# Clinical assessment incorporating a personal genome

*Euan A Ashley, Atul J Butte, Matthew T Wheeler, Rong Chen, Teri E Klein, Frederick E Dewey, Joel T Dudley, Kelly E Ormond, Aleksandra Pavlovic, Alexander A Morgan, Dmitry Pushkarev, Norma F Neff, Louanne Hudgins, Li Gong, Laura M Hodges, Dorit S Berlin, Caroline F Thorn, Katrin Sangkuhl, Joan M Hebert, Mark Woon, Hersh Sagreiya, Ryan Whaley, Joshua W Knowles, Michael F Chou, Joseph V Thakuria, Abraham M Rosenbaum, Alexander Wait Zaranek, George M Church, Henry T Greely, Stephen R Quake, Russ B Altman*

## Summary

**Background** The cost of genomic information has fallen steeply, but the clinical translation of genetic risk estimates remains unclear. We aimed to undertake an integrated analysis of a complete human genome in a clinical context.

**Methods** We assessed a patient with a family history of vascular disease and early sudden death. Clinical assessment included analysis of this patient's full genome sequence, risk prediction for coronary artery disease, screening for causes of sudden cardiac death, and genetic counselling. Genetic analysis included the development of novel methods for the integration of whole genome and clinical risk. Disease and risk analysis focused on prediction of genetic risk of variants associated with mendelian disease, recognised drug responses, and pathogenicity for novel variants. We queried disease-specific mutation databases and pharmacogenomics databases to identify genes and mutations with known associations with disease and drug response. We estimated post-test probabilities of disease by applying likelihood ratios derived from integration of multiple common variants to age-appropriate and sex-appropriate pre-test probabilities. We also accounted for gene-environment interactions and conditionally dependent risks.

**Findings** Analysis of 2·6 million single nucleotide polymorphisms and 752 copy number variations showed increased genetic risk for myocardial infarction, type 2 diabetes, and some cancers. We discovered rare variants in three genes that are clinically associated with sudden cardiac death—*TMEM43*, *DSP*, and *MYBPC3*. A variant in *LPA* was consistent with a family history of coronary artery disease. The patient had a heterozygous null mutation in *CYP2C19* suggesting probable clopidogrel resistance, several variants associated with a positive response to lipid-lowering therapy, and variants in *CYP4F2* and *VKORC1* that suggest he might have a low initial dosing requirement for warfarin. Many variants of uncertain importance were reported.

**Interpretation** Although challenges remain, our results suggest that whole-genome sequencing can yield useful and clinically relevant information for individual patients.

**Funding** National Institute of General Medical Sciences; National Heart, Lung And Blood Institute; National Human Genome Research Institute; Howard Hughes Medical Institute; National Library of Medicine, Lucile Packard Foundation for Children's Health; Hewlett Packard Foundation; Breetwor Family Foundation.

*Lancet* 2010; 375: 1525–35

See Comment page 1497

See Online/Viewpoint
DOI:10.1016/S0140-6736(10)60599-5

**Center for Inherited Cardiovascular Disease, Division of Cardiovascular Medicine** (E A Ashley MRCP, M T Wheeler MD, F E Dewey MD, J W Knowles MD, A Pavlovic BS), **Department of Medicine** (Prof R B Altman MD), **Department of Bioengineering** (S R Quake PhD, D Pushkarev, N F Neff PhD, Prof R B Altman), **Division of Medical Genetics** (Prof L Hudgins MD), **Department of Pediatrics** (A J Butte MD, R Chen PhD, J T Dudley, A A Morgan MS), **Howard Hughes Medical Institute** (S R Quake, D Pushkarev, N F Neff), **and Department of Genetics** (T E Klein PhD, K E Ormond MSc, C F Thorn PhD, M Woon BSE, L Gong PhD, L M Hodges PhD, D S Berlin PhD, K Sangkuhl PhD, J M Hebert MA, H Sagreiya, R Whaley BS, Prof R B Altman), **Stanford University School of Medicine, Stanford, CA, USA; Center for Law and Biosciences, Stanford University School of Law, Stanford, CA, USA** (Prof H T Greely JD); **Division of Clinical and Biochemical Genetics, Massachusetts General Hospital, Boston, MA, USA** (J V Thakuria MD); **and Department of Genetics, Harvard University, Boston, MA, USA** (G M Church PhD, M Chou PhD, J V Thakuria, A M Rosenbaum PhD, A W Zaranek PhD)

Correspondence to:
Dr Euan A Ashley, Falk Cardiovascular Research Building, 300 Pasteur Drive, Stanford, CA, 94305, USA
**euan@stanford.edu**

## Introduction

Technological advance has greatly reduced the cost of genetic information. However, the explanatory power and path to clinical translation of risk estimates for common variants reported in genome-wide association studies remain unclear. Much of the reason lies in the presence of rare and structural genetic variation. Since we are now able to rapidly and inexpensively sequence complete genomes,[1–5] comprehensive genetic risk assessment and individualisation of treatment might be possible.[6] However, present analytical methods are insufficient to make genetic data accessible in a clinical context, and the clinical usefulness of these data for individual patients has not been formally assessed. We aimed to undertake an integrated analysis of a complete human genome in a clinical context.

## Methods

### Patient

A patient with a family history of vascular disease and early sudden death was assessed at Stanford's Center for Inherited Cardiovascular Disease by a cardiologist (EAA) and a board-certified genetic counsellor (KEO). We took the patient's medical history and he was clinically assessed. A four-generation pedigree was drawn. In view of his family history, he underwent electrocardiography, an echocardiogram, and a cardiopulmonary exercise test.

### Genome analysis

Technical details of genome sequencing for this patient have been described previously.[7] In brief, genomic DNA was purified from 2 mL of whole blood and sequenced with a Heliscope (Helicos BioSciences, Cambridge, MA, USA) genome sequencer. We mapped sequence data to the National Center for Biotechnology Information reference human genome build 36 using the open-source aligner IndexDP (Helicos BioSciences, Cambridge, MA, USA).[7] Base calling was done with the UMKA algorithm.[7] A subset of single nucleotide polymorphism calls were independently validated with the Illumina BeadArray (San Diego, CA, USA) and all variants reported here and discussed with the patient were validated with Sanger

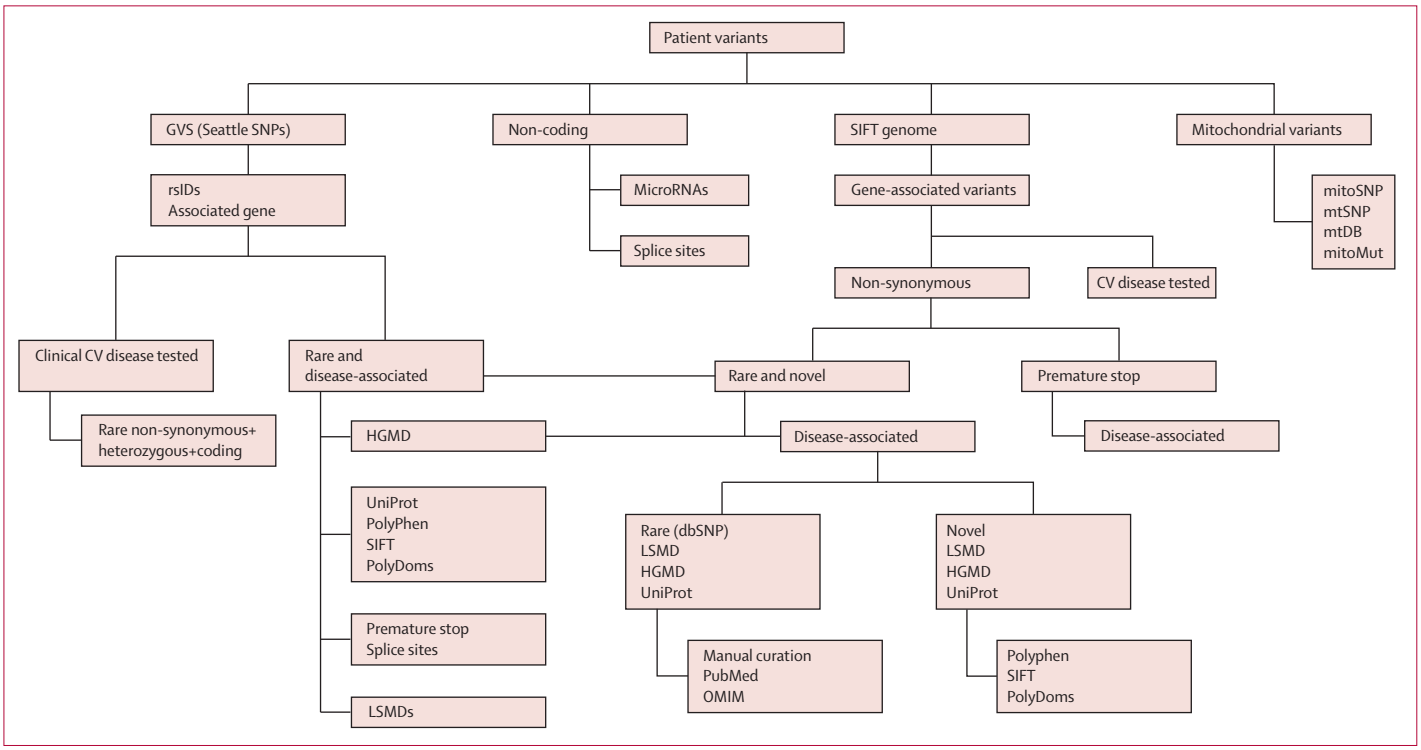



***Figure 1:*** **Approach to rare or novel variants**
CV=cardiovascular. GVS=Genome Variation Server. HGMD=Human Gene Mutation Database.[9] LSMD=locus-specific mutation databases. mtSNP=human mitochondrial genome polymorphism database.[11] OMIM=Online Mendelian Inheritance in Man. PolyDoms=mapping of human coding SNPs onto protein domains.[13] PolyPhen=polymorphism phenotyping.[10] rsID=reference sequence identification number. SIFT=Sorting Intolerant From Tolerant.[8] SNP=single nucleotide polymorphism. UniProt=universal protein resource.[12]

For more on **GVS** see http://gvs.gs.washington.edu/GVS/

sequencing. A subset of copy number variation calls were independently validated with digital PCR.

### Disease and risk analysis

Analysis focused on four areas: (i) variants associated with genes for mendelian disease; (ii) novel mutations; (iii) variants known to modulate response to pharmacotherapy; and (iv) single nucleotide polymorphisms previously associated with complex disease. Database queries, biophysical prediction algorithms, and analyses of non-coding regions were used to screen for rare and novel variants in the genome. We examined disease-specific mutation databases, the human genome mutation database, and Online Mendelian Inheritance in Man to identify genes and mutations with known associations to monogenic diseases. We applied prediction algorithms to weight the likelihood of variant pathogenicity on the basis of allele frequency, conservation, and protein domain disruption. Additionally, we developed algorithms to index variants affecting or creating start sites, stop sites, splice sites, and microRNAs (figure 1; webappendix p 2).[8–13]

For the **Online Mendelian Inheritance in Man** see http://www.ncbi.nlm.nih.gov/omim

For more on the **international HapMap project** see http://hapmap.ncbi.nlm.nih.gov/

For the **human genome mutation database** see http://www.hgvs.org/dblist/glsdb.html

See **Online** for webappendix

The Pharmacogenomics Knowledge Base (PharmGKB)[14] contains data for 2500 variants, of which 650 refer specifically to drug-response phenotypes. PharmGKB curators examined these 650 annotations in the context of the patient's genotype. Key variants were identified on the basis of the relevance of the phenotype in the annotation, the medical and family history, and the study population on which the annotation was based. Since our disease-risk estimation and pharmacogenomic analysis drew on previous reports, we rated the evidence used in one of three categories (webappendix p 2).

To integrate common variant genetic risk across a range of human disease, we built a manually curated disease and single-nucleotide-polymorphism database (webappendix p 2). Diseases and phenotypes were mapped to Unified Medical Language System Concept Unique Identifiers (webappendix p 3). Since strand direction was variably reported between studies, we identified strand direction by comparing with major or minor alleles in the appropriate HapMap population. Odds ratios were available for allele comparisons in most cases (webappendix p 7); however, to generate a medically relevant post-test probability of disease from integrated environmental and genetic risk, we calculated likelihood ratios (LRs) for the most important single nucleotide polymorphism from every haplotype block. Pre-test probability was derived from published sources (webappendix p 16) and the LR was applied to the pre-test odds of disease, which were calculated from age-appropriate and sex-appropriate population prevalence. Investigators did not always provide frequency data for genotype that allowed calculation of the LR.

The study was reviewed by the institutional review board of Stanford University and the patient gave written

| | Patient | Reference range |
|---|---|---|
| Age (years) | 40 | .. |
| Height (cm) | 180 | .. |
| Weight (kg) | 86 | .. |
| Body-mass index (kg/m²) | 26·5 | .. |
| Blood pressure | | .. |
| Systolic (mm Hg) | 128 | .. |
| Diastolic (mm Hg) | 80 | .. |
| Laboratory testing | | |
| Haemoglobin (mmol/L) | 9·7 | 8·4–11·0 |
| Creatinine (μmol/L) | 106·1 | <110 |
| Urea nitrogen (mmol/L) | 7·1 | 1·8–8·9 |
| Leucocyte count ($10^3$ per μL) | 4·9 | 4–11 |
| Cholesterol | | |
| Total (mmol/L) | 5·6 | .. |
| LDL (mmol/L) | 4·0 | .. |
| HDL (mmol/L) | 1·2 | .. |
| Triglycerides (mmol/L) | 0·8 | .. |
| High-sensitivity C-reactive protein (nmol/L) | <2 | <25 |
| Lipoprotein(a) (nmol/L) | 285 | <75 |
| Exercise testing | | |
| Maximum $VO_2$ (mL/kg per min) | 49·6 | .. |
| Maximum external work (W) | 450 | .. |
| $Ve/VCO_2$ slope | 26 | .. |
| Maximum heart rate (bpm) | 191 | .. |
| Resting cardiac output (L/min) | 6·3 | .. |
| Maximum cardiac output (L/min) | 24·5 | .. |
| Electrocardiography | | |
| Heart rate (bpm) | 60 | .. |
| QTc (ms) | 421 | .. |
| Echocardiography | | |
| Interventricular septum diastole (mm) | 10 | 6–11 |
| Left ventricular posterior wall diastole (mm) | 9·7 | 6–11 |
| Left ventricular internal diameter diastole (mm) | 45 | 37–57 |
| Ejection fraction by method of discs (%) | 63% | >55% |
| Aortic root diameter (mm) | 36 | 25–40 |
| Mitral inflow | | |
| E (cm/s) | 84 | .. |
| a (cm/s) | 53 | .. |

bpm=beats per minute. E=early diastolic peak velocity. a=late diastolic peak velocity due to atrial contraction.

***Table 1:*** **Clinical characteristics of the patient**

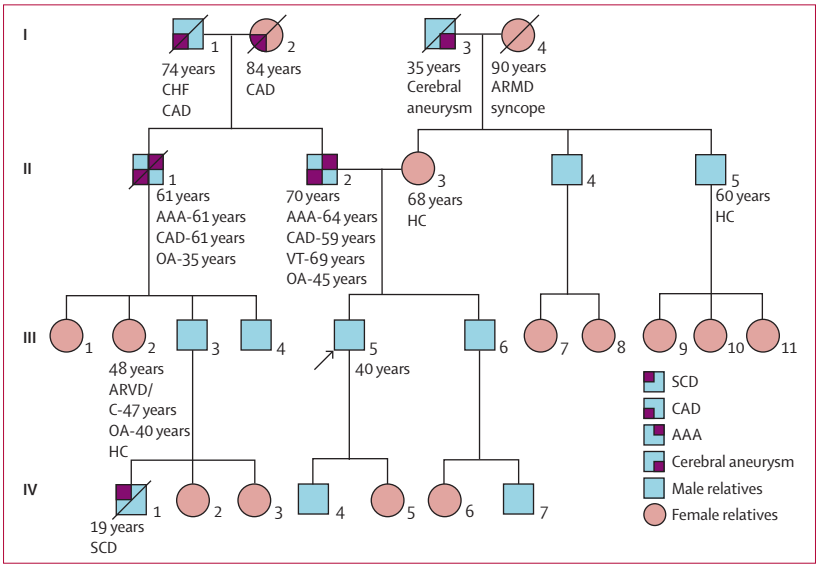



***Figure 2:*** **Patient pedigree**

The arrow shows the patient. Diagonal lines show relatives who are deceased. Years are age at death or diagnosis. AAA=abdominal aortic aneurysm. ARMD=age-related macular degeneration. ARVD/C=arrhythmogenic right-ventricular dysplasia or cardiomyopathy. CAD=coronary artery disease. CHF=congestive heart failure. HC=hypercholesterolaemia. OA=osteoarthritis. SCD=sudden cardiac death (presumed). VT=paroxysmal ventricular tachycardia.

consent. The patient received education and counselling before signing the consent form and throughout testing and follow-up.

### Role of the funding source

The study sponsors had no role in the design, data collection, data analysis, data interpretation, or writing of the report. EAA had full access to all data in the study and final responsibility for the decision to submit for publication.

## Results

The patient was a 40-year-old man who presented with a family history of coronary artery disease and sudden death. His medical history was not clinically significant and the patient exercised regularly without symptoms. He was taking no prescribed medications and appeared well. Clinical characteristics were within normal limits (table 1). Electrocardiography showed sinus rhythm, normal axis, and high praecordial voltage with early repolarisation. An echocardiogram revealed normal right and left ventricular size, systolic, diastolic, and valvular function. There were no wall motion abnormalities during maximum exercise and 1·5 mm ST depression was upsloping. Maximum oxygen uptake was 50 mL/kg per min. A four-generation family pedigree (figure 2) showed atherosclerotic vascular disease with several manifestations and prominent osteoarthritis. The patient's first cousin once removed (IV-1) died suddenly of an unknown cause.

Sequencing of genomic DNA resulted in an output of 148 GB of raw sequence, with an average read length of 33 bases.[7] Base calling detected 2·6 million single nucleotide polymorphisms and 752 copy number variations.

An important benefit of sequencing compared with DNA chip-based methods of genotyping is the identification of rare or novel variants. We searched for evidence of rare or novel variants that would predispose the patient or his family to disease (table 2; webappendix p 8).[8–10,12,15–27] Specific to cardiovascular disease, we discovered rare variants in three genes that are clinically

associated with sudden cardiac death—*TMEM43*, *DSP*, and *MYBPC3*. The *MYBPC3* variant, encoding an arginine-to-glutamine change at position 326 of the cardiac myosin-binding protein C, was originally associated with late-onset hypertrophic cardiomyopathy.[28] Subsequently, this variant has also been reported in several independent control populations without known hypertrophic cardiomyopathy,[29] suggesting that it might be benign. Mutations in *TMEM43*[30] or *DSP*[31] have been associated with familial arrhythmogenic right-

| | Amino-acid substitution | Gene name | Chromosome number | Position | SNP location | Reference base* | Patient genotype | Associated disease† | Mutation databases‡ | Functional prediction§ | Mode of disease-gene inheritance |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Previously described rare variants in genes associated with common disease** | | | | | | | | | | | |
| *LPA*[15,16] | I4399M¶ | Apolipoprotein A precursor, lipoprotein(a) | 6 | 160881127 | rs3798220 | T | C/T | Coronary artery disease | Associated with high lipoprotein(a) concentrations | Benign | NA |
| *FRZB*[17] | R200W | Frizzled-related protein | 2 | 183411581 | rs288326 | G | A/G | Osteoarthritis | Possibly associated with osteoarthritis‖ | Damaging | NA |
| **Previously described rare variants in genes associated with rare disease** | | | | | | | | | | | |
| *HFE* | H63D | Hereditary haemochromatosis protein precursor | 6 | 26199158 | rs1799945 | C | C/G | Haemochromatosis | Previously described, disease-associated | Damaging | Recessive, incomplete penetrance |
| *BTD*[20] | D444H | Biotinidase precursor | 3 | 15661697 | rs13078881 | G | C/G | Biotinidase deficiency | Previously described, intermediate phenotype | Damaging | Recessive |
| *SLC26A2*[21] | R492W | Solute carrier family 26 (sulphate transporter), member 2 | 5 | 149340823 | None | C | C/T | Diastrophic dysplasia | Disease-associated | Damaging | Recessive |
| *LAMB3*[22] | R635X | Laminin, β3 | 1 | 207865689 | None | G | A/G | Epidermolysis bullosa, junctional | Disease-associated, most common mutation | Truncated protein | Recessive |
| *SLC3A1*[23] | M467T | Solute carrier family 3 (cystine, dibasic, and neutral aminoacid transporters), member 1 | 2 | 44393296 | None | T | C/T | Cystinuria | Disease-associated, most common mutation | Damaging | Recessive |
| **Previously described variants of unknown importance in disease-associated genes** | | | | | | | | | | | |
| *TMEM43*[24] | M41V | Transmembrane protein 43 | 3 | 14146021 | None | A | A/G | ARVD/C | Reported in one of 150 probands with ARVD/C | Benign | Dominant, incomplete penetrance |
| *MYBPC3*[25] | R326Q | Myosin-binding protein C, cardiac-type | 11 | 47324447 | rs34580776 | C | C/T | Familial hypertrophic cardiomyopathy | Variant of unknown importance | Intermediate | Dominant, incomplete penetrance |
| **Novel variants potentially associated with rare disease** | | | | | | | | | | | |
| *DSP*[11] | R1838H | Desmoplakin | 6 | 7528007 | Novel | G | A/G | ARVD/C | Not found | Damaging | Dominant, incomplete penetrance |
| *CDC73*[26] | Q430X | Parafibromin | 1 | 191468879 | Novel | C | C/T | Hyperparathyroidism, jaw tumour | Not found | Truncated protein | Dominant, loss of heterozygosity |
| *CFTR*[27] | G458R | Cystic fibrosis transmembrane conductance regulator | 7 | 116976093 | Novel | G | A/G | Cystic fibrosis | Not found | Damaging | Recessive |
| *HFE2* | H174Y | Haemojuvelin precursor | 1 | 144127058 | Novel | C | C/T | Haemochromatosis, juvenile | Not found | Damaging | Recessive |

SNP=single nucleotide polymorphism. ARVD/C=Arrhythmogenic right-ventricular dysplasia or cardiomyopathy. *Reference allele in the human genome reference sequence, build 36.[7] †Disease associated with inherited mutations in the gene assessed. ‡Mutation databases were assessed for presence of the variant, including UniProt protein variant database,[12] Human Genome Mutation Database,[9] locus-specific mutation databases (curated by the Human Genome Variation Society), Online Mendelian Inheritance in Man, and clinical testing laboratory databases together with associated links. §Prediction of functional effect of mutation, derived from the substitution effect prediction algorithms, Polymorphism Phenotyping[10] and Sorting Intolerant from Tolerant;[8] in-vitro experimental evidence; and assessment of typical mutational mechanisms in other disease gene-associated mutations. ¶Also reported as I1891M; every copy of C *allele* increases lipoprotein(a) concentration 1·8 SD and risk of coronary artery disease two-to-three fold. ‖Inconclusive association in meta-analysis of osteoarthritis-related SNPs, but moderate association with severe hip osteoarthritis.

*Table 2:* **Selected rare non-synonymous variants in genes associated with inherited disease**

ventricular dysplasia or cardiomyopathy. Review of previous clinical assessment of extended family members revealed minor criteria for this disease in one first cousin, whose son died suddenly in his teens. By contrast with the findings for the identified rare *MYBPC3* variant, the *TMEM43* variant, encoding a methionine-to-valine change at position 41 of transmembrane protein 43, has not been previously published, but was seen in one of 150 probands with arrhythmogenic right-ventricular dysplasia or cardiomyopathy.[24] The identified *DSP* variant, encoding an arginine-to-histidine change to aminoacid 1838 of the desmoplakin protein, is entirely novel. Control populations from clinical testing laboratories (more than 1000 total chromosomes) did not contain either the *DSP* or *TMEM43* variants.

Analysis of the patient's genome revealed three novel and potentially damaging variants in two related genes that were previously associated with development of haemochromatosis. Subsequent to these findings, detailed review of personal and family history did not identify a history of haemochromatosis in the patient or family members. Echocardiogram results and liver function tests did not show evidence of the disease. Justification for continued surveillance and testing with serum iron studies was explored with the patient. Additionally, the patient had a novel stop mutation in a gene implicated in hyperparathyroidism and parathyroid

| | Gene name | SNP location | Patient genotype | Drug(s) affected | Summary of effects | Level of evidence |
|---|---|---|---|---|---|---|
| *SLCO1B1* | Solute carrier organic anion transporter family, member 1B1 | rs4149056 | T/T | HMG-CoA reductase inhibitors (statins) | No increased risk of myopathy | High[32–34] |
| *CYP2C19* | Cytochrome P450, family 2, subfamily C, polypeptide 19 | rs4244285 | A/G | Clopidogrel and CYP2C19 substrates | CYP2C19 poor metaboliser; many drugs might need adjustment | High[35] |
| *VKORC1* | Vitamin K epoxide reductase complex, subunit 1 | rs9923231 | C/T | Warfarin | Reduced dose needed | High[36] |
| *CYP4F2* | Cytochrome P450, family 4, subfamily F, polypeptide 2 | rs2108622 | C/C | Warfarin | Reduced dose needed | High[37] |
| *ADRB1* | β1 adrenergic receptor | rs1801252 | A/A | Atenolol, metoprolol | Might be preferable to calcium-channel blockers | High[38,39] |
| *SLCO1B1* | Solute carrier organic anion transporter family, member 1B1 | rs11045819 | A/C | Fluvastatin | Good response | Medium[40] |
| *HMGCR* | HMG-CoA reductase | rs17238540 | T/T | Pravastatin | Patient might have good response | Medium |
| *HMGCR* | HMG-CoA reductase | rs17244841 | A/A | Pravastatin, simvastatin | No reduced efficacy | Medium |
| *ADRB2* | β2 adrenergic receptor, surface | rs1042713 | A/G | β blockers | Other treatment options might be preferable | Medium[41] |
| *ADRB2* | β2 adrenergic receptor, surface | rs1042714 | C/C | β blockers | Other treatment options might be preferable | Medium[41,42] |
| *CYP2D6* | Cytochrome P450, family 2, subfamily D, polypeptide 6 | rs3892097 rs1800716 | C/C | Metoprolol and other CYP2D6 substrates | Normal CYP2D6 metaboliser | Medium[43] |
| *CDKN2A/B* | Cyclin-dependent kinase inhibitor 2A/2B | rs10811661 | T/T | Metformin | Reduced likelihood of response | Medium[44] |
| *CDKN2A/B* | Cyclin-dependent kinase inhibitor 2A/2B | rs10811661 | T/T | Troglitazone | Reduced likelihood of response | Medium[44] |

SNP=single nucleotide polymorphism. HMG-CoA=3-hydroxy-3-methylglutaryl-coenzyme A.

***Table 3:*** **Pharmacogenomic variants with summary of effects and level of evidence**

| | Gene name | SNP location | Patient genotype | Drug(s) affected | Effect type | Coding change |
|---|---|---|---|---|---|---|
| *NOD2* | Nucleotide-binding oligomerisation domain containing 2 | 16:49303700 | A/G | Infliximab | Pharmacodynamic | V793M |
| *NOD2* | Nucleotide-binding oligomerisation domain containing 2 | 16:49302615 | C/T | Infliximab | Pharmacodynamic | S431L |
| *SLC15A1* | Solute carrier family 15 (oligopeptide transporter), member 1 | 13:98176691 | C/T | Atorvastatin, fluvastatin, HMG-CoA reductase inhibitors, lovastatin, pravastatin, rosuvastatin, simvastatin | Pharmacokinetic | Y21C |
| *HLA-DRB5* | MHC class II, DR beta 5 | 6:32593811 | T/T | Clozapine | Pharmacodynamic | T262K |
| *MICA* | MHC class I polypeptide-related sequence A | 6:31484467 | C/T | Mercaptopurine, methotrexate | Pharmacodynamic | I14T |
| *SLC22A8* | Solute carrier family 22 (organic anion transporter), member 8 | 11:62517376 | C/T | Cimetidine, estrone, anti-inflammatory and antirheumatic products, non-steroids, ibuprofen, indometacin, ketoprofen, methotrexate, phenylbutazone, piroxicam, probenecid, atorvastatin, fluvastatin, HMG-CoA reductase inhibitors, lovastatin, pravastatin, rosuvastatin, simvastatin, adefovir dipivoxil, tenofovir, antineoplastic agents, cyanocobalamin, folic acid, folinic acid, pyridoxine | Pharmacokinetic | R534Q |

SNP=single nucleotide polymorphism. HMG-CoA=3-hydroxy-3-methylglutaryl-coenzyme A. *Predicted to be damaging by PhD-SNP algorithm.[45]

***Table 4:*** **Pharmacogenomic rare and novel non-synonymous damaging variants***

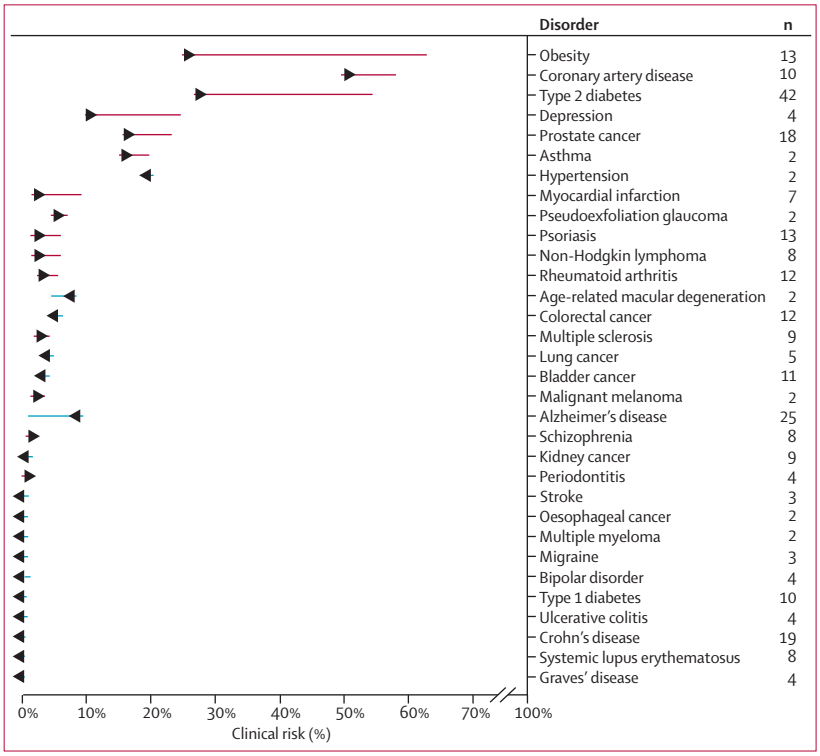



***Figure 3:*** **Clinical risk incorporating genetic-risk estimates for major diseases**
We calculated post-test probabilities by multiplying reported pre-test probabilities or disease prevalence (in white men in the patient's age range; webappendix p 16) with a series of independent likelihood ratios for every patient allele. Only 32 diseases with available pre-test probabilities, more than one associated single nucleotide polymorphism, and with reported genotype frequencies are shown. Disorders such as abdominal aortic aneurysm and progressive supranuclear palsy are not listed, because they have only one available single nucleotide polymorphism. Backs of the arrowheads show pre-test probabilities and arrows point in the direction of change in probability. Blue lines show lowered post-test probabilities, and red increased post-test probabilities. n=number of independent single nucleotide polymorphisms used in calculation of post-test probability for that disorder.

tumours. This variant might increase probability of future development of hyperparathyroidism or parathyroid tumours through a loss-of-heterozygosity mechanism. Consistent with a variant in a gene previously associated with osteoarthritis, there was a family history of osteoarthritis and the patient reported chronic knee pain without a formal diagnosis.

We noted 63 clinically relevant previously described pharmacogenomic variants (table 3, table 4; webappendix p 11)[32–45] and six novel, non-conservative, aminoacid-changing single nucleotide polymorphisms in genes that are important for drug response. There was a heterozygous null mutation in *CYP2C19*, the gene product of which is important for metabolism of many drugs, including proton-pump inhibitors, antiepileptic drugs, and the antiplatelet agent clopidogrel. Notably, the rate of cardiovascular events is raised in patients with *CYP2C19* loss-of-function mutations who take clopidogrel.[46] Additionally, the patient had two types of distinct genetic variations related to decreased maintenance dosing of warfarin. The patient had the single most important variant in *VKORC1* associated with a low maintenance dose,[47] and was homozygous for a *CYP4F2* single nucleotide polymorphism that is associated with reduced dosing.[48] Thus, if prescription of warfarin became necessary, loading could be individually tailored for this patient, with lowered expected doses. The patient had several variants that are associated with good response to statins (including reduced risk of myopathy) and one variant suggesting that he might need a raised dose to achieve a good response. Finally, the patient was wild type (with no copy number variations) for genes for important drug-metabolising enzymes (*CYP2D6*, *CYP2C9*, and *CYP3A4*) affecting hundreds of drug responses.

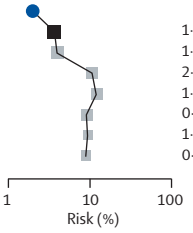
Although genome-wide association studies have provided strong association of many common variants with disease, integration of these small odds ratios in the context of the individual patient remains challenging. In particular, additive or multiplicative models of even strongly associated single nucleotide polymorphisms can add little to the classified status of the patient.[49,50] Furthermore, these approaches take no account of previous probability of disease. To counter some of these concerns, we adopted established methods from within evidence-based medicine that have rarely been applied to clinical genetics. We estimated pre-test probabilities from referenced sources for 121 diseases (webappendix p 7). Of the 55 diseases for which we could estimate a post-test probability, genetic risk was consistently increased (LR >2) for eight diseases and decreased (<0·5) for seven diseases (figure 3). The advantage of plotting pre-test and post-test probabilities is shown by several examples—eg, although the patient has increased genetic risk for Graves' disease, because the pre-test probability of this disease is very low, post-test probability also remains low. Conversely, although the patient has a low genetic contribution to his risk for prostate cancer, his estimated pre-test probability is high, resulting in a high overall post-test probability.

Raised genetic risk did not always translate into high post-test probability. Post-test probabilities that were an order of magnitude higher or lower than pre-test probabilities were rare. Any decision towards acting on these predictions will necessarily be a function of the post-test probability threshold for action (eg, the post-test probability of type 2 diabetes), the consequences of action (eg, regular testing for fasting blood sugar), and the usefulness and effectiveness of action.

***Figure 4:*** **Contribution of individual alleles to overall risk of myocardial infarction (A), type 2 diabetes (B), prostate cancer (C), and Alzheimer's disease (D)**
We ordered single nucleotide polymorphisms (SNPs) with associations established from genome-wide association studies in decreasing order of sample size and number of studies showing association. Darkest colours show polymorphisms with the most studies reporting association with disease, and size of boxes scales with the logarithm of the number of samples used to calculate the likelihood ratio (LR). SNPs at the top of every graph are reported in the most and largest studies, and we have the most confidence in their association with disease. We calculated test probabilities using the pre-test estimate as a starting point, and serially stepping down the list of SNPs and calculating an updated post-test probability including the contribution of that genotype. *Gene related to the SNP, if known. †Number of studies reporting an association. ‡Number of samples used to calculate the LR.

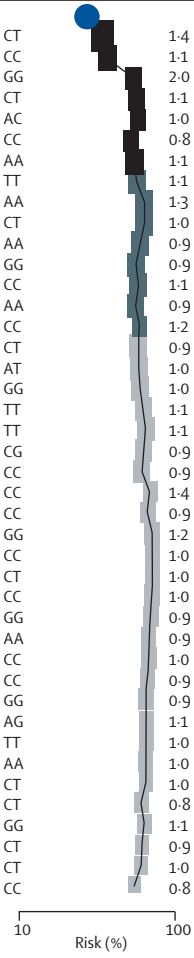
**A** Myocardial infarction

| Gene* | SNP location | Patient genotype | LR | Studies† | Samples‡ | Post-test probability (%) |
|---|---|---|---|---|---|---|
| | | | | | | 2·0% |
| LPA | rs3798220 | CT | 1·86 | 2 | 17031 | 3·7% |
| THBS2 | rs8089 | AC | 1·09 | 1 | 4868 | 4·0% |
| LDLR | rs14158 | GG | 2·88 | 1 | 3542 | 10·6% |
| LIPC | rs11630220 | AG | 1·15 | 1 | 3542 | 12·0% |
| ESR2 | rs1271572 | CC | 0·73 | 1 | 3089 | 9·1% |
| | rs35410698 | GG | 1·03 | 1 | 1094 | 9·4% |
| FXN | rs3793456 | AA | 0·94 | 1 | 1094 | 8·9% |



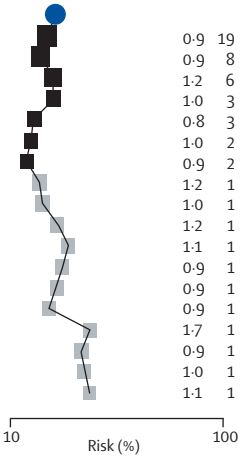

**B** Type 2 diabetes

| Gene* | SNP location | Patient genotype | LR | Studies† | Samples‡ | Post-test probability (%) |
|---|---|---|---|---|---|---|
| | | | | | | 27% |
| TCF7L2 | rs7903146 | CT | 1·4 | 22 | 126642 | 34% |
| SLC30A8 | rs13266634 | CC | 1·1 | 3 | 7629 | 37% |
| KLF11 | rs35927125 | GG | 2·0 | 3 | 6944 | 54% |
| TCF7L2 | rs7901695 | CT | 1·1 | 3 | 4031 | 56% |
| EPO | rs1617640 | AC | 1·0 | 3 | 2572 | 57% |
| PPARGC1A | rs8192678 | CC | 0·8 | 3 | 2386 | 52% |
| RAI1, SREBF1 | rs11868035 | AA | 1·1 | 2 | 8606 | 54% |
| | rs10811661 | TT | 1·1 | 2 | 8147 | 56% |
| CDKAL1 | rs7756992 | AA | 1·3 | 2 | 8019 | 63% |
| | rs564398 | CT | 1·0 | 2 | 8019 | 63% |
| IGF2BP2 | rs1470579 | AA | 0·9 | 2 | 8019 | 60% |
| CDKAL1 | rs7754840 | GG | 0·9 | 2 | 8019 | 56% |
| PPARG | rs1801282 | CC | 1·1 | 2 | 5199 | 57% |
| ENPP1 | rs1044498 | AA | 0·9 | 2 | 4972 | 55% |
| PTGES2 | rs13283456 | CC | 1·2 | 2 | 1665 | 59% |
| GCKR | rs780094 | CT | 0·9 | 1 | 8769 | 57% |
| FTO | rs9939609 | AT | 1·0 | 1 | 8717 | 58% |
| LOC100129623, WFS1 | rs734312 | GG | 1·0 | 1 | 8069 | 59% |
| AHSG | rs2518136 | TT | 1·1 | 1 | 6110 | 62% |
| AHSG | rs2077119 | TT | 1·1 | 1 | 6110 | 64% |
| PYY | rs1058046 | CG | 0·9 | 1 | 5965 | 62% |
| GCK | rs1799884 | CC | 0·9 | 1 | 4433 | 61% |
| LMNA | rs547915 | CC | 1·4 | 1 | 3017 | 68% |
| ADIPOQ | rs266729 | CC | 0·9 | 1 | 2864 | 66% |
| PHF23 | rs222852 | GG | 1·2 | 1 | 2335 | 70% |
| MECR | rs10915239 | CC | 1·0 | 1 | 2335 | 71% |
| PRKAR2B | rs2395836 | CT | 1·0 | 1 | 2335 | 71% |
| CBLB | rs17280845 | CC | 1·0 | 1 | 2335 | 70% |
| | rs11206883 | GG | 0·9 | 1 | 2335 | 69% |
| ARID2 | rs11183212 | AA | 0·9 | 1 | 2335 | 67% |
| SLC2A2 | rs10513684 | CC | 1·0 | 1 | 2335 | 68% |
| PPARGC1A | rs2970871 | CC | 0·9 | 1 | 2335 | 66% |
| PTPN22 | rs2476601 | GG | 0·9 | 1 | 2000 | 64% |
| LOC387761 | rs7480010 | AG | 1·1 | 1 | 1937 | 65% |
| | rs1256517 | TT | 1·0 | 1 | 1937 | 65% |
| MMP26 | rs2499953 | AA | 1·0 | 1 | 1937 | 64% |
| | rs932206 | CT | 1·0 | 1 | 1937 | 65% |
| | rs659366 | CT | 0·8 | 1 | 1686 | 60% |
| | rs10823406 | GG | 1·1 | 1 | 1257 | 63% |
| | rs729287 | CT | 0·9 | 1 | 1129 | 61% |
| KCNJ11 | rs5219 | CT | 1·0 | 1 | 1034 | 60% |
| | rs1884613 | CC | 0·8 | 1 | 531 | 54% |



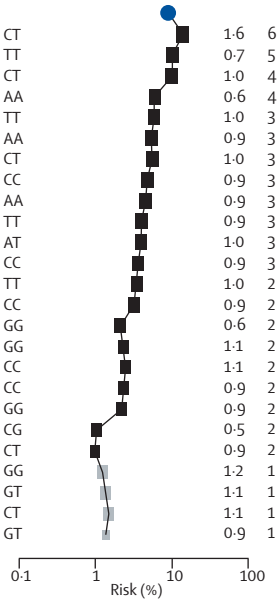

**C** Prostate cancer

| Gene* | SNP location | Patient genotype | LR | Studies† | Samples‡ | Post-test probability (%) |
|---|---|---|---|---|---|---|
| | | | | | | 16% |
| | rs1447295 | CC | 0·9 | 19 | 56485 | 15% |
| TNRC6B | rs9623117 | TT | 0·9 | 8 | 35869 | 14% |
| DAB2IP | rs1571801 | GT | 1·2 | 6 | 13997 | 16% |
| | rs6983267 | GT | 1·0 | 3 | 3985 | 16% |
| CDH1 | rs16260 | CC | 0·8 | 3 | 2238 | 13% |
| | rs6983561 | AA | 1·0 | 2 | 1846 | 12% |
| | rs1551512 | TT | 0·9 | 2 | 1846 | 12% |
| MMP2 | rs1477017 | AG | 1·2 | 1 | 2878 | 13% |
| HIF1A | rs11549465 | CC | 1·0 | 1 | 2878 | 14% |
| MMP2 | rs11639960 | AG | 1·2 | 1 | 2878 | 16% |
| RSR2 | rs2987983 | AG | 1·1 | 1 | 2216 | 18% |
| TLR10 | rs4129009 | TT | 0·9 | 1 | 2163 | 17% |
| TLR10 | rs4274855 | CC | 0·9 | 1 | 2163 | 16% |
| TLR1 | rs5743604 | AA | 0·9 | 1 | 2163 | 15% |
| | rs7837688 | GT | 1·7 | 1 | 2139 | 23% |
| | rs4242382 | GG | 0·9 | 1 | 2139 | 21% |
| | rs10086908 | TT | 1·0 | 1 | 2139 | 22% |
| | rs7000448 | TT | 1·1 | 1 | 1012 | 23% |




**D** Alzheimer's disease

| Gene* | SNP location | Patient genotype | LR | Studies† | Samples‡ | Post-test probability (%) |
|---|---|---|---|---|---|---|
| | | | | | | 9·0% |
| TOMM40 | rs157581 | CT | 1·6 | 6 | 7740 | 13·90% |
| DAPK1 | rs4878104 | TT | 0·7 | 5 | 10397 | 10·19% |
| TRAK2 | rs13022344 | CT | 1·0 | 4 | 6512 | 10·12% |
| DAPK1 | rs4877365 | AA | 0·6 | 4 | 4841 | 5·89% |
| E8F3 | rs11016976 | TT | 1·0 | 3 | 5736 | 5·87% |
| TNK1 | rs1554948 | AA | 0·9 | 3 | 5736 | 5·32% |
| MYH13 | rs2074877 | CT | 1·0 | 3 | 5366 | 5·55% |
| GALP | rs3745833 | CC | 0·9 | 3 | 5366 | 4·82% |
| PCK1 | rs8192708 | AA | 0·9 | 3 | 5366 | 4·47% |
| | rs1859849 | TT | 0·9 | 3 | 5304 | 4·02% |
| | rs11622883 | AT | 1·0 | 3 | 5248 | 3·97% |
| WWC1 | rs17070145 | CC | 0·9 | 3 | 2545 | 3·65% |
| LMNA | rs505058 | TT | 1·0 | 2 | 4646 | 3·49% |
| ACAN | rs2882676 | CC | 0·9 | 2 | 4590 | 3·22% |
| PGBD1 | rs3800324 | GG | 0·6 | 2 | 4590 | 2·11% |
| GOLM1 | rs10868366 | GG | 1·1 | 2 | 2156 | 2·30% |
| GOLM1 | rs7019241 | CC | 1·1 | 2 | 2156 | 2·49% |
| | rs9886784 | CC | 0·9 | 2 | 2156 | 2·36% |
| | rs10519262 | GG | 0·9 | 2 | 2156 | 2·22% |
| | rs463946 | CG | 0·5 | 2 | 1922 | 1·04% |
| PLAU | rs2227564 | CT | 0·9 | 2 | 956 | 0·98% |
| ADAM12 | rs1278279 | GG | 1·2 | 1 | 2320 | 1·23% |
| SORL1 | rs2070045 | GT | 1·1 | 1 | 2031 | 1·36% |
| ABCA1 | rs2230806 | CT | 1·1 | 1 | 1691 | 1·50% |
| PSEN1 | rs165932 | GT | 0·9 | 1 | 170 | 1·37% |



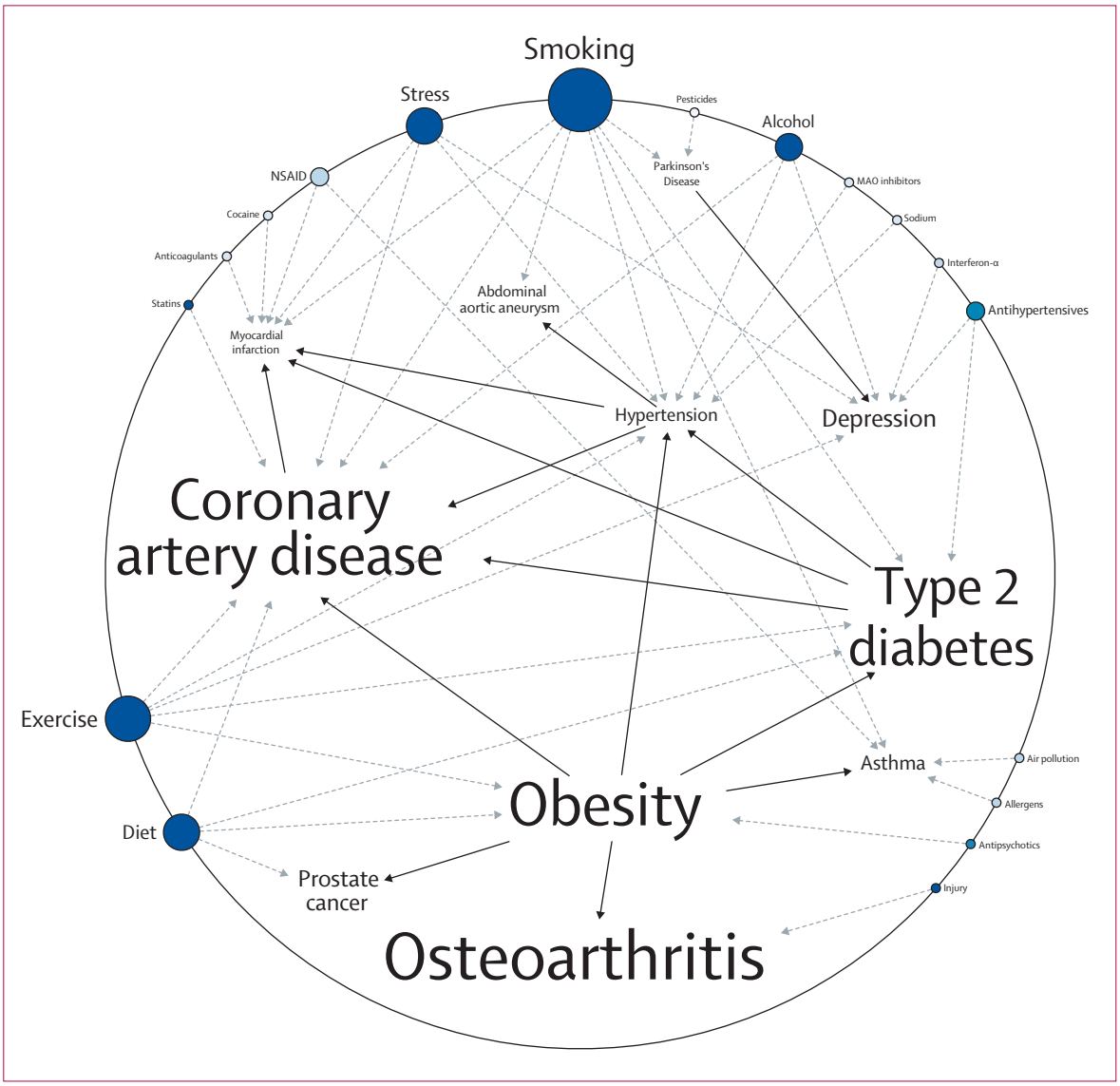



***Figure 5:*** **Gene-environment interaction**
A conditional dependency diagram for diseases represented in the patient's genetic-risk profile. Only diseases for which calculable post-test risk probabilities were greater than 10% are shown. For every disease, text size is proportional to post-test risk probability. Solid black arrows are shown between disease names if one disease predisposes a patient to the other. Environmental factors that are potentially modifiable are shown around the circumference, and dashed arrows are shown between an environmental factor and a disease if the factor has been frequently reported in association with the cause of the disease. Text and circle sizes for environmental factors are proportional to the number of diseases that each factor is associated with in the circuit. Colour intensity of the circle for each environmental factor represents maximum post-test risk probability amongst diseases directly associated with that factor. NSAID=non-steroidal anti-inflammatory drug. MAO=monoamine oxidase.

Increased genetic risk for myocardial infarction took the form of five single nucleotide polymorphisms associated with susceptibility to myocardial infarction and two protective polymorphisms (figure 4). The patient also had risk markers at the locus (9p21) that is most replicated in genome-wide association studies (an example is rs1333049, which is associated with an odds ratio of 1·5 for early onset myocardial infarction[51]—this marker is part of a commercial genetic risk test for myocardial infarction). Furthermore, the patient had one copy of the previously studied variant of *LPA* encoding the apolipoprotein A precursor. Notably, the patient had a very high lipoprotein(a) concentration (285 nmol/L, reference value <75 nmol/L; table 1), which is associated with increased risk of cardiovascular events. This variant is associated with a five-fold increased median plasma lipoprotein(a) concentration, a 1·7 to two-fold[15] increased risk of coronary artery disease, and a three-fold[16] adjusted odds ratio versus non-carriers for severe coronary artery disease. This

polymorphism has been associated with a low number of kringle IV-2 (KIV-2) domain repeats in *LPA*, high lipoprotein(a) concentrations, and adverse cardiovascular events.[52,53] Because of the technical limitations of short-read sequencing, a precise estimate of the number of KIV-2 domains in the patient's genome sequence was not established.

We placed disease-associated genetic risk into the context of environmental and behavioural modifiers, as well as predisposing disorders (figure 5). Diseases that might be independently associated with low genetic risk (eg, abdominal aortic aneurysm) were assessed in the context of others that could be causally related but for which genetic risk might be higher (eg, obesity, which predisposes to type 2 diabetes and hypertension). Thus, overall risk could then be assessed with both direct and conditionally dependent information because they were shown together in the circuit. For example, we predicted a reduced risk probability for hypertension of 16·8% (LR 0·81) relative to the general population; however, the patient had a substantially raised genetic risk of obesity (LR 6·28), imparting a high post-test risk of 56·1% for a predisposing risk factor for hypertension. Furthermore, hypertension is associated with several modifiable environmental factors affecting risk either directly (eg, sodium intake) or conditionally by association with another node in the circuit (eg, antipsychotic drugs). Although no methods exist for statistical integration of such conditionally dependent risks, interpretation of findings in the context of the causal circuit diagram allows assessment of the combined effect of environmental and genetic risk for every individual.

During genetic counselling, we discussed the possibility that clinical assessment incorporating a personal genome might uncover high risk of a serious disease, including some for which there is no treatment. Additionally, we described the reproductive implications of heterozygous status for autosomal recessive diseases such as cystic fibrosis, which might not be predictable from family history (table 2, figure 1). We also warned of increases or decreases in genetic risk for common diseases. We noted that most of the sequence information is difficult to interpret, and discussed error rates and validation processes. Additionally, we discussed that risk alleles might be discovered that have reproductive or familial importance rather than personal importance (such as those for breast or ovarian cancer). We addressed the possibility of discrimination on the basis of genetics. Although a specialised physician can provide information for a patient seeking a genetic test for a specific disease, patients with whole genome sequence data need information about more diseases with a wide clinical range (table 2). For this reason, we offered extended access to clinical geneticists, genetic counsellors, and clinical lab directors to interpret the information we presented.

## Discussion

We provide an approach to comprehensive analysis of a human genome in a defined clinical context. We assessed whole-genome genetic risk, focusing on variants in genes that are associated with mendelian disease, novel and rare variants across the genome, and variants of pharmacogenomic importance. Additionally, we developed an approach to the integration of disease risk across several common polymorphisms. Although the methods that we used are nascent, the results provide proof of principle that clinically meaningful information can be derived about disease risk and response to drugs in patients with whole genome sequence data.

Prominent aspects of the patient's family history (figure 1) were diagnosis of arrythmogenic right ventricular dysplasia or cardiomyopathy in his first cousin (III-3) and the sudden death of his first cousin once removed (IV-1). Our patient shares 12·5% of his genetic information with his first cousin and 6·25% with that relative's son and, although a diagnostic workup would involve targeted sequencing of DNA from these individuals, our analysis uncovered several variants in genes with potential explanatory value. Most were common variants. One gene variant (in *MYBPC3*) was previously associated with hypertrophic cardiomyopathy, but seems to be a common variant; this exemplifies the limitations of present variant databases. Two rare variants in genes (*TMEM43, DSP*) previously associated with arrythmogenic right ventricular dysplasia or cardiomyopathy were novel.

Our patient reported a prominent family history of vascular disease including aortic aneurysm and coronary artery disease (figure 2; individuals II-1, II-2, I-1, I-2). During estimation of the risk of coronary artery disease, we integrated the most replicated risk associations, likelihood ratio projections from published work, and a known variant in *LPA* that might not have been identified with chip-based genotyping. According to adult treatment panel III guidelines,[54] our patient does not currently have major risk factors for coronary artery disease and would need an LDL concentration higher than 4·9 mmol/L to qualify for lipid-lowering therapy in the USA. However, he is borderline for three major risk factors (one of which is age) and any two of these would lower the threshold for treatment to 4·1 mmol/L (his measured LDL concentration was 4·0 mmol/L). Although no standards yet exist for the incorporation of global genetic risk in cardiovascular risk assessment, physicians are accustomed to incorporating many sources of information in clinical decision making. In this case, the patient's physician considered lifetime genetic risk and likely response to therapy when making the clinical decision to recommend a lipid-lowering drug. The patient's genome includes variants (table 3, table 4) that predict increased likelihood of beneficial effect for statins and reduced risk of the adverse effect of skeletal myopathy. Additionally, attributable risk was substantially reduced in carriers of the *LPA* risk allele who took aspirin,[15] leading to a discussion between the physician

and his patient about the threshold for primary prevention with aspirin therapy.

In view of a predisposition to coronary artery disease and other diseases on which risk is conditionally dependent (figure 5), understanding of the patient's potential response to clopidogrel and warfarin might be important for individualisation of future medical therapy. The patient is at risk of clopidogrel resistance as a result of his *CYP2C19* loss-of-function mutation, and his physician might recommend either an increased dose of clopidogrel in the event of future use, or consideration of new agents with alternative metabolism. By contrast, should the patient develop an indication for warfarin, his genotype at the *VKORC1* and *CYP4F2* loci suggests that he should take reduced initial doses of warfarin.

By contrast, our patient did not report a family history of haemochromatosis or parathyroid tumours, yet has some genetic risk for these disorders. In consideration of future screening studies, integrated clinical and genetic risks were assessed.

Important limitations remain in our ability to comprehensively integrate genetic information into clinical care. For example, a comprehensive database of rare mutations is needed. Since risk estimates change as studies are completed, a continually updated pipeline is necessary. There are imperfections in all human genomes published to date—false positive and false negative SNP calls, incomplete measurement of structural variation, and little direct haplotype data. Finally, gene-environment interactions are challenging to quantify and have been little studied.

As whole-genome sequencing becomes increasingly widespread, availability of genomic information will no longer be the limiting factor in application of genetics to clinical medicine. Development of methods integrating genetic and clinical data will assist clinical decision making and represent a large step towards individualised medicine. The transition to a new era of genome-informed medical care will need a team approach incorporating medical and genetics professionals, ethicists, and health-care delivery organisations.

#### Contributors

EAA and HTG conceived of the study. All authors contributed to data collection. EAA, AJB, MTW, RC, TEK, FED, JTD, KEO, LH, AAM, LG, LMH, DSB, KS, CFT, JMH, HS, JWK, MC, JT, AR, AWZ, GC, HTG, SRQ, and RBA participated in data interpretation. EAA, AJB, MTW, RC, TEK, JTD, KEO, LH, JWK, HTG, SRQ, and RBA prepared the report. All authors provided critical review of the draft and approved the final version.

#### Conflicts of interest

RBA is consultant to a direct-to-consumer genetic testing company, 23andme, and has received consultancy fees from Novartis. GMC is an adviser to several sequencing and direct-to-consumer companies (23andme, Knome, Helicos; full list at the time of publication is available in the webappendix, p 17). KEO was a paid consultant as a member of the Genetic Counseling Task Force for Navigenics from June, 2007, to August, 2009. SRQ is a founder, consultant, and equity holder in Helicos BioSciences. DP is an equity holder in Helicos BioSciences. AWZ is a founder, consultant, and equity holder in Scalable Computing Experts Inc. AJB is a scientific advisory board member and founder for NuMedii and Genstruct, is a scientific advisory board member for Johnson and Johnson, has received consultancy fees from Lilly, NuMedii, Johnson and Johnson, Genstruct, Tercica, and Prevendia and honoraria from Lilly and Siemens, and holds stock in NuMedii and Genstruct. EAA, DSB, RC, MFC, FED, JTD, LG, HTG, JMH, LMH, LH, TEK, JWK, AAM, NFN, AP, AMR, HS, KS, JVT, CFT, RW, MTW, and MW declare that they have no conflicts of interest.

#### Acknowledgments

We thank Josephine Puryear, Joshua Spin, Emidio Capriotti, and Connie Oshiro, and Priyanka Korde and Prajkta Bhide from Optra Systems for the curation of disease-associated single nucleotide polymorphisms from literature. This work was supported by grants from the National Institute of General Medical Sciences (GM61374 and associated ARRA supplement, GM079719), National Heart, Lung And Blood Institute (F32HL097462, K08 HL083914), National Human Genome Research Institute (HG003389), Howard Hughes Medical Institute, National Library of Medicine (LM007033, LM009719), Lucile Packard Foundation for Children's Health, Hewlett Packard Foundation, and Breetwor Family Foundation.

#### References

1 Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 2009; **106:** 19096–101.

2 Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009; **461:** 272–76.

3 Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008; **452:** 872–76.

4 Kim JI, Ju YS, Park H, et al. A highly annotated whole-genome sequence of a Korean individual. *Nature* 2009; **460:** 1011–15.

5 Levy S, Sutton G, Ng PC, et al. The diploid genome sequence of an individual human. *PLoS Biol* 2007; **5:** e254.

6 Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet* 2009; **85:** 142–54.

7 Pushkarev D, Neff NF, Quake SR. Single-molecule sequencing of an individual human genome. *Nat Biotechnol* 2009; **27:** 847–52.

8 Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31:** 3812–14.

9 Stenson PD, Mort M, Ball EV, et al. The human gene mutation database: 2008 update. *Genome Med* 2009; **1:** 13.

10 Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002; **30:** 3894–900.

11 Tanaka M, Takeyasu T, Fuku N, et al. Mitochondrial genome single nucleotide polymorphisms and their phenotypes in the Japanese. *Ann N Y Acad Sci* 2004; **1011:** 7–20.

12 The Uniprot Consortium. The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res* 2009; **37:** D169–74.

13 Jegga AG, Gowrisankar S, Chen J, Aronow BJ. PolyDoms: a whole genome database for the identification of non-synonymous coding SNPs with the potential to impact disease. *Nucleic Acids Res* 2007; **35:** D700–06.

14 Klein TE, Chang JT, Cho MK, et al. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. *Pharmacogenomics J* 2001; **1:** 167–70.

15 Chasman DI, Shiffman D, Zee RY, et al. Polymorphism in the apolipoprotein(a) gene, plasma lipoprotein(a), cardiovascular disease, and low-dose aspirin therapy. *Atherosclerosis* 2009; **203:** 371–76.

16 Luke MM, Kane JP, Liu DM, et al. A polymorphism in the protease-like domain of apolipoprotein(a) is associated with severe coronary artery disease. *Arterioscler Thromb Vasc Biol* 2007; **27:** 2030–36.

17 Evangelou E, Chapman K, Meulenbelt I, et al. Large-scale analysis of association between GDF5 and FRZB variants and osteoarthritis of the hip, knee, and hand. *Arthritis Rheum* 2009; **60:** 1710–21.

18 Allen KJ, Gurrin LC, Constantine CC, et al. Iron-overload-related disease in HFE hereditary hemochromatosis. *N Engl J Med* 2008; **358:** 221–30.

19 Tomatsu S, Orii KO, Fleming RE, et al. Contribution of the H63D mutation in HFE to murine hereditary hemochromatosis. *Proc Natl Acad Sci USA* 2003; **100:** 15788–93.

20 Hymes J, Stanley CM, Wolf B. Mutations in BTD causing biotinidase deficiency. *Hum Mutat* 2001; **18:** 375–81.

21 Rossi A, Superti-Furga A. Mutations in the diastrophic dysplasia sulfate transporter (DTDST) gene (SLC26A2): 22 novel mutations, mutation review, associated skeletal phenotypes, and diagnostic relevance. *Hum Mutat* 2001; **17:** 159–71.

22 Pulkkinen L, Christiano AM, Gerecke D, et al. A homozygous nonsense mutation in the beta 3 chain gene of laminin 5 (LAMB3) in Herlitz junctional epidermolysis bullosa. *Genomics* 1994; **24:** 357–60.

23 Calonge MJ, Gasparini P, Chillaron J, et al. Cystinuria caused by mutations in rBAT, a gene involved in the transport of cystine. *Nat Genet* 1994; **6:** 420–25.

24 van der Zwaag PA, Jongbloed JD, van den Berg MP, et al. A genetic variants database for arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Hum Mutat* 2009; **30:** 1278–83.

25 Maron BJ, Niimura H, Casey SA, et al. Development of left ventricular hypertrophy in adults in hypertrophic cardiomyopathy caused by cardiac myosin-binding protein C gene mutations. *J Am Coll Cardiol* 2001; **38:** 315–21.

26 Shattuck TM, Valimaki S, Obara T, et al. Somatic and germ-line mutations of the HRPT2 gene in sporadic parathyroid carcinoma. *N Engl J Med* 2003; **349:** 1722–29.

27 Cuppens H, Legius E, Cabello P, et al. Association between XV2c/CS7/KM19/D9 haplotypes and the delta F508 mutation. A study of 57 Belgian families. *Hum Genet* 1990; **85:** 402–03.

28 Morner S, Richard P, Kazzam E, et al. Identification of the genotypes causing hypertrophic cardiomyopathy in northern Sweden. *J Mol Cell Cardiol* 2003; **35:** 841–49.

29 Van Driest SL, Vasile VC, Ommen SR, et al. Myosin binding protein C mutations and compound heterozygosity in hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2004; **44:** 1903–10.

30 Merner ND, Hodgkinson KA, Haywood AF, et al. Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. *Am J Hum Genet* 2008; **82:** 809–21.

31 Yang Z, Bowles NE, Scherer SE, et al. Desmosomal dysfunction due to mutations in desmoplakin causes arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circ Res* 2006; **99:** 646–55.

32 Nishizato Y, Ieiri I, Suzuki H, et al. Polymorphisms of OATP-C (SLC21A6) and OAT3 (SLC22A8) genes: consequences for pravastatin pharmacokinetics. *Clin Pharmacol Ther* 2003; **73:** 554–65.

33 Kivistö KT, Niemi M. Influence of drug transporter polymorphisms on pravastatin pharmacokinetics in humans. *Pharm Res* 2007; **24:** 239–47.

34 SEARCH Collaborative Group, Link E, Parish S, Armitage J, et al. SLCO1B1 variants and statin-induced myopathy–a genomewide study. *N Engl J Med* 2008; **359:** 789–99.

35 Mega JL, Close SL, Wiviott SD, et al. Cytochrome p-450 polymorphisms and response to clopidogrel. *N Engl J Med* 2009; **360:** 354–62.

36 Yuan HY, Chen JJ, Lee MT, et al. A novel functional VKORC1 promoter polymorphism is associated with inter-individual and inter-ethnic differences in warfarin sensitivity. *Hum Mol Genet* 2005; **14:** 1745–51.

37 Pérez-Andreu V, Roldán V, Antón AI, et al. Pharmacogenetic relevance of CYP4F2 V433M polymorphism on acenocoumarol therapy. *Blood* 2009; **113:** 4977–79.

38 Pacanowski MA, Gong Y, Cooper-Dehoff RM, et al, INVEST Investigators. Beta-adrenergic receptor gene polymorphisms and beta-blocker treatment outcomes in hypertension. *Clin Pharmacol Ther* 2008; **84:** 715–21.

39 Johnson JA, Zineh I, Puckett BJ, McGorray SP, Yarandi HN, Pauly DF. Beta 1-adrenergic receptor polymorphisms and antihypertensive response to metoprolol. *Clin Pharmacol Ther* 2003; **74:** 44–52.

40 Couvert P, Giral P, Dejager S, et al. Association between a frequent allele of the gene encoding OATP1B1 and enhanced LDL-lowering response to fluvastatin therapy. *Pharmacogenomics* 2008; **9:** 1217–27.

41 Lanfear DE, Jones PG, Marsh S, Cresci S, McLeod HL, Spertus JA. Beta2-adrenergic receptor genotype and survival among patients receiving beta-blocker therapy after an acute coronary syndrome. *JAMA* 2005; **294:** 1526–33.

42 Kaye DM, Smirk B, Williams C, Jennings G, Esler M, Holst D. Beta-adrenoceptor genotype influences the response to carvedilol in patients with congestive heart failure. *Pharmacogenetics* 2003; **13:** 379–82.

43 Rau T, Wuttke H, Michels LM, et al. Impact of the CYP2D6 genotype on the clinical effects of metoprolol: a prospective longitudinal study. *Clin Pharmacol Ther* 2009; **85:** 269–72.

44 Moore AF, Jablonski KA, McAteer JB, et al, Diabetes Prevention Program Research Group.Extension of type 2 diabetes genome-wide association scan results in the diabetes prevention program. *Diabetes* 2008; **57:** 2503–10.

45 Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* 2006; **22:** 2729–34.

46 Shuldiner AR, O'Connell JR, Bliden KP, et al. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA* 2009; **302:** 849–57.

47 Klein TE, Altman RB, Eriksson N, et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 2009; **360:** 753–64.

48 Caldwell MD, Awad T, Johnson JA, et al. CYP4F2 genetic variant alters required warfarin dose. *Blood* 2008; **111:** 4106–12.

49 Jakobsdottir J, Gorin MB, Conley YP, Ferrell RE, Weeks DE. Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. *PLoS Genet* 2009; **5:** e1000337.

50 Kathiresan S, Melander O, Anevski D, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 2008; **358:** 1240–49.

51 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447:** 661–78.

52 Kamstrup PR, Tybjaerg-Hansen A, Steffensen R, Nordestgaard BG. Pentanucleotide repeat polymorphism, lipoprotein(a) levels, and risk of ischemic heart disease. *J Clin Endocrinol Metab* 2008; **93:** 3769–76.

53 Clarke R, Peden JF, Hopewell JC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med* 2009; **361:** 2518–28.

54 Grundy SM, Cleeman JI, Merz CN, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. *Circulation* 2004; **110:** 227–39.