Published in final edited form as:
Circulation. 2010 November 30; 122(22): 2323–2334. doi:10.1161/CIRCULATIONAHA.109.909309.

Genetic Cardiovascular Risk Prediction – Will We Get There?

George Thanassoulis, MD and **Ramachandran S. Vasan, MD**
NIH/NHLBIs Framingham Heart Study and Faculty of Medicine Boston University

Introduction

Major advances in genetics, including the sequencing of the human genome in 20011›2 and the publication of the HapMap in 2005[3], have paved the way for a revolution in our understanding of the genetics of complex diseases, including cardiovascular disease (CVD). After years of inconsistent results and failures to replicate putative candidate gene associations, high throughput technologies (that genotype over 500,000 genetic markers, known as single nucleotide polymorphisms [SNPs]) and novel statistical tools have led to a literal explosion of novel genetic markers associated with complex human diseases. In the context of CVD, these advances have been remarkably successful in uncovering many novel genetic associations with myocardial infarction and cardiovascular risk factors such as lipids, blood pressure, diabetes and obesity. A major objective of these studies has always been to provide new insights into the biology of cardiovascular disease. However, a highly touted additional aim of these discoveries has been to use these genetic markers to usher in a new era of personalized medicine by incorporating genetic information into risk prediction (including for the primary prevention of CVD). In fact, direct-to-consumer testing of recently discovered genetic markers has proliferated despite a lack of evidence for clinical use.[4]

As with all nascent technologies, many fundamental questions remain to be answered: Can genetic markers or gene scores improve CVD risk prediction, over and above, validated risk algorithms such as the Framingham risk score and a family history of CVD? How many SNPs are responsible for the genetic component of CVD, and how many genetic markers will we need to discover to reliably improve risk prediction? What are the implications of the allelic architecture of CVD and other complex diseases for risk prediction? And, finally, what steps will be needed prior to bringing this information to patients? In this review, we will examine each of these questions with regards to risk prediction of coronary artery disease (CAD) and myocardial infarction (MI) in a primary prevention setting.

Cardiovascular Risk Prediction – Is there a need for improving currently used algorithms?

For over five decades, the major cardiovascular risk factors, namely male sex, hypertension, cholesterol, smoking and diabetes have been well known.5 Based on these factors, a number of risk prediction algorithms scores have been developed, including the Framingham risk

Address Correspondence to: Ramachandran S. Vasan, MD Framingham Heart Study 73 Mt Wayte, #2 Framingham, MA 1703 Tel. 508-935-3450 Fax. 508-626-1262 vasan@bu.edu.

Disclosures: None

score, that provide an estimate of the 10-year risk (and recently, the 30-year risk) of CVD.[6-9] Generally speaking, the metrics used to assess risk scores include an assessment of their performance for risk discrimination, calibration, risk reclassification, and clinical utility (change in management and patient outcomes associated with their implementation) (reviewed by Cook et al.10). Discrimination is summarized by the c-statistic, which represents the area under the receiver–operating-characteristic curve (plotting 'sensitivity' in relation to '1-specificity'). The numerical value of the c-statistic represents the probability of correctly ranking two randomly selected individuals (ine likely to develop disease and one not so predisposed) based on their predicted risk from the prediction model under evaluation. Calibration compares the predicted risk with the observed risk in groups of individuals classified by risk level and provides a measure of the overall accuracy of the risk estimates derived by the model. Reclassification, a newer metric, estimates the improvement (or lack threreof) in risk classification of individuals using a novel marker compared to a standard model without the marker of interest. Most currently used risk scores have been validated in many populations and have been shown to have good discriminatory capacity and calibration. However, risk reclassification and clinical utility of many of these scores has been less well studied and remains an area of active investigation.

One common reason offered for pursuing newer risk factors/markers is that current risk scores explain a modest proportion of CVD incidence in the community. Indeed, a common misconception is that only 50% of the incidence of CVD is explained by the traditional risk factors and, therefore, novel markers of pre-clinical disease are needed to refine contemporary risk prediction algorithms.[11] In fact, the major risk factors explain a large proportion of the risk of CVD.[12] However, it is estimated that nearly 15-20% of MI patients have none of the traditional risk factors and would be considered "low risk" by current risk prediction scores.[13]

Whereas the importance of the traditional risk factors and the utility of current risk prediction algorithms cannot be ignored, efforts to improve risk prediction are needed given that CVD is preventable, the first manifestation may be sudden cardiac death, the occurrence of a MI is associated with a high early mortality, and survivors may suffer considerable morbidity and a reduction in the quality of their lives. This had led to an intense search for novel biomarkers that can enhance the currently available risk scores. However, the majority of studies that have claimed to identify novel biomarkers that enhance risk prediction beyond the Framingham risk score have been noted to have flaws in their design, analysis or interpretation.[14] Recently, an expert group has proposed criteria for appropriate assessment of the clinical utility of novel biomarkers for the purpose of enhancing CVD risk prediction.[15]

One category of biomarkers that has evoked extensive study recently is that of genetic variants. As for other biomarkers, the case for risk prediction using genetic polymorphisms must be held up to the same standards. Incremental improvement of the addition of a genetic biomarker must be shown over and above well-validated risk scores using standard metrics to evaluate their clinical performance. Therefore, useful genetic markers for risk prediction will need to be sufficiently uncorrelated with known CVD risk factors as to provide independent information regarding risk. It can also be argued that any genetic marker should also provide incremental risk information, over and above, a model which incorporates family history[8,16,17], given that such information is often readily available.

The interest in incorporating genetic information into risk prediction algorithms stems from the fact that many MI patients have a family history of the CVD, which confers a nearly doubling of the risk of MI among family members, even after adjusting for traditional cardiovascular risk factors.[18,19] Moreover, the heritability of MI, which provides an estimate

of the genetic variance in MI risk, has been estimated at 40-60%.[20,21] These findings suggest that genetic factors play an important role in cardiovascular disease and could potentially refine risk prediction for CVD. Indeed, recent genetic studies have confirmed that several genetic factors are associated with MI and CAD, which has led to much excitement regarding the possibility of their use for risk prediction.

Genome-Wide Association Studies - Initial Successes in gene discovery

To date, genome-wide association studies (GWAS) have been remarkably successful in uncovering many novel genetic loci not previously implicated in cardiovascular disease. Among the first major discoveries for CVD were the simultaneous reports of the association of variants at the 9p21 locus with MI.[22,23] This association represents one of the most consistent and robust SNP-disease associations in the GWAS era, having been replicated in several independent samples in numerous ethnicities. A recent large meta-analysis reported an allele relative risk of 1.27 (95% CI 1.23-1.31, $p = 1.62 \times 10^{-12}$) for CAD.[24] Despite the strength and consistency of the associations of 9p21 variants with CVD traits, little is known about the biological role of this locus and how it may confer increased CVD risk. This SNP is located in a gene desert, with the nearest gene, *ANRIL*, being several thousand kilobases away. Despite significant effort, the function of *ANRIL*, a non-coding RNA, has not yet been elucidated but continues to be an area of active investigation. In a recent study, the mouse ortholog of the 9p21 locus (CDKN2A/2B) associated with CAD was successfully knocked out. These mice had reduced survival, a more rapid increase in body weight, and a hyperproliferative smooth muscle cell phenotype suggesting altered vascular cell dynamics.[25] If these findings are confirmed this would represent a novel mechanism for myocardial infarction that is unrelated to conventional risk factors. Several additional GWAS for MI have also been completed identifying a number of novel loci (Table 1).[26-30] In addition, GWAS for blood pressure,[33-35] and lipid traits[36-38] have also discovered additional loci that are associated with these traits. “To date, over 100 new genetic variants have been discovered that relate to MI or MI risk factors using GWAS[39] (13 SNPs have been replicated for MI; 26 for high density lipoprotein levels, 16 for low density lipoprotein levels; 26 for triglycerides levels; 42 for diabetes and fasting glucose; 10 for hypertension; 6 for C-reactive protein levels and 16 for body mass index).” These SNPs have been rigorously replicated in one or multiple additional independent studies confirming that they represent genuine true associations with CAD. After years of inconsistent results,[40,41] these studies have provided an early glimpse at the underlying genetic risk of CAD. However, these initial studies represent the first steps towards understanding the complete allelic architecture of CAD and it is likely that many more genetic variants remain to be discovered[42]. Yet, despite our limited understanding of genetic risk for CAD, a number of studies have attempted to incorporate these newly discovered genetic risk variants into CVD risk prediction tools with limited initial success (as reviewed below; Table 2).

Use of Genetic Information for Cardiovascular risk prediction – Overview of initial experience with single genetic variants and genetic risk scores

Several studies have evaluated the predictive power of the addition of single SNPs and combinations of risk SNPs into genetic risk scores on MI risk based on pre- and post-GWAS results (Table 2). Additions of single SNPs at 9p21 to the Framingham risk score have not been found to consistently improve risk prediction.[44,47] A genetic risk score incorporating 9 CAD associated SNPs resulted in a >2-fold higher odds ratio for MI in subjects in the highest quintile of the risk score compared to those in the lowest quintile but did not evaluate the incremental value of the addition of such a score to traditional risk factors.[28] In a separate report, a genetic risk score using SNPs strongly associated with lipid levels conferred a 15% increase in CAD risk per lipid-associated SNP allele.[46] Despite the

increased CAD risk per allele, the genetic score did not improve discriminative ability over and above traditional risk factors, and showed only modest improvement in risk reclassification. The limited success of these initial studies has led to the development of more elaborate genetic risk scores comprising many SNPs encompassing both MI risk alleles and SNPs associated with other cardiovascular risk factors in an effort to increase the genetic risk explained and to improve the predictive performance of genetic risk scores. A genetic risk score comprising 101 validated SNPs from large-scale GWAS of MI and other cardiovascular risk factors was evaluated for cardiovascular risk prediction in over 18 000 women. After adjustment for traditional risk factors, the genetic risk score was not associated with CVD events and the addition of the genetic risk score to a standard risk prediction model did not significantly improve discrimination or reclassification.

Despite incorporating multiple CAD associated SNPs, genetic risk scores to date have explained less than 5% of the inter-individual variance in risk[53] and have not led to clinically meaningful improvements in risk prediction. However, the modest improvements in risk reclassification seen in some of these studies[45,46] highlight the future potential for the use of genetic markers for risk prediction, as additional genetic variants are discovered.

Although the results of these initial studies of genetic risk prediction have been underwhelming, several important insights regarding the future of cardiovascular genetic risk scores can be gleaned from these initial studies of genetic risk prediction. First, the incremental predictive utility of genetic risk scores that explain a small fraction of the heritability will likely be marginal. In order to significantly improve risk prediction, genetic risk assessment will need to be markedly refined. It has been suggested that approximately 20% of the heritability needs to be explained to provide similar discrimination as obtained from standard risk prediction models.54 Second, the addition of predictors with relative risks <10 will have limited effect on risk discrimination,[55] although their contribution to risk reclassification warrants further study. Therefore, it is not surprising that the addition of recently discovered genetic markers for MI, with relative risks ranging from 1.1-1.3, have had led to limited improvements in risk prediction. For useful risk prediction, genetic risk scores with many additional markers will likely be needed to improve contemporary CVD risk stratification algorithms. It must also be emphasized that relative risks across extreme comparisons (i.e. top quintile of genetic risk to bottom quintile of genetic risk), which are often reported for genetic risk scores, are not relevant for risk prediction. For translation to risk prediction, the reported risks should be compared with ‘average’ risks found in the general population where risk prediction will likely be used in a primary prevention setting.

‘Effect’ Estimates from Initial Discovery GWAS – Caveats and Implications for risk prediction

An important additional consideration explaining the limited success of genetic risk prediction is that effect estimates for associations from discovery GWAS may be biased and of limited utility in risk prediction. First, genetic effect estimates from GWAS are likely inflated due to the “winner’s curse”, where early reports of relatively large effect sizes become attenuated with further replication in studies of increasing sample sizes.56,57 Second, estimates from meta-analysis in the genetic literature frequently assume ‘fixed effects’ despite important between study heterogeneity. It has been demonstrated that some markers become statistically non-significant at a genome-wide level when ‘random effects’ models are used and therefore may have poor generalizability across populations. 53,58 Third, several GWAS have used extreme subjects to identify genetic associations by sampling high genetic risk cases (i.e “hypercases” that are frequently younger with less risk factors and a positive family history) and low genetic risk controls (i.e. hypercontrols who lack such factors) further inflating effect estimates.[59] Moreover, the odds ratios generated

from case-control GWAS to date (which have used prevalent cases and controls) are likely overestimating the true risk ratio as MI is not a "rare disease" and controls have not been sampled using an appropriate sampling strategy (e.g. incidence-density sampling) to provide odds ratio estimates that approximate risk ratios.[60,61] Lastly, most GWAS to date may also suffer from major potential survival biases as enrollment into the study is conditional upon survival post-MI. Given that 30-70% of MI patients die prior to admission to the hospital,[62-64] analyses of prevalent MI cases are poorly representative of most incident MIs.

While these practices are acceptable for gene discovery, the relative risk estimates associated with putative genetic variants are unlikely to be applicable to the general population where risk prediction is applied for a future time horizon. Accordingly, it is likely that such estimates will perform poorly in prospective assessments of CVD prediction in a primary care setting. These observations highlight an important point – that if risk prediction is the objective, then GWAS of incident CVD are needed in large prospective cohort studies of representative populations to complement currently available studies. To date, and to our knowledge, there have been no published GWAS for incident CVD using a prospective cohort design despite calls stressing the importance of such a study design in genetic epidemiology.[65]

How Many SNPs do We Need for CVD prediction? – Theoretical predictions using the c-statistic

Despite criticisms that the c-statistic is insensitive to most biomarkers studied in the "-omics era",[66] it still represents an important, yet perhaps overly conservative, starting point for evaluating the possibility of using genetic variants or risk scores in CVD risk assessment. A number of simulation studies have provided important information regarding the feasibility of genetic risk prediction in CVD using the c-statistic as the metric for discrimination. We review these investigations, acknowledging that similar simulations are warranted using additional newer metrics such as risk reclassification.

For genetic studies, the c-statistic is a function of the heritability, the genetic variance explained by the genetic variants, the prevalence of the disease condition, and the minor allele frequency in the population.[67,68] For CVD, it can be estimated that the upper bounds of the c-statistic is ~0.90, for populations with 10% prevalence of disease, and ~0.85 for populations with a higher disease prevalence.[67,69] While achieving the maximum c-statistics would provide excellent discrimination, this would require identifying all the genetic variants associated with CVD. A more reasonable goal would be to achieve c-statistics ~0.80-0.85, which would still represent an improvement over current risk prediction models. It has been estimated that to achieve this level of discrimination, ~100 uncorrelated genetic variants (i.e., in linkage equilibrium) with relative risks of ~1.5, minor allele frequencies of 10%, and explaining ~20% of the heritability of CVD would be needed.[67] However, few CVD genetic variants identified in the GWAS or pre-GWAS eras with relative risks in this range. If we assume that CVD genetic variants will have mean relative risks of 1.1-1.2, then even 100 genetic variants would only explain 1.0-9.1% of the variance of CVD and provide c-statistics ~0.75,[68] which would be similar but not much better than current prediction models. To achieve a higher level of discrimination based on the genetic relative risks observed most frequently to date (i.e., RR 1.1-1.25), it can be estimated that 150 to >400 genetic variants would be needed depending on the frequency of these genetic variants (i.e., >5% minor allele frequency [MAF]).[67]

While the large number of genetic variants required for a useful genetic risk score may seem daunting, it can be argued that the these simulations may artificially inflate the estimated number of genetic variants needed for reliable CVD risk prediction due to reliance on the c-

statistic, which may be insensitive to the small effects of genetic variants. Using a method that relied on the accuracy of genetic risk prediction (i.e. genetic variance explained by the genetic score) instead of the c-statistic, it has been reported that many fewer genetic variants may be needed for useful CVD risk prediction.70 Under the assumption that a reasonable number of genetic variants (i.e. <1000) with weak effects explain the heritability of CVD, the authors estimate that a large case-control study with 20,000 subjects and a 1:1 case to control ratio, could identify 80-120 genetic variants explaining >50% of the genetic variance of the disease.70 They also demonstrate that this set of genetic variants would likely have good to excellent calibration in a validation dataset and would have relative risks of ~4.0 for the top 5% of subjects based on the genetic risk score compared to the average risk in the population. However, it should be noted that this inference is based on simulation modeling an assumed risk of disease (prevalence of 5-10% with a sibling relative risk of 1.45-2.90). In addition, published empirical studies have not reported on the proportion variance explained as a potential metric for risk prediction, thereby underscoring the need for additional research. Although there is some debate regarding the optimal metric for assessing the utility of genetic markers in simulation studies (and the appropriate genetic models for these simulations), these studies provide valuable insight on the future feasibility of genetic risk prediction. Similar evaluations of novel metrics, such as the net reclassification index, have not been reported as yet and would provide important new information for assessing genetic risk prediction.

While apparently encouraging for genetic risk prediction, these simulations depend on assumptions regarding the total number of genetic variants needed to explain the totality of genetic variation in CVD. If the heritability of CVD can be explained by 100-1000 SNPs, then risk prediction may be possible with a relatively small number of SNPs (~100) in the near future. However, the total number of SNPs that explain the heritability of CVD is currently unknown. Recent studies examining the allelic architecture of CVD using simulated data has shed some light on this problem and have suggested that the genetic component of CVD (and other complex diseases) may be under the influence of many more genetic variants (or other factors) than anticipated,[42,71] which could have profound implications for risk prediction.

GWAS and the Allelic Architecture of CVD – How many SNPs explain the genetic risk of CVD?

GWAS were designed based on the theory that the genetic architecture of complex diseases would follow the “common disease-common variant” hypothesis that predicts that common diseases, such as CAD, are caused by many common genetic variants, each explaining a small portion of the variance in the risk of disease.[72] Most detected genetic variants have allele frequencies >5%, have small to very small effect sizes (i.e. relative risks of 1.1-1.3), and each explain <1% of the variance in risk of disease. While theory posits low effect sizes for complex disease, the very weak effect sizes and the low variance explained by recently uncovered SNPs has been somewhat unexpected. These findings have profound implications for risk prediction as the total number of genetic variants needed to explain the heritability of a disease is proportional to the proportion of variance explained by each genetic variant. It has been argued that if indeed current GWAS have detected the common SNPs that explain the highest fraction of the genetic variance, which may be quite likely, the remaining variants to be found will explain *exponentially smaller* proportions of the remaining genetic variance.42 GWAS data for complex traits (such as diabetes and height) have been shown to follow this pattern quite convincingly.42,73 Many prior simulations studies estimating the total number of SNPs required to explain heritability or the number of genetic variants required for accurate risk prediction have not considered these recent insights into the genetic architecture of CVD (and related traits) and have frequently oversimplified the

models by assuming fixed genetic relative risks for each genetic variant that remains to be identified. Such simplifications would markedly underestimate the number of SNPs that explain the heritability of CVD.

Given that the strongest common SNP associated with CAD has a relative risk of 1.3 and other recently identified variants have relative risks of 1.1 to 1.2, the remaining undiscovered variants are expected to have small to very small effect sizes, and therefore it is possible that hundreds to thousands of genetic variants be needed to explain the relatively high heritability of CVD. This may pose a challenge for risk prediction for a number of reasons. First, to demonstrate robust disease associations with thousands of genetic variants with weak effects would require studies with >100,000 individuals, which would be larger than even the largest genetic consortia currently in place.[74] Second, as the number of genetic variants implicated in CVD increases, the possible combinations of risk alleles become unmanageable such that every individual would have a unique genetic signature with little overlap between individuals, making genetic risk prediction very challenging if not nearly impossible. Higher risk genetic profiles requiring many hundred genetic variants would also be exceedingly rare.[75] Third, if many thousand SNPs are required to explain a substantial proportion of the variance of CVD, it has been argued that most individuals would have many of these “risk variants,” which could seriously hamper their usefulness for risk prediction. As one author has said presciently: “In pointing at everything, genetics would point at nothing”.[42] To date, such dire pronouncements have rung true as genetic risk prediction has not been successful across the spectrum of complex disease. However, a recent study using a novel approach to genetic risk scores, has provided some evidence that these initial concerns may be unfounded and that genetic risk prediction could still be possible even if thousands of genetic variants are needed.[76,77]

Using GWAS data from the Wellcome Trust Case Control Consortium, [76] a genetic risk score was constructed by including the top SNPs ranked by statistical significance. Instead of limiting the genetic risk score to genome-wide significant SNPs, they lowered the statistical threshold to include many more SNPs acknowledging that some would likely represent false positives. In doing so they also captured many unidentified true positives increasing the genetic variance explained. Using this genetic risk score on a validation sample, they demonstrate that as the p-value threshold for inclusion into the risk score is lowered, the c-statistic increased for many (but not all) common diseases, including CVD. Despite low c-statistics (~0.70) and low levels of genetic variance explained (~3%) for risk scores that included thousands of SNPs, they showed by simulation that as the sample size of the discovery data set increased, the top SNPs would be enriched for true positives compared to false positives. Using a discovery set with 10,000 cases and 10,000 controls, the top 5% associated SNPs would be expected to explain up to 20% of the genetic variance. This analysis is noteworthy for a number of reasons. First, it provides empirical evidence that there may be many SNPs with weak effects that are beyond the detection limit of current GWAS (using contemporary statistical thresholds for genome-wide significance). In particular for CVD, even when the threshold was relaxed to as low as $p<0.80$, the c-statistic continued to increase, suggesting that the genetic architecture of CVD may consist of numerous SNPs with very weak effects that are currently undetectable. Second, this study implies that it may be possible to construct useful genetic risk scores without actually identifying most of the true positive associations at a genome-wide significance level. Whether empirical evidence will confirm these findings and prove the utility of this method remains to be seen, emphasizing the need for additional research.

The Potential Causes of 'Missing Heritability' in initial GWAS

Although GWAS has uncovered many novel genetic associations in several complex diseases, the overall genetic variance explained by these associations has been lower than expected. For CVD, the very low variance explained by recently discovered SNPs, despite large studies with substantial statistical power, implies that either hundreds to thousands of genetic variants with very low effects are needed to explain most of the heritability of CVD, as discussed above, or that other factors explain the so-called "missing heritability" (which has been dubbed "the dark matter" of human genetics).[71]

Explanations for the "missing heritability" in CVD and other complex diseases include overestimation of heritability using conventional methods, measurement imprecision of phenotypes, gene-gene and gene-environment interactions, linkage disequilibrium of associated SNPs with 'true' causal variants, existence of low penetrant variants, and the potential contributions of structural variation, epigenetic modifications and rare genetic variation to disease risk.[71]

A primary concern, prior to evaluating the missing heritability, is the accuracy of current estimates of heritability. Heritability represents the proportion of the total variance in the phenotype explained by genetic factors. Given that familial clustering of disease is due to both a shared environment and shared genes, and that genetic factors are subject to much less misclassification than environmental exposures, heritability estimates (using current family-based methods) can be confounded by poorly measured shared environmental factors.[71,78] Recent studies in quantitative traits using novel methods to estimate heritability suggest that current methods are likely unbiased79 and, therefore, incorrect heritability estimates are unlikely to explain a major component of the missing heritability in complex disease.

A large component of the missing heritability in complex disease may be due to interactions – both gene-gene and gene-environment interactions. Interactions can be viewed simplistically as combinations of risk predictors where the combined presence of two predictors leads to larger (or smaller) effects than expected for either predictor alone. Failure to incorporate interactions into GWAS has likely led to an underestimation of the true genetic effects and reduced statistical power to identify novel genetic variants. Given that populations may be made up of genetically "susceptible" and "null" subpopulations based on their environmental co-exposures (so-called context-specific genetic effects) and that genetic risk estimates are weighted averages of the risks in both sub-populations, ignoring interactions will often bias genetic exposures to the null when they are present.[80] Interactions are, however, difficult to study as they can occur at many levels and add to the already large number of statistical tests performed in GWAS. However, larger sample sizes, novel analytic methods and a "systems biology" approach may soon uncover important interactions which could explain a significant portion of the heritability.81,82 Improvements in our understanding of interactions would also be expected to improve genetic risk prediction models.

Linkage disequilibrium of associated SNPs with true causal variants and the existence of poorly penetrant genetic variants may also explain some additional genetic variance. Deep resequencing efforts are ongoing in an effort to identify these causal variants.[83] Copy number variants, a form of structural genetic variation including small deletions, insertions or inversions in the genome, have also been proposed as an additional source of missing heritability. Fortunately, most common copy number variants are relatively well 'tagged' by HapMap SNPs and are well represented in GWAS.[84] In the MIGen consortium, no CNVs were found to be associated with MI despite good coverage of common CNVs.[28] Rare

CNVs could potentially represent an important source of genomic variation that would not be detected by GWAS, a possibility that has been recently demonstrated for obesity.85 While this may also be true for CVD, this possibility invokes an important consideration - that most genetic variation in the genome, whether they represent SNPs or CNVs, may be rare. If most of the genetic variation resides in rare variants not captured by HapMap SNPs, this variation would not be detected by GWAS. Indeed, studies conducted on lipid traits have demonstrated that a significant portion of the genomic variability was attributable to rare variants detectable only by sequencing.86 Current efforts to sequence the genome in large cohorts of individuals are underway.87,88 It is very likely that these strategies will uncover many new genetic variants that are biologically linked with disease; however, unless they identify rare variants that explain relatively large proportions of genetic variance (i.e. with strong effects), their value for risk prediction will likely be limited. Furthermore, because rare variants may be specific to certain groups (i.e., founder populations or families) their impact on risk prediction in the general population may be very modest. While the current trend of pooling samples in increasing larger consortia to maximize statistical power for GWAS has led to some additional SNP discovery, it is likely that this approach will soon be exhausted, and more refined approaches will need to be prioritized in future efforts (Table 2). Several additional strategies to identify the missing heritability have recently been suggested.[71] These include assessment of analytical pooling strategies for rare variants, targeting recent African American samples with narrow LD for resequencing efforts (to uncover rare variants) and studies of family-based cohorts (in which susceptibility alleles/ risk variants are likely enriched).[71]

A final point must be made regarding the missing heritability of CVD and the phenotypes that are currently being evaluated, which may represent “crude” representations of disease. Disease classifications have evolved in clinical medicine by fitting similar patterns of symptoms into categories, often organized by organ system, in order to reduce complexity and simplify diagnosis. While this practice has been extremely helpful for managing and treating disease, it is increasingly apparent that these crude divisions of diseases may be suboptimal for etiologic genetic research. In CVD, it is very conceivable that MIs, like many other diseases, could be classified by primary pathophysiological process, for example, by increased propensity for endothelial dysfunction, accelerated atherosclerosis or for thrombosis. MIs may represent the culmination of multiple different causal pathways, with each pathway having its own set of genetic associations. Using new tools from the “–omics” toolbox, disease phenotypes based on causal pathways could be extended to cellular and molecular profiles. This concept of refining disease phenotypes to produce distinct phenotypes of increasing homogeneity has been described as “deep phenotyping”.[89] Deep phenotyping disease could lead to improvements in the resolution of genetic signals and provide increasingly specific genetic insights that may enhance future genetic research.

Using family history of CVD as a marker of genetic risk for CVD

A parental history of premature CVD is a well-established risk factor for incident CVD.[18,19] While a portion of the familial aggregation of CVD is mediated by non-genetic factors, a positive family history of premature CAD is thought to represent a good surrogate for an increased genetic risk. Whereas individual genetic variants or genetic risk scores have not yet led to significant improvements in risk prediction, the addition of family history improves risk reclassification and has been formally added to risk prediction models.8,16,19 In a recent study incorporating over 100 CVD risk SNPs to a prediction model that includes family history, the genetic risk score was not associated with incident CVD, but the association with family history remained strongly associated.43 In fact, the magnitude of the association between family history and CVD was not attenuated when the genetic variants were added suggesting that current genetic variants explain only a very small fraction of the

familial risk. It is possible that as additional common variants associated with CVD are uncovered, a larger proportion of the familial risk will be explained. However, an alternative possibility is that a significant proportion of the familial genetic risk is related to other shared environmental factors (including behavioral and lifestyle factors) or to rare familial genetic variants. Indeed, it possible that a significant proportion of the familial genetic risk may, in fact, be specific to a given family (i.e. “private genetic epidemiology”90), which could significantly hamper risk genetic risk prediction in the general population, unless a few rare genetic variants (with large effects) explained a large portion of the familial risk (as seen with BRCA1 and BRCA2 in breast cancer).

Although family history may currently represent the best marker for the genetic risk of complex disease, a number of important limitations exist in using family history as a marker of genetic risk. A major limitation stems from the fact that family history predicts the same risk for all members of the immediate family despite the fact that 50% of the genetic variation occurs within families. Furthermore, even under ideal conditions of complete ascertainment of family history over 3 generations, up to 55% of CVD cases are expected to have no family history.[91] In a recent simulation study, the maximal value of AUC for family history of CVD under such idealized conditions was only 0.71 (as compared to >0.90 for a genetic risk score explaining 100% of the genetic variance of CVD)[69]. However, a complete multigenerational family history would still explain 16% of the genetic variance of CVD which is currently significantly better than any reported genetic risk score for CVD. This underscores the importance and clinical utility of ascertaining family history of CVD as a marker of genetic risk at the current time.

Translating Genetic Risk Prediction to Clinical Use

Using a combination of approaches outlined above, it is plausible that a greater portion of the genetic variance of CVD will be explained in the near future. Whether this will lead to genetic risk prediction that can be useful in clinical practice remains to be determined. While high levels of prognostic performance for genetic risk prediction (i.e. discrimination, reclassification, calibration) are important, many additional considerations exist (Table 3).

First, any new predictive model will require validation in independent cohorts. It is currently unclear whether genetic risk scores will be transportable across ethnicities and races with varying allele frequencies and environmental exposures92. Independent validation and recalibration may be required for use in such populations. Second, appropriate cut-offs for risk scores are needed to optimize false positive and false negative rates depending on their relative importance[53,59,93]. Furthermore, similar to many other cardiovascular tests, genetic risk scores will also be subject to Bayes’ theorem implying that inappropriate use of these tests will lead to poor predictive accuracy. Reports of predictive values and likelihood ratios for populations of varying cardiovascular risks would be needed to provide guidance as to which patients would benefit most from screening. Whether genetic risk scores should target only individuals with a family history of CVD or those with intermediate CVD risk (where the prevalence of CVD and the likelihood of true positives may be higher) rather than the general population, warrants further study.[94,95] Cost will also represent an important consideration in deciding which segments of the population should have such testing[93,96]. Third, the level of evidence required for genetic risk scores prior to clinical use will need to be robust. Clinically significant improvements in predictive performance (that are also cost-effective) should represent the threshold for clinical utility. Marginal improvements, that meet an arbitrary threshold for statistical significance, will not suffice for translation to clinical use. Fourth, the type of evidence required for clinical use will also need to be clarified. Whether genetic risk scores will require “biomarker trials” where a management strategy using a risk score is compared to a strategy without the use of the risk score[97], or

whether a well validated risk score with good prognostic performance in prospective cohorts will be sufficient evidence for clinical use will need to be determined. Fifth, education of patients and clinicians on the use and interpretation of such risk scores will also be needed to limit genetic determinism. The advantage of genetic risk scores is that they remain stable throughout life and can predict the genetic risk of disease at any age, however their intransience may also be a liability as it may be challenging to use them to assess efficacy of treatment or other risk reduction strategies. In fact, it could be argued that for any genetic variant that has a measurable product, either in blood or other accessible tissue, it would be much more useful to track the product than to obtain a genotype, obviating the need for genetic risk scores altogether. Whether genomic risk scores will eventually become a reality or will be complemented or superseded by proteomic or other -omic risk models remains to be seen.

Conclusions

At this early stage in the GWAS era, many questions remain regarding the feasibility and utility of genetic risk prediction for CVD. While genetic information is far from ready for clinical use in CVD prediction, genetics have made important clinical inroads in other areas, such as pharmacogenomics for predicting efficacy and adverse events of common cardiovascular drugs (reviewed in 98). Whether “we will get there” for genetic CVD prediction, as we have asked in the title of our review, remains an open question. Cardiovascular disease may represent a particularly difficult phenotype for genetic risk prediction. Nonetheless, we hope that the many challenges faced for genetic CVD risk prediction will not be insurmountable and that novel strategies will lead to a greater understanding of the heritability of CVD. However, we must also appreciate the complexity of the human genome, and the challenges inherent in achieving the goal of personalized medicine in CVD risk prediction. At this stage, clinicians should continue to inquire about family history for risk prediction, as this continues to represent a simple, cheap and clinically useful risk factor for CVD which likely represents the net effect of hundreds of genetic risk variants which have yet to be discovered. Regardless, of whether genetic information will be used clinically in CVD risk prediction, GWAS have been a resounding success for cardiovascular medicine by identifying many genetic variants not previously linked with CVD that will undoubtedly provide novel mechanistic insights in the years to come.

Acknowledgments

Funding sources: Dr. Thanassoulis is supported by a Research Fellowship from the Canadian Institute of Health Research and the Fonds de Recherche en Santé du Québec. Dr. Vasan is supported in part by NIH contract NO1-HC-25195 and RO1-HL93328.

References

1. Initial sequencing and analysis of the human genome. Nature. 2001; 409(6822):860–921. [PubMed: 11237011]
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Miklos GL Gabor, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Francesco VD, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji R-R, Ke Z, Ketchum KA, Lai Z, Lei Y,

Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang ZY, Wang A, Wang X, Wang J, Wei M-H, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu SC, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers Y-H, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang Y-H, Coyne M, Dahlke C, Mays AD, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The Sequence of the Human Genome. Science. 2001; 291:1304–1351. [PubMed: 11181995]

3. International HapMap C. A haplotype map of the human genome. Nature. 2005; 437:1299–1320. [PubMed: 16255080]
4. Kolor K, Liu T, St Pierre J, Khoury MJ. Health care provider and consumer awareness, perceptions, and use of direct-to-consumer personal genomic tests, United States, 2008. Genet Med. 2009; 11:595. [PubMed: 19680046]
5. Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes J 3rd. Factors of risk in the development of coronary heart disease--six year follow-up experience. The Framingham Study. Ann Intern Med. 1961; 55:33–50. [PubMed: 13751193]
6. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ. 2007; 335:136. [PubMed: 17615182]
7. Pencina MJ, D'Agostino RB Sr. Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: the framingham heart study. Circulation. 2009; 119:3078–3084. [PubMed: 19506114]
8. Ridker PM, Paynter NP, Rifai N, Gaziano JM, Cook NR. C-reactive protein and parental history improve global cardiovascular risk prediction: the Reynolds Risk Score for men. Circulation. 2008; 118:2243–2251. 2244p following 2251. [PubMed: 18997194]
9. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. 1998; 97:1837–1847. [PubMed: 9603539]
10. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med. 2009; 150:795–802. [PubMed: 19487714]
11. Kannel WB, Vasan RS. Adverse consequences of the 50% misconception. Am J Cardiol. 2009; 103:426–427. [PubMed: 19166702]
12. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L, Investigators IS. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004; 364:937–952. [PubMed: 15364185]
13. Khot UN, Khot MB, Bajzer CT, Sapp SK, Ohman EM, Brener SJ, Ellis SG, Lincoff AM, Topol EJ. Prevalence of conventional risk factors in patients with coronary heart disease. JAMA. 2003; 290:898–904. [PubMed: 12928466]

14. Tzoulaki I, Liberopoulos G, Ioannidis JP. Assessment of claims of improved prediction beyond the Framingham risk score. JAMA. 2009; 302:2345–2352. [PubMed: 19952321]

15. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MS, Go AS, Harrell FE Jr. Hong Y, Howard BV, Howard VJ, Hsue PY, Kramer CM, McConnell JP, Normand SL, O'Donnell CJ, Smith SC Jr. Wilson PW, American Heart Association Expert Panel on Subclinical Atherosclerotic D; Emerging Risk F; the Stroke C. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. Circulation. 2009; 119:2408–2416. [PubMed: 19364974]

16. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. JAMA. 2007; 297:611–619. [PubMed: 17299196]

17. Scheuner MT, Setodji CM, Pankow JS, Blumenthal RS, Keeler E. General Cardiovascular Risk Profile identifies advanced coronary artery calcium and is improved by family history: the multiethnic study of atherosclerosis. Circ Cardiovasc Genet. 3:97–105. [PubMed: 20160201]

18. Lloyd-Jones DM, Nam BH, D'Agostino RB Sr. Levy D, Murabito JM, Wang TJ, Wilson PW, O'Donnell CJ. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA. 2004; 291:2204–2211. [PubMed: 15138242]

19. Murabito JM, Pencina MJ, Nam BH, D'Agostino RB Sr. Wang TJ, Lloyd-Jones D, Wilson PW, O'Donnell CJ. Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. JAMA. 2005; 294:3117–3123. [PubMed: 16380592]

20. Fischer M, Broeckel U, Holmer S, Baessler A, Hengstenberg C, Mayer B, Erdmann J, Klein G, Riegger G, Jacob HJ, Schunkert H. Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction. Circulation. 2005; 111:855–862. [PubMed: 15710764]

21. Nora JJ, Lortscher RH, Spangler RD, Nora AH, Kimberling WJ. Genetic--epidemiologic study of early-onset ischemic heart disease. Circulation. 1980; 61:503–508. [PubMed: 7353240]

22. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007; 316:1488–1491. [PubMed: 17478681]

23. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007; 316:1491–1493. [PubMed: 17478679]

24. Schunkert H, Gotz A, Braund P, McGinnis R, Tregouet D-A, Mangino M, Linsel-Nitschke P, Cambien F, Hengstenberg C, Stark K, Blankenberg S, Tiret L, Ducimetiere P, Keniry A, Ghori MJR, Schreiber S, El Mokhtari NE, Hall AS, Dixon RJ, Goodall AH, Liptau H, Pollard H, Schwarz DF, Hothorn LA, Wichmann H-E, Konig IR, Fischer M, Meisinger C, Ouwehand W, Deloukas P, Thompson JR, Erdmann J, Ziegler A, Samani NJ, for the Cardiogenics Consortium. Repeated Replication and a Prospective Meta-Analysis of the Association Between Chromosome 9p21.3 and Coronary Artery Disease. Circulation. 2008; 117:1675–1684. [PubMed: 18362232]

25. Visel A, Zhu Y, May D, Afzal V, Gong E, Attanasio C, Blow MJ, Cohen JC, Rubin EM, Pennacchio LA. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. Nature. 464:409–412. [PubMed: 20173736]

26. Consortium W. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]

27. Erdmann J, Groszhennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, Linsel-Nitschke P, Kathiresan S, Wright B, Tregouet D-A, Cambien F, Bruse P, Aherrahrou Z, Wagner AK, Stark K, Schwartz SM, Salomaa V, Elosua R, Melander O, Voight BF, O'Donnell CJ, Peltonen L, Siscovick DS, Altshuler D, Merlini PA, Peyvandi F, Bernardinelli L, Ardissino D, Schillert A, Blankenberg S, Zeller T, Wild P, Schwarz DF, Tiret L, Perret C, Schreiber S, Mokhtari NEE,

Schafer A, Marz W, Renner W, Bugert P, Kluter H, Schrezenmeir J, Rubin D, Ball SG, Balmforth AJ, Wichmann HE, Meitinger T, Fischer M, Meisinger C, Baumert J, Peters A, Ouwehand WH, Deloukas P, Thompson JR, Ziegler A, Samani NJ, Schunkert H. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet. 2009; 41:280–282. [PubMed: 19198612]

28. Myocardial Infarction Genetics C; Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Merlini PA, Berzuini C, Bernardinelli L, Peyvandi F, Tubaro M, Celli P, Ferrario M, Fetiveau R, Marziliano N, Casari G, Galli M, Ribichini F, Rossi M, Bernardi F, Zonzin P, Piazza A, Yee J, Friedlander Y, Marrugat J, Lucas G, Subirana I, Sala J, Ramos R, Meigs JB, Williams G, Nathan DM, MacRae CA, Havulinna AS, Berglund G, Hirschhorn JN, Asselta R, Duga S, Spreafico M, Daly MJ, Nemesh J, Korn JM, McCarroll SA, Surti A, Guiducci C, Gianniny L, Mirel D, Parkin M, Burtt N, Gabriel SB, Thompson JR, Braund PS, Wright BJ, Balmforth AJ, Ball SG, Hall AS, Wellcome Trust Case Control C. Linsel-Nitschke P, Lieb W, Ziegler A, Konig I, Hengstenberg C, Fischer M, Stark K, Grosshennig A, Preuss M, Wichmann HE, Schreiber S, Ouwehand W, Deloukas P, Scholz M, Cambien F, Li M, Chen Z, Wilensky R, Matthai W, Qasim A, Hakonarson HH, Devaney J, Burnett MS, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Epstein SE, Scheffold T, Berger K, Huge A, Martinelli N, Olivieri O, Corrocher R, McKeown P, Erdmann E, Konig IR, Holm H, Thorleifsson G, Thorsteinsdottir U, Stefansson K, Do R, Xie C, Siscovick D. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet. 2009; 41:334–341. [PubMed: 19198609]

29. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann H-E, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet D-A, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H, WTCCC t, the Cardiogenics Consortium. Genomewide Association Analysis of Coronary Artery Disease. N Engl J Med. 2007; 357:443–453. [PubMed: 17634449]

30. Tregouet DA, Konig IR, Erdmann J, Munteanu A, Braund PS, Hall AS, Grosshennig A, Linsel-Nitschke P, Perret C, DeSuremain M, Meitinger T, Wright BJ, Preuss M, Balmforth AJ, Ball SG, Meisinger C, Germain C, Evans A, Arveiler D, Luc G, Ruidavets JB, Morrison C, van der Harst P, Schreiber S, Neureuther K, Schafer A, Bugert P, El Mokhtari NE, Schrezenmeir J, Stark K, Rubin D, Wichmann HE, Hengstenberg C, Ouwehand W, Wellcome Trust Case Control C. Cardiogenics C, Ziegler A, Tiret L, Thompson JR, Cambien F, Schunkert H, Samani NJ. Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. Nat Genet. 2009; 41:283–285. [PubMed: 19198611]

31. McPherson R. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007; 316:1488–1491. [PubMed: 17478681]

32. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, Thorleifsson G, Helgadottir H, Steinthorsdottir V, Stefansson H, Williams C, Hui J, Beilby J, Warrington NM, James A, Palmer LJ, Koppelman GH, Heinzmann A, Krueger M, Boezen HM, Wheatley A, Altmuller J, Shin HD, Uh S-T, Cheong HS, Jonsdottir B, Gislason D, Park C-S, Rasmussen LM, Porsbjerg C, Hansen JW, Backer V, Werge T, Janson C, Jonsson U-B, Ng MCY, Chan J, So WY, Ma R, Shah SH, Granger CB, Quyyumi AA, Levey AI, Vaccarino V, Reilly MP, Rader DJ, Williams MJA, van Rij AM, Jones GT, Trabetti E, Malerba G, Pignatti PF, Boner A, Pescollderungg L, Girelli D, Olivieri O, Martinelli N, Ludviksson BR, Ludviksdottir D, Eyjolfsson GI, Arnar D, Thorgeirsson G, Deichmann K, Thompson PJ, Wjst M, Hall IP, Postma DS, Gislason T, Gulcher J, Kong A, Jonsdottir I, Thorsteinsdottir U, Stefansson K. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. Nat Genet. 2009; 41:342–347. [PubMed: 19198610]

33. Adeyemo A, Gerry N, Chen G, Herbert A, Doumatey A, Huang H, Zhou J, Lashley K, Chen Y, Christman M, Rotimi C. A genome-wide association study of hypertension and blood pressure in African Americans. PLoS Genet. 2009; 5:e1000564. [PubMed: 19609347]

34. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T, Aulchenko Y, Lumley T, Kottgen A, Vasan RS, Rivadeneira F, Eiriksdottir G, Guo X, Arking DE, Mitchell GF, Mattace-Raso FUS, Smith AV, Taylor K, Scharpf RB, Hwang S-J, Sijbrands EJG, Bis J, Harris TB, Ganesh SK, O'Donnell CJ, Hofman A, Rotter JI, Coresh J, Benjamin EJ, Uitterlinden AG, Heiss G, Fox CS, Witteman JCM, Boerwinkle E, Wang TJ, Gudnason V, Larson MG, Chakravarti A, Psaty BM, van Duijn CM. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009; 41:677–687. [PubMed: 19430479]

35. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, Najjar SS, Zhao JH, Heath SC, Eyheramendy S, Papadakis K, Voight BF, Scott LJ, Zhang F, Farrall M, Tanaka T, Wallace C, Chambers JC, Khaw K-T, Nilsson P, van der Harst P, Polidoro S, Grobbee DE, Onland-Moret NC, Bots ML, Wain LV, Elliott KS, Teumer A, Luan Ja, Lucas G, Kuusisto J, Burton PR, Hadley D, McArdle WL, Brown M, Dominiczak A, Newhouse SJ, Samani NJ, Webster J, Zeggini E, Beckmann JS, Bergmann S, Lim N, Song K, Vollenweider P, Waeber G, Waterworth DM, Yuan X, Groop L, Orho-Melander M, Allione A, Di Gregorio A, Guarrera S, Panico S, Ricceri F, Romanazzi V, Sacerdote C, Vineis P, Barroso I, Sandhu MS, Luben RN, Crawford GJ, Jousilahti P, Perola M, Boehnke M, Bonnycastle LL, Collins FS, Jackson AU, Mohlke KL, Stringham HM, Valle TT, Willer CJ, Bergman RN, Morken MA, Doring A, Gieger C, Illig T, Meitinger T, Org E, Pfeufer A, Wichmann HE, Kathiresan S, Marrugat J, O'Donnell CJ, Schwartz SM, Siscovick DS, Subirana I, Freimer NB, Hartikainen A-L, McCarthy MI, O'Reilly PF, Peltonen L, Pouta A, de Jong PE, Snieder H, van Gilst WH, Clarke R, Goel A, Hamsten A, Peden JF, Seedorf U, Syvanen A-C, Tognoni G, Lakatta EG, Sanna S, Scheet P, Schlessinger D, Scuteri A, Dorr M, Ernst F, Felix SB, Homuth G, Lorbeer R, Reffelmann T, Rettig R, Volker U, Galan P, Gut IG, Hercberg S, Lathrop GM, Zelenika D, Deloukas P, Soranzo N, Williams FM, Zhai G, Salomaa V, Laakso M, Elosua R, Forouhi NG, Volzke H, Uiterwaal CS, van der Schouw YT, Numans ME, Matullo G, Navis G, Berglund G, Bingham SA, Kooner JS, Connell JM, Bandinelli S, Ferrucci L, Watkins H, Spector TD, Tuomilehto J, Altshuler D, Strachan DP, Laan M, Meneton P, Wareham NJ, Uda M, Jarvelin M-R, Mooser V, Melander O, Loos RJF, Elliott P, Abecasis GR, Caulfield M, Munroe PB. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009; 41:666–676. [PubMed: 19430483]

36. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, Penninx BW, Janssens AC, Wilson JF, Spector T, Martin NG, Pedersen NL, Kyvik KO, Kaprio J, Hofman A, Freimer NB, Jarvelin MR, Gyllensten U, Campbell H, Rudan I, Johansson A, Marroni F, Hayward C, Vitart V, Jonasson I, Pattaro C, Wright A, Hastie N, Pichler I, Hicks AA, Falchi M, Willemsen G, Hottenga JJ, de Geus EJ, Montgomery GW, Whitfield J, Magnusson P, Saharinen J, Perola M, Silander K, Isaacs A, Sijbrands EJ, Uitterlinden AG, Witteman JC, Oostra BA, Elliott P, Ruokonen A, Sabatti C, Gieger C, Meitinger T, Kronenberg F, Doring A, Wichmann HE, Smit JH, McCarthy MI, van Duijn CM, Peltonen L, Consortium E. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet. 2009; 41:47–55. [PubMed: 19060911]

37. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU, Crawford G, Surti A, Guiducci C, Burtt NP, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ, Kuusisto J, Bergman RN, Sundvall J, Laakso M, Ferrucci L, Scheet P, Sanna S, Uda M, Yang Q, Lunetta KL, Dupuis J, de Bakker PI, O'Donnell CJ, Chambers JC, Kooner JS, Hercberg S, Meneton P, Lakatta EG, Scuteri A, Schlessinger D, Tuomilehto J, Collins FS, Groop L, Altshuler D, Collins R, Lathrop GM, Melander O, Salomaa V, Peltonen L, Orho-Melander M, Ordovas JM, Boehnke M, Abecasis GR, Mohlke KL, Cupples LA. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet. 2009; 41:56–65. [PubMed: 19060906]

38. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varilo T, Kaakinen M, Sovio U, Ruokonen A, Laitinen J, Jakkula E, Coin L, Hoggart C, Collins A, Turunen H, Gabriel S, Elliot P, McCarthy MI, Daly MJ, Jarvelin MR, Freimer NB, Peltonen L. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet. 2009; 41:35–46. [PubMed: 19060910]

39. Hindorff, LA.; Junkins, HA.; Mehta, JP.; Manolio, TA. A Catalog of Published Genome-Wide Association Studies. [Accessed March 26]. 2010

40. Morgan TM, Krumholz HM, Lifton RP, Spertus JA. Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. JAMA. 2007; 297:1551–1561. [PubMed: 17426274]
41. Ntzani EE, Rizos EC, Ioannidis JP. Genetic effects versus bias for candidate polymorphisms in myocardial infarction: case study and overview of large-scale evidence. Am J Epidemiol. 2007; 165:973–984. [PubMed: 17293603]
42. Goldstein DB. Common genetic variation and human traits. N Engl J Med. 2009; 360:1696–1698. [PubMed: 19369660]
43. 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. 303:631–637. [PubMed: 20159871]
44. Paynter NP, Chasman DI, Buring JE, Shiffman D, Cook NR, Ridker PM. Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. Ann Intern Med. 2009; 150:65–72. [PubMed: 19153409]
45. Brautbar A, Ballantyne CM, Lawson K, Nambi V, Chambless L, Folsom AR, Willerson JT, Boerwinkle E. Impact of adding a single allele in the 9p21 locus to traditional risk factors on reclassification of coronary heart disease risk and implications for lipid-modifying therapy in the Atherosclerosis Risk in Communities study. Circ Cardiovasc Genet. 2009; 2:279–285. [PubMed: 20031596]
46. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt 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. [PubMed: 18354102]
47. Talmud PJ, Cooper JA, Palmen J, Lovering R, Drenos F, Hingorani AD, Humphries SE. Chromosome 9p21.3 coronary heart disease locus genotype and prospective risk of CHD in healthy middle-aged men. Clin Chem. 2008; 54:467–474. [PubMed: 18250146]
48. Morrison AC, Bare LA, Chambless LE, Ellis SG, Malloy M, Kane JP, Pankow JS, Devlin JJ, Willerson JT, Boerwinkle E. Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. Am J Epidemiol. 2007; 166:28–35. [PubMed: 17443022]
49. Humphries SE, Cooper JA, Talmud PJ, Miller GJ. Candidate gene genotypes, along with conventional risk factor assessment, improve estimation of coronary heart disease risk in healthy UK men. Clin Chem. 2007; 53:8–16. [PubMed: 17130180]
50. Junyent M, Tucker KL, Shen J, Lee YC, Smith CE, Mattei J, Lai CQ, Parnell LD, Ordovas JM. A composite scoring of genotypes discriminates coronary heart disease risk beyond conventional risk factors in the Boston Puerto Rican Health Study. Nutrition, metabolism, and cardiovascular diseases : NMCD. 20:157–164.
51. Trichopoulou A, Yiannakouris N, Bamia C, Benetou V, Trichopoulos D, Ordovas JM. Genetic predisposition, nongenetic risk factors, and coronary infarct. Arch Intern Med. 2008; 168:891–896. [PubMed: 18443266]
52. Yamada Y, Izawa H, Ichihara S, Takatsu F, Ishihara H, Hirayama H, Sone T, Tanaka M, Yokota M. Prediction of the risk of myocardial infarction from polymorphisms in candidate genes. N Engl J Med. 2002; 347:1916–1923. [PubMed: 12477941]
53. Ioannidis JP. Prediction of cardiovascular disease outcomes and established cardiovascular risk factors by genome-wide association markers. Circ Cardiovasc Genet. 2009; 2:7–15. [PubMed: 20031562]
54. Amin N, van Duijn CM, Janssens AC. Genetic scoring analysis: a way forward in genome wide association studies? Eur J Epidemiol. 2009; 24:585–587. [PubMed: 19728114]
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. [PubMed: 15105181]
56. Ioannidis JP. Why most discovered true associations are inflated. Epidemiology. 2008; 19:640–648. [PubMed: 18633328]

57. Zollner S, Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case-control data. Am J Hum Genet. 2007; 80:605–615. [PubMed: 17357068]
58. Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. PLoS One. 2007; 2:e841. [PubMed: 17786212]
59. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: methodological considerations for future research. Genome Med. 2009; 1:20. [PubMed: 19341491]
60. Lubin JH, Gail MH. Biased selection of controls for case-control analyses of cohort studies. Biometrics. 1984; 40:63–75. [PubMed: 6375751]
61. Wang MH, Shugart YY, Cole SR, Platz EA. A simulation study of control sampling methods for nested case-control studies of genetic and molecular biomarkers and prostate cancer progression. Cancer Epidemiol Biomarkers Prev. 2009; 18:706–711. [PubMed: 19258478]
62. Goldstein P, Lapostolle F, Steg G, Danchin N, Assez N, Montalescot G, Charpentier S, Wiel E, Juliard JM. Lowering mortality in ST-elevation myocardial infarction and non-ST-elevation myocardial infarction: key prehospital and emergency room treatment strategies. Eur J Emerg Med. 2009; 16:244–255. [PubMed: 19318957]
63. Norris RM. Fatality outside hospital from acute coronary events in three British health districts, 1994-5. United Kingdom Heart Attack Study Collaborative Group. BMJ. 1998; 316:1065–1070. [PubMed: 9552910]
64. Lloyd-Jones D. Heart disease and stroke statistics[mdash]2009 update. A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation. 2009; 119:e21–e181. [PubMed: 19075105]
65. Manolio TA. Cohort studies and the genetics of complex disease. Nat Genet. 2009; 41:5–6. [PubMed: 19112455]
66. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007; 115:928–935. [PubMed: 17309939]
67. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM. Predictive testing for complex diseases using multiple genes: fact or fiction? Genet Med. 2006; 8:395–400. [PubMed: 16845271]
68. van derNet JB, Janssens AC, Sijbrands EJ, Steyerberg EW. Value of genetic profiling for the prediction of coronary heart disease. Am Heart J. 2009; 158:105–110. [PubMed: 19540399]
69. Wray NR, Yang J, Goddard ME, Visscher PM. The genetic interpretation of area under the ROC curve in genomic profiling. PLoS Genet. 6:e1000864. [PubMed: 20195508]
70. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res. 2007; 17:1520–1528. [PubMed: 17785532]
71. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff 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. [PubMed: 19812666]
72. Chakravarti A. Population genetics--making sense out of sequence. Nat Genet. 1999; 21:56–60. [PubMed: 9915503]
73. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk of complex disease. Curr Opin Genet Dev. 2008; 18:257–263. [PubMed: 18682292]
74. Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J, Elliott P. Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. Int J Epidemiol. 2009; 38:263–273. [PubMed: 18676414]
75. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: advances and prospects. Hum Mol Genet. 2008; 17:R166–173. [PubMed: 18852206]
76. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. Hum Mol Genet. 2009; 18:3525–3531. [PubMed: 19553258]
77. International Schizophrenia C. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460:748–752. [PubMed: 19571811]

78. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet. 2008; 9:255–266. [PubMed: 18319743]
79. Visscher PM, Medland SE, Ferreira MA, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet. 2006; 2:e41. [PubMed: 16565746]
80. Khoury MJ, Wacholder S. Invited commentary: from genome-wide association studies to gene-environment-wide interaction studies--challenges and opportunities. Am J Epidemiol. 2009; 169:227–230. [PubMed: 19022826]
81. Ordovas JM, Tai ES. Why study gene-environment interactions? Curr Opin Lipidol. 2008; 19:158–167. [PubMed: 18388696]
82. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet. 2009; 10:392–404. [PubMed: 19434077]
83. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 11:31–46. [PubMed: 19997069]
84. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitch J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet. 2008; 40:1166–1174. [PubMed: 18776908]
85. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, Saeed S, Hamilton-Shield J, Clayton-Smith J, O'Rahilly S, Hurles ME, Farooqi IS. Large, rare chromosomal deletions associated with severe early-onset obesity. Nature. 2009; 463:666–70. [PubMed: 19966786]
86. Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nat Genet. 2007; 39:513–516. [PubMed: 17322881]
87. Collins FS, Barker AD. Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. Sci Am. 2007; 296:50–57. [PubMed: 17348159]
88. 1000 genomes. [Accessed Jan 15th]. 2010 Available at: http://www.1000genomes.org
89. Tracy RP. 'Deep phenotyping': characterizing populations in the era of genomics and systems biology. Curr Opin Lipidol. 2008; 19:151–157. [PubMed: 18388695]
90. Ioannidis JPA. Commentary: Grading the credibility of molecular evidence for complex diseases. Int. J. Epidemiol. 2006; 35:572–578. [PubMed: 16540537]
91. Yang J, Visscher PM, Wray NR. Sporadic cases are the norm for complex disease. Eur J Hum Genet. 2009; 17:1–4. [PubMed: 18941473]
92. Ioannidis JP. Population-wide generalizability of genome-wide discovered associations. J Natl Cancer Inst. 2009; 101:1297–1299. [PubMed: 19726754]
93. Greenland S. The need for reorientation toward cost-effective prediction: comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine. Stat Med. 2008; 27:199–206. [PubMed: 17729377]
94. Kraft P, Hunter DJ. Genetic risk prediction--are we there yet? N Engl J Med. 2009; 360:1701–1703. [PubMed: 19369656]
95. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet. 2002; 31:33–36. [PubMed: 11984562]
96. Rogowski WH, Grosse SD, Khoury MJ. Challenges of translating genetic tests into clinical and public health practice. Nat Rev Genet. 2009; 10:489–495. [PubMed: 19506575]
97. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical Trial Designs for Predictive Marker Validation in Cancer Treatment Trials. J Clin Oncol. 2005; 23:2020–2027. [PubMed: 15774793]
98. Marín F, González-Conejero R, Capranzano P, Bass TA, Roldán V, Angiolillo DJ. Pharmacogenetics in Cardiovascular Antithrombotic Therapy. Journal of the American College of Cardiology. 2009; 54:1041–1057. [PubMed: 19744613]

Table 1

Loci associated with myocardial infarction from GWAS

Locus	Nearest Gene	Risk Allele Frequency	P-value	Relative risk for MI	References
3q22.3	MRAS	0.15	7×10^{-13}	1.15 [1.11-1.19]	27
12q24.31	HNF1A	0.36	5×10^{-7}	1.08 [1.05-1.11]	27
9p21.3	*CDKN2A, CDKN2B*	0.56	3×10^{-44}	1.29 [1.25-1.34]	23,26,28,29,31
1p13.3	*CELSR2, PSRC1, SORT1*	0.81	8×10^{-12}	1.19 [1.13-1.26]	28,29
21q22.11	*SLC5A3, MRPS6, KCNE2*	0.13	6×10^{-11}	1.20 [1.14-1.27]	28
1q41	*MIA3*	0.72	1×10^{-9}	1.14 [1.10-1.19]	28,29
6p24.1	*PHACTR1*	0.65	1×10^{-9}	1.12 [1.08-1.17]	28
19p13.2	*LDLR*	0.75	2×10^{-9}	1.15 [1.10-1.20]	28
10q11.21	*CXCL12*x	0.84	7×10^{-9}	1.17 [1.11-1.24]	28,29
1p32.3	*PCSK9*	0.81	1×10^{-8}	1.15 [1.10-1.21]	28
2q33.1	*WDR12*	0.14	1×10^{-8}	1.17 [1.11-1.23]	28
6q25.3	*SLC22A3,LPAL2,LPA*	0.02	4×10^{-15}	1.82 [1.57-2.12]	30
6q25.3	*SLC22A3,LPAL2,LPA*	0.16	1×10^{-9}	1.20 [1.13-1.28]	30
12q24.12	*SH2B3*	0.38	9×10^{-8}	1.13 [1.08, 1.18]	32

Loci were selected from the GWAS catalog based on a search for the following phenotypes: coronary disease, coronary artery disease, major CVD, myocardial infarction, myocardial infarction (early onset). We also included the 12q24.12 locus which has been associated with eosinophil levels, asthma and myocardial infarction. Reported associations were limited to those with $p<5\times10^{-8}$ and replication in at least one independent cohort.

Table 2

Studies Using Genetic Information in CVD Risk Prediction

Author	Markers Used	Source of Genetic Markers	Outcome	Covariates		Metric to Assess Incremental Utility	Evidence for Clinical Utility	Comments
				Conventional Risk Factors	Family History			
Paynter et al.[43]	Multiple	GWAS	Incident CVD	Yes	Yes	AUC, NRI	No	Women only
Paynter et al. [44]	single (9p21)	GWAS	incident CVD	Yes	Yes	AUC, NRI	No	Women only
Brautbar et al.[45]	Single (9p21)	GWAS	Incident CVD	Yes	No	AUC, NRI	Minimal	
Kathiresan et al[28]	Multiple	GWAS	Incident CVD	Yes	Yes	AUC, NRI	Minimal	Lipid SNPs only
Kathiresan et al[46]	Multiple	GWAS	Prevalent CVD	No	No	none	No	-
Talmud et al. [47]	single (9p21)	GWAS	incident CVD	Yes	No	AUC	Minimal	Diabetes not in model
Morrison et al. [48]	multiple	candidate genes	incident CVD	Yes	No	AUC	Minimal (in blacks only)	-
Humphries et al. [49]	multiple	candidate genes	incident CVD	Yes	No	AUC	Minimal	GxE interactions considered; diabetes not in model
Junyent et al.[50]	mutiple	candidate genes and GWAS	prevalent CHD	Yes	No	OR	No	Gene score associated with CHD case status
Trichopoulos et al. [51]	multiple	candidate genes	Incident CHD	Yes	No	OR	No	Gene score associated with CHD case status; GxE interactions considered
Yamada et al. [52]	mutliple	candidate genes	prevalent CVD	Yes	No	OR	No	Gene score associated with CHD case status

GWAS – genome wide association study, CVD – cardiovascular disease, CHD – Coronary heart disease, AUC – area under the curve, NRI – net reclassification index, OR – odds ratio, GxE – gene environment interactions. Studies were identified using a PubMed search using the following key words: cardiovascular disease, coronary disease, genetics, risk prediction. We also reviewed the references of relevant articles to find additional articles of interest. From these studies, we included only studies which evaluated the utility (or made positive or negative claims regarding utility) of genetic information in cardiovascular risk prediction.

Table 3

Summary of Challenges Facing Genetic CVD Risk Prediction, Their Implications and Potential Solutions

Challenges for risk prediction	Possible Issues and/or Implications	Potential solutions
General considerations for CVD prediction		
	Conventional risk factors explain a large proportion of the risk for CVD	Genetic risk must be incremental to standard factors and family history
	Family history information is predictive, easily obtained and free	
	Determining predictive performance of genetic information	Use of a combination of c-statistic and reclassification measures
Biases in genetic effect sizes from GWAS		
	Use of extreme case and extreme controls	GWAS for incident CVD in population-based cohorts
	Incidence-Prevalence bias	
	Survivor bias	
Allelic architecture of CVD		
	Small to very small effect sizes	Larger sample sizes
	Hundreds to thousands genes may underlie CVD risk	
Missing heritability		
	Inaccurate estimates of heritability	Heritability by identity-by-descent methods
	Gene-gene and gene-environment studies	Case-only and family-based studies
	Poorly penetrant SNPs	Larger sample sizes
	Identifying causal variants	Sequencing, studies in population with narrow LD (e.g. African-Americans)
	Structural variants (i.e. CNVs)	
	Rare variants	Exome and whole genome sequencing, studies in populations with narrow LD, Family-based studies, Founder populations
	Imprecise phenotypes	Deep phenotyping using -omics methods
Large number of genes explain genetic risk		
	Unique genetic signature for each individual	Larger sample sizes
	High genetic risk will be rare	
Translation of genetic risk prediction to clinical practice		
	External validation	Cohort studies in appropriate populations
	Generalizability across ethnicities	Cohort studies in diverse ethnicities, re-calibration
	Optimize false postive and false negative rates using appropriate cut-offs	
	Assessment of predictive values and likelihood ratios in populations with differing baseline risks	Evaluation of prediction in individuals of varying baseline risk.
	Efficacy and effectiveness (i.e. need for screening RCTs)	Randomized screening trials
	Cost-effectiveness	Cost-effectivenes studies

Challenges for risk prediction	Possible Issues and/or Implications	Potential solutions
	Clinical utility over other -omic approaches	Evaluation of genomic predictors vs other -omics predictors