whereas the related endonucleases of this protein superfamily (including AGO) were missing the active-site residue corresponding to Glu 1013 (ref. 48). Our findings revising this view imply that some other proteins for which only a catalytic triad has hitherto been identified (for example, bacterial UvrC DNA repair protein) might also use the conditional insertion of a 'missing' catalytic residue to impart specificity.

METHODS SUMMARY

KpAGO was overexpressed in *Escherichia coli* as a His–SUMO-tagged fusion. Native and SeMet-substituted crystals were obtained by sitting-drop vapour diffusion at 20 °C. The phase was determined by the single-wavelength anomalous dispersion method with selenium anomalous signals. Cleavage assays were performed with synthetic or transcribed RNA in 30 mM Tris-HCl pH 7.5, 130 mM KCl, 10 mM NaCl, 1.1 mM MgCl$_2$, 0.1 mM EDTA, 1.3 mM dithiothreitol (DTT) and 5% glycerol, including 0.1% Triton X-100 where indicated. Yeast manipulations, *in vivo* assays and high-throughput sequencing were essentially as described previously[29,32]. Details of all procedures are listed in Methods.

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

1. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116,** 281–297 (2004).
2. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431,** 343–349 (2004).
3. Malone, C. D. & Hannon, G. J. Small RNAs as guardians of the genome. *Cell* **136,** 656–668 (2009).
4. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409,** 363–366 (2001).
5. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15,** 188–200 (2001).
6. Hutvágner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293,** 834–838 (2001).
7. Matranga, C., Tomari, Y., Shin, C., Bartel, D. P. & Zamore, P. D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123,** 607–620 (2005).
8. Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. & Siomi, M. C. Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev.* **19,** 2837–2848 (2005).
9. Rand, T. A., Petersen, S., Du, F. & Wang, X. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123,** 621–629 (2005).
10. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **19,** 517–529 (2005).
11. Carthew, R. W. & Sontheimer, E. J. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136,** 642–655 (2009).
12. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305,** 1434–1437 (2004).
13. Yuan, Y. R. *et al.* Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell* **19,** 405–419 (2005).
14. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305,** 1437–1441 (2004).
15. Parker, J. S., Roe, S. M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **23,** 4727–4737 (2004).
16. Ma, J. B. *et al.* Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434,** 666–670 (2005).
17. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456,** 209–213 (2008).
18. Wang, Y. *et al.* Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456,** 921–926 (2008).
19. Wang, Y. *et al.* Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461,** 754–761 (2009).
20. Parker, J. S. How to slice: snapshots of Argonaute in action. *Silence* **1,** 3 (2010).
21. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4,** 29 (2009).
22. Meister, G. *et al.* Identification of novel argonaute-associated proteins. *Curr. Biol.* **15,** 2149–2155 (2005).
23. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature* **426,** 465–469 (2003).
24. Yan, K. S. *et al.* Structure and conserved RNA binding of the PAZ domain. *Nature* **426,** 468–474 (2003).
25. Ma, J. B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429,** 318–322 (2004).
26. Boland, A., Tritschler, F., Heimstädt, S., Izaurralde, E. & Weichenrieder, O. Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep.* **11,** 522–527 (2010).
27. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465,** 818–822 (2010).
28. Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E. & Weichenrieder, O. Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proc. Natl Acad. Sci. USA* **108,** 10466–10471 (2011).
29. Drinnenberg, I. A. *et al.* RNAi in budding yeast. *Science* **326,** 544–550 (2009).
30. Weinberg, D. E., Nakanishi, K., Patel, D. J. & Bartel, D. P. The inside-out mechanism of Dicers from budding yeasts. *Cell* **146,** 262–276 (2011).
31. Rivas, F. V. *et al.* Purified Argonaute2 and an siRNA form recombinant human RISC. *Nature Struct. Mol. Biol.* **12,** 340–349 (2005).
32. Addo-Quaye, C., Eshoo, T. W., Bartel, D. P. & Axtell, M. J. Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr. Biol.* **18,** 758–762 (2008).
33. German, M. A. *et al.* Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nature Biotechnol.* **26,** 941–946 (2008).

34. Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20,** 6877–6888 (2001).

35. Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nature Struct. Mol. Biol.* **11,** 599–606 (2004).

36. Martinez, J. & Tuschl, T. RISC is a 5′ phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18,** 975–980 (2004).

37. Ameres, S. L., Martinez, J. & Schroeder, R. Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* **130,** 101–112 (2007).

38. Förstemann, K., Horwich, M. D., Wee, L., Tomari, Y. & Zamore, P. D. *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130,** 287–297 (2007).

39. Frazão, C. *et al.* Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature* **443,** 110–114 (2006).

40. Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434,** 663–666 (2005).

41. Pan, Y. & MacKerell, A. D. Jr. Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. *Nucleic Acids Res.* **31,** 7131–7140 (2003).

42. Schwarz, D. S., Hutvágner, G., Haley, B. & Zamore, P. D. Evidence that siRNAs function as guides, not primers, in the *Drosophila* and human RNAi pathways. *Mol. Cell* **10,** 537–548 (2002).

43. Chiu, Y. L. & Rana, T. M. siRNA function in RNAi: a chemical modification analysis. *RNA* **9,** 1034–1048 (2003).

44. Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol. Cell* **33,** 204–214 (2009).

45. Mallory, A. C. *et al.* MicroRNA control of *PHABULOSA* in leaf development: importance of pairing to the microRNA 5′ region. *EMBO J.* **23,** 3356–3364 (2004).

46. Nowotny, M., Gaidamakov, S. A., Crouch, R. J. & Yang, W. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121,** 1005–1016 (2005).

47. Hall, T. M. Structure and function of argonaute proteins. *Structure* **13,** 1403–1408 (2005).

48. Nowotny, M. Retroviral integrase superfamily: the structural perspective. *EMBO Rep.* **10,** 144–151 (2009).

49. Schirle, N. T. & Macrae, I. J. The crystal structure of human Argonaute2. *Science* http://dx.doi.org/10.1126/science.1221551 (26 April 2012).

METHODS

Protein purification. DNA encoding *K. polysporus* AGO1(Thr 207–Ile 1251) was cloned into a modified pRSFDuet vector (Novagen) containing an amino-terminal Ulp1-cleavable His$_6$–SUMO tag. Protein was overexpressed in *E. coli* BL21(DE3) Rosetta2 (Novagen). Cell extract was prepared using a French press in buffer A (10 mM phosphate buffer pH 7.3, 1.5 M NaCl, 25 mM imidazole, 10 mM β-mercaptoethanol, 1 mM phenylmethylsulphonyl fluoride) and cleared by centrifugation. The supernatant was loaded onto a nickel column (GE Healthcare) and then washed with buffer A. The target protein was eluted with a linear gradient of 0.025–1.5 M imidazole. After mixing with Ulp1 protease, the eluted sample was dialysed against buffer B (10 mM phosphate buffer pH 7.3, 500 mM NaCl, 20 mM imidazole, 10 mM β-mercaptoethanol) overnight. The digested protein was loaded onto a nickel column to remove the cleaved His$_6$–SUMO tag. The flow-through sample was dialysed against buffer C (5 mM phosphate buffer pH 7.3, 10 mM β-mercaptoethanol) and then loaded onto an SP column (GE Healthcare). The protein was eluted with a linear gradient of 0.0–2.0 M NaCl, mixed with ammonium sulphate (2 M final concentration) and then centrifuged. The supernatant was loaded onto a phenyl-Sepharose hydrophobic interaction column (GE Healthcare) in buffer D (10 mM phosphate buffer pH 7.3, 2 M ammonium sulphate, 10 mM β-mercaptoethanol), and the protein was eluted with a linear gradient of 2.0–0.0 M ammonium sulphate. The eluted protein was dialysed against buffer E (300 mM sodium dihydrogen phosphate, 10 mM β-mercaptoethanol) and then loaded onto a MonoQ column (GE Healthcare) in buffer E. The protein was eluted with a linear gradient of 0.0–2.0 M NaCl. The eluted sample was concentrated by ultrafiltration and loaded onto a HiLoad 200 16/60 column (GE Healthcare) in buffer F (10 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM DTT). Purified KpAGO was concentrated to approximately 40 mg ml^{-1} using ultrafiltration and stored at −80 °C in protein storage buffer (10 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM DTT).

Structure determination and refinement. Initial crystals of recombinant KpAGO diffracted poorly but could be improved by addition of 1,4-dioxane to the crystallization buffer. Native crystals of KpAGO were obtained at 20 °C by sitting-drop vapour diffusion in 100 mM MIB buffer pH 5.0 (molar ratio, 2 Na-malonate:3 imidazole:3 boric acid), 3% 1,4-dioxane, 19% PEG3350, 12 mM MnCl$_2$ and 3% ethanol. SeMet-substituted crystals were grown at 20 °C by sitting-drop vapour diffusion in 100 mM MIB pH 5.0, 3% 1,4-dioxane, 19% PEG3350, 12 mM MnCl$_2$, 3% ethanol and 9 mM sarcosine. The native and SeMet-substituted crystals of KpAGO were soaked in collection buffer (1.2-fold concentrated reservoir solution) and cryoprotected with 20% glycerol. Both derivative data sets were collected at the Advanced Photon Source NE-CAT beamlines. Data were processed with HKL2000[50]. Data collection and refinement statistics are listed (Supplementary Table 1). A total of 33 selenium sites were found using peak data with HKL2MAP[51] and were used for phase calculation at 4.2 Å resolution with Phaser-EP[52]. The initial phases were improved by solvent flattening, electron density histogram and non-crystallographic symmetry averaging with Parrot and DM[53]. The initial model was built manually with Coot[54] and was improved by iterative cycles of refinement with Phenix[55]. Molecular replacement was performed with MOLREP[56] using the SeMet structure as a search model. The final model was improved using the native data processed at 3.2 Å. The Ramachandran plot analysis by PROCHECK[53] showed 82.0%, 17.3% and 0.8% of the protein residues in the most favourable, additionally allowed and generously allowed regions, respectively, with no residues in disallowed regions. The simulated-annealing omit map was calculated by CNS[57]. All figures of structures were generated with PyMOL[58].

RNAs. A list of RNA oligonucleotide sequences is provided (Supplementary Table 4). To generate cap-labelled target RNAs, RNA was transcribed *in vitro* with T7 RNA polymerase using DNA oligonucleotide templates. DNase-treated transcripts were purified on a denaturing gel and capped using the ScriptCap m^{7}G Capping System (CellScript) according to the manufacturer's directions, except that high-specific-activity RNA was prepared by omitting GTP and including 5 μl [α-^{32}P]GTP (6,000 Ci mmol^{-1}), and low-specific-activity RNA was prepared by using a 1,500:1 molar ratio of GTP:[α-^{32}P]GTP (6,000 Ci mmol^{-1}). Cap-labelled RNA was gel-purified and quantified by scintillation counting, and 10× stocks were prepared in water supplemented with 1 μM DNA carrier oligonucleotide.

5′-phosphorylated guide RNA and its 2′-deoxy-substituted variants were chemically synthesized (IDT) and gel purified. To prepare 5′ end-labelled RNAs, 5′-OH RNAs were chemically synthesized (Dharmacon), deprotected, purified on a denaturing gel, phosphorylated with [γ-^{32}P]ATP (6,000 Ci mmol^{-1}) using T4 polynucleotide kinase (PNK, NEB) and gel purified again. To prepare 3′ end-labelled RNAs, 5′-phosphorylated RNAs lacking the terminal nucleotide (that is, 22-nucleotide variants) were chemically synthesized (IDT), gel purified, extended using cordycepin 5′-[α-^{32}P]triphosphate (5,000 Ci mmol^{-1}) and yeast poly(A) polymerase (USB) and gel purified again.

siRNA duplexes were prepared by annealing synthetic ssRNAs. Complementary RNAs designed to hybridize to generate 21-base-pair duplexes with two-nucleotide 3′ overhangs, were combined (using at least threefold excess unlabelled RNA) in dsRNA annealing buffer (30 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA) and slow-cooled from 90 °C to room temperature over >2 h. Annealed RNAs were separated from ssRNAs on native 20% polyacrylamide gels, and duplexes were eluted from gel slices in 0.3 M NaCl overnight at 4 °C, ethanol-precipitated and stored in dsRNA storage buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.1 mM EDTA). RNA was quantified by scintillation counting, and 10× stocks were prepared in dsRNA storage buffer supplemented with 1 μM DNA carrier oligonucleotide.

AGO activity assays. For all biochemical assays, KpAGO was diluted and stored at −20 °C in protein dilution buffer (5 mM Tris-HCl pH 7.5, 100 mM NaCl, 2.5 mM DTT, 50% glycerol). The concentration of KpAGO was determined by absorbance at 280 nm. For the slicing assay in Fig. 1c, 1.1 μM KpAGO was pre-incubated with 110 nM guide RNA in 1.1× reaction buffer (1× reaction buffer: 30 mM Tris-HCl pH 7.5, 130 mM KCl, 1.1 mM MgCl$_2$, 1 mM DTT, 0.1 mM EDTA) for 1 h at 25 °C. To initiate the slicing reaction, 1 μl cap-labelled target RNA (final concentration, 200 nM) was added to 9 μl of the pre-incubated mixture. Reactions were incubated at 30 °C, and 3-μl aliquots were removed at the indicated time and quenched by addition to 12 μl formamide loading buffer (95% formamide, 18 mM EDTA, 0.025% sodium dodecyl sulphate, 0.025% xylene cyanol, 0.025% bromophenol blue). The slicing assay in Fig. 2h was conducted similarly except pre-incubation was performed with 110 nM KpAGO and 50 pM guide RNA, and target was subsequently added to a final concentration of 100 pM. The slicing assay in Fig. 2f was guided by copurifying RNA and thus did not involve pre-incubation. These reactions contained 1× reaction buffer supplemented with 0.1% Triton X-100, 100 nM KpAGO (or an equal volume of protein dilution buffer) and 100 pM target RNA. Reactions were incubated at 30 °C, and 5 μl aliquots were removed at the indicated time and quenched by addition to 15 μl formamide loading buffer. The passenger-strand cleavage reactions in Fig. 3e contained 1× reaction buffer, 100 μg ml^{-1} Ultrapure BSA (Ambion), 10 nM KpAGO and 50 pM substrate. All other passenger-strand cleavage reactions contained 1× reaction buffer supplemented with 0.1% Triton X-100, 100 nM KpAGO (or an equal volume of protein dilution buffer) and 50 pM substrate. Reactions were incubated at 30 °C, and 5 μl aliquots were removed at the indicated time and quenched by addition to 10–15 μl formamide loading buffer.

To monitor cleavage, RNAs were resolved on denaturing (7.5 M urea) polyacrylamide gels (15% gel for target cleavage using synthetic guide RNA, 20% for target cleavage using copurifying guide RNA or 22.5% for passenger-strand cleavage), and radiolabelled products were visualized by phosphorimaging (Fujifilm BAS-2500) and quantified using Multi Gauge (Fujifilm). For kinetic analyses, at each time point (*t*) the fraction product was measured as F_P = product/(product + substrate). Data in Fig. 3e were fit with a smoothed curve using the cubic spline method implemented in KaleidaGraph.

Analysis of copurifying RNA. To avoid loss of especially small fragments, polynucleotides were extracted without subsequent precipitation. For analysis of pooled crystals, approximately 100 crystals were collected, stored in harvest buffer (100 mM MIB buffer pH 5.0, 3% 1,4-dioxane, 20% PEG 3350, 12 mM MnCl$_2$, 3% ethanol, 6 mM sarcosine, 25% glycerol) and immediately frozen. After thawing, the mixture was diluted with an equal volume of water and extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) followed by extraction with chloroform. The aqueous phase was retained and diluted 1:20 in water for use in labelling reactions or used undiluted to prepare sequencing libraries. For analysis of soluble protein, polynucleotides were similarly extracted from 1.5 nmol KpAGO. For analysis of individual crystals, each single crystal was collected, stored in 1 mM EDTA and immediately frozen. After thawing, the mixture was heated at 90 °C for 3 min, chilled on ice for 5 min and extracted with an equal volume of phenol:-chloroform:isoamyl alcohol (25:24:1, Sigma) followed by two extractions with chloroform. The aqueous phase was retained and used undiluted in labelling reactions.

Prior to 5′ labelling, polynucleotides were dephosphorylated in a 20 μl reaction containing 2 μl diluted polynucleotides (or water) and 1× PNK buffer (NEB) in the presence or absence of 2 units of thermosensitive alkaline phosphatase (TSAP, Promega) for 30 min at 37 °C. To inactivate TSAP, the reaction was quenched with 1 μl 220 mM EDTA and incubated at 74 °C for 15 min. 5′ phosphorylation was performed in a 30 μl reaction containing 21 μl heat-inactivated TSAP reaction, 3 units T4 PNK (NEB), 0.04 μl [γ-^{32}P]ATP (8,000 Ci mmol^{-1}), 0.8 μl 10× PNK buffer and 1 μl 240 mM MgCl$_2$ for 1 h at 37 °C. Reactions were quenched with an equal volume of 2× urea loading buffer, and products were resolved on a denaturing 22.5% polyacrylamide gel. For analysis of nuclease sensitivity, 15 μl aliquots were removed from PNK reactions and incubated with 2 μl RNase I (Ambion) or RQ1 RNase-free DNase (Promega) for 30 min at 37 °C before gel analysis.

To monitor preparation of sequencing libraries, trace amounts of synthetic 3′-pCp[5′-^{32}P]-labelled 7- and 23-nucleotide RNA internal standards were added to 2 μl undiluted polynucleotides isolated from soluble or crystalline KpAGO (or a water-only mock control). Dephosphorylation was performed in a 30 μl reaction containing 3 units TSAP (Promega) and 1× PNK buffer (NEB) supplemented with 2 μl manganese-chelating mix (10 mM MgCl$_2$, 10 mM EDTA) for 30 min at 37 °C. To inactivate TSAP, the reaction was quenched with 1.5 μl 240 mM EDTA and incubated at 74 °C for 15 min. RNA was ligated to pre-adenylated adaptor DNA in a 50 μl reaction containing 32 μl heat-inactivated TSAP reaction, 100 pmol adaptor DNA[59], 45 units T4 RNA ligase 1 (NEB), 10% PEG8000 (NEB), 2 μl 10× PNK buffer and 1 μl 390 mM MgCl$_2$ for 2.5 h at room temperature. After phenol extraction and precipitation, 28–50-nucleotide ligation products were gel-purified and 5′ phosphorylated in a 50 μl reaction containing 20 units T4 PNK (NEB) and 1× PNK buffer supplemented with 1 μl [γ-^{32}P]ATP (6,000 Ci mmol^{-1}) for 30 min at 37 °C, followed by a chase with 10 μl cold reaction mixture (1× PNK buffer, 28 units T4 PNK, 6 mM ATP) and incubation for an additional 30 min at 37 °C. After desalting, phenol extraction and precipitation, RNA was ligated to a 5′-adaptor RNA, gel-purified, converted to complementary DNA, amplified 10 cycles (soluble) or 12 cycles (crystalline) and sequenced using the Illumina SBS platform. The library prepared without input polynucleotides did not yield an observable PCR product, indicating minimal contamination from polynucleotides that might copurify with the enzymes used for library construction.

Sequencing reads were filtered by requiring that they contain a perfect match to the first 12 nucleotides of the 3′ adaptor and that every nucleotide up to the beginning of the 3′ adaptor have a Phred+64 quality score of at least '∧'. After removing the internal-standard reads and trimming away the adaptor sequences, reads representing the small RNAs were collapsed to a non-redundant set of 8–24-nucleotide sequences. To examine the origins of the copurifying RNAs, 15–24-nucleotide sequences were mapped sequentially to the KpAGO expression plasmid, the chloramphenicol-resistance gene found on pRARE2, and the BL21(DE3) genome[60], allowing no mismatches and recovering all hits. (The 15-nucleotide lower bound for mapping was chosen because this was the minimum read length that achieved a <1% genome-mapping rate for random or shuffled small-RNA sequences.) Because there were fewer fortuitous matches to the KpAGO expression plasmid, analysis of 12-nucleotide sequences was performed on reads that mapped to the plasmid. For mapping-independent analyses, sequences with <10 reads were not considered.

For analysis of nucleotide composition, information content was calculated by determining the relative frequency of each nucleotide at position X compared to the relative frequency at all other positions combined. The selectivity for a given nucleotide n at position X was calculated using the following equation:

$$S_n = [\Sigma_{i=A,U,G,C}([f(i,X)/f(n,X)]/[f(i,\sim X)/f(n,\sim X)])]^{-1}$$

where $f(i,X)$ is the frequency of nucleotide i at position X and $f(i,\sim X)$ is the frequency of nucleotide i at all other positions.

Information content scores were then calculated using the following equation:

$$I_n = S_n \times [\log_2(S_n) + 2]$$

For phasing analysis, the frequency of distances separating 5′-end pairs (i, j) mapping to opposite DNA strands was calculated using the following equation:

$$\text{Frequency}_D = \Sigma_{i,j} \min(\text{Reads}_i, \text{Reads}_j)_D$$

where D = (distance between small-RNA 5′ ends) + 1

Yeast manipulations. *S. castellii* and *S. cerevisiae* were grown at 25 °C and 30 °C, respectively, on standard *S. cerevisiae* plate and liquid media (for example, YPD and SC). Transformations of *S. castellii* were performed as described previously[29]. Transformations of *S. cerevisiae* were performed as described[61]. For FACS analyses, strains were inoculated in SC, in either non-inducing (2% glucose) or inducing (1% galactose and 1% raffinose) conditions, and grown overnight. Fresh cultures were then seeded from the overnight cultures, and cells were grown to log phase. Cells were analysed using FACSCalibur (BD Biosciences); data were processed with CellQuest Pro (BD Biosciences) and FlowJo (Tree Star).

Plasmids and strains used and generated in this study are listed (Supplementary Tables 5 and 6). Vectors pRS404CYC1-KpAGO1 and pRS405TEF-KpDCR1 were constructed by insertion of the coding sequencing of the respective *K.*

polysporus genes between the *CYC1* or *TEF* promoter and *CYC1* terminator (cloned from p416CYC or p416TEF[62]) of the appropriate vector[63] using SpeI and XhoI sites (KpAGO1) or BamHI and XhoI sites (KpDCR1). Vector pRS404CYC1-KpAGO1(207–1251) was constructed similarly, with the insertion of an 'ATG' codon upstream of amino acid 207. Vector pRS404CYC1-FLAG$_3$-KpAGO1 was generated by PCR-based insertion of the sequencing encoding the Flag$_3$ epitope downstream of the 'ATG' codon of pRS404CYC1-KpAGO1. Point mutations were introduced by PCR-based mutagenesis to generate vectors encoding mutant Flag-tagged Ago1. pRS402GPD-GFP(S65T) was constructed by insertion of the coding sequence of GFP(S65T) (amplified from pFA6a-GFP(S65T)-kanMX6[64]) between the *GPD* promoter and *CYC1* terminator (cloned from p416GPD[62]) of pRS402[63] using SpeI and XhoI sites. To reconstitute RNAi in *S. cerevisiae*, GFP(S65T), KpAGO1 and KpDCR1 expression vectors were integrated into W303-1B variants already containing other components of the GFP-silencing system[29], using standard protocols[61]. To generate *S. castellii* strains DPB267 and DPB268 for degradome sequencing, *XRN1* was deleted in DPB005 and DPB007, respectively, using the kanMX6 cassette[64].

Degradome sequencing and analysis. Total RNA was isolated from mid-log phase ($D_{600} \approx 0.6$) cultures of strains DPB267 and DPB268 using the hot-phenol method. Degradome libraries were constructed from 5 μg poly(A)$^+$ RNA essentially as described[32] and sequenced on the Illumina SBS platform. After removing adaptor sequences and generating each reverse complement, reads representing degradome-cleavage tags were collapsed to a non-redundant set. To analyse tags deriving from Y′-element loci, 20–21-nucleotide sequences were mapped to a consensus *S. castellii* Y′ element as described previously[29], and 49% of reads in the *AGO1* library were randomly sampled (to normalize for higher sequencing yield, Supplementary Fig. 2b) and used for subsequent analyses. Mapping data were then used to generate a single-nucleotide-resolution plot of the consensus Y′ element. For phasing analysis, the frequency of distances separating opposite-strand pairs of 5′ ends of 20–21-nucleotide degradome tags (i), and 5′ ends of 22–23-nucleotide small RNAs (j) was calculated using the following equation:

$$\text{Frequency}_D = \Sigma_{i,j} \text{Reads}_i \times \text{Reads}_j/\text{Norm}_i$$

where D = position of 5′ end of degradome tag with respect to 5′ end of small RNA, and Norm$_i$ = number of reads for all small RNAs in which the 5′ end of degradome tag i falls. Fractional frequencies were calculated for each D by dividing Frequency$_D$ by the total number of reads corresponding to degradome-tag 5′ ends that map opposite 22–23-nucleotide small RNAs.

50. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276,** 307–326 (1997).
51. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37,** 843–844 (2004).
52. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
53. Collaborative Computational Project. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **50,** 760–763 (1994).
54. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
55. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr.* **58,** 1948–1954 (2002).
56. Vagin, A. & Teplyakov, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr.* **56,** 1622–1624 (2000).
57. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.* **54,** 905–921 (1998).
58. DeLano, W. L. & Lam, J. W. PyMOL: A communications tool for computational models. *Abstracts of Papers of the American Chemical Society* **230** (2005).
59. Grimson, A. *et al.* Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455,** 1193–1197 (2008).
60. Jeong, H. *et al.* Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol.* **394,** 644–652 (2009).
61. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols* **2,** 31–34 (2007).
62. Mumberg, D., Muller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156,** 119–122 (1995).
63. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122,** 19–27 (1989).
64. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14,** 953–961 (1998).

LETTER

An abundance of small exoplanets around stars with a wide range of metallicities

Lars A. Buchhave[1,2], David W. Latham[3], Anders Johansen[4], Martin Bizzarro[2], Guillermo Torres[3], Jason F. Rowe[5], Natalie M. Batalha[6], William J. Borucki[7], Erik Brugamyer[8], Caroline Caldwell[8], Stephen T. Bryson[7], David R. Ciardi[9], William D. Cochran[8], Michael Endl[8], Gilbert A. Esquerdo[3], Eric B. Ford[10], John C. Geary[3], Ronald L. Gilliland[11], Terese Hansen[1], Howard Isaacson[12], John B. Laird[13], Philip W. Lucas[14], Geoffrey W. Marcy[12], Jon A. Morse[15], Paul Robertson[8], Avi Shporer[16,17], Robert P. Stefanik[3], Martin Still[18] & Samuel N. Quinn[3]

The abundance of heavy elements (metallicity) in the photospheres of stars similar to the Sun provides a 'fossil' record of the chemical composition of the initial protoplanetary disk. Metal-rich stars are much more likely to harbour gas giant planets[1–4], supporting the model that planets form by accumulation of dust and ice particles[5]. Recent ground-based surveys suggest that this correlation is weakened for Neptunian-sized planets[4,6–9]. However, how the relationship between size and metallicity extends into the regime of terrestrial-sized exoplanets is unknown. Here we report spectroscopic metallicities of the host stars of 226 small exoplanet candidates discovered by NASA's Kepler mission[10], including objects that are comparable in size to the terrestrial planets in the Solar System. We find that planets with radii less than four Earth radii form around host stars with a wide range of metallicities (but on average a metallicity close to that of the Sun), whereas large planets preferentially form around stars with higher metallicities. This observation suggests that terrestrial planets may be widespread in the disk of the Galaxy, with no special requirement of enhanced metallicity for their formation.

In February 2011, the Kepler mission[10] announced its discovery of 1,235 planet candidates, of which more than half have radii smaller than that of Neptune[11]: $R_P < 4R_\oplus$, where $R_\oplus$ is the Earth radius. We used reconnaissance spectra obtained by the Kepler Follow-up Observing Program (FOP) to derive metallicities for several hundred of the brighter planet candidates, and used the results to explore the relationship between planet size and host-star metallicity. Metallicity, denoted [m/H], is defined as the proportion of a star's outer layers made up of chemical elements other than hydrogen and helium and expressed on a logarithmic scale where zero is the Sun's metallicity. Thousands of spectra have been gathered by the Kepler FOP, but the majority of the spectra have signal-to-noise ratios too low to extract precise stellar parameters using traditional methods. To take full advantage of this large observational effort, we have developed a tool (stellar parameter classification (SPC); see Supplementary Information) that uses a library of synthetic spectra to determine stellar parameters from spectra with modest signal-to-noise ratios (signal-to-noise per pixel >15). Using this approach, we derived metallicities in a consistent and homogeneous manner for the entire sample of Kepler FOP spectra, thus avoiding the systematic differences that can occur when comparing metallicities derived by different techniques. Only the most robust classifications are presented here (Supplementary Information), yielding precise stellar parameters for 152 stars harbouring 226 planet

candidates mostly in orbits within 0.5 AU of the host star. We used the stellar parameters from SPC and the Yonsei–Yale stellar evolutionary models[12] to estimate the radii of the host stars, which we couple with the photometric data from the Kepler mission[11] to infer the planet radii (Supplementary Information).

Previous studies[4,6–9] have suggested that the observed correlation between metallicity and the likelihood that solar-type stars host gas giants is weaker for Neptunian-sized planets. However, it is unclear whether this correlation extends into the regime of terrestrial-sized planets, which is important for a better understanding of planet-formation processes. The number of host stars with planets smaller than Neptune in our sample (175 planets) is significantly larger than in earlier studies and includes much smaller planets (as small as Earth). This allows us to compare a statistically significant sample of homogenously derived spectroscopic metallicities of solar-type stars hosting small and large planets. By contrast, a recent study used metallicity indicators based on photometry[13]. In Fig. 1, we show that the average metallicity of stars hosting planets with radii smaller than that of Neptune ($R_P < 4.0R_\oplus$) is lower ([m/H] = -0.01 ± 0.02) than that of the stars harbouring gas giant planets ([m/H] = $+0.15 \pm 0.03$). We find that smaller planets are observed at a wide range of host-star metallicities ($-0.6 < $ [m/H] $< +0.5$), whereas larger planets are detected preferentially around stars with higher metallicity (Figs 2 and 3). To investigate the statistical significance of the difference in metallicity, we perform the two-sample Kolmogorov–Smirnov test of the two subsamples of host stars and find the probability that the two distributions are not drawn randomly from the same parent population is 99.96% (over 3.5σ). An F-test shows that fitting the data in Fig. 3 with a metallicity that increases linearly, as opposed to being constant, as a function of radius yields a better fit with a confidence level of 99.99995% ($\sim 5\sigma$).

Figures 2 and 3 reveal that the population of small planets has a wide range of host-star metallicities, but on average the metallicity of the stars hosting the smaller planets is lower than that of the larger planets. The Kepler-11 system[14] demonstrates that small planets can possess a wide range of mean densities, much like their Jupiter-sized counterparts, and the low mean density of exoplanets Kepler-11d, e and f implies that these planets formed before the gas in the system dissipated completely. The metallicity of the protoplanetary disk may have a key role in how quickly planetary cores can form and, thus, in whether they are able to accrete a gaseous envelope before the gas in the system dissipates. However, additional data, including dynamical masses, are

[1]Niels Bohr Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark. [2]Centre for Star and Planet Formation, Natural History Museum of Denmark, University of Copenhagen, DK-1350 Copenhagen, Denmark. [3]Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. [4]Lund Observatory, Lund University, Box 43, 221 00 Lund, Sweden. [5]SETI Institute/NASA Ames Research Center, Moffett Field, California 94035, USA. [6]San Jose State University, San Jose, California 95192, USA. [7]NASA Ames Research Center, Moffett Field, California 94035, USA. [8]University of Texas, Austin, Texas 78712, USA. [9]NASA Exoplanet Science Institute/California Institute of Technology, Pasadena, California 91109, USA. [10]University of Florida, 211 Bryant Space Sciences Center, Gainesville, Florida 32611, USA. [11]Space Telescope Science Institute, Baltimore, Maryland 21218, USA. [12]University of California, Berkeley, California 94720, USA. [13]Bowling Green State University, Bowling Green, Ohio 43403, USA. [14]Centre for Astrophysics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK. [15]Department of Physics, Applied Physics and Astronomy, Rensselaer Polytechnic Institute, CII 9015, 110 8th Street, Troy, New York 12180, USA. [16]Las Cumbres Observatory, Global Telescope Network, Santa Barbara, California 93117, USA. [17]Department of Physics, University of California, Santa Barbara, California 93106, USA. [18]Bay Area Environmental Research Institute/NASA Ames Research Center, Moffett Field, California 94035, USA.
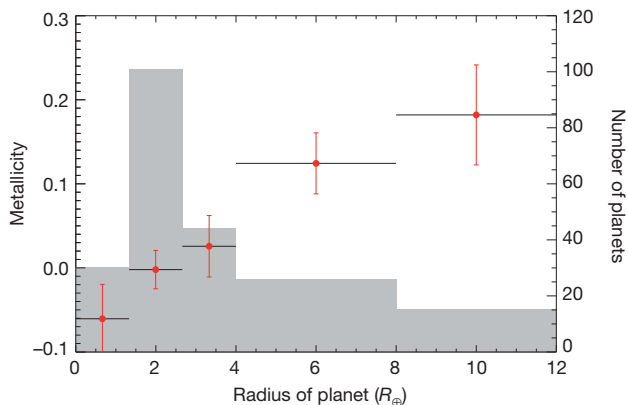
Figure 1 | Average host-star metallicities. Stellar metallicity is defined as $[m/H] = \log_{10}(N_m/N_H)_{star} - \log_{10}(N_m/N_H)_{Sun}$, where N_m and N_H are respectively the number densities of metal atoms (all elements more massive than helium) and hydrogen atoms. Red points represent the average metallicity of the host stars with planets of different radii grouped in $1.33R_\oplus$ and $4R_\oplus$ bins. The bin size is indicated by the length of the horizontal line and the uncertainty in the average metallicity is given by the standard error. The shaded grey histogram shows the number of planets in each bin, and illustrates the large number of small planets in the Kepler sample. The average metallicity of host stars with smaller planets ($R_P < 4R_\oplus$) is lower ($[m/H] = -0.01 \pm 0.02$) than that of host stars with larger planets ($[m/H] = +0.15 \pm 0.03$). Some of the planetary candidates in the Kepler sample are expected to be false positives that do not turn out to be transiting planets, such as occurs when the reduced signal from a background eclipsing binary is by chance contained within the photometric aperture of the foreground target star. The false-positive rate of the candidates that pass the standard vetting procedures applied by the Kepler team has been estimated to be less than 10% (ref. 26). Therefore, such a low false-positive rate is not expected to impact our results and interpretation. We have thus ignored possible contamination by false positives. We do not derive absolute probabilities or occurrence rates of planets and therefore do not attempt to eliminate the many strong bias and selection effects that, for example, completeness studies (for example ref. 27) must take into account. We have explored the possibility that correlations between planet size and parameters such as orbital semi-major axis are the source of the apparent dependence on metallicity, but find no evidence for such an effect (Supplementary Information).
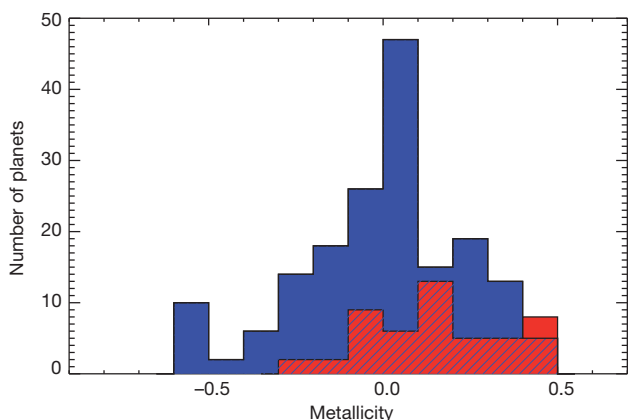
Figure 2 | Comparison of host-star metallicities for small and large planets. The histograms compare the metallicities of two samples of stars hosting planets by dividing the sample at $R_P = 4R_\oplus$. The host stars of the gas giant planets ($R_P \gtrsim 4R_\oplus$; red histogram) are clearly more metal rich than those of the smaller planets ($R_P < 4R_\oplus$; blue histogram), which have a much wider range of metallicities. The hatched area represents the area where the histograms overlap. A Kolmogorov–Smirnov test shows that the probability that the two distributions are not drawn randomly from the same parent population is greater than 99.96%; that is, the two distributions differ by more than 3.5σ. The average metallicity of the stars with small planets ($[m/H] = -0.01 \pm 0.02$; blue histogram) differs by almost 5σ from that of the larger planets ($[m/H] = +0.15 \pm 0.03$; red histogram).
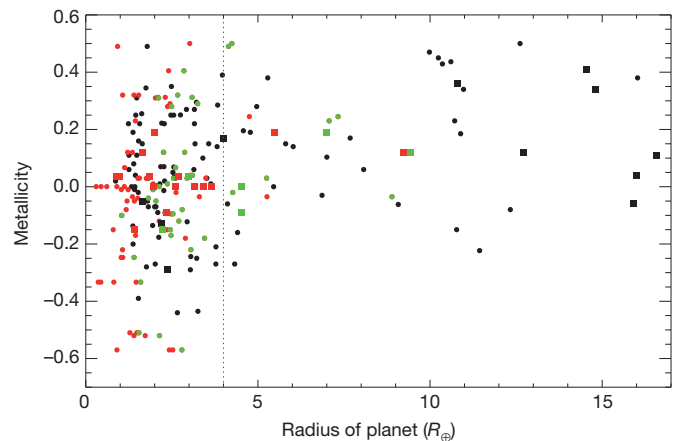
Figure 3 | Individual host-star metallicity as a function of planet radius. The black dots represent single-planet systems, whereas the green dots represent the largest planet and the red dots represent all the smaller planets in multiple-planet systems. The confirmed, published Kepler planets in our samples are plotted as squares with the same colour code as the dots. Planet candidates in multiple systems are each added to the sample with the same host-star metallicity. In Supplementary Information, we consider systems of planets as opposed to individual planets by neglecting all but the largest planet in each system. The vertical dotted line indicates the division of the sample at $R_P = 4.0R_\oplus$. The data show that Kepler detects small planets around stars with a wide range of metallicities ($-0.6 < [m/H] < 0.5$), and that larger planets are found preferentially around stars with solar metallicity or higher. The average uncertainty in the individual measurements in metallicity is 0.08 dex and that in planetary radius is 12%.

needed better to understand the seemingly diverse regime of small planets.

Our data show that the well-established correlation between metallicity and occurrence of giant planets[1–3] does not extend into the smaller planet regime below $R_P < 4R_\oplus$, where the host stars instead have a wide range of metallicities. This observation implies that, by contrast with smaller planets, gas giants require exceptional conditions to trigger their formation. Our findings agree well with the core accretion theory for planet formation, whereby high-metallicity environments allow planetary cores to grow rapidly to reach approximately ten times the mass of the Earth, continue to accrete a gaseous envelope and evolve to gas giants of several hundred Earth masses[5]. Gas disks around young stars are observed to dissipate within a few million years[15], requiring the cores of their planets to reach ten Earth masses within that time if they are to become gas giants. Planets forming in low-metallicity environments, however, may not reach large enough core masses before the dissipation of the gas disk, which could explain why we find very few gas giants around low-metallicity stars. Planetary accretion cannot compete with gas dissipation around low-metallicity stars because the number density of planetesimals is low[16–18] and gas disks dissipate sooner around low-metallicity stars[19,20].

The semi-major axes of the orbits of the majority of the Kepler planets analysed in this work are less than 0.5 AU, so the detected gas giants in our sample were probably brought into orbits within 1 AU by migration[21]. A decreased efficiency of migration in low-metallicity disks could partly explain the observed deficiency of gas giants around the low-metallicity stars. The formation of gas giants late in the lifetime of the protoplanetary gas disk would reduce their subsequent migration because the gas disk is diluted at that stage. This could partly explain why we observe so few gas giants in close orbits. However, late planet formation will in itself suppress formation of gas giants because some cores are formed after the disappearance of the gas disk. Hence, migration cannot be the only reason for the small number of gas giants that we observe around low-metallicity stars.

During the initial stages of planet formation, dust grains collide to form planetesimals, which represent the kilometre-sized building

blocks of planets. In some models, planetesimal formation is only possible in disks with metallicities greater than the solar value[17,18]. However, our results show that small planets are present around stars with a wide range of metallicities. The formation of planetesimals in low-metallicity environments can occur if the metallicity is enhanced by preferential evaporation of gas in the disk, for example by photo-evaporation processes associated with the central star or, alternatively, by external sources such as nearby massive stars. In both cases, the removal of gas by photo-evaporation is expected to occur early in the lifetime of the disk, possibly within one million years[22,23]. Such short timescales are consistent with radiometric age dating of meteorites suggesting that, in the Solar System, planetesimal accretion may have begun as early as a few hundred thousand years following formation of the Sun[24].

Finally, we note that some studies have proposed that a metallicity of at least half that of the Sun is required for the formation of terrestrial planets[25]. However, our analysis based on the Kepler planet candidates indicates that terrestrial planets can form at a wide range of metallicities, including metallicities almost four times lower than that of the Sun ($[m/H] \approx -0.6$). In addition, we find that the frequency of occurrence of small planets ($R_P < 4.0 R_\oplus$) relative to that of large planets ($R_P > 4.0 R_\oplus$) is ~2.7:1 for stars of metallicity greater than that of the Sun but increases to ~5.9:1 for stars of metallicity less than that of the Sun. Therefore, the formation of small, terrestrial planets does not require a metal-rich environment, suggesting that their existence might be widespread in the disk of the Galaxy.

1. Santos, N. C., Israelian, G. & Mayor, M. Spectroscopic [Fe/H] for 98 extra-solar planet-host stars. Exploring the probability of planet formation. *Astron. Astrophys.* **415,** 1153–1166 (2004).
2. Fischer, D. A. & Valenti, J. The planet-metallicity correlation. *Astrophys. J.* **622,** 1102–1117 (2005).
3. Johnson, J. A., Aller, K. M., Howard, A. W. & Crepp, J. R. Giant planet occurrence in the stellar mass-metallicity plane. *Publ. Astron. Soc. Pacif.* **122,** 905–915 (2010).
4. Sousa, S. G., Santos, N. C., Israelian, G., Mayor, M. & Udry, S. Spectroscopic stellar parameters for 582 FGK stars in the HARPS volume-limited sample. Revising the metallicity-planet correlation. *Astron. Astrophys.* **533,** A141 (2011).
5. Pollack, J. B. *et al.* Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124,** 62–85 (1996).
6. Udry, S. *et al.* The HARPS search for southern extra-solar planets. V. A 14 Earth-masses planet orbiting HD 4308. *Astron. Astrophys.* **447,** 361–367 (2006).
7. Sousa, S. G. *et al.* Spectroscopic parameters for 451 stars in the HARPS GTO planet search program. Stellar [Fe/H] and the frequency of exo-Neptunes. *Astron. Astrophys.* **487,** 373–381 (2008).
8. Ghezzi, L. *et al.* Stellar parameters and metallicities of stars hosting Jovian and Neptunian mass planets: a possible dependence of planetary mass on metallicity. *Astrophys. J.* **720,** 1290–1302 (2010).
9. Mayor, M. *et al.* The HARPS search for southern extra-solar planets XXXIV. Occurrence, mass distribution and orbital properties of super-Earths and Neptune-mass planets. Preprint at http://arxiv.org/abs/1109.2497 (2011).
10. Borucki, W. J. *et al.* Kepler planet-detection mission: introduction and first results. *Science* **327,** 977–980 (2010).
11. Borucki, W. J. *et al.* Characteristics of planetary candidates observed by Kepler. II. Analysis of the first four months of data. *Astrophys. J.* **736,** 19–41 (2011).
12. Yi, S. *et al.* Toward better age estimates for stellar populations: the Y2 isochrones for solar mixture. *Astrophys. J. Suppl. Ser.* **136,** 417–437 (2001).
13. Schlaufman, K. C. & Laughlin, G. Kepler exoplanet candidate host stars are preferentially metal rich. *Astrophys. J.* **738,** 177–186 (2011).
14. Lissauer, J. J. *et al.* A closely packed system of low-mass, low-density planets transiting Kepler-11. *Nature* **470,** 53–58 (2011).
15. Haisch, K. E., Lada, E. A. & Lada, C. J. Disk frequencies and lifetimes in young clusters. *Astrophys. J.* **553,** L153–L156 (2001).
16. Ida, S. & Lin, D. N. C. Toward a deterministic model of planetary formation. II. The formation and retention of gas giant planets around stars with a range of metallicities. *Astrophys. J.* **616,** 567–572 (2004).
17. Throop, H. B. & Bally, J. Can photoevaporation trigger planetesimal formation? *Astrophys. J.* **623,** L149–L152 (2005).
18. Johansen, A., Youdin, A. & Mac Low, M.-M. Particle clumping and planetesimal formation depend strongly on metallicity. *Astrophys. J.* **704,** L75–L79 (2009).
19. Ercolano, B. & Clarke, C. J. Metallicity, planet formation and disc lifetimes. *Mon. Not. R. Astron. Soc.* **402,** 2735–2743 (2010).
20. Yasui, C., Kobayashi, N., Tokunaga, A. T., Saito, M. & Tokoku, C. Short lifetime of protoplanetary disks in low-metallicity environments. *Astrophys. J.* **723,** L113–L116 (2010).
21. Lin, D. N. C., Bodenheimer, P. & Richardson, D. C. Orbital migration of the planetary companion of 51 Pegasi to its present location. *Nature* **380,** 606–607 (1996).
22. Gorti, U. & Hollenbach, D. Photoevaporation of circumstellar disks by far-ultraviolet, extreme-ultraviolet and X-ray radiation from the central star. *Astrophys. J.* **690,** 1539–1552 (2009).
23. Hester, J. J., Desch, S. J., Healy, K. R. & Leshin, L. A. The cradle of the Solar System. *Science* **304,** 1116–1117 (2004).
24. Schiller, M. *et al.* Rapid timescales for magma ocean crystallization on the howardite-eucrite-diogenite parent body. *Astrophys. J.* **740,** L22–L28 (2011).
25. Gonzalez, G., Brownlee, D. & Ward, P. The galactic habitable zone: galactic chemical evolution. *Icarus* **152,** 185–200 (2001).
26. Morton, T. D. & Johnson, J. A. On the low false positive probabilities of Kepler planet candidates. *Astrophys. J.* **738,** 170–182 (2011).
27. Howard, A. W. *et al.* Planet occurrence within 0.25 AU of Solar-type stars from Kepler. Preprint at http://arxiv.org/abs/1103.2541 (2011).

Author Contributions L.A.B. led the project and developed the classification tools for the metallicity analysis. D.W.L., A.J. and M.B. contributed to the discussion of the theoretical implications of the data. G.T. supplied the isochrone fitting tools. J.F.R. provided the planet radii from the Kepler photometry. D.W.L., C.C., W.D.C., G.A.E., E.B., M.E., J.C.G., T.H., G.W.M., P.R., R.P.S. and S.N.Q. worked on gathering the spectroscopic observations. D.W.L., W.J.B., S.T.B., N.M.B., D.R.C., W.D.C., R.L.G., P.W.L., G.W.M., A.S., M.S., H.I., E.B.F. and S.N.Q. worked on identifying the Kepler planets and eliminating false positives. W.J.B. led the Kepler mission. J.B.L., J.A.M. and D.W.L. worked on the synthetic stellar model library. All authors discussed the results and commented on the manuscript. L.A.B. wrote the paper with equal input from D.W.L., A.J. and M.B.

LETTER

Constraints on the volatile distribution within Shackleton crater at the lunar south pole

Maria T. Zuber[1], James W. Head[2], David E. Smith[1], Gregory A. Neumann[3], Erwan Mazarico[1], Mark H. Torrence[4], Oded Aharonson[5], Alexander R. Tye[2], Caleb I. Fassett[2], Margaret A. Rosenburg[5] & H. Jay Melosh[6]

Shackleton crater is nearly coincident with the Moon's south pole. Its interior receives almost no direct sunlight and is a perennial cold trap[1,2], making Shackleton a promising candidate location in which to seek sequestered volatiles[3]. However, previous orbital and Earth-based radar mapping[4–8] and orbital optical imaging[9] have yielded conflicting interpretations about the existence of volatiles. Here we present observations from the Lunar Orbiter Laser Altimeter on board the Lunar Reconnaissance Orbiter, revealing Shackleton to be an ancient, unusually well-preserved simple crater whose interior walls are fresher than its floor and rim. Shackleton floor deposits are nearly the same age as the rim, suggesting that little floor deposition has occurred since the crater formed more than three billion years ago. At a wavelength of 1,064 nanometres, the floor of Shackleton is brighter than the surrounding terrain and the interiors of nearby craters, but not as bright as the interior walls. The combined observations are explicable primarily by downslope movement of regolith on the walls exposing fresher underlying material. The relatively brighter crater floor is most simply explained by decreased space weathering due to shadowing, but a one-micrometre-thick layer containing about 20 per cent surficial ice is an alternative possibility.

Detailed study of the topography of Shackleton (Fig. 1a) offers the opportunity to improve understanding of processes that operate in permanently shadowed regions (Fig. 1b). Crater geometry, age and preservation state are relevant for understanding the accumulation and preservation of volatiles as well as the processes that modify the lunar surface over geologic timescales.

Our analysis uses observations from the Lunar Orbiter Laser Altimeter (LOLA)[10], an instrument on NASA's Lunar Reconnaissance Orbiter (LRO) mission. LOLA is a five-beam laser altimeter that operates at a wavelength of 1,064.4 nm with a 28-Hz pulse repetition rate. From LRO's mapping orbit at ~50 km altitude, the instrument illuminates 5-m-diameter spots on the lunar surface, returning up to 140 measurements of elevation per second; the five profiles enable characterization of bi-directional slopes over various baselines, and roughness from averaging of pulse elevations. In addition, from the spreading of backscattered laser pulses, LOLA obtains the root-mean-square (RMS) roughness of the surface within laser footprints. Finally, from the ratio of received to transmitted laser energy, LOLA measures the reflectance of the lunar surface at zero phase angle at the laser wavelength within laser spots.

As of 1 December 2011, the LOLA instrument has accumulated more than 5.1 billion elevation measurements[11]. Because Shackleton lies nearly at a pole, where the LOLA coverage is densest, it is possible to construct a digital elevation model of unprecedented spatial resolution and radial accuracy. More than 5,000 LOLA tracks, referenced to the Moon's centre of mass via precision orbits determined from radio tracking[12] aided by Earth-based laser tracking[13,14], were converted

to topography. Track segments within the area of interest were geometrically corrected at orbit crossover points[15].

Figure 1a shows the topography of Shackleton crater sampled at 10-m spatial resolution; individual measurements have an accuracy of ~1 m with respect to the Moon's centre of mass. The 40 km × 40 km topographic model of Shackleton is derived from 5.358 million elevation measurements with an average of 0.34 altimeter measurements in each 10-m square area; the resolution is comparable to or better than other studies of Shackleton's interior by images[9], Earth-based radar[6] and orbital synthetic aperture radar[16]. The topography reveals the near-axisymmetric bowl-shaped nature of the crater, in which the crater rim and interior walls are well preserved. The depth/diameter ratio of the crater is 0.195 ± 0.025 (Table 1), which is consistent with other fresh simple craters[17].

Figure 1c shows bi-directional surface slopes over 10-m baselines that quantify the uniform, steep inner walls; slopes approach the angle of repose. Slopes are greatest in the mid-levels of walls, which is in contrast to many crater walls on Mars, where near-vertically oriented cliffs of outwardly dipping coherent cap rock are exposed in the upper walls[18].

Surface roughness at a scale of 20–50 m is shown in Fig. 1d. These data indicate that crater walls are smoother, within bounds of measurement uncertainty, at this spatial scale than the floor or rim, especially portions of lower walls aligned with mounds on the crater floor. Table 1 lists estimated roughness values of various crater components. Similarly, the average RMS roughness derived from spreading of individual laser footprints (Supplementary Information) is lower on the crater walls. Pulses returned from the steep walls of Shackleton are spread in time by >10 ns, but after correcting for the effect of local slope on a longer baseline, the pulse spreading due to surface roughness is somewhat less than on the crater floor or surrounding terrain. The floor can be divided into two regions, a flat portion and an elevated terrain. The roughness of the mound unit increases at the largest scales due to its hummocky character, but it is smoother than the flat region at smaller scales, due to its paucity of craters.

Figure 2 shows a more detailed view of the topography of Shackleton's floor, which highlights the irregularly distributed deposits and numerous small craters (see also Fig. 1e). The largest mound of material has a relief of ~210 m (Table 1) and the highest-local slope of any of the floor deposits is ~25°, which is below the angle of repose. Two areas of the floor show fan-shaped structures consisting of material that has been transported downslope from the interior walls in a manner commonly observed in craters of this size range[17]. The limited fan material around the margins of the crater floor, combined with the asymmetric distribution and slope properties of deposits, suggest that the predominant contribution to the fill is ejecta fallback with a secondary contribution from slumped wall deposits.

[1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA. [3]Solar System Exploration Division, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [4]Stinger Ghaffarian Technologies, Greenbelt, Maryland 20770, USA. [5]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. [6]Department of Earth and Atmospheric Sciences, Purdue University, West Lafayette, Indiana 47907, USA.
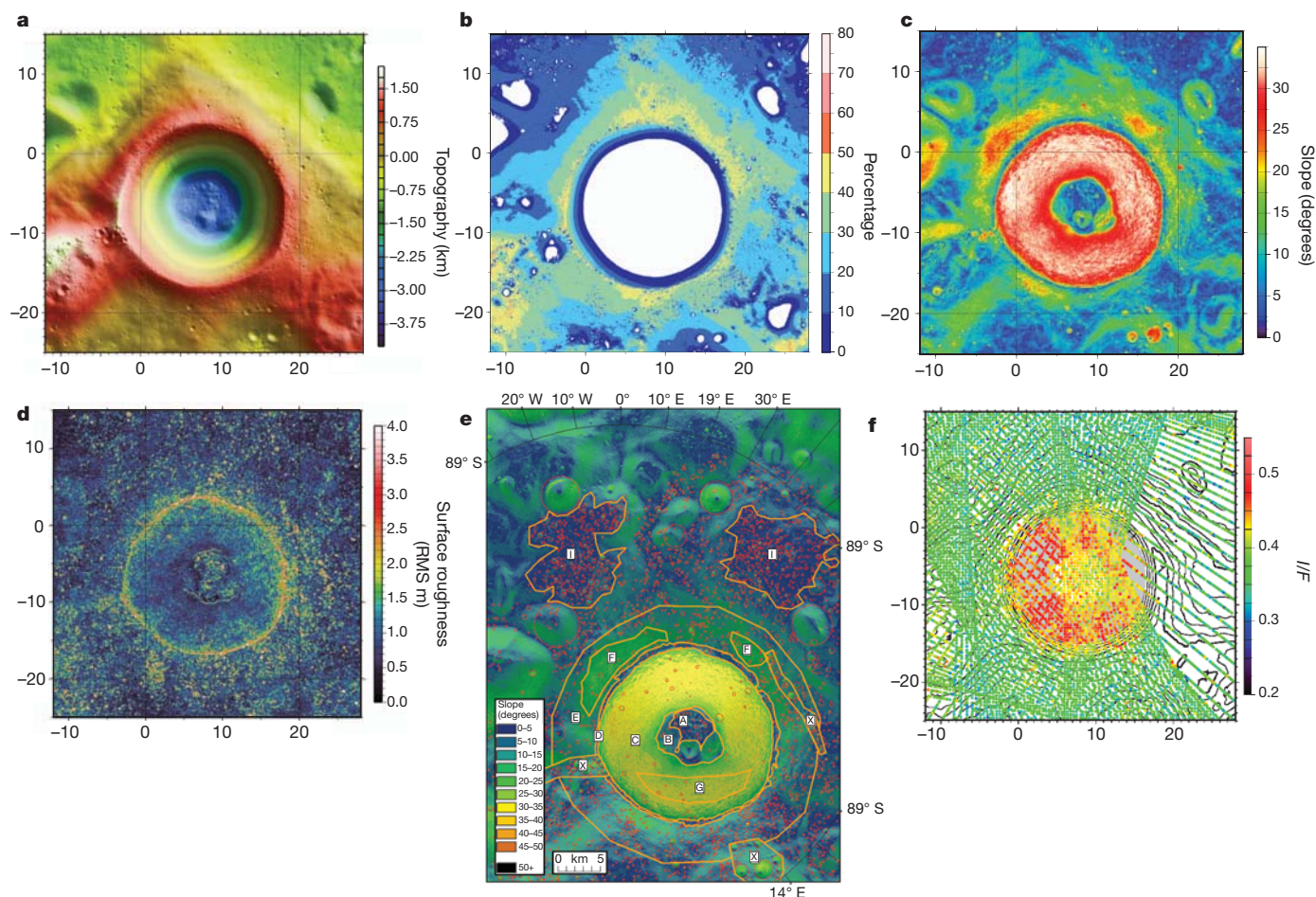
Figure 1 | Detailed characterization of Shackleton crater. a, Topography in km; **b**, percentage of time illuminated; **c**, 10-m baseline slopes in degrees; **d**, surface roughness shown as RMS residual in m; **e**, locations of crater counts used to determine relative ages; and **f**, zero-phase, 1,064-nm reflectance shown as I/F. Topography, slopes and roughness are based on a 10-m spatial resolution grid of all available LOLA profiles. In **a–d** and **f**, x and y axes indicate spatial scale, where (0, 0) is the lunar south pole and colour scales show magnitude of plotted quantity. White regions in **b** correspond to zero illumination. Panel **e** shows locations of craters counted to estimate relative age, plotted over 10-m slopes (colour coded as in inset). Crater regions in **e** correspond to: A, flat region of crater floor; A/B, entire crater floor; C, crater wall; D, crater rim crest; E/F, inner rim annulus ($\sim$5.5 km); E, inner rim annulus excluding steep region (F); F, steep rim region within annulus; G, crater wall section; I, Shackleton crater deposits north of rim in flat areas; and X, secondary crater chains and clusters (removed from analysis). In **f**, reflectance is expressed as a radiance factor (I/F), which is defined as the ratio of the measured radiance I to the radiance F of an ideal diffusive surface in vacuum with 100% reflectance under the same illumination. Each dot represents a 0.4×0.4 km pixel median average of LOLA's spot 3 reflectance. Contours show topography at 0.2 km intervals. The grey annulus shows the 17-km diameter of the steepest portion of the walls and the 7-km diameter of the floor.

Table 1 | Parameters describing Shackleton crater

Parameter	Value
Areocentric latitude of centre of rim (degrees)	−89.655
Areocentric longitude of centre of rim (degrees)	129.174
Lunar radius at floor centre (km)	1,734.63
Mean crater diameter at rim (km)	21
Mean depth, rim to floor (km)	4.1 ± 0.05
Mean rim height above datum (km)	1.3
Range of floor topography (km)	$\sim$0.210
Area of crater at rim (km^2)	$\sim$346
Area of crater floor (km^2)	$\sim$38
Estimated fill depth (km)	$\sim$0.75
Crater volume (km^3)	640 ± 10
Fill volume, including mounds (km^3)	12 ± 1
Maximum wall slope (degrees)	35
Average wall slope (degrees)	30.5
RMS roughness* of crater exterior (m)	$\sim$1
RMS roughness* of crater walls (m)	<1
RMS roughness* of crater floor (m)	$\sim$1
RMS roughness* of crater rim (m)	$\sim$1
I/F of crater exterior	0.32 ± 0.04
I/F of interior walls	0.46 ± 0.03
I/F of interior floor	0.43 ± 0.02
Ratio of average depth/average rim diameter, d/D	0.195 ± 0.025

See Fig. I legend for definition of I/F.
* Within 5-m spots.

Shackleton was previously assigned an Eratosthenian age[19] (middle lunar history; in the approximate interval 1–3.2 Gyr before present) on the basis of its relatively fresh morphology, its lack of rays, and counts of superposed craters[3] using AMIE image data (50 m per pixel) and Arecibo radar data (20 m per pixel). Craters were counted within a crater diameter ($\sim$20 km) of the rim crest, avoiding obvious secondary craters. This analysis was subsequently revisited and resulted in an older, Imbrian age[3] (in the approximate interval 3.2–3.8 Gyr before present).

Here we use a LOLA shaded relief map to advance previous work by individually dating different parts of the crater (see Fig. 1e and Supplementary Information) to investigate the processes that have operated since crater formation. LOLA observations permit dating of shadowed regions in the crater interior and allow spatially unbiased measurements of crater density due to uniformity in illumination conditions. In addition, illumination can be varied over the topographic model to enhance crater detection. On the basis of comparison of the several different areas of the rim of the crater, it is clear that the variable slopes of the rough crater rim have an influence on crater retention. For example, two areas of very flat terrain on the Shackleton flank within one crater diameter of the rim crest (I; Fig. 1e) yield modelled crater ages of $\sim$3.69 Gyr, whereas areas closer to the rim crest (within $\sim$5.5 km)
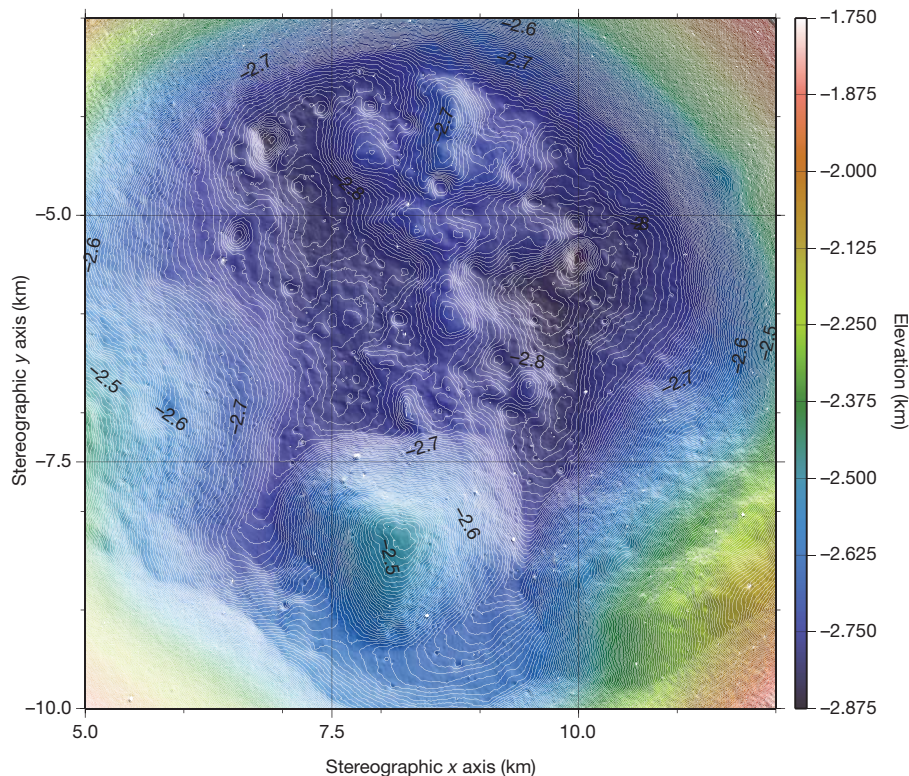
Figure 2 | High-resolution elevation map in stereographic projection of the floor of Shackleton. Elevations are contoured at 5-m intervals with colours indicating elevation with respect to 1,737.4 km. The axes indicate spatial scales.

yield ages of ~1.21 Gyr (F, steeper slopes; Fig. 1e) and ~2.91 Gyr (E, fewer steep slopes; Fig. 1e). The flat areas of the Shackleton crater deposit (I; Fig. 1e) indicate an age of ~3.69 Gyr (Supplementary Table 1), older than the originally estimated age of 1.3–3.3 Gyr (ref. 19) but close to the Upper Imbrian age of ~3.6 Gyr estimated subsequently[3]. Determination of ages of these individual regions permits quantitative investigation of how the crater has been modified.

A critical question is the age of the permanently shadowed portions of the interior walls and floor of Shackleton. Analysis indicates that the lunar spin axis has been at its approximate current orientation for ~2 Gyr (ref. 20). If the crater had been accumulating volatiles in the permanently shadowed areas over a period of this order, it is reasonable to hypothesize that its interior was resurfaced, covering and burying craters and thus producing a younger relative age. Examination of the steep crater wall (C; Fig. 1e) yields a much younger crater retention age of ~1.44 Gyr, which could be consistent with either volatile mantling or downslope mass wasting (Fig. 1c)[21]. Examination of the permanently shadowed parts of the crater floor (A; Fig. 1e), the area where volatiles plausibly accumulated, reveals a crater retention age of ~3.60 Gyr, essentially identical to the flat areas of the crater rim. Including the rougher parts of the crater floor (B; Fig. 1e) produces an age of ~3.29 Gyr, which almost certainly reflects the influence of downslope transport, as observed on the inner part of the Shackleton rim, on the retention of craters. The similar age of the Shackleton crater exterior and floor (~3.69 and 3.60 Gyr) is evidence that if volatiles accumulated in this cold trap for ~10[9] years, they were not in sufficient quantity to alter in a statistically significant sense the size–frequency distribution of superposed craters in the size range counted.

Knowledge of the crater size–frequency distribution (Supplementary Fig. 1) permits an assessment of the minimum amount of deposition that could occur without disrupting this distribution. The craters counted on Shackleton's floor ranged from ~250 m up to ~500 m in diameter, representing a range of fresh crater depths of ~50–100 m (ref. 22). Accumulation of volatiles to thicknesses in excess of 20–50 m would significantly alter the crater size–frequency distribution and

should be observed in terms of a deficit in the number of small craters. A rollover in the curve at small crater diameters is observed, but is very similar to that seen outside the crater (I; Fig 1e and Supplementary Fig. 1), and thus is more likely to be due to the typical destruction effects of superposed craters and other diffusive processes (such as micrometeorite bombardment or seismic shaking associated with moonquakes triggered by stresses associated with impacts or tides) at these diameters. On the other hand, if volatiles were cold-trapped by vapour diffusion into the regolith, as opposed to deposition in surface layers, then muting of superposed craters may not have been as significant.

Figure 1f shows profiles of 1,064-nm reflectance of Shackleton crater and its surroundings. A previous study[9] obtained images of the floor of Shackleton from the Kaguya Terrain Camera at the time of maximum scattering illumination at the lunar south pole and observed no evidence for brightening; results were interpreted to indicate an absence of pure ice deposits on the crater floor. In the current study, LOLA profiles assembled from numerous orbital passes at the most favourable conditions for obtaining reliable measurements of reflectance show that the crater walls are anomalously bright relative to the surrounding terrain. Similarly to other impact craters in this size range, this brightness could be due to downslope movement of material caused by micrometeorite and small projectile bombardment on steep slopes, or by seismic shaking. The cascading of regolith material downslope exposes optically less mature surfaces than those developed and retained on lower slopes.

At a wavelength of 1,064 nm, the floor of Shackleton crater is darker than its interior walls, but both floors and walls are considerably brighter than the surrounding terrain, including the interiors of nearby craters that are both shadowed and sunlit. The relative brightness of the floor relative to surroundings requires explanation. Micrometeorite bombardment and impingement of the solar wind produce 'space weathering' of exposed geologic materials that reddens and darkens their surfaces[23]. The former (bombardment) would be less significant in the shadowed interior of Shackleton than in the surrounding region

because the interior of the crater has not been exposed to the Sun for more than 2 Gyr (ref. 20), whereas the latter (solar wind impingement) might be enhanced in permanently shadowed craters[24]. Thus the floor brightness enhancement could be explained by a dearth of space weathering by micrometeorite bombardment.

Volatile deposition is an alternative possibility. Under the conservative assumption that water ice has a 1,064-nm reflectance twice that of the lunar regolith[25] and that both are observed at zero phase, the measured reflectance of the floor can be explained by a micrometre-thick surface layer (the depth over which the laser backscatter measurement is sensitive) of 22% ice mixed with rock[26]. Greater ice contents distributed throughout a thicker layer are possible but cannot be constrained from LOLA's reflectivity measurement. For comparison, far-ultraviolet reflectance of permanently shadowed regions from the LRO Lyman Alpha Mapping Project (LAMP) is consistent with ~1–2% surface water frost[27].

Results from the LRO Mini-RF orbital radar for the interior of Shackleton[16] provide additional insight. Regions with thick ice deposits are expected to have circular polarization ratios (CPR) >1, but such high ratios can also be explained by surface roughness. High-resolution images from ground-based radar[6] show that some areas with high CPR lie inside the rim of Shackleton. Mini-RF data[16] reveal that CPR values decrease with depth within the crater, and CPR values on the floor of Shackleton crater are predominantly <1. The pixels with CPR values in excess of unity are distributed heterogeneously throughout the crater walls, correlate generally with regions of high roughness observed by LOLA (Fig. 1d) and include some sunlit areas (Fig. 1b). Although some contribution to high CPR values from volatiles is possible, and would imply a process in which volatile deposition operates very rapidly in comparison to the rate of removal, the combined data suggest that the higher floor reflectance is due primarily to the dearth of space weathering in this shadowed environment.

In considering why Shackleton's interior walls have a higher reflectance than its floor, it is instructive to note that wall brightening is not restricted to areas that are continuously shadowed but extends to the upper illuminated portions. Consequently, a higher concentration of surface volatiles on the walls than present on the floor is an unlikely explanation. More likely is downslope movement of regolith material on the steep crater walls that has exposed brighter underlying material; downslope movement is consistent with the observed slopes near the angle of repose and the roughness and morphology of Shackleton's interior walls (Fig. 1c, d), as well as with Earth-based radar backscatter[6] and Mini-RF[16] measurements.

METHODS SUMMARY

LOLA, an instrument on board the LRO spacecraft, outputs five beams per laser pulse that are backscattered from the lunar surface and detected in the instrument's receiver. The relevant measurement is the time of flight of each individual laser pulse, which can be converted to a range of the spacecraft to the lunar surface given knowledge of the position of the spacecraft with respect to the Moon's centre of mass. Radial range errors were minimized by geometric adjustment of altimetric tracks within the study area. In practice, profiles of one-way range from the LOLA instrument to the lunar surface along the spacecraft ground track were converted to lunar radius at each bounce point using the reconstructed orbit of LRO. Topography was determined by subtracting a sphere of 1,737.4 km from each radius measurement.

Slopes were calculated from a two-laser-spot fit at 20–50-m length scales, and RMS roughness represents a standard deviation about a plane fitted by least-squares to two laser shots along-track, from which at least four valid spots are returned out of a possible total of ten. The plane has dimensions of 90–100 m in the longest axis and 10–40 m in the shortest axis, depending on the positions of the spots returned.

The ratio of returned to transmitted pulse energy is a measure of surface reflectance at the laser wavelength. The transmitted and returned pulse energies were measured by integrating the area under the pulses. Observations of 1,064-nm reflectance were derived from LOLA tracks crossing Shackleton for days 130–149 in 2010.

1. Watson, K., Murray, B. C. & Brown, H. The behavior of volatiles on the lunar surface. *J. Geophys. Res.* **66**, 3033–3045 (1961).
2. Arnold, J. R. Ice in the lunar polar regions. *J. Geophys. Res.* **84**, 5659–5668 (1979).
3. Spudis, P. D., Plescia, J., Josset, J.-L & Beauvivre, S. Geology of Shackleton Crater and the south pole of the Moon. *Geophys. Res. Lett.* 36, L14201, http://dx.doi.org/10.1029/2008GL034468 (2008).
4. Nozette, S. *et al.* The Clementine bistatic radar experiment. *Science* **274**, 1495–1498 (1996).
5. Stacy, N. J. S., Campbell, D. B. & Ford, P. G. Arecibo radar mapping of the lunar poles: a search for ice deposits. *Science* **276**, 1527–1530 (1997).
6. Campbell, D. B., Campbell, B. A., Carter, L. M., Margot, J.-L. & Stacy, N. J. S. No evidence for thick deposits of ice at the lunar south pole. *Nature* **443**, 835–837 (2006).
7. Simpson, R. & Tyler, G. L. Reanalysis of Clementine bistatic radar data from the lunar south pole. *J. Geophys. Res.* **104**, 3845–3862 (1999).
8. Nozette, S. *et al.* Integration of lunar polar remote-sensing data sets: evidence for ice at the lunar south pole. *J. Geophys. Res.* **106** (E10), 23253–23266 (2001).
9. Haruyama, J. *et al.* Lack of exposed ice inside lunar south pole Shackleton crater. *Science* **322**, 938–939 (2008).
10. Smith, D. E. *et al.* The Lunar Orbiter Laser Altimeter investigation on the Lunar Reconnaissance Orbiter mission. *Space Sci. Rev.* **150**, 209–241 (2010).
11. Smith, D. E. *et al.* Results from the Lunar Orbiter Laser Altimeter (LOLA): global, high resolution topographic mapping of the Moon. *Lunar Planet. Sci. Conf.* **XLII**, 2350 (2011).
12. Mazarico, E. *et al.* Orbit determination of the Lunar Reconnaissance Orbiter. *J. Geod.* **86**, 193–207 (2012).
13. Sun, X. *et al. The Laser Ranging Subsystem on the Lunar Reconnaissance Orbiter.* Report No. GSC-15884-1 (NASA New Technology Report, Washington DC, 2009).
14. Zuber, M. T. *et al.* The Lunar Reconnaissance Orbiter laser ranging investigation. *Space Sci. Rev.* **150**, 63–80 (2010).
15. Neumann, G. A., Rowlands, D. D., Lemoine, F. G., Smith, D. E. & Zuber, M. T. Crossover analysis of MOLA altimetric data. *J. Geophys. Res.* **106** (E10), 23753–23768 (2001).
16. Thomson, B. J. *et al.* The interior of Shackleton crater as revealed by Mini-RF orbital radar. *Lunar Planet. Sci. Conf.* **XLII**, 1626 (2011).
17. Pike, R. J. in *Impact and Explosion Cratering* (eds Roddy, D. J., Pepin, R. O. & Merrill, R. B.) 489–509 (Pergamon, 1977).
18. Squyres, S. W. *et al.* Exploration of Victoria crater by the Mars Rover Opportunity. *Science* **324**, 1058–1061 (2009).
19. Wilhelms, D. E., Howard, K. A. & Wilshire, H. G. *Geologic Map of the South Side of the Moon* (Map I-1162, US Geological Survey, 1979).
20. Ward, W. R. Past orientation of the lunar spin axis. *Science* **189**, 377–379 (1975).
21. Howard, K. A. Fresh lunar impact craters — review of variations with size. In *Proc. 5th Lunar Sci. Conf.* 61–69 (Pergamon, 1974).
22. Pike, R. J. Depth/diameter relations of fresh lunar craters: revision from spacecraft data. *Geophys. Res. Lett.* **1**, 291–294 (1974).
23. Hapke, B. Space weathering from Mercury to the asteroid belt. *J. Geophys. Res.* **106** (E5), 10039–10073 (2001).
24. Zimmerman, M. I., Farrell, W. M., Stubbs, T. J., Halekas, J. S. & Jackson, T. L. Solar wind access to polar craters: feedback between surface charging and plasma expansion. *Geophys. Res. Lett.* 38, L19202, http://dx.doi.org/10.1029/2011GL048880 (2011).
25. Kwok, R., Cunningham, G. F., Zwally, H. J. & Yi, D. ICESat over Arctic sea ice: Interpretation of altimetric and reflectivity profiles. *J. Geophys. Res.* 111, C06006, http://dx.doi.org/10.1029/2005JC003175 (2006).
26. Pieters, C. M. *et al.* Character and spatial distribution of OH/H_2O on the surface of the Moon seen by M3 on Chandrayaan-1. *Science* **326**, 568–572 (2009).
27. Gladstone, G. R. *et al.* Far-ultraviolet reflectance properties of the Moon's permanently shadowed regions. *J. Geophys. Res.* 117, E00H04, http://dx.doi.org/10.1029/2011JE003913 (2012).

METHODS

The Lunar Orbiter Laser Altimeter (LOLA), an instrument aboard the Lunar Reconnaissance Orbiter (LRO) spacecraft, outputs five beams per laser pulse that are backscattered from the lunar surface and detected in the instrument's receiver. The relevant measurement is the time of flight of each individual laser pulse, which can be converted to a range of the spacecraft to the lunar surface given knowledge of the position of the spacecraft with respect to the Moon's centre of mass. The timing of LOLA instrument events was derived from the LRO ultrastable oscillator, which is monitored by ground tracking stations. Time systems on board LRO used Coordinated Universal Time (UTC) to correlate spacecraft Mission Elapsed Time (MET) to ground time. The analysis of LOLA data used Barycentric Dynamical Time as its primary time system. Spacecraft states relative to the Solar System Barycentre (SSB) at the laser transmit and detector receive times were projected along the instrument boresight and return path vectors to match the observed time of flight, correcting for the aberration of light and general-relativistic time delays. SSB states were determined in the Earth Mean Equator of 2000 (J2000) inertial reference frame using lunar spacecraft trajectories and the DE421 planetary ephemeris[28].

During its lunar mapping mission, the LRO spacecraft is tracked using S-band Doppler and range data by the Universal Space Network, Deep Space Network and White Sands Missile Range. The precise reconstruction of LRO orbits used Doppler tracking observations from these stations as well as laser ranging to LRO[14] from the Goddard Space Flight Center and participating members of the International Laser Ranging Service. Precision orbit determination was accomplished using the NASA/Goddard Space Flight Center's GEODYN system of programs[29] using the GLGM-3 gravity model[30] as a reference. GEODYN numerically integrates the spacecraft Cartesian state and force-model partial derivatives by employing a high-order Cowell predictor-corrector model. In addition to a model of the lunar gravity field, the force modelling included point mass representations for the Sun and planets. Solar radiation pressure, measurement and timing biases, and tracking station coordinates were also estimated.

Radial range errors were minimized by geometric adjustment of altimetric tracks within the study area. In practice, profiles of one-way range from the LOLA instrument to the lunar surface along the spacecraft ground track were converted to lunar radius at each bounce point using the reconstructed orbit of LRO. Topography (Fig. 1a) was determined by subtracting a sphere of 1,737.4 km from each radius measurement. Topographic measurements were binned and interpolated within a 40 km × 40 km area at 10-m spatial resolution.

The LOLA digital elevation model (DEM) used in this analysis was combined with a lunar ephemeris[31] to characterize the solar illumination conditions of Shackleton and surroundings (Fig. 1b). A polar gnomonic projection on which great circle paths plot as straight lines was applied to the DEM to calculate lighting conditions that are accurate over geological timescales.

The confluence of orbit ground tracks in the vicinity of the lunar poles, combined with LOLA's multi-beam profiling capability, permitted slopes over a range of baselines and directions to be determined. In this study, slopes (Fig. 1c) were calculated from a two-laser-spot fit at 20–50-m length scales. In similar fashion, RMS roughness (Fig. 1d) was calculated as a standard deviation about a plane fitted by least-squares to two laser shots along-track, from which at least four valid spots were returned out of a possible total of ten. The plane had dimensions of 90–100 m in the longest axis and 10–40 m in the shortest axis, depending on the positions of the spots returned. An independent measure of surface roughness, discussed in Supplementary Information and shown in Supplementary Fig. 2, used the spreading in time of backscattered pulses. The spreading in time of LOLA's backscattered pulses provides a measure of the RMS roughness of the surface at a smaller scale—the 5-m diameter of the laser footprints on the lunar surface. As with plane-deviation roughness shown in Fig. 1d, the pulse-spread-derived roughness featured a correction for local slopes.

The ratio of returned to transmitted pulse energy is a measure of surface reflectance at the laser wavelength of 1,064 nm. The transmitted and returned pulse energies were measured by integrating the area under the pulses. LOLA's measurement of reflectance is calibrated only in a relative sense, with respect to pre-launch testing, as the instrument lacks a source with known brightness in flight. Observations of reflectance were derived from LOLA tracks crossing Shackleton and environs from day of year 130–149 2010, which represented the most favourable time period for stable reflectance measurements due to the geometry of spacecraft terminator crossings.

28. Folkner, W. M., Williams, J. G. & Boggs, D. H. *The Planetary and Lunar Ephemeris DE421* (Jet Propulsion Laboratory, Pasadena, 2008).
29. Pavlis, D. E., Poulouse, S. G. & McCarthy, J. J. *GEODYN Operations Manuals* (SGT, Inc., Greenbelt, 2009).
30. Mazarico, E., Lemoine, F. G., Han, S.-C. & Smith, D. E. GLGM-3, a degree-150 lunar gravity model from the historical tracking data of NASA Moon orbiters. *J. Geophys. Res.* 115, E05001, http://dx.doi.org/10.1029/2009JE003472 (2010).
31. Williams, J. G., Boggs, D. H. & Folkner, W. M. *Lunar Orbit, Physical Librations and Surface Coordinates* (Jet Propulsion Laboratory, Pasadena, 2008).

LETTER

Electronic nematicity above the structural and superconducting transition in BaFe$_2$(As$_{1-x}$P$_x$)$_2$

S. Kasahara[1,2], H. J. Shi[1], K. Hashimoto[1]†, S. Tonegawa[1], Y. Mizukami[1], T. Shibauchi[1], K. Sugimoto[3,4], T. Fukuda[5,6,7], T. Terashima[2], Andriy H. Nevidomskyy[8] & Y. Matsuda[1]

Electronic nematicity, a unidirectional self-organized state that breaks the rotational symmetry of the underlying lattice[1,2], has been observed in the iron pnictide[3–7] and copper oxide[8–11] high-temperature superconductors. Whether nematicity plays an equally important role in these two systems is highly controversial. In iron pnictides, the nematicity has usually been associated with the tetragonal-to-orthorhombic structural transition at temperature T_s. Although recent experiments[3–7] have provided hints of nematicity, they were performed either in the low-temperature orthorhombic phase[3,5] or in the tetragonal phase under uniaxial strain[4,6,7], both of which break the 90° rotational C_4 symmetry. Therefore, the question remains open whether the nematicity can exist above T_s without an external driving force. Here we report magnetic torque measurements of the isovalent-doping system BaFe$_2$(As$_{1-x}$P$_x$)$_2$, showing that the nematicity develops well above T_s and, moreover, persists to the non-magnetic superconducting regime, resulting in a phase diagram similar to the pseudogap phase diagram of the copper oxides[8,12]. By combining these results with synchrotron X-ray measurements, we identify two distinct temperatures—one at T^*, signifying a true nematic transition, and the other at T_s ($<T^*$), which we show not to be a true phase transition, but rather what we refer to as a 'meta-nematic transition', in analogy to the well-known meta-magnetic transition in the theory of magnetism.

Magnetic torque measurements provide a stringent test of nematicity for systems with tetragonal symmetry[13]. The torque $\tau = \mu_0 V \mathbf{M} \times \mathbf{H}$ is a thermodynamic quantity, a differential of the free energy with respect to angular displacement. Here μ_0 is the permeability of vacuum, V is the sample volume, and $\mathbf{M}$ is the magnetization induced in the magnetic field $\mathbf{H}$. When $\mathbf{H}$ is rotated within the tetragonal a–b plane (Fig. 1a, b), τ is a periodic function of 2ϕ, where ϕ is the azimuthal angle measured from the a axis:

$$\tau_{2\phi} = \frac{1}{2}\mu_0 H^2 V[(\chi_{aa} - \chi_{bb})\sin 2\phi - 2\chi_{ab}\cos 2\phi] \quad (1)$$

where the susceptibility tensor χ_{ij} is defined by $M_i = \Sigma_j \chi_{ij} H_j$. In a system maintaining tetragonal symmetry, $\tau_{2\phi}$ should be zero, because $\chi_{aa} = \chi_{bb}$ and $\chi_{ab} = 0$. Finite values of $\tau_{2\phi}$ appear if a new electronic or magnetic state emerges that breaks the C_4 tetragonal symmetry. In such a case, rotational symmetry breaking is revealed by $\chi_{aa} \neq \chi_{bb}$ and/or $\chi_{ab} \neq 0$, depending on the direction of the nematicity.

BaFe$_2$(As$_{1-x}$P$_x$)$_2$ is a prototypical family of iron pnictides[14–18], whose phase diagram is displayed in Fig. 1c. The temperature evolution of the torque $\tau(\phi)$ for the optimally doped compound ($x = 0.33$) is depicted in the upper panels of Fig. 1d. The two- and four-fold oscillations, $\tau_{2\phi}$ and $\tau_{4\phi}$, obtained from the Fourier analysis are shown respectively in the middle and lower panels of Fig. 1d. The distinct two-fold oscillations appear at low temperatures, whereas they are absent at high temperatures
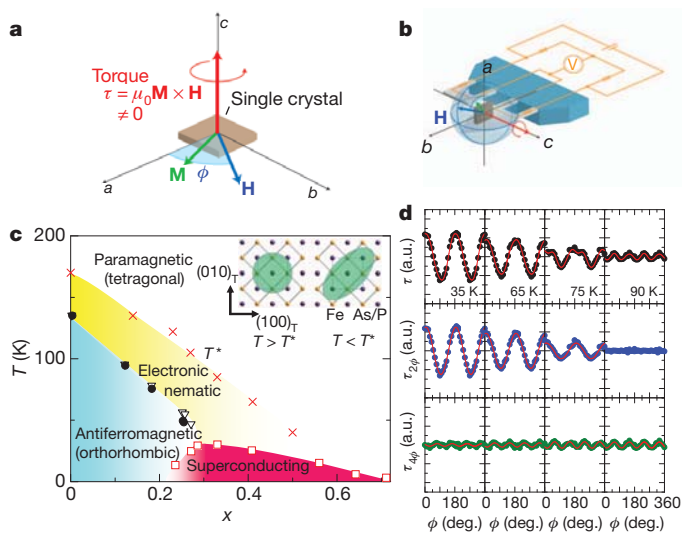
Figure 1 | Torque magnetometry and the doping–temperature phase diagram of BaFe$_2$(As$_{1-x}$P$_x$)$_2$. a, b, Schematic representations of the experimental configuration for torque measurements under in-plane field rotation. In a nematic state, domain formation with different preferred directions in the a–b plane ('twinning') will occur. We used very small single crystals with typical size $\sim$70 µm $\times$ 70 µm $\times$ 30 µm, in which a significant difference in volume between the two types of domains enables the observation of uncompensated $\tau_{2\phi}$ signals. The equation given in the figure for τ assumes unit volume; see text for details. A single-crystalline sample (brown block) is mounted on the piezo-resistive lever which is attached to the base (blue block) and forms an electrical bridge circuit (orange lines) with the neighbouring reference lever. A magnetic field $\mathbf{H}$ can be rotated relative to the sample, as illustrated by a blue arrow on a sphere. In this experiment, the field is precisely applied in the a–b plane. **c,** Phase diagram of BaFe$_2$(As$_{1-x}$P$_x$)$_2$. This system is clean and homogeneous[14,16,17], as demonstrated by the quantum oscillations observed over a wide x range[16]. The antiferromagnetic transition at T_N (filled circles)[15] coincides or is preceded by the structural transition at T_s (open triangles)[18]. The superconducting dome extends over a doping range $0.2 < x < 0.7$ (open squares), with maximum $T_c = 31$ K. Crosses indicate the nematic transition temperature T^* determined by the torque and synchrotron X-ray diffraction measurements. The insets illustrate the tetragonal FeAs/P layer. $\chi_{ab} = 0$ above T^* yielding an isotropic torque signal (green-shaded circle), whereas $\chi_{ab} \neq 0$ below T^*, indicating the appearance of the nematicity along the $[110]_T$ (Fe–Fe bond) direction, illustrated with the green-shaded ellipse. **d,** The upper panels depict the temperature evolution of the raw torque $\tau(\phi)$ at $\mu_0 H = 4$ T for BaFe$_2$(As$_{0.67}$P$_{0.33}$)$_2$ ($T_c = 30$ K). All torque curves are reversible with respect to the field rotation. $\tau(\phi)$ can be decomposed as $\tau(\phi) = \tau_{2\phi} + \tau_{4\phi} + \tau_{6\phi} + \cdots$, where $\tau_{2n\phi} = A_{2n\phi}\sin 2n(\phi - \phi_0)$ has $2n$-fold symmetry with integer n. The middle and lower panels display the two- and four-fold components obtained from Fourier analysis. The four-fold oscillations $\tau_{4\phi}$ (and higher-order terms) arise primarily from the nonlinear susceptibilities[13]. a.u., arbitrary units.

[1]Department of Physics, Kyoto University, Kyoto 606-8502, Japan. [2]Research Center for Low Temperature and Materials Sciences, Kyoto University, Kyoto 606-8501, Japan. [3]Research and Utilization Division, JASRI SPring-8, Sayo, Hyogo 679-5198, Japan. [4]Structural Materials Science Laboratory, RIKEN SPring-8, Sayo, Hyogo 679-5148, Japan. [5]Quantum Beam Science Directorate, JAEA SPring-8, Sayo, Hyogo 679-5148, Japan. [6]Materials Dynamics Laboratory, RIKEN SPring-8, Sayo, Hyogo 679-5148, Japan. [7]JST, Transformative Research-Project on Iron Pnictides (TRIP), Chiyoda, Tokyo 102-0075, Japan. [8]Department of Physics and Astronomy, Rice University, 6100 Main Street, Houston, Texas 77005, USA. †Present address: Institute for Materials Research, Tohoku University, Sendai 980-8577, Japan.

(middle panels of Fig. 1d). As shown in the upper panel of Fig. 2a, the amplitude of the two-fold oscillation $|A_{2\phi}|$ is nearly zero at high temperatures, and grows rapidly below $T^* \approx 85$ K. These results clearly indicate that the tetragonal C_4 symmetry, which is preserved at high temperatures, is broken below T^*, demonstrating the formation of the electronic nematic phase at T^*. The two-fold oscillations below T^* follow the functional form $\tau_{2\phi} = A_{2\phi}\cos 2\phi$, meaning that $\chi_{aa} = \chi_{bb}$ and $\chi_{ab} \neq 0$, which indicates the nematicity along the tetragonal [110] direction, that is, the Fe–Fe bond direction (Fig. 1c, inset). Anomalies at T^* can also be seen in the synchrotron X-ray diffraction measurements (middle panel of Fig. 2a), in which the full-width at half-maximum (FWHM) of the high-order Bragg peaks at $T < T^*$ (Fig. 2a, red circles) grows more quickly as temperature is reduced than does the linear extrapolation from above T^*, and is accompanied by the suppression of the peak intensity (green circles). This indicates a broadening (or small splitting) of the Bragg peak below T^*, implying domain formation due to the nematicity which, to some extent, couples to the orthorhombic lattice distortion (as discussed later).

Figure 2b–e shows the results for the lower-concentration samples ($x = 0.27, 0.23, 0.14, 0$) exhibiting both the tetragonal-to-orthorhombic and magnetic transitions. In these crystals, $|A_{2\phi}|$ is finite even at 200 K (grey circles in upper panels), and initially increases with decreasing temperature, exhibiting a cusp-like peak at T^*. We attribute T^* to the electronic nematic transition temperature for the following reasons. Similarly to the optimally doped compound, the X-ray FWHM and intensity change their slope at T^* (middle panels in Fig. 2b–e). Moreover, the two-fold term in these compounds can be clearly separated into two components (see Supplementary Information) as $\tau_{2\phi} = A_{2\phi}^{\mathrm{nem}} \cos 2\phi + A_{2\phi}^{\mathrm{ext}} \sin 2(\phi - \phi_{\mathrm{ext}})$, in which ϕ_{ext} is temperature independent and $A_{2\phi}^{\mathrm{ext}}$ has a smooth temperature dependence with the phase ϕ_{ext} that is sample dependent. This demonstrates that the

observed two-fold oscillations contain the [110] nematic term $A_{2\phi}^{\mathrm{nem}}\cos 2\phi$ in addition to some other, extrinsic two-fold term. The obtained nematic amplitude $A_{2\phi}^{\mathrm{nem}}(T)$ appears below T^* (blue circles in upper panels); this behaviour is well reproduced for different crystals of the same composition (Fig. 2b and Supplementary Fig. 3). The origin of the non-zero two-fold signal at temperatures above T^* is not clear, but the sample-dependent phase ϕ_{ext} points to the possibility of impurity-induced in-plane magnetic anisotropy[19]. In contrast to previous experiments, the present measurements have been performed in the absence of uniaxial stress, providing thermodynamic evidence for the electronic nematic transition at T^*, well above the structural lattice transition temperature T_s. The continuous doping dependence of $T^*(x)$ displayed in Fig. 1c indicates that nematicity persists over a wide range of doping, covering the non-magnetic superconducting regime.

Clearly, there cannot be two nematic phase transitions at both T_s and T^*, because the C_4 rotational symmetry can only be broken once. Our measurements show that the temperature T^* ($>T_s$) marks the onset of the true phase transition, accompanied by the nematic two-fold torque component $A_{2\phi} \neq 0$. This raises the question as to what happens at the structural transition temperature, T_s. This question can be answered straightforwardly if one considers the Landau free energy expansion in terms of two order parameters, the orthorhombic lattice distortion $\delta = (a - b)/(a + b)$ and the phenomenological electronic nematic parameter ψ (proportional to the measured torque amplitude $A_{2\phi}$), which can be written as follows:

$$F[\delta, \psi] = \left[t_s\delta^2 - u\delta^4 + v\delta^6\right] + \left[t_p\psi^2 + w\psi^4 + \mathcal{O}(\psi^6)\right] - g\psi\delta \quad (2)$$

with the terms in the first set of square brackets on the right-hand side describing the first-order structural phase transition and the second bracket responsible for the (second-order) nematic phase transition.
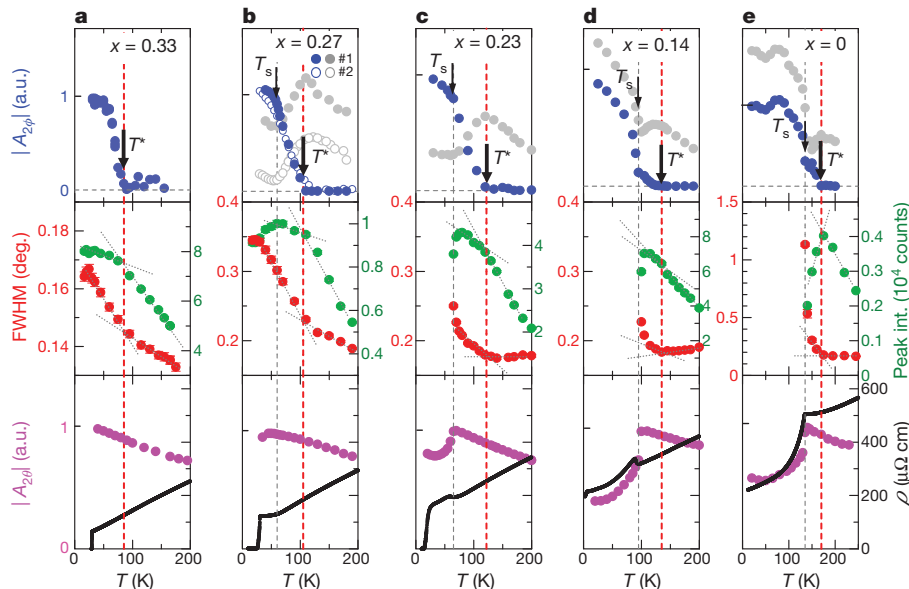
Figure 2 | Nematic and meta-nematic transitions. Temperature dependence of the two-fold oscillation amplitude $|A_{2\phi}|$ of the torque (blue circles in upper panels), the FWHM and peak intensity (red and green circles in middle panels, respectively) of the synchrotron X-ray Bragg reflection, the in-plane resistivity ρ (black lines in lower panels), and the torque amplitude for polar-angle rotation (pink circles in lower panels) are shown for x values of 0.33 (**a**), 0.27 (**b**), 0.23 (**c**), 0.14 (**d**) and 0 (**e**). The nematic transition temperature T^* and the meta-nematic transition temperature T_s are defined by red and grey vertical dashed lines, respectively, as well as by arrows in the upper panels. For underdoped ($x \lesssim 0.27$) and parent ($x = 0$) crystals, the intrinsic nematic signal for $|A_{2\phi}|$ is extracted from the raw data (grey circles in upper panels) with subtraction of the smooth background of the two-fold oscillations, $A_{2\phi}^{\mathrm{ext}}\sin 2(\phi - \phi_{\mathrm{ext}})$, above T^* (Supplementary Fig. 2). For $x = 0.27$, the sample

dependence is also shown. The X-ray data have been analysed for several Bragg peaks in the following tetragonal directions of the reciprocal space: $[10,0,0]_T$ (**a**), $[7,7,0]_T$ (**b**, **e**) or $[8,8,0]_T$ (**c**, **d**). For $x \lesssim 0.23$ a clear splitting of the Bragg peaks is observed below T_s (Supplementary Fig. 5). The dotted lines are guides to the eye. For $x = 0, 0.27$ and 0.33, the same crystal has been used for the torque and X-ray measurements. We find no clear anomaly in the resistivity $\rho(T)$ and the torque signal $|A_{2\theta}|$ for polar-angle rotation (which would detect the a–c anisotropy of susceptibility; lower panel), indicating that the anomalies are associated only with the in-plane nematicity. The magnitude of $|A_{2\phi}|$ has only a weak field dependence (Supplementary Fig. 4), implying that the nematic anisotropy prevails in the zero-field limit, which is consistent with the X-ray results taken at zero field.

Here u, v and w are phenomenological Landau coefficients describing these transitions, and g is the coupling strength between the two order parameters. The temperature-dependent coefficients $t_s = \left(T - T_s^{(0)}\right)/T_s^{(0)}$ and $t_p = \left(T - T_p^{(0)}\right)/T_p^{(0)}$ were chosen such that in the absence of the coupling between the two order parameters, the structural transition occurs at lower temperature $T_s^{(0)}\left(< T_p^{(0)}\right)$.

Thanks to the linear coupling between the two order parameters, as expressed in the last term in equation (2), both ψ and δ develop non-zero values below the nematic transition temperature T^* (Fig. 3a, b). On the other hand, T_s ceases to be a true phase transition, since the C_4 symmetry is already broken on either side of T_s, and the lattice distortion δ is non-zero over the entire temperature range (Fig. 3a). Instead, both δ and ψ undergo a finite jump at T_s, as illustrated in Fig. 3. We call this a 'meta-nematic transition', in analogy to the meta-magnetic transition in the theory of magnetism, where the magnetization undergoes a jump as a function of temperature or applied magnetic field, but remains non-zero on both sides of the transition. The analysis of the free energy below T^* shows that it exhibits a maximum at $\psi = 0$ and a single minimum at finite ψ, as in the second-order Landau phase transition (Supplementary Fig. 6).

To quantify the lattice distortion δ experimentally, we have analysed the X-ray data using a two-peak fitting procedure (Supplementary Fig. 5), which reveals that the data in the region $T_s < T < T^*$ can be fitted with very small but finite δ. The results, $\delta(T)$, can be well reproduced within

the framework of equation (2), as Fig. 3a demonstrates. Moreover, the same set of Landau parameters also fits the temperature dependence of the nematic order parameter $\psi \propto A_{2\phi}$ well, except below the Néel temperature T_N, where additional effects of antiferromagnetism may enter the magnetic torque anisotropy (Fig. 3b). We have thus established that the true thermodynamic transition occurs at $T = T^*$, and is accompanied by the development of non-zero values of both the nematic order parameter ψ and the lattice distortion δ. We note that similarly small but non-zero values of δ have been recently reported[20] in the powder diffraction measurements of SmFeAs($O_{1-x}F_x$).

Under applied uniaxial strain, the lattice distortion δ is expected to develop a non-zero value even above the tetragonal-to-orthorhombic structural transition temperature, and by virtue of linear coupling to the nematic order parameter in equation (2), this will lead to an increase in the value of the electronic anisotropy ψ. This reasoning is consistent with the observation of resistivity anisotropy in the detwinned samples above the structural transition[6,7].

We note that the above explanation in terms of coupling between the electronic and lattice degrees of freedom is very generic, and does not depend on the precise microscopic nature of the nematic order parameter ψ, be it caused by Z_2 spin-nematic ordering[21–24], or by orbital ordering[25–29]. In fact, both mechanisms can cooperate to break the C_4 symmetry spontaneously. In the spin-nematic scenario, the instability is driven by thermal spin fluctuations above the antiferromagnetically ordered phase[24], potentially applicable to the $x \lesssim 0.30$ regime in BaFe$_2$(As$_{1-x}$P$_x$)$_2$ (ref. 15). In the absence of long-range magnetic ordering ($x > 0.30$), quantum spin fluctuations can in principle still drive the instability; however, the fact that the nematic transition at T^* occurs even for superconducting samples far away from the antiferromagnetic phase (see Fig. 1c) indicates that a different mechanism may be at play. In particular, orbital ordering may turn out to be more important, where the electronic nematic transition naturally occurs as a result of orbital polarization between the Fe d_{xz} and d_{yz} orbitals[29], $\psi \propto (n_{xz} - n_{yz})$, where n_{xz} (n_{yz}) denotes the occupation of the iron d_{xz} (d_{yz}) orbital. This mechanism of nematicity is supported by the recent angle-resolved photoemission spectroscopy (ARPES)[4] and quadrupolar resonance measurements[30].

There is a growing body of evidence that entanglement of the spin and orbital degrees of freedom leads to emergent novel electronic phases in the iron pnictides. The present temperature-doping phase diagram bears a resemblance to that of the high-transition-temperature (high-T_c) copper oxides, in that the suppression of the antiferromagnetic ground state leads to the emergence of high-T_c superconductivity, and the electronic nematic instability occurs well above the magnetic and superconducting transitions. Recent infrared studies of charge dynamics report the formation of a pseudogap in the excitation spectrum of optimally doped BaFe$_2$(As$_{1-x}$P$_x$)$_2$ below ~100 K (S. J. Moon et al., unpublished results). Further studies are therefore needed to clarify how the nematic transition we report here is related to the pseudogap formation in the iron pnictides.

Figure 3 | Temperature dependence of the nematic order parameter and lattice distortion. a, Lattice distortion δ ($= (a - b)/(a + b)$) and **b**, the nematic order parameter ψ, proportional to measured $A_{2\phi}$ component of the torque in the paramagnetic temperature region, are fitted to the theory (lines) based on the Landau free energy expansion (equation (2)) using the same set of parameters (Supplementary Information). For the parent compound ($x = 0$, red): $u = 3.28$, $v = 4.08$, $w = 4.52$, $g = 1.91$, $T_s^{(0)} = 52$ K, $T_p^{(0)} = 117$ K; for $x = 0.14$ (green): $u = 9.48$, $v = 31.4$, $w = 31.5$, $g = 2.72$, $T_s^{(0)} = 35$ K, $T_p^{(0)} = 79$ K; for $x = 0.23$ (black): $u = 8.57$, $v = 27.3$, $w = 22.4$, $g = 2.34$, $T_s^{(0)} = 26$ K, $T_p^{(0)} = 79$ K; and for $x = 0.27$ (blue): $u = 4.37$, $v = 10.07$, $w = 42.58$, $g = 3.11$, $T_s^{(0)} = 24$ K, $T_p^{(0)} = 56$ K. Note that because of the coupling between the order parameters, both T_s and T^* get renormalized compared to their initial values $T_s^{(0)}$ and $T_p^{(0)}$. The experimental errors in **a** (error bars, s.e.m.) are estimated in the two-peak fitting procedure from the width of the Bragg peaks (Supplementary Information). In the antiferromagnetic phase below T_N, the amplitude $A_{2\phi}(T)$ shows deviations from $\psi(T)$ (dashed lines in **b**) for $x \leq 0.14$, where effects of the antiferromagnetic moment on torque may be large.

1. Kivelson, S. A., Fradkin, E. & Emery, V. J. Electronic liquid-crystal phases of a doped Mott insulator. *Nature* **393,** 550–553 (1998).
2. Fradkin, E., Kivelson, S. A., Lawler, M. J., Eisenstein, J. P. & Mackenzie, A. P. Nematic Fermi fluids in condensed matter physics. *Annu. Rev. Condens. Matter Phys.* **1,** 153–178 (2010).
3. Zhao, J. et al. Spin waves and magnetic exchange interactions in CaFe$_2$As$_2$. *Nature Phys.* **5,** 555–560 (2009).
4. Yi, M. et al. Symmetry-breaking orbital anisotropy observed for detwinned Ba(Fe$_{1-x}$Co$_x$)$_2$As$_2$ above the spin density wave transition. *Proc. Natl Acad. Sci. USA* **108,** 6878–6883 (2011).
5. Chuang, T.-M. et al. Nematic electronic structure in the "parent" state of the iron-based superconductor Ca(Fe$_{1-x}$Co$_x$)$_2$As$_2$. *Science* **327,** 181–184 (2010).
6. Chu, J.-H. et al. In-plane resistivity anisotropy in an underdoped iron arsenide superconductor. *Science* **329,** 824–826 (2010).
7. Tanatar, M. et al. Uniaxial-strain mechanical detwinning of CaFe$_2$As$_2$ and BaFe$_2$As$_2$ crystals: optical and transport study. *Phys. Rev. B* **81,** 184508 (2010).
8. Lawler, M. J. et al. Intra-unit-cell electronic nematicity of the high-T_c copper-oxide pseudogap states. *Nature* **466,** 347–351 (2010).

9. Hinkov, V. *et al.* Electronic liquid crystal state in the high-temperature superconductor $YBa_2Cu_3O_{6.45}$. *Science* **319,** 597–600 (2008).
10. Ando, Y., Segawa, K., Komiya, S. & Lavrov, A. N. Electrical resistivity anisotropy from self-organized one dimensionality in high-temperature superconductors. *Phys. Rev. Lett.* **88,** 137005 (2002).
11. Daou, R. *et al.* Broken rotational symmetry in the pseudogap phase of a high-T_c superconductor. *Nature* **463,** 519–522 (2010).
12. Damascelli, A., Hussain, Z. & Shen, Z.-X. Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75,** 473–541 (2003).
13. Okazaki, R. *et al.* Rotational symmetry breaking in the hidden-order phase of URu_2Si_2. *Science* **331,** 439–442 (2011).
14. Kasahara, S. *et al.* Evolution from non-Fermi- to Fermi-liquid transport via isovalent doping in $BaFe_2(As_{1-x}P_x)_2$ superconductors. *Phys. Rev. B* **81,** 184519 (2010).
15. Nakai, Y. *et al.* Unconventional superconductivity and antiferromagnetic quantum critical behavior in the isovalent-doped $BaFe_2(As_{1-x}P_x)_2$. *Phys. Rev. Lett.* **105,** 107003 (2010).
16. Shishido, H. *et al.* Evolution of the Fermi surface of $BaFe_2(As_{1-x}P_x)_2$ on entering the superconducting dome. *Phys. Rev. Lett.* **104,** 057008 (2010).
17. Hashimoto, K. *et al.* Line nodes in the energy gap of superconducting $BaFe_2(As_{1-x}P_x)_2$ single crystals as seen via penetration depth and thermal conductivity. *Phys. Rev. B* **81,** 220501(R) (2010).
18. Böhmer, A. E. *et al.* Thermodynamic phase diagram and phase competition in $BaFe_2(As_{1-x}P_x)_2$ studied by thermal expansion. Preprint at http://arxiv.org/abs/1203.2119 (2012).
19. Inoue, Y., Yamakawa, Y. & Kontani, H. Impurity-induced electronic nematic state in iron-pnictide superconductors. Preprint at ⟨http://arxiv.org/abs/1110.2401⟩ (2011).
20. Martinelli, A. *et al.* Retention of the tetragonal to orthorhombic structural transition in F-substituted SmFeAsO: a new phase diagram for $SmFeAs(O_{1-x}F_x)$. *Phys. Rev. Lett.* **106,** 227001 (2011).
21. Fang, C., Yao, H., Tsai, W.-F., Hu, J. & Kivelson, S. A. Theory of electron nematic order in LaFeAsO. *Phys. Rev. B* **77,** 224509 (2008).
22. Xu, C., Müller, M. & Sachdev, S. Ising and spin orders in the iron-based superconductors. *Phys. Rev. B* **78,** 020501(R) (2008).
23. Fernandes, R. M. *et al.* Effects of nematic fluctuations on the elastic properties of iron arsenide superconductors. *Phys. Rev. Lett.* **105,** 157003 (2010).
24. Fernandes, R. M., Chubukov, A. V., Knolle, J., Eremin, I. & Schmalian, J. Preemptive nematic order, pseudogap, and orbital order in the iron pnictides. *Phys. Rev. B* **85,** 024534 (2012).
25. Singh, R. R. P. Exchange constants and neutron spectra of iron pnictide materials. Preprint at http://arxiv.org/abs/0903.4408 (2009).
26. Lee, C.-C., Yin, W.-G. & Ku, W. Ferro-orbital order and strong magnetic anisotropy in the parent compounds of iron-pnictide superconductors. *Phys. Rev. Lett.* **103,** 267001 (2009).
27. Lv, W., Wu, J. & Phillips, P. Orbital ordering induces structural phase transition and the resistivity anomaly in iron pnictides. *Phys. Rev. B* **80,** 224506 (2009).
28. Chen, C.-C. *et al.* Orbital order and spontaneous orthorhombicity in iron pnictides. *Phys. Rev. B* **82,** 100504 (2010).
29. Nevidomskyy, A. H. Interplay of orbital and spin ordering in the iron pnictides. Preprint at http://arxiv.org/abs/1104.1747 (2011).
30. Lang, G. *et al.* Nanoscale electronic order in iron pnictides. *Phys. Rev. Lett.* **104,** 097001 (2010).

Author Contributions S.K. and H.J.S. performed magnetic torque measurements. S.K., H.J.S., K.H., S.T., Y. Mizukami, T.S., K.S. and T.F. contributed to the synchrotron X-ray measurements. S.K. grew the single crystals and performed transport measurements. A.H.N. carried out theoretical modelling and calculations. T.S. and Y. Matsuda conceived and designed the project. T.S., A.H.N. and Y. Matsuda wrote the manuscript with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y. Matsuda (matsuda@scphys.kyoto-u.ac.jp) or T.S. (shibauchi@scphys.kyoto-u.ac.jp).

LETTER

Multiscale gigapixel photography

D. J. Brady[1], M. E. Gehm[2], R. A. Stack[3], D. L. Marks[1], D. S. Kittle[1], D. R. Golish[2], E. M. Vera[2] & S. D. Feller[1]

Pixel count is the ratio of the solid angle within a camera's field of view to the solid angle covered by a single detector element. Because the size of the smallest resolvable pixel is proportional to aperture diameter and the maximum field of view is scale independent, the diffraction-limited pixel count is proportional to aperture area. At present, digital cameras operate near the fundamental limit of 1–10 megapixels for millimetre-scale apertures, but few approach the corresponding limits of 1–100 gigapixels for centimetre-scale apertures. Barriers to high-pixel-count imaging include scale-dependent geometric aberrations, the cost and complexity of gigapixel sensor arrays, and the computational and communications challenge of gigapixel image management. Here we describe the AWARE-2 camera, which uses a 16-mm entrance aperture to capture snapshot, one-gigapixel images at three frames per minute. AWARE-2 uses a parallel array of microcameras to reduce the problems of gigapixel imaging to those of megapixel imaging, which are more tractable. In cameras of conventional design, lens speed and field of view decrease as lens scale increases[1], but with the experimental system described here we confirm previous theoretical results[2–6] suggesting that lens speed and field of view can be scale independent in microcamera-based imagers resolving up to 50 gigapixels. Ubiquitous gigapixel cameras may transform the central challenge of photography from the question of where to point the camera to that of how to mine the data.

AWARE-2 is a monocentric, multiscale camera with 120°-by-50° field of view (FOV) and a 38-μrad instantaneous FOV (the angular extent of a single pixel). A monocentric, multiscale camera consists of a spherically symmetric objective lens[3] surrounded by an array of secondary microcameras[2,4,6–9]. AWARE-2 includes 98 microcameras, each with a 14-megapixel sensor. It was constructed as part of the US Defense Advanced Research Projects Agency AWARE programme, which focuses on creating a microcamera platform for scalable, 1–100-gigapixel cameras. Just as the microprocessor is a platform for scalable parallel computing, the microcamera is a platform for scalable parallel cameras with diverse applications in wide-field microscopy, event capture, persistent surveillance and space awareness. As with microprocessors, the designer of microcameras must address granularity and performance trade-offs in selecting aperture and focal plane size, materials and components, and interconnection and processing architecture. Details of the AWARE-2 system design are presented in Supplementary Information, sections 2 and 3. In this Letter, we compare AWARE-2 with previous gigapixel-scale imaging systems, illustrate the capture of a large-scale dynamic event, demonstrate the capacity for high-dynamic-range (HDR) imaging in microcamera systems and analyse the camera's optical resolution.

A gigapixel camera requires a lens system capable of resolving more than 10^9 elements and detectors containing more than 10^9 elements. Designs that address these challenges may be segmented into terrestrial cameras with horizontal 90–120° FOVs (AWARE[4] and Asymmagon; see http://www.gigapxl.org/), airborne surveillance cameras with cylindrical 60–70° FOVs (ARGUS-IS[10] and the multilens array[11]) and astronomical systems with cylindrical 3–4° FOVs (LSST[12] and Pan-STARRS[13]). Design metrics for these systems are provided in

Supplementary Table 1.1. Lens design strategies differ between arrays of narrow-field cameras (the multilens array) and single objectives with curved focal planes. Although arbitrarily large pixel counts may be obtained using arrays of conventional cameras, design comparisons confirm, as predicted[1], that the volume of a system with a flat focal plane scales much faster than does that of a design with a curved focal plane. Multiscale design captures the advantage of camera arrays (off-the-shelf focal plane arrays) while avoiding the disadvantage (cost and volume of multiple objective lenses). AWARE uses microcamera arrays to create a large virtual-focal-plane array, in place of precision-mosaicked sensor arrays as used in ARGUS, LSST and Pan-STARRS. The most important points of comparison are that the AWARE design is scalable to larger pixel counts, as illustrated by the AWARE-40 design[4], and that AWARE provides operational advantages over the comparison systems because the focus, gain and exposure of each microcamera can be independently controlled. The disadvantage of the AWARE approach is that a stop must be introduced in the microcamera to balance relative illumination and modulation transfer, thereby increasing system f number (the ratio of lens focal length to effective aperture, which is a measure of lens speed) and volume[4,9]. Bare mosaicked arrays may be preferred for fixed-focus, cost-insensitive airborne and astronomical systems. Multiscale design, however, uniquely makes possible compact, low-cost, terrestrial imaging systems focused at finite range.

The potential for novel science using AWARE-2 is illustrated in Fig. 1, which is a gigapixel snapshot of tundra swans on Pungo Lake in the Pocosin Lakes National Wildlife Refuge, USA. Microcamera data was registered and stitched onto a rectangular grid with 38-μrad instantaneous FOV, to produce a 0.96-gigapixel version of Fig. 1. Non-uniformity correction and logarithmic scaling were applied to improve visual display. Figure 1 is downsampled to 4 megapixels for publication. Raw microcamera images of Fig. 1 regions a, b, c, d and e are shown in Fig. 2. The gigapixel snapshot provides information, such as exactly how many swans are on Pungo Lake (656) or in the air above it (27) at the instant the snapshot was taken, that would be unobtainable using a scanned panoramic camera shooting from the same position. Also, the image and image sequences can be mined to analyse signalling behaviour across the flock and to track individual birds. AWARE-2 microcamera control modules use custom electronic modules (Supplementary Information, section 3) to buffer approximately ten image frames locally and to capture asynchronously and transfer individual microcamera images at up to ten frames per second.

An example HDR image is shown in Fig. 3. The brightness of this scene varies from regions of fully sunlit flat surfaces and bright sky to areas of deep shadow. An auto-exposure algorithm adjusts the exposure time for each individual microcamera to maximize the usage of the 8-bit dynamic range in each sensor. When the images are composited into the final scene, knowledge of the exposure setting is combined with the measured pixel values to estimate the source radiance with a global 32-bit dynamic range. Displays are typically limited to 8-bit output, so this HDR result must be mapped onto this smaller dynamic range. Figure 3 was generated by tone-mapping the HDR image.

[1]Fitzpatrick Institute for Photonics, Duke University, PO Box 90291, Durham, North Carolina 27708, USA. [2]ECE Department, University of Arizona, PO Box 210104, Tucson, Arizona 85721, USA. [3]Distant Focus Corporation, PO Box 7857, Champaign, Illinois 61826, USA.

Figure 1 | Pungo Lake as captured using AWARE-2. The total FOV is 120° by 50° and the composite image here consists of 0.96 gigapixels, where each pixel has an instantaneous FOV of 38 μrad. The measured values are logarithmically mapped to make better use of the display dynamic range. The

Figure 2 | Details of Fig. 1. a–e, Labelled regions in Fig. 1. The swans in **c** are 114 pixels long and are 310–350 m from AWARE-2. Each pixel corresponds to 13 mm at the position of the swans. In **d**, the most distant bird is 17 pixels long and the closest is 70 pixels long. The limited depth of focus of the camera is illustrated in **e**, where regions of the foreground foliage are in sharpest focus.

relative responsivities of the individual microcameras were adjusted iteratively to minimize variation across sub-image boundaries. The labelled regions are referred to in Fig. 2.

Tone mapping creates an individual 32-bit/8-bit conversion for each pixel in the scene, but ensures that the mappings vary smoothly from pixel to pixel[14]. The majority of the display dynamic range is used on shadows and highlights, with mid-tones compressed. The tone-mapped image more accurately matches human visual processing because our vision is foveated and our pupils adjust as we examine different regions of a wide field. More details on the compositing process are provided in Supplementary Information.

The maximum pixel count for an imager with aperture diameter A and the mean operating wavelength λ is $S = \pi A^2 \sin^2(\text{FOV}/2)/\lambda^2$ (ref. 2). For AWARE-2, FOV = 120° and A = 16 mm, corresponding to a limit of two gigapixels at an operating wavelength of 550 nm. At its design capacity of 220 microcameras, a fully populated AWARE-2 would capture three gigapixels but the estimated field would decrease to two gigapixels after sensor regions with limited illumination had been removed and overlapping regions had been merged. AWARE-2's resolution is illustrated by the star field shown in Fig. 4, which shows details of a gigapixel sky survey with an exposure time of 1.85 s. Faint stars in this image illuminate two to four pixels after median filtering to remove hot pixels and logarithmic intensity mapping to fill the display dynamic range. We anticipate that higher-resolution multiscale systems with terrestrial motion compensation and microcamera adaptive optics to remove atmospheric blur may be developed for space situational awareness.

As shown in systematic resolution images and modulation transfer function measurements presented in Supplementary Information, section 4b, most of the blur in the star field image is due to defects in AWARE-2's microcamera lenses. To allow low-cost gigapixel array integration, AWARE-2 uses injection-moulded plastic relay optics. These lenses may be moulded with aspheric surfaces, requiring fewer elements and consequently less volume and mass than a spherical glass-element camera with similar performance. However, birefringence in the fabricated optics due to residual stresses introduced during moulding degrades AWARE-2's image quality and resolution relative to fundamental limits. We expect newer high-index plastics that minimize birefringence to enable the camera to approach these theoretical limits. We have also built glass microcamera lenses and, as shown in Supplementary Information, section 4c, AWARE-2 achieves pixel-limited optical resolution with these lenses.
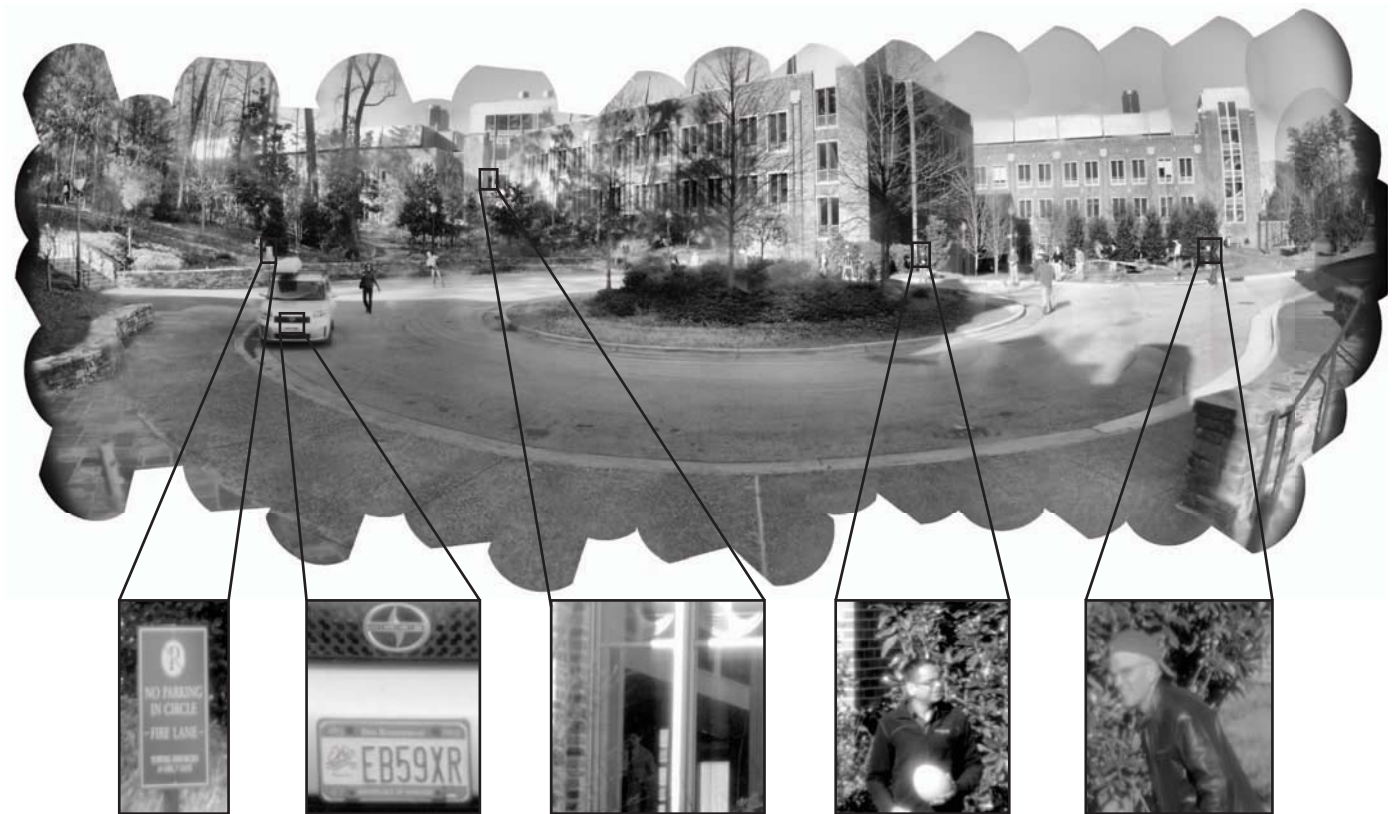
Figure 3 | Traffic circle captured using AWARE-2. Insets are digitally magnified by a factor of 13. Distances to the inset regions range from 15 m ('no parking' sign; first from left) to 92 m (detail of building; third from left). The exposure time for each microcamera was set independently of the others, and a tone-mapping algorithm was used to convert the resulting HDR image for display. Global distortion associated with mapping the 120° horizontal field onto a flat image is apparent.

AWARE-2 demonstrates that the age of ever-increasing pixel count is far from over. Although development of high-performance, low-cost microcamera optics and optomechanics have been the main challenge in the present stage of multiscale camera development, integrated circuits, rather than optics, remain the primary barrier to ubiquitous high-pixel-count imaging. To accommodate the electronics and allow for heat dissipation (the camera expends 430 W during image acquisition), AWARE-2 is mounted in a 0.75 m × 0.75 m × 0.5 m frame. The optical system occupies less than 3% of the system volume. The size of the camera is dictated both by the size of the electronic control boards and the need to cool them effectively. As more efficient and compact electronics are developed, hand-held gigapixel photography may become an everyday reality.
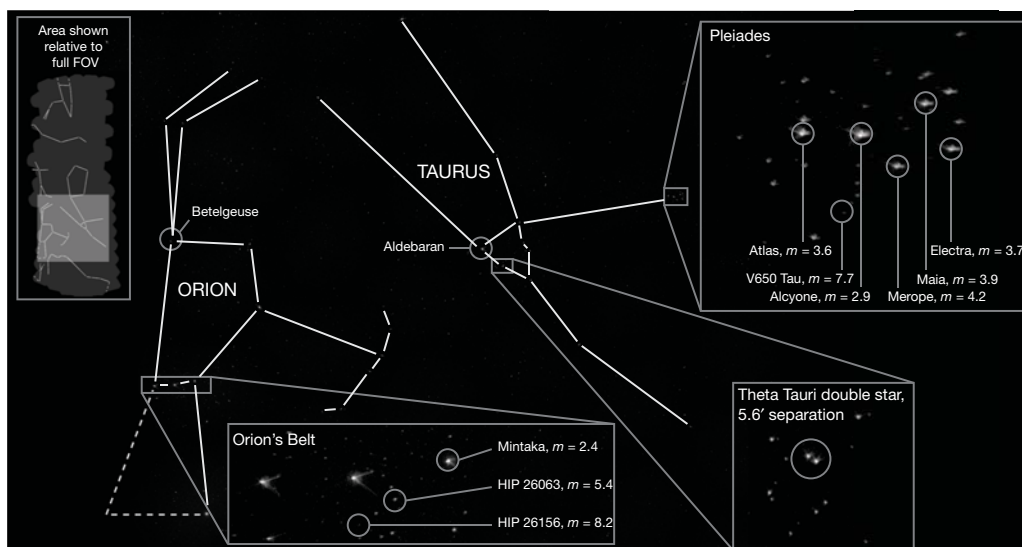
Figure 4 | Details of a star field captured using AWARE-2 with a 1.85-s exposure time. Image data was logarithmically mapped to display values to make better use of the available display dynamic range. Stars with apparent magnitudes of $m < 8.2$ mag are visible in the image. However, those with $m \leq 3.5$ mag saturate the detector at this exposure time.

METHODS SUMMARY

In Figs 1 and 2, all microcameras were set to focus at infinity and had a fixed exposure time of 232 µs. The image was captured on 5 December 2011 at 10:43. Figure 3 was captured using independent auto-exposure and auto-focus in each microcamera. The image shows the Fitzpatrick Center at Duke University on 18 January 2012. In Fig. 4, all microcameras were set to focus at infinity and had a fixed exposure time of 1.85 s. The camera was pointed at a right ascension of 4 h 59 min 2 s and a declination of $37° 05' 45''$ and was located at $36° 00' 50.24''$ N, $79° 00' 13.38''$ W. Cloud cover was near zero. The typical sky brightness at this location is 19.50–18.38 mag arcsec^{-2} in the V band (Bortle scale of approximately 6). The image was captured on 15 January 2012 at 21:44:54 local time.

Auto-exposure initialization requires several seconds and updates in every third frame thereafter. The algorithm steps exposure up and down in logarithmically spaced increments until 1–3% of the pixels are saturated. After exposure is set, the focus motor is stepped in fixed increments. The focus metric is the soft-thresholded sum of the absolute value of the horizontal and vertical gradients independently evaluated over a region of interest, typically a $1,024 \times 1,024$-pixel image centre. Soft-thresholding sets gradient values below a fixed bias to zero, thus emphasizing sharper features. The step direction reverses when the metric at the present position exceeds the metric at the last position. Focal adjustment stops when reverses occur in three subsequent steps or when eight reverses occur in total. Focus resumes if the metric exceeds 30% of the stored minimum value. If the number of saturated pixels leaves the 1–3% target range, auto-exposure and focus reset.

AWARE-2 was pointed at clear sky with a diffuse plastic dome over the gigagon lens location, to calibrate each pixel's gain and illumination. In all figures shown, variable pixel gain and illumination was removed by 'flat-field correction'.

1. Lohmann, A. W. Scaling laws for lens systems. *Appl. Opt.* **28**, 4996–4998 (1989).
2. Brady, D. J. & Hagen, N. Multiscale lens design. *Opt. Express* **17**, 10659–10674 (2009).
3. Marks, D. L. & Brady, D. J. in *Imaging Systems 2010*, paper ITuC2 (Optical Society of America, 2010).
4. Marks, D. L., Tremblay, E. J., Ford, J. E. & Brady, D. J. Microcamera aperture scale in monocentric gigapixel cameras. *Appl. Opt.* **50**, 5824–5833 (2011).
5. Son, H. S., Marks, D. L., Hahn, J., Kim, J. & Brady, D. J. Design of a spherical focal surface using close-packed relay optics. *Opt. Express* **19**, 16132–16138 (2011).
6. Tremblay, E. J., Marks, D. L., Brady, D. J. & Ford, J. E. Design and scaling of monocentric multiscale imagers. *Appl. Opt.* (in the press).
7. Cossairt, O., Miau, D. & Nayar, S. K. in *Proc. IEEE Internat. Conf. Comput. Photogr.* http://dx.doi.org/10.1109/ICCPHOT.2011.5753115 (IEEE, 2011).
8. Ford, J. E. & Tremblay, E. in *Imaging Systems 2010*, paper IMC2 (Optical Society of America, 2010).
9. Marks, D. L. & Brady, D. J. Close-up imaging using microcamera arrays for focal plane synthesis. *Opt. Eng.* **50**, 033205–033209 (2011).
10. Leininger, B. *et al.* Autonomous real-time ground ubiquitous surveillance-imaging system (ARGUS-IS). *Proc. SPIE* **6981**, 69810H (2008).
11. Pollock, D. B. *et al.* Multi-lens array system and method. US patent application 11/456 745 (2006).
12. Kahn, S. M. *et al.* Design and development of the 3.2 gigapixel camera for the Large Synoptic Survey Telescope. *Proc. SPIE* **7735**, 77350J (2010).
13. Kaiser, N. Pan-STARRS: a wide-field optical survey telescope array. *Proc. SPIE* **5489**, 11–22 (2004).
14. Reinhard, E., Ward, G., Pattanaik, S. & Debevec, P. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting* 187–191 (Morgan Kaufmann, 2006).

First dairying in green Saharan Africa in the fifth millennium BC

Julie Dunne[1], Richard P. Evershed[1], Mélanie Salque[1], Lucy Cramp[1], Silvia Bruni[2], Kathleen Ryan[3], Stefano Biagetti[4] & Savino di Lernia[4,5]

In the prehistoric green Sahara of Holocene North Africa—in contrast to the Neolithic of Europe and Eurasia—a reliance on cattle, sheep and goats emerged as a stable and widespread way of life, long before the first evidence for domesticated plants or settled village farming communities[1–3]. The remarkable rock art found widely across the region depicts cattle herding among early Saharan pastoral groups, and includes rare scenes of milking; however, these images can rarely be reliably dated[4]. Although the faunal evidence provides further confirmation of the importance of cattle and other domesticates[5], the scarcity of cattle bones makes it impossible to ascertain herd structures via kill-off patterns, thereby precluding interpretations of whether dairying was practiced. Because pottery production begins early in northern Africa[6] the potential exists to investigate diet and subsistence practices using molecular and isotopic analyses of absorbed food residues[7]. This approach has been successful in determining the chronology of dairying beginning in the 'Fertile Crescent' of the Near East and its spread across Europe[8–11]. Here we report the first unequivocal chemical evidence, based on the $\delta^{13}C$ and $\Delta^{13}C$ values of the major alkanoic acids of milk fat, for the adoption of dairying practices by prehistoric Saharan African people in the fifth millennium BC. Interpretations are supported by a new database of modern ruminant animal fats collected from Africa. These findings confirm the importance of 'lifetime products', such as milk, in early Saharan pastoralism, and provide an evolutionary context for the emergence of lactase persistence in Africa.

It is widely accepted that African pastoralism with cattle, sheep and goats emerged long before plant domestication[2], in contrast to the process of 'neolithization' in the Near East, characterized by the transition from a mobile hunter-gatherer lifestyle to an increasingly settled, agricultural way of life. In Saharan Africa, during the Early Holocene, largely sedentary and pottery-producing hunters, fishers and gatherers became nomadic cattle herders[3], dynamically adapting to, and exploiting, different environments and resources.

Today, it seems impossible that cattle could survive in such a hostile environment as the arid desert land of the Sahara, but this region enjoyed vastly more favourable climatic and environmental conditions[12] during the Holocene African Humid Period, which began around 10,000 years ago[13]. Here, faunal evidence demonstrates that by the early sixth millennium BC, cattle, sheep and goats were found together across the savannas of what is now the Sahara[1,5]. This suggests that the inception of dairying practices in North Africa and an early and independent 'secondary products' economy[14] seems plausible given what we now know of the first appearance of milking in the Near East[8].

Compelling evidence of prehistoric cattle herding in northern Africa comes from the remarkable rock paintings and engravings of the Sahara (Fig. 1), possibly the world's largest concentration of prehistoric art, long known for their rich and vivid portrayal of scenes from everyday life[4,15,16]. The extensive rock art demonstrates that cattle played an important part in the lives and ideology of ancient human groups living in this region during the Holocene. This pictorial record contains countless scenes with representations of cattle, some emphasizing the female's full udders and, in a few cases, depictions of the actual milking of a cow, such as at Wadi Teshuinat II[15] in the Acacus or Wadi Tiksatin in the Messak[16]. However, reliable dates for this rock art can rarely be ascertained[4].

Faunal remains from securely dated contexts indicate that domesticated animals (cattle, sheep or goats) were present in the area from the early sixth millennium BC, becoming much more common in the fifth millennium BC. Unfortunately, these remains are highly fragmented and poorly preserved, precluding herd reconstructions, and thus even indirect evidence of dairying is missing[5].

Direct evidence for the practice of dairying, beginning in the seventh millennium BC in northwestern Anatolia[8], appearing in the sixth millennium BC in eastern Europe[11] and reaching Britain in the fourth millennium BC[9,10], has been established through the compound-specific stable carbon isotope analysis of animal fat residues preserved in archaeological pottery. Notably, this research on the antiquity of dairying practices has largely been confined to Europe, the Near East and Eurasia, with no attempt yet being made to identify the inception of dairying practices in the African continent.

Here we present direct chemical evidence for early dairying practices within the central Sahara through the use of gas chromatography (GC), gas chromatography–mass spectrometry (GC–MS) and gas chromatography–combustion–isotope ratio mass spectrometry (GC–C–IRMS) analyses carried out on organic residues extracted from archaeological pottery sampled from the Takarkori rock shelter located in the southwest Fezzan, Libyan Sahara, an area licensed to Sapienza University of Rome (Supplementary Fig. 1). Four seasons of fieldwork identified evidence of Late Acacus (hunter-gatherer) occupation followed by Early, Middle and Late Pastoral remains (Supplementary Fig. 2), dating between approximately 8100 and 2600 BC (Supplementary Table 1). This long Pastoral period, between approximately 6000 and 2600 BC, denotes the adoption of cattle together with sheep and goats, combined with intensive exploitation of wild cereals[17,18].

Analyses of absorbed organic residues focused on 81 potsherds (Supplementary Table 2), covering a wide range of decorative techniques and motifs found on Saharan ceramics[18,19] (Supplementary Fig. 3). These vessels were mainly excavated from securely dated Middle Pastoral ($n = 56$) levels (approximately 5200–3800 BC), with a small number originating from the Late Acacus ($n = 8$) and Early ($n = 14$) and Late Pastoral ($n = 3$) periods. The lipids were extracted using established protocols[8–10]. Many potsherds demonstrated extraordinary preservation of lipids, containing concentrations of up to $6\,mg\,g^{-1}$ (mean $1.2\,mg\,g^{-1}$), with one particular potsherd (TAK

Figure 1 | Rock art image and tracing from Teshuinat II rock shelter, South West Libya. a, b, Rock art image (**a**) and tracing (**b**) showing Saharan pastoralists with their pots and cattle (adapted with permission from ref. 15).

443) having a concentration of $17\,mg\,g^{-1}$. It is noteworthy that lipids were observed in every potsherd, in contrast to European archaeological sites, where generally <40% of potsherds contain extractable lipids with mean concentrations of approximately $0.1\,mg\,g^{-1}$ (refs 10, 20). This remarkable preservation is likely to be related to the extremely arid conditions prevailing in the region.

Lipid biomarker analyses by GC–MS showed that residues fall into three broad categories (Fig. 2). The most common distribution (Fig. 2a) was dominated by high abundances of the $C_{16:0}$ and $C_{18:0}$ fatty acids, which derive from degraded animal fats. Also abundant were branched-chain fatty acids, C_{13} to C_{18}, components of bacterial origin diagnostic of ruminant animal fats[21]. The second most common type of residue (Fig. 2c) contained a relatively low abundance of the $C_{18:0}$ alkanoic acid, with several extracts showing high abundances of C_{12} and C_{14} homologues. Such distributions have rarely been seen in European pottery and are more diagnostic of plant oils[22]. Also present, and again rarely seen in European prehistoric ceramics, are a homologous series of long-chain n-alkanes from C_{16}–C_{33} (odd-over-even carbon number predominance), usually maximising at C_{25}, regarded as originating from epicuticular waxes of vascular plants[23]. A third, intermediate category of residue (Fig. 2b) is characterized by a series of α,ω-dicarboxylic acids (diacids), in the C_5 to C_{18} carbon-chain-length range (Fig. 2a, b) with C_9 (azelaic acid) the most abundant homologue, the latter commonly deriving from the 'drying reaction' of plant oils[24]. Such residues also contained long-chain alkyl lipids of plant origin. Together such mixtures probably reflect either processing of both plant and animal products in the vessels or the multi-use of vessels.

Of the 81 potsherds, only those residues unambiguously assigned as degraded animal fats (Table 1)—that is, those dominated by palmitic ($C_{16:0}$) and stearic ($C_{18:0}$) alkanoic acids (for example, Fig. 2a)—were selected for GC–C–IRMS analysis to determine the $\delta^{13}C$ values for $C_{16:0}$ and $C_{18:0}$, with the aim of establishing their origins. Differences in the $\delta^{13}C$ values of $C_{16:0}$ and $C_{18:0}$ alkanoic acids are due to the

differential routing of dietary carbon and fatty acids during the synthesis of adipose and dairy fats in ruminant animals, thus allowing ruminant milk fatty acids to be distinguished from carcass fats by calculating $\Delta^{13}C$ values ($\delta^{13}C_{18:0}$–$\delta^{13}C_{16:0}$) and plotting that against the $\delta^{13}C$ value of the $C_{16:0}$ alkanoic acid. Previous research has shown that by plotting $\Delta^{13}C$ values, variations in C_3 versus C_4 plant consumption are removed, thereby emphasizing biosynthetic and metabolic characteristics of the fat source[9,10]. We have now confirmed this through the GC–C–IRMS of a new reference collection of modern ruminant animal fats from Africa collected to encompass the range of carbon isoscapes[25] likely to have been encountered by early Saharan pastoralists. The $\delta^{13}C$ values of the $C_{16:0}$ and $C_{18:0}$ components of these modern fats, presented in Fig. 3, show $\delta^{13}C$ values of goat dairy fats from the Acacus region, Libya ($n = 9$), together with cattle dairy fats and cattle, sheep and goat adipose fats ($n = 9$, 12, 7 and 12, respectively) from Kenya. The $\delta^{13}C_{16:0}$ values for the $C_{16:0}$ alkanoic acids of the African reference fats plot in the range from −35 to −15‰, indicating diets ranging from predominantly C_3 to C_4. These results confirm the global applicability of the $\Delta^{13}C$ proxy.

Of the 29 animal fat residues selected for GC–C–IRMS analyses, 22 originate from Middle Pastoral levels, 3 from the Late Acacus, 2 from the Early Pastoral and the remaining 2 from the Late Pastoral period (Table 1). The comparison of the $\Delta^{13}C$ values of the modern reference animal fats with those of the archaeological pottery residues from the Middle Pastoral period (approximately 5200–3800 BC) show that 50% of these plot within, or on the edge of, the isotopic ranges for dairy fats, with a further 33% falling within the range for ruminant adipose fats and the remainder corresponding to non-ruminant carcass fats (Fig. 3). Notably, the residues originating from earlier periods do not contain dairy fats, and plot in the non-ruminant fat range, probably deriving from wild fauna found locally. The unambiguous conclusion is that the appearance of dairy fats in pottery correlates with the more abundant presence of cattle bones in the cave deposits, suggesting
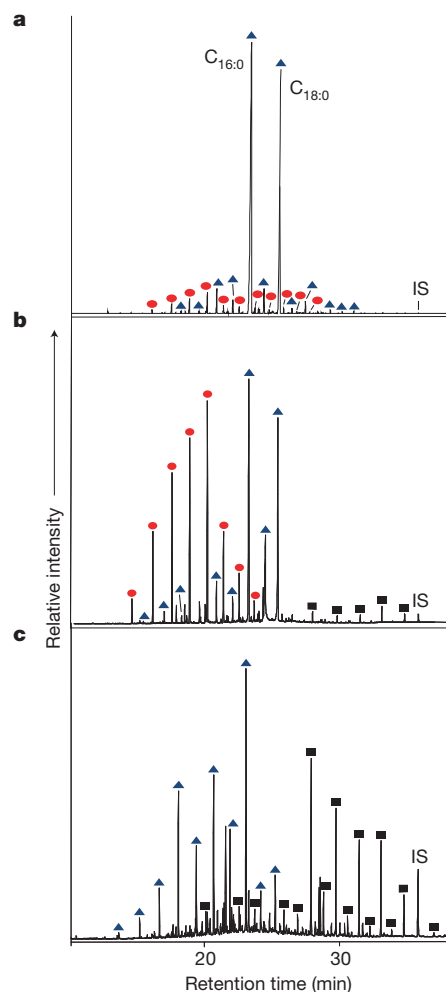
Figure 2 | Partial gas chromatograms displaying the trimethylsilylated lipid extract from potsherds excavated from Middle Pastoral levels in the Takarkori rock shelter. a–c, The distributions are characteristic of degraded animal fat (**a**), a mixture of animal and plant fats (**b**) and plant material (**c**). Chromatographic peak identities denoted by filled triangles comprise straight-chain fatty acids in the carbon chain range $C_{9:0}$ to $C_{29:0}$, maximizing at $C_{16:0}$; filled squares represent *n*-alkanes in the carbon-chain range $C_{20:0}$ to $C_{35:0}$; and filled circles indicate α,ω-dicarboxylic acids (diacids) in the carbon-chain range $C_{5:0}$ to $C_{16:0}$. IS, internal standard, C_{34} *n*-tetratriacontane.

a full pastoral economy as the cattle were intensively exploited for their secondary products.

Of particular note is the wide range of $\delta^{13}C$ values exhibited by the alkanoic acids, plotting across the range −25‰ to −10‰, which is broader even than the reference fats range (maximum −15‰). This suggests that the animals giving rise to these fats had subsisted on an extensive range of different forages either composed completely of C_3 plants, varying combinations of C_3 and C_4 plants, or a diet comprising wholly C_4 plants. The wide range of alkanoic acid $\delta^{13}C$ values found for these African potsherds is unprecedented and points to differing pastoral modes of subsistence (such as vertical transhumance, which is still practised today) by these prehistoric Saharan groups. This is supported by their settlement pattern based on summer sites in the lowland sand seas and winter sites (such as Takarkori) in the mountains[3], which was probably in response to seasonal weather patterns.

Our findings provide unequivocal evidence for extensive processing of dairy products in pottery vessels in the Libyan Sahara during the Middle Pastoral period (approximately 5200–3800 BC), confirming that milk played an important part in the diet of these prehistoric pastoral people. The findings are notable for three other reasons: (1) they confirm that domesticated cattle, used as part of a dairying economy, were present in North Africa during the fifth millennium BC, thus supporting the idea of an earlier ingression into the central Sahara[1–3] and suggesting a local process of pastoral development, based on the exploitation of secondary products; (2) the finding of dairy fat residues in pottery is consistent with milk being processed, thereby

Table 1 | Subset of potsherds selected for isotopic analyses

Potsherd number	Laboratory code	Period	Wall decoration technique	Diameter at mouth (cm)	Part of vessel	Lipid concentration (µg g⁻¹)	$\delta^{13}C_{16.0}$	$\delta^{13}C_{18.0}$	$\Delta^{13}C$	Classification
21	TAK21A	Middle Pastoral	Plain edge fishnet	Not known	Wall + base	5,830.6	−14.7	−20.5	−5.8	Dairy
26	TAK1	Middle Pastoral	APS return	20	Rim + wall	760.7	−14.2	−15.0	−0.9	Ruminant adipose
45	TAK45	Middle Pastoral	APS return	Not known	Wall	639.8	−21.9	−24.1	−2.1	Ruminant adipose
120	TAK120	Middle Pastoral	APS return	Not known	Wall	5,592.7	−15.2	−18.7	−3.5	Dairy
124	TAK124	Middle Pastoral	Ridged APS return	Not known	Wall	1,615.5	−18.1	−20.1	−2.0	Ruminant adipose
197	TAK197	Middle Pastoral	APS return	Not known	Wall	151.5	−20.9	−21.1	−0.2	Non-ruminant adipose
420	TAK420	Middle Pastoral	APS return, triangles	Not known	Wall	1,119.3	−18.3	−21.5	−3.2	Dairy
443	TAK443	Middle Pastoral	APS return	Not known	Wall	17,217.6	−16.9	−23.7	−6.8	Dairy
576	TAK6	Middle Pastoral	Ridged plain edge	30	Rim + wall	800.2	−22.0	−21.7	0.3	Non-ruminant adipose
748	TAK9	Middle Pastoral	Ridged APS return	22	Rim + wall	5,650.5	−13.7	−19.0	−5.2	Dairy
824	TAK11	Late Pastoral	Continuous plain edge and impressed dashes	Not known	Wall + base	4,994.2	−20.5	−24.9	−4.4	Dairy
873	TAK873	Middle Pastoral	APS return	Not known	Wall	71.8	−18.5	−17.7	0.8	Non-ruminant adipose
896	TAK896	Middle Pastoral	APS return	Not known	Wall	218.0	−23.6	−25.0	−1.5	Ruminant adipose
987	TAK987	Middle Pastoral	APS return	Not known	Rim + wall	4,442.6	−13.6	−19.3	−5.7	Dairy
997	TAK15	Middle Pastoral	APS return	16	Rim + wall	1,117.4	−13.3	−17.4	−4.1	Dairy
1009	TAK1009	Middle Pastoral	APS paired lines, banded	Not known	Wall	1,555.7	−11.0	−11.0	0.0	Non-ruminant adipose
1012	TAK1012	Middle Pastoral	Irregular APS return	Not known	Wall	3,591.2	−14.9	−16.5	−1.7	Ruminant adipose
1572	TAK1572	Middle Pastoral	APS return	Not known	Wall	3,148.5	−23.7	−28.2	−4.5	Dairy
1693	TAK21	Late Acacus	Undecorated	Not known	Rim + wall	20.1	−23.1	−19.8	3.3	Non-ruminant adipose
1797	TAK24	Early Pastoral	Combined plain edge with semilunar-motif impressed dots	30	Rim + wall	1,674.6	−21.9	−21.0	0.9	Non-ruminant adipose
1846	TAK25	Middle Pastoral	Plain edge continuous	Not known	Wall + base	819.2	−15.6	−19.7	−4.1	Dairy
1863	TAK26	Middle Pastoral	Ridged APS return	16	Rim + wall	175.0	−22.3	−26.2	−4.0	Dairy
1903	TAK27	Early Pastoral	Cord impression	20	Rim + wall	308.5	−22.8	−21.7	1.1	Non-ruminant adipose
2028	TAK2028	Middle Pastoral	APS, paired lines of dashes	Not known	Wall + base	1,931.0	−24.5	−28.9	−4.4	Dairy
2251	TAK28	Middle Pastoral	APS (row of impressed dots)	14	Rim	96.9	−21.5	−24.0	−2.5	Ruminant adipose
2523	TAK29	Late Pastoral	Undecorated	Not known	Rim + wall	445.6	−18.5	−19.7	−1.2	Ruminant adipose
2588	TAK30	Late Acacus	UNCL (impressed dots)	30	Rim + wall	823.3	−13.9	−13.8	0.1	Non-ruminant adipose
2817	TAK32	Late Acacus	Rocker packed	Not known	Rim	6,882.8	−19.3	−17.5	1.8	Non-ruminant adipose
2857	TAK35	Middle Pastoral	APS return continuous, triangles	Not known	Rim + wall	238.3	−20.1	−22.9	−2.8	Ruminant adipose

APS, alternately pivoting stamp.
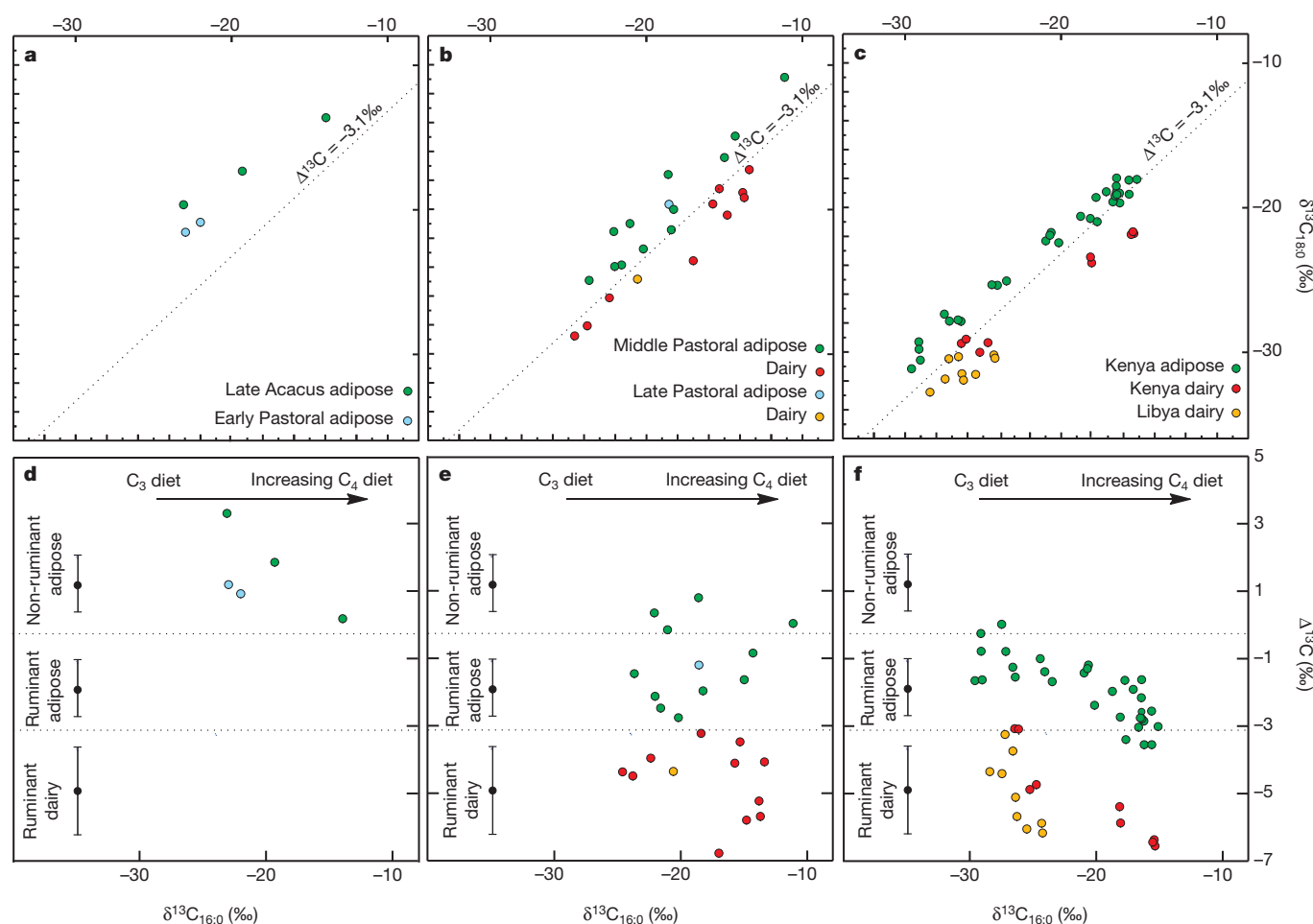
Figure 3 | Plots of δ¹³C values and Δ¹³C values of alkanoic acids in modern reference ruminant fats and archaeological animal fat residues in prehistoric Saharan pottery. a–c, Plots of the $\delta^{13}C_{16:0}$ and $\delta^{13}C_{18:0}$ values for archaeological animal fat residues in Late Acacus (hunter-gatherer) and Early Pastoral (Neolithic) pottery (**a**), archaeological animal fat residues in Middle and Late Pastoral Neolithic pottery (**b**), and modern reference animal fats collected from Libya and Kenya (**c**). **d–f,** Plots denote $\Delta^{13}C$ values for the archaeological fat residues (Late Acacus/Early Pastoral) (**d**) and Middle/Late Pastoral (**e**) and modern reference animal fats (**f**). Notably, the residues originating from the Late Acacus and Early Pastoral periods (**d**) do not contain dairy fats, and plot in the non-ruminant range, probably deriving from wild

fauna. **e,** The extensive processing of dairy products in pottery vessels from this region begins in the Middle Pastoral period approximately 5200–3800 BC. The broad array of values suggests that the animals giving rise to these ruminant fats subsisted on an extensive range of different diets either composed completely of C_3 plants, varying amounts of C_3 and C_4 plants or, for some of the archaeological samples, a diet comprising wholly C_4 plants. The ranges shown here represent the mean ± 1 standard deviation of the $\Delta^{13}C$ values for a global database comprising modern reference ruminant animal fats from Africa, the UK (animals raised on a pure C_3 diet)[10], Kazakhstan[28], Switzerland[29] and the Near East[30], published elsewhere.

providing an explanation of how, in spite of lactose intolerance, milk products could be consumed by these people with the practice being adopted quickly; (3) they are consistent with the finding of the $-13910*T$ allele, associated with the lactase persistence trait in Europeans, across some Central African groups such as the Fulbe from northern Cameroon[26], supporting arguments for some movement of people, together with their cattle, from the Near East into eastern Africa in the Early to Middle Holocene; and (4) they provide a context for understanding the origins and spread of other, independently arising LP-associated gene variants in sub-Saharan Africa[27].

METHODS SUMMARY

A total of 81 potsherds were sampled from the Takarkori rock shelter, Tadrart Acacus Mountains, Libyan Sahara, of which 56 were excavated from the Middle Pastoral period and the remainder originated from the Late Acacus (*n* = 8), and Early (*n* = 14) and Late Pastoral (*n* = 3) periods (Supplementary Table 2). Of the 81 potsherds analysed, 29 were selected for GC–C–IRMS analysis; of these, 18 showed clear evidence of pure animal fat origin, with the remaining 11 comprising lipid profiles suggestive of the mixing of animal and plant fats.

Lipid analysis and interpretations were performed using established protocols described in detail in earlier publications[8–10]. Briefly, ~2 g of potsherd was

sampled and surfaces were cleaned with a modelling drill to remove any exogenous lipids. The potsherds were then ground to a powder, an internal standard was added, and solvent was extracted by ultrasonication (chloroform/methanol 2:1 v/v, 2 × 10 ml). The solvent was evaporated under a gentle stream of nitrogen to obtain the total lipid extract (TLE). Aliquots of the TLE were trimethylsilylated (*N,O*-bis(trimethylsilyl)trifluoroacetamide 80 μl, 70 °C, 60 min), and submitted to analysis by GC and GC–MS. Further aliquots of the TLE were treated with NaOH/H_2O (9:1 w/v) in methanol (5% v/v, 70 °C, 1 h). Following neutralization, lipids were extracted into chloroform and the excess solvent was evaporated under a gentle stream of nitrogen. Fatty acid methyl esters (FAMEs) were prepared by reaction with BF_3-methanol (14% w/v, 70 °C, 1 h). The FAMEs were extracted with chloroform and the solvent removed under nitrogen. The FAMEs were re-dissolved into hexane for analysis by GC and GC–C–IRMS.

FAMEs of freeze-dried reference fats (typically using 5 mg of TLEs) were prepared exactly as above.

1. Gifford-Gonzalez, D. & Hanotte, O. Domesticating animals in Africa: implications of genetic and archaeological findings. *J. World Prehist.* **24,** 1–23 (2011).
2. Marshall, F. & Hildebrand, E. Cattle before crops: the beginnings of food production in Africa. *J. World Prehist.* **16,** 99–143 (2002).

3. di Lernia, S. in *Droughts, Food and Culture: Ecological Change and Food Security in Africa's Later Prehistory* (ed. Hassan, F. A.) 225–250 (Kluwer Academic/Plenum, 2002).
4. di Lernia, S. & Gallinaro, M. The date and context of Neolithic rock art in the Sahara: engravings and ceremonial monuments from Messak Settafet (south-west Libya). *Antiquity* **84,** 954–975 (2010).
5. Gautier, A. in *Droughts, Food and Culture: Ecological Change and Food Security in Africa's Later Prehistory* (ed. Hassan, F. A.) 195–208 (Kluwer Academic/Plenum, 2002).
6. Huysecom, E. *et al.* The emergence of pottery in Africa during the tenth millennium cal BC: new evidence from Ounjougou (Mali). *Antiquity* **83,** 905–917 (2009).
7. Evershed, R. P. Organic residue analysis in archaeology: the archaeological biomarker revolution. *Archaeometry* **50,** 895–924 (2008).
8. Evershed, R. P. *et al.* Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* **455,** 528–531 (2008).
9. Copley, M. S. *et al.* Direct chemical evidence for widespread dairying in prehistoric Britain. *Proc. Natl Acad. Sci. USA* **100,** 1524–1529 (2003).
10. Dudd, S. N. & Evershed, R. P. Direct demonstration of milk as an element of archaeological economies. *Science* **282,** 1478–1481 (1998).
11. Craig, O. E. *et al.* Did the first farmers of central and eastern Europe produce dairy foods? *Antiquity* **79,** 882–894 (2005).
12. Cremaschi, M., Zerboni, A., Spötl, C. & Felletti, F. The calcareous tufa in the Tadrart Acacus Mt. (SW Fezzan, Libya) An early Holocene palaeoclimate archive in the central Sahara. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **287,** 81–94 (2010).
13. deMenocal, P. *et al.* Abrupt onset and termination of the African Humid Period: rapid climate responses to gradual insolation forcing. *Quat. Sci. Rev.* **19,** 347–361 (2000).
14. Sherratt, A. The secondary exploitation of animals in the Old World. *World Archaeol.* **15,** 90–104 (1983).
15. Gallinaro, M. *et al.* in *La Memoria dell'arte. Le pitture rupestri dell'Acacus tra passato e futuro* (eds di Lernia, S. & Zampetti, D.) 73–255 (All'Insegna del Giglio, 2008).
16. Lutz, R. & Lutz, G. *The Secret of the Desert: the Rock Art of the Messak Settafet and Messak Mellet, Libya* (Golf, 1995).
17. Biagetti, S. & di Lernia, S. in *On Shelter's Ledge: Histories, Theories and Methods of Rockshelter Research* (eds Kornfeld, M., Vasil'ev, S. & Miotti, L.) 125–132 (British Archaeological Reports S1655, Archaeopress, 2007).
18. Biagetti, S., Merighi, F. & di Lernia, S. Decoding an Early Holocene Saharan stratified site. Ceramic dispersion and site formation processes in the Takarkori rock-shelter, Acacus Mountains, Libya. *J. Afric. Archaeol.* **2,** 3–21 (2004).
19. Caneva, I. in *Archaeology and Environment in the Libyan Sahara. The Excavations in the Tadrart Acacus, 1978–1983* (ed. Barich, B.) 231–254 (British Archaeological Reports S368, Archaeopress, 1987).
20. Charters, S. *et al.* Quantification and distribution of lipid in archaeological ceramics: implications for sampling potsherds for organic residue analysis. *Archaeometry* **35,** 211–223 (1993).
21. Christie, W. W. *Lipid Metabolism in Ruminant Animals* (Pergamon, 1981).
22. Copley, M. S. *et al.* Detection of palm fruit lipids in archaeological pottery from Qasr Ibrim, Egyptian Nubia. *Proc. R. Soc. Lond. B* **268,** 593–597 (2001).
23. Tulloch, A. P. in *Chemistry and Biochemistry of Natural Waxes* (ed. Kolattukudy, P. E.) 235–287 (Elsevier, 1976).
24. Mills, J. S. & White, R. *The Organic Chemistry of Museum Objects* 2nd edn (Elsevier, 1987).
25. West, J. B., Bowen, G. J., Dawson, T. E. & Tu, K. P. *Isoscapes: Understanding Movement, Pattern and Process on Earth through Isotope Mapping* (Springer, 2010).
26. Mulcare, C. A. *et al.* The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (*LCT*) (*C−13.9kbT*) does not predict or cause the lactase-persistence phenotype in Africans. *Am. J. Hum. Genet.* **74,** 1102–1110 (2004).
27. Itan, Y., Jones, B. L., Ingram, C. J., Swallow, D. M. & Thomas, M. G. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol. Biol.* **10,** 36 (2010).
28. Outram, A. K. *et al.* The earliest horse harnessing and milking. *Science* **323,** 1332–1335 (2009).
29. Spangenberg, J. E., Jacomet, S. & Schibler, J. Chemical analyses of organic residues in archaeological pottery from Arbon Bleiche 3, Switzerland—evidence for dairying in the late Neolithic. *J. Archaeol. Sci.* **33,** 1–13 (2006).
30. Gregg, M. W., Banning, E. B., Gibbs, K. & Slater, G. F. Subsistence practices and pottery use in Neolithic Jordan: molecular and isotopic evidence. *J. Archaeol. Sci.* **36,** 937–946 (2009).

LETTER

The clonal and mutational evolution spectrum of primary triple-negative breast cancers

Sohrab P. Shah[1,2]*, Andrew Roth[1,2]*, Rodrigo Goya[3]*, Arusha Oloumi[1,2]*, Gavin Ha[1,2]*, Yongjun Zhao[3]*, Gulisa Turashvili[1,2]*, Jiarui Ding[1,2]*, Kane Tse[3]*, Gholamreza Haffari[1,2]*, Ali Bashashati[1,2]*, Leah M. Prentice[1,2], Jaswinder Khattra[1,2], Angela Burleigh[1,2], Damian Yap[1,2], Virginie Bernard[4], Andrew McPherson[1,2], Karey Shumansky[1,2], Anamaria Crisan[1,2], Ryan Giuliany[1,2], Alireza Heravi-Moussavi[1,2], Jamie Rosner[1,2], Daniel Lai[1,2], Inanc Birol[3], Richard Varhol[3], Angela Tam[3], Noreen Dhalla[3], Thomas Zeng[3], Kevin Ma[3], Simon K. Chan[3], Malachi Griffith[3], Annie Moradian[3], S.-W. Grace Cheng[3], Gregg B. Morin[3,5], Peter Watson[1,6], Karen Gelmon[6], Stephen Chia[6], Suet-Feung Chin[7,8], Christina Curtis[7,8,9], Oscar M. Rueda[7,8], Paul D. Pharoah[7], Sambasivarao Damaraju[10], John Mackey[10], Kelly Hoon[11], Timothy Harkins[11], Vasisht Tadigotla[11], Mahvash Sigaroudinia[12], Philippe Gascard[12], Thea Tlsty[12], Joseph F. Costello[13], Irmtraud M. Meyer[5,14,15], Connie J. Eaves[16], Wyeth W. Wasserman[4,5], Steven Jones[3,5,17], David Huntsman[1,2,18], Martin Hirst[3,15,19], Carlos Caldas[7,8,20,21], Marco A. Marra[3,5] & Samuel Aparicio[1,2]

Primary triple-negative breast cancers (TNBCs), a tumour type defined by lack of oestrogen receptor, progesterone receptor and *ERBB2* gene amplification, represent approximately 16% of all breast cancers[1]. Here we show in 104 TNBC cases that at the time of diagnosis these cancers exhibit a wide and continuous spectrum of genomic evolution, with some having only a handful of coding somatic aberrations in a few pathways, whereas others contain hundreds of coding somatic mutations. High-throughput RNA sequencing (RNA-seq) revealed that only approximately 36% of mutations are expressed. Using deep re-sequencing measurements of allelic abundance for 2,414 somatic mutations, we determine for the first time—to our knowledge—in an epithelial tumour subtype, the relative abundance of clonal frequencies among cases representative of the population. We show that TNBCs vary widely in their clonal frequencies at the time of diagnosis, with the basal subtype of TNBC[2,3] showing more variation than non-basal TNBC. Although *p53* (also known as *TP53*), *PIK3CA* and *PTEN* somatic mutations seem to be clonally dominant compared to other genes, in some tumours their clonal frequencies are incompatible with founder status. Mutations in cytoskeletal, cell shape and motility proteins occurred at lower clonal frequencies, suggesting that they occurred later during tumour progression. Taken together, our results show that understanding the biology and therapeutic responses of patients with TNBC will require the determination of individual tumour clonal genotypes.

To understand the patterns of somatic mutation in TNBC, we enumerated genome aberrations at all scales from 104 cases of primary TNBC (Affymetrix SNP6.0, 104 cases; RNA-seq, 80 cases; genome/exome sequencing, 65 cases: 54 exomes, 15 genomes with 4 overlapping) (Supplementary Table 1 and Supplementary Fig. 1), annotated with clinical information (Supplementary Table 2). We revalidated 2,414 somatic single nucleotide variants[4,5] (SNVs) (Supplementary Table 3) with targeted deep sequencing to a median of 20,000× coverage, including 43 non-coding splice site dinucleotide mutations (Supplementary Table 4) and 104 genes with 107 indels (Supplementary Table 5 and Supplementary Methods). Notably, the distribution of somatic mutation abundance varies in a continuous distribution among tumours (Fig. 1a) and seems to be unrelated to the proportion of the genome altered by copy number alterations (CNAs) (Fig. 1b) or tumour cellularity (Supplementary Fig. 2b). Although this distribution could be partially explained by a false-negative rate in mutation discovery, others have noted similar distributions in epithelial cancers[6], suggesting that the total mutation content of individual tumours may be shaped by biological processes or differential exposure to mutagenic influences in the population.

The overall pattern (Supplementary Fig. 3a, b) of CNA abundance appears similar (Supplementary Fig. 4) to that seen in a larger, independent series of ~2,000 SNP6.0 profiled breast tumours[7]. Among the most frequently observed CNA events (Supplementary Table 6) are the tumour suppressor and oncogenes *PARK2* (6%), *RB1* (5%), *PTEN* (3%) and *EGFR* (5%). Here we report intragenic deletions (Supplementary Fig. 5) in the *PARK2* tumour suppressor[8,9], specifically linking *PARK2* with TNBC for the first time. Consistent with previous reports in breast cancer[10], we did not observe frequent recurrent structural rearrangements (Supplementary Fig. 3d and Supplementary Table 7), although we revalidated many individual fusion events involving known oncogenes or tumour suppressors (for example, *KRAS*, *RB1*, *IDH1*, *ETV6*) (Supplementary Tables 8–10).

A comparison of RNA-seq data with genomes/exomes data revealed that only 36% of validated somatic SNVs were observed in the transcriptome sequence (Supplementary Table 3 and Supplementary Fig. 2b). In a recent lymphoma study, similar proportions were observed (137 of 329 somatic mutations expressed in RNA-seq)[11]. As expected, the proportion of low-abundance somatic SNVs observed in RNA is reflected in the distribution of wild-type, heterozygous and homozygous expressed mutations (Supplementary Fig. 2b), consistent with the notion that

[1]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. [2]Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia V5Z 1L3, Canada. [3]Canada's Michael Smith Genome Sciences Centre, Vancouver, British Columbia V5Z 1L3, Canada. [4]Centre for Molecular Medicine and Therapeutics, 950 West 28th Avenue, Vancouver, British Columbia V5Z 4H4, Canada. [5]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. [6]British Columbia Cancer Agency, 600 West 10th Avenue, Vancouver, British Columbia V5Z 4E6, Canada. [7]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. [8]Department of Oncology, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK. [9]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. [10]Departments of Oncology and Laboratory Medicine and Pathology, University of Alberta, 11560 University Avenue, Cross Cancer Institute, Edmonton, Alberta T6G 1Z2, Canada. [11]Life Technologies, 101 Lincoln Centre Dr., Foster City, California 94404, USA. [12]Department of Pathology and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California 94143, USA. [13]Brain Tumor Research Center, Department of Neurosurgery, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California 94143, USA. [14]Department of Computer Science, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [15]Centre for High-Throughput Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [16]Terry Fox Laboratory, BC Cancer Agency, 675 W 10th Avenue, Vancouver, British Columbia V5Z 1L3, Canada. [17]Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A1S6, Canada. [18]Centre for Translational and Applied Genomics, BC Cancer Agency, 600 West 10th Ave, Vancouver, British Columbia V5Z 4E6, Canada. [19]Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. [20]Cambridge Breast Unit, Addenbrookes Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. [21]Cambridge Experimental Cancer Medicine Centre (ECMC), Cambridge CB2 0RE, UK.
*These authors contributed equally to this work.

a Somatic mutation content by case

Basal
Other
p53
PIK3CA
PTEN
RB1
EGFR

b Relationship of somatic mutations and CNA by case

Number of HOMDs and HLAMPs
Percentage genome altered

HOMD (0.33%)
HETD (16.3%)
NEUT (67.9%)
GAIN (14.56%)
AMP (0.87%)
HLAMP (0.04%)
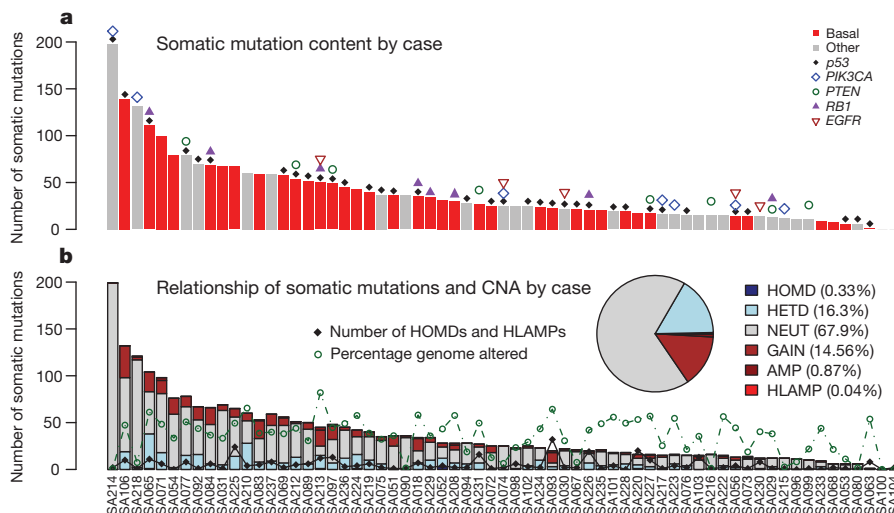
Figure 1 | Distribution of number of validated somatic mutations by case over 65 cases.
a, Mutation frequency (basal, red; other, grey). Patients harbouring known driver gene mutations are indicated. **b**, Case-specific and overall (inset) distributions of mutations in CNA classes. AMP, amplification; GAIN, single copy gain; HETD, hemizygous deletion; HLAMP, high-level amplification; HOMD, homozygous deletion; NEUT, no copy number change. The number of (HOMD, HLAMP) CNAs (black diamonds) and percentage genome altered (green circles) are indicated.

low-abundance alleles may represent rarer clones in the primary tumour. We found 43 splice junction mutations with evidence for an impact on splicing patterns (Supplementary Table 4), encompassing several known tumour suppressors (*p53*, *PIK3R1*; Supplementary Fig. 6) as well as many genes not yet implicated in carcinogenesis. Analysis of 72 somatic mutations in the non-coding space of experimentally determined human regulatory regions[12] showed (Supplementary Table 11) a significant overrepresentation (31.9% versus expected 2.5%, Fisher exact test $P = 2 \times 10^{19}$) of mutations within retinoblastoma-associated protein (RB)-binding sites. Six mutations were predicted to be damaging to RB binding (Supplementary Methods and Supplementary Fig. 7). This is consistent with observations of frequent functional disruption of the RB-regulated cell cycle network[13] in TNBC.

We next searched for mutation enrichment patterns in three ways: by single gene mutation frequency over multiple cases; by the mutation frequency over multiple members of a gene family; and by correlating mutation status with expression networks. First, similar to other studies[14,15], *p53* is the most frequently mutated gene (Supplementary Table 12) with 62% of basal TNBC (determined by gene expression classification with PAM50 (ref. 16) analysis on RNA-seq expression profiles) and 43% of non-basal TNBC cases harbouring a validated somatic mutation. We also observed frequent mutations in *PIK3CA* at 10.2% (7/65), *USH2A* (Usher syndrome gene, implicated in actin cytoskeletal functions) at 9.2% (6/65), *MYO3A* at 9.2%, *PTEN* and *RB1* at 7.7% (5/65) and a further eight genes (including *ATR*, *UBR5* (also known as *EDD1*), *COL6A3*) at 6.2% (4/65) of cases in the cohort

(Fig. 2a). Considering background mutation rates[17], *p53*, *PIK3CA*, *RB1*, *PTEN*, *MYO3A* and *GH1* showed evidence of single gene selection ($q < 0.1$) (Supplementary Table 13). Additional recurrent mutations of note occurred in the synuclein genes (*SYNE1* and *SYNE2*, 9.2% 6/65, recently implicated in squamous head and neck cancers[18,19]), *BRCA2* (three cases), and several other well known oncogenes (*BRAF*, *NRAS*, *ERBB2* and *ERBB3*) with mutations in two cases each. Approximately 20% of cases contained examples of potentially 'clinically actionable' somatic aberrations, including *BRAF* V600E, high-level *EGFR* amplifications and *ERBB2* and *ERBB3* mutations.

In the second approach we searched for statistically overrepresented gene families and protein functions using the Reactome functional protein interaction database[20] (Supplementary Methods). This analysis quantifies gene family involvement through sparse mutation patterns in functionally connected genes, which would be statistically underrepresented by single gene recurrent mutation analysis. The overrepresented pathways (false discovery rate (FDR) < 0.001) included *p53*-related pathways along with chromatin remodelling, PIK3 signalling, ERBB signalling, integrin signalling and focal adhesion, WNT/cadherin signalling, growth hormone and nuclear receptor co-activators, and ATM/RB-related pathways (Fig. 3a and Supplementary Table 14). We note that the candidate 'driver' *MYO3A*, a cytoskeleton motor protein involved in cell shape and motility, relates to several pathways upstream and downstream of integrin signalling. The mutated genes include extracellular matrix (ECM) interactions (laminins, collagens), ECM receptors (integrins), several proteins

a

HLAMP
HOMD
Missense
Truncating
Indel
Splice

b

ER−
ER+
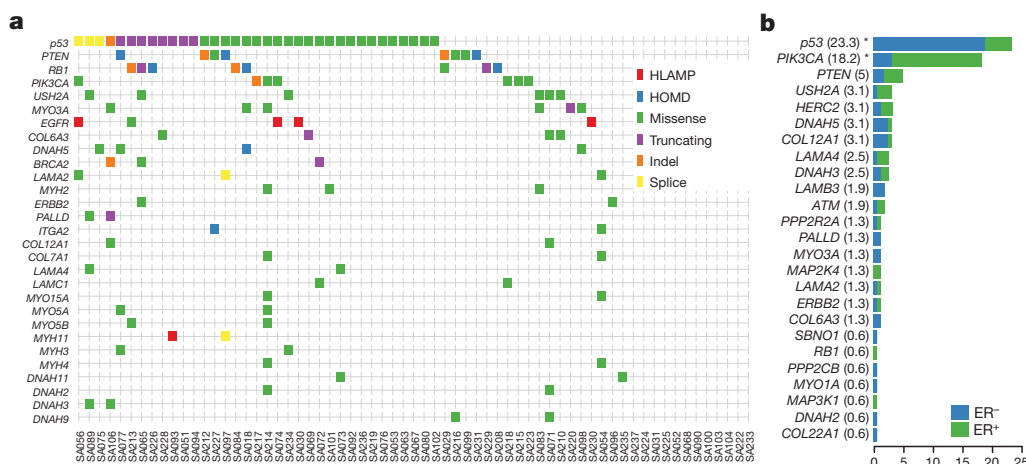
Figure 2 | Population patterns of co-occurrence and mutual exclusion of genomic aberrations in TNBC. a, Case-specific mutations in known driver genes, plus genes from integrin signalling and ECM-related proteins (laminins, collagens, integrins, myosins and dynein) derived from all aberration types: high-level amplifications (HLAMP), homozygous deletions (HOMD), missense, truncating, splice site and indel somatic mutations are depicted in genes with at least two aberrations in the population. **b**, Distribution of somatic mutations in 25 genes across all exons of 159 additional breast cancers (relative proportion of ER+ cases in green, and ER− in blue), shown as a percentage of cases (in parentheses) with one or more mutations. *$P < 0.05$.

regulating actin cytoskeleton dynamics (usherin, palladin, multiple myosins) and microtubule motor proteins (kinesins) (Fig. 2a). All of these contribute to cellular processes that have been functionally implicated in cancer progression; however, a signature of somatic mutation associated with these proteins has not been previously noted in TNBC.

To confirm the mutational spectrum in the general breast cancer population we re-sequenced all exons of 29 genes in an additional 159 breast cancers (82 oestrogen receptor (ER)[+] and 77 ER[−], tumour and matched normal) (Fig. 2b), and confirmed that many of the genes found in the discovery cohort were recurrently mutated in an additional population. Whether this pattern of mutation represents the occurrence of disease-modifying mutations, or possibly selection from other processes (for example, transcription-related hypermutation) is unknown. Interestingly, the enrichment of cytoskeletal functions in the somatic aberration landscape is also evident from the copy number and alternative splicing landscapes (Supplementary Fig. 8).

Third, we integrated both the CNA and mutation data with expression data to reveal genomic events associated with extreme changes in the transcription of interacting genes[20] (Table 1), using a bipartite graph-based method (driverNet; Supplementary Methods). The somatic aberrations showing statistically significant association with extreme expression in this analysis ($P < 0.05$) (Table 1 and Supplementary Table 15) implicate well known oncogenes and tumour suppressors (TP53, PIK3CA, NRAS, EGFR, RB1, ATM) and suggest several new genes of interest, including PRPS2 (a nucleotide biosynthesis enzyme, rank 7), harbouring homozygous deletions in three cases, NRC31 (a glucocorticoid receptor, rank 10) with SNVs in three cases, four PKC-related genes, PRKCZ, PRKCQ, PRKG1 and PRKCE. The gene networks show a partial overlap with driverNet applied to the TCGA ovarian high-grade serous data[21] (Supplementary Table 16).

Having identified candidate driver genes and significantly over-represented pathways, we asked how these are distributed among individual tumours by clustering a pathway–patient-mutation matrix (Supplementary Fig. 9). The abundance of implicated pathways can be

Table 1 | Analysis of the top somatically aberrated genes influencing expression

Rank	Gene	gband	SNV or indel	HLAMP	HOMD	Events	P value
1	TP53	17p13.1	35	0	0	2242	0
2	PIK3CA	3q26.32	7	0	0	441	1×10^{-4}
3	NRAS	1p13.2	2	0	0	271	4×10^{-4}
4	EGFR	7p11.2	1	5	0	220	4×10^{-4}
5	RB1	13q14.2	5	0	5	184	5×10^{-4}
6	PGM2	4p14	1	0	1	172	5×10^{-4}
7	PRPS2	23p22.2	0	0	3	171	5×10^{-4}
8	PTEN	10q23.31	5	0	3	150	5×10^{-4}
9	PRKCE	2p21	0	0	1	136	7×10^{-4}
10	NR3C1	5q31.3	3	0	0	130	7×10^{-4}
11	CREBBP	16p13.3	1	0	1	119	8×10^{-4}
12	CS	12q13.2	1	0	0	108	0.0011
13	MAN2A2	15q26.1	2	0	1	104	0.0012
14	HMGCS2	1p12	1	2	0	100	0.0013
15	HEXA	15q24.1	2	1	0	97	0.0013
16	ADCY9	16p13.3	2	1	0	91	0.0017
17	OR4N4	15q11.2	0	0	5	90	0.0017
18	LCLAT1	2p23.1	0	0	1	85	0.002
19	DGKI	7q33	2	0	0	82	0.0022
20	CYP2A6	19q13.2	1	0	0	80	0.0024
21	JAK1	1p31.3	1	0	0	78	0.0026
22	POLR1A	2p11.2	2	0	0	78	0.0026
23	PLD1	3q26.31	1	0	0	69	0.0038
24	IDH3B	20p13	1	0	1	68	0.004
25	PAPSS2	10q23.2	0	0	3	67	0.0041
26	PRKX	23p22.33	0	0	2	65	0.0046
27	TPH2	12q21.1	1	0	0	65	0.0046
28	UGT2B17	4q13.2	0	0	1	63	0.0053
29	RRM2	2p25.1	1	0	0	57	0.0072
30	ATM	11q22.3	1	0	0	55	0.0084
31	CLCA1	1p22.3	2	0	0	54	0.009
32	PRKCZ	1p36.33	1	0	0	53	0.0095

Rank, derived by the driverNet algorithm (see Supplementary Methods); gene, somatically aberrated gene; gband, chromosomal band containing gene; SNV or indel, the number of cases harbouring an SNV or indel in the gene; HLAMP, the number of cases harbouring a predicted high-level amplification; HOMD, the number of cases harbouring a predicted homozygous deletion; events, number of gene expression outliers (see Supplementary Methods) coincident with a genomic aberration and where the outlying gene is connected to the aberrated gene; P value, statistical significance based on a randomly generated background distribution (Supplementary Methods).

Figure 3 | Network analysis of 254 recurrently mutated genes by somatic point mutations and indels. a, Case-specific mutations shaded according to clonal frequencies in known driver genes, plus genes from integrin signalling and ECM-related proteins (laminins, collagens, integrins, myosins and dyneins). **b**, Significantly overrepresented pathways (FDR < 0.001) from recurrently mutated genes (see Supplementary Methods). Node shading encodes the adjusted P value (q value) of the comparison of the distribution of clonal frequencies of mutations in a given pathway to the overall distribution of clonal frequencies. A spectrum of higher (red) and lower (yellow) clonal frequencies is evident. Letters in parentheses indicate database sources.

seen to be only partially related to the total number of mutations in a case, groups 1 and 2 having on average fewer mutations per case. The frequent involvement of pathways with *p53*, *PTEN* and *PIK3CA* as members, is noted (Supplementary Fig. 9); however, the case groupings also vary by the progressive inclusion of additional pathways (for example, WNT signalling, integrin signalling, ERBB signalling, hypoxia and PI3K). More than two thirds of cases contained one or more mutations in the actin/cytoskeletal functions group of genes (Supplementary Fig. 9). Some 12% of cases did not contain somatic aberrations in any of the frequent drivers or cytoskeletal genes (Supplementary Table 12). This suggests that primary TNBCs are mutationally heterogenous from the outset, with some patients' tumours having a small number of implicated pathways and few mutations, whereas other patients present with tumours containing extensive mutation burdens and multiple pathway involvement.

Motivated by the observation that early primary TNBCs show a wide variation of mutation content, we asked whether the clonal composition of these primary cancers is similarly varied. We and others have shown[22,23] how deep-frequency measurements of allelic abundance can be used to study tumour clonal evolution. Clonal mutation frequency, a compound measure of clonal complexity, (Fig. 4a) can be estimated from allele abundance, once the influence of copy number states, regional loss of heterozygosity (LOH state) and tumour cellularity have been considered (although we note that approximately 68% of SNVs in this study are in diploid, neutral regions). To extend allelic abundance measurements to estimation of clonal frequencies, we implemented a Dirichlet process clustering model (pyclone; Supplementary Methods and Supplementary Fig. 10) that simultaneously estimates the genotype and clonal frequency given a list of deeply sequenced mutations and their local copy number and heterozygosity contexts.

Using the set of deeply sequenced (median 20,000×), validated SNVs, our analysis revealed (Fig. 4b) that groups of mutations within individual cases have different clonal frequencies, indicative of distinct clonal genotypes. Remarkably, the tumours exhibit a wide spectrum of modes over clonal frequencies (Fig. 4b and Supplementary Fig. 11), with some cases showing only one or two frequency modes (Fig. 4b), indicating a smaller number of clonal genotypes, whereas other tumours exhibit multiple clonal frequency modes, indicating more extensive clonal evolution. Consistent with early 'driver gene' status, mutations in known tumour suppressors such as p53 tend to occur in the highest clonal frequency group in most tumours. However, in some cases (for example, SA219, SA236; Fig. 4b, Supplementary Fig. 11) p53 resides in lower-abundance clonal frequency groups (Supplementary Fig. 12 and Fig. 3a), suggesting that it was not the founding event. Although the number of clonal frequency modes tends to increase with the number of mutations, the relationship is not strictly linear (Fig. 4c). To determine whether basal and non-basal cancers differ in their clonality, we compared the distribution of clonal modes (clusters) by case and as an overall distribution, and note that basal TNBCs have more clonal frequency modes than non-basal TNBCs (Fig. 4c). Both of these distributions emphasize a key observation; namely, that at the time of diagnosis TNBCs already display a widely varying clonal evolution that mirrors the variation in mutational evolution.

Finally, we asked where key pathways appear in the distribution of clonal frequency groups. We examined the clonal frequency of genes in each pathway and ascertained if there was a deviation away from the distribution of clonal frequency for all mutations. As expected, pathways involving p53 and *PIK3CA* showed significantly skewed distributions (Wilcoxon, $q < 0.01$; Fig. 3b and Supplementary Fig. 12) towards higher clonal frequencies, consistent with their roles in early tumorigenesis (Fig. 3a and Supplementary Table 17). Intriguingly, pathways with cytoskeletal genes such as myosins, laminins, collagens and integrins tend to have lower median clonal frequencies, suggesting that somatic mutations in these genes are acquired much later (Fig. 3b). Notably, the median clonal frequency for Reactome pathway 'p53
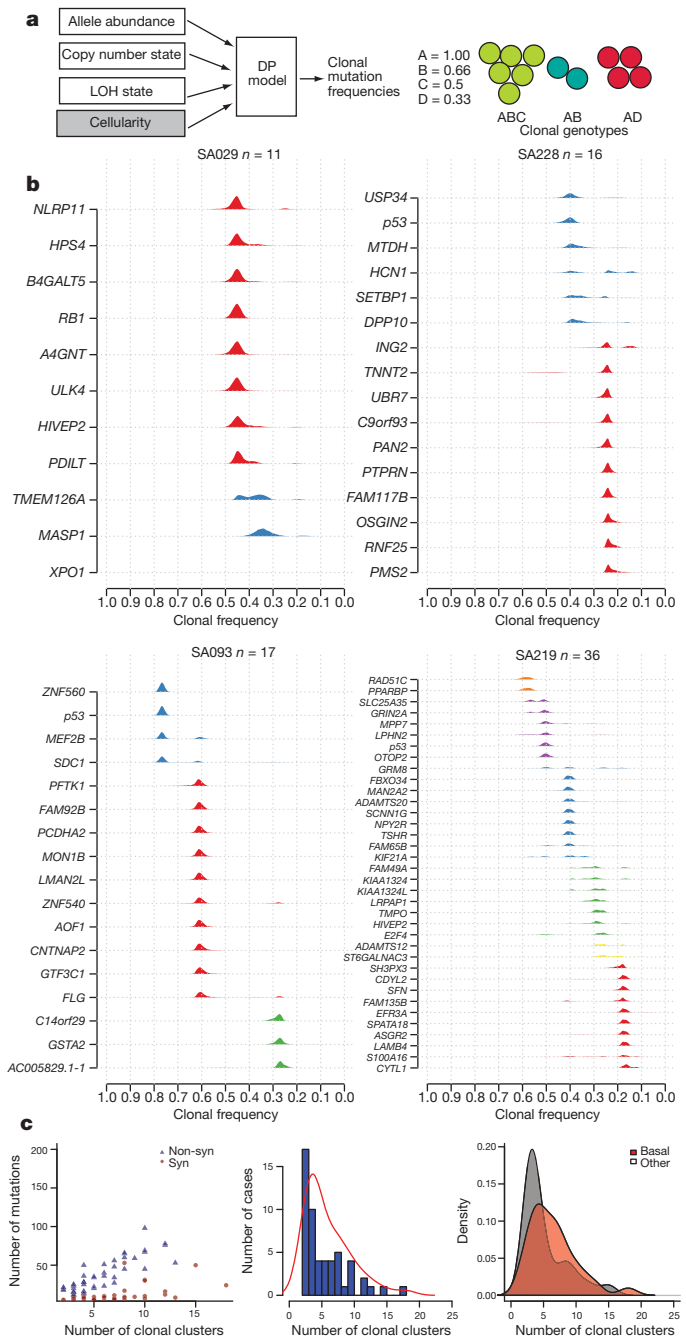
Figure 4 | Clonal evolution in TNBC. a, Schematic representation of integration of CNA, LOH, allelic abundance measurements and normal cell contamination for clonal frequency estimation using a Dirichlet process (DP) model (left). Example of a mixture of three clonal genotypes composed of four mutations (A, B, C, D) and their resulting clonal frequencies. **b**, Estimated clonal frequencies for four cases are shown as the distribution of posterior probabilities from the pyclone model (Supplementary Methods). Clonal frequency distributions are coloured by their frequency group membership. **c**, Left, relationship of mutation abundance (synonymous (Syn) and non-synonymous (Non-syn)) and the inferred number of clonal clusters. Middle, distribution and kernel density (red line) of the number of inferred clonal clusters over 54 TNBCs. Right, kernel density distribution of clonal clusters for basal (red) and non-basal (grey) tumours.

pathway feedback loops', including 46 mutations in *ATM*, *ATR*, *NRAS*, *PIK3CA*, *PTEN*, *SIAH1* and *p53*,was 73% (Wilcoxon, $q = 0.0007$), whereas 'integrin cell surface interactions', including 23 mutations in integrin, laminin and collagen genes, had a median clonal frequency of 42% (Wilcoxon, $q = 0.9569$).

Primary TNBCs are still treated as if they were a single disease entity, yet it is clear they do not behave as a single entity in response to current therapies. Here we show for the first time, using next-generation sequencing mutational profiling methods, that treatment-naive TNBCs display a complete spectrum of mutational and clonal evolution, with some patients' tumours showing only a few somatic coding sequence point mutations with a limited number of molecular pathways implicated, whereas other patients' tumours exhibit considerable additional mutational involvement. Moreover, the clonal heterogeneity of these cancers is also a continuum, with some patients presenting with low-clonality cancers and other cases exhibiting more extensive clonal evolution at diagnosis. In this respect, the basal expression subtype of TNBCs also tends to show higher clonality at diagnosis, although the relationship is not exact.

In clonally evolving tumours, identification of genes by single gene mutation frequency measurements will probably favour early driver genes, because the subsequent involvement of multiple additional pathways during tumour progression is unlikely to be observed as a frequent single gene mutation. The clonality analysis emphasizes this point: known drivers such as *p53*, *PIK3CA* and *PTEN* have among the highest clonal frequencies, whereas mutations in cell shape/motility and ECM-signalling genes appear in the lower clonal frequency groups, distributed over many genes. Although p53 somatic mutations are clearly early events, the clonal frequencies observed in some TNBC suggest that they are not always the first event, raising a question about what drives early clonal expansion in some of these cancers. Our findings suggest that each TNBC at the time of primary diagnosis may be at a very different phase of molecular progression, with possible implications for approaches to the biology of low clonality versus high clonality primary tumours.

METHODS SUMMARY

The genomes and transcriptomes of 104 TNBCs were profiled with Affymetrix SNP6.0 arrays (all cases), RNA-seq (80 cases; Illumina GAII), and whole exome/genome sequencing (65 cases; tumour and normal DNA). Exomes were obtained using Agilent's Human All Exon SureSelect Target Enrichment System v.1 followed by Illumina GAII sequencing, and whole genomes were sequenced using Life Technologies SOLiD system. Data were analysed using computational approaches to detect somatic SNVs[4,5], indels, copy number alterations, gene fusions and gene expression patterns. Predictions were then validated using orthogonal experimental assays, including targeted ultra-deep amplicon sequencing of SNVs to ~20,000× redundancy. We determined single genes under selection using a statistical approach that considers patient-specific background mutation and transition/transversion rates. Mutations predicted to alter transcriptional profiles were determined using an integrated bipartite graph-based method (driverNet) that associates genomic aberrations with outlying expression patterns informed by pre-defined pathway gene sets. Disrupted pathways were determined using the Reactome FI Cytoscape plugin. Clonal analysis was performed (cases with >10 mutations) using a Dirichlet process statistical model that simultaneously estimates clonal frequencies and mutation genotype given deeply sequenced somatic SNVs and copy number estimates. Experimental assays and analytical methodology are detailed in the Supplementary Information.

1. Blows, F. M. *et al.* Subtyping of breast cancer by immunohistemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med.* **7,** e1000279 (2010).
2. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).
3. Sørlie, T. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98,** 10869–10874 (2001).
4. Roth, A. *et al.* JointSNVmix: A probabilistic model for accurate detection of somatic mutations in normal/tumour paired next generation sequencing data. *Bioinformatics* http://dx.doi.org/10.1093/bioinformatics/bts053 (27 January 2012).
5. Ding, J. *et al.* Feature-based classifiers for somatic mutation detection in tumour—normal paired sequencing data. *Bioinformatics* **28,** 167–175 (2012).
6. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455,** 1069–1075 (2008).
7. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* http://dx.doi.org/10.1038/nature10983 (this issue).
8. Poulogiannis, G. *et al.* PARK2 deletions occur frequently in sporadic colorectal cancer and accelerate adenoma development in *Apc* mutant mice. *Proc. Natl Acad. Sci. USA* **107,** 15145–15150 (2010).
9. Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463,** 893–898 (2010).
10. Stephens, P. J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462,** 1005–1010 (2009).
11. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476,** 298–303 (2011).
12. Chicas, A. *et al.* Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17,** 376–387 (2010).
13. Herschkowitz, J. I., He, X., Fan, C. & Perou, C. M. The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas. *Breast Cancer Res.* **10,** R75 (2008).
14. Langerød, A. *et al.* TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. *Breast Cancer Res.* **9,** R30 (2007).
15. Børresen-Dale, A.-L. TP53 and breast cancer. *Hum. Mutat.* **21,** 292–300 (2003).
16. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27,** 1160–1167 (2009).
17. Youn, A. & Simon, R. Identifying cancer driver genes in tumor genome sequencing studies. *Bioinformatics* **27,** 175–181 (2011).
18. Agrawal, N. *et al.* Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333,** 1154–1157 (2011).
19. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333,** 1157–1160 (2011).
20. Wu, G., Feng, X. & Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.* **11,** R53 (2010).
21. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615 (2011).
22. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
23. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464,** 999–1005 (2010).

Author Contributions S.A., S.P.S., C. Caldas and M.A.M. designed and implemented the research plan and wrote the manuscript. S.P.S., A.R., R. Goya, G. Ha, J.D., G. Haffari, A. Bashashati, A. McPherson, K.S., A.C., R. Giuliany, A.H.-M., J.R., D.L., I.B., R.V., S.W.C., M.G., I.M.M., S.J., C. Curtis, O.M.R., P.D.P., V.B. and W.W.W. conducted bioinformatic analyses of the data and/or gave advice on analytic methodology. G.T. conducted histopathological review and immunohistochemistry. A.O., Y.Z., G.T., K.T., L.M.P., J.K., A.B., D.Y., A.T., N.D., T.Z., S.-F.C., K.M. and M.H. conducted sequencing or experimental validation of somatic aberrations. D.Y., A. Moradian, S.-W.G.C. and G.B.M. conducted proteome validation of splicing. P.W., K.G., S.C., S.-F.C., G.T., J.M., C. Caldas, P.D.P. and D.H. collected and interpreted clinical data. S.D., J.F.C., T.T., M.S., P.G. and C.J.E. contributed materials or reagents. K.H., V.T., T.H., M.H. and M.A.M. generated sequence data.

Author Information Aligned exome/genome sequence data, RNA-seq data and Affymetrix SNP6.0 data sets are available at the European Genome-phenome Archive (http://www.ebi.ac.uk/ega/) under study accession number EGAS00001000132. Normal reference RNA-seq datasets are available at the NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/Traces) under study accession number SRP000930. 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 S.A. (saparicio@bccrc.ca), C. Caldas (carlos.caldas@cancer.org.uk), S.P.S. (sshah@bccrc.ca) or M.M. (mmarra@bcgsc.ca).

LETTER

The landscape of cancer genes and mutational processes in breast cancer

Philip J. Stephens[1]*, Patrick S. Tarpey[1]*, Helen Davies[1], Peter Van Loo[1,2], Chris Greenman[1,3,4], David C. Wedge[1], Serena Nik-Zainal[1], Sancha Martin[1], Ignacio Varela[1], Graham R. Bignell[1], Lucy R. Yates[1,5,6], Elli Papaemmanuil[1], David Beare[1], Adam Butler[1], Angela Cheverton[1], John Gamble[1], Jonathan Hinton[1], Mingming Jia[1], Alagu Jayakumar[1], David Jones[1], Calli Latimer[1], King Wai Lau[1], Stuart McLaren[1], David J. McBride[1], Andrew Menzies[1], Laura Mudie[1], Keiran Raine[1], Roland Rad[1], Michael Spencer Chapman[1], Jon Teague[1], Douglas Easton[7,8], Anita Langerød[9], OSBREAC†, Ming Ta Michael Lee[10], Chen-Yang Shen[10], Benita Tan Kiat Tee[11], Bernice Wong Huimin[12], Annegien Broeks[13], Ana Cristina Vargas[14], Gulisa Turashvili[15,16], John Martens[17], Aquila Fatima[18], Penelope Miron[18], Suet-Feung Chin[19], Gilles Thomas[20], Sandrine Boyault[20], Odette Mariani[21], Sunil R. Lakhani[14,22,23], Marc van de Vijver[24], Laura van 't Veer[13], John Foekens[17], Christine Desmedt[25], Christos Sotiriou[25], Andrew Tutt[5], Carlos Caldas[19,26], Jorge S. Reis-Filho[27], Samuel A. J. R. Aparicio[15,16], Anne Vincent Salomon[21,28], Anne-Lise Børresen-Dale[9,29], Andrea L. Richardson[18,30], Peter J. Campbell[1,31,32], P. Andrew Futreal[1] & Michael R. Stratton[1]

All cancers carry somatic mutations in their genomes. A subset, known as driver mutations, confer clonal selective advantage on cancer cells and are causally implicated in oncogenesis[1], and the remainder are passenger mutations. The driver mutations and mutational processes operative in breast cancer have not yet been comprehensively explored. Here we examine the genomes of 100 tumours for somatic copy number changes and mutations in the coding exons of protein-coding genes. The number of somatic mutations varied markedly between individual tumours. We found strong correlations between mutation number, age at which cancer was diagnosed and cancer histological grade, and observed multiple mutational signatures, including one present in about ten per cent of tumours characterized by numerous mutations of cytosine at TpC dinucleotides. Driver mutations were identified in several new cancer genes including *AKT2*, *ARID1B*, *CASP8*, *CDKN1B*, *MAP3K1*, *MAP3K13*, *NCOR1*, *SMARCD1* and *TBX3*. Among the 100 tumours, we found driver mutations in at least 40 cancer genes and 73 different combinations of mutated cancer genes. The results highlight the substantial genetic diversity underlying this common disease.

The coding exons of 21,416 protein coding genes and 1,664 microRNAs were sequenced and copy number changes examined in 100 primary breast cancers, 79 of which were oestrogen receptor positive (ER+) and 21 of which were oestrogen receptor negative (ER−) (Supplementary Table 1). We sequenced normal DNAs from the same individuals to exclude inherited sequence variation. We identified 7,241 somatic point mutations: 6,964 were single-base substitutions, of which 4,737 were predicted to generate missense; 422, nonsense; 158, an essential splice site; 8, stop codon read-through; and 1,637, silent changes in protein sequence. Two substitutions were found in microRNAs. There were 277 small insertions or deletions (71 and 206, respectively), of which 231 introduced translational frameshifts and 46 were in-frame (Supplementary Table 2). Analyses of copy number yielded 1,712 homozygous deletions and 1,751 regions of increased copy number (amplification) (Supplementary Table 3).

Somatic driver substitutions and small insertions/deletions (indels) were identified in cancer genes previously implicated in breast cancer development, including *AKT1*, *BRCA1*, *CDH1*, *GATA3*, *PIK3CA*, *PTEN*, *RB1* and *TP53* (Supplementary Table 4; see also http://www.sanger.ac.uk/genetics/CGP/Census). Likely drivers were also found in cancer genes involved in other cancer types, including *APC*, *ARID1A*, *ARID2*, *ASXL1*, *BAP1*, *KRAS*, *MAP2K4*, *MLL2*, *MLL3*, *NF1*, *SETD2*, *SF3B1*, *SMAD4* and *STK11*.

To identify new cancer genes, we searched for non-random clustering of somatic mutations in each of the 21,416 protein-coding genes[2,3] and sequenced a subset of genes highlighted by this analysis in a follow-up series of 250 breast cancers (Supplementary Tables 5 and 6). Persuasive evidence was found for nine new cancer genes (Fig. 1a and Supplementary Fig. 1). Of these *ARID1B*, *CASP8*, *MAP3K1*, *MAP3K13*, *NCOR1*, *SMARCD1* and *CDKN1B* had the truncating

[1]Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. [2]Human Genome Laboratory, Department of Human Genetics, VIB and University of Leuven, Herestraat 49 Box 602, B-3000 Leuven, Belgium. [3]School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK. [4]The Genome Analysis Centre, Norwich Research Park, Norwich NR4 7UH, UK. [5]Breakthrough Breast Cancer Research Unit, Research Oncology, 3rd Floor Bermondsey Wing, Guy's Hospital Campus, Kings College London School of Medicine, London SE1 9RT, UK. [6]Department of Clinical Oncology, Ground floor, Lambeth Wing, Guys and St Thomas' NHS Trust, Westminster Bridge Road, London SE1 7EH, UK. [7]Centre for Cancer Genetic Epidemiology, Department of Oncology, Strangeways Research Laboratory, Cambridge CB1 8RN, UK. [8]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Strangeways Research Laboratory, Cambridge CB1 8RN, UK. [9]Department of Genetics, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, O310 Oslo, Norway. [10]National Genotyping Center, Institute of Biomedical Sciences, Academia Sinica, 128 Academia Road, Sec 2, Nankang, Taipei 115, Taiwan, China. [11]Department of General Surgery, Singapore General Hospital, 169608, Singapore. [12]NCCS-VARI Translational Research Laboratory, National Cancer Centre Singapore, 11 Hospital Drive, 169610, Singapore. [13]Department Experimental Therapy, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. [14]The University of Queensland Centre for Clinical Research, The Royal Brisbane & Women's Hospital, Herston, Brisbane, Queensland 4029, Australia. [15]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. [16]Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia V5Z 1L3, Canada. [17]Department of Medical Oncology, Erasmus University Medical Center, Daniel den Hoed Cancer Center and Cancer Genomics Center, Postbus 2040, 3000 CA Rotterdam, Netherlands. [18]Department of Cancer Biology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02215, USA. [19]Department of Oncology, University of Cambridge and Cancer Research UK Cambridge Research Institute, Li Ka Shin Centre, Cambridge CB2 0RE, UK. [20]Universite Lyon 1, INCa-Synergie, Centre Leon Berard, 28 rue Laennec, Lyon CEDEX 08, France. [21]Institut Curie, Department of Tumor Biology, 26 rue d'Ulm, 75248 Paris CEDEX 05, France. [22]The University of Queensland School of Medicine, Herston Road, Herston, Brisbane, Queensland 4006, Australia. [23]Anatomical Pathology, Pathology Queensland, The Royal Brisbane and Women's Hospital, Herston, Brisbane, Queensland 4029, Australia. [24]Department of Pathology, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. [25]Breast Cancer Translational Laboratory, Université Libre de Bruxelles, Jules Bordet Institute, Boulevard de Waterloo 121, 1000 Brussels, Belgium. [26]NIHR Cambridge Biomedical Research Centre and Cambridge Experimental Cancer Medicine Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 2QQ, UK. [27]The Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London SW3 6JB, UK. [28]Institut Curie, INSERM Unit 830, 26 rue d'Ulm, 75248 Paris CEDEX 05, France. [29]K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, O318 Oslo, Norway. [30]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St, Boston, Massachusetts 02115, USA. [31]Department of Haematology, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. [32]Department of Haematology, University of Cambridge, Hills Road, Cambridge CB2 2XY, UK.
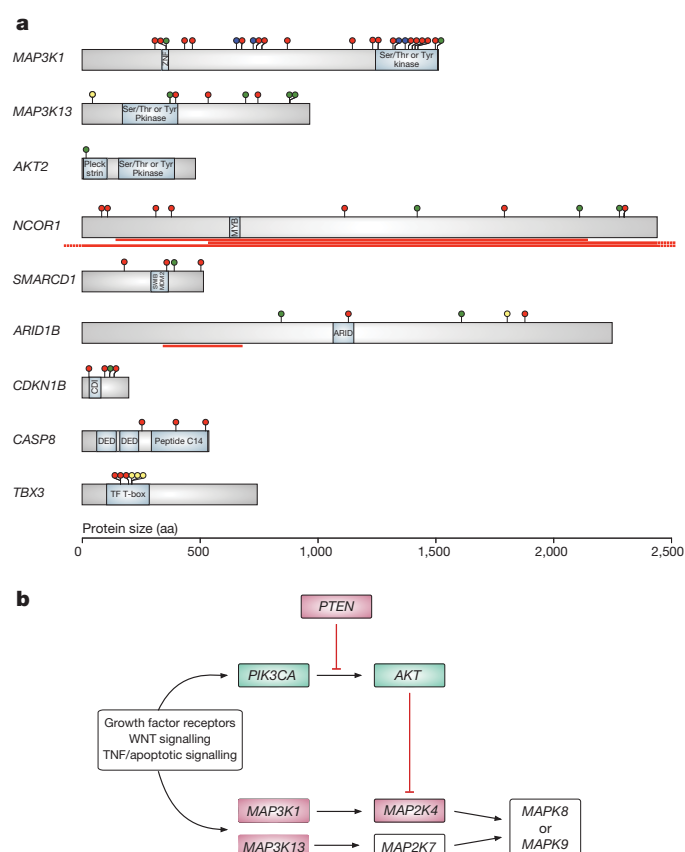
*These authors contributed equally to this work.
†Lists of participants and their affiliations appear at the end of the paper.

mutations and often biallelic inactivation characteristic of inactivated, potentially recessive cancer genes (Supplementary Table 4). *AKT2* is probably an activated, dominantly acting cancer gene. The effects of *TBX3* mutations on its function are unclear.

MAP3K1 encodes a serine/threonine protein kinase that regulates the activity of the ERK MAP kinase (the extracellular signal-regulated mitogen-activated protein kinase), JUN kinase and p38 signalling pathways implicated in control of cell proliferation and death[4]. Somatic mutations in *MAP3K1* were observed in 6% of breast cancers, predominantly in ER+ cases. Most were protein truncating. MAP3K1 phosphorylates and activates the protein encoded by *MAP2K4*, a known recessive cancer gene with inactivating mutations in breast and other cancers[5]. In turn, MAP2K4 phosphorylates and activates the JUN kinases MAPK8 (also known as JNK1) and MAPK9 (also known as JNK2), which phosphorylate JUN, TP53 and other transcription factors mediating cellular responses to stress[4]. Truncating mutations and other non-synonymous mutations were also found in *MAP3K13*, which encodes a kinase that phosphorylates and activates MAP2K7. MAP2K7 phosphorylates and activates MAPK8 and MAPK9 (ref. 4). Thus, in breast cancer, inactivating mutations in *MAP3K1*, *MAP2K4* and *MAP3K13* are predicted to abrogate signalling pathways that activate JUN kinases (Fig. 1b).

In the serine/threonine kinase gene *AKT2*, we identified a single somatic missense mutation, Glu 17 Lys, that is identical to the recurrent, activating mutation in *AKT1* previously reported in breast cancer[6].

a, **b**

Thus, *AKT2* is also probably a cancer gene, albeit one infrequently implicated in breast cancer development. Because AKT phosphorylates and inhibits MAP2K4 (ref. 7) and mutations in *PIK3CA* and *PTEN* can result in AKT activation[8], about half of breast cancers may have abrogation of JUN kinase signalling (Fig. 1b). The biological consequences of the reduction in JUN kinase activity are likely to be diverse and complex, but may include destabilization and consequent inactivation of TP53 with disruption of pro-apoptotic cellular signalling in response to stress[9].

We observed truncating mutations and homozygous deletions of *NCOR1*. In addition to mediating repression of thyroid-hormone and retinoic-acid receptors by promoting chromatin condensation and preventing access of the transcription machinery[10], NCOR1 participates in ligand-dependent transcriptional repression by oestrogen receptor alpha[11]. We also identified inactivating mutations in *SMARCD1* and *ARID1B*, further implicating aberrant chromatin regulation. The encoded proteins of both are components of the SWI/SNF chromatin modelling complex, which incorporates the products of several established recessive cancer genes, including *PBRM1*, *ARID1A*, *SMARCB1* and *SMARCA4* (refs 3, 12–14).

We found three truncating mutations and a missense mutation in *CDKN1B*. Two truncating mutations in *CDKN1B* in cancer have previously been reported[15,16], and collectively the results confirm that *CDKN1B* is a cancer gene. CDKN1B (also known as p27 or KIP1) normally inhibits activation of cyclin E/CDK2 and cyclin D/CDK4 complexes, thus preventing cell cycle progression at phase G1[17].

Three truncating mutations were observed in *CASP8*. CASP8 is a member of the cysteine/aspartic acid protease family that forms a complex with the FAS cell surface receptor to promote programmed cell death. Inactivation of CASP8 in these cancers is therefore predicted to abrogate apoptosis in response to a variety of signals.

Six tumours had mutations in *TBX3*, which encodes a T-box transcription factor that regulates stem cell pluripotency-associated and reprogramming factors and is involved in normal breast development[18,19]. Constitutional inactivating mutations in *TBX3* cause ulnar-mammary syndrome, in which there is failure of breast and apocrine development coupled with abnormalities of limb morphogenesis[20]. Three breast cancers had in-frame deletions, one of Thr 210 and the other two of Asn 212, a residue through which the T-box domain binds to DNA. Despite the presence of truncating mutations in three further cases, the recurrent and clustered in-frame deletions and the finding that all mutations were heterozygous suggests that they may not simply result in loss of function. Indeed, recent reports suggest that increased activity of TBX3 is likely to contribute to oncogenesis. The proportion of stem-like cells in breast cancers is increased by oestrogen-dependent activation of the TBX3 pathway[21]. Moreover, TBX3 overexpression increases the efficiency of the derivation of induced pluripotent stem cells[18] and the ability of cancer cells to form tumours[21].

Further supporting their role in oncogenesis, three of the nine newly identified somatically mutated cancer genes, *MAP3K1*, *CASP8* and *TBX3*, carry inherited common variants, identified by genome-wide association studies, that confer small increased risks of breast cancer[22,23]. Several additional genes showed truncating mutations and are biologically plausible candidate cancer genes contributing infrequently to breast cancer development. Some, including *ASXL2*, *ARID5B*, *KDM3A*, *SETD1A*, *CHD1*, *NCOR2*, *HDAC9* and *CTCF*, encode proteins that regulate chromatin structure, whereas others, including *FANCA* and *ATR*, are involved in DNA repair.

Cancers arise through successive waves of clonal expansion dependent on the sequential acquisition of driver mutations. A central parameter of cancer development is therefore the number of driver mutations required for conversion of a normal cell into a symptomatic cancer. Estimates based on cancer age–incidence curves have indicated that approximately five rate-limiting steps underlie the development of common adult solid tumours[24]. Experimental studies have similarly indicated that a limited number of key genetic changes are required for

Figure 1 | New cancer genes established in this study and involvement of the JUN kinase signalling pathway. a, Representations of the protein-coding sequences and major domains in cancer genes established in this study. Somatic mutations are shown as circles: truncating (red), essential splice site (blue), missense (green) and in-frame indel (yellow). The red lines indicate the positions of large homozygous deletions. aa, amino acids. **b**, Pathways regulating the JUN kinases MAP2K7 and MAP2K8, indicating genes with mutations in this series. Genes in green are activated by mutations, whereas genes in red are inactivated.

Figure 2 | The landscape of driver mutations in breast cancer. Each of the 40 cancer genes in which a driver mutation or copy number change has been identified is listed down the left-hand side. The number of mutations in each gene in the 100 tumours is shown (rows), as is the number of driver mutations in each breast cancer (columns). Point mutations and copy number changes are coloured red and blue, respectively.

neoplastic transformation of human cells[25]. Our systematic genome analysis now provides a direct survey of the landscape of driver mutations in breast cancer.

Somatic driver point mutations and/or copy number changes in at least 40 cancer genes were implicated in the development of the 100 breast cancers (Fig. 2, Supplementary Tables 3 and 4, and Supplementary Methods). The maximum number of mutated cancer genes in an individual cancer was 6, but 28 cases only showed a single driver. Thus, there seems to be substantial variation in the number of drivers. In some cases, the presence of multiple drivers was associated with subclonal evolution of the cancer (Supplementary Statistical Analyses). However, in others multiple drivers were in the root cancer clone. Seven of the 40 cancer genes (*TP53*, *PIK3CA*, *ERBB2*, *MYC*, *FGFR1/ZNF703*, *GATA3* and *CCND1*) were mutated in more than 10% of cases. Collectively these contributed 58% of driver mutations (144 of 250). Therefore, 33 mutated cancer genes, each contributing relatively infrequently, were responsible for the remaining 42% of driving genetic events. We observed 73 different combinations of mutated cancer genes. Thus, most breast cancers differed from all others (Fig. 2 and Supplementary Fig. 2). This assessment of the genetic diversity of breast cancer is probably conservative because, for several reasons, it underestimates the number of mutated cancer genes in each case.

At present, we know little about the mutational processes responsible for the generation of somatic mutations in breast and other cancers. In the 100 breast cancers analysed here, there was substantial variation in the total numbers of base substitutions and indels between individual cases (Fig. 3a). There was also considerable diversity of mutational pattern, ranging from cases in which C•G → T•A transitions predominated to cases in which all transitions and transversions made equal contributions (Fig. 3b and Supplementary Fig. 3). Taken together, the results suggest that multiple distinct mutational processes are operative. For most of these processes, the underlying mechanism is unknown.

To illustrate one mutational signature in detail, we selected the ER+ breast cancer with the largest number of base substitutions in the series, PD4120 (Fig. 3a, asterisk; Fig. 4). The mutation spectrum of this case was distinctive, featuring C•G → T•A, C•G → G•C and C•G → A•T mutations and very few mutations at A•T base pairs (Fig. 4a). To characterize this process further, we examined the sequence context in which the mutations occurred (in the following discussion, mutations at C•G base pairs are represented as the change at the C base) and found pronounced overrepresentation of thymine

immediately 5′ to the mutated cytosines. Thus, in PD4120 the large majority of mutations were of cytosine at TpC dinucleotides (Fig. 4b).

To obtain further insight into the underlying mechanism in this case, we looked for differences in mutation prevalence between the transcribed and untranscribed strands of the 21,416 genes analysed ('strand bias') and found a higher prevalence of C → T, C → G and C → A mutations on transcribed strands ($P = 0.02$) (Fig. 4c and Supplementary Table 7). This strand bias raises the possibility that transcription-coupled nucleotide excision repair (NER) has been operative. NER removes bulky DNA adducts that distort the DNA
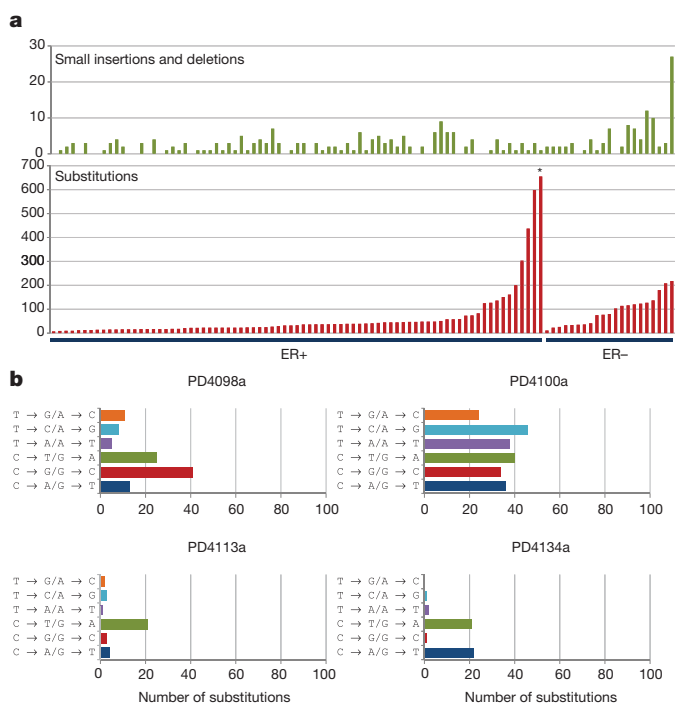
Figure 3 | The variation in numbers and types of mutation between individual breast cancers. a, Numbers of small indels and base substitutions in the protein-coding exons of each of the 100 breast cancers studied. The cases are ranked according to the number of base substitutions. *Breast cancer PD4120 (see main text). **b**, Mutation spectrum of four primary tumours with diverse mutational patterns.
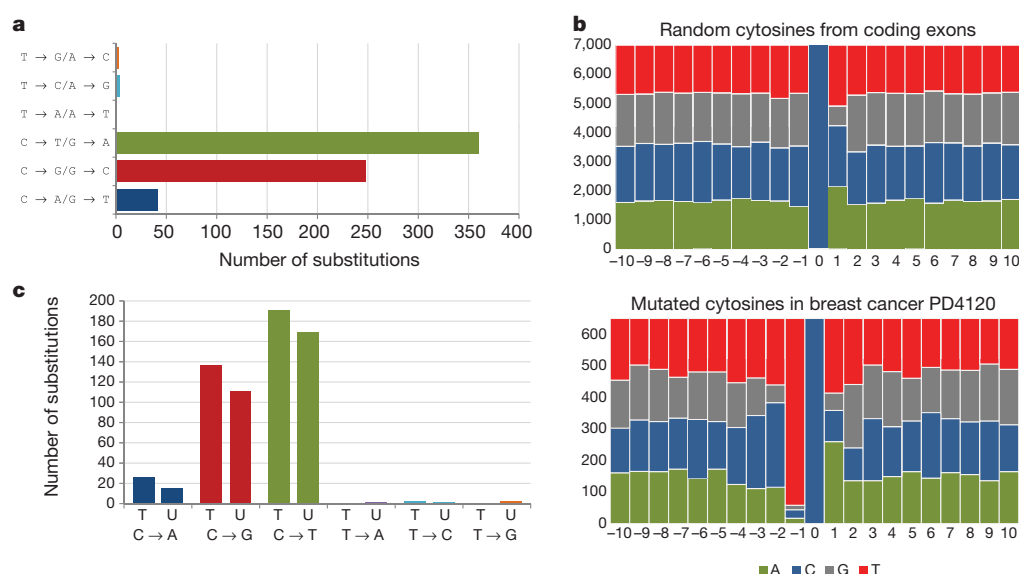
Figure 4 | The mutational signature of ER+ breast cancer PD4120. a, The mutational spectrum. b, The sequence context of $C \rightarrow T$, $C \rightarrow G$ and $C \rightarrow A$ mutations. The central blue bar indicates the position of the mutated cytosine and the bases 5′ and 3′ are numbered on the horizontal axis. c, Strand bias of mutations showing substitutions at C bases and at T bases according to whether they are on the transcribed (T) or untranscribed (U) strands of the genes screened.

double helix, notably pyrimidine dimers due to ultraviolet light exposure or adducts due to mutagens in tobacco smoke[26]. There is a form of NER, recruited by RNA polymerase II, that is operative only on the transcribed strand of each gene and thus introduces a strand bias for mutations[27]. Therefore, one hypothesis to account for the strand bias in PD4120 is past involvement of NER, in turn implicating exposure to a bulky DNA-damaging agent, either of endogenous or exogenous origin. However, we cannot exclude the possibility that other DNA damage or repair processes generate a strand bias. At least eight additional cancers in this series had a very similar mutational spectrum, sequence context and strand bias (Supplementary Fig. 4 and Supplementary Statistical Analysis). None had been treated before excision of the cancer.

The somatic mutations in a cancer genome accumulate over a patient's lifetime, during the lineage of mitotic divisions from the fertilized egg to the cancer cell. Some are acquired while cells in the lineage are biologically normal, whereas others are acquired after acquisition of the neoplastic phenotype. However, the relative proportions accumulated in these two phases are unknown. To explore this question, we examined the relationship between the total numbers of somatic base substitutions and the age at diagnosis in the 100 tumours (Fig. 5). In both ER+ and ER− cancers, no correlation was observed ($P = 0.33$ and 0.14 respectively). If most somatic mutations in a cancer genome are acquired in normal tissues before neoplastic transformation, the later the onset of the cancer the longer this part

of the lineage is likely to have been and, consequently, the higher the number of mutations. The absence of a correlation therefore suggests that most mutations in breast cancer genomes occur after the initiating driver event.

We then considered separately the subset of somatic mutations constituted by C•G → T•A substitutions at CpG dinucleotides, because this mutational pattern is observed in non-diseased tissues, manifesting prominently in normal germline variation. This subset showed a strong positive correlation with the age at cancer diagnosis in ER− cancers ($P = 1.2 \times 10^{-7}$), supporting the proposition that it is enriched in mutations occurring in normal tissues and that, overall, other mutation classes occur later. By contrast, ER+ cancers showed no correlation between C•G → T•A substitutions at CpG dinucleotides and age at diagnosis ($P = 0.27$). The basis for this pronounced difference is unclear, but potentially highlights a profound divergence in the dynamics of mutation acquisition between these two major subclasses of breast cancer.

In clinical practice, breast cancers are graded microscopically on the basis of mitotic counts, pleomorphism of cancer cell nuclei and extent of tubule formation, which are then collected into an overall grade score. High scores indicate large numbers of mitoses, substantial tumour cell pleomorphism and little tubule formation, and are generally associated with more rapid progression. Significant correlations were not observed between numbers of driver mutations and grade scores (Supplementary Statistical Analysis). However, there were strong positive correlations between the total number of substitutions (that is, drivers and passengers) and mitosis and tubule scores ($P = 0.0002$ and 0.002 respectively), which remained significant after multiple testing corrections. The causal relationships between these features are unclear. However, because most substitutions are likely to be biologically inert passengers, it is possible that the biological state of high-grade breast cancers may be responsible for generating increased numbers of mutations, rather than the converse.

The panorama of mutated cancer genes and mutational processes in breast cancer is becoming clearer, and a sobering perspective on the complexity and diversity of the disease is emerging. Driver mutations are operative in many cancer genes. A few are commonly mutated, but many infrequently mutated genes collectively make a substantial contribution in myriad different combinations. Multiple somatic mutational processes have been operative. Ultimately, characterization of the genomes of breast cancer, and others, will provide a robust and biologically meaningful classification generating insights into the clinical heterogeneity of the disease and influencing strategies to find new modes of prevention and treatment.
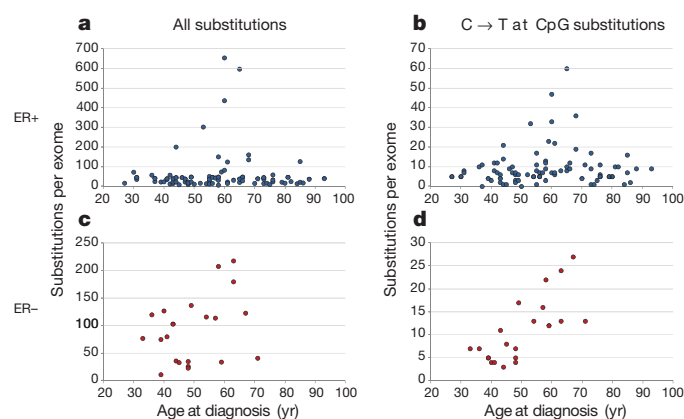
Figure 5 | The relationship between age at breast cancer diagnosis and all substitutions, and for C → T substitutions at CpG sites. a, b, Data from the 79 ER+ breast cancers. c, d, Data from the 21 ER− breast cancers.

METHODS SUMMARY

DNAs from breast cancers and normal tissues from the same individuals were subjected to in-solution enrichment for the coding exons of 21,416 genes and 1,664 micro RNAs (Agilent 50 Mb Exome) and subsequently sequenced on Illumina GAIIX machines. The average exome coverage (at a minimum depth of ×30) was 70%. Following alignment to the reference genome using BWA[28], somatic substitutions and small indels were identified using CaVEMan and Pindel calling algorithms, respectively[3,29]. DNAs from cancers were also hybridized to Affymetrix SNP6 arrays and analysed using the ASCAT algorithm[30] for copy number and zygosity changes. Confirmation by orthogonal sequencing technologies was attempted for all putative somatic mutations and follow-up of a subset of genes in additional case series was undertaken by targeted PCR and Illumina sequencing. Identification of genes showing evidence of selection was conducted as previously described[3].

Informed consent was obtained from all subjects and ethical approval obtained from Cambridgeshire 3 Research Ethics Committee (ref 09/H0306/36).

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

1. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458,** 719–724 (2009).
2. Greenman, C., Wooster, R., Futreal, P. A., Stratton, M. R. & Easton, D. F. Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics* **173,** 2187–2198 (2006).
3. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **469,** 539–542 (2011).
4. Keshet, Y. & Seger, R. The MAP kinase signaling cascades: a system of hundreds of components regulates a diverse array of physiological functions. *Methods Mol. Biol.* **661,** 3–38 (2010).
5. Su, G. H. *et al.* Alterations in pancreatic, biliary, and breast carcinomas support MKK4 as a genetically targeted tumor suppressor gene. *Cancer Res.* **58,** 2339–2342 (1998).
6. Carpten, J. D. *et al.* A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* **448,** 439–444 (2007).
7. Park, H. S. *et al.* Akt (protein kinase B) negatively regulates SEK1 by means of protein phosphorylation. *J. Biol. Chem.* **277,** 2573–2578 (2002).
8. Carnero, A., Blanco-Aparicio, C., Renner, O., Link, W. & Leal, J. F. The PTEN/PI3K/AKT signalling pathway in cancer, therapeutic implications. *Curr. Cancer Drug Targets* **8,** 187–198 (2008).
9. Wu, G. S. The functional interactions between the MAPK and p53 signaling pathways. *Cancer Biol. Ther.* **3,** 146–151 (2004).
10. Horlein, A. J. *et al.* Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* **377,** 397–404 (1995).
11. Merrell, K. W. *et al.* Differential recruitment of nuclear receptor coregulators in ligand-dependent transcriptional repression by estrogen receptor-alpha. *Oncogene* **30,** 1608–1614 (2010).
12. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330,** 228–231 (2010).
13. Reisman, D., Glaros, S. & Thompson, E. A. The SWI/SNF complex and cancer. *Oncogene* **28,** 1653–1668 (2009).
14. Wiegand, K. C. *et al.* ARID1A mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* **363,** 1532–1543 (2010).
15. Spirin, K. S. *et al.* p27/Kip1 mutation found in breast cancer. *Cancer Res.* **56,** 2400–2404 (1996).
16. Tigli, H., Buyru, N. & Dalay, N. Molecular analysis of the p27/kip1 gene in breast cancer. *Mol. Diagn.* **9,** 17–21 (2005).
17. Chu, I. M., Hengst, L. & Slingerland, J. M. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nature Rev. Cancer* **8,** 253–267 (2008).
18. Han, J. *et al.* Tbx3 improves the germ-line competency of induced pluripotent stem cells. *Nature* **463,** 1096–1100 (2010).
19. Howard, B. & Ashworth, A. Signalling pathways implicated in early mammary gland morphogenesis and breast cancer. *PLoS Genet.* **2,** e112 (2006).
20. Bamshad, M. *et al.* Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nature Genet.* **16,** 311–315 (1997).
21. Fillmore, C. M. *et al.* Estrogen expands breast cancer stem-like cells through paracrine FGF/Tbx3 signaling. *Proc. Natl Acad. Sci. USA* **107,** 21737–21742 (2010).
22. Ghoussaini, M. *et al.* Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nature Genet.* **44,** 312–318 (2012).
23. Varghese, J. S. & Easton, D. F. Genome-wide association studies in common cancers–what have we learnt? *Curr. Opin. Genet. Dev.* **20,** 201–209 (2010).
24. Miller, D. On the nature of susceptibility to cancer. *Cancer* **46,** 1307–1318 (1980).
25. Schinzel, A. C. & Hahn, W. C. Oncogenic transformation and experimental models of human cancer. *Front. Biosci.* **13,** 71–84 (2008).
26. Shuck, S. C., Short, E. A. & Turchi, J. J. Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res.* **18,** 64–72 (2008).
27. Fousteri, M. & Mullenders, L. H. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res.* **18,** 73–84 (2008).
28. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26,** 589–595 (2010).
29. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25,** 2865–2871 (2009).
30. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl Acad. Sci. USA* **107,** 16910–16915 (2010).

The Oslo Breast Cancer Consortium (OSBREAC) Rolf Karesen[1,2], Ellen Schlichting[1], Bjorn Naume[2,3], Torill Sauer[2,4] & Lars Ottestad[3]

[1]Department of Breast and Endocrine Surgery, Oslo University Hospital, 0424 Oslo, Norway. [2]Medical Faculty, University of Oslo, 0424 Oslo, Norway. [3]Department of Oncology, Oslo University Hospital, 0424 Oslo, Norway. [4]Department of Pathology, Oslo University Hospital, 0424 Oslo, Norway.

METHODS

Patient samples. Informed consent was obtained from all subjects and ethical approval obtained from Cambridgeshire 3 Research Ethics Committee (ref 09/H0306/36). Collection and use of patient samples were approved by the appropriate IRB of each Institution. In addition, this study and usage of its collective materials had specific IRB approval.

Exome enrichment and sequencing. Genomic libraries were prepared using the Illumina Paired End Sample Prep Kit following the manufacturer's instructions. Enrichment was performed as described previously[31], using the Agilent SureSelect Human All Exon 50Mb kit following the manufacturer's recommended protocol but excluding pre-enrichment PCR amplification. Each exome was sequenced using the 75 or 76-bp paired-end protocol, on an Illumina GAII or HiSeq DNA Analyser, to produce approximately 10 Gb of sequence per exome. Sequencing reads were aligned to the human genome (NCBI build 37) using the BWA algorithm on default settings[32]. Reads which were unmapped, PCR-derived duplicates or outside the targeted region of the genome were excluded from the analysis. The remaining uniquely mapping reads (~60%) provided 60–80% coverage over the targeted exons at a minimum depth of ×30.

Sequencing of pooled PCR amplimers. Selected genes were targeted for follow-up investigations in 250 additional breast cancers by sequencing of pooled PCR products. An 8-bp index was introduced during amplification to enable sequence data from individual tumours to be identified in downstream analyses.

For each amplimer, a primary PCR was performed using gene-specific primers modified with the inclusion of a common upstream adaptor sequence. A secondary PCR was performed using primers complementary to the common adaptor sequences. The reverse secondary primer contained the internal index, and 96 different indexed primers were used to enable 96 different DNAs to be pooled before sequencing. The primary and secondary PCR amplifications were performed as a simultaneous multiplex reaction. Primer sequences are available on request.

For each amplimer, PCR was performed in batches of 96 DNA samples. Following amplification, the 96 PCR products were pooled, purified using a QiaQuick column (Qiagen) and quantified on a Bioanalyser (Agilent). Pooled reactions from different amplimers (up to 50) were normalized for concentration and subsequently also pooled to produce the final template used for sequencing on a single lane of an Illumina GAII DNA Analyser (~5,000 amplimers per lane). Amplimers which failed PCR were excluded from the pooling experiments. The subsequent sequence reads were aligned with BWA and resulted in coverage typically exceeding ×500 per individual sample amplimer.

Variant detection. The CaVEMan (cancer variants through expectation maximization) algorithm was used to call single nucleotide substitutions[31]. This uses a naive Bayesian classifier to estimate the posterior probability of each possible genotype (wild type, germline, somatic mutation) at each base. We applied several post-processing filters to the set of initial CaVEMan mutation calls to remove variants reported in poor-quality sequence and increase the specificity of the output.

To call insertions and deletions, we used split-read mapping implemented as a modification of the Pindel algorithm[33]. This algorithm searches for reads where one end is anchored on the genome and the other end can be mapped with high confidence in two (split) portions, spanning a putative indel. Post-processing filters were applied to the output to improve specificity.

Mutations were annotated to Ensembl version 58.

Variant validation. Validation of all 7,241 putative somatic variants in the primary screen of 100 tumours and all variants found in the follow-up of 250 cases was attempted by either capillary resequencing or 454 pyrosequencing of PCR products spanning the mutation in the tumour and the normal pair. Where independent validation failed (approximately 20%) variants were reported to be somatic if manual inspection of the aligned sequence reads provided strong evidence to support their validity.

Identification of likely driver base substitutions and indels. A subset of the 7,241 substitution and indel somatic mutations identified in the exome screen were classified as 'likely driver mutations' using conservative criteria. To do this, we identified the established cancer genes from the Cancer Gene Census (http://www.sanger.ac.uk/genetics/CGP/Census/) that are known to be mutated by base substitutions and indels to contribute to cancer development. We then classified as likely driver mutations those that conformed to the known patterns of cancer-causing mutation for each cancer gene. Thus, for recessive cancer genes truncating mutations, essential splice site mutations and homozygous deletions were included. Missense mutations were also included where they had been seen previously or conformed to the known pattern of missense mutation in each gene (COSMIC database; http://www.sanger.ac.uk/genetics/CGP/cosmic/). For established, dominantly acting cancer genes, we included mutations that had been previously registered in COSMIC. For the new cancer genes established in this study, we applied essentially the same rules. However, for the recessive cancer genes, to be conservative we did not include missense variants (other than the single

variant in *MAP3K1*, which is almost certainly disruptive to the function of the protein). We included the variant in *AKT2* because it is identical in nature to the recurrent variant in *AKT1*, and we included all *TBX3* mutations. As indicated in the main text, we may have both underestimated and overcalled some somatic variants as drivers using this approach. However, the number of erroneous calls is likely to be small and overall we have probably underestimated the number of driver mutations. For the calling procedure for likely driver copy number variants, see below.

Detection of copy number variation. Single nucleotide polymorphism (SNP) array hybridization on the SNP6.0 platform was done according to Affymetrix Protocols and as described at http://www.sanger.ac.uk/cgi-bin/genetics/CGP/cghviewer/CghHome.cgi.

Copy number analysis was performed using ASCAT (version 2.1) taking into account non-neoplastic cell infiltration and tumour aneuploidy[34], and resulted in integral allele-specific copy number profiles for the tumour cells. Amplifications in the 100 samples analysed were called if copy number was ≥5 (for diploid tumours, with ASCAT ploidy <2.7 n) or ≥9 (for tumours with evidence of a whole-genome duplication, with ASCAT ploidy ≥2.7 n). Homozygous deletions were called if there were zero copies in the tumour cells.

Identification of likely driver copy number variants. To identify likely driver copy number variants, we derived a conservatively generated list of frequently amplified regions in breast cancer from a previous study[35]. From the amplified regions in breast cancer obtained by GISTIC analysis of that study, those with a GISTIC Q-value of less than 10^{-5} were selected. Regions within 40 Mb of amplified regions with more significant Q-values were excluded, as many of these probably point to the same amplified target gene. This process generated seven focal, highly significantly amplified regions. These regions were annotated with their putative target genes where additional biological studies have indicated that they are the likely targets (*ERBB2*, *CCND1*, *MYC*, *FGFR1/ZNF703*, *ZNF217*, *MDM2*). Only the amplified region on chromosome 15 was not annotatable. Driver amplification of these seven focal regions in the 100 samples was called using the criteria above. Driver homozygous deletions were called if part or all of a homozygous deletion overlapped with a known recessive cancer gene from the Cancer Gene Census[36] or a newly discovered gene from this study.

Estimation of the number of mutated copies. Allele-specific copy number estimates for point mutations and indels were obtained by integrating copy number and sequencing data. In a sample containing only tumour cells, the number of reads, r, with a mutation can be expressed as

$$r = \frac{n_{mut}R}{n_{locus}} \tag{1}$$

In equation (1), n_{locus} is the copy number of the locus, n_{mut} is the number of mutated copies and R is the total number of reads from that locus. In case of a tumour sample consisting of a fraction of tumour cells ρ, infiltrated with a fraction of normal cells $1-\rho$ (assumed to have two copies), equation (1) becomes

$$r = \frac{n_{mut}R\rho}{\rho n_{locus} + 2(1-\rho)}$$

Hence, allele-specific copy number estimates for point mutations and indels can be obtained as

$$n_{mut} = f_s \frac{1}{\rho}(\rho n_{locus} + 2(1-\rho)) \tag{2}$$

In equation (2), $f_s = r/R$ is the frequency of mutated reads observed in the sequencing data, and ρ and n_{locus} can be obtained from the ASCAT copy number analysis.

These copy number estimates of mutations were used to determine which mutations are likely subclonal: if $n_{mut} \geq 0.8$, the mutation is called likely clonal and if $n_{mut} < 0.8$, the mutation is called likely subclonal.

In the case of indels, reads with an insertion or deletion may not map as well as reads without insertions and deletions. Therefore, a procedure was followed to estimate f_s for indels that was independent of ease of mapping. Reads were obtained by matching flanking sequence (10 bp on each side) around the indel, further filtered to exclude spurious matches. The mutated read frequency was subsequently calculated, accounting for the difference in sequence lengths with and without the indel:

$$f_s = \frac{r_{indel}/(l_s - l_{indel} + 1)}{r_{indel}/(l_s - l_{indel} + 1) + r_{normal}/(l_s - l_{normal} + 1)} \tag{3}$$

In equation (3), r_{indel} and r_{normal} are the respective numbers of reads with and without the indel, l_s is the read length (76 bp), and l_{indel} and l_{normal} are the respective lengths of the matching fragment in sequences with and without the indel.

Detection of selection and oncogenicity in protein-coding genes. The overall significance of an excess of non-silent mutations was determined using the methods previously described[37]. The ranking of gene significances was determined

using the following model. We let s_{kg}^i denote the number of silent mutations, where k indexes mutation type (C•G → A•T, C•G → T•A, C•G → G•C, T•A → A•T, T•A → C•G or T•A → G•C) in gene g, where $i = 1$ for the primary screen and $i = 2$ for the follow-up screen. We also have counts m_{kg}^i and n_{kg}^i of missense and nonsense mutations, respectively. Finally we have counts l_g^i of indels. The numbers of screened bases, S_{kg}^i, M_{kg}^i and N_{kg}^i, in each gene for each mutation type were also calculated. The total number of screened bases was L_g^i. We let ρ_k represent the per-base passenger mutation prevalence and use γ to denote the per-base passenger rate of indels.

Next we assume that genes can be neutral to cancer, oncogenically triggered by missense mutations or inactivated by truncating mutations. Genes are not precluded from belonging to both of the last two categories. We assume that proportions α and β of genes belong to the missense group and truncating group, respectively. Genes that belong to these groups have mutation rates that increase by factors λ and μ, respectively. These terms quantify the selection pressure for missense and truncating variants, respectively. This results in a mixture model with the following likelihood:

$$L(\rho_k, \gamma, \alpha, \beta, \lambda, \mu | s_{kg}^i, m_{kg}^i, n_{kg}^i, l_g^i)$$

$$= \prod_{g \in G} \prod_k^K Po_{s_{kg}^1}(S_{kg}^1 \rho_k) \left(\sum_{j=1}^2 \alpha_m \prod_k^K Po_{m_{kg}^1}(M_{kg}^1 \rho_k \lambda_m) \right)$$

$$\times \left(\sum_{j=1}^2 \beta_m Po_{l_g^1}(L_g^1 \gamma \mu_m) \prod_k^K Po_{n_{kg}^1}(N_{kg}^1 \rho_k \mu_m) \right)$$

$$\times \prod_{g \in G_F} \prod_k^K Po_{s_{kg}^2}(S_{kg}^2 \rho_k) \left(\sum_{j=1}^2 \alpha_m \prod_k^K Po_{m_{kg}^2}(M_{kg}^2 \rho_k \lambda_m) \right)$$

$$\times \left(\sum_{j=1}^2 \beta_m Po_{l_g^2}(L_g^2 \gamma \mu_m) \prod_k^K Po_{n_{kg}^2}(N_{kg}^2 \rho_k \mu_m) \right)$$

Here $\alpha_1 = 1 - \alpha$, $\alpha_2 = \alpha$, $\beta_1 = 1 - \beta$, $\beta_2 = \beta$, $\lambda_1 = 1$, $\lambda_2 = \lambda$, $\mu_1 = 1$ and $\mu_2 = \mu$, and $Po_c(r)$ indicates the Poisson probability of obtaining value c from a Poisson process with rate parameter r. G_F denotes the set of genes in the follow-up study. The parameters for this model were then estimated with the expectation–maximization algorithm. Confidence intervals for these parameters were obtained using parametric bootstrapping. Conditional on these parameter estimates, we can then use Bayes' law to calculate the probability that each gene belongs to the neutral, the missense or the truncating group. Specifically, if $\phi_g, \psi_g \in \{1, 2\}$ index whether the gene g does or does not belong to the missense or the truncating group, respectively, we have

$$\Pr(\varphi_g = m, \psi_g = n)$$

$$\propto Po_{l_g^1}(L_g^1 \gamma \mu_n) \prod_k^K Po_{s_{kg}^1}(S_{kg}^1 \rho_k) Po_{m_{kg}^1}(M_{kg}^1 \rho_k \lambda_m)$$

$$\times Po_{n_{kg}^1}(N_{kg}^1 \rho_k \mu_n)$$

$$\times Po_{l_g^2}(L_g^2 \gamma \mu_n) \prod_k^K Po_{s_{kg}^2}(S_{kg}^2 \rho_k) Po_{m_{kg}^2}(M_{kg}^2 \rho_k \lambda_m)$$

$$\times Po_{n_{kg}^2}(N_{kg}^2 \rho_k \mu_n)$$

The probability of belonging to either the missense or the truncating group, $1 - \Pr(\phi_g = 1, \psi_g = 1)$, was then used to rank the genes.

Generalized linear models. Generalized linear models (GLMs) are extensions to ordinary linear regression that model underlying distributions using members of the exponential family[38]. The response variable is related to the linear model by a link function using maximum-likelihood estimates of the parameters. Because they are not restricted to modelling normally distributed data, GLMs have particular utility in modelling count data such as, in this manuscript, the number of mutations.

If mutations were generated by a random process, with a constant probability of occurring at any point throughout an individual's life, we would expect the number of mutations to have a Poisson distribution, dependent only on the (unknown) rate of mutation and the age of the individual. Where goodness-of-fit tests indicated that the Poisson distribution was an appropriate model for the number of mutations, we used this distribution. However, in the models where goodness-of-fit tests indicated that mutation numbers were overdispersed, we used negative binomial distributions in place of Poisson distributions, as the negative binomial distribution incorporates an additional parameter that allows the adjustment of the variance of the distribution independently of its mean.

GLMs were implemented using the glm and glm.nb functions in R. The predictor variables were {age, tumour grade, tubule score, pleomorphism score, mitotic score, mitotic count}, each of which was used within a two-factor model, with oestrogen receptor status as the second predictor variable. The response variable was the number of mutations of a particular type, from the set {substitutions + indels, substitutions, indels, copy number amplifications, C → T at CpG mutations, all driver mutations}.

Evaluation of strand bias in tumours displaying the mutator phenotype. To assess whether there was a strand bias of C → X (C → T, C → G and C → A) mutations in PD4120 and the other tumours showing the mutator phenotype, we first estimated the expected ratio of cytosines found in transcribed and untranscribed strands, by random sampling of 20,000 CCDS exons from Ensembl version 61. A χ-squared test was then used to examine whether the C → X mutations observed in each sample differed significantly from this ratio. Similar tests were conducted on the combined mutations from all mutator phenotype samples and on all mutator phenotype samples except PD4120.

31. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **469,** 539–542 (2011).
32. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26,** 589–595 (2010).
33. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25,** 2865–2871 (2009).
34. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl Acad. Sci. USA* **107,** 16910–16915 (2010).
35. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463,** 899–905 (2010).
36. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4,** 177–183 (2004).
37. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–158 (2007).
38. Nelder, J. A., & Wedderburn, R. Generalized linear models. *J. R. Stat. Soc. A* **135,** 370–384 (1972).

LETTER

Sequence analysis of mutations and translocations across breast cancer subtypes

Shantanu Banerji[1,2,3]†*, Kristian Cibulskis[1]*, Claudia Rangel-Escareno[4]*, Kristin K. Brown[5]*, Scott L. Carter[1], Abbie M. Frederick[1], Michael S. Lawrence[1], Andrey Y. Sivachenko[1], Carrie Sougnez[1], Lihua Zou[1], Maria L. Cortes[1], Juan C. Fernandez-Lopez[4], Shouyong Peng[2], Kristin G. Ardlie[1], Daniel Auclair[1], Veronica Bautista-Piña[6], Fujiko Duke[1], Joshua Francis[1], Joonil Jung[1], Antonio Maffuz-Aziz[6], Robert C. Onofrio[1], Melissa Parkin[1], Nam H. Pho[1], Valeria Quintanar-Jurado[4], Alex H. Ramos[1], Rosa Rebollar-Vega[4], Sergio Rodriguez-Cuevas[6], Sandra L. Romero-Cordoba[4], Steven E. Schumacher[1,2], Nicolas Stransky[1], Kristin M. Thompson[1], Laura Uribe-Figueroa[4], Jose Baselga[3,7], Rameen Beroukhim[1,2,3,8], Kornelia Polyak[2,3,9], Dennis C. Sgroi[3,10], Andrea L. Richardson[2,3,11], Gerardo Jimenez-Sanchez[4]†, Eric S. Lander[1,3,12], Stacey B. Gabriel[1], Levi A. Garraway[1,2,3], Todd R. Golub[1,3,13,14], Jorge Melendez-Zajgla[4], Alex Toker[3,5], Gad Getz[1], Alfredo Hidalgo-Miranda[4] & Matthew Meyerson[1,2,3,8]

Breast carcinoma is the leading cause of cancer-related mortality in women worldwide, with an estimated 1.38 million new cases and 458,000 deaths in 2008 alone[1]. This malignancy represents a heterogeneous group of tumours with characteristic molecular features, prognosis and responses to available therapy[2–4]. Recurrent somatic alterations in breast cancer have been described, including mutations and copy number alterations, notably *ERBB2* amplications, the first successful therapy target defined by a genomic aberration[5]. Previous DNA sequencing studies of breast cancer genomes have revealed additional candidate mutations and gene rearrangements[6–10]. Here we report the whole-exome sequences of DNA from 103 human breast cancers of diverse subtypes from patients in Mexico and Vietnam compared to matched-normal DNA, together with whole-genome sequences of 22 breast cancer/normal pairs. Beyond confirming recurrent somatic mutations in *PIK3CA*[11], *TP53*[6], *AKT1*[12], *GATA3*[13] and *MAP3K1*[10], we discovered recurrent mutations in the *CBFB* transcription factor gene and deletions of its partner *RUNX1*. Furthermore, we have identified a recurrent *MAGI3–AKT3* fusion enriched in triple-negative breast cancer lacking oestrogen and progesterone receptors and *ERBB2* expression. The MAGI3–AKT3 fusion leads to constitutive activation of AKT kinase, which is abolished by treatment with an ATP-competitive AKT small-molecule inhibitor.

Breast cancers are classified according to gene-expression subtypes: luminal A, luminal B, Her2-enriched (Her2 is also known as ERBB2), and basal-like[14]. Luminal subtypes are associated with expression of oestrogen and progesterone receptors and differentiated luminal epithelial cell markers. The subtypes differ in genomic complexity, key genetic alterations and clinical prognosis[2–4,15]. To discover genomic alterations in breast cancers, we performed whole-genome and whole-exome sequencing of 108 primary, treatment-naive, breast carcinoma/normal DNA pairs from all major expression subtypes (Table 1 and Supplementary Tables 1–3), 17 cases by whole-exome and whole-genome sequencing, 5 cases by whole-genome sequencing alone, and 86 cases by whole-exome sequencing alone.

In total, whole-exome sequencing was performed on 103 tumour/normal pairs, 54 from Mexico and 49 from Vietnam, targeting 189,980 exons comprising 33 megabases (Mb) of the genome and with a median of 85.1% of targeted bases covered at least 30-fold across the sample set. This analysis revealed a total of 4,985 candidate somatic substitutions (see https://confluence.broadinstitute.org/display/CGATools/MuTect for methods and data sets) and insertions/deletions (indels, see https://confluence.broadinstitute.org/display/CGATools/Indelocator for methods) in the target protein-coding regions and the adjacent splice sites, ranging from 14 to 307 putative events in individual samples (Supplementary Table 4). These mutations represented 3,153 missense, 1,157 silent, 242 nonsense, 97 splice site, 194 deletions, 110 insertions and 32 other mutations (Supplementary Table 5). The total mutation rate was 1.66 per Mb (range 0.47–10.5) with a non-silent mutation rate of 1.27 per Mb (range 0.31–8.05), similar to previous reports in breast carcinoma[6–9]. The mutation rate in breast cancer exceeds that of haematologic malignancies and prostate cancer, but is significantly lower than in lung cancer and melanoma[10,16–19]. The most common

Table 1 | Sample collections successfully completed sequencing and analysis

Patients	Mexico $N = 56$	Vietnam $N = 52$
Median age (range)	54 (37–92)	48 (31–81)
Source of normal DNA	Blood	Adjacent tissue
Pathology subtype (percent)		
Ductal	46 (82%)	41 (79%)
Lobular	4 (7%)	0 (0%)
DCIS	0 (0%)	9 (17%)
Other	6 (11%)*	2 (4%)†
Stage		
0	0 (0%)	9 (17%)
I	8 (14%)	3 (6%)
II	36 (64%)	31 (60%)
III	12 (21%)	9 (17%)
Expression subtype (per cent)‡		
Luminal A	24 (43%)	14 (27%)
Luminal B	13 (23%)	9 (17%)
Her2	9 (16%)	12 (23%)
Basal	5 (9%)	8 (15%)
Unknown	2 (4%)	3 (6%)
Normal-like	3 (5%)	6 (11%)

* Includes tubular carcinoma, medullary carcinoma, mucinous carcinoma and mixed carcinoma (3).
† Includes mucinous carcinoma (2).
‡ Based on PAM-50 classification.
DCIS, ductal carcinoma *in situ*.

[1]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [3]Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Instituto Nacional de Medicina Genómica, Mexico City 01900, Mexico. [5]Department of Pathology, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, Massachusetts 02215. [6]Instituto de Enfermedades de la Mama FUCAM, Mexico City 04980, Mexico. [7]Division of Hematology and Oncology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [8]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [9]Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [10]Derpartment of Pathology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [11]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [12]Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [13]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [14]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. †Present addresses: Department of Medical Oncology, CancerCare Manitoba, Winnipeg, Manitoba R3E 0V9, Canada (S.B.); Global Biotech Consulting Group, Mexico City 01900, Mexico (G.J.-S.).
*These authors contributed equally to this work.
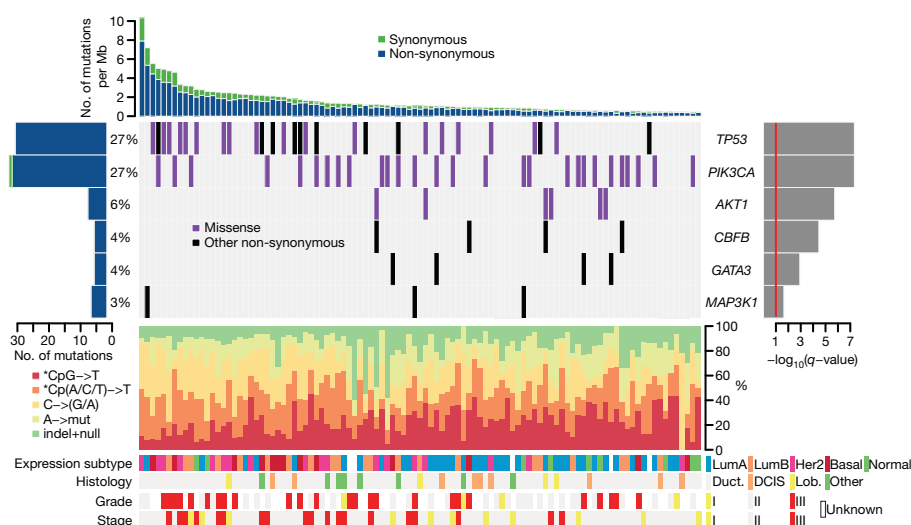
Figure 1 | Most significantly mutated genes in breast cancer as determined by whole-exome sequencing ($n = 103$). Upper histogram, rates of sample-specific mutations (substitutions and indels). Green, synonymous; blue, non-synonymous. Left histogram, number of mutations per gene and percentage of samples affected (colour coding as in upper histogram). Central heat map, distribution of significant mutations across sequenced samples ('Other non-synonymous' mutations: nonsense, indel and splice-site). Right histogram, $-\log_{10}$ score of MutSig q value. Red line at $q = 0.1$. Lower chart: top, rates of non-silent mutations within categories indicated by legend; bottom, key molecular features of samples in each column. DCIS, ductal carcinoma *in situ*; Duct., infiltrating ductal carcinoma; Lob., infiltrating lobular carcinoma; Lum, luminal.

mutation events observed are C to T transition events in CpG dinucleotides (Fig. 1 and Supplementary Fig. 4).

We performed validation experiments on 494 candidate mutations (representing all significantly mutated genes and genes in significantly mutated gene sets) using a combination of mass-spectrometric genotyping, 454 pyrosequencing, Pacific Biosciences sequencing and Illumina sequencing of matched formalin-fixed paraffin-embedded tissue, and confirmed the presence of 94% of protein-altering point mutations (Supplementary Table 4 and Supplementary Fig. 5); this validation rate is consistent with previous results that 95% of point mutations can be validated with orthogonal methods[16,17]. Only 18 out of 39 (46%) indels among significantly mutated genes were confirmed.

Six genes were found to be mutated with significant recurrence in the 103 whole-exome sequenced samples, by analysis with the MutSig algorithm[16,17] (https://confluence.broadinstitute.org/display/CGATools/MutSig) at a false discovery rate (FDR) < 0.1 after correction for multiple hypothesis testing (Supplementary Table 6a), manual review of reads, and subsequent orthogonal confirmation of somatic events (Fig. 1 and Supplementary Fig. 6). One gene, *CBFB*, is identified for the first time as a significantly mutated gene in breast cancer or any other epithelial cancer, to our knowledge, whereas the other five genes (*TP53*, *PIK3CA*, *AKT1*, *GATA3* and *MAP3K1*) have previously been reported as mutated in breast cancer[7,10,13]. This significantly mutated genes list, as any list produced by a statistical method, is probably incomplete and reflects the statistical power of our cohort size—larger sample sets will provide further statistical power.

Somatic mutations in *TP53* and *PIK3CA* were each present in 27% of samples, consistent with published frequencies[10,20] (Fig. 1). *TP53* mutations occur in samples with a higher mutation rate (*t*-test $P = 0.0079$ comparing samples with mutation rates greater than or less than the median 1.66 mutations per Mb) and were distributed across the gene in sites reported in COSMIC (http://www.sanger.ac.uk/genetics/CGP/cosmic/). Also, using the ABSOLUTE algorithm for determining allele-specific copy number[21], we observed that 21 out of 31 *TP53* mutations were homozygous (Supplementary Table 4). *PIK3CA* mutations were clustered in the helical (amino acids 542/545; 40%) and kinase domains (amino acid 1047; 47%)[20]. Six samples harboured the *AKT1* E17K mutation that alters the pleckstrin-homology (PH) domain and leads to activation of the kinase[12].

AKT1 and *PIK3CA* mutations, which activate the phosphatidylinositol-3-kinase (PI3K) pathway, were mutually exclusive in our data set. *MAP3K1*, recently reported as mutated in oestrogen-receptor-positive breast cancers[10], harboured five mutations in three patients with oestrogen-receptor-positive disease, and followed a pattern consistent with positive selection for recessive inactivation of the gene. In total, two frameshift, two nonsense and one missense mutation, combined with a homozygous deletion spanning the coding region were observed. Although the point mutations seemed to be heterozygous by copy-number analysis, two patients harboured dual mutations, consistent with compound heterozygous inactivation, although confirmatory phasing data were not available. The *GATA3* transcription factor gene harboured mutations in four patients with luminal tumours, including three previously unknown frameshift mutations near the 3′-end of the coding sequence. We also identified one previously described splice-site mutation that disrupts zinc-finger domains in GATA3 required for DNA binding[13].

CBFB, encoding the core-binding-factor beta subunit, was mutated in four oestrogen-receptor-positive samples, with one nonsense mutation and three truncating frameshift mutations (Fig. 2a). *CBFB* somatic mutations have been noted in isolated cases of breast cancer[6,10]. This is the first report of these mutations recurring at a significant rate above background; the sample size is not sufficient to determine whether these mutations are specific for oestrogen-receptor-positive subtypes. *CBFB* encodes the non-DNA-binding component of a heterodimeric protein complex, together with the DNA-binding RUNX proteins encoded by *RUNX1*, *RUNX2* and *RUNX3*. Copy-number analysis, using the ABSOLUTE algorithm[21], provides further evidence for loss of function of the RUNX1/CBFB complex in breast cancer: the cases with *CBFB* mutations seem to have hemizygous deletions of one parental allele, whereas two additional cases harbour homozygous deletions of *RUNX1* (Fig. 2b, c and Supplementary Figs 7 and 8). Oncogenic rearrangements of *RUNX1* or *CBFB* are common in acute myeloid leukaemia[22,23] (including the *CBFB*–*MYH11* translocation believed to have dominant negative function[22]). This is to our knowledge the first report of inactivation of this transcription factor complex in epithelial cancers.

Significance analysis restricted to somatic mutations in genes reported in COSMIC revealed three significantly mutated genes,
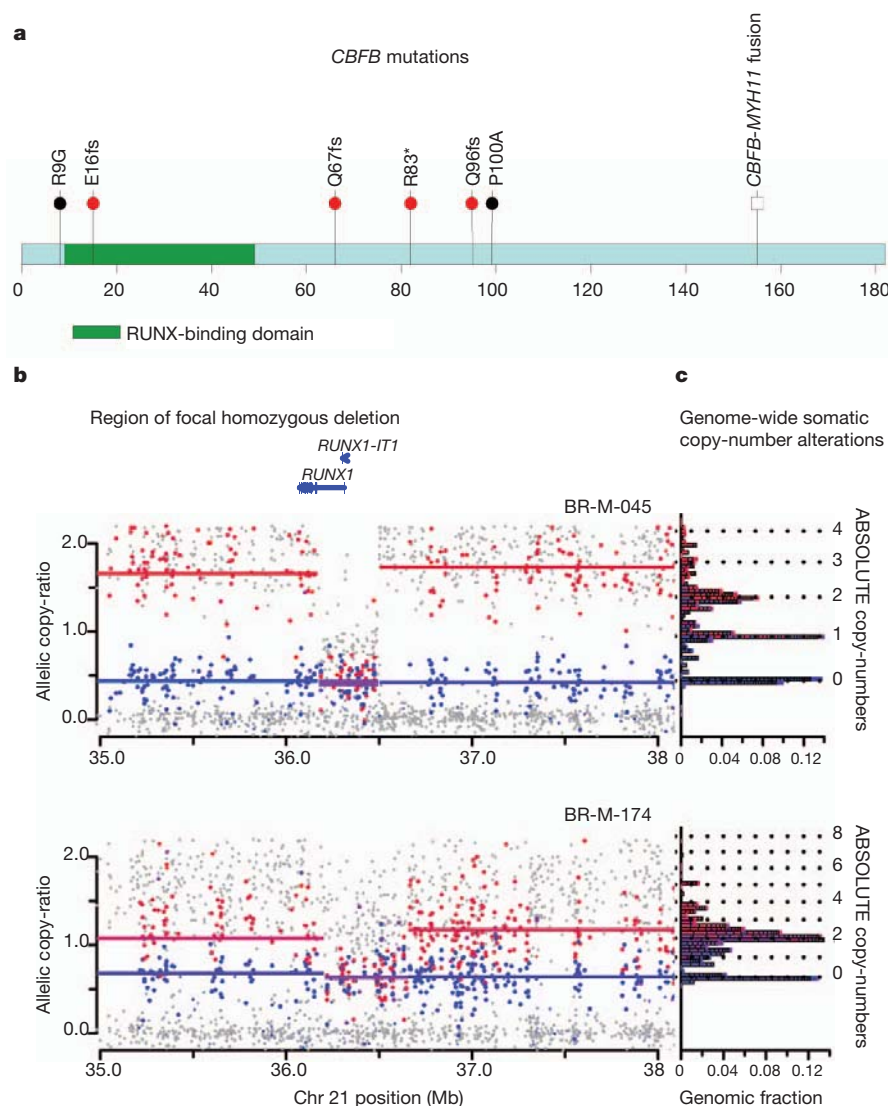
Figure 2 | CBFB mutations and RUNX1 deletions. a, *CBFB* coding region diagram, RUNX-binding domain in green. Mutations identified in this study (red bullets), previously identified mutations[6,10] (black bullets), and known *CBFB–MYH11* fusion indicated. **b**, Allelic copy ratios for the 3-Mb region surrounding *RUNX1* in samples BR-M-045 and BR-M-174. Dots indicate copy-ratios for individual SNP alleles. Red, higher copy-ratio allele for informative SNPs that are heterozygous in matched normal DNA; blue, lower-copy ratio SNPs; grey, uninformative SNPs (homozygous in matched normal). Lines indicate inferred segmental copy-ratios. Red, higher-copy segment; blue, lower-copy segment; purple, equal-copy segment. **c**, Histogram depicting bins of segmented copy number (*y* axis), with inferred integral copies shown by dotted lines; the length of each horizontal block corresponds to the fraction of the haploid genome at the copy number level, or 'genomic fraction' (*x* axis).

PIK3CA, *TP53* and *ERBB2*, the latter below the significance threshold in the complete analysis (Supplementary Table 7). *ERBB2* contained somatic mutations in three samples, with two being identical S310F mutations (these two samples are distinct on the basis of their germline and somatic genotypes). The S310F mutation can activate ERBB2 and is transforming *in vitro* (personal communication from H. Greulich). Neither sample with the S310F activating mutation has *ERBB2* amplification (Supplementary Fig. 9). The two samples belong to the Her2-enriched and luminal B subtypes, which typically have *ERBB2* amplification; this supports the notion that the observed mutations have a driving role in these tumours[10,24].

To identify candidate genomic rearrangements, we applied the dRanger algorithm[16,17] to the 22 cases with paired tumour/normal whole-genome sequencing data (Supplementary Table 8). The rate of rearrangements ranged from a median of 30 rearrangements per sample in the luminal A subtype (range 0–218) to the basal-like and Her2-enriched subtypes with a median of 237 and 246 rearrangements, respectively (Supplementary Fig. 10); the rates are similar to a recent report[15]. We performed polymerase chain reaction (PCR)

amplification on a subset of the candidate rearrangements (Supplementary Methods) and confirmed 89 out of 165 events (54%). No rearrangement was seen in more than one sample (Supplementary Table 8). In addition, we did not identify rearrangements previously observed by DNA sequencing[15] nor by complementary DNA (cDNA)-sequencing, including MAST and NOTCH family-gene fusions[25].

The discovery of recurrent driver rearrangements in other epithelial cancers[26,27] led to a closer examination of the list of confirmed rearrangements. In a triple-negative, basal-like subtype tumour, we observed a rearrangement between the genes *MAGI3* (membrane-associated guanylate kinase, WW and PDZ domain containing 3) on chromosome 1p and *AKT3* (v-akt murine thymoma viral oncogene homologue 3) on chromosome 1q, resulting in a balanced transloca-tion from intron 9 in *MAGI3* to intron 1 of *AKT3* (Fig. 3a). The previously unknown fusion genes were confirmed in tumour DNA by sequencing the product of PCR amplification (Fig. 3b). The *MAGI3* disruption is complemented by a hemizygous deletion of the other allele (Supplementary Fig. 11a). The expression levels of indi-vidual exons of *MAGI3* and *AKT3* correspond to the predicted
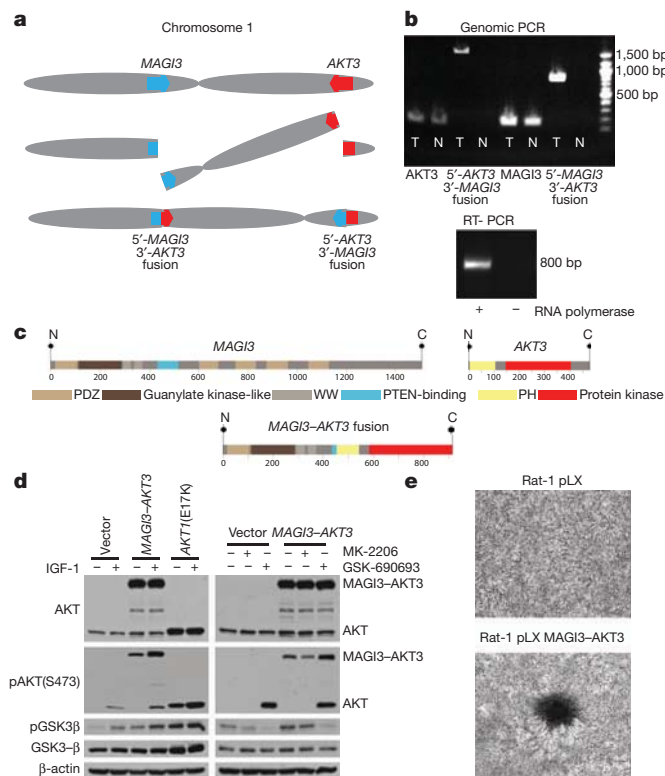
Figure 3 | MAGI3–AKT3 fusion gene. a, Diagram of balanced translocation between *MAGI3* and *AKT3*. **b**, Top, genomic DNA PCR for *AKT3*, *MAGI3* and both fusion products in tumour (T) and normal (N). Bottom, cDNA PCR of fusion gene in tumour. **c**, Above, MAGI3 and AKT3 protein domains; below, putative fusion protein. **d**, Immunoblots of lysates from ZR-75 cells transfected with vector, *MAGI3–AKT3* fusion, or *AKT1* E17K mutant, grown in low-serum media, for the indicated antibodies. Left, infected cells with and without insulin growth factor 1 (IGF-1) stimulation; right, treatment of vector or *MAGI3–AKT3* overexpressing cells with AKT inhibitors MK-2206 and GSK-690693. **e**, Focus formation assays with Rat-1 cells expressing pLX control or *MAGI3–AKT3*, and stained with crystal violet.

5′-*MAGI3–AKT3*-3′ fusion (Supplementary Fig. 11b), with this sample having the highest *AKT3* expression in the data set. Expression of the fusion gene was confirmed in the tumour sample by PCR amplification of the cDNA (Fig. 3b).

The rearrangement produces an in-frame fusion gene with a predicted MAGI3–AKT3 fusion protein that combines MAGI3 lacking the second PDZ domain, reported to bind to PTEN and be required for the inhibitory effect of PTEN on the PI3K pathway[28], together with an AKT3 region that retains an intact kinase domain but has a disruption of the pleckstrin homology domain before the glutamate at position 17 (Fig. 3c). *AKT3* shares significant homology to *AKT1* and is reported to be the dominant AKT family member expressed in hormone-receptor-negative breast cancers[29]. Together, the *MAGI3–AKT3* translocation and deletion of *MAGI3* could result in the combined loss of function of a tumour suppressor gene (*PTEN*) and activation of an oncogene (*AKT3*).

To evaluate oncogenic activity of the *MAGI3–AKT3* fusion, we expressed the fusion gene ectopically in ZR-75 cells. The MAGI3–AKT3 fusion protein is constitutively phosphorylated at serine 473 in the AKT3 kinase domain (numbered according to the wild-type protein) in the absence of growth factors (Fig. 3d); ectopically expressed AKT1 with an engineered E17K mutation is likewise constitutively phosphorylated (Fig. 3d), as previously reported[12]. Constitutive activation of the MAGI3–AKT3 kinase in turn activates downstream pathways as demonstrated by phosphorylation of GSK3β, an AKT substrate (Fig. 3d). Phosphorylation of GSK3β by the MAGI3–AKT3 fusion can be inhibited with an ATP-competitive

small molecule AKT inhibitor, GSK-690693, but not with an allosteric AKT inhibitor, MK-2206, that interacts with the PH domain of AKT (Fig. 3d). Overexpression of the *MAGI3–AKT3* fusion gene in Rat-1 fibroblast cell lines led to loss of contact inhibition and focus formation (Fig. 3e).

We screened 235 additional breast cancer samples for the presence of the 5′-*MAGI3–AKT3*-3′ fusion event by PCR with reverse transcription (RT–PCR) of cDNA followed by Sanger sequencing of breakpoints. The fusion was present in 8 of the 235 samples, including 5 out of 72 triple-negative (oestrogen-receptor-, progesterone-receptor- and Her2-negative) samples (Supplementary Fig. 12).

The power provided by whole-genome and whole-exome sequencing of a relatively large and diverse breast cancer sample set has enabled several significant discoveries, including the identification of recurrent inactivating mutations in *CBFB* and of a recurrent translocation of *MAGI3–AKT3*. The mutations in *CBFB*, *RUNX1* and *GATA3* suggest the importance of understanding epithelial cell differentiation and its regulatory transcription factors in breast cancer pathogenesis. The recurrent genomic fusion involving *AKT3* suggests that the use of ATP-competitive AKT inhibitors should be evaluated in clinical trials for the treatment of fusion-positive triple-negative breast cancers, a subtype where limited therapeutic options exist beyond systemic cytotoxic chemotherapy.

METHODS SUMMARY

All samples were obtained under institutional IRB approval and with documented informed consent. Breast cancer specimens from Mexico were paired with peripheral blood normal DNA whereas the Vietnamese samples were paired with DNA from normal adjacent breast tissue. Tumour RNA for each case was analysed on exon arrays to determine breast cancer expression subtype using the PAM50 classification method, whereas tumour/normal DNA pairs were analysed for copy number, allelic imbalance, and ancestry using single nucleotide polymorphism (SNP) arrays. A total of 108 samples, 17 both whole-genome sequencing and whole-exome sequencing, 86 whole-exome sequencing only, and 5 whole-genome sequencing only, passed initial qualification metrics, library construction, and successfully achieved desired sequencing depth (100× whole-exome sequencing; 30× whole-genome sequencing) on the Illumina sequencing platform (Supplementary Figs 1–3, Supplementary Tables 2 and 3). Tumour-specific point mutations, small insertions/deletions (indels), and rearrangements were detected by comparing tumour DNA to its paired normal DNA and using a series of algorithms to identify somatic events (Supplementary Fig. 2)[16,17]. Additional mutation calling was performed separately on tumour and normal DNA to identify germline mutation events that may confer susceptibility to breast carcinoma. Allele-specific copy number of each gene/mutation was determined using the HAPSEG and ABSOLUTE analysis methods. Confirmation of point mutations and indels was performed using mass-spectrometry-based genotyping and orthogonal next-generation sequencing methods, whereas putative in-frame genomic rearrangements were PCR-amplified from DNA to confirm the presence of the event.

A complete description of the materials and methods is provided in the Supplementary Information. Access to the data and computational algorithms used in this study can be found at https://confluence.broadinstitute.org/display/CGATools/Home.

1. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J. Clin.* **61,** 69–90 (2011).
2. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98,** 10869–10874 (2001).
3. Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10,** 529–541 (2006).
4. Gatza, M. L. *et al.* A pathway-based classification of human breast cancer. *Proc. Natl Acad. Sci. USA* **107,** 6994–6999 (2010).
5. King, C. R., Kraus, M. H. & Aaronson, S. A. Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science* **229,** 974–976 (1985).
6. Sjöblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314,** 268–274 (2006).
7. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318,** 1108–1113 (2007).
8. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
9. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464,** 999–1005 (2010).

10. Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466,** 869–873 (2010).
11. Samuels, Y. *et al.* High frequency of mutations of the *PIK3CA* gene in human cancers. *Science* **304,** 554 (2004).
12. Carpten, J. D. *et al.* A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* **448,** 439–444 (2007).
13. Usary, J. *et al.* Mutation of *GATA3* in human breast tumors. *Oncogene* **23,** 7669–7678 (2004).
14. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl Acad. Sci. USA* **100,** 8418–8423 (2003).
15. Stephens, P. J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462,** 1005–1010 (2009).
16. Berger, M. F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470,** 214–220 (2011).
17. Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471,** 467–472 (2011).
18. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463,** 191–196 (2010).
19. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463,** 184–190 (2010).
20. Bachman, K. E. *et al.* The *PIK3CA* gene is mutated with high frequency in human breast cancers. *Cancer Biol. Ther.* **3,** 772–775 (2004).
21. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnol.* doi:10.1038/nbt.2203 (29 April 2012).
22. Cameron, E. R. & Neil, J. C. The *Runx* genes: lineage-specific oncogenes and tumor suppressors. *Oncogene* **23,** 4308–4314 (2004).
23. Shigesada, K., van de Sluis, B. & Liu, P. P. Mechanism of leukemogenesis by the inv(16) chimeric gene *CBFB/PEBP2B-MHY11*. *Oncogene* **23,** 4297–4307 (2004).
24. Stephens, P. *et al.* Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* **431,** 525–526 (2004).
25. Robinson, D. R. *et al.* Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nature Med.* **17,** 1646–1651 (2011).
26. Soda, M. *et al.* Identification of the transforming *EML4–ALK* fusion gene in non-small-cell lung cancer. *Nature* **448,** 561–566 (2007).
27. Tomlins, S. A. *et al.* Recurrent fusion of *TMPRSS2* and ETS transcription factor genes in prostate cancer. *Science* **310,** 644–648 (2005).
28. Wu, Y. Interaction of the tumor Suppressor PTEN/MMAC with a PDZ domain of MAGI3, a novel membrane-associated guanylate kinase. *J. Biol. Chem.* **275,** 21477–21485 (2000).
29. Nakatani, K. *et al.* Up-regulation of Akt3 in estrogen receptor-deficient breast cancers and androgen-independent prostate cancer lines. *J. Biol. Chem.* **274,** 21528–21532 (1999).

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

LETTER

Astrocyte glypicans 4 and 6 promote formation of excitatory synapses via GluA1 AMPA receptors

Nicola J. Allen[1]†, Mariko L. Bennett[1], Lynette C. Foo[1], Gordon X. Wang[2], Chandrani Chakraborty[1], Stephen J. Smith[2] & Ben A. Barres[1]

In the developing central nervous system (CNS), the control of synapse number and function is critical to the formation of neural circuits. We previously demonstrated that astrocyte-secreted factors powerfully induce the formation of functional excitatory synapses between CNS neurons[1]. Astrocyte-secreted thrombospondins induce the formation of structural synapses, but these synapses are postsynaptically silent[2]. Here we use biochemical fractionation of astrocyte-conditioned medium to identify glypican 4 (Gpc4) and glypican 6 (Gpc6) as astrocyte-secreted signals sufficient to induce functional synapses between purified retinal ganglion cell neurons, and show that depletion of these molecules from astrocyte-conditioned medium significantly reduces its ability to induce postsynaptic activity. Application of Gpc4 to purified neurons is sufficient to increase the frequency and amplitude of glutamatergic synaptic events. This is achieved by increasing the surface level and clustering, but not overall cellular protein level, of the GluA1 subunit of the AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) glutamate receptor (AMPAR). Gpc4 and Gpc6 are expressed by astrocytes in vivo in the developing CNS, with Gpc4 expression enriched in the hippocampus and Gpc6 enriched in the cerebellum. Finally, we demonstrate that Gpc4-deficient mice have defective synapse formation, with decreased amplitude of excitatory synaptic currents in the developing hippocampus and reduced recruitment of AMPARs to synapses. These data identify glypicans as a family of novel astrocyte-derived molecules that are necessary and sufficient to promote glutamate receptor clustering and receptivity and to induce the formation of postsynaptically functioning CNS synapses.

To understand how astrocytes regulate functional synapse formation, we examined postsynaptic function, AMPAR levels and AMPAR localization at synapses between purified retinal ganglion cells (RGCs) cultured alone or with a feeder layer of astrocytes (Supplementary Fig. 1a–d). Astrocytes strengthen individual excitatory glutamatergic synapses in RGCs, as shown by increased frequency and amplitude of miniature excitatory postsynaptic currents[1] (mEPSCs) (Fig. 1a–c). In RGCs, mEPSCs are mediated purely by AMPARs, composed of combinations of four subunits, GluA1 to GluA4, forming tetramers[1,3]. Astrocytes do not greatly alter total AMPAR levels in RGCs (except for a small significant increase in GluA4), and thus do not induce the synthesis of new AMPARs or block the degradation of existing receptors (Fig. 1d, e). Astrocytes do, however, increase surface levels of all AMPAR subunits on RGCs by a factor of three, as shown by surface biotinylation and quantitative western blotting (Fig. 1f, g). Surface staining for GluA1-containing AMPARs demonstrated that the increased surface receptors are clustered together in puncta throughout the dendrites (Fig. 1h–j). These results demonstrate that astrocyte-derived signals lead to increased surface levels and clustering of pre-existing AMPARs.

We previously identified thrombospondins and hevin as astrocyte-secreted proteins sufficient to induce structural synapse formation, but

the synapses so formed are postsynaptically silent because they lack AMPARs[2,4]. Therefore, we used biochemistry to identify the astrocyte-secreted factor that is sufficient to induce functional synapse formation. Astrocytes were maintained in minimal medium and the factors they secreted collected as astrocyte-conditioned medium (ACM), which was concentrated and fed to RGCs, and was sufficient to induce functional synapse formation as assessed by electrophysiological recording of total synaptic activity and mEPSCs (Supplementary Fig. 1c, e–g). Analysis of ACM by two-dimensional electrophoresis revealed that hundreds of proteins were present, and size-exclusion experiments demonstrated the activity factor to be relatively large, between 100 and 300 kDa (Supplementary Fig. 2). To narrow down candidate factors present in ACM, we conducted affinity column fractionation, initially using individual columns and then combining them in series (Supplementary Fig. 3). In the final fractionation scheme, the unbound proteins were taken from a heparin column, bound and eluted from an anion column, and then bound and eluted from a hydrophobic interaction column. This final eluted protein fraction was fed to RGCs, and as it was unclear whether the activity factor would be sufficient to induce synapses directly, thrombospondin was included to induce structural synapse formation. This final fraction was sufficient to induce a large increase in synaptic activity (Fig. 1k), contained 1% of the starting protein (Supplementary Table 1) and was sixfold enriched for functional activity. This fraction was analysed by mass spectrometry and contained approximately 25 candidate factors (Supplementary Table 2).

To identify which candidate protein was sufficient to enhance synaptic activity, we overexpressed them in COS-7 cells (which do not secrete endogenous synaptogenic factors (Supplementary Fig. 4)) and fed the conditioned medium to RGCs along with thrombospondin. Most of the candidates lacked activity, whereas medium conditioned by Gpc4-expressing COS-7 cells was sufficient to induce a large increase in synaptic activity (Supplementary Fig. 4 and Fig. 1l). Western blotting of conditioned media from RGCs, cultured astrocytes and immuno-panned astrocytes (that closely resemble in vivo mature astrocytes[5]) demonstrated that astrocytes and not neurons secrete Gpc4 in vitro (Supplementary Fig. 5). Glypicans are a conserved family of heparan sulphate proteoglycans, with six members (Gpc1 to Gpc6) in mammals, and Gpc4 is homologous to Drosophila Dally-like[6]. The 63-kDa core protein is heavily glycosylated and can be in excess of 200 kDa in mass. Glypicans are tethered to the extracellular face of the plasma membrane of cells by GPI (glycosyl phosphoinositide) linkages[7,8], which can be cleaved by endogenous phospholipases, thus releasing the protein.

Having demonstrated that Gpc4 is sufficient to enhance total synaptic activity, we asked whether this was due to effects on postsynaptic strengthening of synapses. In subsequent experiments, RGCs were treated with purified Gpc4 (Supplementary Fig. 6a, b), in the absence of thrombospondin, to assess specific effects of Gpc4 on synapse formation and function. Gpc4 was sufficient to strengthen individual

[1]Stanford University School of Medicine, Department of Neurobiology, 299 Campus Drive, Fairchild Science Building D231, Stanford, California 94305-5125, USA. [2]Stanford University School of Medicine, Department of Molecular and Cellular Physiology, 279 Campus Drive, Beckman Center B100, Stanford, California 94305-5125, USA. †Present address: Molecular Neurobiology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.
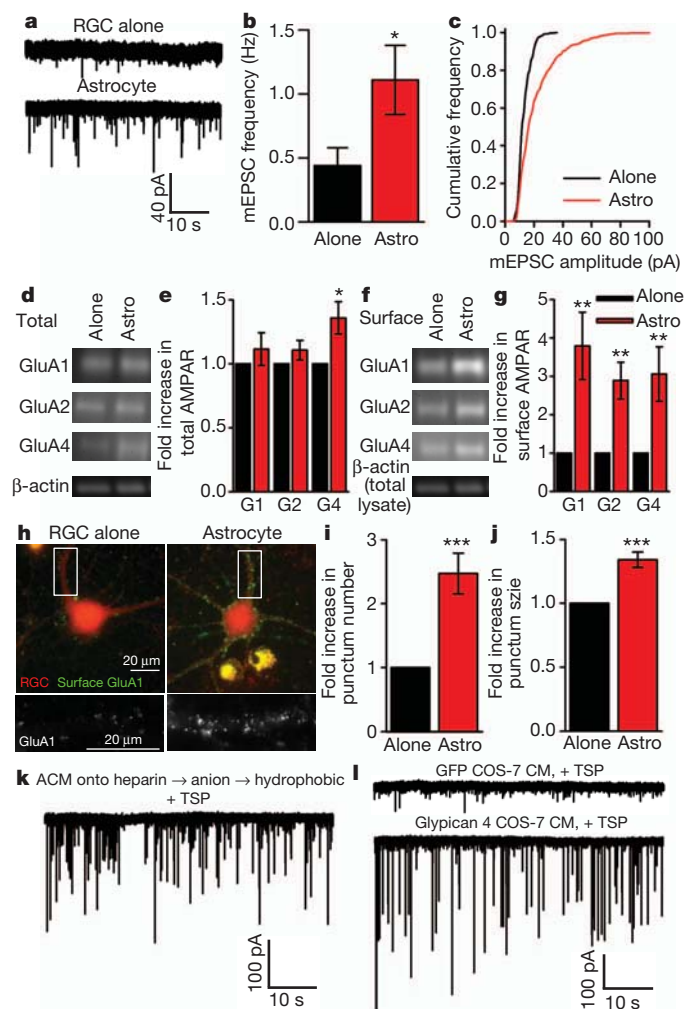
Figure 1 | Astrocyte signals strengthen synapses by recruitment of surface AMPARs. a–c, Example mEPSC recordings (**a**) show that frequency (**b**) and amplitude (**c**) are significantly increased in RGCs cultured with astrocytes. Average mEPSC amplitude: 12.2 ± 0.5 pA (RGCs alone, N = 13 cells), 20.4 ± 1.7 pA (astrocytes, N = 14, P < 0.0002). **d, e,** Astrocytes do not alter total AMPAR levels in RGCs. Western blots (**d**) of RGC lysates for AMPAR subunits GluA1 (G1), GluA2 (G2) and GluA4 (G4), and β-actin loading control, and quantification of band intensity relative to RGC alone (**e**). N = 7 experiments. **f, g,** Astrocytes increase surface AMPARs in RGCs. Western blots (**f**) of surface AMPAR subunits GluA1, GluA2 and GluA4, and β-actin loading control from total lysate (same experiment as **d**), and quantification of band intensity relative to RGC alone (**g**). N = 16 experiments. **h–j,** Astrocytes cluster GluA1-containing AMPARs on the RGC surface. Example images (**h**) of surface GluA1 (green) and RGC processes (red); bottom panel, enlargement of GluA1 (boxed white). Quantification of number (**i**) and size (**j**) of GluA1 clusters. N = 10 experiments. **k,** Total synaptic activity induced by final protein fraction from column fractionation of ACM. **l,** Total synaptic activity in RGCs cultured in COS-7-cell-conditioned medium (COS-7 CM) transfected with a control protein (green fluorescent protein (GFP)) or Gpc4, plus thrombospondin (TSP). *P < 0.05, **P < 0.01, ***P < 0.001; error bars, s.e.m.

Figure 2 | Gpc4 is sufficient to strengthen glutamatergic synapses and increase surface GluA1-containing AMPARs. a–d, Example mEPSC recordings (**a**) show that frequency (**b**) and amplitude (**c**) are significantly increased in RGCs cultured with Gpc4. Average traces aligned by rise time (**d**). Average mEPSC amplitude: 13.8 ± 0.7 pA (RGC alone, N = 16 cells), 20.0 ± 1.2 pA (astrocyte, N = 16, P < 0.05), 17.8 ± 1.4 pA (Gpc4, N = 13, P < 0.05). **e, f,** Gpc4 increases surface GluA1 AMPARs in RGCs. Western blots (**e**) of surface AMPAR subunits GluA1, GluA2 and GluA4 and neuron-specific enolase (NSE) loading control from total lysate, and quantification of band intensity relative to RGC alone (**f**). For full blot results, see Supplementary Fig. 9. N = 6 experiments for GluA1 and GluA2; N = 3 for GluA4. **g–i,** Gpc4 clusters GluA1-containing AMPARs on the RGC surface. Example images (**g**) of surface GluA1 (green) and RGC processes (red); bottom panel, enlargement of GluA1 (boxed white). Quantification of number (**h**) and size (**i**) of GluA1 clusters. N = 7 experiments. **j, k,** Gpc4 induces structural synapses. Example images (**j**) of presynaptic (bassoon, red) and postsynaptic (homer, green) staining; bottom panels, enlargements of the respective markers (boxed white). Quantification of synapse number (**k**) (co-localization of pre- and postsynaptic puncta). N = 6 experiments. *P < 0.05; error bars, s.e.m.

synapses, as shown by increased mEPSC frequency and amplitude, although the very large-amplitude events induced by astrocytes were absent (Fig. 2a–d and Supplementary Fig. 7). To determine whether increased synaptic activity was due to increased numbers of AMPARs on the cell surface, we isolated surface receptors and analysed them by western blotting. Gpc4 induced a 2.5-fold increase in surface GluA1, comparable to that observed in the presence of astrocytes; however, there was less of an increase in surface levels of the other subunits (Fig. 2e, f and Supplementary Fig. 9). Thus, Gpc4 specifically recruits GluA1-containing AMPARs, and astrocytes release additional factors

that bring GluA2, GluA3 and GluA4 AMPAR subunits to the synapse. Surface staining for GluA1 revealed a 2.5-fold increase in the number of receptor clusters on RGCs exposed to Gpc4, and a 20% increase in size, comparable to astrocytes (Fig. 2g–i). Gpc4-induced clustering of GluA1 on RGCs is dose dependent, being effective at 0.1–10 nM (comparable to levels in ACM) and ineffective at higher concentrations (Supplementary Fig. 6c, d). These experiments demonstrate that Gpc4 is sufficient to strengthen pre-existing synapses by increasing mEPSC amplitude; however, we also observed an increase in the number of mEPSCs and GluA1 clusters on RGCs, suggesting that Gpc4 can induce new structural synapses. We assessed synapse number by counting co-localization of pre- and postsynaptic markers, and observed a significant threefold increase in synapse number in RGCs treated with Gpc4

compared with RGCs grown alone (Fig. 2j, k). This demonstrates that Gpc4 is synaptogenic, although to a lesser degree than astrocytes.

The glypicans are a gene family with six members in mouse and human; *Gpc6* is the most homologous to *Gpc4* (ref. 9). We identified both Gpc4 and Gpc6 in the fractionation positive fraction (Supplementary Table 2), so we assessed whether Gpc6 is also synaptogenic. Gpc6 was sufficient to recruit GluA1 to the neuronal surface, and also induced structural synapse formation to the same extent as Gpc4 (Supplementary Fig. 8). To determine whether Gpc4 and Gpc6 are necessary for ACM to enhance postsynaptic activity, we used short interfering RNA (siRNA) to reduce expression of both in astrocytes, and confirmed protein reduction in ACM by western blotting (Supplementary Fig. 10a). ACM with reduced Gpc4 and Gpc6 was unable to

significantly increase mEPSC amplitude in RGCs (Fig. 3a–c) and unable to cluster GluA1 on the neuronal surface (Fig. 3d and Supplementary Fig. 10c, d). Overexpression of the human form of Gpc4, which is resistant to the siRNA, rescued the ability of ACM to cluster GluA1 receptors on the neuronal surface (Fig. 3d). The use of siRNA against individual glypicans was not as effective in reducing the ability of ACM to enhance total synaptic activity (Supplementary Fig. 10b). ACM with reduced levels of Gpc4 and Gpc6 still induced a significant increase in structural synapse formation, presumably owing to the presence of other synaptogenic proteins such as thrombospondin and hevin[2,4] (Fig. 3e). These results show that Gpc4 and Gpc6 are necessary components of ACM for enhancing postsynaptic activity, and that there is functional redundancy between family members.

To examine the time course of Gpc4 effects, we assayed surface clustering of GluA1 after 4 h and 18 h of treatment. There was no increase in receptor recruitment after 4 h, but after 18 h there was a significant increase in surface GluA1, to the same level as after 6 d (time point previously used; Fig. 3f and Supplementary Fig. 11a, b). Thus, Gpc4 does not immediately capture GluA1 AMPARs on the surface of the cell, which suggests that downstream signalling cascades are involved. We examined structural synapse formation at 4 h, 18 h, 1 d, 2 d, 3 d and 6 d after Gpc4 addition, and found that synapse number only significantly increased after 3 d (Fig. 3g and Supplementary Fig. 11c, d). Therefore, Gpc4 first clusters GluA1-containing AMPARs on the cell surface, and only then recruits postsynaptic scaffolding molecules. This suggests that clustering of GluA1 is a necessary step in Gpc4-induced synapse formation. To test this hypothesis, we used siRNA to decrease GluA1 in RGCs, and asked whether Gpc4 could still induce structural synapse formation (Supplementary Fig. 11e). Whereas astrocytes still increased synapse formation in RGCs lacking GluA1, Gpc4 did not (Fig. 3h). This mechanism of synaptogenesis is distinct from that of thrombospondin, which induces structural synapses that are postsynaptically silent[2]. Thus, astrocytes can induce excitatory synapse formation by at least two distinct mechanisms (Supplementary Fig. 19).

Because they are heparan sulphate proteoglycans, glypicans can interact with signalling receptors and morphogens (for example Wnt, fibroblast growth factor, Hedgehog and bone morphogenetic protein) via their sugar chains[10]. We generated a 'mutant' form of Gpc4 lacking glycosylation sites to investigate whether glycosylation is necessary for the synaptogenic effects of Gpc4 (Supplementary Fig. 12a). Glycosylation-deficient Gpc4 could not cluster GluA1 or induce structural synapse formation (Supplementary Fig. 12b, c and Supplementary Fig. 6d). Comparison by mass spectrometry of co-purified factors between full and mutant Gpc4 produced by HEK cells showed few differences between them, and these factors were not detected when Gpc4 was purified from astrocytes (data not shown), which can induce an increase in synaptic activity (Fig. 2a–d). We nevertheless tested a number of known glypican interactors for their ability to cluster surface GluA1 on RGCs and found none to be effective (Supplementary Fig. 12d). Thus, modification of Gpc4 by heparan sulphate is necessary for its synaptogenic activity, and this may be due to the necessity of heparan sulphate for structural interaction with a receptor rather than delivery of associated morphogens, as has been shown for Dally-like binding to its receptor LAR[11].

We next investigated the roles of Gpc4 and Gpc6 in developmental synapse formation and maturation *in vivo*. *In situ* hybridization revealed overlapping messenger RNA expression throughout the brain during postnatal development, particularly in the cortex, with Gpc4 enriched in the hippocampus and Gpc6 enriched in the cerebellum (Fig. 4a and Supplementary Fig. 13a). Prior gene profiling of purified forebrain CNS cells showed that messenger RNA for both Gpc4 and Gpc6 is highly enriched in astrocytes compared with neurons[12] (Supplementary Fig. 14a), which we confirmed using quantitative real-time PCR of purified astrocytes and neurons from the cortex and hippocampus (Supplementary Fig. 14b, c). Combining *in situ*
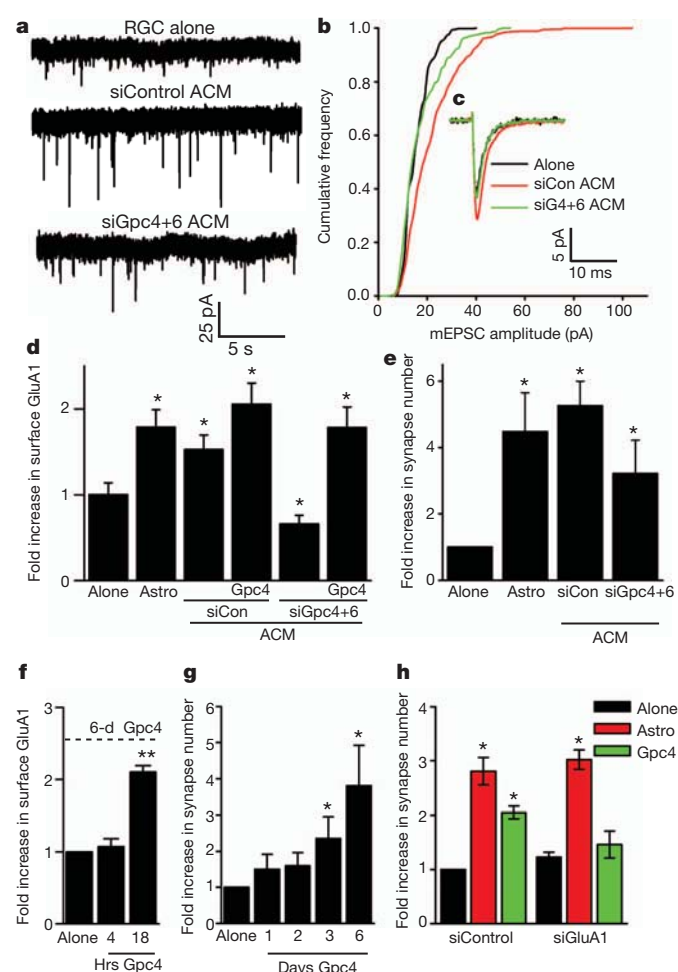
Figure 3 | Gpc4 and Gpc6 are necessary for ACM to cluster surface GluA1, and mechanism of action. a–c, Reduction of Gpc4 and Gpc6 levels in ACM reduces its ability to increase mEPSC amplitude in RGCs. Example mEPSC recordings (**a**), cumulative amplitude plot (**b**) and average traces aligned by rise time (**c**). Average mEPSC amplitude: 14.6 ± 0.6 pA (RGC alone, $N = 8$ cells), 20.6 ± 0.9 pA (siRNA control ACM, $N = 10$, $P < 0.05$), 16.0 ± 1.0 pA (siRNA Gpc4 and Gpc6 ACM, $N = 11$, $P = 0.3$). **d**, Reduction of Gpc4 and Gpc6 levels in ACM prevents GluA1 surface clustering, which is rescued by expression of siRNA-resistant Gpc4. $N = 30$ cells per condition. **e**, Reduction of Gpc4 and Gpc6 levels in ACM does not prevent ACM-induced structural synapse formation. $N = 3$ experiments. **f**, Time course of Gpc4-induced surface clustering of GluA1 shows 18 h of treatment is required to increase surface GluA1 levels. Dashed line, 6-d data from Fig. 2h. $N = 3$ experiments (4 h), 4 (18 h). **g**, Gpc4 does not rapidly induce structural synapse formation and requires 3 d of treatment. $N = 5$ experiments (1 d), 3 (2 d), 6 (3 d), 3 (6 d). **h**, Surface clustering of GluA1 is necessary for Gpc4-induced synapse formation, shown by the inability of Gpc4 to induce synapse formation in RGCs lacking GluA1. $N = 5$ experiments. *$P < 0.05$, **$P < 0.0001$; error bars, s.e.m.
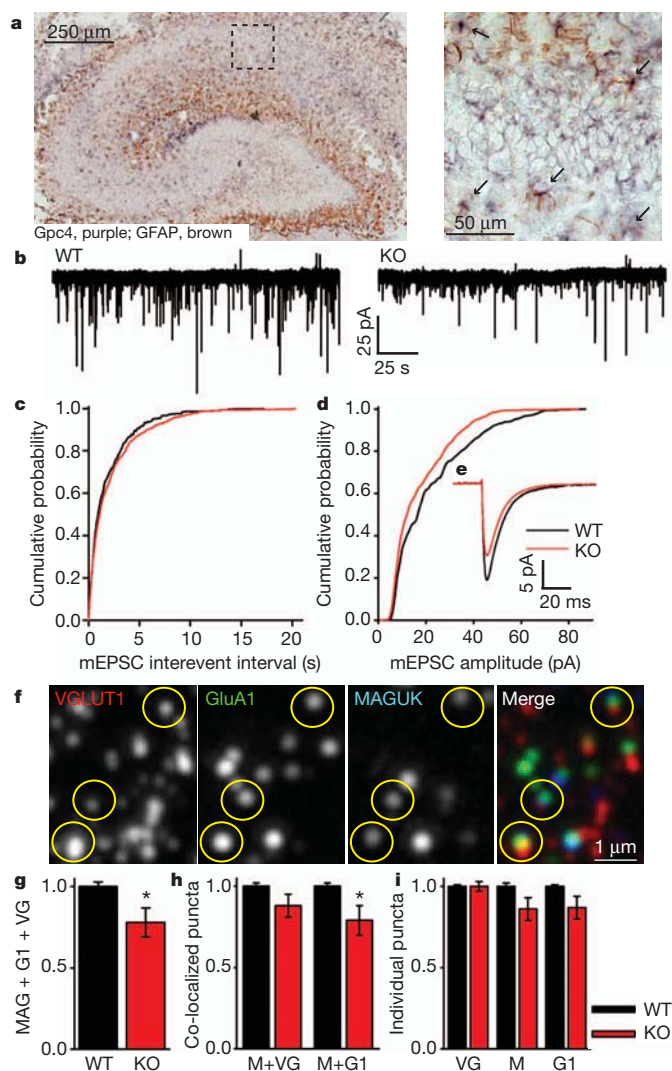
Figure 4 | Mice deficient in Gpc4 have weaker excitatory synapses *in vivo*.
a, *In situ* hybridization of Gpc4 messenger RNA in P6 mouse hippocampus.
Left: Gpc4 (purple) is expressed in synaptic regions and co-localizes with
astrocytes (glial fibrillary acidic protein (GFAP), brown). Right: enlargement of
boxed region; arrows mark Gpc4-positive astrocytes. **b–e**, Mice lacking Gpc4
(KO) have weaker excitatory synapses in CA1 pyramidal neurons at P12.
Example mEPSC recordings (**b**), cumulative probability plot of mEPSC
interevent interval (**c**) (no significant difference in frequency: wild type,
0.88 ± 0.25 Hz; *Gpc4*-knockout, 0.55 ± 0.08 Hz; $P = 0.16$), cumulative
probability plot of mEPSC amplitude (**d**) (significant decrease in amplitude:
wild type, 20.67 ± 2.16 pA; *Gpc4*-knockout, 16.07 ± 0.96 pA; $P < 0.05$),
average traces aligned by rise time (**e**). $N = 9$ cells (*Gpc4*-knockout, five mice), 6
(wild type, five mice). **f–i**, Mice lacking Gpc4 recruit fewer GluA1 AMPARs to
synaptic sites in hippocampal CA1 at P12. Example array tomography image
(**f**) from wild type; VGLUT1 (red), GluA1 (green), MAGUK (blue), triple co-
localization (yellow circles). There is a significant decrease in triple co-
localization of VGLUT1 + MAGUK + GluA1 in *Gpc4*-knockout mice
(**g**). There is no significant difference in structural synapse number (VGLUT1
+ MAGUK), but there is a significant decrease in GluA1 association with
MAGUK in *Gpc4*-knockout mice (**h**) and no difference in individual synaptic
markers (**i**). $N = 10$ arrays per genotype from four wild-type and four *Gpc4*-
knockout mice. $*P < 0.05$; error bars, s.e.m.

hybridization for Gpc4 with immunostaining for the astrocyte marker
GFAP demonstrated that hippocampal astrocytes express Gpc4 at
early postnatal periods (P6–P14) during the initiation of synapto-
genesis[13], and that astrocytic expression decreases with maturation
(P21) and switches to subsets of neurons (Fig. 4a and Supplemen-
tary Fig. 13b–d). Therefore, developing astrocytes express different
glypicans with distinct but overlapping regional patterns.

To determine the *in vivo* function of Gpc4, we examined synapse
formation and function in hippocampal area CA1 of *Gpc4*-knockout
mice, and chose this region because Gpc4 is enriched here (Sup-
plementary Fig. 13a) and Gpc4 and Gpc6 have redundant functions
in vitro. We recorded mEPSCs from CA1 pyramidal neurons in acute
hippocampal slices from *Gpc4*-knockout and wild-type littermate
controls at P12, during functional synapse formation, and at P24,
when more mature synapses are present. We observed a significant
(22%) decrease in the amplitude of mEPSCs at P12, and at P24 a
significant shift in amplitude distribution remained (Fig. 4b–e and
Supplementary Figs 15 and 16). These effects were not due to gross
differences in dendritic architecture or neuronal membrane properties
(Supplementary Fig. 17), or to changes in total levels of synaptic
proteins (Supplementary Fig. 18). We used array tomography to
address whether the decreased mEPSC amplitude was due to a defect
in structural synapse formation or to reduced recruitment of GluA1
receptors to synapses in hippocampal area CA1. Functional excitatory
synapses were classified by triple co-localization of the presynaptic
vesicular marker VGLUT1 with the postsynaptic density marker
MAGUK, plus GluA1 at the postsynaptic side (Fig. 4f). There is a
significant (22%) decrease in this class of synapse at P12 in *Gpc4*-
knockout mice (Fig. 4g). This decrease is not due to a difference in
the number of individual puncta (Fig. 4i) or to less structural synapse
formation (VGLUT1 + MAGUK is not altered), but to a significant
decrease in the co-localization of GluA1 with MAGUK (Fig. 4h).
Therefore, the same number of synapses form in *Gpc4*-knockout mice
as in wild type, but these synapses recruit fewer GluA1 AMPARs,
probably leading to the observed physiological defect of smaller
mEPSCs. The amplitude decrease observed is similar to that seen when
the GluA1 subunit of the AMPAR is removed from CA1 pyramidal
neurons[14,15], and this is the subunit we have shown to be recruited to
the surfaces of RGCs by Gpc4 *in vitro* (Fig. 2e, f).

Here we have identified glypicans as a novel family of astrocyte-
secreted proteins that regulate glutamate receptor clustering and
excitatory synapse formation in neurons. Other secreted factors
have been described that can induce structural synapses that are post-
synaptically silent (thrombospondins and hevin[2,4]), increase surface
AMPAR levels (neuronal pentraxins and tumour necrosis factor-α
(refs 16, 17)) or alter AMPAR mobility and synaptic plasticity
(extracellular matrix molecules[18]). Glypicans are the first non-
neuronal secreted factors to be identified that are sufficient to induce
functional synapses that cluster pre- and postsynaptic density proteins,
are postsynaptically active and contain surface AMPARs. We have
demonstrated that astrocytes increase GluA1 to GluA4 on the surfaces
of neurons and that Gpc4 and Gpc6 specifically regulate GluA1 surface
expression, indicating that astrocytes produce multiple factors that
regulate the surface expression of different AMPAR subunits.
Notably, specific AMPAR subunits may be used for the developmental
initiation of synapse formation and also for the strengthening of
synapses during long-term potentiation, with GluA1 used during
synapse initiation and GluA2 and GluA3 used during synapse
maturation[19–22]. This would be consistent with a role for glypicans in
initiating nascent synapse formation by recruiting postsynaptic
GluA1, followed by the action of additional factors that induce synapse
maturation. Astrocytes may control these events by releasing distinct
factors, including glypicans and thrombospondin, with spatiotemporal
specificity, providing a new mechanism by which astrocytes control the
formation and maturation of neural circuitry.

An attractive candidate receptor for inducing the actions of Gpc4
and Gpc6 is LAR, a protein tyrosine phosphatase receptor. *Drosophila*
Dally-like can bind to and signal through LAR[11], and removal of LAR
family members from mammalian hippocampal neurons reduces their
ability to form synapses and to recruit AMPARs to synapses[23]. Mice
lacking Gpc6 die shortly after birth[24] from apparent breathing
difficulties (N.J.A. and B.A.B., unpublished observations), suggesting
neural dysfunction, which is reminiscent of the phenotype observed in

neuroligin-triple-knockout mice[25]. In addition, mutations in Gpc6 have been associated with attention deficit hyperactivity disorder[26] and neuroticism[27] in humans, disorders in which synaptic dysfunction has been implicated. Thus, the identification of glypicans as regulators of functional synapse formation has important implications for the formation of appropriate neuronal circuits in development and disease.

METHODS SUMMARY

For detailed methods, see Methods.

Purification and culture of RGCs and astrocytes. RGCs were purified by sequential immunopanning to greater than 99% purity from P5–P7 Sprague-Dawley rats (Charles Rivers) and cultured in serum-free medium containing BDNF, CNTF and forskolin on laminin-coated coverslips at 50,000 cells per well, as previously described[1,2]. Cortical astrocytes (MD astrocytes) were prepared as described in ref. 1. RGCs were cultured for 7–10 d to allow robust process outgrowth, and were then cultured with astrocyte inserts, ACM, COS-7-conditioned media, thrombospondin, Gpc4 or Gpc6 for an additional 6 d unless stated otherwise.

Column fractionation procedure. ACM (3.3 mg), collected from 10×15 cm plates of astrocytes, was diluted in ethanolamine buffer and passed over a heparin column to which the protein of interest did not bind. The unbound fraction was collected and passed over an anion column (Q column), to which the protein of interest did bind. The NaCl (salt) concentration was increased to 0.5 M to remove irrelevant proteins, and then to 2 M to elute proteins of interest. The 0.5–2 M anion column eluate was passed over a hydrophobic interaction column (phenyl (low sub)), to which the protein of interest did bind. The bound proteins were eluted by decreasing the NaCl concentration from 2 M to 0 M. This final eluate contained 32 μg protein (Supplementary Table 1) and was sixfold enriched for functional activity, and was analysed by mass spectrometry.

Mice. Gpc4-knockout mice were obtained from Genentech/Lexicon and were generated by homologous recombination targeting exon 3, confirmed by Southern blotting[24]. No Gpc4 was detected by western-blotting-conditioned media prepared from Gpc4-knockout astrocytes, demonstrating that no functional protein was being secreted (Supplementary Fig. 5c).

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

1. Ullian, E. M., Sapperstein, S. K., Christopherson, K. S. & Barres, B. A. Control of synapse number by glia. *Science* **291,** 657–661 (2001).
2. Christopherson, K. S. *et al.* Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis. *Cell* **120,** 421–433 (2005).
3. Traynelis, S. F. *et al.* Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62,** 405–496 (2010).
4. Kucukdereli, H. *et al.* Control of excitatory CNS synaptogenesis by astrocyte-secreted proteins Hevin and SPARC. *Proc. Natl Acad. Sci. USA* **108,** E440–E449 (2011).
5. Foo, L. C. *et al.* Development of a method for the purification and culture of rodent astrocytes. *Neuron* **71,** 799–811 (2011).
6. Ybot-Gonzalez, P., Copp, A. J. & Greene, N. D. Expression pattern of glypican-4 suggests multiple roles during mouse development. *Dev. Dyn.* **233,** 1013–1017 (2005).
7. Song, H. H. & Filmus, J. The role of glypicans in mammalian development. *Biochim. Biophys. Acta* **1573,** 241–246 (2002).
8. Watanabe, K., Yamada, H. & Yamaguchi, Y. K-glypican: a novel GPI-anchored heparan sulfate proteoglycan that is highly expressed in developing brain and kidney. *J. Cell Biol.* **130,** 1207–1218 (1995).
9. De Cat, B. & David, G. Developmental roles of the glypicans. *Semin. Cell Dev. Biol.* **12,** 117–125 (2001).
10. Van Vactor, D., Wall, D. P. & Johnson, K. G. Heparan sulfate proteoglycans and the emergence of neuronal connectivity. *Curr. Opin. Neurobiol.* **16,** 40–51 (2006).
11. Johnson, K. G. *et al.* The HSPGs Syndecan and Dally-like bind the receptor phosphatase LAR and exert distinct effects on synaptic development. *Neuron* **49,** 517–531 (2006).
12. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28,** 264–278 (2008).
13. Hsia, A. Y., Malenka, R. C. & Nicoll, R. A. Development of excitatory circuitry in the hippocampus. *J. Neurophysiol.* **79,** 2013–2024 (1998).
14. Lu, W. *et al.* Subunit composition of synaptic AMPA receptors revealed by a single-cell genetic approach. *Neuron* **62,** 254–268 (2009).
15. Andrasfalvy, B. K., Smith, M. A., Borchardt, T., Sprengel, R. & Magee, J. C. Impaired regulation of synaptic strength in hippocampal neurons from GluR1-deficient mice. *J. Physiol. (Lond.)* **552,** 35–45 (2003).
16. O'Brien, R. J. *et al.* Synaptic clustering of AMPA receptors by the extracellular immediate-early gene product Narp. *Neuron* **23,** 309–323 (1999).
17. Beattie, E. C. *et al.* Control of synaptic strength by glial TNF-α. *Science* **295,** 2282–2285 (2002).
18. Frischknecht, R. *et al.* Brain extracellular matrix affects AMPA receptor lateral mobility and short-term synaptic plasticity. *Nature Neurosci.* **12,** 897–904 (2009).
19. Kumar, S. S., Bacci, A., Kharazia, V. & Huguenard, J. R. A developmental switch of AMPA receptor subunits in neocortical pyramidal neurons. *J. Neurosci.* **22,** 3005–3015 (2002).
20. Friedman, H. V., Bresler, T., Garner, C. C. & Ziv, N. E. Assembly of new individual excitatory synapses: time course and temporal order of synaptic molecule recruitment. *Neuron* **27,** 57–69 (2000).
21. Groc, L., Gustafsson, B. & Hanse, E. AMPA signalling in nascent glutamatergic synapses: there and not there! *Trends Neurosci.* **29,** 132–139 (2006).
22. Malinow, R. & Malenka, R. C. AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.* **25,** 103–126 (2002).
23. Dunah, A. W. *et al.* LAR receptor protein tyrosine phosphatases in the development and maintenance of excitatory synapses. *Nature Neurosci.* **8,** 458–467 (2005).
24. Tang, T. *et al.* A mouse knockout library for secreted and transmembrane proteins. *Nature Biotechnol.* **28,** 749–755 (2010).
25. Varoqueaux, F. *et al.* Neuroligins determine synapse maturation and function. *Neuron* **51,** 741–754 (2006).
26. Lesch, K.-P. *et al.* Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J. Neural Transm.* **115,** 1573–1585 (2008).
27. Calboli, F. C. *et al.* A genome-wide association study of neuroticism in a population-based sample. *PLoS ONE* **5,** e11504 (2010).

Author Contributions N.J.A. and B.A.B. designed the experiments and wrote the manuscript; N.J.A., M.L.B., L.C.F. and C.C. performed and analysed the experiments; and G.X.W. and S.J.S. assisted with data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. 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 N.J.A. (nallen@salk.edu).

METHODS

Purification and culture of RGCs and astrocytes. RGCs were purified by sequential immunopanning to greater than 99% purity from P5–P7 Sprague-Dawley rats (Charles Rivers) and cultured in serum-free medium containing BDNF, CNTF and forskolin on laminin-coated coverslips at 50,000 cells per well, as previously described[1,2]. Cortical astrocytes (MD astrocytes) were prepared as described in ref. 1. In some experiments, astrocytes were prepared by immuno-panning (IP astrocytes), as recently described in ref. 5. RGCs were cultured for 7–10 d to allow robust process outgrowth, and were then cultured with astrocyte inserts, ACM, COS-7-conditioned media, thrombospondin, Gpc4 or Gpc6 for an additional 6 d unless stated otherwise.

RGC transfection. RGC transfection with siRNA, along with GFP to mark transfected cells, was carried out using Lipofectamine as described in ref. 28. An OnTarget Plus siRNA pool against GluA1 was obtained from Dharmacon (catalogue number L-097755-01). In companion control experiments, RGCs were transfected with the same amount of either a targeting control (siCyclophilin B) or a non-targeting control (siControl) pool from Dharmacon; results obtained were the same with both.

Mice. *Gpc4*-knockout mice were obtained from Genentech/Lexicon and were generated by homologous recombination targeting exon 3, confirmed by Southern blotting[24]. No Gpc4 was detected by western-blotting-conditioned media prepared from *Gpc4*-knockout astrocytes, demonstrating that no functional protein was being secreted (Supplementary Fig. 5c). Mice were maintained on a mixed genetic background (129/C57bl6/J, crossed three generations onto C57bl6/J), because they seemed less viable when repeatedly crossed onto a pure C57Bl6/J background (N.J.A, unpublished observations). All experiments were carried out using male wild-type and knockout littermates (glypican 4 is on the X chromosome, so experiments compared $Gpc4^{+/y}$ with $Gpc4^{-/y}$). To trace dendrites, *Gpc4*-knockout mice were crossed with Thy1-GFP-M mice obtained from JAX.

Preparation of ACM and siRNA. Cortical astrocytes were cultured in 10-cm plates in serum-containing medium until confluent. Cells were then washed three times with warm DPBS to remove serum, and placed in minimal conditioning medium for 4 d. Conditioning media contained phenol-red-free Neurobasal, glutamine, pyruvate, penicillin and streptomycin. ACM was collected and first centrifuged to pellet dead cells and debris, then placed in centrifugal concentrators (Sartorius) with a size cut-off filter of 5 kDa unless otherwise stated. ACM was concentrated 50-fold. Protein concentration was determined by Bradford assay, and ACM was fed to RGCs at 50–80 μg ml^{-1}. This method produced ACM that induced synaptic activity when fed to RGCs the majority of the time, as long as astrocytes were used soon after isolation and not passaged extensively.

For siRNA experiments, OnTarget Plus siRNA pools against rat glypican 4 (catalogue number, L-098055-01) and glypican 6 (catalogue number, L-106892-01) were obtained from Dharmacon. Following validation, individual siRNAs against glypican 4 (sequence 9) and glypican 6 (sequence 6) were selected and used in all experiments (Supplementary Fig. 6). In companion control experiments, astrocytes were transfected with the same amount of either a targeting control (siCyclophilin B) or a non-targeting control (siControl) pool from Dharmacon; results obtained were the same with both. In rescue experiments, astrocytes were co-transfected with complementary DNA (cDNA) for human glypican 4, which is resistant to the siRNA. Confluent astrocytes were put into single-cell suspension using trypsin, and Amaxa nucleofection was used to introduce siRNA into the astrocytes. Astrocytes were plated into RGC growth medium minus B27 and growth factors for 24 h to recover from the transfection and to allow for knockdown of the RNA and subsequent reduction in secretion of the proteins of interest. Astrocytes were then washed three times with warm DPBS and placed in minimal conditioning medium for 3 d before collection of ACM as outlined above. Comparable results were seen when Lipofectamine, rather than nucleofection, was used to introduce siRNA to astrocytes (data not shown). Knockdown of glypican 4 and glypican 6 was validated by western blotting of ACM for glypican 4 (Proteintech rabbit anti-glypican 4) and glypican 6 (RandD goat anti-glypican 6). RGCs were fed with equal amounts of protein from control and siGlypican4+6 ACM, between 30–40 μg ml^{-1}. ACM was fed to RGCs in a range that was at the low end of the effective dose for control ACM, so that the effect of reducing the levels of glypican 4 and glypican 6 in ACM could be observed. This was necessary owing to incomplete knockdown of glypican 4 and glypican 6 in astrocytes, so the levels present in ACM were reduced rather than eliminated.

Two-dimensional gel electrophoresis. ACM was concentrated as outlined above, loaded onto an IPG strip, pH 3–10, and first separated by isoelectric point followed by separation by mass on an SDS gel following manufacturers' instructions (Biorad). The separated proteins were visualized using silver staining following manufacturers' instructions (Invitrogen).

Column fractionation procedure. ACM was prepared as outlined above and diluted into column loading buffer appropriate for each column. Columns (GE Healthcare) were prepacked and 1 ml in size, and fractionation procedures were carried out at room temperature (18–22 °C) in a tissue culture hood to keep the ACM sterile, using a syringe to apply manual pressure. Ethanolamine, pH 9.5 (20 mM), was used as the buffer in all experiments. Initially each column was tested in isolation and then the columns were combined in series in the following procedure (Supplementary Fig. 3). ACM (3.3 mg), collected from 10×15 cm plates of astrocytes, was diluted in ethanolamine buffer and passed over a heparin column to which the protein of interest did not bind. The unbound fraction was collected and passed over an anion column (Q column), to which the protein of interest did bind. The NaCl (salt) concentration was increased to 0.5 M to remove irrelevant proteins, and then to 2 M to elute proteins of interest. The 0.5–2 M anion column eluate was passed over a hydrophobic interaction column (phenyl (low sub)), to which the protein of interest did bind. The bound proteins were eluted by decreasing the NaCl concentration from 2 M to 0 M. This final eluate contained 32 μg protein (Supplementary Table 1) and was sixfold enriched for functional activity, and was analysed by mass spectrometry. Column fractions were tested by feeding them to RGCs at known protein concentrations, in the presence of thrombospondin to induce structural synapses, followed by electrophysiological recording of total synaptic activity. Before they were fed to RGCs, the column fractions went through a buffer exchange from ethanolamine buffer to Neurobasal, to remove high levels of salt. Pierce Zeba desalt spin columns were used for buffer exchange following manufacturer's instructions.

Mass spectrometry analysis. Both the positive fraction (phenyl column eluate) and the negative fraction (phenyl column unbound fraction) were analysed for comparison. Proteins were analysed in solution, reduced with DTT and alkylated with iodoacetamide, and then digested with trypsin. The liquid peptide mixture was passed over an HPLC column to separate individual peptides, which were then spotted onto a MALDI plate. Individual spots on the MALDI plate were analysed using MS/MS on an Applied Biosystems machine. Peptides were identified using the manufacturer's software and the NCBI database

COS-7 cell overexpression screen. Expression constructs for the majority of the proteins identified from the column fractionation procedure were obtained from Open Biosystems or Origene. The identity of the cDNA was verified by sequencing before use. COS-7 cells were used as a 'negative cell line' in which to express candidates, as COS-7 CM induced little synaptic activity in RGCs (Supplementary Fig. 4). COS-7 cells were transfected with cDNA using Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions. Three hours after transfection, cells were washed three times with warm DPBS and placed into minimal conditioning medium for 3 d, and conditioned media was collected as described for ACM. COS-7 CM was fed to RGCs in the presence of thrombospondin to induce structural synapse formation.

Recombinant proteins and DNA constructs. Full-length cDNA for glypican 4 (rat cDNA from Open Biosystems, clone no. 7124728) or glypican 6 (mouse cDNA from Open Biosystems, clone no. 5008374) with a 6-histidine tag at the amino terminus was cloned into pAPtag5 vector (GenHunter) between the SfiI and XhoI sites. These were expressed in HEK293 cells or astrocytes, which were transfected using Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions. The secreted recombinant protein was then purified from conditioned culture media by Ni-chelating chromatography using Ni-NTA resin (Qiagen) following the manufacturer's instructions. The purity of the protein was assessed by running a sample on an SDS gel and staining with Coomassie blue (Pierce). Protein concentration was assessed by Bradford assay and western blotting for glypican 4 (rabbit anti-glypican 4, Proteintech) or glypican 6 (goat anti-glypican 6, RandD) and the 6-histidine tag (mouse anti-histidine, Abcam), and glypican 4 and glypican 6 were fed to neurons at concentrations between 0.1 and 10 nM. Purified human platelet TSP1 was obtained from Haematologic Technologies and fed to neurons at 5 μg ml^{-1}. For testing candidate glypican 4 interactors, recombinant proteins were purchased and fed at the concentrations indicated in the figure legend.

Glycosylation-deficient glypican 4 was generated by site-directed mutagenesis against the presumed glycosylation attachment sites, as in ref. 29. Mutagenesis was confirmed by DNA sequencing, and by the loss of high-molecular-weight smearing when the purified protein was run on a western blot.

Synapse assay on RGCs. For synapse quantification of RGC cultures, cells were fixed for 7 min with 4% paraformaldehyde (PFA), washed three times in phosphate-buffered saline (PBS) and blocked in 200 μl of a blocking buffer containing 50% normal goat serum and 0.1% Triton X-100 for 30 min. After blocking, coverslips were washed three times in PBS, and 200 μl of primary antibody solution, consisting of mouse anti-bassoon (1:500, Stressgen) and postsynaptic antibody against all isoforms of homer (1:1,000, Chemicon), was added to each coverslip. Coverslips were incubated overnight at 4 °C, washed three times in PBS and incubated with

200 μl of Alexa-conjugated secondary antibodies (Invitrogen) diluted 1:1,000 in antibody buffer. Following incubation for 2 h, coverslips were washed three or four times in PBS and mounted in Vectashield mounting medium with DAPI (Vector Laboratories Inc.) on glass slides (VWR Scientific). Secondary-only controls were routinely performed and revealed no significant background staining.

Mounted coverslips were imaged using a Nikon Eclipse E800 epifluorescence microscope (Nikon). Healthy cells that were at least two cell diameters from their nearest neighbours were identified and selected at random by eye using DAPI fluorescence. Eight-bit digital images of the fluorescence emission at both 594 and 488 nm were recorded for each selected cell using a monochrome charge-coupled-device (CCD) camera and SPOT image capture software (Diagnostic Instruments, Inc.). Set exposures were used for each experiment, and were calculated from the positive control condition (RGC + astrocyte). Merged images were analysed for co-localized puncta by using a custom plug-in (written by Barry Wark and available on request from nallen@salk.edu) for the NIH image-processing package IMAGEJ. In each experiment, 15–30 cells across three coverslips were imaged per condition. Graphs are averages of values obtained from multiple experiments, as detailed in the legend, and are normalized to the 'RGC alone' condition.

GluA1 live surface staining. Rabbit anti-GluA1 (Calbiochem), an antibody recognizing the extracellular region of GluA1, was added to the cell culture medium for 15 min at 37 °C at 5 μg ml^{-1} to label surface AMPA receptors. Cell Tracker Red CMPTX dye (Invitrogen) was added at 0.5 μM during the incubation to label the whole cell. Cells were washed three times with DPBS to remove unbound antibody, fixed for 5 min in 4% PFA (the short fixation time prevents permeabilization of the cell), washed three times with PBS and blocked in 50% goat serum for 30 min at room temperature. After blocking, cells were washed three times with PBS and incubated in goat anti-rabbit Alexa 488 secondary antibody 1:500 for 1 h at room temperature. Coverslips were washed three or four times in PBS and mounted in Vectashield mounting medium with DAPI (Vector Laboratories Inc.) on glass slides (VWR Scientific). Secondary-only controls were routinely performed and revealed no significant background staining. Rabbit polyclonal antibodies against intracellular proteins gave no staining using this method.

Mounted coverslips were imaged using a Nikon Eclipse E800 epifluorescence microscope (Nikon). Healthy cells that were at least two cell diameters from their nearest neighbours were identified at random using the Cell Tracker Red channel. Eight-bit digital images of the fluorescence emission at both 594 and 488 nm were recorded for each selected cell using a monochrome CCD camera and SPOT image capture software (Diagnostic Instruments, Inc.). Set exposures were used for each experiment, and were calculated from the positive control condition (RGC + astrocyte). The numbers of puncta per cell were analysed using the Integrated Morphometry application in METAMORPH. In each experiment, 15–30 cells across three coverslips were imaged per condition. Graphs are averages of values obtained from multiple experiments, as detailed in the legend, and are normalized to the 'RGC alone' condition.

Surface biotinylation and western blotting. For surface biotinylation of RGCs, cells were plated in 35-mm wells at 250,000 cells per well, and two or three wells were pooled per condition. Biotinylation was performed using a Pierce Cell Surface Protein Isolation Kit with the following modifications. All steps were performed at 4 °C unless otherwise stated. Cells were placed on ice and washed twice with DPBS, and then biotin at 0.25 mg ml^{-1} in DPBS was added for 20 min at 4 °C to label surface proteins. Cells were washed three times with TBS to remove unbound biotin, and lysed in RIPA buffer. RIPA buffer was added to the well for 5 min, and cells were removed by squirting, collected in an Eppendorf tube and rotated for 1 h to solubilize membranes fully. Tubes were then spun at 13,000 r.p.m. (16,000g) in an Eppendorf microcentrifuge for 15 min to pellet unsolubilized material, and the supernatant was collected. A sample was collected at this stage to analyse for total protein levels. The remainder of the solubilized proteins were mixed with streptavidin-conjugated beads overnight to isolate the biotinylated proteins. Unbound proteins were removed by centrifugation, and bound biotinylated proteins were collected by incubating the beads in SDS buffer plus DTT for 1 h at room temperature.

Both total RGC lysate proteins and surface RGC proteins were separated by size on 4–15% gradient SDS gels (Biorad) and transferred to PVDF membranes. Membranes were blocked in 5% milk for 1 h at room temperature, washed three times in PBS-Tween and incubated overnight at 4 °C in primary antibodies: mouse anti-β-actin 1:5,000 (Sigma), rabbit anti-NSE 1:1,000 (Polyscience), rabbit anti-GluA1 0.25 μg ml^{-1} (Millipore), rabbit anti-GluA2 0.25 μg ml^{-1} (Millipore), rabbit anti-GluA2/3 2.5 μg ml^{-1} (Millipore) and rabbit anti-GluA4 1:1,000 (Millipore). Horseradish-peroxidase-conjugated anti-mouse and anti-rabbit (1:5,000) were used as secondary antibodies (Millipore), and the detection was performed with an ECL-plus kit from Amersham. Signals were acquired using a QImager CCD system and analysed using the software provided by the manufacturer. Band

intensity was normalized to the control (RGC alone) condition in each experiment, and data are presented normalized to RGC alone.

Tissue western blotting. Hippocampi were dissected out in ice-cold PBS, and then each hippocampus was homogenized in 250 μl RIPA buffer containing protease and phosphatase inhibitor cocktails (Pierce). Lysates were solubilized for one hour at 4 °C, spun down to remove unsolubilized material, aliquoted and snap-frozen in liquid nitrogen, and stored at −80 °C until use. Protein concentrations were determined by BCA assay and were typically 5 μg μl^{-1}. For each sample, 5 μg of protein was loaded onto the gel. Blots were probed with antibodies as indicated in the figure legend, and values were normalized to the β-tubulin loading control for each sample.

Electrophysiology in culture. Total synaptic activity and mEPSCs were recorded by whole-cell patch-clamping RGCs at room temperature at a holding potential of −70 mV. The extracellular solution contained 140 mM NaCl, 2.5 mM CaCl$_2$, 2 mM MgCl$_2$, 2.5 mM KCl, 10 mM glucose, 1 mM NaH$_2$PO$_4$ and 10 mM HEPES (pH 7.4), plus TTX (1 μM) when mEPSCs were being recorded. Patch pipettes had resistances of 3–5 MΩ and the internal solution contained 120 mM potassium gluconate, 10 mM KCl, 10 mM EGTA and 10 mM HEPES (pH 7.2). We recorded mEPSCs using PCLAMP software for Windows (Axon Instruments), and analysed them using MINI ANALYSIS PROGRAM (SynaptoSoft). Results from at least four separate experiments were pooled for analysis.

Electrophysiology in hippocampal slices. Experiments were carried out on littermate wild-type and Gpc4-knockout mice aged P12–P14 and P21–P25, and recordings and analysis were both carried out blind to genotype. A chilled choline-based cutting solution was used for dissection and slicing containing 78.3 mM NaCl, 23 mM NaHCO$_3$, 23 mM glucose, 33.8 mM choline chloride, 2.3 mM KCl, 1.1 mM NaH$_2$PO$_4$, 6.4 mM MgCl$_2$ and 0.45 mM CaCl$_2$ (pH 7.4). Parasagittal brain slices of thickness 250 μm containing the hippocampus were cut on a Leica vibratome. Slices were then kept at 31 °C for 25 min in the choline cutting solution and for 30 min in isotonic saline solution (125 mM NaCl, 25 mM NaHCO$_3$, 25 mM glucose, 2.5 mM KCl, 1.25 mM NaH$_2$PO$_4$, 2 mM MgCl$_2$ and 2.5 mM CaCl$_2$ (pH 7.4)). Oxygenation (95% O$_2$, 5% CO$_2$) was continuously supplied during cutting, recovery and recording. Whole-cell voltage-clamp recordings of hippocampal CA1 pyramidal neurons were performed. Patch pipettes with resistances of 2–4 MΩ were pulled from thick-walled borosilicate glass capillaries and filled with an internal solution containing 130 mM CsMeSO$_3$ (Cs$^+$ was used instead of K$^+$ as the main cation to improve voltage uniformity), 4 mM NaCl, 10 mM HEPES, 5 mM EGTA, 0.5 mM CaCl$_2$, 4 mM MgATP, 0.5 mM Na$_2$GTP, 5 mM QX-314 (to suppress voltage-gated sodium currents), with its pH adjusted to 7.2 using CsOH. Access resistance was monitored throughout the recording and was <20 MΩ.

Recordings were carried out at room temperature in flowing isotonic saline containing 1 μM tetrodotoxin to block voltage-gated sodium channels, the GABA$_A$ receptor antagonist bicuculline (40 μM) and 25 μM D-AP5 to block NMDA receptors to isolate AMPA-mediated mEPSCs. In control experiments all events were eliminated by application of the AMPA receptor antagonist NBQX. We recorded mEPSCs for 5 min and analysed them using MINI ANALYSIS PROGRAM. Cumulative probability plots for the interevent interval and the amplitude of mEPSCs were generated for each cell, and plots from wild-type and knockout cells were averaged to produce the plots shown in the figures. In addition, data were analysed by taking the first 100 events from each cell (these particular events were used to prevent bias from very active cells), pooling them for wild-type and knockout condition, and generating cumulative probability plots from the pooled data. The same results were obtained using both analysis methods.

Analysis of dendrite length. In some recordings, Alexa 488 was included in the patch pipette solution. Slices were fixed for 1 h in 4% PFA on ice following recording, and were washed and mounted in Vectashield for imaging. In parallel experiments, vibratome sections were prepared from wild-type and Gpc4-knockout mice that had been crossed with Thy1-GFP-M mice to label sparse subsets of CA1 pyramidal neurons with GFP, and were processed in the same way. Labelled cells were imaged on a Leica confocal microscope. Z-stacks were acquired to encompass all of the dendritic area, and the maximum projection of the dendritic area generated. Total dendrite length was analysed using HCA-vision NEURITE ANALYSIS software.

Purification of cortical and hippocampal neurons and astrocytes and qRT-PCR. C57bl6 mouse pups at P7 were used for purifications. Astrocytes were purified using integrin beta 5 immunopanning as described in ref. 5. Neurons were purified using L1 immunopanning as described in ref. 30. RNA was isolated using a Qiagen RNeasy kit following the manufacturer's instructions.

For expression analysis, template cDNA was prepared from 50–200 ng of total RNA by reverse transcription. Immunopanned cells from either cortex or hippocampus were paired, and purity was assessed by enrichment for astrocyte marker (GFAP) and neuron marker (Tuj1). Gene expression was quantified using

quantitative real-time PCR (qRT–PCR) in combination with gene-specific primers and the SYBR GREEN system (Roche). The reactions were performed on an Eppendorf Realplex4 cycler (Eppendorf). All samples were run in duplicates. For each DNA fragment (primer pair), all biological samples, including the input standard curves, were amplified on a common 96-well plate in the same run. The following primer pairs were used: AGGTCGGTGTGAACGGATTTG (Gapdh forward); TGTAGACCATGTAGTTGAGGTCA (Gapdh reverse); TAGACCC CAGCGGCAACTAT (Tuj-1 forward); TTCCAGGTTCCAAGTCCACC (Tuj-1 reverse); GGGGCAAAAGCACCAAAGAAG (GFAP forward); GGGACAACTT GTATTGTGAGCC (GFAP reverse); CTGGAGGGTCCTTTCAACATT (Gpc4 forward); GACATCAGTAACCAGTCGGTC (Gpc4 reverse); TAGTCCTGTAT TGGCAGCCAC (Gpc6 forward); GGCTAATGTCTATAGCAGGGAA (Gpc6 reverse).

Primers efficiency was determined by generating standard curves. For confirmation, quantitative PCR products were sequenced. Values were normalized relative to Gapdh expression. Relative gene expression between paired samples was estimated using the $2^{-\Delta\Delta C_{T}}$ method.

***In situ* hybridization.** Full-length mouse cDNA for both glypican 4 and glypican 6 was used to generate probes approximately 2 kb in size (Open Biosystems: Gpc4 clone no. 3967797, Gpc6 clone no. 5008374). Digoxigenin-labelled, single-stranded antisense and sense riboprobes were prepared by transcription of the linearized plasmid using either T7 or Sp6 RNA polymerases and a DIG RNA labelling kit (Roche) as per the manufacturer's instructions. *In situ* hybridizations were performed essentially as described in ref. 31 with some modifications for postnatal tissue. Briefly, 10-μm-thick, fresh-frozen sections of mouse brain were air-dried and fixed in 4% paraformaldehyde for 10 min. After washing with PBS three times for 10 min each time, sections were acetylated by incubation in 0.1 M triethanolamine HCl and 0.25% acetic anhydride for 10 min. After washing with PBS three times for 5 min each time, hybridization was performed with probes at concentrations of 200–500 ng ml^{-1} in a hybridization solution (50% formamide, ×5 SSC, ×5 Denhardt's solution, 250 μg ml^{-1} baker's yeast RNA and 100 μg ml^{-1} salmon sperm DNA) at 72 °C for 16 h. After hybridization, slides were washed once in ×0.2 SSC at 72 °C for 60 min, once in ×0.2 SSC at room temperature for 5 min and once in buffer B1 (0.1 M Tris HCl (pH 7.5) and 150 mM NaCl) for 5 min, incubated in B1 with 10% normal goat serum for 1 h, and then incubated in B1 with 1% normal goat serum and 1:1000 anti-digoxygenin alkaline-phosphatase conjugated antibody (Roche) overnight at 4 °C. Slides were then washed in B1 three times for 5 min each time, equilibrated in buffer B3 (0.1 M Tris HCl (pH 9.5), 100 mM NaCl and 50 mM MgCl$_2$) four times for 15 min each time and then developed in B3 with 0.24 mg ml^{-1} levamisole, 0.375 μl ml^{-1} NBT and 3.5 μl ml^{-1} BCIP (Roche) until a colour precipitate was visible. The reaction was stopped by washing once in TBST and four times in water, and slides were mounted with coverslips using Glycergel mounting medium (Dako).

For GFAP immunostaining, anti-mouse GFAP antibody (Sigma) at 1:1000 was added with anti-digoxygenin antibody overnight at 4 °C. After washing in B1 three times for 5 min each time, anti-mouse Fc horseradish-peroxidase-conjugated antibody was added in 1% goat serum in B1 (1:500) for 2 h at room temperature before alkaline phosphatase was developed as above. Following washes in water, horseradish peroxidase was developed using DAB reagent (Vector Labs) for 8 min. Slides were washed six times in water and mounted with coverslips using Glycergel mounting medium.

Array tomography. Array tomography was carried out as described in ref. 32. Hippocampal slices from P12 *Gpc4*-knockout and wild-type littermate controls were prepared in the same way as for electrophysiological analysis (300-μm vibratome sections of live tissue), fixed with 4% PFA overnight at 4 °C, and then dehydrated and embedded in LRWhite resin using the benchtop protocol. Ribbons of 30–35 serial ultrathin sections (70–100 nm) were cut on an ultramicrotome (Leica), mounted on subbed glass coverslips and immunostained using antibodies against VGLUT1 (guinea pig, Millipore), GluA1 (rabbit, Abcam) and pan-MAGUK (mouse, NeuroMabs). Antibody binding was visualized using Alexa-488-, Alexa-594- and Alexa-647-labelled goat secondary antibodies (Invitrogen). Sections were mounted using SlowFade Gold antifade with DAPI (Invitrogen). Images were collected on a Zeiss Axiovert 200M fluorescence microscope with an AxioCam HRm CCD camera, using a Zeiss ×63/1.4 NA Plan Apochromat objective. The hippocampal area CA1 stratum radiatum was imaged, to analyse synapses proximal to the pyramidal neuron cell body.

Tissue volumes were reconstructed and aligned using IMAGEJ (NIH) and the multistackreg plug-in (Brad Busse), and then cropped to include CA1 stratum radiatum so that the same volume was analysed across animals. Image deconvolution was carried out using a custom function in MATLAB (Gordon Wang). Individual puncta were identified for each channel, and then the number of MAGUK + GluA1 co-localized puncta associated with a VGLUT1 puncta was quantified using a separate custom function in MATLAB (Gordon Wang). Puncta numbers were normalized to wild-type values.

Data analysis and presentation. All graphs represent average data with s.e.m. error bars. Statistical analysis was either by Student's *t*-test when only two groups were being compared, or by one-way analysis of variance when three or more groups were compared.

28. Eroglu, C. *et al.* Gabapentin receptor α2δ-1 is a neuronal thrombospondin receptor responsible for excitatory CNS synaptogenesis. *Cell* **139,** 380–392 (2009).
29. Williams, E. H. *et al.* Dally-like core protein and its mammalian homologues mediate stimulatory and inhibitory effects on Hedgehog signal response. *Proc. Natl Acad. Sci. USA* **107,** 5869–5874 (2010).
30. Steinmetz, C. C., Buard, I., Claudepierre, T., Nagler, K. & Pfrieger, F. W. Regional variations in the glial influence on synapse development in the mouse CNS. *J. Physiol. (Lond.)* **577,** 249–261 (2006).
31. Schaeren-Wiemers, N & Gerfin-Moser, A. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: in situ hybridization using digoxigenin-labelled cRNA probes. *Histochem. Cell Biol.* **100,** 431–440 (1993).
32. Micheva, K. D. & Smith, S. J. Array tomography: a new tool for imaging the molecular architecture and ultrastructure of neural circuits. *Neuron* **55,** 25–36 (2007).

LETTER

PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos

Toshinobu Nakamura[1]†, Yu–Jung Liu[2], Hiroyuki Nakashima[2], Hiroki Umehara[2], Kimiko Inoue[3], Shogo Matoba[3], Makoto Tachibana[4], Atsuo Ogura[3], Yoichi Shinkai[4] & Toru Nakano[1,2]

The modification of DNA by 5-methylcytosine (5mC) has essential roles in cell differentiation and development through epigenetic gene regulation[1]. 5mC can be converted to another modified base, 5-hydroxymethylcytosine (5hmC), by the tet methylcytosine dioxygenase (Tet) family of enzymes[2,3]. Notably, the balance between 5hmC and 5mC in the genome is linked with cell-differentiation processes such as pluripotency and lineage commitment[4–7]. We have previously reported that the maternal factor PGC7 (also known as Dppa3, Stella) is required for the maintenance of DNA methylation in early embryogenesis, and protects 5mC from conversion to 5hmC in the maternal genome[8,9]. Here we show that PGC7 protects 5mC from Tet3-mediated conversion to 5hmC by binding to maternal chromatin containing dimethylated histone H3 lysine 9 (H3K9me2) in mice. In addition, imprinted loci that are marked with H3K9me2 in mature sperm are protected by PGC7 binding in early embryogenesis. This type of regulatory mechanism could be involved in DNA modifications in somatic cells as well as in early embryos.

Maternal-genome chromatin bears considerable DNA methylation and contains H3K9me2 (ref. 10). Conversely, little DNA methylation remains and no H3K9me2 is present in the chromatin of the paternal genome. We found that DNA methylation of naked DNA and chromatin DNA did not lead to substantial differences in the binding of PGC7 to DNA (Supplementary Figs 1, 2a and 3 and Supplementary Discussion). By contrast, compared with the nucleosomes purified from wild-type embryonic stem (ES) cells, PGC7 showed significantly weaker binding to the nucleosomes purified from ES cells lacking G9a (also known as Ehmt2)—an H3K9me2-specific lysine methyltransferase—in which H3K9me2 was absent, without affecting DNA methylation status (Fig. 1a, b and Supplementary Figs 2b and 4a, b)[11]. However, PGC7 bound to nucleosomes purified from rescued G9a knockout ($G9a^{-/-}$) ES cells, which expressed the short or long form of G9a ($G9a^{-/-}$-G9a(S) and $G9a^{-/-}$-G9a(L), respectively), with comparable affinity to the nucleosomes of wild-type ES cells (Fig. 1a, b). Next, we conducted a chromatin-immunoprecipitation (ChIP) assay in PGC7-null and PGC7-expressing ES cells using an anti-PGC7 antibody. As shown in Fig. 1c, nucleosomes that were bound by PGC7 specifically contained H3K9me2. H3 with other methylation modifications, however, did not show binding to PGC7. We confirmed the H3K9me2-dependent PGC7 association with chromatin in wild-type, DNA methyltransferase triple-knockout ($Dnmt1^{-/-} Dnmt3a^{-/-} Dnmt3b^{-/-}$) mutants[12] and G9a-null ES cells using a stepwise salt-extraction method[13]. PGC7 and H3 were similarly extracted from the nuclear pellets of wild-type, $Dnmt1^{-/-} Dnmt3a^{-/-} Dnmt3b^{-/-}$ and G9a ($G9a^{-/-}$-G9a(S) ES cells (Fig. 1d and Supplementary Fig. 5a, b). By contrast, PGC7 was extracted from nuclear pellets of $G9a^{-/-}$ ES cells under a lower NaCl concentration compared with that in wild-type and $Dnmt1^{-/-} Dnmt3a^{-/-} Dnmt3b^{-/-}$ cells (Fig. 1d and Supplementary Fig. 5a, b), showing that PGC7 binding to chromatin without

H3K9me2 was weaker. These two experiments indicate that the association between PGC7 and chromatin is dependent on the presence of H3K9me2.

We studied the *in vitro* binding of PGC7 to various histone peptides to see whether PGC7 directly binds to H3K9me2. Although all histone peptides containing residues 1–21 bound to PGC7 to some extent, H3K9me2 bound to PGC7 most strongly (Fig. 1e). The same binding assay using truncated versions of PGC7, namely PGC7ΔC (containing amino acids 1–75) and PGC7ΔN (containing amino acids 76–150), showed that PGC7 bound to H3K9me2 by its amino terminus (Supplementary Fig. 6). Next, we carried out a competitive-binding assay to determine the binding strength of PGC7 and H3K9me2. As shown in Fig. 1f, only the excess amount of H3K9me2, but not that of the other peptides, outcompeted the binding of PGC7 to H3K9me2, indicating that the binding of PGC7 to H3K9me2 was stronger than that of other histone modifications.

PGC7 binding was higher at two representative H3K9me2-enriched loci, *Magea2* and *Wfdc15a*, but not at the control *Pou5f1* locus (Fig. 1g). By contrast, such enrichment was not observed under the G9a-null condition (Fig. 1g). These data clearly demonstrate that PGC7 specifically binds to the loci marked with H3K9me2 *in vivo*. Meanwhile, micrococcal nuclease (MNase) activity was inhibited significantly by the enforced expression of PGC7 and PGC7ΔNES, a nuclear export signal deleted mutant of PGC7 (ref. 8), in wild-type and $Dnmt1^{-/-} Dnmt3a^{-/-} Dnmt3b^{-/-}$ ES cells (Supplementary Fig. 5c–f and Supplementary Discussion). However, neither PGC7 nor PGC7ΔNES reduced MNase sensitivity in $G9a^{-/-}$ ES cells, indicating that the protective function of PGC7 required H3K9me2 (Supplementary Fig. 5a, b, g and h). These results indicate that the protective function of PGC7 is dependent on H3K9me2, but not DNA methylation.

We next asked whether the H3K9me2-dependent binding of PGC7 occurs under physiological conditions during early embryogenesis. Both the paternal and maternal pronuclei stained positively with anti-PGC7 antibody after conventional paraformaldehyde (PFA) fixation (PFA–Triton (PT) condition in Fig. 2a, b and Supplementary Fig. 7). However, the staining pattern was completely different when zygotes were treated with Triton X-100 before PFA fixation[14] (Triton–PFA (TP) condition in Fig. 2a, b and Supplementary Fig. 7). Under TP conditions, only the maternal pronucleus was labelled with the anti-PGC7 antibody, indicating that PGC7 in the paternal pronucleus was eluted by Triton X-100. In other words, PGC7 was tightly attached to a type of 'architecture' that was present in the maternal, but not the paternal, pronuclei.

Considering the maternal pronucleus-specific localization of H3K9me2 (Fig. 2b) and the results of our experiments using ES cells (Fig. 1), we hypothesized that chromatin containing H3K9me2 is the crucial structure to which PGC7 strongly binds in the maternal pronucleus. To test this hypothesis we expressed Jhdm2a (also known as
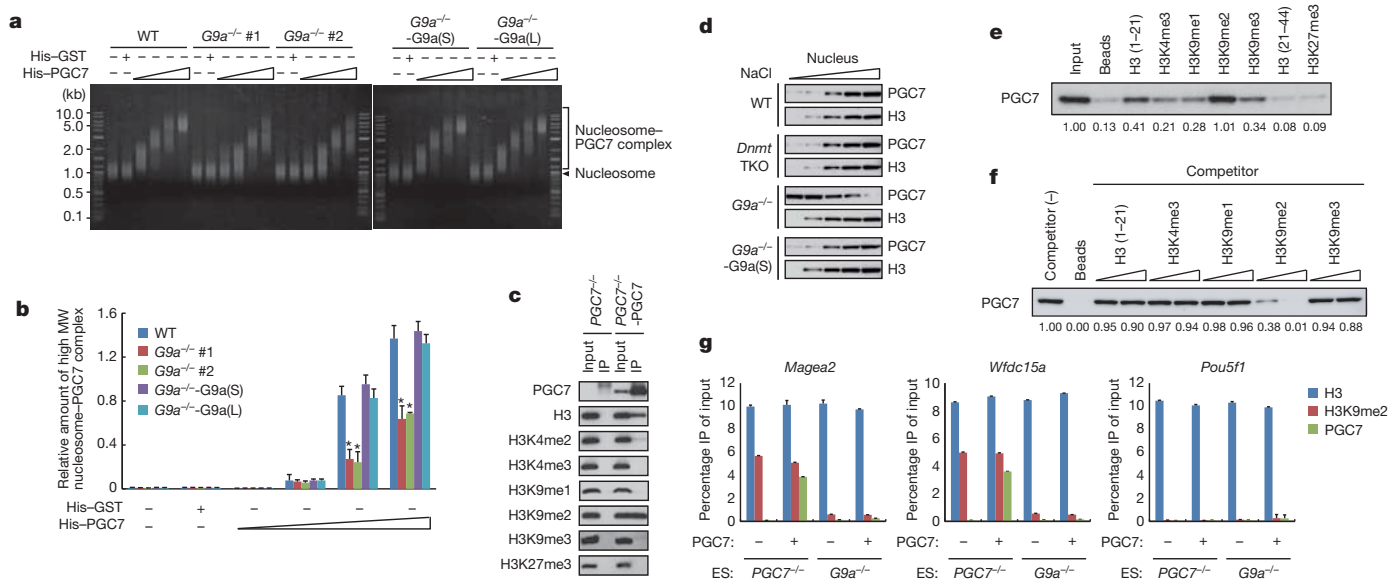
Figure 1 | Preferential binding of PGC7 to H3K9me2-marked chromatin.
a, **b**, Electrophoretic gel-mobility shift assay of His–PGC7. Nucleosomes (approximately 10 μg) purified from wild-type (WT) and $G9a^{-/-}$ ES cells, and $G9a^{-/-}$ ES cells in which $G9a^{-/-}$-G9a(S) and $G9a^{-/-}$-G9a(L) forms of G9a were expressed were incubated with increasing concentrations of His–PGC7 (2.5, 5, 10 and 20 μg). The binding mixtures were analysed on agarose gels (**a**). Ratios of high-molecular-weight (MW) PGC7–nucleosome complex (>5 kb) were determined using NIH Image J (**b**). Error bars indicate s.d. ($n = 3$). Binding affinities were significantly different between wild-type and $G9a^{-/-}$ ES cells (*$P < 0.005$, t-test). His–glutathione S-transferase (GST; 20 μg) was used as a negative control in both experiments. Similar results were obtained in at least three independent experiments and representative results are shown. **c**, Histone-methylation status of PGC7-containing chromatin. Anti-PGC7 antibody was used to immunoprecipitate chromatin from $PGC7^{-/-}$ ES cells stably expressing PGC7. Immunoprecipitates (IP) and aliquots of the input protein were analysed by immunoblotting with antibodies against various histone modifications. Essentially the same results were obtained in two independent experiments. **d**, Chromatin-binding status of PGC7 in ES cells. Nuclei were isolated from stable-PGC7-expressing wild-type, $Dnmt1^{-/-} Dnmt3a^{-/-} Dnmt3b^{-/-}$ triple knockout ($Dnmt$ TKO), $G9a^{-/-}$ or $G9a^{-/-}$-G9a(S) ES cells and were treated with DNase I under various

concentrations of NaCl (100, 200, 300, 400 and 500 mM) to separate nuclear extract from nuclear debris. Equivalent amount of aliquots were analysed by immunoblotting with anti-PGC7 or anti-H3 antibodies. **e**, In vitro assay of the binding of recombinant PGC7 to various H3 tail peptides. Recombinant His–PGC7 and histone H3 tail peptides with various methylation modifications were used in an in vitro peptide-binding assay. Binding characteristics were analysed by immunoblotting with anti-PGC7 antibody. Similar results were obtained in three independent experiments. **f**, PGC7 competitive peptide-binding assay. N-terminal biotinylated H3K9me2 peptide was immobilized on streptavidin–sepharose beads and incubated with recombinant His–PGC7 in the presence of increasing amounts of unmodified H3 or the indicated methylated H3 peptides. Binding characteristics were analysed by immunoblotting with anti-PGC7 antibody. Similar results were obtained in two independent experiments. **g**, Concomitant binding of PGC7 to the chromatin loci marked with H3K9me2 in ES cells. A ChIP analysis using the indicated antibodies was conducted in the $PGC7^{-/-}$ and $G9a^{-/-}$ ES cells with or without enforced PGC7 expression. The percentages of each PCR product in the immunoprecipitated sample per those of the input samples are shown (mean and s.d., $n = 3$). Anti-mouse or -rabbit IgG was used for the negative controls and the signals per total input of the negative controls were <0.12% in all genes examined (data not shown).

Kdm3a), an H3K9 methylation/dimethylation-specific demethylase[15], in zygotes to erase H3K9 dimethylation. This involved injection of *Jhdm2a* polyadenylated messenger RNA and its inactive mutant, which contains a histidine to alanine point mutation (*Jhdm2a(H1122A)*) in the JmjC domain and does not possess histone-demethylase function (Supplementary Figs 8, 9 and Supplementary Discussion).

As expected, Jhdm2a expression abolished H3K9 methylation and dimethylation (Fig. 2c, d and Supplementary Fig. 10) but did not affect H3K9 trimethylation (Supplementary Fig. 11). Microinjection of *Jhdm2a* mRNA abolished PGC7 staining of the maternal pronuclei (Fig. 2e and Supplementary Fig. 12), resulting in the well-correlated PGC7 and H3K9me2 staining patterns under the TP condition (Supplementary Fig. 13). These data clearly indicate that the structure to which PGC7 strongly binds in the maternal pronucleus is H3K9me2-containing chromatin. Next, we examined whether DNA methylation was maintained after Jhdm2a expression. As shown in Fig. 3f, g, DNA methylation was not retained after H3K9me2 demethylation by Jhdm2a. In addition, the H3K9me2 and methylated-cytosine staining intensities were well correlated (Supplementary Fig. 14), and DNA methylation in the maternal pronucleus was not maintained at the pronuclear (PN) 5 stage without PGC7, as reported previously[8]. Embryonic development was impaired by reducing methylated H3K9, and the accessibility of the antibodies to chromatin is shown in Supplementary Figs 15 and 16 and Supplementary Discussion. Taken together,

these observations indicate that PGC7 protects methylated DNA by binding to chromatin regions containing H3K9me2.

Next, we confirmed the importance of PGC7 binding to H3K9me2-containing chromatin in a different context. DNA methylation of paternally imprinted genes such as the differentially methylated regions (DMRs) of the *Dlk1–Gtl2* domain (where *Gtl2* is also known as Meg3), *H19* and *Rasgrf1* is protected during the active demethylation in normal mouse development[16]. However, DNA methylation of *H19* and *Rasgrf1* DMRs was not maintained, whereas that of the *Dlk1–Gtl2* DMR was maintained in PN5-stage embryos derived from PGC7-null oocytes. It is reasonable to assume that epigenetic modifications of *H19* and *Rasgrf1* differ from those of *Dlk1–Gtl2*, and that epigenetic modifications residing only in the former genes are responsible for the protective function of PGC7. Although the histones in nucleosomes are replaced by protamine during spermiogenesis, a recent epigenomic analysis of human and mouse sperm revealed that histone-containing nucleosomes are preferentially retained at loci of developmental importance, including imprinted gene clusters, homeobox-containing (*Hox*) gene clusters and Nanog homeobox (*Nanog*), but not in the intergenic regions of protein kinases, cyclic-AMP-dependent catalytic β (*Prkacb*), or zona pellucida glycoprotein 4 pseudogene (*Zp4-ps*)[17,18]. Thus, we carried out a ChIP analysis of the above-mentioned paternally imprinted genes (*Dlk1–Gtl2*, *H19* and *Rasgrf1*) and three maternally imprinted genes (paternally expressed genes 1, 3 and 5 (*Peg1* (*Mest*), *Peg3*, *Peg5* (*Nnat*)),
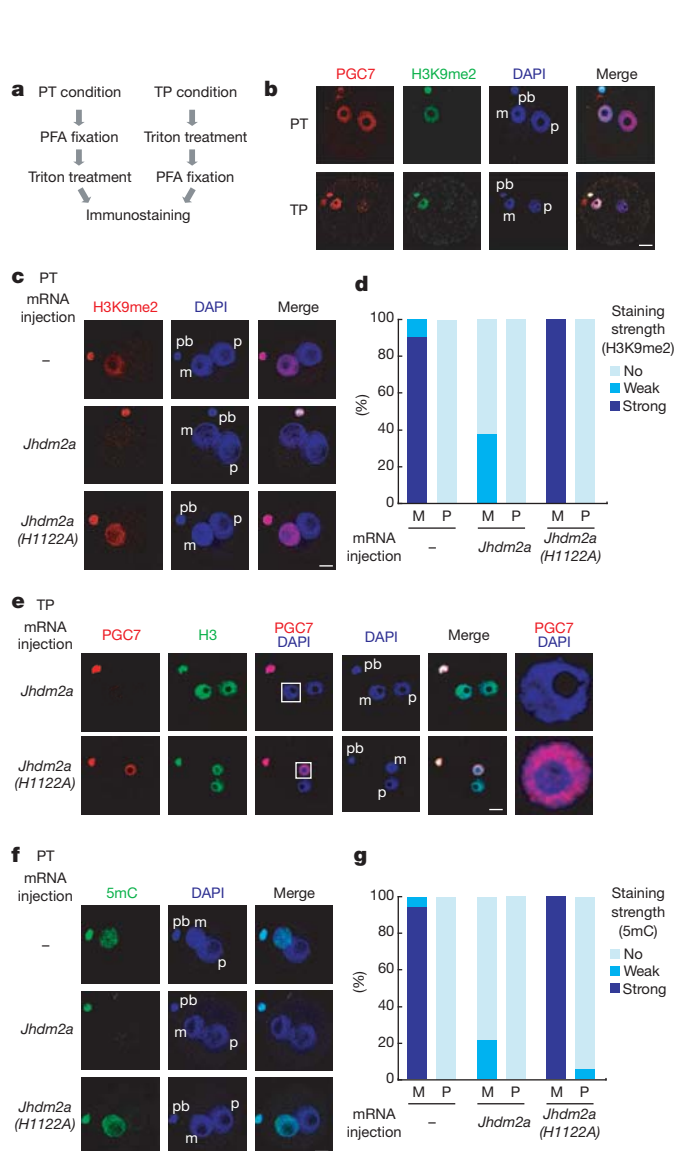
Figure 2 | Protection of the maternal genome from DNA demethylation by PGC7 through H3K9me2-containing chromatin in early embryos.
a, Schematic diagram of the two pre-treatment procedures: PT and TP conditions. **b**, After PT or TP treatment, immunostaining was performed with the indicated antibodies (m, maternal pronuclei; p, paternal pronuclei; pb, polar body). PGC7 and H3K9me2 are shown in red and green, respectively; nuclei were stained with DAPI (blue). A total of 38 and 42 zygotes were stained, under PT and TP conditions, respectively. **c, d**, H3K9 dimethylation after microinjecting *Jhdm2a* or *Jhdm2a* mRNA coding for the H1122A mutation (*Jhdm2a (H1122A)*). Zygotes were injected with *Jhdm2a* or *Jhdm2a(H1122A)* mRNA and cultured for 4.5 h in potassium-enriched simplex optimized medium (KSOM). H3K9 dimethylation was analysed using anti-H3K9me2 antibody (H3K9me2, red; DAPI, blue) (**c**). A total of 32 and 34 zygotes were injected with *Jhdm2a* and *Jhdm2a(H1122A)*, respectively. A total of 55 non-injected zygotes were analysed as controls. The strength of the H3K9me2 staining in maternal (M) and paternal (P) pronuclei was analysed (**d**). **e**, Analysis of PGC7 localization after microinjecting *Jhdm2a* or *Jhdm2a(H1122A)* mRNA. Zygotes were stained after TP fixation. PGC7 and histone H3 are shown in red and green, respectively; nuclei are stained with DAPI (blue). **f, g**, The methylation status of the parental genome after microinjecting *Jhdm2a* or *Jhdm2a(H1122A)* mRNA. Zygotes were injected with *Jhdm2a* mRNA and cultured for 4.5 h in KSOM. Demethylation of the parental genome was analysed using anti-5mC antibody (5mC, green; DAPI, blue) (**f**). A total of 33 and 34 zygotes were injected with *Jhdm2a* and *Jhdm2a(H1122A)* mRNA, respectively, and 41 non-injected zygotes were analysed as controls. The strength of 5mC staining in the maternal and paternal pronuclei was analysed (**g**). Scale bar, 20 µm.
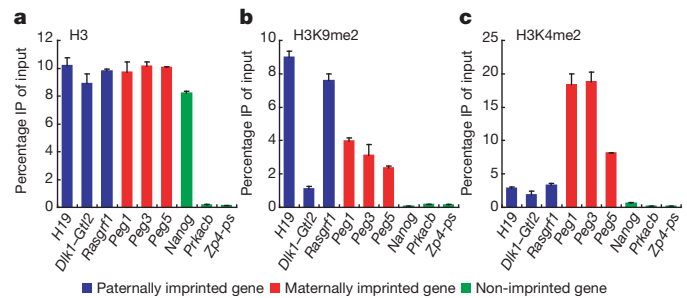
Figure 3 | Remaining H3K9me2 at the DMRs of two paternally imprinted genes, *H19* and *Rasgrf1*, in mature sperm. a–c, A ChIP–quantitative (q)PCR analysis of mature sperm was conducted using anti-H3 (**a**), -H3K9me2 (**b**) and -H3K4me2 (**c**) antibodies. *H19*, *Dlk1–Gtl2*, *Rasgrf1*, *Peg1*, *Peg3*, *Peg5*, *Nanog*, *Prkacb* and *Zp4-ps* were analysed by qPCR. The mean and s.d. (*n* = 3) of the percentage of each PCR product in the immunoprecipitated sample compared with that in the input sample is shown. Anti-mouse or -rabbit IgG was used as a negative control. The signals per total input of the negative controls were <0.16% for all genes examined (data not shown).

as well as the promoter and intergenic regions of three non-imprinted genes (*Nanog*, *Prkacb* and *Zp4-ps*) in mature mouse sperm.

Histone enrichment was detected at all DMRs of the analysed imprinted genes and at the *Nanog* promoter region, but not at the *Prkacb* and *Zp4-ps* intergenic regions (Fig. 3a) as reported for human sperm. Notably, H3K9me2 was enriched at the *H19* and *Rasgrf1* DMRs (Fig. 3b), but enrichment was not observed at the DMR of the paternally imprinted domain *Dlk1–Gtl2*. Considering that the *Nanog*-promoter region from the paternal genome was demethylated in wild-type zygotes, even in the presence of PGC7 (ref. 19), H3K9me2 should be crucial for the PGC7 protective function. These observations strongly suggest that the remaining H3K9me2 is important for protecting the *H19* and *Rasgrf1* DMRs against active DNA demethylation after fertilization. It is reasonable to conclude that this protection is mediated by PGC7 binding at the loci containing H3K9me2. H3K4me2 was enriched at all maternally imprinted genes examined but not at the paternally imprinted genes (Fig. 3c and Supplementary Discussion. Taking the dynamics of H3K9me2 into account (see Supplementary Discussion), in the maternal genome and two paternally imprinted loci, the binding of PGC7 to H3K9me2-containing chromatin is critical for the protection of DNA methylation.

We and another group found that the Tet3-mediated conversion of 5mC to 5hmC takes place in the paternal genome but not in the maternal genome[9,20,21]. We also showed that this conversion occurred in the maternal genome of zygotes derived from PGC7-null oocytes. These observations prompted us to analyse the effects of H3K9me2 on 5hmC status as an analogy to the PGC7-dependent regulation of zygotic 5mC. As shown in Fig. 4a, eradicating H3K9me2 by expressing Jhdm2a induced hydroxymethylation of the maternal genome, and the levels of H3K9me2 exhibited an inverse correlation with those of 5hmC (Supplementary Fig. 17). In addition, RanBP5–mER (RanBP5 fused to a mutated oestrogen receptor) expression, which inhibits PGC7 function by driving the subcellular protein out of the nucleus[8] (Supplementary Fig. 18), also induced maternal-genome hydroxymethylation (Fig. 4a). Although it has been reported that acetylation of H3K9 decreases in *Jhdm2a*[−/−] testes[22], this effect was negligible in our experimental system (Supplementary Fig. 19). Taken together, these results indicate that PGC7 protected 5mC from the Tet3-mediated conversion to 5hmC through binding to H3K9me2-containing chromatin, which is essentially identical to the model in which 5mC is protected from active demethylation.

Tet3 was detectable in both paternal and maternal pronuclei under PT conditions (Fig. 4b, c). However, Tet3 was detected only in paternal pronuclei under TP conditions (Fig. 4b, c), indicating that Tet3 was tightly bound to a structure that existed in the paternal but not the maternal chromatin. Furthermore, Tet3 was detectable in the maternal pronucleus after microinjecting *Jhdm2a* or *RnaBP5–mER* mRNA and
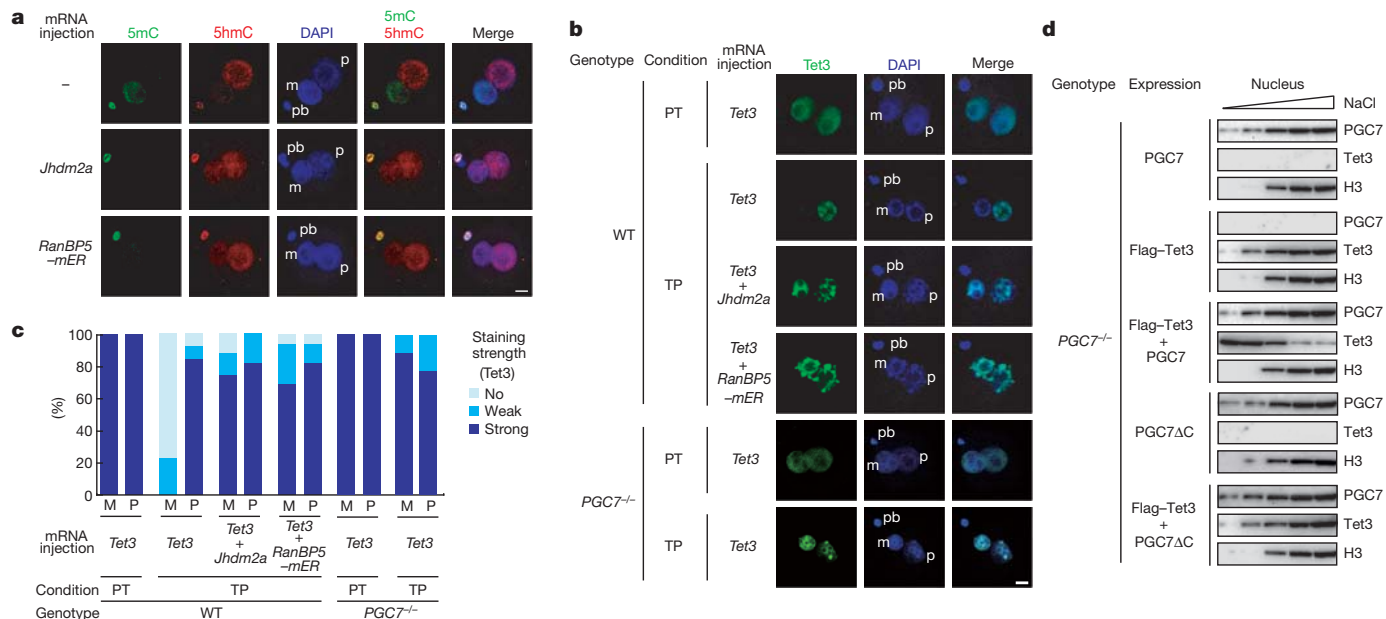
Figure 4 | Protection of the conversion of 5mC to 5hmC by PGC7 through H3K9me2-containing chromatin in early embryos. a, 5mC and 5hmC status after the microinjection of *Jhdm2a* or *RanBP5–mER* mRNA. Zygotes were injected with *Jhdm2a* or *RanBP5–mER* mRNA and cultured for 4.5 h in KSOM. The 5mC and 5hmC states were analysed using anti-5mC and anti-5hmC antibodies (5mC, green; 5hmC, red; DAPI, blue). Scale bar, 20 μm. **b, c**, Flag-tagged *Tet3* mRNA was injected into zygotes obtained from wild-type or *PGC7*⁻/⁻ female mice with or without *Jhdm2a* or *RanBP5–mER* mRNA and cultured for 4.5 h in KSOM. After PT or TP treatment as described in the Fig. 2 legend, immunostaining was performed using an anti-Flag antibody. Tet3 is shown in green; nuclei were stained with DAPI (blue). Ten zygotes obtained from wild-type female mice injected with Flag-tagged *Tet3* (Flag–*Tet3*) mRNA were stained with the anti-Flag antibody under PT conditions; 13 zygotes were injected with Flag–*Tet3* mRNA alone and 16 zygotes were injected with *Tet3* and *Jhdm2a* mRNA or with *Tet3* and *RanBP5–mER* mRNA. These embryos were stained with anti-Flag antibody under TP conditions. Zygotes obtained from *PGC7*⁻/⁻ female mice were injected with Flag–*Tet3* mRNA. A total of 6 and 9 zygotes were stained with the anti-Flag antibody under PT and TP conditions, respectively (**b**) and the percentage of Flag staining in maternal (M) and paternal (P) pronuclei is shown (**c**). Scale bar, 20 μm. **d**, Effect of PGC7 on the chromatin binding of Tet3 in ES cells. Nuclei were isolated from *PGC7*⁻/⁻ ES cells transfected with full-length PGC7, Flag–*Tet3*, PGC7 and Flag–*Tet3*, PGC7ΔC, and both PGC7ΔC and Flag–*Tet3*. These nuclei were treated with DNase I under various concentrations of NaCl (100, 200, 300, 400 and 500 mM) to separate nuclear extract from nuclear debris. Equivalent amount of aliquots were analysed by immunoblotting with anti-PGC7, anti-Flag and anti-H3 antibodies as described.

was also detectable both in maternal and paternal pronuclei of PGC7-null fertilized eggs under TP conditions (Fig. 4b, c and Supplementary Fig. 20). These data clearly indicate that the tight binding of Tet3 to chromatin was inhibited by intranuclear PGC7 through its binding to H3K9me2-containing chromatin. Although we examined direct binding of PGC7 to Tet3, no obvious binding was observed (data not shown).

To examine the effect of PGC7 on Tet3 in more detail, we carried out a stepwise salt-extraction analysis using the *PGC7*⁻/⁻ ES cells. As shown in Fig. 4d, PGC7 considerably reduced the binding of Tet3 to chromatin, which was as tight as that of H3 without PGC7. By contrast, PGC7ΔC did not show this effect despite having a chromatin-binding affinity similar to that of full-length PGC7. Moreover, the expression of PGC7ΔC did not inhibit the endonuclease activity in the MNase assay, like Tet3 binding (Supplementary Fig. 21), indicating that the inhibitory effect of PGC7 on Tet3 was not caused by competitive binding. PGC7 inhibited MNase digestion of the linker part of chromatin by binding to H3K9me2, a modified histone tail, as described in the discussion of the MNase assay of *G9a*⁻/⁻ ES cells (Supplementary Fig. 5). PGC7 (~17 kDa) is a relatively small protein compared to the histone octamer (~100 kDa) and the deleted carboxy-terminal part of PGC7 (~8 kDa), which is essential for the inhibitory effect on both Tet3 binding and MNase activity. Therefore, it is probably too small to have a marked steric effect. Meanwhile, it is conceivable that the distribution of H3K9me2, through which PGC7 binds to chromatin, is not sufficiently dense. Considering these points, although we cannot exclude it completely, the possibility of a steric effect seems improbable. Therefore, we prefer the hypothesis that PGC7 inhibits the activity of the enzyme(s) acting on DNA, such as Tet3 and MNase, by means of a change in chromatin configuration.

Although 5mC has been the only recognized DNA modification for many years in mammals, several reports have described a novel 5hmC DNA modification[2,3]. The discovery of 5hmC raises numerous questions, including its function, tissue localization and regulatory mechanisms of modification. The tight differential regulatory patterns of 5mC and 5hmC in paternal and maternal pronuclei imply that 5hmC is important[9,20,21]. Here, we showed the molecular function of PGC7 during the modification as a first step towards understanding the regulatory mechanisms of 5-hydroxymethylation in DNA. It is noteworthy that regulation of reciprocal DNA methylation and hydroxymethylation was controlled by the same protein, PGC7, through binding to H3K9me2-containing chromatin. Two recent studies have raised the possibility that 5hmC is an intermediate during DNA demethylation and that the Tet family are the critical enzymes for this process[23,24]. Our conclusion is consistent with this notion, and the inhibition of Tet3 activity through the binding of PGC7 to H3K9me2 would explain the regulatory function of PGC7 in global DNA demethylation during early embryonic development (Supplementary Fig. 22).

METHODS SUMMARY

Details of cell culture, gel-shift assay, ChIP–western blotting, sperm collection, chromatin preparation, Plasmids, global DNA methylation analysis, MNase assay, histone peptide-binding assay, stepwise salt extraction, zygote collection and culture, triton treatment of zygotes, immunohistochemistry, ChIP–quantitative PCR and sperm chromatin preparation can be found in Methods.

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

1. Feng, S., Jacobsen, S. E. & Reik, W. Epigenetic reprogramming in plant and animal development. *Science* **330,** 622–627 (2010).
2. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).

3.   Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).
4.   Koh, K. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8,** 200–213 (2011).
5.   Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473,** 398–402 (2011).
6.   Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473,** 343–348 (2011).
7.   Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473,** 389–393 (2011).
8.   Nakamura, T. *et al.* PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature Cell Biol.* **9,** 64–71 (2007).
9.   Wossidlo, M. *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Commun.* **2,** 241 (2011).
10. Santos, F., Peters, A. H., Otte, A. P., Reik, W. & Dean, W. Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Dev. Biol.* **280,** 225–236 (2005).
11. Tachibana, M. *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16,** 1779–1791 (2002).
12. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11,** 805–814 (2006).
13. Ge, Y. Z. *et al.* Chromatin targeting of *de novo* DNA methyltransferases by the PWWP domain. *J. Biol. Chem.* **279,** 25447–25454 (2004).
14. Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329,** 78–82 (2010).
15. Yamane, K. *et al.* JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125,** 483–495 (2006).
16. Olek, A. & Walter, J. The pre-implantation ontogeny of the H19 methylation imprint. *Nature Genet.* **17,** 275–276 (1997).
17. Hammoud, S. S. *et al.* Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460,** 473–478 (2009).
18. Brykczynska, U. *et al.* Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nature Struct. Mol. Biol.* **17,** 679–687 (2010).
19. Farthing, C. R. *et al.* Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* **4,** e1000116 (2008).
20. Iqbal, K., Jin, S. G., Pfeifer, G. P. & Szabo, P. E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl Acad. Sci. USA* **108,** 3642–3647 (2011).
21. Gu, T. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477,** 606–610 (2011).
22. Liu, Z. *et al.* Jmjd1a demethylase-regulated histone modification is essential for cAMP-response element modulator-refulated gene expression and spermatogenesis. *J. Biol. Chem.* **285,** 2758–2770 (2010).
23. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).
24. He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333,** 1303–1307 (2011).

Author Contributions T. Nakamura and T. Nakano conceived the project and wrote the manuscript. T. Nakamura designed and performed the all experiments and evaluated the results. Y.-J.L., H.N. and H.U. propagated *PGC7*$^{-/-}$ mice. K.I., S.M. and A.O. performed some of the experiments. M.T. and Y.S. provided *G9a*$^{-/-}$ ES cells and related materials.

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 T.Nakamura (tnakamura@nagahama-i-bio.ac.jp) or T.Nakano (tnakano@patho.med.osaka-u.ac.jp).

METHODS

Cell culture. $G9a^{-/-}$, $Dnmt1^{-/-}$ $Dnmt3a^{-/-}$ $Dnmt3b^{-/-}$ and $PGC7^{-/-}$ ES cells and their clones were used. $PGC7^{-/-}$ ES cells were derived from blastocysts. These cells were maintained as described previously[8].

Gel-shift assay. Purified mononucleosomes were incubated with various amounts of purified histidine-tagged PGC7 (His–PGC7) on ice for 30 min in 50 mM NaCl, 20 mM Tris buffer–HCl (pH 7.5), 2 mM EDTA buffer, 5 mg ml^{-1} BSA buffer and 5% glycerol. The samples were electrophoresed on a 1% native agarose gel and the chromatin was visualized using ethidium bromide.

ChIP–western blotting. Mononucleosomes prepared from $PGC7^{-/-}$ ES cells and a cell line stably expressing PGC7 were incubated with anti-PGC7 antibody for 3 h at 4 °C. After adding protein G-Sepharose 6 Fast Flow (GE Healthcare), the samples were incubated for another hour at 4 °C, washed and eluted from the beads using SDS sample buffer. The immunoprecipitates were analysed by immunoblotting as described previously[8].

Sperm collection. Sperm was obtained from ICR mice aged 10–20 weeks. The cauda epididymides were partially cut open in human tubal fluid (Millipore) and incubated for 1 h at 37 °C in 5% CO_2 to allow the sperm to swim out. The sperm pellet was washed with PBS buffer and ChIP–quantitative (q)PCR analysis was performed as described in the Supplementary Information.

Chromatin preparation. Cell pellets were re-suspended in buffer I (300 mM sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA buffer, 15 mM Tris–HCl (pH 7.5), 0.4% tergitol-type NP-40 (NP-40) and 0.5 mM DTT), and an equal volume of buffer II (300 mM sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 15 mM Tris–HCl (pH 7.5) and 0.5 mM dithiothreitol (DTT)) was added. After incubation on ice for 10 min, the cell suspensions were layered over buffer III (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 15 mM Tris–HCl (pH 7.5) and 0.5 mM DTT) and the nuclear pellets were collected by centrifugation (10,000g, 4 °C, 20 min). To avoid carry over of NP-40, the supernatant was carefully removed at least three times using a new Pasteur pipette at each time. Nuclear pellets were washed in MNase digestion buffer (320 mM sucrose, 50 mM Tris–HCl (pH 7.5), 4 mM $MgCl_2$ and 1 mM $CaCl_2$) and then re-suspended in MNase digestion buffer. Chromatin was released from the nuclear preparations by digestion with 20 U ml^{-1} MNase (Takara) at 37 °C for 9 min. Digestion was stopped by adding 0.2 mM EDTA to a final concentration of 5 mM on ice. Chromatin preparations were analysed by native agarose gel electrophoresis (Supplementary Fig. 1) and the mononucleosome was used immediately for the gel-shift assay.

Plasmids. The $Jhdm2a$ complementary DNA was cloned into pcDNA4mycHisA (Invitrogen). The $Jhdm2a(H1122A)$ mutant was generated by PCR-based mutagenesis and confirmed by sequencing. The primers used to generate the mutant are described in Supplementary Table 2.

Global DNA methylation analysis. A previous protocol for global DNA methylation analysis[25] was used with slight modifications. Genomic DNA was isolated from various ES cells with proteinase K and RNase A, followed by phenol/chloroform extraction and ethanol precipitation. A 2-µg aliquot of genomic DNA was digested with 50 U of methylation-sensitive HpaII or the methylation-insensitive isoschisomer MspI for 16–18 h at 37 °C. The digested genomes were purified by phenol/chloroform extraction and ethanol precipitation, and 250 ng of purified DNA was labelled with ^{3}H-deoxycytidine triphosphate (dCTP) at 56 °C for 1 h using a single-nucleotide extension reaction. Undigested genomic DNA served as a background control. The samples were applied to DE-81 ion-exchange filters and washed three times with 0.5 M Na_3PO_4 buffer (pH 7.0) at room temperature. The filters were then dried and processed for scintillation counting.

MNase assay. Living cells were permeabilized on ice with 0.02% L-α-lysolecithin (Sigma) in 150 mM sucrose, 35 mM HEPES–NaOH (pH 7.4), 5 mM $KHPO_4$, 5 mM $MgCl_2$ and 0.5 mM $CaCl_2$ for 90 s, followed by digestion with 3 U ml^{-1} MNase (Takara) in 150 mM sucrose, 50 mM Tris–HCl (pH 7.5), 50 mM NaCl and 2 mM $CaCl_2$ at room temperature for 0, 2, 4, 6, 8, 10, 30 and 60 min. Digestion was stopped by adding EDTA to a final concentration 5 mM, on ice. DNA was purified by phenol/chloroform extraction and electrophoresed on 0.8% native agarose gels.

Histone peptide-binding assay. Biotinylated histone peptides were purchased from Upstate Biotechnology. In brief, biotinylated histone peptides (0.5 µg) were incubated with streptavidin beads (GE Healthcare) in binding buffer (50 mM Tris–HCl (pH 8.0), 300 mM NaCl and 0.1% NP-40) for 1 h at 4 °C with rotation. After extensive washing, the beads were incubated with 1 µg recombinant-purified His–PGC7 for 2 h at 4 °C with rotation. 10- and 100-fold unbiotinylated histone peptides were added for the competitive-binding assay. After extensive washing, bound PGC7 protein was analysed by SDS–PAGE and immunoblotting with anti-PGC7 antibody.

Stepwise salt extraction. ES cells were suspended in nuclear isolation buffer (10 mM Tris–HCl (pH 7.5), 60 mM KCl, 15 mM NaCl, 1 mM DTT, 1.5 mM $MgCl_2$, 1 mM $CaCl_2$, 250 mM sucrose, 10% glycerol, 1 mM DTT and 0.15% NP-40) on ice for 10 min. Nuclear pellets were treated with 100 U ml^{-1} DNase I (Takara) in nuclear isolation buffer with increasing amounts of NaCl (100, 200, 300, 400 and 500 mM) at 25 °C for 20 min and on ice for 10 min. After the incubation, EDTA was added to a final concentration of 5 mM, and the sample was incubated on ice for 10 min. Nuclear extracts were separated from pellets by centrifugation.

Zygote collection and culture. Female B6D2F1 mice >8 weeks old were super-ovulated by injecting 5 U of human chorionic gonadotropin 48 h after injecting 5 U of pregnant mare serum gonadotropin, and then mated with male mice. Fertilized eggs were collected from the oviduct, placed in 100-µl drops of KSOM (Millipore) and cultured at 37 °C in an atmosphere of 5% CO_2.

Triton treatment of zygotes. Zygotes were treated with Triton X-100, similar to a previous report with minor modifications[14]. Zygotes were treated with 0.2% Triton X-100 in PBS for 45–60 s until the perivitelline space was eliminated. Immediately after Triton treatment, the zygotes were washed with PBS at least five times and then fixed in 4% PFA. After washing with PBS, immunostaining was performed as described below.

Immunohistochemistry. Fertilized eggs were washed with PBS, fixed for 15 min in 4% PFA in PBS at room temperature and permeabilized with 0.2% Triton X-100 in PBS for 20 min at room temperature. The eggs were blocked for 1 h in 5% normal goat serum in PBS at room temperature and incubated overnight at 4 °C with primary antibodies as shown in Supplementary Table 1. The following day, the eggs were washed three times with 0.05% Tween20 in PBS and staining was detected by incubating the eggs with secondary antibodies as shown in Supplementary Table 1. Nuclei were stained with 1 µg ml^{-1} DAPI. 5mC, 5hmC and DNA staining was performed as described previously[8]. Immunofluorescence was visualized using an LSM510 confocal laser scanning microscope (Carl Zeiss).

ChIP–qPCR. $PGC7^{-/-}$, $PGC7^{-/-}$–PGC7, $G9a^{-/-}$ and $G9a^{-/-}$–PGC7 ES cells (3×10^7) were treated with 1% PFA for 8 min at room temperature. After quenching the PFA crosslinking reaction with 200 mM glycine, the fixed cells were washed with PBS. The cells were suspended in radio immunoprecipitation assay (RIPA) buffer (50 mM Tris–HCl (pH 8) 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.5% deoxycholate and 0.1% SDS) and sonicated to an average fragment size of 200–1000 bp. Solubilized chromatin was clarified by centrifugation for 10 min at 15,000 r.p.m. and 4 °C. The supernatant was pre-cleared with protein G-Sepharose beads, which were pre-blocked with salmon sperm DNA and BSA, at 4 °C for 1 h. The pre-cleared chromatin was incubated with anti-PGC7, anti-H3 and anti-H3K9me2 antibodies for 14–18 h at 4 °C. Immune complexes were bound to pre-blocked protein G-Sepharose beads for 2 h at 4 °C. The beads were washed with RIPA buffer, high-salt wash buffer (20 mM Tris–HCl (pH 8), 500 mM NaCl, 1 mM EDTA, 1% NP-40, 0.5% deoxycholate and 0.1% SDS), LiCl wash buffer (250 mM LiCl, 20 mM Tris–HCl (pH 8), 1 mM EDTA, 1% NP-40 and 0.5% deoxycholic acid (DOC)), and once with Tris–EDTA. Immune complexes bound to protein G beads were suspended in elution buffer (20 mM Tris–HCl (pH 8), 300 mM NaCl, 1 mM EDTA and 0.5% SDS) and incubated for 6 h at 65 °C. After incubation, the samples were treated with 30 µg ml^{-1} RNase A for 1 h at 37 °C, and 100 µg ml^{-1} proteinase K for 8 h at 56 °C. DNA was extracted with phenol/chloroform and precipitated with ethanol plus Dr.GenTLE (Takara) as a carrier. Precipitated DNA was re-suspended in 40 µl of water and analysed by qPCR using the specific primers shown in Supplementary Table 2.

Sperm chromatin preparation. Sperm pellets were washed with PBS, suspended in lysis buffer (0.1% SDS, 0.5% Triton X-100 in PBS) and incubated on ice for 20 min. After centrifugation, the sperm pellets were washed with PBS, suspended in 0.05% L-α-lysolecithin (Sigma) in PBS and incubated on ice for 15 min. The sperm pellets were collected by centrifugation, suspended in 10 mM DTT in PBS and incubated on ice for 10 min. After centrifugation, sperm pellets were fixed with 1% PFA in PBS for 8 min at room temperature. Next, 2.5 M glycine was added to a final concentration of 0.2 M, and incubation was continued for an additional 10 min at room temperature. After washing with PBS, the pellet was suspended in RIPA buffer (20 mM HEPES–NaOH (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.5% DOC and 0.1% SDS) and disrupted by sonication using a Bioruptor (Diagenode). ChIP–qPCR was conducted as described above using the specific primers shown in Supplementary Table 2.

25. Pogribny, I., Yi, P. & James, S. A sensitive new method for rapid detection of abnormal methylation patterns in global DNA and within CpG islands. *Biochem. Biophys. Res. Commun.* **262,** 624–628 (1999).

LETTER

Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets

Masaki Imai[1], Tokiko Watanabe[1,2], Masato Hatta[1], Subash C. Das[1], Makoto Ozawa[1,3], Kyoko Shinya[4], Gongxun Zhong[1], Anthony Hanson[1], Hiroaki Katsura[5], Shinji Watanabe[1,2], Chengjun Li[1], Eiryo Kawakami[2], Shinya Yamada[5], Maki Kiso[5], Yasuo Suzuki[6], Eileen A. Maher[1], Gabriele Neumann[1] & Yoshihiro Kawaoka[1,2,3,5]

Highly pathogenic avian H5N1 influenza A viruses occasionally infect humans, but currently do not transmit efficiently among humans. The viral haemagglutinin (HA) protein is a known host-range determinant as it mediates virus binding to host-specific cellular receptors[1–3]. Here we assess the molecular changes in HA that would allow a virus possessing subtype H5 HA to be transmissible among mammals. We identified a reassortant H5 HA/H1N1 virus—comprising H5 HA (from an H5N1 virus) with four mutations and the remaining seven gene segments from a 2009 pandemic H1N1 virus—that was capable of droplet transmission in a ferret model. The transmissible H5 reassortant virus preferentially recognized human-type receptors, replicated efficiently in ferrets, caused lung lesions and weight loss, but was not highly pathogenic and did not cause mortality. These results indicate that H5 HA can convert to an HA that supports efficient viral transmission in mammals; however, we do not know whether the four mutations in the H5 HA identified here would render a wholly avian H5N1 virus transmissible. The genetic origin of the remaining seven viral gene segments may also critically contribute to transmissibility in mammals. Nevertheless, as H5N1 viruses continue to evolve and infect humans, receptor-binding variants of H5N1 viruses with pandemic potential, including avian–human reassortant viruses as tested here, may emerge. Our findings emphasize the need to prepare for potential pandemics caused by influenza viruses possessing H5 HA, and will help individuals conducting surveillance in regions with circulating H5N1 viruses to recognize key residues that predict the pandemic potential of isolates, which will inform the development, production and distribution of effective countermeasures.

Although H5N1 viruses continue to cause outbreaks in poultry and there are cases of human infection in Indonesia, Vietnam, Egypt and elsewhere (http://www.who.int/influenza/human_animal_interface/H5N1_cumulative_table_archives/en/index.html), they have not acquired the ability to cause human-to-human transmission. Investment in H5N1 vaccines has therefore been questioned. However, because humans lack immunity to influenza viruses possessing an H5 HA, the emergence of a transmissible H5-HA-possessing virus would probably cause a pandemic. To prepare better for such a scenario, it is critical that we understand the molecular changes that may render H5-HA-possessing viruses transmissible in mammals. Such knowledge would allow us to monitor circulating or newly emerging variants for their pandemic potential, focus eradication efforts on viruses that already have acquired subsets of molecular changes critical for transmission in mammals, stockpile antiviral compounds in regions where such viruses circulate, and initiate vaccine generation and large-scale production

before a pandemic. Therefore, we studied the molecular features that would render H5-HA-possessing viruses transmissible in mammals.

Previous studies suggested that HA has a major role in host-range restriction of influenza A viruses[1–3]. The HA of human isolates preferentially recognizes sialic acid linked to galactose by α2,6-linkages (Siaα2,6Gal), whereas the HA of avian isolates preferentially recognizes sialic acid linked to galactose by α2,3-linkages (Siaα2,3Gal)[3]. A small number of avian H5N1 viruses isolated from humans show limited binding to human-type receptors, a property conferred by several amino acid changes in HA[4–9]. None of the H5N1 viruses tested transmitted efficiently in a ferret model[10–13], although, while our paper was under review, one study[14] reported that a virus with a mutant H5 HA and a neuraminidase (NA) of a human virus in the H5N1 virus background caused respiratory droplet transmission in one of two contact ferrets.

To identify novel mutations in avian H5 HAs that confer human-type receptor-binding preference, we introduced random mutations into the globular head (amino acids 120–259 (H3 numbering), which includes the receptor-binding pocket) of A/Vietnam/1203/2004 (H5N1; VN1203) HA (Supplementary Fig. 1). Although this virus was isolated from a human, its HA retains avian-type receptor-binding properties[6,15]. We also replaced the multibasic HA cleavage sequence with a non-virulent-type cleavage sequence, allowing us to perform studies in biosafety level 2 containment (http://www.who.int/csr/resources/publications/influenza/influenzaRMD2003_5.pdf). The mutated polymerase chain reaction (PCR) products were cloned into RNA polymerase I plasmids[16] containing the VN1203 HA complementary DNA, which resulted in *Escherichia coli* libraries representing the randomly generated HA variants. Sequence analysis of 48 randomly selected clones indicated an average of 1.0 amino acid changes per HA globular head (data not shown). To generate an H5N1 virus library, plasmids for the synthesis of the mutated HA gene and the unmodified NA gene of VN1203 were transfected into human embryonic kidney (293T) cells together with plasmids for the synthesis of the six remaining viral genes of A/Puerto Rico/8/34 (H1N1; PR8), a laboratory-adapted human influenza A virus.

Turkey red blood cells (TRBCs; which possess both Siaα2,6Gal and Siaα2,3Gal on their surface (data not shown)) were treated with *Salmonella enterica* serovar Typhimurium LT2 sialidase, which preferentially removes α2,3-linked sialic acid (that is, avian-type receptors), creating TRBCs that predominantly possess Siaα2,6Gal on the cell surface (Siaα2,6-TRBCs; Supplementary Fig. 2). The virus library was then adsorbed to Siaα2,6-TRBCs at 4 °C and extensively washed to remove nonspecifically or weakly bound viruses. Bound viruses were eluted by incubation at 37 °C for 30 min, and then diluted to approximately ~0.5 viruses per well (on the basis of a pilot experiment that

[1]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, Wisconsin 53711, USA. [2]ERATO Infection-Induced Host Responses Project, Saitama 332-0012, Japan. [3]Department of Special Pathogens, International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. [4]Department of Microbiology and Infectious Diseases, Kobe University, Hyogo 650-0017, Japan. [5]Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. [6]Health Science Hills, College of Life and Health Sciences, Chubu University, Kasugai, Aichi 487-8501, Japan.

assessed the approximate number of eluted viruses). We screened one-third of the library (that is, 2.1×10^6 viruses) in three separate selection experiments (that is, 0.7×10^6 viruses per experiment) and isolated 370 viruses that bound to Siaα2,6-TRBCs (Supplementary Fig. 2). Individual viruses were then grown in Madin-Darby canine kidney (MDCK) cells modified to overexpress Siaα2,6Gal (AX4 cells[17]), and screened again for their ability to agglutinate Siaα2,6-TRBCs (Supplementary Fig. 2). The parental control virus (designated VN1203/PR8) with avian-type receptor-binding specificity agglutinated untreated TRBCs (which possess both human- and avian-type receptors on their surface), but not TRBCs possessing predominantly human-type receptors (Siaα2,6-TRBCs; Supplementary Table 1). By contrast, of the 370 viruses originally isolated, nine agglutinated Siaα2,6-TRBCs, albeit with different efficiencies (Supplementary Table 1). All nine viruses possessed mutations in the region targeted for random mutagenesis; one mutant also possessed an additional mutation (E119G) in an area that was not targeted for mutation. Most of the mutations clustered around the receptor-binding pocket (Fig. 1a). Several of the selected viruses possessed mutations known to increase binding to human-type receptors, including N186K (ref. 9), S227N (ref. 5) and Q226L (which confers human-type receptor binding together with G228S)[15] (all shown in blue in Fig. 1a). The identification of known determinants of human-type receptor-binding specificity from a library of random mutants validates our approach. Notably, our screen also identified mutations not previously associated with receptor-binding specificity.

Although viruses were diluted to ~0.5 viruses per well for amplification in AX4 cells, we cannot exclude the possibility that some wells were infected with more than one virus, resulting in mixed populations. To confirm the significance of the identified mutations in HA for human-type receptor binding, the mutations were engineered into a VN1203/PR8 virus (possessing an avirulent HA cleavage site sequence, as described earlier). All nine mutants were generated; however, after two passages in MDCK cells, the S136N mutation reverted to the wild-type sequence. This mutant was excluded from further evaluation.

First, we confirmed the binding of the remaining eight variants to Siaα2,6-TRBCs (Supplementary Table 1). For comparison, we included a VN1203/PR8 virus with two changes in its HA (Q226L and G228S) previously shown to have increased binding to Siaα2,6Gal[6,15]. Indeed, compared to the wild-type VN1203/PR8 virus, the Q226L/G228S mutant displayed an increased ability to bind to human-type receptors. For the recreated variants, haemagglutination titres were higher and slightly different from the initial characterization, which we attribute to biological differences (the initial characterization was carried out with non-concentrated cell culture supernatant and potentially mixed virus populations, whereas the recreated viruses were concentrated and purified) and to experimental differences (that is, differences between the TRBC batches or the efficiency of α2,3-sialidase treatment, or both). Collectively, however, these experiments demonstrate that this random mutagenesis approach allows the identification of hitherto unrecognized amino acid substitutions that permit avian virus HAs to bind to human-type receptors.

To characterize further the receptor-binding properties of the selected variants, we used solid-phase binding assays in which sialylglycopolymers were absorbed to plates, which were then incubated with virus (Fig. 2a). A virus possessing the HA and NA genes of the seasonal human A/Kawasaki/173/2001 (H1N1; K173) virus and the remaining genes from PR8 (K173/PR8) served as a control virus with typical human-type receptor specificity. Indeed, K173/PR8 preferentially bound to Siaα2,6Gal. In contrast, VN1203/PR8 bound to only Siaα2,3Gal. As reported elsewhere[6,15], the Q226L/G228S mutations led to increased binding to Siaα2,6Gal. Variants I202T/R220S, W153R/T160I, N169I/H184L/I217M and H130Q/K157E resembled VN1203/PR8 in their binding to glycans, despite the fact that these mutants weakly agglutinated Siaα2,6-TRBCs (see Supplementary Table 1). These viruses may have bound to glycans on TRBCs that were different from Siaα2,6Galβ1,4GlcNAc used in this study. However, variants N186K/M230I, S227N/G228A and Q226L/E231G showed an appreciable increase in binding to Siaα2,6Gal but also retained binding capacity for Siaα2,3Gal. Of all of the variants tested, only E119G/V152I/N224K/Q226L exhibited specificity for only Siaα2,6Gal. Thus, only one H5 HA variant with receptor-binding capability akin to that of seasonal influenza viruses was isolated from the library screen of 2.1×10^6 viruses. To identify the amino acid change(s) responsible for the conversion from Siaα2,3Gal to Siaα2,6Gal recognition in the E119G/V152I/N224K/Q226L virus HA, we tested the amino acid changes at positions 119, 152, 224 and 226 individually and in various combinations. Solid-phase binding assays demonstrated that the N224K/Q226L combination is critical for the shift from Siaα2,3Gal to Siaα2,6Gal recognition (Fig. 2b); Q226L in combination with V152I also conferred weak binding to α2,6-glycans.
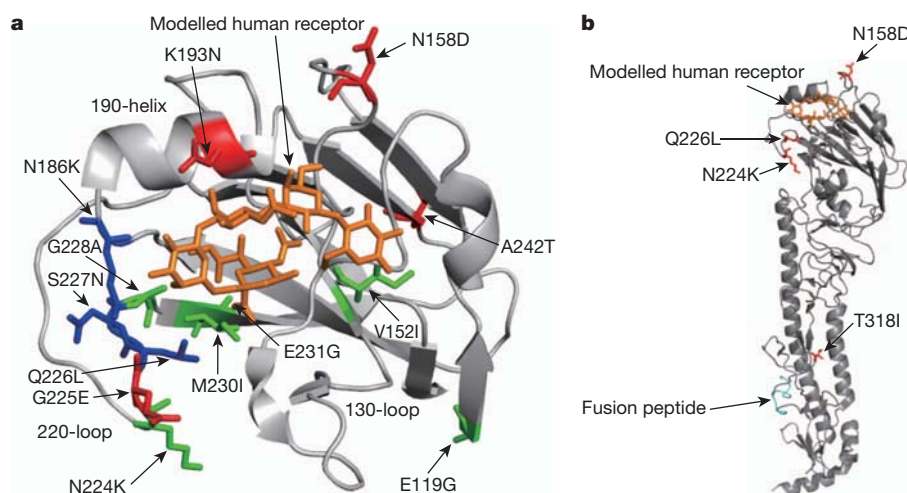
Figure 1 | Localization of amino acid changes identified in this study on the three-dimensional structure of the monomer of VN1203 HA (Protein Data Bank accession 2FK0)[15]. a, Close-up view of the globular head of VN1203 HA. Mutations known to increase affinity to human-type receptors are shown in blue. Amino acid changes not previously known to affect receptor binding are shown in green. Additional mutations that occurred in the HA of H5 avian–human reassortant viruses during replication and/or transmission in ferrets are shown in red. b, The positions of four mutations in the HA of H5 transmissible reassortant mutant virus, HA(N158D/N224K/Q226L/T318I)/CA04, are highlighted in red. The fusion peptide of HA is shown in cyan. All mutations are shown with H3 numbering. Images were created with MacPymol (http://www.pymol.org/).
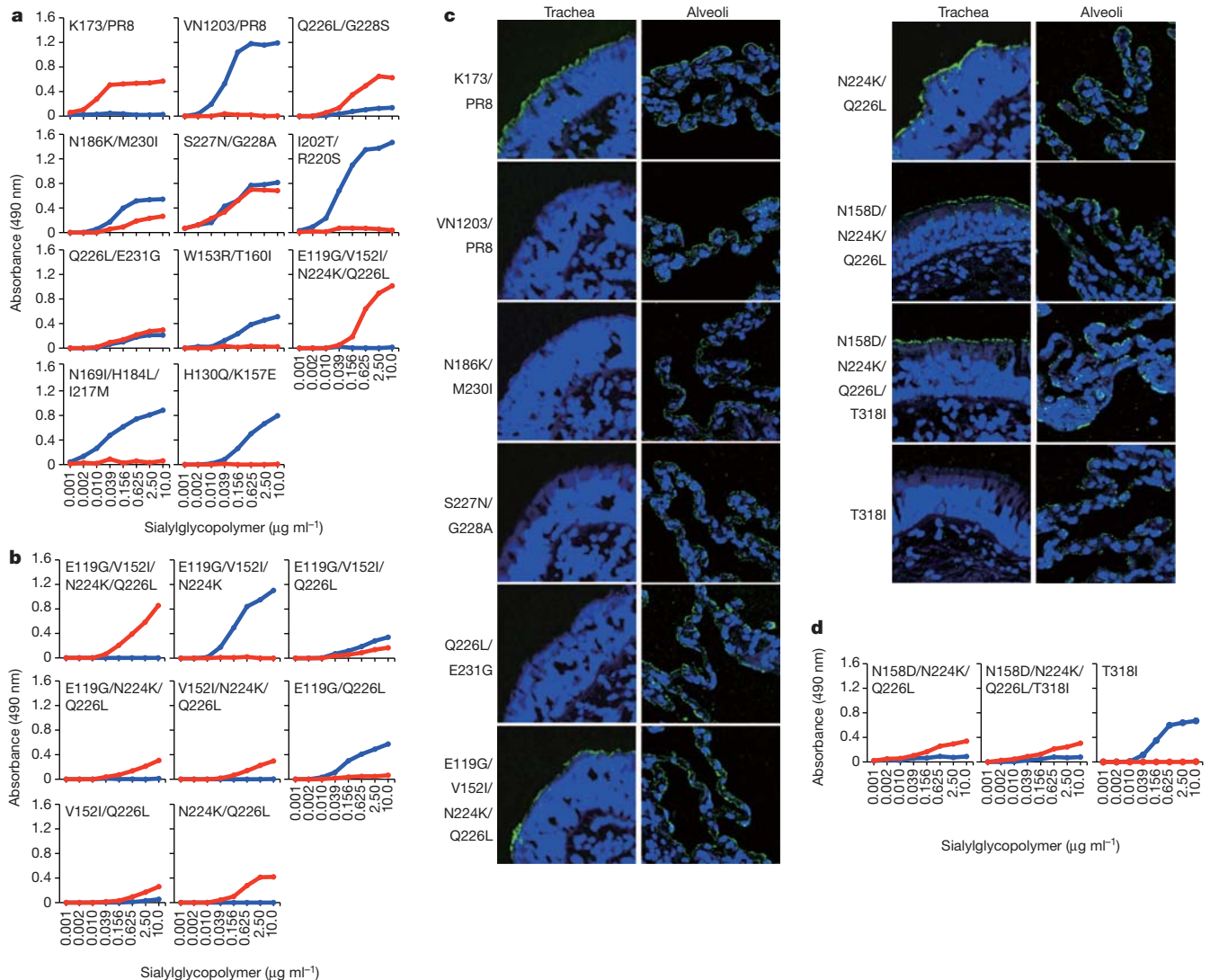
Figure 2 | Characterization of the receptor-binding properties of isolated viruses. a, Binding of VN1203 mutants to sialylglycopolymers in solid-phase binding assays. A human virus (K173/PR8), an avian virus (VN1203/PR8) and mutant VN1203/PR8 viruses were compared for their ability to bind to sialylglycopolymers containing either α2,3-linked (blue) or α2,6-linked (red) sialic acids. **b**, Identification of mutations that confer binding to human-type receptors. **c**, Binding of VN1203 mutant viruses to human respiratory tissues. K173/PR8, VN1203/PR8 and mutant VN1203/PR8 viruses were incubated with human tissue sections and then stained with either anti-K173 antiserum (green) or anti-VN1203 HA antibodies (green). All sections were subsequently incubated with labelled secondary antibodies and Hoechst dye (blue). **d**, Characterization of the receptor-binding properties of N158D/N224K/Q226L, N158D/N224K/Q226L/T318I and T318I viruses. The direct binding of virus to sialylglycopolymers containing either α2,3-linked (blue) or α2,6-linked (red) sialic acids was determined as described in panel **a**.

To assess the effect of enhanced α2,6-glycan recognition on the attachment of viruses to human respiratory tracts, sections of tracheal and lung tissues were exposed to K173/PR8 (human-type receptor binder), VN1203/PR8 (avian-type receptor binder) and mutant VN1203/PR8 viruses (Fig. 2c). Because the N186K/M230I, S227N/G228A, Q226L/E231G, E119G/V152I/N224K/Q226L and N224K/Q226L mutants exhibited appreciable binding to Siaα2,6Gal (Fig. 2a, b), the attachment of these mutants was also tested. On tracheal sections, the K173/PR8 virus bound extensively to ciliated epithelial cells (Fig. 2c and Supplementary Fig. 3), whereas the VN1203/PR8 virus bound poorly. By contrast, on lung sections, both viruses bound extensively to the alveolar epithelial surface (both type I and II pneumocytes; Fig. 2c and Supplementary Fig. 4). The binding patterns of these viruses correlate with the distribution of Siaα2,3Gal (that is, avian-type receptors; present in lung epithelia) and Siaα2,6Gal (that is, human-type receptors; present in both trachea and lung epithelia) on the tissues, as observed with lectin staining[18] (Supplementary Fig. 5). Like the human K173/PR8 virus, the E119G/V152I/N224K/Q226L and N224K/Q226L

mutants exhibited strong binding to the ciliated epithelial cells of the trachea (Fig. 2c and Supplementary Fig. 3). By contrast, the N186K/M230I, S227N/G228A and Q226L/E231G mutants displayed little-to-no binding to tracheal epithelia (Fig. 2c), despite their binding to Siaα2,6Gal (Fig. 2a). A number of sialylated oligosaccharides with differing branching patterns and chain lengths are thought to be present on the cell surface[19]. We therefore speculate that the mutants can recognize a short glycan structure such as Siaα2,6Galβ1,4GlcNAc, but may not recognize longer, more complex glycan structures, which are possibly required for binding to human tracheal epithelium. On the other hand, all mutants bound to alveolar epithelial cells (both type I and II pneumocytes; Fig. 2c and Supplementary Fig. 4). When the tissue sections were pre-treated with *Arthrobacter ureafaciens* sialidase (which cleaves all non-reducing terminally branched and unbranched sialic acids), virus binding to the tissues was substantially reduced (Supplementary Fig. 6a–c), confirming the sialic acid binding specificity of the virus. These data indicate that alterations in the receptor specificity of the E119G/V152I/N224K/Q226L and N224K/Q226L

mutants have profound effects on virus attachment to human respiratory epithelium.

In an avian H3 HA, the Q226L mutation changed the binding preference from avian- to human-type[20]. A previous study found that the Q226L mutation on an H5 HA does not confer efficient binding to α2,6-glycans in a glycan array[15]; however, when tested in combination with G228S, increased binding to human-type receptors, but not a complete switch from avian- to human-type receptor-binding specificity, was observed[15]. By contrast, here we found that Q226L in combination with N224K resulted in a switch from Siaα2,3Gal to Siaα2,6Gal binding in an H5 HA and allowed virus binding to human tracheal epithelia (Fig. 2c). The receptor-binding domain of HA is formed by the 190-helix at the top of HA, the 220-loop at the edge of the globular head, and the 130-loop at the other edge of the globular head (Fig. 1a). Crystal structure analysis revealed that the 220-loop of avian H5 HA is closer to the opposing 130-loop than in human H3 HA, indicating that a wider binding site for human H3 HA, compared to that of avian H5 HA, may be required to optimize contacts with the larger Siaα2,6-glycans[21]. N224 lies on the turn leading into the 220-loop, adjacent to position 226 (Fig. 1a). Replacement of N224 may alter the orientation of the 220-loop and thus optimize contacts between L226 and Siaα2,6Gal-containing receptors, thereby increasing the preference for α2,6 linkages.

Recent studies reported that 2009 pandemic H1N1 and H5N1 viruses show high genetic compatibility[22,23]. These two viruses have been isolated from pigs[24-28], which have been considered as 'mixing vessels' for the reassortment of avian, swine and human strains. Thus, the coexistence of H5N1 and 2009 pandemic H1N1 viruses could provide an opportunity for the generation of transmissible H5 avian–human reassortants in mammals. Therefore, we generated reassortant viruses possessing the mutant VN1203 HAs generated above, and the seven remaining gene segments from a prototype 2009 pandemic H1N1 virus (A/California/04/2009, CA04). Experiments with viruses possessing the wild-type HA cleavage site were performed in enhanced biosafety level 3 (BSL3+) containment laboratories approved for such use by the Centers for Disease Control and Prevention (CDC) and the United States Department of Agriculture (USDA). Because efficient human-to-human transmission is a critical feature of pandemic influenza viruses, we examined the growth and transmissibility of reassortant viruses in ferrets, which are widely accepted as an animal model for influenza virus transmissibility and pathogenesis studies. Because the E119G/V152I/N224K/Q226L and N224K/Q226L variants bound extensively to human tracheal epithelia (Fig. 2c), we generated by reverse genetics (rg) three H5 reassortant viruses possessing the VN1203 HA or mutant HAs (all with the wild-type multibasic cleavage site) and the remaining genes from the CA04 virus. The VN1203 HA mutants tested included the one containing four mutations, E119G, V152I, N224K and Q226L (designated rg(E119G/V152I/N224K/Q226L)/CA04), and another containing two mutations, N224K and Q226L (designated rg(N224K/Q226L)/CA04).

To determine whether the introduced HA mutations affected the replication of the H5 reassortant viruses, six ferrets were inoculated intranasally with 10^6 plaque-forming units (p.f.u.) of virus. On day 3 after infection, a recombinant virus whose genes all came from CA04, rgCA04, replicated efficiently in the respiratory organs of infected animals, and was isolated from the colon, but not from any other organs tested (Fig. 3 and Supplementary Table 2). A virus possessing H5 VN1203 HA and the remaining genes from CA04 (designated rgVN1203/CA04) replicated to titres comparable to those of rgCA04 in nasal turbinates, but substantially less in the lungs. By contrast, the two H5 reassortant viruses with HA mutations (rg(E119G/V152I/N224K/Q226L)/CA04 and rg(N224K/Q226L)/CA04) were severely limited in their replicative ability in trachea. Although virus titres in nasal turbinates and lung were not statistically different between rg(N224K/Q226L)/CA04 and rgCA04, the virus titre in nasal turbinates was significantly lower in animals inoculated with rg(E119G/
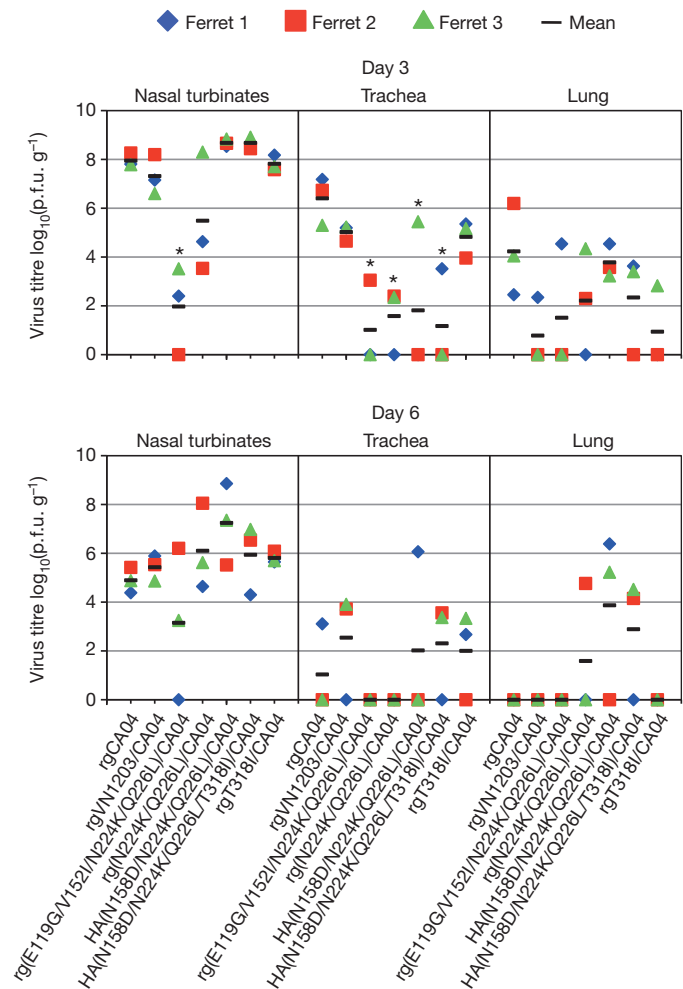
Figure 3 | Virus replication in respiratory organs. Ferrets were infected intranasally with 10^6 p.f.u. of virus. Three ferrets per group were killed on days 3 and 6 after infection for virus titration. Virus titres in nasal turbinates, trachea and lung were determined by use of a plaque assay on MDCK cells. Horizontal bars show the mean. Asterisks indicate virus titres significantly different from that of rgCA04 (Dunnett's test; $P < 0.05$).

V152I/N224K/Q226L)/CA04 than in animals inoculated with rgCA04 (Dunnett's test; $P = 0.0002$; Fig. 3). Notably, rgVN1203/CA04 (avian-type receptor binder) replicated efficiently in nasal turbinates of ferrets, which have a similar sialic acid receptor distribution pattern to that of the human respiratory tract[29,30]. The reason for this discrepancy is unclear; however, replication of avian H5N1 viruses in ferret nasal turbinates has been reported[12,13].

Although virus titres in respiratory organs were generally lower on day 6 after infection than on day 3 after infection, rg(N224K/Q226L)/CA04 still showed high levels of replication at day 6 after infection; titres in nasal turbinates ranged from $10^{4.6}$ to $10^{8.1}$ p.f.u. g^{-1} (Fig. 3). Sequence analysis of viruses in nasal turbinates on day 6 after infection revealed that viruses in ferret 2 and ferret 3 possessed N158D and N158K mutations in their HA (in addition to the original two mutations), respectively, leading to the loss of the glycosylation site at position 158 (that is, 158N-S-T to 158D-S-T or 158K-S-T; Fig. 1a and Supplementary Table 3). In nasal turbinates on day 6 after infection, the titre of the virus with the N158D/N224K/Q226L mutations ($10^{8.1}$ p.f.u. g^{-1}; see Fig. 3, ferret 2 of rg(N224K/Q226L)/CA04) was approximately four orders of magnitude higher than that of the original rg(N224K/Q226L)/CA04 ($10^{4.6}$ p.f.u. g^{-1}; Fig. 3, ferret 1 of rg(N224K/Q226L)/CA04), whereas the virus with the N158K/N224K/Q226L mutations ($10^{5.6}$ p.f.u. g^{-1}; Fig. 3, ferret 3 of rg(N224K/Q226L)/CA04) grew to one order of magnitude higher than

the original mutant. These data indicate that the additional mutation N158D improved the replication of rg(N224K/Q226L)/CA04 in ferrets. To test the effect of this mutation on the replication of H5 reassortant viruses in ferrets, we examined the replicative ability of a virus with the triple N158D/N224K/Q226L HA substitutions in ferrets. This HA(N158D/N224K/Q226L)/CA04 virus replicated efficiently in infected animals, except in the trachea (Fig. 3 and Supplementary Table 2). On day 3 after infection, this virus was isolated from the brain of two of the three animals tested, although we did not observe neurological signs in these animals. These results indicate that the N158D mutation contributed to the efficient growth in the nasal turbinates of ferrets of an H5 reassortant virus with the N224K/Q226L mutations. Removal of the glycosylation site at position 158 has been reported to result in enhanced binding of H5N1 viruses to human-type receptors in combination with the Q226L/G228S mutations[7]. A previous study showed that H5N1 viruses lacking this glycosylation site transmit efficiently by direct contact among guinea-pigs[31]. By contrast, H5N1 viruses that acquire this glycosylation site lose the ability to transmit among guinea-pigs. Therefore, we speculated that the loss of the glycosylation site in HA(N158D/N224K/Q226L)/CA04 virus may affect its transmissibility in ferrets.

To assess the ability of H5 reassortant viruses with human-type receptor specificity to transmit between ferrets, we placed naive ferrets in wireframe cages next to ferrets inoculated with 10^6 p.f.u. of rgCA04, rgVN1203/CA04, rg(N224K/Q226L)/CA04, or HA(N158D/N224K/Q226L)/CA04 (Supplementary Fig. 7). Similar to previous experiments[32], rgCA04 was efficiently transmitted via respiratory droplets to all three contact ferrets, as evidenced by the detection of virus in nasal washes and haemagglutination inhibition (HI) antibody in these animals (Table 1 and Fig. 4). By contrast, rgVN1203/CA04 and rg(N224K/Q226L)/CA04 were not transmitted; neither virus shedding nor seroconversion was detected in any contact animals, despite the binding of the latter to Siaα2,6Gal. This result was consistent with that of previous studies in which human-type receptor recognition was shown to be necessary but not sufficient for respiratory droplet transmission of an H5N1 virus in a ferret model[12,14]. In the HA(N158D/N224K/Q226L)/CA04-inoculated group, virus was recovered from two of the six contact ferrets (pairs 1 and 2) between days 5 and 7 after contact. Moreover, seroconversion was detected in five animals including those from which virus was recovered. No animals died in the course of these transmission experiments. This finding demonstrates the generation of an H5 HA that supports virus transmission by respiratory droplets among ferrets.

To determine whether additional mutations occurred in the HA of HA(N158D/N224K/Q226L)/CA04 during transmission, viral RNA was analysed from nasal washes of inoculated and contact ferrets (Fig. 4 and Supplementary Table 4). On day 5 after infection, the A242S and T318I mutations in HA were present in five (pairs 1, 3, 4, 5 and 6) and one (pair 2) of the six inoculated animals, respectively. Viruses derived from the contact animals of pair 1 on day 7 after contact had two changes in HA

(K193N and A242S) (Fig. 1a), whereas those derived from the contact animals of pair 2 contained a single change in HA (T318I) (Fig. 1b), indicating that additional changes in HA occurred during the infection of ferrets with HA(N158D/N224K/Q226L)/CA04. No mutations in the remaining genes were detected in any of these viruses from nasal washes compared with the CA04 virus sequences.

Because HA(N158D/N224K/Q226L)/CA04 was isolated from only one-third of the contact animals, we isolated a virus from the nasal wash of the contact ferret that shed a high titre ($10^{7.5}$ p.f.u. ml^{-1}) of virus on day 7 after contact (pair 2) (Fig. 4d) to evaluate the replication and transmissibility of that virus in ferrets. This mutant virus, designated HA(N158D/N224K/Q226L/T318I)/CA04, replicated efficiently in the nasal turbinates and was isolated from brain tissue (Fig. 3 and Supplementary Table 2). In the transmission study, four of the six contact ferrets were positive for virus between days 3 and 7 after contact, and all contact animals were seropositive; no animals died in the course of the transmission experiments (Table 1; Fig. 4e and Supplementary Fig. 8). Notably, this transmission pattern is comparable to that of the 1918 pandemic H1N1 virus when tested under the same experimental conditions; the 1918 pandemic virus was recovered from the nasal wash of two of three contact animals (our own unpublished data). Sequence comparison of viruses from inoculated and contact animals identified mutations at positions 225 and 242 as well as a reversion at position 224 (Fig. 1a and Supplementary Table 5) (in addition to the original four mutations) although the 224 reversion was found only in viruses from inoculated ferrets. Collectively, these findings demonstrate that four amino acid substitutions (N158D/N224K/Q226L/T318I) in H5 HA confer efficient respiratory droplet transmission in ferrets to a virus possessing an H5 HA in a 2009 pandemic H1N1 backbone. We also confirmed that recombinant viruses possessing the three HA mutations N158D, N224K and Q226L, or the four HA mutations N158D, N224K, Q226L and T318I, and the NA of VN1203 in a PR8 backgroud (designated N158D/N224K/Q226L or N158D/N224K/Q226L/T318I, respectively) preferentially bind to Siaα2,6Gal and attach to human tracheal epithelia (Fig. 2c, d).

HA(N158D/N224K/Q226L/T318I)/CA04 transmitted by respiratory droplet more efficiently than HA(N158D/N224K/Q226L)/CA04, raising the possibility that the T318I mutation is involved in the efficient transmission of avian H5N1/pandemic H1N1 reassortants. To explore the functional role of this mutation in respiratory droplet transmission, we generated an H5 reassortant expressing the H5 HA with the T318I mutation and examined its receptor-binding specificity and transmissibility. This reassortant (designated rgT318I/CA04) bound to only Siaα2,3Gal and showed little binding to human tracheal epithelia (Fig. 2c, d). rgT318I/CA04 did not transmit via respiratory droplet among ferrets (Table 1 and Fig. 4f), although it replicated in nasal turbinates and trachea as efficiently as rgCA04 (Fig. 3 and Supplementary Table 2). These results indicate that the T318I mutation alone is not sufficient for H5 reassortant viruses to transmit efficiently among ferrets.

**Table 1 | Transmission in ferrets inoculated with H5 avian–human reassortant viruses**

Virus	Inoculated ferrets			Contact ferrets	
	Weight loss (%)*	Peak virus titre in nasal wash (mean log$_{10}$(p.f.u. ml^{-1})) (days after inoculation)	Seroconversion (positive and total numbers) (HI titre)†	Virus detection in nasal wash (positive and total numbers)	Seroconversion (positive and total numbers) (HI titre)
rgCA04	3 of 3 (15.1)	7.5 (1)	3 of 3 (≥1,280, ≥1,280, ≥1,280)	3 of 3	3 of 3 (≥1,280, ≥1,280, ≥1,280)
rgVN1203/CA04	3 of 3 (5.9)	5.3 (5)	3 of 3 (80, 40, 80)	0 of 3	0 of 3 (<10, <10, <10)
rg(N224K/Q226L)/CA04	2 of 3 (7.8)‡	3.9 (5)	3 of 3 (≥1,280, ≥1,280, ≥1,280)	0 of 3	0 of 3 (<10, <10, <10)
HA(N158D/N224K/Q226L)/CA04	6 of 6 (5.7)	6.7 (3)	6 of 6 (640, ≥1,280, ≥1,280, 640, ≥1,280, ≥1,280)	2 of 6	5 of 6 (160, 320, 20, 160, 40, <10)
HA(N158D/N224K/Q226L/T318I)/CA04	6 of 6 (9.8)	6.1 (5)	6 of 6 (≥1,280, ≥1,280, 640, ≥1,280, ≥1,280, ≥1,280)	4 of 6	6 of 6 (640, 640, ≥1280, 80, ≥1,280, 320)
rgT318I/CA04	3 of 5 (1.5)§	5.6 (3)	5 of 5 (40, 20, 20, 40, 40)	0 of 5	0 of 5 (<10, <10, <10, <10, <10)

* Maximum percentage weight loss is shown.
† Haemagglutination inhibition (HI) assays were carried out with homologous virus and turkey red blood cells.
‡ One animal did not lose any body weight.
§ Two animals did not lose any body weight.
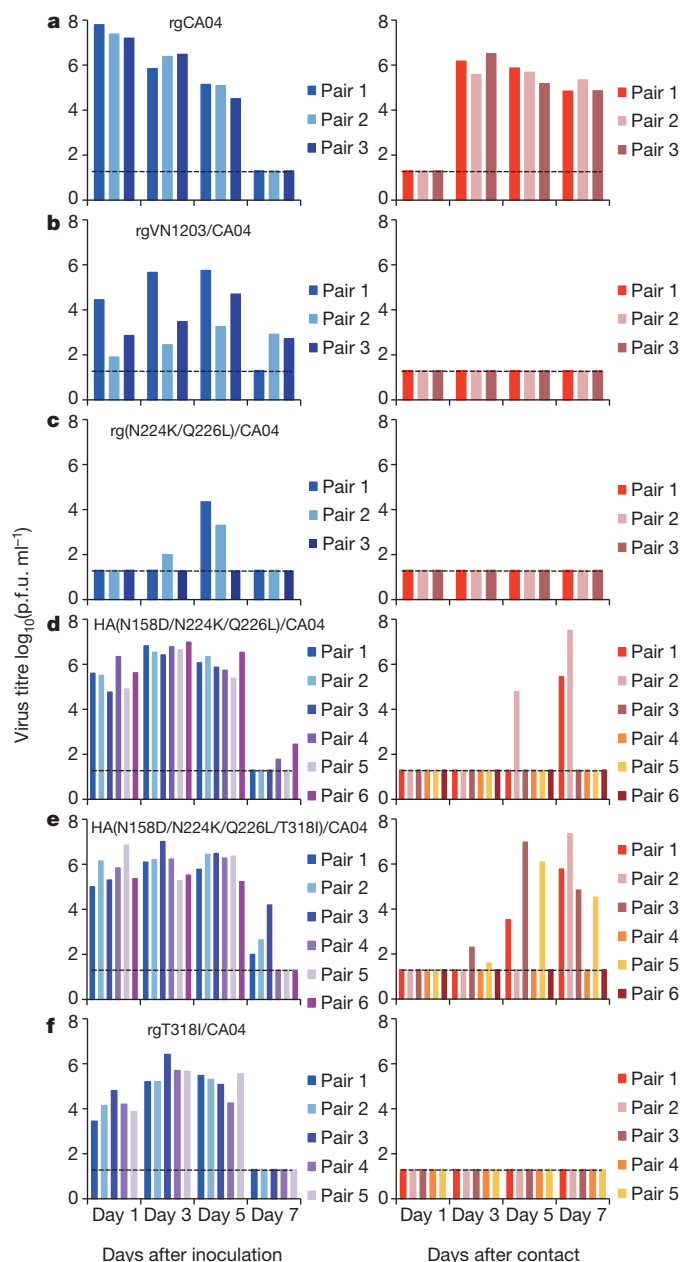
Figure 4 | Respiratory droplet transmission of H5 avian–human reassortant viruses in ferrets. a–f, Groups of three, five, or six ferrets were inoculated intranasally with 10^6 p.f.u. of rgCA04 (**a**), rgVN1203/CA04 (**b**), rg(N224K/Q226L)/CA04 (**c**), HA(N158D/N224K/Q226L)/CA04 (**d**), HA(N158D/N224K/Q226L/T318I)/CA04 (**e**), or rgT318I/CA04 (**f**). One day after infection, three, five, or six naive ferrets were placed in adjacent cages. Nasal washes were collected every other day from both inoculated (left panel) and contact (right panel) animals for virus titration. Virus titres in organs were determined by using a plaque assay on MDCK cells. The lower limit of detection is indicated by the horizontal dashed line.

Influenza virus HA protein has membrane-fusion as well as receptor-binding activity. Notably, in the three-dimensional model of influenza A virus HA, residue 318 is located proximally to the fusion peptide (Fig. 1b), which has key roles in the membrane fusion process. To assess the effect of HA mutations on low-pH-induced membrane fusion activity, we examined the pH at which the fusion activity of wild-type and mutant HA was activated (Fig. 5). The wild-type HA had a threshold for membrane fusion of pH 5.7; the N224K/Q226L and N158D/N224K/Q226L mutations raised the threshold for fusion to >pH 5.9, whereas the T318I mutation reduced the threshold for fusion to pH 5.5. The N158D/N224K/Q226L/T318I mutations showed
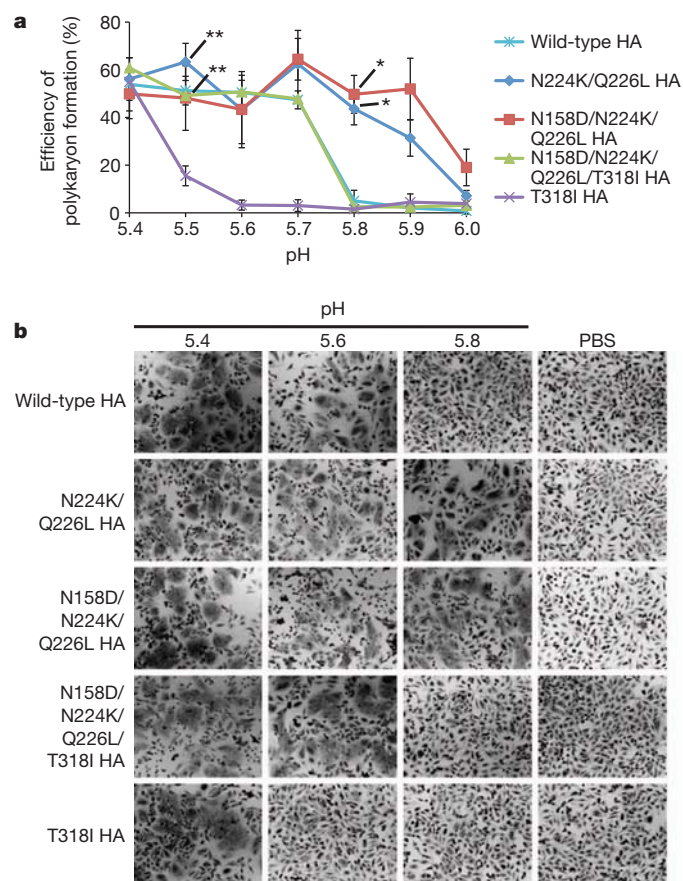

Figure 5 | Polykaryon formation by HeLa cells expressing wild-type or mutant HAs after acidification at low pH. a, The efficiency of polykaryon formation over a pH range of 5.4–6.0 was estimated from the number of nuclei in polykaryons divided by the total number of nuclei in the same field. The mean and standard deviations determined from five randomly chosen fields of cell culture are shown. Single asterisks indicate values significantly different between the wild-type HA and the N224K/Q226L or N158D/N224K/Q226L HA (Tukey test; $P < 0.05$). The double asterisk indicates values significantly different between the T318I HA and the N224K/Q226L or N158D/N224K/Q226L HA (Tukey test; $P < 0.05$). **b,** Representative fields of cells expressing the indicated HAs and exposed to pH 5.4, 5.6, or 5.8 are shown. Images were taken at ×10 magnification.

wild-type fusogenic properties (that is, a threshold at pH 5.7). The HA of influenza virus undergoes a low-pH-dependent conformational change, which is required for fusion of the viral envelope with the target membrane[33]. Such a conformational change to a fusion-active form can also lead to viral inactivation. Therefore, sustained and efficient human-to-human transmission of virus may require a certain level of stability of the HA protein in an acidic environment, as the pH of human nasal mucosa, where human influenza viruses replicate primarily, is approximately pH 5.5–6.5 (ref. 34). Our findings suggest that an increase in the pH threshold for fusion as a result of the N224K/Q226L mutations that shift the HA receptor recognition from avian-type to human-type may reduce HA protein stability; however, the T318I mutation decreases the pH threshold for fusion activity, resulting in a stable mutant HA.

Because heat treatment at neutral pH is also known to promote a fusogenic form of HA protein[35,36] and serve as a surrogate assay for HA stability[37], we next tested whether the HA mutations described above affect the heat stability of the HA protein. Wild-type and mutant HA viruses were incubated at 50 °C for various times, after which the loss of infectivity and haemagglutination activity were determined. The wild-type and N224K/Q226L viruses lost most of their infectivity by heating for 60 min (>5.5-$\log_{10}$ decrease in titre; Fig. 6a), whereas the
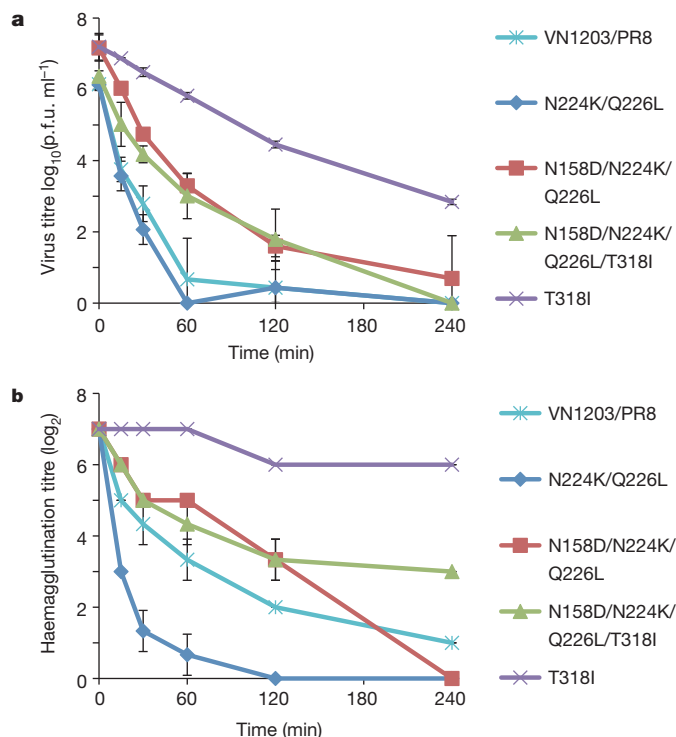
Figure 6 | Effect of heat treatment on the infectivity and haemagglutination activity of viruses. Aliquots of a virus stock containing 128 HA units were incubated for the times indicated at 50 °C. **a**, Virus titres in heat-treated samples were determined by plaque assays on MDCK cells. **b**, Haemagglutination titres in heat-treated samples were determined by using haemagglutination assays with 0.5% TRBCs. Each point represents the mean ± standard deviation from triplicate experiments.

N158D/N224K/Q226L and N158D/N224K/Q226L/T318I mutants exhibited considerable tolerance to high temperature (3.9- and 3.4-$\log_{10}$ decrease after a 60-min incubation, respectively) and the T318I mutant was most resistant (only a 1.4-$\log_{10}$ decrease under the same conditions). In haemagglutination assays, the N224K/Q226L mutant HA lost activity more rapidly than did the wild-type HA, and N158D/N224K/Q226L lost activity more rapidly than did the N158D/N224K/Q226L/T318I mutant (Fig. 6b). Thus, addition of the N158D mutation to the N224K/Q226L HA increased HA stability and subsequent addition of the fourth mutation, T318I, rendered the HA protein even more stable. Taken together, these results suggest that the addition of the T318I mutation to H5 HAs that preferentially recognize human-type receptors restores HA protein stability, thereby allowing a virus carrying the N158D/N224K/Q226L/T318I mutations in HA to transmit efficiently via respiratory droplet among ferrets. In conclusion, a fine balance of mutations affecting different functions in HA (such as receptor-binding specificity and HA stability) may be critical to confer transmissibility in ferrets.

We next compared the pathogenicity in ferrets of H5 avian–human reassortants with that of the pandemic H1N1 virus CA04 (Fig. 7, Supplementary Information and Supplementary Figs 9–11). The control virus, rgCA04, caused substantial body weight loss (15.1%) (Table 1 and Supplementary Fig. 9). By contrast, the four reassortant viruses caused only modest weight loss (<10%) in most of the animals. However, no statistically significant differences in body weight loss were found between the reassortant viruses and rgCA04. Pathological examination revealed similar histological changes and levels of viral antigens in the nasal mucosa of rgCA04-, HA(N158D/N224K/Q226L)/CA04- and HA(N158D/N224K/Q226L/T318I)/CA04-infected ferrets (Fig. 7a, b). In the rgVN1203/CA04 and rg(N224K/Q226L)/CA04 groups, however, less tissue damage was found in the nasal mucosa compared with the rgCA04 group on day 3 after infection (Dunnett's test; $P = 0.0057$ and 0.0175, respectively; Fig. 7b). In addition, all three
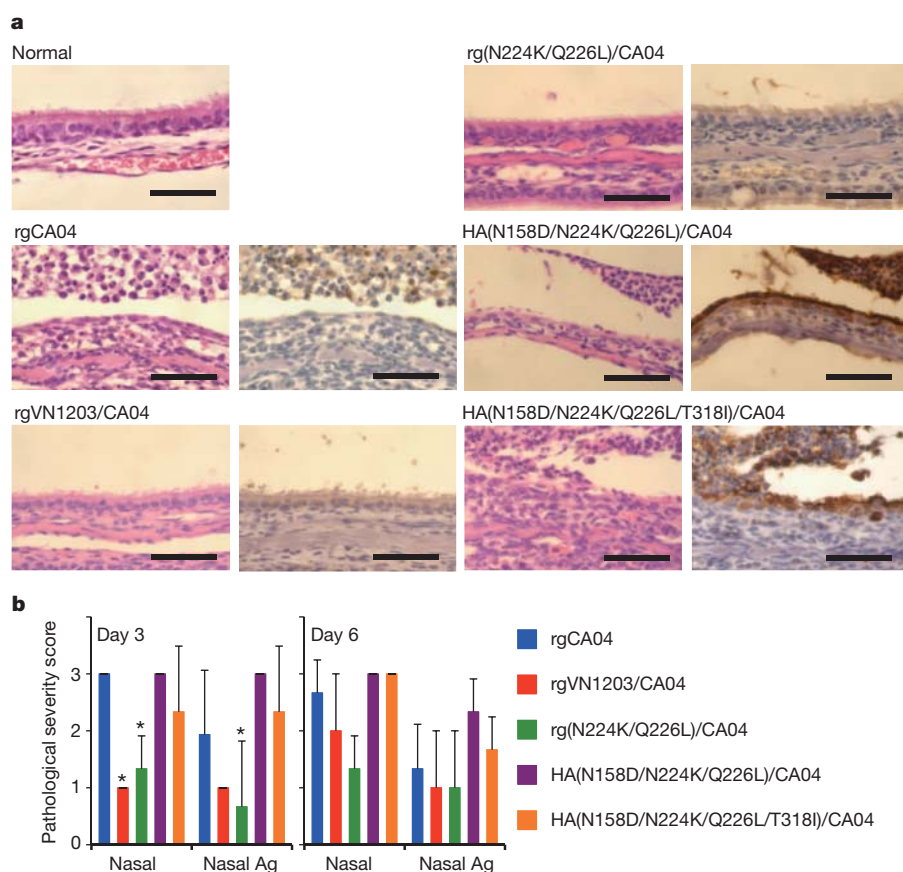
Figure 7 | Pathological analyses of H5 avian–human reassortant viruses. a, Representative histological changes in nasal turbinates from influenza-virus-infected ferrets. Three ferrets per group were infected intranasally with 10^6 p.f.u. of virus, and tissues were collected on day 3 after infection for pathological examination. Uninfected ferret tissues served as negative controls (normal). Left panel, haematoxylin-and-eosin staining. Right panel, immunohistochemical staining for viral antigen detection (brown staining). Scale bars, 50 μm. **b**, Pathological severity scores in infected ferrets. To represent comprehensive histological changes, respiratory tissue slides were evaluated by scoring the pathological changes and viral antigen expression levels. The pathological scores were determined for each animal in each group ($n = 3$ per group on days 3 and 6 after infection) using the following scoring system: 0, no pathological change/antigen negative; 1, affected area (<30%) or only interstitial lesion/rare viral antigens; 2, affected area (<80%, ≥30%)/moderate viral antigens; 3, severe lesion (≥80%)/many viral antigens. Nasal, pathological changes in the nasal mucosa; nasal Ag, viral antigens in the nasal mucosa. Asterisks indicate virus pathological scores significantly different from that of rgCA04 (Dunnett's test; $P < 0.05$). Error bars denote standard deviation.

viruses caused lung lesions (Supplementary Information and Supplementary Figs 10 and 11).

To assess whether current control measures may be effective against the H5 transmissible reassortant mutant virus, we examined the reactivity of sera from individuals vaccinated with an H5N1 prototype vaccine[38] against a virus possessing the N158D/N224K/Q226L/T318I mutations in HA. We found that pooled human sera from individuals immunized with this vaccine reacted with the virus possessing the mutant H5 HA (N158D/N224K/Q226L/T318I) at a higher titre than with a wild-type H5 HA virus (VN1203/PR8; Supplementary Table 6), indicating that current H5N1 vaccines would be efficacious against the H5 transmissible reassortant mutant virus. In addition, the H5 transmissible reassortant mutant virus (HA(N158D/N224K/Q226L/T318I)/CA04) was highly susceptible to a licensed NA inhibitor, oseltamivir (Supplementary Table 7). These experiments show that appropriate control measures would be available to combat the transmissible virus described in this study.

Currently, we do not know whether the mutations that we identified in this study that allowed the HA(N158D/N224K/Q226L/T318I)/CA04 virus to be transmissible in ferrets would also support sustained human-to-human transmission. In particular, we wish to emphasize that the transmissible HA(N158D/N224K/Q226L/T318I)/CA04 virus possesses seven segments (all but the HA segment) from a human pandemic 2009 H1N1 virus. Human-virus-characteristic amino acids in these seven segments may have critically contributed to the respiratory droplet transmission of the HA(N158D/N224K/Q226L/T318I)/CA04 virus in ferrets. Examples include amino acids in the PB2 polymerase protein that confer efficient replication in mammalian, but not avian, cells[39–43]. As the PB2 gene of the HA(N158D/N224K/Q226L/T318I)/CA04 virus is of human virus origin, the virus possesses high replicative ability in mammalian cells. In contrast, most avian virus PB2 proteins lack these human-type amino acids, although one of these changes (a glutamic-acid-to-lysine mutation at position 627) is found in highly pathogenic avian H5N1 viruses circulating in the Middle East[44]. As a second example, the viral NA gene may contribute to viral transmissibility. The NA protein cleaves α-ketosidic linkages between a terminal sialic acid and an adjacent sugar residue, an activity that balances the sialic-acid-binding activity of HA. A recent study found that a human virus NA gene was critical to confer limited transmissibility to a mutant H5 avian–human reassortant virus[14]. In general, a human-type receptor recognizing H5 HA alone may not be sufficient to confer transmissibility in mammals, but may have to act together with other human-virus-characteristic traits (in PB2, NA, and/or other viral proteins). Therefore, at this point we cannot predict whether the four mutations in the H5 HA identified here would render a wholly avian H5N1 virus transmissible.

Three of the residues identified here (N224, Q226 and T318) have been strictly conserved among H5 HA proteins isolated since 2003. However, as H5N1 viruses continue to evolve and infect people, receptor-binding variants of H5N1 viruses, including avian–human reassortant viruses as tested here, may emerge. One of the four mutations we identified in our transmissible virus, the N158D mutation, results in loss of a glycosylation site. Many H5N1 viruses isolated in the Middle East, Africa, Asia and Europe do not have this glycosylation site. Therefore, only three nucleotide changes are needed for the HA of these viruses to support efficient transmission in ferrets. In addition, the H5N1 viruses circulating in these geographic areas also possess a glutamic-acid-to-lysine mutation at position 627 in the PB2 protein, which promotes viral replication in certain mammals, including humans[40,45]. Therefore, these viruses may be several steps closer to those capable of efficient transmission in humans and are of concern.

Our study highlights the pandemic potential of viruses possessing an H5 HA. Although current vaccines may protect against a virus similar to that tested here, the continued evolution of H5N1 viruses reinforces the need to prepare and update candidate vaccines to H5 viruses. The amino acid changes identified here will help individuals conducting surveillance in regions with circulating H5N1 viruses (for example, Egypt, Indonesia, Vietnam) to recognize key residues that predict the pandemic potential of isolates. Rapid responses in a potential pandemic situation are essential in order to generate appropriate vaccines and initiate other public health measures to control infection. Furthermore, our findings are of critical importance to those making public health and policy decisions.

Our research answers a fundamental question in influenza research: can H5-HA-possessing viruses support transmission in mammals? Moreover, our findings have suggested that different mechanisms (that is, receptor-binding specificity and HA stability) may act in concert for efficient transmissibility in mammals. This knowledge will facilitate the identification of additional mutations that affect viral transmissibility; the monitoring of this expanded set of changes in natural isolates may improve our ability to assess the pandemic potential of H5N1 viruses. Thus, although a pandemic H5N1 virus may not possess the amino acid changes identified in our study, the findings described here will advance our understanding of the mechanisms and evolutionary pathways that contribute to avian influenza virus transmission in mammals.

METHODS SUMMARY

Viruses. All recombinant viruses were generated by using reverse genetics essentially as described previously[16]. All experiments with the viruses possessing the wild-type HA cleavage site were performed in an enhanced biosafety level 3 (BSL3+) containment laboratory approved for such use by the CDC and the USDA.

Infection and transmission in ferrets. Six–ten-month-old female ferrets (Triple F Farms) were intramuscularly anaesthetized and intranasally inoculated with 10^6 p.f.u. (500 µl) of virus. On days 3 and 6 after infection, ferrets were killed for virological and pathological examinations. The virus titres in various organs were determined by use of plaque assays in MDCK cells.

For transmission studies in ferrets, animals were housed in adjacent transmission cages that prevented direct and indirect contact between animals but allowed spread of influenza virus through the air (Showa Science; Supplementary Fig. 7). Ferrets were intranasally inoculated with 10^6 p.f.u. (500 µl) of virus (inoculated ferrets). Twenty-four hours after infection, naive ferrets were each placed in a cage adjacent to an inoculated ferret (contact ferrets). To assess viral replication in the nasal turbinates, we determined viral titres in nasal washes collected from virus-inoculated and contact ferrets on day 1 after inoculation or co-housing, respectively, and then every other day. Animal studies were performed in accordance with Animal Care and Use Committee guidelines of the University of Wisconsin-Madison.

Biosafety and biosecurity. All recombinant DNA protocols were approved by the University of Wisconsin-Madison's Institutional Biosafety Committee after risk assessments were conducted by the Office of Biological Safety, and by the University of Tokyo's Subcommittee on Living Modified Organisms, and, when required, by the competent minister of Japan. In addition, the University of Wisconsin-Madison Biosecurity Task Force regularly reviews the research program and ongoing activities of the laboratory. The task force has a diverse skill set and provides support in the areas of biosafety, facilities, compliance, security and health. Members of the Biosecurity Task Force are in frequent contact with the principal investigator and laboratory personnel to provide oversight and assure biosecurity. Experiments with viruses possessing the wild-type HA cleavage site were performed in enhanced BSL3 containment laboratories approved for such use by the CDC and the USDA. Ferret transmission studies were conducted by three scientists with both DVM and PhD degrees who each had more than a minimum of 6 years of experience with highly pathogenic influenza viruses and animal studies with highly pathogenic influenza viruses. Our staff wear powered air-purifying respirators that filter the air, and disposable coveralls; they shower out on exit from the facility. The containment facilities at University of Wisconsin-Madison were designed to exceed standards outlined in Biosafety in Microbiological and Biomedical Laboratories (5th edition; http://www.cdc.gov/biosafety/publications/bmbl5/BMBL.pdf). Features of the BSL3-enhanced suites include entry/exit through a shower change room, effluent decontamination, negative air-pressure laboratories, double-door autoclaves, double HEPA-filtered exhaust air, and gas decontamination ports. The BSL3-Agriculture suite features include all those listed for BSL3-enhanced plus HEPA-filtered supply and double-HEPA-filtered exhaust air, double-gasketed watertight and airtight seals, airtight dampers on all ductwork, and the structure was pressure-decay tested during commissioning. The University of Wisconsin-Madison facility has a dedicated alarm system that monitors all building controls and sends alarms (~500 possible alerts). Redundancies and emergency resources are built-in to the facility including two

air handlers, two compressors, two filters each place filters are needed, two effluent sterilization tanks, two power feeds to the building, an emergency generator in case of a power failure and other physical containment measures in the facility that operate without power. Biosecurity monitoring of the facility is ongoing. All personnel undergo Select Agent security risk assessment by the United States Criminal Justice Information Services Division and complete rigorous biosafety, BSL3 and Select Agent training before participating in BSL3-level experiments. Refresher training is scheduled on a regular basis. The principal investigator participates in training sessions and emphasizes compliance to maintain safe operations and a responsible research environment. The laboratory occupational health plan is in compliance with the University of Wisconsin-Madison Occupational Health Program. Select agent virus inventory is checked monthly and submitted to the University of Wisconsin-Madison Research Compliance Specialist. Virus inventory is submitted 1–2 times per year to the file holder in the Select Agent branch of the CDC. The research program, procedures, occupational health plan, documentation, security and facilities are reviewed annually by the University of Wisconsin-Madison Responsible Official and at regular intervals by the CDC and the Animal and Plant Health Inspection Service (APHIS) as part of the University of Wisconsin-Madison Select Agent Program.

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

1. Glaser, L. et al. A single amino acid substitution in 1918 influenza virus hemagglutinin changes receptor binding specificity. J. Virol. **79,** 11533–11536 (2005).
2. Matrosovich, M. et al. Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. J. Virol. **74,** 8502–8512 (2000).
3. Rogers, G. N. & Paulson, J. C. Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. Virology **127,** 361–373 (1983).
4. Auewarakul, P. et al. An avian influenza H5N1 virus that binds to a human-type receptor. J. Virol. **81,** 9950–9955 (2007).
5. Gambaryan, A. et al. Evolution of the receptor binding phenotype of influenza A (H5) viruses. Virology **344,** 432–438 (2006).
6. Stevens, J. et al. Recent avian H5N1 viruses exhibit increased propensity for acquiring human receptor specificity. J. Mol. Biol. **381,** 1382–1394 (2008).
7. Wang, W. et al. Glycosylation at 158N of the hemagglutinin protein and receptor binding specificity synergistically affect the antigenicity and immunogenicity of a live attenuated H5N1 A/Vietnam/1203/2004 vaccine virus in ferrets. J. Virol. **84,** 6570–6577 (2010).
8. Watanabe, Y. et al. Acquisition of human-type receptor binding specificity by new H5N1 influenza virus sublineages during their emergence in birds in Egypt. PLoS Pathog. **7,** e1002068 (2011).
9. Yamada, S. et al. Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. Nature **444,** 378–382 (2006).
10. Jackson, S. et al. Reassortment between avian H5N1 and human H3N2 influenza viruses in ferrets: a public health risk assessment. J. Virol. **83,** 8131–8140 (2009).
11. Maines, T. R. et al. Lack of transmission of H5N1 avian-human reassortant influenza viruses in a ferret model. Proc. Natl Acad. Sci. USA **103,** 12121–12126 (2006).
12. Maines, T. R. et al. Effect of receptor binding domain mutations on receptor binding and transmissibility of avian influenza H5N1 viruses. Virology **413,** 139–147 (2011).
13. Yen, H. L. et al. Inefficient transmission of H5N1 influenza viruses in a ferret contact model. J. Virol. **81,** 6890–6898 (2007).
14. Chen, L. M. et al. In vitro evolution of H5N1 avian influenza virus toward human-type receptor specificity. Virology **422,** 105–113 (2012).
15. Stevens, J. et al. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science **312,** 404–410 (2006).
16. Neumann, G. et al. Generation of influenza A viruses entirely from cloned cDNAs. Proc. Natl Acad. Sci. USA **96,** 9345–9350 (1999).
17. Hatakeyama, S. et al. Enhanced expression of an α2,6-linked sialic acid on MDCK cells improves isolation of human influenza viruses and evaluation of their sensitivity to a neuraminidase inhibitor. J. Clin. Microbiol. **43,** 4139–4146 (2005).
18. Shinya, K. et al. Avian flu: influenza virus receptors in the human airway. Nature **440,** 435–436 (2006).
19. Varki, A. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature **446,** 1023–1029 (2007).
20. Rogers, G. N. et al. Host-mediated selection of influenza virus receptor variants. Sialic acid-α2,6Gal-specific clones of A/duck/Ukraine/1/63 revert to sialic acid-α2,3Gal-specific wild type in ovo. J. Biol. Chem. **260,** 7362–7367 (1985).
21. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. Proc. Natl Acad. Sci. USA **98,** 11181–11186 (2001).
22. Cline, T. D. et al. Increased pathogenicity of a reassortant 2009 pandemic H1N1 influenza virus containing an H5N1 hemagglutinin. J. Virol. **85,** 12262–12270 (2011).
23. Octaviani, C. P., Ozawa, M., Yamada, S., Goto, H. & Kawaoka, Y. High level of genetic compatibility between swine-origin H1N1 and highly pathogenic avian H5N1 influenza viruses. J. Virol. **84,** 10918–10922 (2010).
24. Li, H. et al. Isolation and characterization of H5N1 and H9N2 influenza viruses from pigs in China [in Chinese]. Chin. J. Prev. Vet. Med. **26,** 1–6 (2004).
25. Nidom, C. A. et al. Influenza A (H5N1) viruses from pigs, Indonesia. Emerg. Infect. Dis. **16,** 1515–1523 (2010).
26. Pasma, T. & Joseph, T. Pandemic (H1N1) 2009 infection in swine herds, Manitoba, Canada. Emerg. Infect. Dis. **16,** 706–708 (2010).
27. Pereda, A. et al. Pandemic (H1N1) 2009 outbreak on pig farm, Argentina. Emerg. Infect. Dis. **16,** 304–307 (2010).
28. Welsh, M. D. et al. Initial incursion of pandemic (H1N1) 2009 influenza A virus into European pigs. Vet. Rec. **166,** 642–645 (2010).
29. Xu, Q., Wang, W., Cheng, X., Zengel, J. & Jin, H. Influenza H1N1 A/Solomon Island/3/06 virus receptor binding specificity correlates with virus pathogenicity, antigenicity, and immunogenicity in ferrets. J. Virol. **84,** 4936–4945 (2010).
30. van Riel, D. et al. Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. Am. J. Pathol. **171,** 1215–1223 (2007).
31. Gao, Y. et al. Identification of amino acids in HA and PB2 critical for the transmission of H5N1 avian influenza viruses in a mammalian host. PLoS Pathog. **5,** e1000709 (2009).
32. Itoh, Y. et al. In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. Nature **460,** 1021–1025 (2009).
33. Skehel, J. J. & Wiley, D. C. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu. Rev. Biochem. **69,** 531–569 (2000).
34. England, R. J., Homer, J. J., Knight, L. C. & Ell, S. R. Nasal pH measurement: a reliable and repeatable parameter. Clin. Otolaryngol. Allied Sci. **24,** 67–68 (1999).
35. Carr, C. M., Chaudhry, C. & Kim, P. S. Influenza hemagglutinin is spring-loaded by a metastable native conformation. Proc. Natl Acad. Sci. USA **94,** 14306–14313 (1997).
36. Haywood, A. M. & Boyer, B. P. Time and temperature dependence of influenza virus membrane fusion at neutral pH. J. Gen. Virol. **67,** 2813–2817 (1986).
37. Krenn, B. M. et al. Single HA2 mutation increases the infectivity and immunogenicity of a live attenuated H5N1 intranasal influenza vaccine candidate lacking NS1. PLoS ONE **6,** e18577 (2011).
38. Treanor, J. J., Campbell, J. D., Zangwill, K. M., Rowe, T. & Wolff, M. Safety and immunogenicity of an inactivated subvirion influenza A (H5N1) vaccine. N. Engl. J. Med. **354,** 1343–1351 (2006).
39. Bussey, K. A., Bousse, T. L., Desmet, E. A., Kim, B. & Takimoto, T. PB2 residue 271 plays a key role in enhanced polymerase activity of influenza A viruses in mammalian host cells. J. Virol. **84,** 4395–4406 (2010).
40. Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. Science **293,** 1840–1842 (2001).
41. Li, Z. et al. Molecular basis of replication of duck H5N1 influenza viruses in a mammalian mouse model. J. Virol. **79,** 12058–12064 (2005).
42. Mehle, A. & Doudna, J. A. Adaptive strategies of the influenza virus polymerase for replication in humans. Proc. Natl Acad. Sci. USA **106,** 21312–21316 (2009).
43. Yamada, S. et al. Biological and structural characterization of a host-adapting amino acid in influenza virus. PLoS Pathog. **6,** e1001034 (2010).
44. Salzberg, S. L. et al. Genome analysis linking recent European and African influenza (H5N1) viruses. Emerg. Infect. Dis. **13,** 713–718 (2007).
45. Subbarao, E. K., London, W. & Murphy, B. R. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. J. Virol. **67,** 1761–1764 (1993).

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

Author Contributions M.I., T.W., M.H., S.C.D., M.O., K.S., G.Z., A.H., H.K., S.W., C.L., S.Y., M.K., Y.S., E.A.M., G.N. and Y.K. designed the experiments; M.I., T.W., M.H., S.C.D., M.O., K.S., G.Z., A.H., H.K., S.W., C.L., S.Y. and M.K. performed the experiments; M.I., T.W., M.H., S.C.D., M.O., K.S., G.Z., A.H., H.K., S.W., C.L., E.K., S.Y., M.K., Y.S., E.A.M., G.N. and Y.K. analysed the data; M.I., T.W., M.H., S.C.D., K.S., E.A.M., G.N. and Y.K. wrote the manuscript; M.I., T.W. and M.H. contributed equally to this work.

METHODS

Cells. Madin–Darby canine kidney (MDCK) cells and MDCK cells overexpressing Siaα2,6Gal (AX4 cells[17]) were maintained in Eagle's minimal essential medium (MEM) containing 5% newborn calf serum. Human embryonic kidney 293T cells were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum (FBS). HeLa cells were maintained in MEM containing 10% FBS. All cells were maintained at 37 °C in 5% CO_2.

Plasmid construction and reverse genetics. Plasmid constructs for viral RNA production (pPolI)—containing the genes of the A/Vietnam/1203/2004 (H5N1; VN1203), A/Puerto Rico/8/34 (H1N1; PR8), A/Kawasaki/173/2001 (H1N1; K173) and A/California/04/2009 (H1N1; CA04) viruses flanked by the human RNA polymerase I promoter and the mouse RNA polymerase I terminator—were constructed as described[16]. The multibasic amino acids at the haemagglutinin (HA) cleavage site (RERRRKKR↓G) of the reassortant viruses between VN1203 and PR8 were changed to RETR↓G by site-directed mutagenesis. All transfectant viruses were generated by using reverse genetics essentially as described previously[16]. Recombinant viruses were amplified in MDCK or AX4[17] cells and stored at −80 °C until use. The HA segment of all viruses was sequenced to ensure the absence of unwanted mutations. All experiments with the reassortant viruses between VN1203 and CA04 were performed in enhanced biosafety level 3 containment laboratories approved for such use by the CDC and the USDA.

To introduce random mutations into the globular head of the VN1203 HA protein, a 143-amino-acid region spanning residues 120–259 (H3 numbering) was selected. This region was subjected to PCR-based random mutagenesis by use of the GeneMorph II kit (Stratagene) following the manufacturer's instructions. The targeted mutation rate (1–2 amino acid replacements per molecule) was achieved through optimization of the template quantity, and was confirmed by sequence analysis of 48 individual clones. By using a PCR-based cloning strategy, we inserted the mutagenized region into its respective vector containing the VN1203 HA gene between the human RNA polymerase I promoter and mouse RNA polymerase I terminator sequences. The composition of the plasmid library was confirmed by sequencing. The plasmid library was then used to generate an influenza virus library, essentially as described[16]. The size of the virus library was 7×10^6 p.f.u.

Preparation of sialidase-treated TRBCs. Turkey red blood cells (TRBCs) were washed three times with phosphate-buffered saline (PBS), and diluted to 20% (vol/vol) in PBS. TRBCs (1 ml) were incubated with 500 U of α2,3-sialidase from *Salmonella enterica* serovar Typhimurium LT2 (NEB) for 20–24 h at 37 °C, washed three times in PBS, and re-suspended in PBS or MEM containing 1% bovine serum albumin (BSA) (MEM/BSA).

Haemagglutination assay. Viruses (50 μl) were serially diluted with 50 μl of PBS in a microtitre plate. An equal volume (that is, 50 μl) of a 0.5% (vol/vol) TRBC suspension was added to each well. The plates were kept at room temperature and haemagglutination was assessed after a 1-h incubation.

Virus library screening. To select VN1203 HA variants that had acquired the ability to recognize human-type receptors, three parallel experiments were carried out, each with 0.7×10^6 viruses. The virus library was first incubated with 0.1 ml of 10% (vol/vol) α2,3-sialidase-treated TRBCs for 10 min at 4 °C. After this incubation, the TRBCs and bound viruses were pelleted at 1,000 r.p.m. for 1 min, and the pellets then washed ten times in MEM/BSA containing 313 mM NaCl. Bound viruses were eluted by incubation at 37 °C for 30 min and then diluted to approximately 0.5 virus per well (determined by virus titration in a pilot study). Individual viruses were then amplified in AX4 cells, which overexpress Siaα2,6Gal[17]. Individual viruses were re-screened by using haemagglutination assays with α2,3-sialidase-treated TRBCs.

Solid-phase binding assay. Viruses were grown in MDCK cells, clarified by low-speed centrifugation, laid over a cushion of 30% sucrose in PBS, and ultracentrifuged at 25,000 r.p.m. for 2 h at 4 °C. Virus stocks were aliquoted and stored at −80 °C. Virus concentrations were determined by using haemagglutination assays with 0.5% (vol/vol) TRBCs. The direct receptor-binding capacity of viruses was examined by use of a solid-phase binding assay as previously described[9]. Microtitre plates (Nunc) were incubated with the sodium salts of sialylglycopolymers (poly-L-glutamic acid backbones containing *N*-acetylneuraminic acid linked to galactose through either an α2,3 (Neu5Acα2,3Galβ1,4GlcNAcβ1-pAP) or an α2,6 (Neu5Acα2,6Galβ1,4GlcNAcβ1-pAP) bond) in PBS at 4 °C overnight. After the glycopolymer solution was removed, the plates were blocked with 0.15 ml of PBS containing 4% BSA at room temperature for 1 h. After four successive washes with ice-cold PBS, the plates were incubated in a solution containing influenza virus (8–32 HA units in PBS) at 4 °C overnight. After washing as described above, the plates were incubated for 2 h at 4 °C with rabbit polyclonal antiserum to either K173 or VN1203 virus. The plates were then washed again as before and incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG antiserum for 2 h at 4 °C. After washing, the plates were incubated with *O*-phenylenediamine

(Sigma) in PBS containing 0.01% H_2O_2 for 10 min at room temperature, and the reaction was stopped with 0.05 ml of 1 M HCl. The optical density at 490 nm was determined in a plate reader (Infinite M1000; Tecan).

Virus binding to human airway tissues. Paraffin-embedded normal human trachea (US Biological) and lung (BioChain) tissue sections were deparaffinized and rehydrated. Sections were then blocked by using 4% BSA in PBS and covered with virus suspensions (64 HA units in PBS) at 4 °C overnight. After being washed four times in ice-cold PBS, the sections were incubated with primary antibodies for 3 h at 4 °C. The primary antibodies used were as follows: a pool of mouse anti-VN1203 HA monoclonal antibodies (15A3, 3G2, 7A11, 8A3, 14C5 and 18E1; Rockland); rabbit anti-K173 polyclonal antibody; rabbit anti-surfactant protein A polyclonal antibody (Millipore); and mouse anti-surfactant protein A monoclonal antibody (Abcam). Antibody binding was detected by using an IgG secondary antibody conjugated with Alexa Fluor 488 or Alexa Fluor 633 (Molecular Probes). Sections were also counterstained with Hoechst 33342, trihydrochloride, trihydrate (Molecular Probes). The samples were examined by using confocal laser scanning microscopy (model LSM 510; Carl Zeiss).

To confirm sialic-acid-specific virus binding, tissue sections were treated, before incubation with viruses, with *Arthrobacter ureafaciens* sialidase (Sigma) for 3 h at 37 °C. Viruses bound to tissue were detected as described above.

Experimental infection of ferrets. Animal studies were performed in accordance with the Animal Care and Use Committee guidelines of the University of Wisconsin-Madison. We used 6–10-month-old female ferrets (Triple F Farms) that were serologically negative by haemagglutination inhibition (HI) assay for currently circulating human influenza viruses. Six ferrets per group were anaesthetized intramuscularly with ketamine and xylazine (5–30 mg and 0.2–6 mg kg^{-1} of body weight, respectively) and inoculated intranasally with 10^6 p.f.u. (500 μl) of viruses. On days 3 and 6 after infection, three ferrets per group were killed for virological and pathological examinations. The virus titres in various organs were determined by use of plaque assays in MDCK cells.

Excised tissue samples of nasal turbinates, trachea, lungs, brain, liver, spleen, kidney and colon from euthanized ferrets were preserved in 10% phosphate-buffered formalin. Tissues were then trimmed and processed for paraffin embedding and cut into 5-μm-thick sections. One section from each tissue sample was stained by using a standard haematoxylin-and-eosin procedure, whereas another one was processed for immunohistological staining with a mixture of two anti-influenza virus rabbit antibodies (1:2,000; R309 and anti-VN1203; both prepared in our laboratory) that react with CA04 and VN1203, respectively. Specific antigen–antibody reactions were visualized by using an indirect two-step dextran-polymer technique (Dako EnVision system; Dako) and 3,3′-diaminobenzidine tetrahydrochloride staining (Dako).

Ferret transmission study. For transmission studies in ferrets, animals were housed in adjacent transmission cages that prevented direct and indirect contact between animals but allowed spread of influenza virus through the air (Showa Science; Supplementary Fig. 7). Three, five, or six ferrets were inoculated intranasally with 10^6 p.f.u. (500 μl) of virus (inoculated ferrets). Twenty-four hours after infection, three, five, or six naive ferrets were each placed in a cage adjacent to an inoculated ferret (contact ferrets). The ferrets were monitored for changes in body weight and the presence of clinical signs. To assess viral replication in nasal turbinates, we determined viral titres in nasal washes collected from virus-inoculated and contact ferrets on day 1 after inoculation or co-housing, respectively, and then every other day.

Serological tests. Serum samples were collected between days 14 and 20 after infection, treated with receptor-destroying enzyme, heat-inactivated at 56 °C for 30 min, and tested by use of an HI assay with 0.5% TRBCs (http://www.wpro.who.int/entity/emerging_diseases/documents/docs/manualonanimalaidiagnosisandsurveillance.pdf). Viruses bearing homologous HA were used as antigens for the HI tests.

Polykaryon formation representing membrane fusion activity. Monolayers of HeLa cells grown in 12-well plates were transfected with the protein expression vector pCAGGS[46] encoding wild-type or mutant HA. At 24 h after transfection, cells transiently expressing HA protein were treated with trypsin (1 μg ml^{-1}) in MEM containing 0.3% BSA for 30 min at 37 °C to cleave the HA into its HA1 and HA2 subunits. Polykaryon formation was induced by exposing the cells to low-pH buffer (145 mM NaCl, 20 mM sodium citrate (pH 6.0–5.4)) for 2 min at 37 °C. After this exposure, the low-pH buffer was replaced with MEM containing 10% FBS and the cells were incubated for 3 h at 37 °C. The cells were then fixed with methanol and stained with Giemsa's solution and photographed with a digital camera mounted on an inverted microscope (Nikon, Eclipse Ti). For quantitative analyses, cell nuclei were counted in five randomly chosen fields of cell culture. Polykaryon formation activity was calculated from the number of nuclei in polykaryons divided by the total number of nuclei in the same field.

Thermostability. Viruses (128 HA units in PBS) were incubated for the times indicated at 50 °C. Subsequently, infectivity and haemagglutination activity were determined by use of plaque assays in MDCK cells and haemagglutination assays using 0.5% TRBCs, respectively.

Neuraminidase (NA) inhibition assay. To assess the sensitivity of viruses to the NA inhibitor oseltamivir, NA inhibition assays were performed as described previously[32].

Statistical analysis. All statistical analyses were performed using JMP 9.0.0 (SAS Institute Inc.). The statistical significance of differences between rgCA04 and H5 avian/human reassortant viruses was determined by using a Dunnett's test. Comparisons of polykaryon formation between wild-type and mutant HAs were done using Tukey's test. P values of <0.05 were considered significant.

46. Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108,** 193–199 (1991).

CAREERS

| |

S. YAKOV/SHUTTERSTOCK

LIFE SCIENCES

All jazzed up

Bioscience is thriving in New Orleans as the city bounces back from multiple disasters.

BY AMANDA MASCARELLI

In New Orleans, Louisiana, in 2009, three men hatched a plan to start a business. One man, Mark Heiman, was then a chief scientific officer at pharmaceutical company Eli Lilly, based in Indianapolis, Indiana; another, Dale Pfost, was a serial biotechnology entrepreneur; and the third, John Elstrott, was a professor of entrepreneurship at Tulane University in New Orleans and chair of Whole Foods Market, a natural-foods supermarket chain based in Austin, Texas. Their meeting led to the creation in 2010 of New Orleans biotechnology start-up NuMe Health, which develops food compounds that promote the growth of healthy microbes in the gut. The founders attribute the partnership's creation not only to the pooling of their respective strengths, but also, says Heiman, to "the magic" of New Orleans: the city's innovative spirit, already known to inspire greatness in food and music.

Increasingly, bioscience researchers are trying to capture the same inspiration.

This bioscience renaissance is powered by a programme of investment by the state and federal governments and by the private sector, much of which was spurred by the effort to help the region to recover from the ravages of Hurricane Katrina in 2005. The area is also benefiting from research funds resulting from the Deepwater Horizon oil spill in 2010 (see 'Investment after the spill'). "This part of the country has become very attractive to people who are very idealistic and want to make a difference," says Vijay John, a chemical engineer at Tulane. "A lot of risk-taking young people have moved in. There's a sense of commitment, of community."

BIOSCIENCE ON THE BAYOU

Not long ago, New Orleans and the surrounding area had virtually no start-up companies. But in the past few years, nearly 60 have sprung up, many in the biosciences sector, attracting

the attention of big-time investors, out-of-state venture-capital funds and pharmaceutical companies. Although the city is certainly not yet an established biotechnology hub, onlookers see the potential for significant growth.

"I'm beginning to see a critical mass of these elements of an entrepreneurial environment come together," says Tom Dickerson, adviser to Louisiana Fund I, a US$26-million healthcare venture-capital fund based in Baton Rouge. The region offers a cost of living that is below the US average, as well as robust tax incentives for investors and start-ups and a strong pool of talent from research and educational institutions.

Although start-ups often rely on just a handful of employees, the smattering of new businesses is already creating opportunities for research jobs and internships at all levels, from undergraduates to junior and senior researchers. The biotechnology boom is also offering scientists a chance to see their ▶

work commercialized, says Dickerson, noting that technology-transfer opportunities have become a tool for attracting top-quality researchers to universities.

The heart of the local biotechnology movement is BioDistrict New Orleans, a 6-square-kilometre state-supported economic-development area established in the business district in 2005. It draws together intellectual resources from the city's research institutions, including Tulane University, Xavier University of Louisiana, Louisiana State University (LSU) Health Sciences Center and Delgado Community College. One big project in the district is the Louisiana Cancer Research Center, a $102-million facility that has created 235 jobs — mostly in research — and drawn nearly $155 million in research grants. In 2011, officials also broke ground on the state's new teaching hospital, the University Medical Center, and construction has begun on the New Orleans Veterans Affairs Medical Center.

INCUBATING INNOVATION

Another key component of the BioDistrict is the 6,000-square-metre New Orleans Bio-Innovation Center, a non-profit life-sciences business incubator focused on university research and located on Canal Street near the French Quarter — close to the hub of New Orleans' economic activity, and to Tulane and the LSU Health Sciences Center.

The idea for the innovation centre took root in 2002, when Louisiana's governor was seeking ways to spur economic development, forge a 'knowledge economy' rather than relying on tourism, and revive an area full of empty hotels and office buildings. "Everyone was looking for ways to rejuvenate the area while at the same time providing this kind of resource to the universities," says Aaron Miscenich, who became president of the centre in 2004.

At the time, there was little local precedent for translating research into products. Miscenich didn't see much of a culture of entrepreneurship. "There was brilliant research being done in the city, but none was being brought to market," he says. "Historically, the technologies were either leaving the state or just sitting in filing cabinets." And before Katrina, there was little economic investment to support such ventures, he adds.

"As some of these ventures succeed, hopefully they can spool up into larger-scale operations serving a national and global market."

When the hurricane hit, plans for the facility ground to a halt. "The level of uncertainty in our community was just staggering," says Miscenich. In late 2005 and early 2006, centre principals re-evaluated the business model, says Miscenich, and cut the projected size of the centre by half, but the state of Louisiana stepped in to provide $47 million in construction costs, and building went ahead. The BioInnovation Center opened in 2011 and the state currently supports most of the facility's annual operational costs.

The centre houses around 20 tenants, from biotechnology start-ups and a clinical-research organization (CRO) to intellectual-property attorneys and venture capitalists. Ultimately, it will provide space for up to 50 companies, which Miscenich hopes will draw well-established businesses and CROs to the city to create a wealth of research positions. "As some of these ventures succeed, hopefully they can spool up into larger-scale operations serving a national and global market," says Michael Bernstein, provost at Tulane. He is optimistic that companies will remain and generate more jobs.

THE SEEDS OF A HUB

Collaborations are already developing. NuMe Health is a tenant of the BioInnovation Center, and Heiman, who says the centre helps spur alliances, has built promising relationships with two companies in the building. "That led to a volley of brainstorming and an outline of potential new areas for our companies," he says. Heiman notes that Louisiana offers several incentives for start-ups, including an angel-investor tax credit and a 'phase zero' grant that funds companies to draw up proposals for higher-level funding; such schemes helped his company get off the ground, he says, and should draw other businesses.

Sudhir Sinha, a scientist turned entrepreneur, sold one biotechnology company in 2008 and was planning to retire when he learned about the opportunities at the BioInnovation Center. He rented space in the incubator to launch a new business, InnoGenomics, which specializes in techniques for extracting forensic information from trace amounts of degraded DNA samples found at crime scenes. He got the company under way with start-up capital, including a $150,000

ENVIRONMENTAL SCIENCE

Investment after the spill

Environmental researchers in New Orleans and the surrounding area are benefiting from the silver lining of a rather large cloud: an influx of research dollars after the Deepwater Horizon oil spill in the Gulf of Mexico in 2010.

BP, the operator of the well that spewed almost 5 million barrels of oil into the Gulf, pledged to provide US$500 million over ten years to support research related to the spill and its impacts on ecosystems. The company established the Gulf of Mexico Research Initiative (GoMRI), which so far has awarded eight grants to consortia made up of research institutions from around the region.

"With the federal and state budgets declining so drastically, it's one of the few sources of external research money right now," says Nancy Rabalais, director of the Louisiana Universities Marine Consortium based in Chauvin, which has received one of the grants. The disaster brought national attention to some of the Gulf region's pressing long-term problems, such as land loss and hypoxia in coastal waters. Ultimately, she says, this has produced a shift in conservation: from small-scale mitigation projects to whole-ecosystem restoration.

The Louisiana Universities Marine Consortium's GoMRI project focuses on assessing the effects of the spill on coastal ecosystems and will create 25–30 positions, says Rabalais. Graduate students, postdoctoral researchers and research assistants will study marsh degradation and the effects of the oil on organisms living in shallow waters near the continental shelf. The 12 institutions comprising the consortium will collectively receive $12 million, divided among all of them, over three years.

The demand for researchers is also growing in the private sector, says Ralph Portier, an environmental scientist at Louisiana State University (LSU) in Baton Rouge. LSU has just approved a PhD programme in environmental sciences, set to begin in autumn 2012, which will focus on research related to the oil spill. The university also offers a master's degree in environmental sciences, the "scientific equivalent of an MBA", says Portier: companies send policy specialists and other employees without research backgrounds to develop expertise in areas such as soils and coastal restoration. "Employees go back to their companies with the environmental training needed to function as environmental-scientist planners and managers," says Portier.

The region is a hotspot of pressing environmental issues, adds Portier, noting its constellation of ecological attributes — including the Mississippi River, wetlands and agricultural land — and development challenges resulting from industrialization of oil and gas resources in the Gulf. "There are lots of issues, lots of problems and lots of opportunities," he says. **A.M.**

grant from the US National Science Foundation, and received help in developing his commercialization plan from graduating MBA students and interns at the centre. In the past year, InnoGenomics has created five full-time lab-research positions; Sinha anticipates adding several more in the next 12–18 months, including jobs for undergraduates and postdocs in biochemistry and molecular biology.

Another tenant, energy-technology company ReactWell, has taken advantage of the centre's non-profit status, core laboratory facility and access to Tulane's office of technology transfer, says founder Brandon Iglesias. In April, Iglesias won $20,000 in start-up capital from the Tulane Business Plan Competition and the Domain Companies New Orleans Entrepreneur Challenge, to develop ways to use underground geothermal reactors to create synthetic crude oil from algal biomass. "It's quite scary to be in a start-up because the risk is pretty high," he says. "But I can't think of a more important issue to be working on than energy security." Iglesias says that New Orleans' location on the major waterway of the Mississippi River, close to the oil and gas industry and aquaculture in the Gulf of Mexico, gave him ready access to experts in his field.

The city still faces formidable challenges. More start-up capital is needed for new companies, and researchers at local universities aren't necessarily trained to meet businesses' needs.

"We have these institutions in place to help with training, but we just need to make sure that the programming they're providing matches the needs of the companies that we're putting together," says Miscenich. He adds that, for the city to succeed as a biohub, it must be able to retain talent. Louisiana must maintain its tax incentives and business infrastructure, and encourage an influx of established pharmaceutical firms and CROs to buoy the young companies and bring new jobs to the area.

"We need to come up with a business rationale to keep these companies in New Orleans," says Miscenich. "It's not just going to be the good food or the music; it's going to be because it makes economic sense to the company." ■

Amanda Mascarelli *is a freelance science writer based in Denver, Colorado.*

COLUMN
Piled too high

A passion for science is admirable, but can have unwanted outcomes, argues **Mariano A. Loza-Coll**.

An independent film about the pursuit of a science PhD became a hit last year — at least among the fledgling scientists that it represents. Thousands flocked to see *The PhD Movie*, based on the hugely popular comic strip *Piled Higher and Deeper* by Jorge Cham, a former mechanical engineer, at hundreds of on-campus screenings in several countries. Not bad for a film produced, directed and acted in by graduate students and researchers at the California Institute of Technology in Pasadena, and funded on a shoe-string budget.

Why the popularity? The film puts the plight of the PhD student on the big screen, giving student audiences a chance to laugh at themselves in recognition of the years of schooling, hard work and frustration that they are undertaking.

But I would argue that there is another reason: the film tackles some of the negative aspects of pursuing a PhD and a science career. At its climax, *The PhD Movie* raises a question that crosses many students' minds: why bother? The answer it provides resonates with the audience: "Everybody is here because they want to be here … You have to embrace the things you're passionate about."

Yet the passion that drives many scientists to investigate the natural and physical world can also get them into trouble. It compels them to push the boundaries of science, but not always with healthful long-term consequences.

Hard work and a stubborn can-do attitude are admirable and rewarding. We put hours, days and months into experiments that don't work, because we're passionate about the exhilarating moments when they do. But the trade-offs and downsides are many. The romantic ideal of the scientist as an independent academic investigator uncovering knowledge can convince some very smart people to accept mediocre pay and delays in starting a family, settling down, buying a house and planning for retirement — with no guarantees that these sacrifices are worthwhile. That is the drama so nicely portrayed in *The PhD Movie*.

The film could be seen as a warning, showing budding researchers that they need to be realistic about their career expectations. But I see it as much more. It is a wake-up call for scientists and non-scientists alike. After all, it is not scientists, but the rest of society that may suffer most if scientists end up burning out. To ensure that scientists can cure diseases, feed the hungry and prevent environmental catastrophes, we should
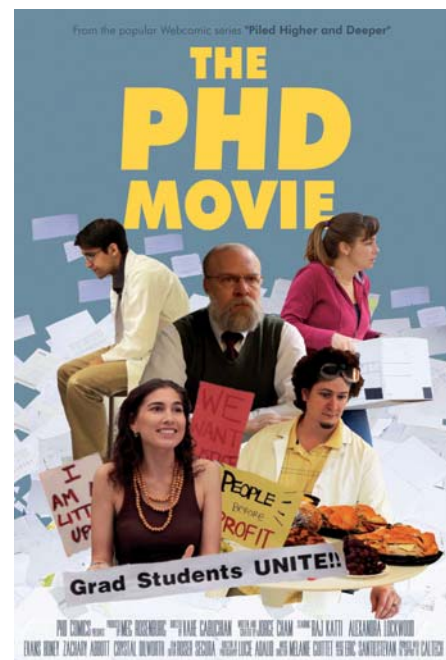
select them on the basis of their intelligence, effectiveness and focus, not their personal sacrifices, obsession and stamina.

Why, then, isn't society at large demanding changes? I suspect that most people just don't know or understand what it takes to advance as a scientist. In the interests of science outreach, *The PhD Movie* and similar films should be distributed more widely to society at large (it can be viewed for a small fee at www.phdmovie.com).

We can use such films to start a conversation about how to nurture the practitioners and purveyors of science. With a good dose of sarcasm, *The PhD Movie* shows how graduate students spend much of their day second-guessing why they would spend another minute doing science. Another film, the 2009 documentary *Naturally Obsessed* (www.naturallyobsessed.com), shows how capable, aspiring science graduate students become 'drop-outs' and 'sell-outs' for reasons that have nothing to do with intelligence and drive.

Scientists complain about how we are portrayed in popular culture. Maybe it is time to start sharing who we are, and what we do. ■

Mariano A. Loza-Coll *is a postdoc in genetics at the Salk Institute for Biological Studies in La Jolla, California.*