

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

Katherine Beaney<sup>1\*</sup>, Fotios Drenos<sup>1,2</sup>, and Steve E Humphries<sup>1</sup>

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](mailto: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  $\geq 20\%$  [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  $\geq 7.5\%$  [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  $\geq 10\%$  [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 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

[17]. 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 therapeutic 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,  $1.42 \times 10^{-3}$ ), 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<sup>15</sup>, 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 possibly 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  $\geq 60$  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 a 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 *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 GRSs 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 GRSs 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 (Odds 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” .

- [1] Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circ*. 1998; 97: 1837-1847
- [2] D'Agostino RB, Grundy S, Sullivan LM, Wilson P, Grp CHDRP. Validation of the Framingham Coronary Heart Disease prediction scores - Results of a multiple ethnic groups investigation. *JAMA-J Am Med Assoc*. 2001;286(2):180-7.
- [3] Barzi F, Patel A, Gu D, Sritara P, Lam TH, Rodgers A, et al. Cardiovascular risk prediction tools for populations in Asia. *J Epidemiol Community Health*. 2007;61(2):115-21. Epub 2007/01/20.
- [4]. Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T, et al. Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. *BMJ*. 2003;327(7426):1267. Epub 2003/12/04.
- [5] Hense HW, Schulte H, Lowel H, Assmann G, Keil U. Framingham risk function overestimates risk of coronary heart disease in men and women from Germany--results from the MONICA Augsburg and the PROCAM cohorts. *Eur Heart J*. 2003;24(10):937-45. Epub 2003/04/26.
- [6] Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, et al. (2003) Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal* 24: 987-1003.
- [7] 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(7611):136. Epub 2007/07/07.
- [8] Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ*. 2008;336(7659):1475-82. Epub 2008/06/25.
- [9] Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T, et al. Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. *BMJ*. 2003;327(7426):1267. Epub 2003/12/04.
- [10] Cholesterol Treatment Trialists' (CTT) Collaboration., Baigent C, Blackwell L, Emberson J, Holland LE, Reith C, Bhalra N, Peto R, Barnes EH, Keech A, Simes J, Collins R. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet*. 2010 Nov 13;376(9753):1670-81.

- [11] Cooper A, O'Flynn N; Guideline Development Group.. Risk assessment and lipid modification for primary and secondary prevention of cardiovascular disease: summary of NICE guidance. *BMJ*. 2008 May 31;336(7655):1246-8.
- [12] Goff DC, Jr., Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2014;129(25 Suppl 2):S49-73.
- [13] National Institute for Health and Care Excellence : Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. 2014.
- [14] Goldacre B, Smeeth L. Mass treatment with statins. *BMJ*. 2014 Jul 22;349:g4745.
- [15] Usher-Smith JA, Pritchard J, Poole S, et al. Offering statins to a population attending health checks with a 10-year cardiovascular disease risk between 10% and 20. *International journal of clinical practice* 2015.
- [16] Rose G. Strategy of prevention: lessons from cardiovascular disease. *Br Med J (Clin Res Ed)*. 1981;282(6279):1847-51. Epub 1981/06/06
- [17] Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. *BMJ*. 2010;340:c2442. Epub 2010/05/15.
- [18] Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med*. 2002 Sep;252(3):247-54.
- [19] Wienke A, Herskind AM, Christensen K, Skytthe A, Yashin AI. The heritability of CHD mortality in danish twins after controlling for smoking and BMI. *Twin Res Hum Genet*. 2005 Feb;8(1):53-9.
- [20] Casas JP, Cooper J, Miller GJ, Hingorani AD, Humphries SE. Investigating the genetic determinants of cardiovascular disease using candidate genes and meta-analysis of association studies. *Ann Hum Genet*. 2006; 70: 145-169
- [21] 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
- [22] Helgadóttir A, Thorleifsson G, Manolescu A, Gretarsdóttir S, Blondal T, Jonasdóttir A, Jonasdóttir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson D, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdóttir S, Jonsdóttir T, Palsson S, Einarsdóttir H, Gunnarsdóttir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdóttir U, Kong A, Stefansson K. A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction. *Science*. 2007; 316: 1491-1493
- [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

[24] Humphries SE, Drenos F, Ken-Dror G, Talmud PJ. Coronary heart disease risk prediction in the era of genome-wide association studies: current status and what the future holds. *Circulation*. 2010 May 25;121(20):2235-48

[25] CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, König IR, Cazier JB, Johansson A, Hall AS, Lee JY, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lyytikäinen LP, Rafelt S, Shungin D, Strawbridge RJ, Thorleifsson G, Tikkanen E, Van Zuydam N, Voight BF, Waite LL, Zhang W, Ziegler A, Absher D, Altshuler D, Balmforth AJ, Barroso I, Braund PS, Burgdorf C, Claudi-Boehm S, Cox D, Dimitriou M, Do R; DIAGRAM Consortium; CARDIOGENICS Consortium, Doney AS, El Mokhtari N, Eriksson P, Fischer K, Fontanillas P, Franco-Cereceda A, Gigante B, Groop L, Gustafsson S, Hager J, Hallmans G, Han BG, Hunt SE, Kang HM, Illig T, Kessler T, Knowles JW, Kolovou G, Kuusisto J, Langenberg C, Langford C, Leander K, Lokki ML, Lundmark A, McCarthy MI, Meisinger C, Melander O, Mihailov E, Maouche S, Morris AD, Müller-Nurasyid M; MuTHER Consortium, Nikus K, Peden JF, Rayner NW, Rasheed A, Rosinger S, Rubin D, Rumpf MP, Schäfer A, Sivananthan M, Song C, Stewart AF, Tan ST, Thorgeirsson G, van der Schoot CE, Wagner PJ; Wellcome Trust Case Control Consortium, Wells GA, Wild PS, Yang TP, Amouyel P, Arveiler D, Basart H, Boehnke M, Boerwinkle E, Brambilla P, Cambien F, Cupples AL, de Faire U, Dehghan A, Diemert P, Epstein SE, Evans A, Ferrario MM, Ferrières J, Gauguier D, Go AS, Goodall AH, Gudnason V, Hazen SL, Holm H, Iribarren C, Jang Y, Kähönen M, Kee F, Kim HS, Klopp N, Koenig W, Kratzer W, Kuulasmaa K, Laakso M, Laaksonen R, Lee JY, Lind L, Ouwehand WH, Parish S, Park JE, Pedersen NL, Peters A, Quertermous T, Rader DJ, Salomaa V, Schadt E, Shah SH, Sinisalo J, Stark K, Stefansson K, Trégouët DA, Virtamo J, Wallentin L, Wareham N, Zimmermann ME, Nieminen MS, Hengstenberg C, Sandhu MS, Pastinen T, Syvänen AC, Hovingh GK, Dedoussis G, Franks PW, Lehtimäki T, Metspalu A, Zalloua PA, Siegbahn A, Schreiber S, Ripatti S, Blankenberg SS, Perola M, Clarke R, Boehm BO, O'Donnell C, Reilly MP, März W, Collins R, Kathiresan S, Hamsten A, Kooner JS, Thorsteinsdottir U, Danesh J, Palmer CN, Roberts R, Watkins H, Schunkert H, Samani NJ. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013 Jan;45(1):25-33. doi: 10.1038/ng.2480. Epub 2012 Dec 2.

\*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.

[26] Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjorndal A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang J, Hwang SJ, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikäinen LP, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith AV, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P,

Gauguier D, Goodall AH, Gottesman O, Haber M, Han BG, Huang J, Jalilzadeh S, Kessler T, König IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki ML, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon FU, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardissino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, Esko T, Feitosa MF, Franco OH, Franzosi MG, Granger CB, Gu D, Gudnason V, Hall AS, Hamsten A, Harris TB, Hazen SL, Hengstenberg C, Hofman A, Ingelsson E, Iribarren C, Jukema JW, Karhunen PJ, Kim BJ, Kooner JS, Kullo IJ, Lehtimäki T, Loos RJ, Melander O, Metspalu A, März W, Palmer CN, Perola M, Quertermous T, Rader DJ, Ridker PM, Ripatti S, Roberts R, Salomaa V, Sanghera DK, Schwartz SM, Seedorf U, Stewart AF, Stott DJ, Thiery J, Zalloua PA, O'Donnell CJ, Reilly MP, Assimes TL, Thompson JR, Erdmann J, Clarke R, Watkins H, Kathiresan S, McPherson R, Deloukas P, Schunkert H, Samani NJ, Farrall M; CARDIoGRAMplusC4D Consortium.. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015 Oct;47(10):1121-30

- [27] Dehghan A, Bis JC, White CC, Smith AV, Morrison AC, Cupples LA, Trompet S, Chasman DI, Lumley T, Völker U, Buckley BM, Ding J, Jensen MK, Folsom AR, Kritchevsky SB, Girman CJ, Ford I, Dörr M, Salomaa V, Uitterlinden AG, Eiriksdottir G, Vasani RS, Franceschini N, Carty CL, Virtamo J, Demissie S, Amouyel P, Arveiler D, Heckbert SR, Ferrières J, Ducimetière P, Smith NL, Wang YA, Siscovick DS, Rice KM, Wiklund PG, Taylor KD, Evans A, Kee F, Rotter JI, Karvanen J, Kuulasmaa K, Heiss G, Kraft P, Launer LJ, Hofman A, Markus MR, Rose LM, Silander K, Wagner P, Benjamin EJ, Lohman K, Stott DJ, Rivadeneira F, Harris TB, Levy D, Liu Y, Rimm EB, Jukema JW, Völzke H, Ridker PM, Blankenberg S, Franco OH, Gudnason V, Psaty BM, Boerwinkle E, O'Donnell CJ. Genome-Wide Association Study for Incident Myocardial Infarction and Coronary Heart Disease in Prospective Cohort Studies: The CHARGE Consortium. *PLoS One.* 2016 Mar 7;11(3):e0144997.
- [28] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature.* 2009 Oct 8;461(7265):747-53
- [29] Talmud PJ, Cooper JA, Palmieri 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 Mar;54(3):467-74.
- [30] Brautbar A, Ballantyne CM, Lawson K, Nambi V, Chambless L, et al. (2009) 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* 2: 279-285.

- [31] Paynter NP, Chasman DI, Buring JE, Shiffman D, Cook NR, et al. (2009) Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. *Annals of Internal Medicine* 150: 65-72
- [32] Drenos F. The use of genetic information in the prediction of Type 2 diabetes. *Pers Med.* 2015;12(5):483-96.
- [33] 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 Jan;53(1):8-16
- [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
- [35] Beaney KE, Cooper JA, Ullah Shahid S, Ahmed W, Qamar R, Drenos F, Crockard MA, Humphries SE. Clinical Utility of a Coronary Heart Disease Risk Prediction Gene Score in UK Healthy Middle Aged Men and in the Pakistani Population. *PLoS One.* 2015 Jul 2;10(7):e0130754.
- \*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.
- [36] Kraft P. Curses--winner's and otherwise--in genetic epidemiology. *Epidemiology.* 2008;19(5):649-51;
- [37] Beaney KE, Cooper JA, Drenos F and Humphries SE. Assessment of the Clinical Utility of Adding Common Single Nucleotide Polymorphism Genetic Scores to Classical Risk Factor Algorithms in Coronary Heart Disease Risk Prediction in UK men. *Clinical Chemistry and Laboratory Medicine*, 2017 in press
- [38] Morris RW, Cooper JA, Shah T, Wong A, Drenos F, Engmann J, McLachlan S, Jefferis B, Dale C, Hardy R, Kuh D, Ben-Shlomo Y, Wannamethee SG, Whincup PH, Casas JP, Kivimaki M, Kumari M, Talmud PJ, Price JF, Dudbridge F, Hingorani AD, Humphries SE; UCLEB Consortium.. Marginal role for 53 common genetic variants in cardiovascular disease prediction. *Heart.* 2016;102(20):1640-7.
- \* 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.
- [39] de Vries PS<sup>1</sup>, Kavousi M<sup>1</sup>, Ligthart S<sup>1</sup>, Uitterlinden AG<sup>2</sup>, Hofman A<sup>1</sup>, Franco OH<sup>1</sup>, Dehghan A<sup>3</sup> Incremental predictive value of 152 single nucleotide polymorphisms in the 10-year risk prediction of incident coronary heart disease: the Rotterdam Study. *Int J Epidemiol.* 2015 Apr;44(2):682-8. doi: 10.1093/ije/dyv070. Epub 2015 May 6.

\*\* 5899 subjects, 152 CAD cases from the CARDIoGRAMplusC4D consortium. Constructed three weighted GRSs based on 49 genome-wide significant SNPs, 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.

[40] Ganna A, Magnusson PK, Pedersen NL, de Faire U, Reilly M, Arnlöv J, Sundström J, Hamsten A, Ingelsson E. Multilocus genetic risk scores for coronary heart disease prediction. *Arterioscler Thromb Vasc Biol.* 2013 Sep;33(9):2267-72

\*\*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.

[41] Tada H, Melander O, Louie JZ, Catanese JJ, Rowland CM, Devlin JJ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur Heart J.* 2016;37(6):561-7. Epub 2015/09/24.

[42] Kullo IJ, Jouni H, Austin EE, Brown SA, Kruisselbrink TM, Isseh IN, Haddad RA, Marroush TS, Shameer K, Olson JE, Broeckel U, Green RC, Schaid DJ, Montori VM, Bailey KR. Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). *Circulation.* 2016 Mar 22;133(12):1181-8

\*\* 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.

[43] Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, Chasman DI, Baber U, Mehran R, Rader DJ, Fuster V, Boerwinkle E, Melander O, Orho-Melander M, Ridker PM, Kathiresan S. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med.* 2016;375(24):2349-2358.

[44] Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 2013 ;9(3):e1003348.

[45] Abraham G, Havulinna AS, Bhalala OG, Byars SG, De Livera AM, Yetukuri L, Tikkanen E, Perola M, Schunkert H, Sijbrands EJ, Palotie A, Samani NJ, Salomaa V, Ripatti S, Inouye M. Genomic prediction of coronary heart disease. *Eur Heart J.* 2016 Nov 14;37(43):3267-3278.

\*\*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  $\geq 60$  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.

- [46] Dehghan A, Bis JC, White CC, Smith AV, Morrison AC, Cupples LA, Trompet S, Chasman DI, Lumley T, Völker U, Buckley BM, Ding J, Jensen MK, Folsom AR, Kritchevsky SB, Girman CJ, Ford I, Dörr M, Salomaa V, Uitterlinden AG, Eiriksdottir G, Vasana RS, Franceschini N, Carty CL, Virtamo J, Demissie S, Amouyel P, Arveiler D, Heckbert SR, Ferrières J, Ducimetière P, Smith NL, Wang YA, Siscovick DS, Rice KM, Wiklund PG, Taylor KD, Evans A, Kee F, Rotter JI, Karvanen J, Kuulasmaa K, Heiss G, Kraft P, Launer LJ, Hofman A, Markus MR, Rose LM, Silander K, Wagner P, Benjamin EJ, Lohman K, Stott DJ, Rivadeneira F, Harris TB, Levy D, Liu Y, Rimm EB, Jukema JW, Völzke H, Ridker PM, Blankenberg S, Franco OH, Gudnason V, Psaty BM, Boerwinkle E, O'Donnell CJ. Genome-Wide Association Study for Incident Myocardial Infarction and Coronary Heart Disease in Prospective Cohort Studies: The CHARGE Consortium. *PLoS One*. 2016 Mar 7;11(3):e0144997.
- [47] ENCODE Project Consortium., Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korb J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaöz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Löytynoja A, Whelan S, Pardi F, Masingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA; NISC Comparative Sequencing Program.; Baylor College of Medicine Human Genome Sequencing Center.; Washington

University Genome Sequencing Center.; Broad Institute.; Children's Hospital Oakland Research Institute., Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameer A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyraes E, Hallgrímsson IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007 Jun 14;447(7146):799-816

[48] Xia K, Shabalín AA, Huang S, Madar V, Zhou YH, Wang W, Zou F, Sun W, Sullivan PF, Wright FA. seeQTL: a searchable database for human eQTLs. *Bioinformatics*. 2012 Feb 1;28(3):451-2.

[49] Smith AJ, Humphries SE, Talmud PJ. Identifying functional noncoding variants from genome-wide association studies for cardiovascular disease and related traits. *Curr Opin Lipidol*. 2015 Apr;26(2):120-6.

\*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.

[50] MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, Conrad DF, Cooper GM, Cox NJ, Daly MJ, Gerstein MB, Goldstein DB, Hirschhorn JN, Leal SM, Pennacchio LA, Stamatoyannopoulos JA, Sunyaev SR, Valle D, Voight BF, Winckler W, Gunter C. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014 Apr 24;508(7497):469-76.



- [51] Fairwozy RH, White J, Palmén J, Kalea AZ, Humphries SE. Identification of the Functional Variant(s) that Explain the Low-Density Lipoprotein Receptor (LDLR) GWAS SNP rs6511720 Association with Lower LDL-C and Risk of CHD. *PLoS One*. 2016 Dec 14;11(12):e0167676.
- [52] Larifla L, Beaney KE, Foucan L, Bangou J, Michel CT, Martino J, Velayoudom-Cephise FL, Cooper JA, Humphries SE. Influence of Genetic Risk Factors on Coronary Heart Disease Occurrence in Afro-Caribbeans. *Can J Cardiol*. 2016 Aug;32(8):978-85
- [53] Bennet, A. M., E. Di Angelantonio, et al. (2007). "Association of apolipoprotein E genotypes with lipid levels and coronary risk." *JAMA* **298**(11): 1300-1311.
- [54] Benn, M., B. G. Nordestgaard, et al. (2010). "PCSK9 R46L, low-density lipoprotein cholesterol levels, and risk of ischemic heart disease: 3 independent studies and meta-analyses." *Journal of the American College of Cardiology* 55(25): 2833-2842
- [55] Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardisino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buyschaert I; Cardiogenics., Carlquist JF, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Diemert P, Do R, Doering A, Eifert S, Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de Faire U, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA, Kaplan LM, Kastelein JJ, Khaw KT, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Loley C, Lotery AJ, Mannucci PM, Maouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Mühleisen TW, Muhlestein JB, Münzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nöthen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schäfer A, Schillert A, Schreiber S, Schrezenmeir J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoop JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WH, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann HE, Wild PS, Willenborg C, Wittteman JC, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, März W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J; CARDIoGRAM Consortium., Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011 Mar 6;43(4):333-8.

**Table 1:** Top 25 CARDIoGRAMplusC4D CHD risk loci [25] ranked by ln(OR) multiplied by RAF

Chromosome	Lead SNP	Gene/Locus	Odds Ratio (OR)	Risk allele frequency (RAF)	ln(OR) x RAF
19	rs445925	<b>ApoE-ApoC1*</b>	1.13	0.9	0.110
4	rs7692387	<i>GUCY1A3</i>	1.13	0.81	0.099
9	rs1333049	9p21*	1.23	0.47	0.097
1	rs17114036	<i>PPAP2B</i>	1.11	0.91	0.095
1	rs602633	<b>SORT1*</b>	1.12	0.77	0.087
10	rs12413409	<i>CYP17A1-CNNM2-NT5C2</i>	1.1	0.89	0.085
19	rs1122608	<b>LDLR</b>	1.1	0.76	0.072
13	rs9515203	<i>COL4A1-COL4A2</i>	1.08	0.74	0.057
9	rs3217992	9p21	1.16	0.38	0.056
10	rs501120	<i>CXCL12*</i>	1.07	0.83	0.056
6	rs9369640	<i>PHACTR1</i>	1.09	0.65	0.056
7	rs11556924	<i>ZC3HC1</i>	1.09	0.65	0.056
8	rs264	<b>LPL*</b>	1.06	0.86	0.050
13	rs4773144	<i>COL4A1-COL4A2</i>	1.07	0.74	0.050
6	rs4252120	<i>PLG</i>	1.07	0.73	0.049
1	rs11206510	<b>PCSK9*</b>	1.06	0.84	0.049
1	rs17464857	<i>MIA3*</i>	1.05	0.87	0.042
1	rs4846525	<i>IL6R</i>	1.09	0.47	0.041
15	rs7173743	<i>ADAMTS7</i>	1.07	0.58	0.039
17	rs12936587	<i>RAI1-PEMT-RASD1</i>	1.06	0.59	0.034
6	rs12205331	<i>ANKS1A</i>	1.04	0.81	0.032
13	rs9319428	<i>FLT1</i>	1.1	0.32	0.030
2	rs1561198	<i>VAMP5-VAMP8-GGCX</i>	1.07	0.45	0.030
12	rs3184504	<i>SH2B3</i>	1.07	0.4	0.027
8	rs2954029	<b>TRIB1</b>	1.05	0.55	0.027

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**

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

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.**

<b>SNPs included in the 19 SNP GS</b>							
Gene/Locus	SNP	Risk Allele	OR	OR in original score	Frequency	p-value*	Source
<i>APOE*</i>	rs7412	C	1.25	0.80	0.87	-	[53]
<i>APOE*</i>	rs429358	C	1.06	1.06	0.26	-	[53]
<i>PCSK9*</i>	rs11591147	G	1.39	1.43	0.99	-	[54]
<i>CDKN2A/9p21*</i>	rs10757274	G	1.23	1.29	0.47	1.39x10 <sup>-52</sup>	[25]
<i>SORT1*</i>	rs599839	A	1.11	1.19	0.77	3.8x10 <sup>-15</sup>	[25]
<i>LPA*</i>	rs10455872	G	1.32	1.70	0.06	3.80x10 <sup>-13</sup>	[55]
<i>MIA3*</i>	rs17465637	C	1.14	1.14	0.74	1.36x10 <sup>-8</sup>	[55]
<i>MRAS*</i>	rs9818870	T	1.07	1.15	0.14	2.62x10 <sup>-9</sup>	[25]
<i>CXCL12*</i>	rs1746048	C	1.07	1.17	0.83	1.79x10 <sup>-8</sup>	[25]
<i>LPL*</i>	rs328	C	1.09	1.25	0.91	2.34x10 <sup>-4</sup>	[25]
<i>DAB2IP*</i>	rs7025486	A	1.04	1.16	0.29	2.14x10 <sup>-3</sup>	[25]
<i>LPL*</i>	rs1801177	A	1.10	1.33	0.06	4.04x10 <sup>-4</sup>	[25]
<i>LPA*</i>	rs3798220	C	1.28	1.92	0.01	4.90x10 <sup>-5</sup>	[25]
<i>APOA5*</i>	rs662799	G	1.05	1.19	0.06	0.01	[25]
<i>CETP</i>	rs708272	C	1.04	1.28	0.56	0.04	[55]
<i>ACE</i>	rs4341	G	1.01	1.22	0.52	0.43	[55]
<i>APOB</i>	rs1042031	A	1.01	1.73	0.18	0.80	[25]
<i>NOS3</i>	rs1799983	G	1.00	1.31	0.67	0.90	[55]
<i>SMAD3</i>	rs17228212	C	1.01	1.21	0.31	0.94	[25]
<b>Additional SNPs included in the 21 SNP GS</b>							
Gene/Locus	SNP	Risk Allele	OR	OR in original score	Frequency	p-value*	Source
<i>GUCY1A3</i>	rs7692387	G	1.13	1.13	0.81		[25]
<i>PPAP2B</i>	rs17114036	T	1.11	1.11	0.91		[25]
<i>CYP17A1-CNNM2-NT5C2</i>	rs12413409	G	1.10	1.10	0.89		[25]
<i>LDLR</i>	rs1122608	G	1.10	1.10	0.76		[25]
<i>COL4A1-COL4A2</i>	rs9515203	T	1.08	1.08	0.74		[25]
<i>PHACTR1</i>	rs9369640 <sup>f</sup>	A	1.09	1.09	0.65		[25]
<i>ZC3HC1</i>	Rs11556924	C	1.09	1.09	0.65		[25]

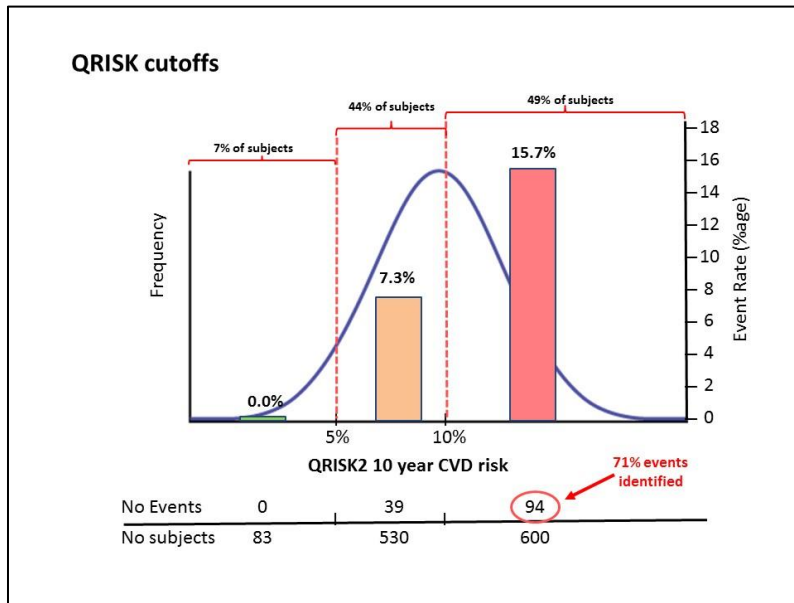
\* Included in both the 19 SNP GS and the 21 SNP GS. For detailed methods see [35, 37]. OR = Odds Ratio. SNPs in highlighted section were those excluded in the 21 SNP score

**Table 4 in separate file**

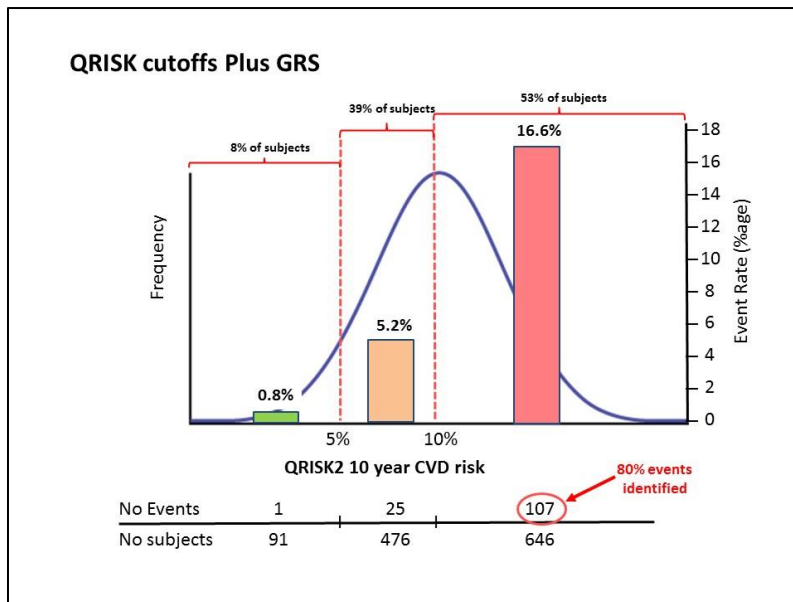
## Figure Legends

**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

1A



1B



**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].

