How close are we to implementing a genetic risk score for coronary heart disease?

Katherine Beaney¹*, Fotios Drenos¹,², and Steve E Humphries¹

1 Centre for Cardiovascular Genetics, BHF Laboratories, Institute of Cardiovascular Science, University College London, London, UK

2 MRC Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, UK

*Current address: MRC Clinical Trials Unit at UCL, Institute of Clinical Trials & Methodology, Aviation House, 125 Kingsway, London WC2B 6NH

Corresponding author: Professor Steve E Humphries, Centre for Cardiovascular Genetics, British Heart Foundation Laboratories, Institute of Cardiovascular Science, The Rayne Building University College London, London, UK WC1E 6JF, Tel 0207 679 6962 (internal 46962), email: steve.humphries@ucl.ac.uk

Key Words: Genome wide association studies, Genetic Risk Score, Classical risk factors,
Abstract

Genome-wide association meta-analysis have now identified more than 150 loci where common variants (SNPs) are significantly associated with cardiovascular disease (CVD) and CVD end points. These can be used in combination, and “scaled” by their effect size, to create a “weighted” Genetic risk Score (GRS), which, in combination with an individual’s classical CVD risk factors, can be used to identify those at overall low, intermediate and high future risk. Those at highest risk can be offered life-style and therapeutic options to reduce their risk and those at intermediate levels can be monitored. Here we discuss the selection of the best variants to be included in the GRS, and the potential utility of such scores in different clinical settings. 115 words

Risk prediction scores using classical risk factors (CRFs)

A large proportion of CHD events are preventable, therefore, predicting those at highest risk of developing the disease is an important public health consideration. To take advantage of the combined knowledge of how CRFs affect CHD risk, risk scores have been developed. The first risk score for CHD that gained widespread use was developed from the Framingham Heart Study and thus is referred to as the Framingham score [1] Included in it were age, total cholesterol, HDL-cholesterol, systolic blood pressure, diabetes and smoking (with separate equations for men and women). The score showed good predictively ability in some cohorts similar to that from which it was derived [2], but it was found to overestimate risk in other ethnic groups [3] and in other populations of European ethnicity where there was a lower incidence of CHD [4, 5].

In response to this, region-specific scores have been developed such as SCORE which was derived using data from 12 prospective European cohorts [6]. The development of large primary care electronic records has enabled risk scores to be derived from large population cohorts. In England the QRISK score was derived from the QRESEARCH database, (which contains 1.2 million individuals) to estimate risk of CVD (rather than CHD) [7]. This score was updated (QRISK2) to include a number of other risk factors, most notably self-reported ethnicity [8]. QRISK2 also includes measures of social deprivation. This was prompted by the observation that the
Framingham score underestimated risk in socially deprived individuals and thus could re-enforce social gradients in disease [9]. The QRISK2 model is updated annually (http://www.qrisk.org/).

**Primary prevention strategies and Current clinical guidance**

Many of the CRFs for CHD are modifiable and thus lifestyle interventions such as use of smoking cessation services and dietary review, form an important part of the strategy to reduce CHD risk. Prescription of lipid-lowering therapies, primarily statins, has also been used to compliment this. Statin use has been found to reduce risk of CVD events by approximately one fifth per 1mmol/l of LDL-cholesterol reduction, in a wide range of individuals [10]. A benefit has also been found in those with low CVD risk.

When risk scores were first introduced into clinical practice, the Framingham risk score was recommended for use in both the USA and the UK, with the high risk group being defined as those having a ten-year risk of CHD ≥20% [11]. Those who fell into that category were then recommended for intensive lifestyle changes and prescription of lipid-lowering medications (usually statins). However, the joint guidelines issued by the American College of Cardiology (ACC) and the American Heart Association (AHA) developed new risk equations and lowered the high-risk cut-off to ≥7.5% [12]. Similarly, in the UK the National Institute of Health and Clinical Excellence (NICE) updated their guidelines to recommend use of QRISK2 and lowered the high-risk threshold to ≥10% [13]. However, given the shortcomings in the available data for statin use there have been concerns particularly regarding the “medicalisation of healthy individuals” and the numbers of adverse events observed in certain groups [14]. There is also evidence that uptake of statins in the those classified in the 10-20 % risk group is much lower than estimated by NICE [15] although larger studies are required to confirm this.

However, the majority of cases of CHD/CVD come from individuals classified with average risk using the CRF risk scores – the so-called prevention paradox [16]. For example, when QRISK2 (2010 version) was validated with data from the health improvement network (THIN), (using a 20% high-risk cut-off), 14% of men and 6% of women were identified as being at high risk. This captured 40% of the cardiovascular events in men and 26% of the cardiovascular events in women.
This leaves scope for refinement of the risk score to discriminate better between those who do and do not go on to develop CVD.

**Use of genetics in risk prediction**

Data from twin studies has estimated the heritability of CHD mortality to be 40-60% [18,19] suggesting that inherited factors are likely to be making a strong contribution to an individual’s future risk of premature CVD. The identification of robustly associated CHD risk loci would be potentially of interest in risk prediction, since as being fixed at conception, they have a lifelong impact and need only be determined once. The first question is thus what genes and what common single nucleotide polymorphisms (SNPs) should be included in the genetic risk profile? Before 2007, when technology improvements enabled genome-wide association studies (GWAS) to address this question in a “hypothesis-free” way, studies in the field used knowledge of the pathophysiology of CVD to identify potential “candidate” genes. Combining published small studies using meta-analysis [20] lead to the identification of a number of potential candidates particularly in the areas of lipid metabolism (eg APOE, APOC3 and LPL). The post-2007 era of GWAS identified many previously unknown genomic loci with statistically significant associations with CVD, most strikingly with the simultaneous publication in 2007 of three GWAS studies that identified a single locus on chromosome 9p.21 [21-23]. This effect has been replicated in countless other studies and is an independent CHD risk factor in different ethnic groups [24].

To exploit the potential of GWASs to identify risk loci for CHD, the CARDIoGRAMplusC4D consortium was set up [25]. A meta-analysis fine-mapping GWAS results for CHD including over 60,000 cases and 130,000 controls was published identifying more than 50 SNPs from 46 loci that were robustly associated with CHD. Reassuringly, many of the original candidate genes have been confirmed as GWAS “hits” for CVD, including APOE, APOC3 and LPL. This has been recently expanded and a GWAS meta-analysis investigating over 9 million SNPs using haplotype data from the 1000 genomes project in approximately 185,000 individuals was recently published [26] Ten new CHD risk loci were identified (eight from an additive model and two from a recessive model). Overall, combining published data using meta-analysis, has now led to more than 80 loci that are
robustly associated with CHD [27], and for example 30 associated with CHD traits such as coronary calcification. (http://www.ebi.ac.uk/gwas/).

As has been pointed out previously [28], because of the polygenic nature of CVD, a common SNP is usually associated with only a modest risk effect size, while a SNP with a much larger risk effect is usually much less common. Table 1 shows a list of 25 GWAS CVD hits with the loci ranked by the sum of their effect size (given as an odds ratio (OR)) and their risk allele frequency (RAF) to give an overall estimate of their potential clinical utility. While the ORs range from 1.23 to 1.04 per risk allele carried, the RAFs for all of these SNPs are high, and overall there is a fourfold difference in score from the top to the lowest on this list. Six of these loci are known to be involved in lipid metabolism, as would be expected from our knowledge of the importance of dyslipidaemia in the development of CVD.

It is unlikely that there are any additional common SNPs to be found by GWAS meta-analysis that will outrank these 25, since SNPs with larger effects that are common would have been detected by the size of the datasets currently available. This means that any CVD SNP panel should therefore include all or at least have a good representation of these 25 SNPs or their proxies.

**Construction of a “Genetic Risk Score”**

Given the relatively small effect sizes associated with these risk loci, it is unsurprising that the addition of one variant into a CRF risk score does not result in improved predictive ability [29-31]. This has led to the development of so-called “genetic risk scores” (GRSs) where SNPs at independent loci are combined. A GRS can be unweighted, where simply the number of risk alleles carried by an individual at each locus is summed. This model assumes that the risk effect associated with each SNP is equal (and additive) and this clearly is not the case. A more accurate GRS can be constructed if carriage of the individual SNPs are weighted using the published effect size [32]. An individual’s GRS can be adjusted for the population GRS (based on the RAFs present in the population) and then be combined with a CRF score such as the Framingham score or QRISK2 [8] to give an overall CHD risk estimate.
Using a SNP Genetic Risk Score in men in the UK

To illustrate the development and use of such a GRS we present data from a UK prospective cohort of ~2700 middle-aged men (the Second Northwick Park Hospital study -NPHSII), whose baseline characteristics are presented in Table 2. As expected, the men who went on to develop CHD were older, had higher BMI, higher systolic blood pressure, higher total cholesterol, LDL cholesterol, and a higher proportion were smokers and had diabetes, at baseline. As expected, those who subsequently developed CHD had a higher ten-year CHD risk as calculated using the Framingham risk score and those who subsequently developed CVD had a higher ten-year CVD risk as calculated using the QRISK2 score.

In 2007 we started to develop a multi-SNP panel using 12 SNPs in candidate genes that, and showed that, when used in combination with the Framingham CRF algorithm they had the potential to identify individuals at high future risk of CHD [33, 34]. We next showed that, in this same cohort, the addition of a single SNP from the chromosome 9p21 GWAS CHD locus, improved the CRF area under the receiver operator characteristic curve (AROC) by 3% but that this effect was not statistically significant [29]. By mathematical modelling, we estimated that an additional three SNPs with similar risk size and RAF would be needed to have a significant improvement [29]. We therefore developed a 19 SNP GRS shown in Table 3 comprised of the 9p21 SNP plus six other GWAS loci identified at that time, supplemented by 12 common SNPs in candidate genes where published meta-analyses, mainly of case-control studies, had demonstrated robust albeit modest risk effects. The RAFs for these SNPs varied between 0.01 to 0.9 and the published Odds Ratios from 1.1-1.7. The mean of the GRS was significantly higher (p=0.01) in those who went on to develop CHD over 13.5 years of follow-up, but there was only a modest and non-significant improvement in discrimination (increase in AROC curve p=0.48) and risk classification (net reclassification improvement (NRI), p=0.28) compared to the Framingham algorithm alone [35].

Similar results were seen if only a 13-SNP GRS including SNPs in loci identified by the CARDIoGRAMplusC4D consortium was used. However, the key aspect of clinical utility is in the subjects at intermediate risk, since those with a Framingham 10 year CHD risk over 20% qualified for statin therapy (under the NICE guidelines at the time), while those at low risk (>10%) did not require any intervention. If the addition of the GRS can better discriminate those at intermediate
risk who subsequently go on to develop a CHD event then these subjects can be offered
erapeutic advice and can reduce their subsequent risk. When considering only individuals who
moved up a risk category with inclusion of the GRS, the improvement in risk classification was
statistically significant (p=0.01) [35]. Overall these data suggest a modest clinical utility of this 19
SNP GRS, with its main utility being confined to those at intermediate risk.

The GSs were originally weighted using the effect sizes determined in meta-analyses of candidate
gene studies or GWASs. However in the CARDIoGRAMplusC4D meta-analyses the ORs for several
of the included SNPs had reduced in part because of the so-called “winners curse” [36], and in
some cases had become non-significant, and the use of these more robust risk estimates should
improve the clinical utility of the gene score. Also NICE now recommended the use of the QRISK
algorithm [13] and we found that QRISK2 was better at predicating cardiovascular outcome in
NPHSII compared to the Framingham score, with the Framingham score overestimating risk and
showing poor calibration in the NPHSII men [37]. This is consistent with the literature where even
the NICE-adjusted Framingham risk equations have been found to overestimate ten-year CHD risk
in the UK population, particularly in men [17]. The superior performance of QRISK2 compared to
the Framingham score is unsurprising given that QRISK2 was derived from a very large British
cohort while Framingham was developed from the Framingham study based in Massachusetts,
USA [1]. As shown in Figure 1A in the NPHSII men 49% were above the 10% 10 year CVD cut-off
and the event rate in this group was 15.7%. Since 71% of all events occurring in the next 10 years
were in this high risk group, treatment of all of these subjects is likely to be cost effective. By
contrast in the 7% of men below 5% 10 year risk there were no CVD events during follow-up. The
intermediate group, which constitutes those where clinical judgement is required, represented
44% of men with an event rate of 7.3%, and it is in this group where better risk stratification would
be helpful.

Updating the weightings to those determined in the CARDIoGRAMplusC4D analysis [25] improved
the performance of the 19 SNP GRS. The weighted 19 SNP GRS was associated with CHD after
adjustment for CRFs, (OR=1.31 per standard deviation, p=0.03). Not surprisingly since it includes
seven loci involved in lipid metabolism, the GRS was significantly associated with higher total- and
LDL-cholesterol and modestly with a reported family history of early CHD (p=0.03). As shown in
Figure 1B, 53% of men were now in the high risk group, with 80% of all events being correctly identified. Although 1 man in the intermediate risk group incorrectly was assigned to a lower risk, 116 men (24% of group) moved up from the intermediate to the high risk category, of whom 13 (11.2%) experienced an event. Overall, addition of the 19 SNP GRS to QRISK2 showed improved discrimination (AROC 0.68 v 0.70 p=0.02), and a positive net reclassification index (7%, p=0.04) compared to QRISK2 alone.

Finally, one aim of our work is to identify a minimum SNP data set that will have clinical utility in CHD risk stratification, and thus replacing those SNPs which have non-significant CHD risk effects with SNPs with larger risk effects should improve clinical utility. To assess if the gene score could be improved, we removed five SNPs which had shown little evidence of an association with CHD in the CARDIoGRAMplusC4D meta-analysis. We then selected the top seven ranked SNPs from those robustly associated with CHD in CARDIoGRAMplusC4D shown in Table 1, discounting those in loci already included in the GRS. These were then genotyped in NPHSII and added into the GRS to create a 21 SNP GRS. This GRS was strongly associated with CHD even after adjustment for CRFs (OR=1.39 per standard deviation, 1.42x10^{-3}), but the combined QRISK2 plus GRS score was poorly calibrated (p=0.03) and showed no improvement in discrimination (p=0.55) or reclassification (p=0.10) compared to QRISK2 alone [37]. While it is possible that the addition of SNPs who rank below these seven may improve clinical utility, the data suggest this improvement is at best likely to be modest. However, our results indicate that including the updated weighted 19 SNP GRS along with QRISK2 may have clinical utility in the UK population.

**Use of a SNP Genetic Risk Score in other studies.**

A number of CHD SNP GRSs have been assessed by others and are shown in Table 4. In the UCLEB consortium of over 25,000 participants and ~2200 events for incident and prevalent CHD, a combined QRISK2 plus CARDIoGRAMplusC4D GRS score was used, although it should be noted that QRISK2 was poorly calibrated in this cohort [38]. Although overall the combined score did not show additional benefit over and above QRISK2 alone, the GRS appeared to carry some utility when applied only to those who, according to conventional risk scoring, would have been classified at
intermediate risk, by moving some individuals who subsequently had an event into the high-risk category. The authors estimated that among 100,000 people from a population represented by the UCLEB studies, 29,445 would be classed as of intermediate risk according to the QRISK equation, and of these, 5,434 would then be reclassified as high risk once the GRS was applied, and 1,082 would suffer a CVD event if untreated with statins. Treatment with statins according to guidelines could postpone the CVD event in 20% of these, leaving a number needed to screen of 462 to postpone one CVD event. Recent evidence suggests the risk reductions from statin therapy might be enhanced for those at highest risk \(^1\), so this figure may be conservative.

The results of including a GRS constructed using the results of the CARDIoGRAMplusC4D meta-analysis in CHD risk prediction have been somewhat disappointing. In the prospective Rotterdam study [39] found very limited benefit in the population-wide inclusion of the GRS in risk prediction. However, when 103 “suggestively” associated SNPs were included (giving a GRS of 152 SNPs) this second score improved the AUC compared to a model including traditional CRFs (although this was only statistically significant for prevalent CHD), indicating further gene discovery may therefore produce greater improvements. Improvements in both discrimination and reclassification were observed in a meta-analysis of six Swedish prospective cohorts [40] and in the Malmo Diet and Cancer (MDC) study [41]. Only the reclassification analysis in the MDC study was performed using the most recent guidelines however this was based on the US guidelines from the ACC/AHA [12] rather than the 10 % high risk cut-off recommended in the most recent NICE guidelines in the UK [13]. It has been suggested that due to the nature of case selection in GWASs, many of the variants identified in the CARDIoGRAMplusC4D meta-analysis are actually associated with CHD survival rather than an incident CHD event itself. This is supported by data from both the UCLEB consortium and the Rotterdam study where the gene score was more strongly associated with prevalent rather than incident disease [38,39]. This indicates that the weightings used may not accurately reflect the impact of each variant on incident CHD risk and thus effect sizes obtained from a prospective cohort should be used. This strategy was used by Ganna, Magnusson et al. and a better performance was observed with the inclusion of the GRS [40].
Clinical utility of CHD SNP Gene Scores - Statin Adherence

One useful consequence of informing at-risk subjects of their genetic risk would be if such information motivates lifestyle changes or adherence to proscribed medication to a greater extent than CRF information alone. A clinical trial from the US has examined this [42]. 203 subjects (mean age 59 years), at intermediate risk for CHD, and not on statins were randomly assigned to receive their 10-year probability of CHD based either on a Framingham CRF score or CRFs+GRS. The GRS included 28 CHD GWAS hit SNPs (12 of which are in the top 25 shown in Table 1), with each weighted by its published OR effect size. Subjects were told their risk as high or average or low by a genetic counsellor, with a discussion about starting statin therapy with a physician. After 6 months follow-up the +GRS group had 9% lower LDL-C than the CRF group (p=0.04), with particular benefit seen in those given a high overall risk. This was because more subjects in the high risk group started statin therapy than in the CRF only group (39% vs 22%, p<0.01), and was not due to differences between the groups in dietary fat intake or levels of physical activity. These data support the view that in a clinical setting, a high genetic risk may be particularly motivating to start (and possible adhere to) lipid lowering medication.

Clinical utility of CHD SNP Gene Scores - Genetic vs lifestyle

The GRS can be also used as a genetic instrument to compare the relative contribution of genetic and lifestyle factors in determining CHD risk. Using a GRS combining 50 CHD risk variants, Khera et al [43] compared the genetically-determined CHD risk and lifestyle risk in three prospective cohorts in the USA and Sweden (total ~56,000 subjects with 5,103 CHD events). A healthy lifestyle was defined as a weighted composite of no current smoking, no obesity, regular physical activity, and a healthy diet. Results were consistent across the studies and, compared to those in the bottom quintile of the GRS, the relative risk of incident CHD in the top quintile was 91% higher among those in the top quintile. There was a modest but statistically significant association between GRS and a family history of premature CHD (low = 15.1%, intermediate = 16.9% and high 18.7%, p>0.0001). Unsurprisingly, an unhealthy lifestyle (at least three of the four factors)
resulted in a substantially higher (~80%) risk of CHD and the effect of the genetic and lifestyle risk was essentially additive. Among those at high genetic risk, a favourable lifestyle was associated with a 46% lower CHD relative risk than an unfavourable lifestyle. In a sub-study of 4,260 subjects, both genetic and lifestyle factors were significantly associated with coronary artery calcification evaluated with computed tomography. One study (ARIC) included enough African-American participants (2,269 and 350 CHD events) to show that effects were consistent across ancestral groups and were of similar magnitude.

Five year view

There are several ways the GRS approach will change and develop over the next five years. Firstly, there are likely to be improvements in genotyping technology which will allow the use of a very large GRS, where every SNP on a genotyping platform could be included. Constructing large-scale GRSs using tens of thousands of SNPs with CHD risk estimates (from GWASs) has been suggested [44]. While such a GRS may ultimately out-perform those constructed with only a selected subset of robustly associated SNPs, such an approach is not practical for a clinical setting, at least in the short-to-medium term. Abraham et al recently published a study [45] using five prospective population cohorts from Finland and the US, where over 49,000 SNPs were used. Addition of the GRS to CRF scores significantly improved the 10 years risk prediction (P < 0.001), particularly for individuals ≥60 years old (p < 0.001). However, how practical this approach will be in the short-to-medium term remains to be seen. The statistical stringency for inclusion of a SNP could be relaxed only partially, by selecting all SNPs identified in the meta-analysis with an false discovery rate of say >10%. This approach has been assessed by Gana et al [40], but a 353 SNP GRS did not improve performance over the use of 46 robustly associated SNPs. Another approach would be to only include variants known to contribute to the process to atherosclerosis and/or of the onset of myocardial infarction, and thus a GRS could be created that is more robustly associated with CHD [46] incidence, and this may provide a better instrument to estimate genetic risk of CHD. However, this would ignore variants that have been consistently robustly associated with CHD with no known mechanism of action. Ultimately a large-scale fully powered prospective study is required to alleviate the problem of survival bias in genetic association studies. If such data
became available this could be used to improve the precision of the weights for the GRS and that should improve performance.

An alternative strategy to using all of the robustly associated GWAS risk variants is to select only variants known to be functional, for example by using the ENCODE [47] and eQTL [48] databases and protein change algorithms. As can be seen in Figure 2 [49], only 7% of the GWAS hits are exonic altering an amino-acid in the cognate gene product and therefore may be directly affecting gene function, with another 7% being in the promoter region and thus may potentially be affecting gene expression. Over 70% are intronic or intergenic and are therefore of unknown or possibly of no functional consequence. The most notable such locus is on chromosome 9p21, which has the largest effect size of any common variant (see Table 1) but whose mechanism of action in the development of CHD remains obscure, a decade after discovery. In particular, the actual functional SNP or SNPs at this locus has not been definitively identified. Thus most of the SNPs included in the score are GWAS hits where the lead SNP (i.e. that included in the score) is unlikely to be the functional SNP at that risk locus. Linkage disequilibrium (LD) between the lead and functional SNPs may differ between ethnicities, meaning that some SNPs will be better proxies than others. This will reduce the ability of the weighted GRS to accurately reflect CHD risk, particularly in different ethnic groups if the LD is less. Therefore, the second area of progress over the next five years will be the identification of the functional SNP (or possibly SNPs) at each risk locus, and molecular approaches [49] and guidelines for this process have recently been published [50].

As an example of this, the Low-Density Lipoprotein Receptor (LDLR) SNP rs6511720 (G>T), located in intron-1 of the gene, has been identified in GWAS as being associated with ~0.2mmol/l lower plasma levels of LDL-C and a ~12% lower risk of CHD [25]. Whether or not rs6511720 is itself functional or a marker for a functional variant elsewhere in the gene is not known. The SNP is in complete linkage disequilibrium with three intron-1 SNPs (rs141787760, rs60173709, rs57217136), so only one or more than one of these variants may be the functional SNP at this locus. Using luciferase reporter assays in the hepatoma cell line Huh7 cells we showed [51] that the rare alleles of both rs6511720 and rs57217136 (but not rs60173709) caused a significant increase in LDLR expression compared to the common alleles (+29% and +24%, respectively).
Multiplex Competitor-Electrophoretic Mobility Shift Assays identified that the transcription factor serum response element (SRE) binds to rs6511720, while retinoic acid receptor and signal transducer and activator of transcription 1 bind to rs57217136. These data show convincingly that both rs6511720 and rs57217136 are functional variants. Both minor alleles create enhancer-binding protein sites for transcription factors and therefore contribute to increased \textit{LDLR} expression, which is consequently associated with reduced LDL-C levels and 12\% lower CHD risk.

While analysing every GWAS CHD hit locus in this way constitutes a considerable amount of work, ultimately this will provide the most accurate panel of functional SNPs for risk prediction.

Finally we need to be able to extrapolate the use of these GRSs, developed and validated in mainly cohorts of white Caucasians, to subjects of different ethnic backgrounds. In comparison to the genetics of CHD in those of Europeans ethnicity, very little is known about the genetics of CHD in either the South Asian or Afro-Caribbean populations. We have assessed the utility of the 19 SNP GRS in two case-control cohorts from Pakistan (Islamabad and Lahore) [35] and one from Guadeloupe in the Caribbean [52]. The GRS was higher in Afro-Caribbeans with CHD than in those without CHD (13.90 vs 13.17; \( P < 0.001 \)) and was significantly associated with CHD after adjustment for cardiovascular risk factors, with an odds ratio of 1.40 (95\% confidence interval, 1.09-1.80) per standard deviation change. Not surprisingly, there were significant differences in allelic distributions between the NPHSII men and the Afro-Caribbean ethnic groups for 14 of the 19 SNPs, and the GRS was substantially lower in Afro-Caribbean controls compared with white controls (13.17 vs 16.59; \( P < 0.001 \)). By contrast, although the 19 SNP GRS (with the original weightings) were higher in cases compared to controls in the Islamabad sample GRS (2.24 v 2.34, \( p=0.04 \)), the score was not associated with CHD after adjustment for age and sex, and there was no difference in GRS between cases and controls in the Lahore group. As in the Caribbean sample, in the Pakistani samples, risk allele frequencies were significantly lower compared to NPHSII for 19 SNPs studied. The lack of an association is unlikely to be due to low power, rather, the poor performance of the GSs in the Lahore group can be at least partly attributed to the much broader definition of CHD used in recruitment the case group compared to the Islamabad study which used an MI phenotype (more like the “hard” endpoints used in the prospective NPHSII and the Guadeloupe study). Overall, the results indicate that the GSs provide a useful estimate of genetic
CHD risk in Afro-Caribbeans from Guadeloupe at least, but firm conclusions cannot be drawn from the Pakistani data, although they suggest that a different set of risk loci or SNPs may be required for risk prediction in the South Asian population.

**Expert Commentary**

In our view, it is now possible to implement a personal Genetic Risk Score, in combination with the individual’s CRFs to better risk identify subjects at high risk of future CVD. The optimum number of SNPs to be included and which SNPs to choose, still requires further work, but since the OR for individual SNPs are all low, the inclusion or omission of any few SNPs does not appear to materially influence the utility of the GRS. We are unaware of any study that compares utility of scores that systematically drop or include SNPs that could address this. In the future, the technical issues of including a large number of SNPs (with vanishingly small and thus statistically unproven CHD risk) should be overcome using more advanced statistical and computational approaches. The main clinical utility is a modest improvement in risk stratification in subjects at intermediate future CVD risk, where the GRS helps identify subjects at higher than average CVD risk that their CRF risk data suggests. Current data is encouraging that individuals given a high genetic risk find the information motivation to modify their behaviours to reduce their future risk. The clinical utility of the GS described here depends on the context, as pointed out recently for Type 2 Diabetes [34]. If a clinician is trying to predict the risk score of 65 year old men, the GS is irrelevant, since the vast majority will qualify for statin treatment under QRISK2 threshold set in the current NICE guidelines. By contrast for the age of 40 or 30 or even at birth the situation might be different. For example, at birth there will almost never be CRFs of concern but the GS can point much further in the future, suggesting that this individual might need to see a doctor when in their late thirties instead of past 40, as may otherwise be the case.
Acknowledgements

SEH is a British Heart Foundation Professor and is supported by the British Heart Foundation (RG008/08) and by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. KB is supported by a Medical Research Council CASE award (1270920) with Randox Laboratories. FD is supported from the UK Medical Research Council (MC_UU_12013/1–9)

Conflicts of Interest

Steve Humphries is the Medical Director and minority shareholder of a UCL spin-out company called StoreGene, which uses a 20 SNP genetic test, in combination with the classical risk factor profile, for estimating an individual’s future risk of CVD.

Key Issues 8-10 Bullet points

- Classical Risk factor Algorithms such as Framingham and QRISK identify as “high risk” only a proportion of those subjects who subsequently go on to develop Coronary Heart Disease (CHD)
- Candidate genes and common variants (SNPs) in these genes have been used to develop early Genetic Risk Scores (GRS).
- The ability to carry out “hypothesis-free” gene discovery by Genome wide association studies (GWAS) has now led to the identification of more than 80 CHD risk loci, including many of the candidate genes identified earlier
- These CHD risk SNPs can be combined in a GRS by weighting carriage of the risk allele by the meta-analysis effect size (dds Ratio) on CHD.
- The 25 SNPs ranked by the product of their risk allele frequency and CHD Odds Ratio are unlikely to be bettered by any subsequent GWAS findings
- While a GRS may be associated significantly with CHD risk, for it to have clinical utility it must add significantly to the ability of the CRF algorithm to risk stratify.
• Metrics to determine clinical utility include the Hosmer-Lemeshow test (to assess calibration), Area under the Receiver operating curve, (AROC) and improvement in Net Classification Index (NRI).

• GRS with as few as 19 selected SNPs have been shown to have modest clinical utility in middle aged UK men. Inclusion of up to 50-160 additional SNPs gives some small gain in clinical utility.

• Clinical utility is mainly confined to these subjects at intermediate risk as determined by CRFs, where subjects who move up a risk category using the GRS do have a high rate of future events, and could therefore be offered therapeutic advice.
References

Key Refs * of interest or ** of considerable interest”.


[23] The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447: 661-678


*GWAS data from 63,746 CAD cases and 130,681 controls. Identified 15 novel loci, taking the number of statistically robust CAD to 46, with a further 104 SNPs associated with CAD at a 5% false discovery rate (FDR). Together, these variants explain approximately 10.6% of CAD heritability. Of the 46 genome-wide significant lead SNPs, 12 show a significant association with a lipid trait, and 5 show a significant association with blood pressure.


[34] Drenos F, Whittaker JC, Humphries SE. The use of Meta-Analysis Risk Estimates for Candidate Genes in Combination to Predict Coronary Heart Disease Risk. Ann Hum Genet. 2007; 71: 611-619


* Tested the use of a 19 SNP GRS in a study of 2775 healthy UK men (284 cases), and Pakistani case-control studies from Islamabad/Rawalpindi (321 cases/228 controls) and Lahore (414 cases/219 controls). The GRS showed potential clinical utility in European men but much less utility in subjects from Pakistan, suggesting that a different set of risk loci or SNPs may be required for risk prediction in the South Asian population.


* Data were from seven UK prospective studies including 11 851 individuals initially free of CVD or diabetes, with 1444 incident CVD events over 10 years' follow-up. Used 53 CVD GWAS SNPs and QRISK-2. Applying the GRS only to those with QRISK-2 risk of 10%<20% and prescribing statins where risk exceeded 20% suggested that genetic information could prevent one additional event for every 462 people screened. Tailored prediction using genetic information for those at intermediate risk may have clinical utility.

**5899 subjects, 152 CAD CAD** from the CARDIoGRAMplusC4D consortium. Constructed three weighted GRSs based on 49 genome-wide significant SNP, on 103 suggestively associated SNPs; and based on all 152 SNPs. All GRSs were associated with incident CHD despite adjustment for traditional risk factors and family history: participants had a 13% higher rate of CHD per standard deviation increase. While the GRSs did not improve reclassification, GRS(all) strongly improved both discrimination and reclassification of prevalent CHD, even beyond traditional risk factors and family history.


**Used 6 Swedish prospective cohort studies with 10 612 participants free of CHD at baseline and 781 incident events, and a CHD-specific GRS with 46 SNPs. This was significantly associated with CHD risk (hazard ratios for fourth versus first quartile, 1.52; \( P<0.001 \)) and improved risk classification beyond established risk factors (net reclassification improvement, 4.2% and 4.9%; \( P=0.006 \) and 0.017). Using a GRS of 365 SNPs did not improve this. Estimated that 1 additional CHD event for every 318 people screened at intermediate risk could be saved by measuring the CHD-specific genetic score in addition to the established risk factors.


**Participants (mean age, 59 years; 48% men; mean 10-year CHD risk, 8.5%) were randomised to receive either CRF (n=100) or (+)GRS (n=103), risk information. Disclosure of CHD risk estimates that incorporated genetic risk information led to ~9% lower LDL-C levels than disclosure of CHD risk based on conventional risk factors alone.


**Used five prospective population cohorts from Finland and the US and a GRS of over 49,000 SNPs. Addition of the GRS to CRF scores significantly improved the 10 years risk prediction...
(P < 0.001), particularly for individuals ≥60 years old (p < 0.001). The GRS captured substantially different trajectories of absolute risk, with men in the top 20% of the GRS reaching a 10% CHD risk 12-18 y earlier than those in the bottom 20%. High genetic risk was partially compensated for by low systolic blood pressure, low cholesterol level, and non-smoking.


*Review focusing on approaches to identify causal, noncoding variants for CAD, lipid traits and other cardiovascular risk factors. Genetic fine-mapping studies are discussed, along with the novel statistical approaches to produce 'credible sets' of SNPs for further analysis and state-of-the-art molecular methods to explore functionality discussed.


Table 1: Top 25 CARDioGRAMplusC4D CHD risk loci [25] ranked by ln(OR) multiplied by RAF

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Lead SNP</th>
<th>Gene/Locus</th>
<th>Odds Ratio (OR)</th>
<th>Risk allele frequency (RAF)</th>
<th>ln(OR) x RAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>rs445925</td>
<td>ApoE-ApoC1*</td>
<td>1.13</td>
<td>0.9</td>
<td>0.110</td>
</tr>
<tr>
<td>4</td>
<td>rs7692387</td>
<td>GUCY1A3</td>
<td>1.13</td>
<td>0.81</td>
<td>0.099</td>
</tr>
<tr>
<td>9</td>
<td>rs1333049</td>
<td>9p21*</td>
<td>1.23</td>
<td>0.47</td>
<td>0.097</td>
</tr>
<tr>
<td>1</td>
<td>rs17114036</td>
<td>PPAP2B</td>
<td>1.11</td>
<td>0.91</td>
<td>0.095</td>
</tr>
<tr>
<td>1</td>
<td>rs602633</td>
<td>SORT1*</td>
<td>1.12</td>
<td>0.77</td>
<td>0.087</td>
</tr>
<tr>
<td>10</td>
<td>rs12413409</td>
<td>CYP17A1-CONNM2-NT5C2</td>
<td>1.1</td>
<td>0.89</td>
<td>0.085</td>
</tr>
<tr>
<td>19</td>
<td>rs1122608</td>
<td>LDLR</td>
<td>1.1</td>
<td>0.76</td>
<td>0.072</td>
</tr>
<tr>
<td>13</td>
<td>rs9515203</td>
<td>COL4A1-COL4A2</td>
<td>1.08</td>
<td>0.74</td>
<td>0.057</td>
</tr>
<tr>
<td>9</td>
<td>rs3217992</td>
<td>9p21</td>
<td>1.16</td>
<td>0.38</td>
<td>0.056</td>
</tr>
<tr>
<td>10</td>
<td>rs501120</td>
<td>CXCL12*</td>
<td>1.07</td>
<td>0.83</td>
<td>0.056</td>
</tr>
<tr>
<td>6</td>
<td>rs9369640</td>
<td>PHACTR1</td>
<td>1.09</td>
<td>0.65</td>
<td>0.056</td>
</tr>
<tr>
<td>7</td>
<td>rs11556924</td>
<td>ZC3HC1</td>
<td>1.09</td>
<td>0.65</td>
<td>0.056</td>
</tr>
<tr>
<td>8</td>
<td>rs264</td>
<td>LPL*</td>
<td>1.06</td>
<td>0.86</td>
<td>0.050</td>
</tr>
<tr>
<td>13</td>
<td>rs4773144</td>
<td>COL4A1-COL4A2</td>
<td>1.07</td>
<td>0.74</td>
<td>0.050</td>
</tr>
<tr>
<td>6</td>
<td>rs4252120</td>
<td>PLG</td>
<td>1.07</td>
<td>0.73</td>
<td>0.049</td>
</tr>
<tr>
<td>1</td>
<td>rs11206510</td>
<td>PCSK9*</td>
<td>1.06</td>
<td>0.84</td>
<td>0.049</td>
</tr>
<tr>
<td>1</td>
<td>rs17464857</td>
<td>MIA3*</td>
<td>1.05</td>
<td>0.87</td>
<td>0.042</td>
</tr>
<tr>
<td>1</td>
<td>rs4846525</td>
<td>IL6R</td>
<td>1.09</td>
<td>0.47</td>
<td>0.041</td>
</tr>
<tr>
<td>15</td>
<td>rs7173743</td>
<td>ADAMTS7</td>
<td>1.07</td>
<td>0.58</td>
<td>0.039</td>
</tr>
<tr>
<td>17</td>
<td>rs12936587</td>
<td>RAII-PEMT-RASD1</td>
<td>1.06</td>
<td>0.59</td>
<td>0.034</td>
</tr>
<tr>
<td>6</td>
<td>rs12205331</td>
<td>ANKS1A</td>
<td>1.04</td>
<td>0.81</td>
<td>0.032</td>
</tr>
<tr>
<td>13</td>
<td>rs9319428</td>
<td>FLT1</td>
<td>1.1</td>
<td>0.32</td>
<td>0.030</td>
</tr>
<tr>
<td>2</td>
<td>rs1561198</td>
<td>VAMP5-VAMP8-GGCX</td>
<td>1.07</td>
<td>0.45</td>
<td>0.030</td>
</tr>
<tr>
<td>12</td>
<td>rs3184504</td>
<td>SH2B3</td>
<td>1.07</td>
<td>0.4</td>
<td>0.027</td>
</tr>
<tr>
<td>8</td>
<td>rs2954029</td>
<td>TRIB1</td>
<td>1.05</td>
<td>0.55</td>
<td>0.027</td>
</tr>
</tbody>
</table>

RAF = Risk allele frequency. Genes shown in bold are involved in lipid metabolism. * = Loci included in 19 SNP GRS
Table 2: Baseline characteristics in NPHSII for those who did and did not go on to develop CHD during ten-year follow-up

<table>
<thead>
<tr>
<th>Trait</th>
<th>No CHD (n=2491)</th>
<th>CHD (n=284)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>55.91 (3.42)</td>
<td>56.64 (3.60)</td>
<td>4.12x10^{-3}</td>
</tr>
<tr>
<td>Sex (% Male)</td>
<td>100 %</td>
<td>100 %</td>
<td>-</td>
</tr>
<tr>
<td>Smoking</td>
<td>25 %</td>
<td>39 %</td>
<td>2.14x10^{-5}</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>26.38 (3.42)</td>
<td>27.19 (3.44)</td>
<td>9.61x10^{-4}</td>
</tr>
<tr>
<td>Systolic Blood Pressure (mmHg)</td>
<td>137.00 (18.59)</td>
<td>144.09 (20.10)</td>
<td>9.68x10^{-7}</td>
</tr>
<tr>
<td>Total Cholesterol (mmol/l)</td>
<td>5.71 (1.01)</td>
<td>6.13 (1.05)</td>
<td>4.79x10^{-8}</td>
</tr>
<tr>
<td>LDL-cholesterol (mmol/l)</td>
<td>3.07 (1.00)</td>
<td>3.48 (0.97)</td>
<td>2.66x10^{-7}</td>
</tr>
<tr>
<td>HDL-cholesterol (mmol/l)</td>
<td>1.72 (0.59)</td>
<td>1.57 (0.53)</td>
<td>2.60x10^{-4}</td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
<td>2 %</td>
<td>7 %</td>
<td>1.33x10^{-11}</td>
</tr>
<tr>
<td>Framingham ten-year CHD risk</td>
<td>0.12 (0.07-0.15)</td>
<td>0.17 (0.09-0.21)</td>
<td>4.33x10^{-1}</td>
</tr>
<tr>
<td>QRISK2 ten-year CVD risk</td>
<td>0.09 (0.07-0.13)</td>
<td>0.13 (0.09-0.17)</td>
<td>1.93x10^{-14}</td>
</tr>
</tbody>
</table>

All variables are presented as the mean plus standard deviation, unless otherwise stated. Categorical variables were compared using chi-squared tests and continuous variables were compared using Welch’s t-tests, apart from the Framingham and QRISK2 risk scores which were compared using Mann Whitney tests (the median and interquartile range are given).
Table 3: SNPs included in the 19-SNP and 21 SNP CHD risk GRSs.

### SNPs included in the 19 SNP GS

<table>
<thead>
<tr>
<th>Gene/Locus</th>
<th>SNP</th>
<th>Risk Allele</th>
<th>OR</th>
<th>OR in original score</th>
<th>Frequency</th>
<th>p-value*</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>APOE*</td>
<td>rs7412</td>
<td>C</td>
<td>1.25</td>
<td>0.80</td>
<td>0.87</td>
<td>-</td>
<td>[53]</td>
</tr>
<tr>
<td>APOE*</td>
<td>rs429358</td>
<td>C</td>
<td>1.06</td>
<td>1.06</td>
<td>0.26</td>
<td>-</td>
<td>[53]</td>
</tr>
<tr>
<td>PCSK9*</td>
<td>rs11591147</td>
<td>G</td>
<td>1.39</td>
<td>1.43</td>
<td>0.99</td>
<td>-</td>
<td>[54]</td>
</tr>
<tr>
<td>CDKN2A/9p21*</td>
<td>rs10757274</td>
<td>G</td>
<td>1.23</td>
<td>1.29</td>
<td>0.47</td>
<td>1.39x10^{-52}</td>
<td>[25]</td>
</tr>
<tr>
<td>SORT1*</td>
<td>rs599839</td>
<td>A</td>
<td>1.11</td>
<td>1.19</td>
<td>0.77</td>
<td>3.8x10^{15}</td>
<td>[25]</td>
</tr>
<tr>
<td>LPA*</td>
<td>rs10455872</td>
<td>G</td>
<td>1.32</td>
<td>1.70</td>
<td>0.06</td>
<td>3.80x10^{13}</td>
<td>[55]</td>
</tr>
<tr>
<td>MIA3*</td>
<td>rs17465637</td>
<td>C</td>
<td>1.14</td>
<td>1.14</td>
<td>0.74</td>
<td>1.36x10^{-8}</td>
<td>[55]</td>
</tr>
<tr>
<td>MRAS*</td>
<td>rs9818870</td>
<td>T</td>
<td>1.07</td>
<td>1.15</td>
<td>0.14</td>
<td>2.62x10^{-9}</td>
<td>[25]</td>
</tr>
<tr>
<td>CXCL12*</td>
<td>rs1746048</td>
<td>C</td>
<td>1.07</td>
<td>1.17</td>
<td>0.83</td>
<td>1.79x10^{-8}</td>
<td>[25]</td>
</tr>
<tr>
<td>LPL*</td>
<td>rs328</td>
<td>C</td>
<td>1.09</td>
<td>1.25</td>
<td>0.91</td>
<td>2.34x10^{-4}</td>
<td>[25]</td>
</tr>
<tr>
<td>DAB2IP*</td>
<td>rs7025486</td>
<td>A</td>
<td>1.04</td>
<td>1.16</td>
<td>0.29</td>
<td>2.14x10^{-3}</td>
<td>[25]</td>
</tr>
<tr>
<td>LPL*</td>
<td>rs1801177</td>
<td>A</td>
<td>1.10</td>
<td>1.33</td>
<td>0.06</td>
<td>4.04x10^{-4}</td>
<td>[25]</td>
</tr>
<tr>
<td>LPA*</td>
<td>rs3798220</td>
<td>C</td>
<td>1.28</td>
<td>1.92</td>
<td>0.01</td>
<td>4.90x10^{5}</td>
<td>[25]</td>
</tr>
<tr>
<td>APOA5*</td>
<td>rs662799</td>
<td>G</td>
<td>1.05</td>
<td>1.19</td>
<td>0.06</td>
<td>0.01</td>
<td>[25]</td>
</tr>
<tr>
<td>CETP</td>
<td>rs708272</td>
<td>C</td>
<td>1.04</td>
<td>1.28</td>
<td>0.56</td>
<td>0.04</td>
<td>[55]</td>
</tr>
<tr>
<td>ACE</td>
<td>rs4341</td>
<td>G</td>
<td>1.01</td>
<td>1.22</td>
<td>0.52</td>
<td>0.43</td>
<td>[55]</td>
</tr>
<tr>
<td>APOB</td>
<td>rs1042031</td>
<td>A</td>
<td>1.01</td>
<td>1.73</td>
<td>0.18</td>
<td>0.80</td>
<td>[25]</td>
</tr>
<tr>
<td>NOS3</td>
<td>rs1799983</td>
<td>G</td>
<td>1.00</td>
<td>1.31</td>
<td>0.67</td>
<td>0.90</td>
<td>[55]</td>
</tr>
<tr>
<td>SMAD3</td>
<td>rs17228212</td>
<td>C</td>
<td>1.01</td>
<td>1.21</td>
<td>0.31</td>
<td>0.94</td>
<td>[25]</td>
</tr>
</tbody>
</table>

### Additional SNPs included in the 21 SNP GS

<table>
<thead>
<tr>
<th>Gene/Locus</th>
<th>SNP</th>
<th>Risk Allele</th>
<th>OR</th>
<th>Frequency</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUCY1A3</td>
<td>rs7692387</td>
<td>G</td>
<td>1.13</td>
<td>0.81</td>
<td>[25]</td>
</tr>
<tr>
<td>PPAP2B</td>
<td>rs17114036</td>
<td>T</td>
<td>1.11</td>
<td>0.91</td>
<td>[25]</td>
</tr>
<tr>
<td>CYP17A1-CNNM2-NT5C2</td>
<td>rs12413409</td>
<td>G</td>
<td>1.10</td>
<td>0.89</td>
<td>[25]</td>
</tr>
<tr>
<td>LDLR</td>
<td>rs1122608</td>
<td>G</td>
<td>1.10</td>
<td>0.76</td>
<td>[25]</td>
</tr>
<tr>
<td>COL4A1-COL4A2</td>
<td>rs9515203</td>
<td>T</td>
<td>1.08</td>
<td>0.74</td>
<td>[25]</td>
</tr>
<tr>
<td>PHACTR1</td>
<td>rs9369640'</td>
<td>A</td>
<td>1.09</td>
<td>0.65</td>
<td>[25]</td>
</tr>
<tr>
<td>ZC3HC1</td>
<td>Rs11556924</td>
<td>C</td>
<td>1.09</td>
<td>0.65</td>
<td>[25]</td>
</tr>
</tbody>
</table>

* Included in both the 19 SNP GS and the 21 SNP GS. For detailed methods see [35, 37]. OR = Odds Ratio. SNPs in highlighted section were those excluded in the 21 SNP score.

Table 4 in separate file
Figure Legends

Figure 1 Cartoon showing an idealised distribution of the QRISK risk score in the NPHSII men and the risk cut-offs recommend for use in the UK by NICE. 1A) using QRISK, 1B) QRISK + GRS

Foot note. The data is based on a subset of 1213 NPHSII men (mean age 57 years) where complete 19 SNP genotype data and QRISK CRF data was available [37]. There were 133 CVD events over 13 years of follow up (event rate overall 11.0%).
Figure 2
Pie chart showing the Genomic location of lead SNPs for CAD relative to the nearest gene of CHD GWAS hits (from [49]). Lead SNP data derived from CardioGramPLUSC4D meta-analysis [25].