

Predicting disease using genomics

John Bell

University of Oxford, Academic Centre, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK (e-mail: regius@medsci.ox.ac.uk)

Information from the human genome sequence will eventually alter many aspects of clinical practice. It will increase through our understanding of disease mechanisms, and guide the development of new drugs and therapeutic procedures. In the short term, however, knowledge of the genome will have a profound clinical impact on the diagnostic capability of clinical genetics laboratories. Molecular phenotyping using genetic and genomic information will allow early and more accurate prediction and diagnosis of disease and of disease progression. Medicine will become oriented towards disease prevention rather than efforts to cure people at late stages of illness.

Leprosy affects 700,000 people each year and is one of the major health problems in developing countries. *Mycobacterium leprae* attacks the nerves of the hands, feet and face of sufferers, and the lack of feeling in these parts leaves those infected prone to injuries and burns. The longer the disease is left untreated, the more likely it is that deformities will occur. Recently reported genetic research has provided a powerful insight into the cause of this disease, and common genetic risk factors have been identified¹. This information will guide future research aimed at controlling leprosy.

Similar successes in understanding genetic and genomic markers for asthma², cancer^{3–7}, diabetes^{8,9} and cardiovascular disease¹⁰ have been reported. These achievements suggest that new genomic tools will contribute to enhanced clinical practice for many common diseases. Translating this knowledge into routinely applied diagnostics will be a challenge, but will create a new approach to clinical practice with many benefits for patients.

For most diseases, medical treatment doesn't begin until the patient has visited their doctor with various symptoms. By this time it may well be late in the natural history of the disease, and therapeutic intervention may be limited to alleviating symptoms and slowing disease progression (Fig. 1). But in some areas of medicine, this standard approach has progressed to incorporate multiple diagnostic tests that precisely identify disease mechanisms, indicate the most appropriate type of therapeutic intervention, and evaluate therapeutic response and disease outcome. This is particularly evident in the management of infectious diseases such as HIV¹¹ and hepatitis C¹², in which viral load or disease resistance can be rapidly and efficiently measured to help inform therapy, and where diagnostics and therapeutics are closely coupled. In HIV, for example, viral load¹³ is an important indicator of response to therapy or relapse, and mutation screening can help to provide a molecular profile of drug resistance¹¹.

Many non-infectious common diseases may also benefit from a similarly close relationship between diagnostics and therapeutics. New genomic tools emerging from the Human Genome Project will improve our ability to identify individuals at risk or in the presymptomatic phase of diseases, and will more precisely define disease subtypes on the basis of their individual pathophysiology and their responsiveness to therapy. Predictive diagnosis or risk profiling should provide opportunities for environmental modification¹⁴ (such as smoking cessation), early therapy¹⁵ (for example, administering statins for individuals at risk of cardiovascular disease) or targeted cancer screening¹⁶ (for example, the use of colonoscopy in families or individuals at genetic

risk of colorectal cancer). Diagnostic medicine will become increasingly important as our understanding of disease susceptibility and progression markers improves, and as the tools for rapid and effective disease prediction and monitoring are developed.

The human genome has provided a host of opportunities to identify new disease markers. It is already possible to see the impact these tools can have on clinical practice.

Genetics as risk factors

Prediction, prevention and counselling of individuals at risk of genetic diseases have been aimed largely at single-gene disorders that have mendelian patterns of inheritance. These disorders offer the advantage that the identification of genetic variants responsible for disease can lead directly to clinically helpful and reasonably accurate prediction and diagnosis of disease.

Clinical genetics, however, will soon have to move onto aspects of genetics that are less 'deterministic' for a particular disease. Even apparently simple mendelian disorders may prove to have widely variable clinical phenotypes. For example, thalassaemia, an apparently simple genetic disease, has substantial complexities¹⁷. As a result of secondary genetic and environmental factors, individuals with exactly the same globin mutations may suffer either from a severe life-threatening disorder or be relatively unaffected. Another example is when genetic variants identified because of their relatively high penetrance (such as *BRCA1* in breast cancer¹⁸ and *CFTR* in cystic fibrosis¹⁹) have, on closer evaluation, not proved to have consistent levels of penetrance. Even single-gene disorders have a significant level of heterogeneity.

Additional complexity becomes apparent at a molecular level when clinical syndromes are analysed for their genetic causes. A good example of this is the syndrome of sudden cardiac death (SCD), which is often inherited in an autosomal dominant fashion and produces life-threatening cardiac arrhythmias (Fig. 2). Molecular dissection of this phenotype has revealed that multiple mutations of both potassium and sodium channels²⁰, mutations mediating heart muscle disease²¹, or even mutations in proteins anchoring channels to membranes²², can lead to this clinical phenotype. Although comparison across families is difficult, within individual families a single mutation can be tracked, a predictive diagnosis made and appropriate interventions (for example, an implantable cardiac defibrillator or the use of β -blockers to prevent fatal cardiac arrhythmias) can be introduced.

Many common diseases are mechanistically heterogeneous — usually only a small fraction of patients have forms of the disease that are mediated by single genes. For example, the

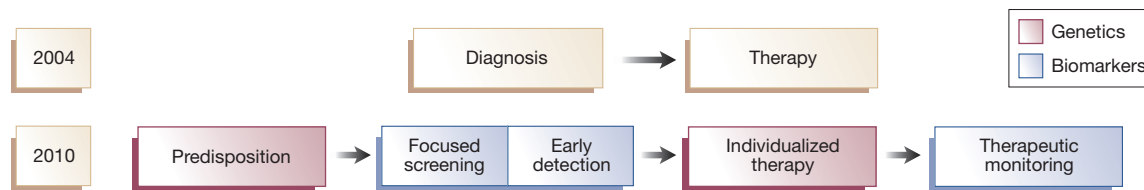


Figure 1 Diagnostic medicine. The role of diagnostics, driven by genetics and the discovery of biomarkers, will grow over the next ten years. Early diagnosis, targeted therapy and disease monitoring will replace the current paradigm of late-stage diagnosis and therapy.

genes for adenomatous polyposis coli (*APC*) and hereditary non-polyposis colorectal cancer (*HNPCC*) account for less than 5% of cases of colorectal cancer; similarly, the diabetes-mediating genes glucokinase (*GCK*), hepatic nuclear factor-1 α (*HNF1A*) and *HNF4A* are responsible for less than 5% of cases of diabetes^{23–26}. Usually, multiple genes and environmental factors mediate disease pathogenesis, and clinical definitions of disease obscure multiple mechanistically distinct subtypes⁸, creating a problem for the application of genetics. Clinical management of these diseases is usually reliant on predicting and managing the risk of disease outcome, the likelihood of a particular disease outcome, or therapeutic efficacy. The evaluation of risk factors defined by physiological, biochemical and radiological parameters is already a critical component of the management of most common diseases. Although the identification and clear validation of a few such risk factors (for example, smoking, hypertension and raised cholesterol levels) and the effect of people acting on that information have had an immense impact on our ability to reduce morbidity and premature death in the population^{9,15}, we still know too few of the risk factors. It is here that genetics may have its greatest impact. Disease-related genetic polymorphisms could be used, in combination with other factors, to define populations and individuals at risk. For example, major susceptibility determinants have been identified in diseases such as asthma², infectious conditions such as leprosy¹, and diabetes^{9,27}. These advantages for patients and practitioners will push clinical genetics towards testing for genetic variants in common diseases, despite their less-than-deterministic contribution to disease.

New approaches to monitoring disease through genomics

Until now, the tools of genetics have been restricted to a clinical setting, but the Human Genome Project has led to the development of substantial new technologies that are capable of defining large sets of biomarkers systematically in biological samples. The systematic approach of analysing all the products of the genome at a messenger RNA or a protein level has emerged from both new detection technology and the availability of the genome sequence^{28,29}. These methodologies, all currently being used in a research setting, will generate data on multiple biomarkers that vary quantitatively very early in disease, with disease onset, with disease progression or with therapeutic response. They may also provide other sets of prognostic factors or biological markers for underlying genetic variants³⁰.

mRNA patterns are strong prognostic markers in cancer, and can be used to define disease subtypes and predict response to therapy^{3–6}. Proteomics appears, even at its existing rudimentary stage, to be able to predict diseases such as cancer⁷, and metabonomics, a systematic approach to analysing small molecules, provides biomarker evidence of vascular disease¹⁰.

Several challenges need to be overcome before these techniques become routine diagnostic tools. Practitioners and regulators are not accustomed to assessing patterns of multiple markers, and so these methods may need to be simplified, as will the technology platforms for routine use. The first chip-based diagnostics are now available for predicting cytochrome P450 metabolism. This enzyme is responsible

for oxidative metabolism of many drugs and is therefore a useful indicator of therapy response. But these techniques will only take off if there is clear evidence of their clinical utility, and if this technology is supported by regulators and practitioners. Ultimately, the education of health care professionals will need to incorporate much more training so that the tools can be used effectively.

Pharmacogenetics

Improving therapeutic efficacy and reducing drug toxicity are two of the most important goals of genomics and genetics in clinical practice. Genetic information may in the future be used widely to more accurately target drugs and to improve their therapeutic usefulness. Consideration of individual variation in metabolism began with Garrod at the beginning of the twentieth century, and a few clear examples (response to suxamethonium, primaquine and debrisoquine) of genetically determined metabolic variation in response to drugs were defined in the 1950s and 1960s (ref. 31). Since then, some metabolic variants have been found to reduce the toxicity of many drugs, not just one (see review in this issue by Evans and Relling, page 464). Genetic variation in cytochrome P450 genes³², acetyltransferase genes, thiopurine methyltransferase³³ and dihydropyrimidine dehydrogenase³⁴ has clinical relevance because in each case it defines patient populations that metabolize drugs at different rates. The availability of genetic and genomic tools will provide important insights into the variation in response and toxicity to many drugs³⁵.

Genetic markers have been used to define some disease subtypes, identify drug targets and predict therapeutic response. For example, characterization of the cytogenetic abnormality that defines chronic myeloid leukaemia has led to the development of a specific drug to combat this disease³⁶ (see review in this issue by Strausberg *et al.*, page 469). The widespread genetic redefinition of disease subtypes is likely to provide many other examples of differential response to therapy. Molecular subtypes of type II diabetes, for example, show differing response to the drug sulphonylureas³⁷, Her2-positive breast cancer is sensitive to the new drug Herceptin³⁸, some molecularly defined subtypes of brain cancer have unusual sensitivity to chemotherapy³⁹, and subtypes of follicular lymphoma defined by transcript profiling are likely to be treatable with the anti-CD20 antibody Rituxan⁴⁰ (see review in this issue by Strausberg *et al.*, page 469).

It may also be possible to predict drug non-responders on the basis of polymorphisms in drug targets or pathways. This approach has been suggested for statin⁴¹, lipo-oxygenase inhibitor⁴² and nicotine-replacement therapy⁴³. Non-metabolic toxicity may be avoided by identifying groups that are particularly vulnerable, because of genetic factors, to complications such as cardiac arrhythmias (long QT genetic variants²⁰), venous thrombosis (factor V Leiden⁴⁴), and potentially fatal systemic drug reactions to abacavir⁴⁵ or carbamazepine³⁵ (HLA). Ultimately, the use of medication to reduce risk (for example, statins could be used to reduce the risk of vascular disease) must be driven by a more complete risk profile, including knowledge of genomic factors (Box 1). Together, these applications of pharmacogenetics will, if established in clinical practice, dramatically alter our use of pharmacological interventions. Few health care

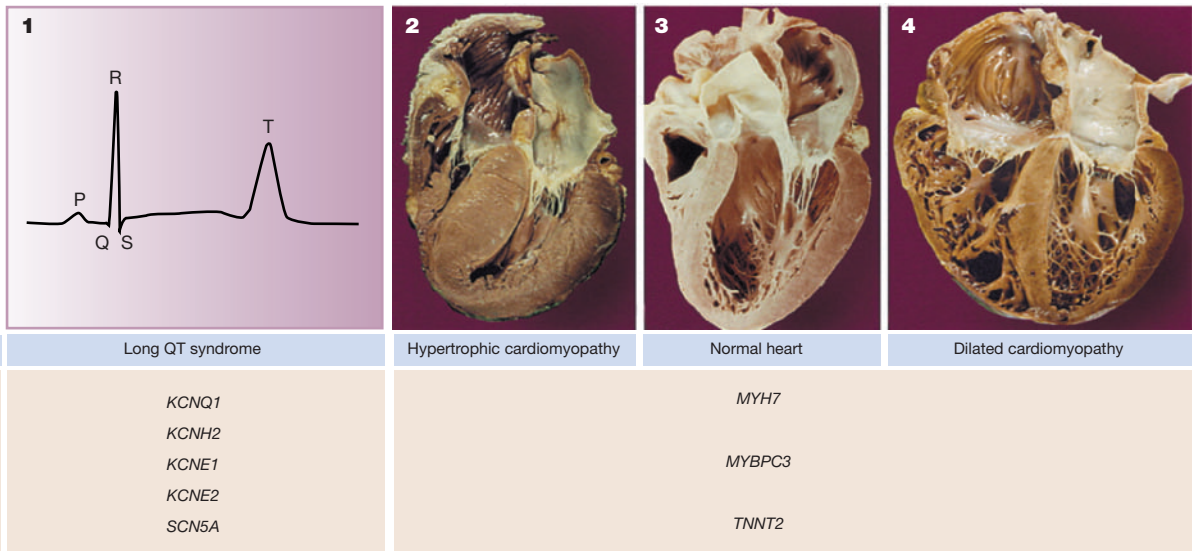


Figure 2 Predictive testing for complex genetic traits in families: sudden cardiac death. SCD arises from three distinct clinical syndromes: long QT syndrome, hypertrophic cardiomyopathy and dilated cardiomyopathy, shown in panels 1, 2 and 4, respectively. Inherited as dominant disorders, individuals at risk can be identified by cascade screening after the causative mutation has been found in a proband.

By screening the limited number of genes related to each of these clinical syndromes, up to 75% of relevant mutations will be detected, and mutation carriers can then be treated with interventions such as β -blockade or implantable cardiac defibrillators to prevent SCD. Panels 2, 3 and 4 reproduced, with permission, from ref. 46.

systems will allow patients to receive drugs that they are unlikely to benefit from, and new information will help to drive genomic diagnostics into routine practice.

From genome to clinical practice

Information valuable to patients and practitioners is emerging from genomic science. This science will need now to be transferred to a clinical setting, creating both technical and cultural challenges for health professionals. Much of the genomic data of clinical relevance generated so far are in a format that is inappropriate for diagnostic testing. The technology has not yet been established for rapid, inexpensive typing of most genomic biomarkers, with the exception of single-nucleotide polymorphisms (SNPs), and too little is known about which SNPs to type for non-mendelian disease. The HapMap, a project to determine common patterns of sequence variation in the genome, is unlikely to contribute to these clinical applications of genetics. This is because varying linkage disequilibrium patterns in different populations make it unsuitable for unsorted patient samples (see review in this issue by Carlson *et al.*, page 446). Simultaneous systematic analysis of larger numbers of biomarkers for disease predic-

tion is still required. These approaches are more suited to research than to routine diagnostic activity. Clear evidence of the predictive strength of these markers in clinical settings remains essential for their implementation.

Very large epidemiological population samples followed prospectively (over a period of years) and characterized for their biomarker and genetic variation will be necessary to demonstrate the clinical utility of these tools. National programmes to develop such genomic epidemiological studies are already in place (Biobank in the United Kingdom) or planned (Canada and China). Obstacles to the routine application of these data in clinical practice include a cultural gap between the approach to clinical practice that is currently employed and that which is possible with these new tools. Diagnostic medicine that includes predisposition testing, early detection, individualized therapy and therapeutic monitoring is neither systematically applied nor well taught in the current health care system. Its implementation will require not just clear data demonstrating its benefits, but also demand by patients and acceptance by health care professionals. This will not come quickly. This approach is also likely to put particular financial pressure on different components of the health care system. The opportunities for clinical genetics to become a mainstream component of clinical medicine are now apparent. This move to the clinic appears to be inevitable, but the transition may take a generation. □

Box 1

Toolbox for clinical genetics/genomics

Genetics/genomics will require the retooling of diagnostic laboratories to incorporate information into disease-prediction or disease-progression diagnostic tests. High-throughput tools will be necessary to test a limited number of markers that have been validated in large clinical studies.

- A DNA sequencing platform
- Comparative genome hybridization (CGH) arrays
- SNP detection technology (for example, dHPLC)
- Multiplex SNP scoring technology
- Transcript profiling technology
- Quantitative multiplex proteomic methodology
- Small-molecule detection assays

doi:10.1038/nature02624

1. Mira, M. T. *et al.* Susceptibility to leprosy is associated with *PARK2* and *PACRG*. *Nature* **427**, 636–640 (2004).
2. Allen, M. *et al.* Positional cloning of a novel gene influencing asthma from chromosome 2q14. *Nature Genet.* **35**, 258–263 (2003).
3. van de Vijver, M. J. *et al.* Gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
4. Weigelt, B. *et al.* Gene expression profiles of primary breast tumours maintained in distant metastases. *Proc. Natl Acad. Sci. USA* **100**, 15901–15905 (2003).
5. Wong, Y. F. *et al.* Expression genomics of cervical cancer: Molecular classification and prediction of radiotherapy response by DNA microarray. *Clin. Cancer Res.* **9**, 5486–5492 (2003).
6. Ross, M. E. *et al.* Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* **102**, 2951–2959 (2003).
7. Petricoin III, E. F. *et al.* Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572–577 (2002).
8. Bell, G. I. & Polonsky, K. S. Diabetes mellitus and genetically programmed defects in β -cell function. *Nature* **414**, 788–791 (2001).

9. Gloyn, A. L. *et al.* Large-scale association studies of variants in genes encoding the pancreatic β -cell KATP channel subunits Kir6.2 (*KCNJ11*) and SUR1 (*ABCC8*) confirm that the *KCNJ11* E23K variant is associated with type 2 diabetes. *Diabetes* **52**, 568–572 (2003).
10. Brindle, J. T. *et al.* Rapid and non-invasive diagnosis of the presence and severity of coronary heart disease using $^1\text{H-NMR}$ -based metabolomics. *Nature Med.* **8**, 1439–1444 (2002).
11. Paul, S. M. & Jorden, V. S. HIV resistance testing. A clinical tool. *N. J. Med.* **100**, 44–49 (2003).
12. Poynard, T., Yeun, M. F., Ratziu, V. & Lai, C. L. Viral hepatitis C. *Lancet* **362**, 2095–2100 (2003).
13. Young, J. *et al.* Stable partnership and progression to AIDS or death in HIV infected patients receiving highly active antiretroviral therapy: Swiss HIV cohort study. *Br. Med. J.* **328**, 15 (2004).
14. The U.S.–Venezuela Collaborative Research Project. Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc. Natl Acad. Sci. USA* **101**, 3498–3503 (2004).
15. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol-lowering with simvastatin in 5963 people with diabetes: a randomised placebo-controlled trial. *Lancet* **361**, 2005–2016 (2003).
16. Ponz de Leon, M. *et al.* Genetic testing among high-risk individuals in families with hereditary nonpolyposis colorectal cancer. *Br. J. Cancer.* **90**, 882–887 (2004).
17. Weatherall, D. J. Phenotype–genotype relationships in monogenic disease: lessons from the thalassaemias. *Nature Rev. Genet.* **2**, 245–255 (2001).
18. Ford, D. *et al.* Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families. *Am. J. Hum. Genet.* **62**, 676–689 (1998).
19. Grody, W. W. Molecular genetic risk screening. *Annu. Rev. Med.* **54**, 473–490 (2003).
20. Keating, M. T. & Sanguinetti, M. C. Molecular and cellular mechanisms of cardiac arrhythmias. *Cell* **104**, 569–580 (2001).
21. Watkins, H. Sudden death in hypertrophic cardiomyopathy. *N. Engl. J. Med.* **342**, 422–444 (2000).
22. Mohler, P. J. *et al.* Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. *Nature* **421**, 634–639 (2003).
23. Hartge, P. Genes, hormones, and pathways to breast cancer. *N. Engl. J. Med.* **348**, 2352–2354 (2003).
24. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
25. Fajans, S. S., Bell, G. I. & Polonsky, K. S. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N. Engl. J. Med.* **345**, 971–980 (2001).
26. Edlund, H. Transcribing diabetes. *Diabetes* **47**, 1817–1823 (1998).
27. Weedon, M. N. *et al.* Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *Am. J. Hum. Genet.* **73**, 1208–1212 (2003).
28. Su, A. I. *et al.* Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl Acad. Sci. USA* **99**, 4465–4470 (2002).
29. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
30. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).
31. Weber, W. W. *Pharmacogenetics* (Oxford Univ. Press, New York, 1997).
32. Evans, D. A. P. *et al.* Genetic control of isoniazid metabolism in man. *Br. Med. J.* **2**, 485–491 (1960).
33. Krynetski, E. Y. *et al.* A single point mutation leads to loss of catalytic activity in human thiopurine S-methyltransferase. *Proc. Natl Acad. Sci. USA* **92**, 949–953 (1995).
34. Harris, B. E., Carpenter, J. T. & Diasio, R. B. Severe 5-fluorouracil toxicity secondary to dihydropyrimidine dehydrogenase deficiency. A potential more common pharmacogenetic syndrome. *Cancer* **68**, 499–501 (1991).
35. Chung, W. H. *et al.* Medical genetics: a marker for Stevens–Johnson syndrome. *Nature* **428**, 486 (2004).
36. Savage, D. G. & Antman, K. H. Imatinib mesylate — A new oral targeted therapy. *N. Engl. J. Med.* **346**, 683–693 (2002).
37. Pearson, E. R. *et al.* Genetic cause of hyperglycaemia in response to treatment in diabetes. *Lancet* **362**, 1275–1281 (2003).
38. Fournier, M., Risio, M., Van Poznak, C. & Siedman, A. HER2 testing and correlation with efficacy of trastuzumab therapy. *Oncology* **16**, 1340–1348; 1351–1352; 1355–1358 (2002).
39. Esteller, M. *et al.* Inactivation of the DNA-repair gene *MGMT* and the clinical response of gliomas to alkylating agents. *N. Engl. J. Med.* **343**, 1350–1354 (2000).
40. Bohlen, S. P. *et al.* Variation in the gene expression patterns in follicular lymphoma and the response to rituximab. *Proc. Natl Acad. Sci. USA* **100**, 1926–1930 (2003).
41. Kuivenhoven, J. A. *et al.* The role of a common variant of the cholesteryl ester transfer protein gene in the progression of coronary atherosclerosis. *N. Engl. J. Med.* **338**, 86–93 (1998).
42. Drazen, J. M. *et al.* Pharmacogenetic association between *ALOX5* promoter genotype and the response to anti-asthma treatment. *Nature Genet.* **22**, 168–170 (1999).
43. Johnstone, E. C. *et al.* Genetic variation in dopaminergic pathways and short-term effectiveness of the nicotine patch. *Pharmacogenetics* (in the press).
44. Seligsohn, U. & Lubetsky, A. Genetic susceptibility to venous thrombosis. *N. Engl. J. Med.* **344**, 1222–1231 (2001).
45. Mallan, S. *et al.* Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* **359**, 727–732 (2002).
46. Seidman, J. G. & Seidman, C. The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms. *Cell* **104**, 557–567 (2001).

Competing interests statement The author declares competing financial interests: details accompany the paper on www.nature.com/nature.