

Frontiers in cardiovascular medicine

Personalized medicine: hope or hype?

Keyan Salari¹, Hugh Watkins², and Euan A. Ashley^{3*}

¹Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA; ²Department of Cardiovascular Medicine, University of Oxford, Oxford, UK; and ³Center for Inherited Cardiovascular Disease, Division of Cardiovascular Medicine, Stanford University School of Medicine, Falk Cardiovascular Research Building, 300 Pasteur Drive, Stanford, CA 94305, USA

Received 21 April 2011; revised 21 March 2012; accepted 3 April 2012; online publish-ahead-of-print 1 June 2012

See page 1553 for the editorial comment on this article (doi:10.1093/eurheartj/ehs089)

Medicine has always been personalized. For years, physicians have incorporated environmental, behavioural, and genetic factors that affect disease and drug response into patient management decisions. However, until recently, the 'genetic' data took the form of family history and self-reported race/ethnicity. As genome sequencing declines in cost, the availability of specific genomic information will no longer be limiting. Rather, our ability to parse these data and our decision whether to use it will become primary. As our understanding of genetic association with drug responses and diseases continues to improve, clinically useful genetic tests may emerge to improve upon our previous methods of assessing genetic risks. Indeed, genetic tests for monogenic disorders have already proven useful. Such changes may usher in a new era of personalized medicine. In this review, we will discuss the utility and limitations of personal genomic data in three domains: pharmacogenomics, assessment of genetic predispositions for common diseases, and identification of rare disease-causing genetic variants.

Keywords

Pharmacogenomics • Common disease risk assessment • Rare genetic variant discovery • Personalized medicine
• Genomic medicine

Introduction

Medicine has always been personalized. For years, physicians have incorporated environmental, behavioural, and genetic factors that affect disease and drug response into patient management decisions. However, until recently, the 'genetic' data took the form of family history and self-reported race/ethnicity. As genome sequencing declines in cost, the availability of specific genomic information will no longer be limiting. Rather, our ability to parse these data and our decision whether to use it will become primary. As our understanding of genetic association with drug responses and diseases continues to improve, clinically useful genetic tests may emerge to improve upon our previous methods of assessing genetic risks. Indeed, genetic tests for monogenic disorders have already proven useful. Such changes may usher in a new era of personalized medicine.

In considering where personal genomic information has begun to, or has the potential to, impact clinical medicine, three domains emerge (Figure 1): pharmacogenomics, assessment of genetic predispositions for common diseases, and identification of rare disease-causing genetic variants. In this review, we will discuss the utility and limitations of personal genomic data in each of these domains.¹

Pharmacogenomics

Pharmacogenomics, the study of how genes modulate drug responses among individuals, is likely to be one of the first direct applications of personal genomics to clinical medicine, with several notable examples already emerging. Genes that underlie differences in drug response can harbour genetic variants involved in the pharmacokinetics of a drug (i.e. how the drug is the absorbed, distributed, metabolized, and excreted) or the pharmacodynamics of a drug (i.e. how the drug interacts with its target and its mechanism of action). Such variation can influence, and hence potentially predict, both efficacy and toxicity. The number of pharmacogenetic associations has steadily increased over the years. Today, evidence for over 2000 genes involved in drug response have been annotated by curators at the Pharmacogenomics Knowledge Base (PharmGKB; <http://www.pharmgkb.org>).

Early successes in pharmacogenetics have arisen in oncology, where somatic genetic changes in a patient's tumour (which often have more substantial effects than variation in an individual's germline DNA) can markedly elevate or repress gene expression (compared with normal tissues) and provide a wide therapeutic window (Figure 2A). In this context, the genetic analysis of a patient's tumour can help predict therapeutic benefit—or lack

* Corresponding author. Tel: +1 650 498 4900, Fax: +1 650 725 1599, Email: euana@stanford.edu

Published on behalf of the European Society of Cardiology. All rights reserved. © The Author 2012. For permissions please email: journals.permissions@oup.com

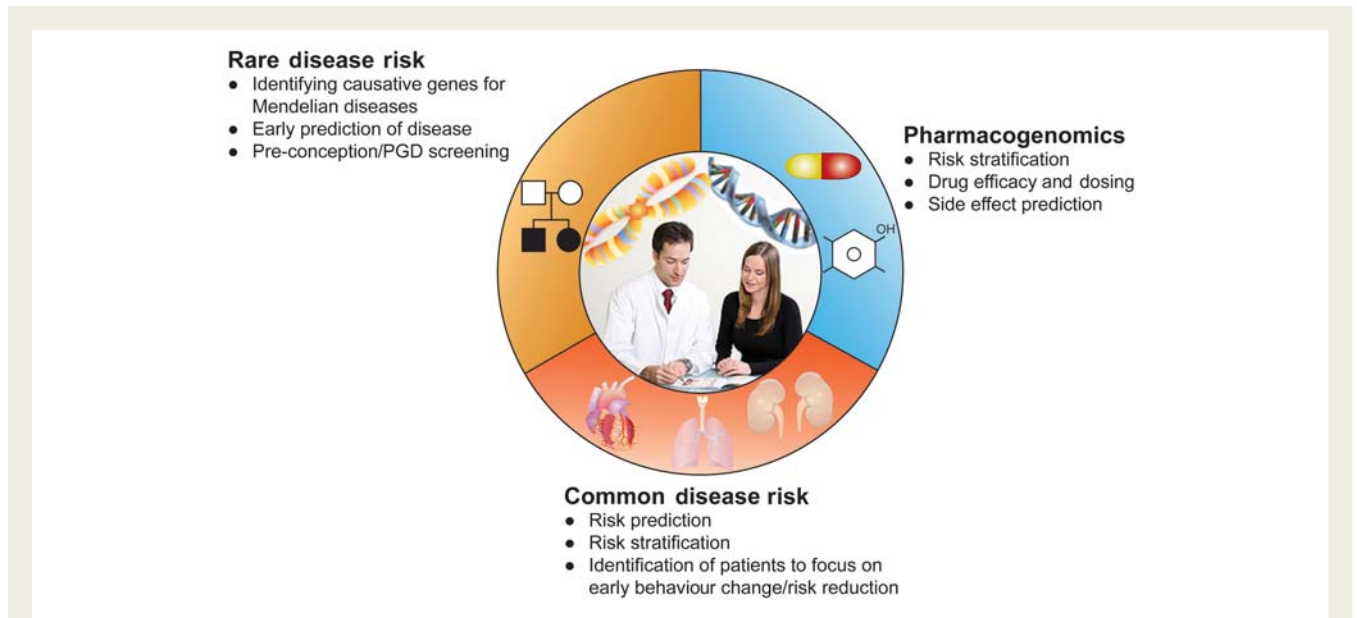


Figure 1 Domains of personalized medicine.

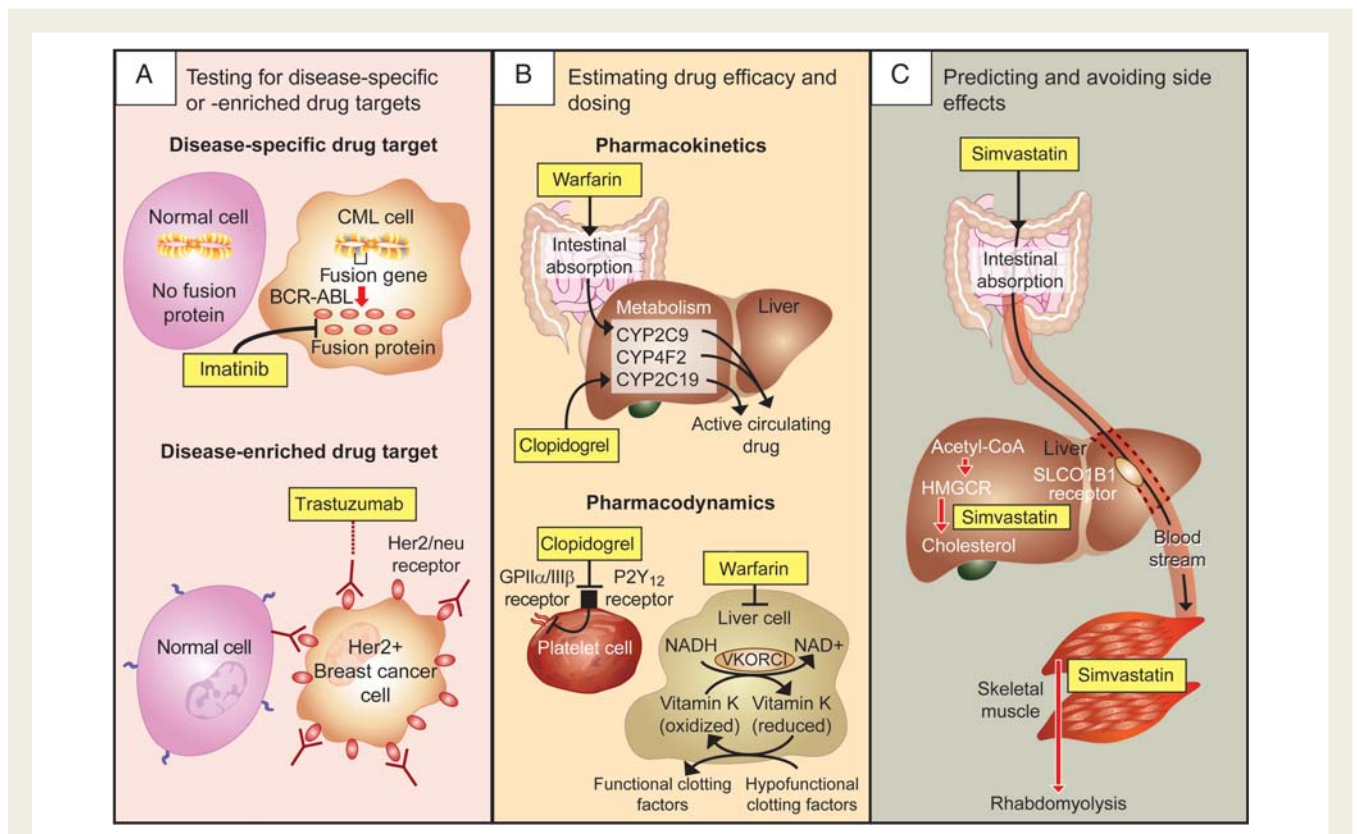


Figure 2 Clinical use of pharmacogenomic information. (A) Patients who test positive for genetic alterations that are disease-specific (e.g. BCR-ABL gene fusion in chronic myelogenous leukemia) or disease-enriched (e.g. Her2/neu receptor amplification in breast cancer) can select biologic drugs targeted specifically to their disease. (B) Patients with genetic variants affecting the pharmacokinetics or pharmacodynamics of a drug can benefit from improved prediction of drug efficacy and required dosing. (C) Patients with genetic variants associated with drug side effects can benefit by predicting and avoiding such side effects.

thereof—of targeted biologics, such as trastuzumab for *ERBB2*(HER2)-amplified breast cancers, erlotinib for epidermal growth factor receptor (EGFR) overexpressing lung cancers, or imatinib for Philadelphia chromosome-positive chronic myelogenous leukaemias. Somatic mutations in tumours can also help predict resistance, as in the case of colorectal cancers where activating mutations in *KRAS* have been established as a predictive marker for resistance to EGFR-specific antibodies cetuximab and panitumumab.^{2–4} Finally, somatic tumour mutations can give rise to ‘synthetic lethal’ interactions with drugs, best exemplified by the profound sensitivity of tumours with *BRCA1* or *BRCA2* dysfunction to the inhibition of poly(ADP-ribose) polymerase (PARP).^{5,6} PARP functions in the repair of DNA single-strand breaks, and in tumours with defective *BRCA1/BRCA2*-mediated repair, PARP inhibition seemingly leads to the persistence of DNA lesions, resulting in chromosomal instability, cell cycle arrest, and subsequent apoptosis.⁶ A recent clinical trial supports the efficacy of PARP inhibition among patients with *BRCA*-mutated tumours,⁷ and trials evaluating PARP inhibitors in the context of sporadic basal-like tumours (which often resemble *BRCA*-mutant tumours) are ongoing.

In cardiovascular medicine, the anti-coagulant drug warfarin represents an informative case study of how germline genetic information might help personalize a patient’s treatment regimen. The appropriate dose of warfarin varies by over 10-fold among patients, and establishing the correct dose for a given patient is critically important because the therapeutic window is narrow. In addition to diet, clinical factors, and demographic variables, genetic variants in three genes—cytochrome P450, family 2, subfamily C, polypeptide 9 (*CYP2C9*), vitamin K epoxide reductase complex, subunit 1 (*VKORC1*), and cytochrome P450, family 4, subfamily F, polypeptide 2 (*CYP4F2*)—contribute significantly to patient variability in response to warfarin^{8–11} (Figure 2B). Recently, the International Warfarin Pharmacogenetics Consortium (IWPC) developed a pharmacogenetic algorithm to help estimate warfarin dosing and showed that it produced recommendations significantly closer to the required stable therapeutic dose than those derived from an algorithm based on only clinical variables or a fixed-dose strategy.¹² Specifically, the pharmacogenetic algorithm correctly predicted low doses for 54% of all patients (when compared with the clinical algorithm which predicted low doses for 33%), while the pharmacogenetic algorithm accurately predicted high doses for 26% of patients who required high doses (vs. 9% for the clinical algorithm). Thus, the pharmacogenetic algorithm significantly improved the dose prediction for those at the tails of the dosage distribution, a group accounting for 46% of the entire cohort. A recent study reported that using patient genotype data for warfarin dosing reduced the risk of hospitalization in outpatients initiating warfarin by nearly one-third.¹³ While further outcome studies of pharmacogenetic-based dosing of warfarin are warranted, and while newer agents metabolized by alternative pathways may supersede these strategies, the IWPC study exemplifies an approach that can likely be generalized to other commonly prescribed drugs where individual variability and/or a narrow therapeutic index are factors.

In addition to estimating efficacy and appropriate drug doses, pharmacogenetic information has the potential to be of a clinical

value when deciding between multiple treatment options to maximize treatment benefit and limit the risk of side effects (Figure 2C). For example, the cholesterol-lowering drug simvastatin can, in rare cases, cause a myopathy when administered at higher doses or in combination with certain other medications. A recent genome-wide association study (GWAS) identified a greater than 16-fold increased risk of statin-induced myopathy in homozygotes of a common variant within *SLCO1B1*, a gene known to regulate the hepatic uptake of statins.¹⁴ Another recent study identified that in statin-treated men with coronary artery disease (CAD), those with intrinsically low levels of cholesteryl ester transfer protein (carriers of the TaqIB-B2 allele) have increased 10-year mortality, suggesting an adverse pharmacogenetic interaction.¹⁵ Finally, several pharmacogenetic variants have been associated with treatment benefit of the angiotensin-converting enzyme-inhibitor perindopril in patients with stable CAD.¹⁶ Not all pharmacogenetic associations, however, are appropriate for population-based screening. In the case of severe statin-induced myopathy, the rarity of the reaction lowers the positive predictive value of this particular variant; nonetheless, the large effect size suggests that genetic testing to avoid serious drug toxicity may have clinical utility for some drugs, thereby helping achieve their benefits in patients more safely and effectively.

Clopidogrel, a thienopyridine that inhibits platelet aggregation, represents another illustrative example for cardiovascular pharmacogenomics. Delivered as an inactive prodrug, it is activated *in vivo* by several cytochrome P450 enzymes in the liver including *CYP2C19* (Figure 2B). Variants in *CYP2C19* can cause the loss of enzymatic function and thus lower the conversion rate to the active drug, thereby reducing efficacy.^{17,18} The clinical relevance of this was initially described in two contemporaneous studies: the Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition with Prasugrel-Thrombolysis in Myocardial Infarction (TRITON-TIMI) 38 and a French registry study of acute MI patients.^{19,20} In the TRITON-TIMI 38 trial, among persons treated with clopidogrel, carriers of a reduced-function *CYP2C19* allele had significantly lower levels of the active metabolite of clopidogrel, diminished platelet inhibition, and a higher rate of major adverse cardiovascular events, including stent thrombosis, than did non-carriers. In the French study, patients undergoing percutaneous coronary intervention for acute myocardial infarction who had two loss of function alleles had >3 times higher risk of death, myocardial infarction or stroke. The genetics substudy of the Platelet Inhibition and Patient Outcomes (PLATO) trial subsequently confirmed a higher event rate in the patient group with any *CYP2C19* loss of function variant.²¹ In response to these findings, the U.S. Food and Drug Administration (FDA) added a black box warning to clopidogrel describing this patient group as one at higher risk. However, debate has continued. Effects were not seen in the analysis of patients from two more recent studies: the Clopidogrel in Unstable Angina to Prevent Recurrent Events (CURE) trial and the Atrial Fibrillation Clopidogrel Trial with Irbesartan for Prevention of Vascular Events (ACTIVE) trial.²² Both studies compared clopidogrel with placebo in combination with aspirin for reducing cardiovascular events and found no difference in the loss of function allele carriers. Interestingly, the gain of function carriers in the CURE group did appear to obtain more benefit.

Adding further to the debate, in the TRITON-TIMI 38 trial, in almost 3000 patients, *ABCB1* and *CYP2C19* variants were significantly independent predictors of CV death, MI and stroke.²³ In sum, most investigators conclude that there is an effect of *CYP2C19* genotype on the platelet response to clopidogrel in the setting of percutaneous coronary intervention but there has not emerged sufficient consensus to recommend routine genetic testing in clinical practice. Indeed, most have used the debate surrounding genotyping as a reason to endorse newer alternatives such as ticagrelor that are not metabolized via this route.

Recognizing the value of pharmacogenetic information, the FDA has already approved, or relabelled, several drugs reflecting the variation in response due to genetic factors. In 2005, the FDA approved the combination drug isosorbide dinitrate/hydralazine (BiDil) for congestive heart failure specifically in African-American patients, emphasizing the principle of targeted therapy, albeit ethnicity-based rather than genetically derived. More recently, the FDA has added pharmacogenetic information to the product labels of warfarin, clopidogrel, and carbamazepine, though they have not included specific guidance for incorporating this information into drug choice or dosing. The FDA has also approved the design of a clinical trial for bucindolol, a β -blocker and mild vasodilator, in a genotype-defined heart failure population.²⁴ This prospective trial highlights the value of pharmacogenetic information from the perspective of pharmaceutical companies, whose drugs may not achieve clinical endpoints needed for FDA approval in the general population, but may be successful in genotype-defined subpopulations.

Despite the large number of well-established and putative pharmacogenetic associations, significant challenges remain in incorporating this information into clinical practice. First, it will be important that only robustly replicated associations are exploited. For example, genotype at the *KIF6* locus has been proposed, and fully commercialized, as a test of responsiveness to statin therapy on the basis of a number of smaller studies,^{25–27} yet the underlying association did not replicate at all in a far larger analysis.²⁸ Even where the genetic association is robust, the clinical utility of each pharmacogenetic test needs to be evaluated, taking into consideration the sensitivity, specificity, and positive and negative predictive values. For example, if a specific side effect from a drug is exceedingly rare, even a genetic variant that predicts the side effect with high sensitivity and specificity may have a sufficiently low positive predictive value to make testing for the variant cost ineffective on a population level. After appropriate cost-effectiveness analysis, there may remain only few examples of agents where germline genetic effects are large enough, therapeutic index low enough, and costs high enough that genetic testing pre-prescription will be warranted. On the other hand, if, as may increasingly be the case in the future, a patient had their genotype or whole-genome data available, several notable pharmacogenetic findings may be of interest to their physicians. In a recent study evaluating one patient's whole-genome sequence, 63 clinically relevant pharmacogenomic variants were noted as well as six novel, non-conservative, amino acid-changing single-nucleotide polymorphisms (SNPs) in genes that are important for drug response.¹ As more pharmacogenetic associations are discovered,

validated, and evaluated for clinical utility, we expect pharmacogenomics to continue to be a fertile ground for clinical translation and the practice of personalized medicine (for a detailed review of pharmacogenetic associations that impact cardiovascular medicine, see Verschuren *et al.*²⁹).

Common disease risk assessment

Over the past two decades, candidate gene association studies and, more recently, GWASs have provided researchers with a large catalogue of genetic variant associations with hundreds of common human diseases and traits.³⁰ While these studies were carried out in large part to identify new biological genes, pathways, and potential drug targets relevant to the studied phenotypes, there has been some interest in extending these findings to the estimation of individual risk and, in some cases, such estimations have been offered as direct to the consumer/patient genetic tests. Here, we discuss four aspects of genetic variants for disease risk estimation: the risk metric, selection of risk variants, integration of multiple risk variants as well as other risk factors of disease, and the clinical utility of the risk predictors.

Risk metric

Most GWASs employ a case–control study design and relate the risk of developing a disease due to a specific genetic variant via an odds ratio (OR). However, the assumption that the OR approximates the true risk only holds for rare diseases (i.e. typically diseases with an incidence <10%). For GWASs evaluating common diseases like cardiovascular disease or type 2 diabetes, the OR will always overestimate the true risk, and the magnitude of the inflation increases as the disease prevalence increases.

A more recently proposed alternative to OR-based risk assessment is the use of likelihood ratios (LRs). In the context of genetic data, an LR is the ratio of the probability of observing a specific genotype in diseased individuals (cases) to that in healthy individuals (controls).³¹ Likelihood ratios from multiple unlinked genetic variants can be multiplied together and used to transform a pre-test to a post-test probability of disease, metrics that are familiar in the framework of evidence-based medicine.

Variant selection

Selecting the genetic variants that might be informative in estimating individual disease risk is a challenge but critically important. Attributes of GWASs that should be considered when selecting disease-associated variants are the population or ethnicity of the study subjects in which the genetic association was discovered, the sample size of the study, the set of variants scanned in the study, and whether the association has been replicated. Most GWASs have evaluated one form of genetic variation—SNPs—and have focused specifically on those SNPs with a minor allele frequency of at least 5% in populations of European descent. This focus is grounded in the ‘common disease–common variant hypothesis’, which posits that common genetic variants in the population underlie the risk for common genetic diseases. The hypothesis further suggests that such diseases are likely caused by a large number of common variants, each contributing only a small risk and

thereby evading negative evolutionary selection. The vast majority of SNP associations confidently identified to date do indeed carry modest effect sizes (i.e. ORs of 1.0–1.5), but collectively they only explain a small proportion of the heritability of most common diseases.³² Thus, even if all the genetic variants discovered to date are selected appropriately for their inclusion in an individual's estimate of disease risk, the predictive power may remain low since much of the genetic variation underlying risk for many diseases remains undiscovered.

Furthermore, most genetic associations have been identified in populations of European descent. Thus, due to differences in the pattern of linkage disequilibrium between different populations, genetic associations identified in these studies may not be applicable across populations. For example, a genetic variant identified to confer the risk of cardiovascular disease among Europeans may be inappropriate to include in the risk estimate for an individual of African descent. Additional studies in diverse populations are thus warranted to identify shared and novel genetic associations in non-European populations.

Caution should also be exercised in selecting genetic variants whose associations with disease have not been replicated. Because of the well-known publication bias of reporting positive results and larger effect sizes (also known as 'winner's curse'),³³ genetic variants are commonly unable to be reproducibly associated with a given disease, and initial reports typically overestimate the effect size of the risk variant.

To find additional determinants of heritability, efforts are being focused towards larger studies of common variation as well as studies of rare variants, other forms of genetic variation aside from SNPs, and interactions between genes or between genes and the environment.³⁴ For example, height is a trait where 80% of phenotypic variation is heritable or attributable to genetic factors; however, the SNPs identified by individual GWASs (each analysing typically 1000–5000 individuals) collectively have only accounted for less than 5% of the variation in height. The largest genetic study to date, a meta-analysis of 46 GWASs (>130 000 individuals), recently identified 180 loci significantly associated with height that extends the proportion of variance explained to 10%.³⁵ Similarly, a recent meta-analysis of 14 GWASs of coronary artery disease (CAD) comprising 22 233 individuals with CAD and 64 762 controls followed by genotyping of top association signals in 56 682 additional individuals identified 13 new susceptibility loci and confirmed 10 of 12 previously reported CAD loci, together explaining ~10% of the genetic variance of CAD.³⁶ As genotyping technologies continue to become cheaper, such larger studies and meta-analyses (i.e. >100 000 subjects) will surely help identify more common variants associated with common diseases.^{35,37}

However, these studies will still fall grossly short of accounting for the total heritability of height or common human diseases, because many of the common SNPs with modest effect sizes are difficult to identify under the stringent statistical thresholds for significance imposed by correction for multiple hypothesis testing. There are some indications that the accepted standards, historically adopted to minimize false positives, may be contributing to false-negative findings. For instance, when Yang et al.³⁸ used all common SNPs (not just those reaching the standard 5×10^{-8} level of

significance) simultaneously to explain variation in height, 45% of the phenotypic variation could be accounted for, indicating that much of the heritability of common traits is not 'missing' but rather has not previously been detected because the individual effects of most SNPs are too small to pass stringent significance tests. The model employed by the study authors estimated the total amount of phenotypic variance accounted for by all common SNPs, but the accuracy of prediction from this model is low because the effects of individual SNPs are estimated with much error. Nonetheless, this study demonstrates that with better estimates of individual SNP effects from ever larger studies, risk prediction using common SNPs could become more robust in the future, with potentially important implications for personalized medicine.

To identify rare variants associated with disease, one promising approach is to sequence across GWAS-associated loci. For example, Johansen et al.³⁹ recently sequenced previously reported GWAS loci in individuals with extreme blood lipid profiles and identified rare variants with large effect sizes. While this finding is so far limited to an intermediate trait (i.e. lipid levels), it may prove generalizable to disease phenotypes. In addition to sequencing, genotyping platforms continue to include more SNPs with lower minor allele frequencies, which should help identify additional loci as well as clarify whether some association signals in GWASs attributed to common SNPs are actually the result of synthetic association or incomplete linkage disequilibrium between common SNPs and rarer causal variants.^{38,40} Preliminary evidence, however, suggests that synthetic associations are unlikely to account for many GWA signals.^{41,42}

Also warranting further investigation are additional forms of genetic variation. Copy number variants (CNVs) and other forms of structural variation affect a substantial portion of the genome and likely contribute to disease phenotypes. Early studies have noted that common CNVs account for only a limited proportion of heritability,^{43,44} yet some notable examples exist.⁴⁵ Future efforts should examine a broader range of structural variation including those that are individually rare yet in aggregate implicate specific genes.⁴⁶

Risk integration

To compute an overall risk of disease conferred by multiple genetic variants, results from several GWASs are often integrated. A common standard is to multiply the ORs corresponding to each risk variant to arrive at a composite OR. This method of integration assumes statistical independence between each of the risk variants, a scenario unlikely to remain true as the number of variants being considered increases, potentially leading to further overestimation of disease risk. One accepted approach has been to simply sum up the number of well-replicated risk alleles for a specific disease that any given individual carries. With this 'allelic dose' scoring approach, several investigators have evaluated the clinical utility of multilocus genotype scores in predicting disease outcomes (discussed below).^{47–52} Summing the number of risk alleles weighted by their estimated effect sizes has also been proposed as a method for computing multilocus genotype risk scores.^{51,53}

Integrating risk conferred by multiple variants and non-genetic factors remains a challenge for personalized medicine. Due to the complex genetic architecture of many common diseases, it is difficult to identify the interactions and dependencies between distinct genetic and non-genetic risk factors. Similarly, given the large role of environmental factors in the development of these diseases, integrating the contribution of such factors to disease risk is important. Formulating the risk contributions of environmental factors as LRs might serve as one method to compatibly integrate genetic and non-genetic risk factors.

Clinical utility

Finally, the clinical utility of the genetic variants used to estimate disease risk should be evaluated. In order for individual disease risk estimates based on personal genetic data to be clinically useful, they must demonstrate robust performance similar to established biomarkers. Performance of prediction models is commonly evaluated using the sensitivity, specificity, positive predictive value, and negative predictive value, often summarized by a receiver operating characteristic (ROC) curve. This approach has been utilized in most of the initial studies examining the clinical utility of common genetic variation in predicting risk for common diseases. Generally, these studies have shown that the inclusion of common SNPs identified in GWASs only modestly improves disease prediction beyond the standard clinical factors.^{47–52} We believe that these results in part reflect the fact that most of the genetic variation underlying the diseases studied remains undiscovered; for diseases with a substantial component of heritability, the inclusion of a more complete set of predictive genetic variants can be expected to improve the overall risk prediction, albeit with the caveat that the definition of individual small effect variants will require progressively larger studies. However, these studies also draw attention to the need for alternative methods for assessing the contribution of new biomarkers

(SNPs or otherwise) to risk prediction. Disease risk prediction is not by nature a binary classification problem, and thus the ROC curve is not as appropriate for evaluating clinical utility of risk predictors as it is for diagnostic tests.^{54,55} For example, a risk factor with an OR of 3 might have limited impact on the ROC curve, but may result in a substantial shift in absolute risk. Indeed, most established clinical risk factors have ORs <3 and fall in this category. The large change in absolute risk may lead to different patient management decisions. This is exemplified by a recent study of the performance of breast cancer risk predictors using clinical and genetic data, where clinical parameters generated a predictor with an area under the ROC curve (AUC) of only 0.58 (only modestly better than an AUC of 0.5 expected by chance).⁵² The addition of 10 common genetic variants increased the AUC to 0.618, but more importantly shifted over 50% of patients to a different quintile of disease risk. Therefore, additional studies of clinical utility of SNPs identified in GWA studies are warranted using more relevant statistical metrics for model performance, such as calibration and clinical risk reclassification.⁵⁶ Along with the continuing development of a more complete catalogue of genetic associations, we anticipate that these efforts are likely to show the clinical utility of genetic variants can be at least comparable with established risk factors, as was shown for breast cancer risk models, and may substantially improve some disease risk predictors.

Rare genetic variant discovery

The identification and characterization of rare genetic variants is a promising approach to elucidating the molecular underpinnings of rare and common genetic diseases. The highest level of 'causality' has been ascribed to single variants in single genes detected by family based linkage studies. This reflects the substantial portion of disease risk explained by those variants, albeit with expressivity

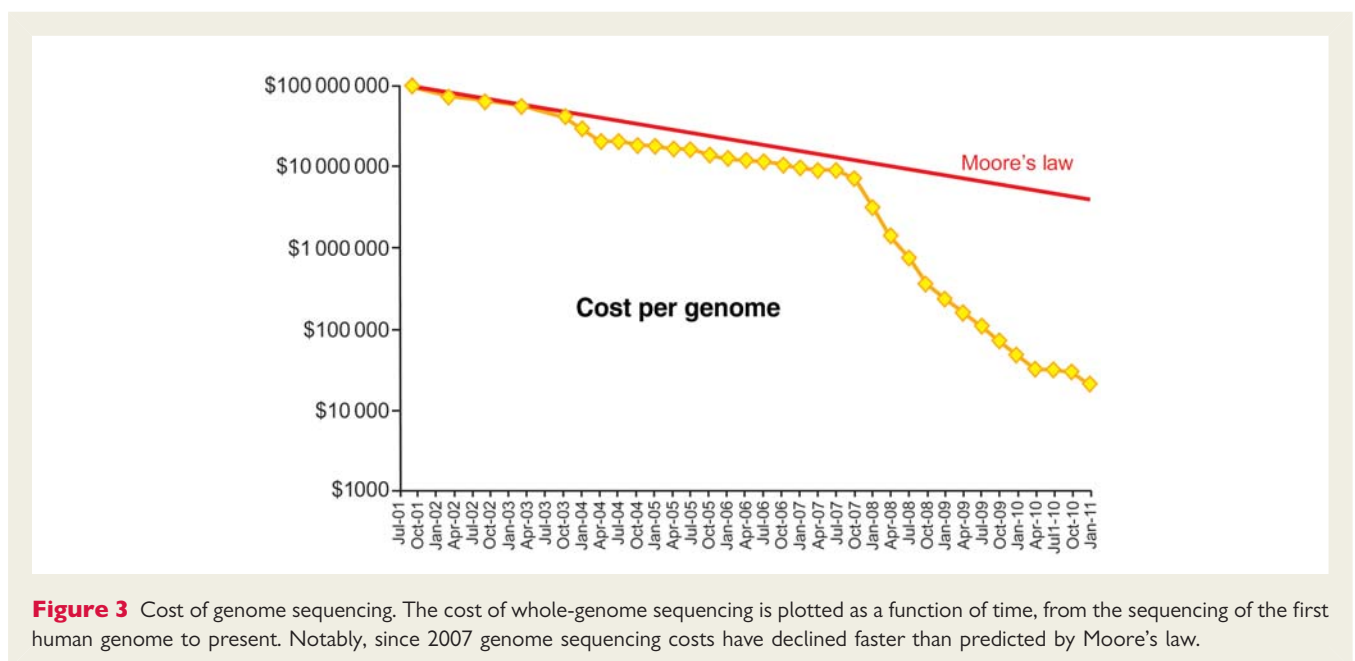


Figure 3 Cost of genome sequencing. The cost of whole-genome sequencing is plotted as a function of time, from the sequencing of the first human genome to present. Notably, since 2007 genome sequencing costs have declined faster than predicted by Moore's law.

modulated by modifying genomes and environment. Indeed, as many publications have already demonstrated,^{57–62} the unparalleled power of sequencing to identify segregating variants even in very small families is likely to bring genetic solutions to many families' rare syndromes. As the cost to perform full exome (the collection of all exons or protein-coding sequences in the genome) or whole-genome sequencing precipitously falls⁶³ (Figure 3), this may be the most immediate impact of the wide availability of high-throughput sequencing. While individually rare by definition, rare genetic diseases are collectively quite common, affecting more than 25 million people worldwide.⁶⁴ Further, if individuals who are carriers for such diseases could be identified early in life, the risk of passing on variants to future generations could be mitigated by pre-conception planning or pre-implantation genetic diagnosis. Indeed, direct-to-consumer genetic testing companies have already begun developing and marketing personal genetic tests for couples planning to conceive.

As discussed above, efforts to identify rare variants are increasingly applied to help uncover determinants of heritability of common diseases. In this approach, the focus is on identifying many rare variants that each contribute to risk of the same disease (i.e. genetic heterogeneity). A primary advantage of using the sequencing-based, rather than chip-based, technology is that the former has the capability to discover novel genetic variants that have never been observed before or have been observed so infrequently as to not warrant inclusion on a genotyping chip. Once thought to be relatively unimportant, these rare or novel genetic variants are now increasingly recognized as a source of genetic variation that underlies several common diseases. Most recently, the first confident estimates of germ line variation have emerged from the 1000 Genomes Project and an earlier study focusing on a family quartet.⁶⁵ At an estimated 10^{-8} new mutations naturally arising per base pair per generation, this leads to ~70–100 novel variants per new human genome. In addition to these newly arising variants, each individual genome sequenced has been found to carry a far larger number (up to 10 000) of very rare, often previously unrecorded, DNA variants, perhaps as many as 100 of which are predicted to result in loss of function of a gene.^{66–68} However, estimating pathogenicity from these data presents many challenges. Specifically, distinguishing causative variants from the large number of non-causative variants discovered by exome or whole-genome sequencing requires prioritizing candidate variants by evidence of their functional impact and validating candidates to establish causality.

There are several approaches investigators have used to prioritize candidate genetic variants. One common and intuitive approach is to search for variants in coding sequences that are predicted to be deleterious. This would include variants that introduce new start or stop codons, frameshift variants, or variants disrupting splicing. However, for many genes, one functional copy is sufficient such that loss of the other copy can be silent, as witnessed by the surprisingly large number of 'loss of function' variants found in each individual genome.^{66–68} To understand more subtle changes in protein structure and function (which may in some cases be more deleterious), biophysical factors may be relevant. A variety of computational tools have been developed to help investigators prioritize candidate genetic variants by predicting their functional impact, such as the Sorting Intolerant from Tolerant (SIFT) algorithm,⁶⁹ Polymorphism

Phenotyping (PolyPhen),⁷⁰ the Universal Protein Resource (UniProt) database,⁷¹ and PolyDoms.⁷² Manual curation of SNPs associated with a known or suspected disease gene can also be done using annotations in the Online Mendelian Inheritance in Man (OMIM) database and Human Gene Mutation Database.⁷³ Despite the availability of tools and databases for the analysis of rare or novel genetic variants, characterizing these variants remains quite challenging due to the fragmented nature of the curated databases, thus making it difficult to annotate novel/rare variants from whole-genome sequence data.

The reference human genome creates another challenge in the confident interpretation of high-throughput sequencing data. As a composite of a small number of individuals' DNA (all likely of European descent), the reference genome sequence by definition contains risk alleles.⁷⁴ In time, with greater computational power or longer reads, *de novo* assembly of genomes will be more easily achievable. This will facilitate the return of a sequence rather than the more computationally convenient method of mapping to a known reference. Further, a less biased view can be obtained by comparing risk alleles in an individual genome to a composite major allele reference genome sequence.⁷⁵

Conclusions

Over the next decade, we will gain an unprecedented appreciation of the genetic variation present within human genomes. This collective undertaking will yield tremendous benefits of new target and pathway discovery related to a multitude of human diseases. Directed to the individual patient, these efforts will provide a basis for refining personal predictors of disease and drug response. Aside from the technological and informatics challenges, the prospect that physicians will, in the next few years, be expected to include individual exome, or whole-genome, data in patient management raises obvious educational,⁷⁶ ethical, and social implications.⁷⁷ In addition, much genetic testing will be marketed direct to consumer. At present, this extends largely to common variants assessed by genotyping SNP arrays; variants identified this way will typically have a modest impact on a common disease and in turn may not be too problematic. However, it will soon become more cost effective to obtain a patient's exome or whole-genome sequence; the expectation that uncommon, highly predictive variants will also be uncovered in a significant minority illustrates the need for a well thought-through and regulated framework.

How the systematic exploration of human genetic variation will transform diagnostic and clinical practices remains to be seen, but tempered by the challenges discussed here, the analyses of human diseases powered by the genomic revolution bolsters hope for sustained and meaningful improvements in the diagnosis, prognosis, and treatment of individual patients.

Funding

K.S. is supported by the Medical Scientist Training Program at Stanford University School of Medicine. H.W. is supported by the NIHR Comprehensive Biomedical Research Centre and the BHF Centre of Excellence, at Oxford. E.A.A. is supported by the NIH NNN (grant no R01: NIH 5R01HL105993 and DP2: NIH DP2 OD004613).

Conflict of interest: none declared.

References

- Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, Altman RB. Clinical assessment incorporating a personal genome. *Lancet* 2010;**375**:1525–1535.
- Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R, Patterson SD, Chang DD. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol* 2008;**26**:1626–1634.
- Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, Simes RJ, Chalchal H, Shapiro JD, Robitaille S, Price TJ, Shepherd L, Au HJ, Langer C, Moore MJ, Zalberg JR. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* 2008;**359**:1757–1765.
- Van Cutsem E, Kohne CH, Hitre E, Zaluski J, Chang Chien CR, Makhson A, D'Haens G, Pinter T, Lim R, Bodoky G, Roh JK, Folprecht G, Ruff P, Stroh C, Tejpar S, Schlichting M, Nippgen J, Rougier P. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med* 2009;**360**:1408–1417.
- Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, Kyle S, Meuth M, Curtin NJ, Helleday T. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* 2005;**434**:913–917.
- Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NM, Jackson SP, Smith GC, Ashworth A. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 2005;**434**:917–921.
- Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, Mortimer P, Swaisland H, Lau A, O'Connor MJ, Ashworth A, Carmichael J, Kaye SB, Schellens JH, de Bono JS. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* 2009;**361**:123–134.
- Caldwell MD, Awad T, Johnson JA, Gage BF, Falkowski M, Gardina P, Hubbard J, Turpaz Y, Langae TY, Eby C, King CR, Brower A, Schmelzer JR, Glurich I, Vidaillet HJ, Yale SH, Qi Zhang K, Berg RL, Burmester JK. CYP4F2 genetic variant alters required warfarin dose. *Blood* 2008;**111**:4106–4112.
- Rieder MJ, Reiner AP, Gage BF, Nickerson DA, Eby CS, McLeod HL, Blough DK, Thummel KE, Veenstra DL, Rettie AE. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* 2005;**352**:2285–2293.
- Sconce EA, Khan TI, Wynne HA, Avery P, Monkhouse L, King BP, Wood P, Kesteven P, Daly AK, Kamali F. The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. *Blood* 2005;**106**:2329–2333.
- Takahashi H, Wilkinson GR, Nutescu EA, Morita T, Ritchie MD, Scordo MG, Pengo V, Barban M, Padriani R, Ieiri I, Otsubo K, Kashima T, Kimura S, Kijima S, Echizen H. Different contributions of polymorphisms in VKORC1 and CYP2C9 to intra- and inter-population differences in maintenance dose of warfarin in Japanese, Caucasians and African-Americans. *Pharmacogenet Genomics* 2006;**16**:101–110.
- Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, Lee MT, Limdi NA, Page D, Roden DM, Wagner MJ, Caldwell MD, Johnson JA. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 2009;**360**:753–764.
- Epstein RS, Moyer TP, Aubert RE, DJ OK, Xia F, Verbrugge RR, Gage BF, Teagarden JR. Warfarin genotyping reduces hospitalization rates results from the MM-WES (Medco-Mayo Warfarin Effectiveness study). *J Am Coll Cardiol* 2010;**55**:2804–2812.
- Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, Gut I, Lathrop M, Collins R. SLCO1B1 variants and statin-induced myopathy—a genome-wide study. *N Engl J Med* 2008;**359**:789–799.
- Regieli JJ, Jukema JW, Grobbee DE, Kastelein JJ, Kuivenhoven JA, Zwinderman AH, van der Graaf Y, Bots ML, Doevendans PA. CETP genotype predicts increased mortality in statin-treated men with proven cardiovascular disease: an adverse pharmacogenetic interaction. *Eur Heart J* 2008;**29**:2792–2799.
- Brugts JJ, Isaacs A, Boersma E, van Duijn CM, Uitterlinden AG, Remme W, Bertrand M, Ninomiya T, Ceconi C, Chalmers J, MacMahon S, Fox K, Ferrari R, Witteman JC, Danser AH, Simoons ML, de Maat MP. Genetic determinants of treatment benefit of the angiotensin-converting enzyme-inhibitor perindopril in patients with stable coronary artery disease. *Eur Heart J* 2010;**31**:1854–1864.
- Hulot JS, Bura A, Villard E, Azizi M, Remones V, Goyenville C, Aiach M, Lechat P, Gaussem P. Cytochrome P450 2C19 loss-of-function polymorphism is a major determinant of clopidogrel responsiveness in healthy subjects. *Blood* 2006;**108**:2244–2247.
- Matetzky S, Shenkman B, Guetta V, Shechter M, Bienart R, Goldenberg I, Novikov I, Pres H, Savion N, Varon D, Hod H. Clopidogrel resistance is associated with increased risk of recurrent atherothrombotic events in patients with acute myocardial infarction. *Circulation* 2004;**109**:3171–3175.
- Mega JL, Close SL, Wiviott SD, Shen L, Hockett RD, Brandt JT, Walker JR, Antman EM, Macias W, Braunwald E, Sabatine MS. Cytochrome p-450 polymorphisms and response to clopidogrel. *N Engl J Med* 2009;**360**:354–362.
- Simon T, Verstuyft C, Mary-Krause M, Quteineh L, Drouet E, Meneveau N, Steg PG, Ferrieres J, Danchin N, Becquemont L. Genetic determinants of response to clopidogrel and cardiovascular events. *N Engl J Med* 2009;**360**:363–375.
- Wallentin L, James S, Storey RF, Armstrong M, Barratt BJ, Horrow J, Husted S, Katus H, Steg PG, Shah SH, Becker RC. Effect of CYP2C19 and ABCB1 single nucleotide polymorphisms on outcomes of treatment with ticagrelor versus clopidogrel for acute coronary syndromes: a genetic substudy of the PLATO trial. *Lancet* 2010;**376**:1320–1328.
- Pare G, Mehta SR, Yusuf S, Anand SS, Connolly SJ, Hirsh J, Simonsen K, Bhatt DL, Fox KA, Eikelboom JW. Effects of CYP2C19 genotype on outcomes of clopidogrel treatment. *N Engl J Med* 2010;**363**:1704–1714.
- Mega JL, Close SL, Wiviott SD, Shen L, Walker JR, Simon T, Antman EM, Braunwald E, Sabatine MS. Genetic variants in ABCB1 and CYP2C19 and cardiovascular outcomes after treatment with clopidogrel and prasugrel in the TRITON-TIMI 38 trial: a pharmacogenetic analysis. *Lancet* 2010;**376**:1312–1319.
- Liggett SB, Miale-Perez J, Thaneemit-Chen S, Weber SA, Greene SM, Hodne D, Nelson B, Morrison J, Domanski MJ, Waggoner LE, Abraham WT, Anderson JL, Carlquist JF, Krause-Steinrauf HJ, Lazzeroni LC, Port JD, Lavrov PW, Bristow MR. A polymorphism within a conserved beta(1)-adrenergic receptor motif alters cardiac function and beta-blocker response in human heart failure. *Proc Natl Acad Sci USA* 2006;**103**:11288–11293.
- Iakoubova OA, Sabatine MS, Rowland CM, Catanese JJ, Blauw GJ, Jukema JW, Murphy MB, Devlin JJ, Ford I, Shepherd J. KIF6 Trp719Arg polymorphism and the effect of statin therapy in elderly patients: results from the PROSPER study. *Eur J Cardiovasc Prev Rehabil* 2010;**17**:455–461.
- Iakoubova OA, Sabatine MS, Rowland CM, Tong CH, Catanese JJ, Ranade K, Simonsen KL, Kirchgessner TG, Cannon CP, Devlin JJ, Braunwald E. Polymorphism in the PROVE IT-TIMI 22 study. *J Am Coll Cardiol* 2008;**51**:449–455.
- Iakoubova OA, Tong CH, Rowland CM, Kirchgessner TG, Young BA, Arellano AR, Shiffman D, Sabatine MS, Campos H, Packard CJ, Pfeffer MA, White TJ, Braunwald E, Shepherd J, Devlin JJ, Sacks FM. Association of the Trp719Arg polymorphism in kinesin-like protein 6 with myocardial infarction and coronary heart disease in 2 prospective trials: the CARE and WOSCOPS trials. *J Am Coll Cardiol* 2008;**51**:435–443.
- Hopewell JC, Parish S, Clarke R, Armitage J, Bowman L, Hager J, Lathrop M, Collins R. No impact of KIF6 genotype on vascular risk and statin response among 18,348 randomized patients in the heart protection study. *J Am Coll Cardiol* 2011;**57**:2000–2007.
- Verschuren JJ, Trompet S, Wessels JA, Guchelaar HJ, de Maat MP, Simoons ML, Jukema JW. A systematic review on pharmacogenetics in cardiovascular disease: is it ready for clinical application? *Eur Heart J* 2012;**33**:165–175.
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009;**106**:9362–9367.
- Morgan AA, Chen R, Butte AJ. Likelihood ratios for genome medicine. *Genome Med* 2010;**2**:30.
- Goldstein DB. Common genetic variation and human traits. *N Engl J Med* 2009;**360**:1696–1698.
- Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;**2**:e124.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747–753.
- Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Magi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Zillikens MC, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Zhao JH, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmaier EL, Amin N, Aspelund T,

- Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, Konig IR, Kwan T, Lawrence RV, Levinson DF, Lorentzon M, McKnight B, Morris AP, Muller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E, Sanna S, Shi J, Sovio U, Thompson JR, Turchin MC, Vandenput L, Verlaan DJ, Vitart V, White CC, Ziegler A, Almgren P, Balmforth AJ, Campbell H, Citterio L, De Grandi A, Dominiczak A, Duan J, Elliott P, Elosua R, Eriksson JG, Freimer NB, Geus EJ, Glorioso N, Haiqing S, Hartikainen AL, Havulinna AS, Hicks AA, Hui J, Igl W, Illig T, Jula A, Kajantie E, Kilpelainen TO, Koivari M, Kolcic I, Koskinen S, Kovacs P, Laitinen J, Liu J, Lokki ML, Marusic A, Maschio A, Meitinger T, Mulas A, Pare G, Parker AN, Peden JF, Petersmann A, Pichler I, Pietilainen KH, Pouta A, Ridderstrale M, Rotter JJ, Sambrook JG, Sanders AR, Schmidt CO, Sinisalo J, Smit JH, Stringham HM, Bragi Walters G, Widen E, Wild SH, Willemssen G, Zagato L, Zgaga L, Zitting P, Alavere H, Farrall M, McArdle WL, Nelis M, Peters MJ, Ripatti S, van Meurs JB, Aben KK, Ardlie KG, Beckmann JS, Beilby JP, Bergman RN, Bergmann S, Collins FS, Cusi D, den Heijer M, Eiriksdottir G, Gejman PV, Hall AS, Hamsten A, Huikuri H, Iribarren C, Kahonen M, Kaprio J, Kathiresan S, Kiemeny L, Kocher T, Launer LJ, Lehtimäki T, Melander O, Mosley TH Jr, Musk AV, Nieminen MS, O'Donnell CJ, Ohlsson C, Oostra B, Palmer LJ, Raitakari O, Ridker PM, Rioux JD, Rissanen A, Rivolta C, Schunkert H, Shuldiner AR, Siscovick DS, Stumvoll M, Tonjes A, Tuomilehto J, van Ommen GJ, Viikari J, Heath AC, Martin NG, Montgomery GW, Province MA, Kayser M, Arnold AM, Atwood LD, Boerwinkle E, Chanock SJ, Deloukas P, Gieger C, Gronberg H, Hall P, Hattersley AT, Hengstenberg C, Hoffman W, Lathrop GM, Salomaa V, Schreiber S, Uda M, Waterworth D, Wright AF, Assimes TL, Barroso I, Hofman A, Mohlke KL, Boomsma DI, Caulfield MJ, Cupples LA, Erdmann J, Fox CS, Gudnason V, Gyllenstein U, Harris TB, Hayes RB, Jarvelin MR, Mooser V, Munroe PB, Ouwehand WH, Penninx BW, Pramstaller PP, Quertermous T, Rudan I, Samani NJ, Spector TD, Volzke H, Watkins H, Wilson JF, Groop LC, Haritunians T, Hu FB, Kaplan RC, Metspalu A, North KE, Schlessinger D, Wareham NJ, Hunter DJ, O'Connell JR, Strachan DP, Wichmann HE, Borecki IB, van Duijn CM, Schadt EE, Thorsteinsdottir U, Peltonen L, Uitterlinden AG, Visscher PM, Chatterjee N, Loos RJ, Boehnke M, McCarthy MI, Ingelsson E, Lindgren CM, Abecasis GR, Stefansson K, Frayling TM, Hirschhorn JN. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010;**467**:832–838.
36. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahou Z, Allayee H, Altschuler D, Anand SS, Andersen K, Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boehnke SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buyschaert I, Carlquist JF, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Diemert P, Do R, Doering A, Eifert S, Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de Faire U, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA, Kaplan LM, Kastelein JJ, Khaw KT, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Loley C, Lotery AJ, Mannucci PM, Mauouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Muhleisen TW, Muhlestein JB, Munzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nothen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schafer A, Schillert A, Schreiber S, Schrezenmeier J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoop JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WH, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann HE, Wild PS, Willenborg C, Witteman JC, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, Marz W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J, Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011;**43**:333–338.
37. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemssen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruukonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, König IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllenstein U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Doring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Althausen D, Althausen M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA Jr, Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JJ, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;**466**:707–713.
38. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010;**42**:565–569.
39. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Hegele RA. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 2010;**42**:684–687.
40. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010;**8**:e1000294.
41. Anderson CA, Soranzo N, Zeggini E, Barrett JC. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol* 2011;**9**:e1000580.
42. Wray NR, Purcell SM, Visscher PM. Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biol* 2011;**9**:e1000579.
43. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, MacArthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurler ME. Origins and functional impact of copy number variation in the human genome. *Nature* 2010;**464**:704–712.
44. Craddock N, Hurler ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatos E, Holmes C, Marchini JL, Stirrups K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Attwood A, Auton A, Ball SG, Balmforth AJ, Barrett JC, Barroso I, Barton A, Bennett AJ, Bhaskar S, Blaszczyk K, Bowes J, Brand OJ, Braund PS, Bredin F, Breen G, Brown MJ, Bruce IN, Bull J, Burren OS, Burton J, Byrnes J, Caesar S, Clee CM, Coffey AJ, Connell JM, Cooper JD, Dominiczak AF, Downes K, Drummond HE, Dudakia D, Dunham A, Ebbs B, Eccles D, Edkins S, Edwards C, Elliot A, Emery P, Evans DM, Evans G, Eyre S, Farmer A, Ferrier IN, Feuk L, Fitzgerald T, Flynn E, Forbes A, Forty L, Franklyn JA, Freathy RM, Gibbs P, Gilbert P, Gokumen O, Gordon-Smith K, Gray E, Green E, Groves CJ, Grozeva D, Gwilliam R, Hall A, Hammond N, Hardy M, Harrison P, Hassanali N, Hebaishi H, Hines S, Hinks A, Hitman GA, Hocking L, Howard E, Howard P, Howson JM, Hughes D, Hunt S, Isaacs JD, Jain M, Jewell DP, Johnson T, Jolley JD, Jones IR, Jones LA, Kirov G, Langford CF, Lango-Allen H, Lathrop GM, Lee J, Lee KL, Lees C, Lewis K, Lindgren CM, Maisuria-Armer M, Maller J, Mansfield J, Martin P, Massey DC, McArdle WL, McGuffin P, McClay KE, Mentzer A, Mimmack ML, Morgan AE, Morris AP, Mowat C, Myers S, Newman W, Nimmo ER, O'Donovan MC, Onipinla A, Onyiah I, Ovington NR, Owen MJ, Palin K, Parnell K, Pernet D, Perry JR, Phillips A, Pinto D, Prescott NJ, Prokopenko I, Quail MA, Rafelt S, Rayner NW, Redon R, Reid DM, Renwick, Ring SM, Robertson N, Russell E, St Clair D, Sambrook JG, Sanderson JD, Schuilenburg H, Scott CE, Scott R, Seal S, Shaw-Hawkins S, Shields BM, Simmonds MJ, Smyth DJ, Somaskantharajah E, Spanova K, Steer S, Stephens J, Stevens HE, Stone MA, Su Z, Symmons DP, Thompson JR, Thomson W, Travers ME, Turnbull C, Valsesia A, Walker M,

- Walker NM, Wallace C, Warren-Perry M, Watkins NA, Webster J, Weedon MN, Wilson AG, Woodburn M, Wordsworth BP, Young AH, Zeggini E, Carter NP, Frayling TM, Lee C, McVean G, Munroe PB, Palotie A, Sawcer SJ, Scherer SW, Strachan DP, Tyler-Smith C, Brown MA, Burton PR, Caulfield MJ, Compston A, Farrall M, Gough SC, Hall AS, Hattersley AT, Hill AV, Mathew CG, Pembrey M, Satsangi J, Stratton MR, Worthington J, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand W, Parkes M, Rahman N, Todd JA, Samani NJ, Donnelly P. Genome-wide association study of CNVs in 16 000 cases of eight common diseases and 3000 shared controls. *Nature* 2010;**464**: 713–720.
45. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, Parish S, Barlera S, Franzosi MG, Rust S, Bennett D, Silveira A, Malarstig A, Green FR, Lathrop M, Gigante B, Leander K, de Faire U, Seedorf U, Hamsten A, Collins R, Watkins H, Farrall M. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med* 2009;**361**:2518–2528.
 46. Fakhro KA, Choi M, Ware SM, Belmont JW, Towbin JA, Lifton RP, Khokha MK, Brueckner M. Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proc Natl Acad Sci USA* 2011;**108**: 2915–2920.
 47. Kathiresan S, Melander O, Anefski D, Guiducci C, Burt NP, Roos C, Hirschhorn JN, Berglund G, Hedblad B, Groop L, Altshuler DM, Newton-Cheh C, Orho-Melander M. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 2008;**358**:1240–1249.
 48. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, Berglund G, Altshuler D, Nilsson P, Groop L. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 2008;**359**:2220–2232.
 49. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB Sr., Cupples LA. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 2008;**359**:2208–2219.
 50. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, Ridker PM. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 2010;**303**:631–637.
 51. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, Kivimaki M, Humphries SE. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *Br Med J* 2010;**340**:b4838.
 52. Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, Diver WR, Thun MJ, Cox DG, Hankinson SE, Kraft P, Rosner B, Berg CD, Brinton LA, Lissowska J, Sherman ME, Chlebowski R, Kooperberg C, Jackson RD, Buckman DW, Hui P, Pfeiffer R, Jacobs KB, Thomas GD, Hoover RN, Gail MH, Chanock SJ, Hunter DJ. Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 2010;**362**:986–993.
 53. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, Guiducci C, Perola M, Jula A, Sinisalo J, Lokki ML, Nieminen MS, Melander O, Salomaa V, Peltonen L, Kathiresan S. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* 2010;**376**: 1393–1400.
 54. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;**115**:928–935.
 55. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 2004;**159**:882–890.
 56. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem* 2008;**54**:17–23.
 57. Bilguvar K, Ozturk AK, Louvi A, Kwan KY, Choi M, Tatli B, Yalinzoglu D, Tuysuz B, Caglayan AO, Gokben S, Kaymakcan H, Barak T, Bakircioglu M, Yasuno K, Ho W, Sanders S, Zhu Y, Yilmaz S, Dincer A, Johnson MH, Bronen RA, Kocer N, Per H, Mane S, Pamir MN, Yalcinkaya C, Kumandas S, Topcu M, Ozmen M, Sestan N, Lifton RP, State MW, Gunel M. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 2010;**467**:207–210.
 58. Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Stehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BB, Veltman JA. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* 2010;**42**:483–485.
 59. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irikat RK, Tom W, Fantin NJ, Muzny DM, Gibbs RA. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 2010;**362**:1181–1191.
 60. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;**42**:790–793.
 61. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 2010;**42**:30–35.
 62. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;**461**:272–276.
 63. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A, Staker B, Pant KP, Baccash J, Borcherding AP, Brownley A, Cedeno R, Chen L, Chernikoff D, Cheung A, Chirita R, Curson B, Ebert JC, Hacker CR, Hartlage R, Hauser B, Huang S, Jiang Y, Karpinchy V, Koenig M, Kong C, Landers T, Le C, Liu J, McBride CE, Morenzoni M, Morey RE, Mutch K, Perazich H, Perry K, Peters BA, Peterson J, Pethiyagoda CL, Pothuraju K, Richter C, Rosenbaum AM, Roy S, Shafto J, Sharanovich U, Shannon KW, Sheppy CG, Sun M, Thakuria JV, Tran A, Vu D, Zaranek AW, Wu X, Drmanac S, Oliphant AR, Banyai WC, Martin B, Ballinger DG, Church GM, Reid CA. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010;**327**:78–81.
 64. Rare Diseases Act. <http://history.nih.gov/research/downloads/PL107-280.pdf> (17 April 2011).
 65. Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010;**328**:636–639.
 66. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC. The diploid genome sequence of an individual human. *PLoS Biol* 2007;**5**:e254.
 67. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Zhang G, Yang B, Yu C, Liang F, Li W, Li S, Ni P, Ruan J, Li Q, Zhu H, Liu D, Lu Z, Li N, Guo G, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C, Yang S, Chen F, Li L, Zhou K, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X, Kristiansen K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X, Yang H. The diploid genome sequence of an Asian individual. *Nature* 2008;**456**:60–65.
 68. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niaz F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XX, Liu Y, Yuan Y, Naz P, Ruan J, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008;**452**: 872–876.
 69. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;**31**:3812–3814.
 70. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002;**30**:3894–3900.
 71. National Center for Biotechnology Information (NCBI). The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res* 2009;**37**(Database issue):D169–D174.
 72. 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**(Database issue):D700–D706.
 73. Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN. The Human Gene Mutation Database: 2008 update. *Genome Med* 2009;**1**:13.
 74. Chen R, Butte AJ. The reference human genome demonstrates high risk of type 1 diabetes and other disorders. *Pac Symp Biocomput* 2010:231–242.
 75. Dewey FE, Chen R, Cordero SP, Ormond KE, Caleshu C, Karczewski KJ, Whirl-Carrillo M, Wheeler MT, Dudley JT, Byrnes JK, Cornejo OE, Knowles JW, Woon M, Sangkuhl K, Gong L, Thorn CF, Hebert JM, Capriotti E, David SP, Pavlovic A, West A, Thakuria JV, Ball MP, Zaranek AW, Rehm HL, Church GM, West JS, Bustamante CD, Snyder M, Altman RB, Klein TE, Butte AJ, Ashley EA. Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. *PLoS Genet* 2011;**7**:e1002280.
 76. Salari K. The dawning era of personalized medicine exposes a gap in medical education. *PLoS Med* 2009;**6**:e1000138.
 77. Ormond KE, Wheeler MT, Hudgins L, Klein TE, Butte AJ, Altman RB, Ashley EA, Greely HT. Challenges in the clinical application of whole-genome sequencing. *Lancet* 2010;**375**:1749–1751.