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

Katherine E Beaney[1], Jackie A Cooper[1], Fotios Drenos[1,2] and Steve E Humphries[1]

1 Centre for Cardiovascular Genetics, British Heart Foundation Laboratories, Institute of Cardiovascular Science, University College London, University Street, London, UK

2 MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

Corresponding author:
Steve Humphries
Cardiovascular Genetics
BHF Laboratories, Rayne Building,
Institute Cardiovascular Sciences
UCL, 5 University Street, London WC1E 6JJ
0207 679 6962
steve.humphries@ucl.ac.uk

Number of Words: 3394
Number of Figures: 2
Number of Tables: 4 (plus one supplementary table)
Number of References : 33

List of Abbreviations:

AUC – area under the receiver operator characteristic curve

CHD – coronary heart disease

CI- confidence interval

CRFs – conventional risk factors

CRP – C-reactive protein

CVD – cardiovascular disease

GS – gene score

GWAS – Genome-wide association study

HDL – high density lipoprotein

LDL- low density lipoprotein

MI – myocardial infarction

NPHSII – second Northwick Park heart study

NRI- net reclassification index

OR – odds ratio

sd – standard deviation

SNP – single nucleotide polymorphism

UCLEB - University College, London School of Hygiene and Tropical Medicine, Edinburgh and Bristol

Abstract

Background :
Risk prediction algorithms for coronary heart disease (CHD) are recommended for clinical use. However, their predictive ability remains modest and the inclusion of genetic risk may improve their performance.

Methods :
QRISK2 was used to assess CHD risk using conventional risk factors (CRFs). The performance of a 19 single nucleotide polymorphism (SNP) gene score (GS) for CHD including variants identified by genome-wide association study and candidate gene studies (weighted using the results from the CARDIoGRAMplusC4D meta-analysis) was assessed using the second Northwick Park Heart Study (NPHSII) of 2775 healthy UK men (284 cases). To improve the GS, five SNPs with weak evidence of an association with CHD were removed and replaced with seven robustly associated SNPs - giving a 21 SNP GS.

Results :
The weighted 19 SNP GS was associated with lipid traits ($p<0.05$) and CHD after adjustment for CRFs, (OR=1.31 per standard deviation, p=0.03). Addition of the 19 SNP GS to QRISK2 showed improved discrimination (area under the receiver operator characteristic curve 0.68 v 0.70 p=0.02), a positive net reclassification index (0.07, p=0.04) compared to QRISK2 alone and maintained good calibration (p=0.17). The 21 SNP GS was also associated with CHD after adjustment for CRFs (OR=1.39 per standard deviation, $1.42x10^{-3}$), but the combined QRISK2 plus GS 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.

Conclusion :
The 19 SNP GS is robustly associated with CHD and showed potential clinical utility in the UK population.

Key Words: Coronary Heart Disease, Genetic Risk, Risk Prediction

Introduction:

While a large proportion of coronary heart disease (CHD) events are preventable (1, 2), CHD remains a common cause of death worldwide. Therefore, predicting those at highest risk of developing the disease to target with lifestyle/therapeutic interventions is an important public health consideration. To take advantage of the combined knowledge of how conventional risk factors (CRFs) predispose individuals to CHD, risk scores have been developed. The first such score was derived from data collected as part of the prospective Framingham study (3) - referred to as the Framingham score. While the score showed good predictive ability in some cohorts similar to that from which it was derived (4, 5), it was found to overestimate risk in other ethnic groups (6) and in other populations of European ethnicity where there was a lower incidence of CHD (7, 8). The development of 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 cardiovascular disease (CVD) (9)) and the most recent National Institute for Health and Clinical Excellence (NICE) guidelines recommend the most recent version of this tool (QRISK2) for clinical use (10).

The majority of cases of CHD come from individuals assigned with average risk using the CRF risk scores – the so-called prevention paradox (11). When use of QRISK2 (2010 version) was validated with data from the health improvement network, 14% of men and 6% of women were identified as being at high risk (as determined by the guidelines at that time). This captured 40% of the cardiovascular events in men and 26% of the cardiovascular events in women (12). This leaves scope for refinement of the risk score to discriminate between those who do and do not go on to develop CVD. In addition, the inclusion of a number of additional risk factors for CHD not currently included has been proposed such as inflammatory markers (13), lipoprotein(a) (14) and genetic information.

It has thus far been unclear whether it is beneficial to include an estimate of genetic risk in CHD risk prediction. The Joint British Societies' consensus recommendations for the prevention of CVD (JBS3) did not advocate its use, concluding that the available evidence showed risk prediction tools including genetic information performed more poorly than CRF based tools (15). This has been underlined by the relatively disappointing performance of risk scores including gene scores (GSs) comprised of the variants identified in the CARDIoGRAMplusC4D meta-analysis of genome-wide association studies (GWASs) (16). Assessment in the prospective Rotterdam study (17) and the University College, London School of Hygiene and Tropical Medicine, Edinburgh and Bristol (UCLEB) consortium data (18)) found only a limited benefit in the population-wide inclusion of the GS in risk prediction, although improvements in both discrimination and reclassification were observed in a meta-analysis of six Swedish prospective cohorts (19) and the Malmo Diet and Cancer (MDC) study (20).

In 2007 we started to develop a multi-single nucleotide polymorphism (SNP) panel using 12 SNPs in candidate genes that, when used in combination with the Framingham classical risk factor algorithm in a cohort of ~2700 middle-aged men from the UK (NPHSII), had the potential to identify individuals at high future risk of CHD (21). We next showed that, in this same cohort, the addition of a single SNP from the first GWAS CHD locus identified on chromosome 9p21, improved the classical risk factor AROC by 3% but that this effect was not statistically significant (22). By modelling, we estimated that an additional three SNPs with similar risk size and risk allele frequencies would be needed to have a significant improvement. We therefore developed a 19 SNP GS and found it to be of potential clinical utility in the same NPHSII cohort (23). The GS comprised 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 risk allele frequencies for these SNPs varied between 0.01 to 0.99 and the published odds ratios from 1.06-1.92 (see table 1 in reference 21 and Table 1 below). However in the meta-analyses of large consortium data sets such as CARDIoGRAMplusC4D the odds ratios for several of the included SNPs had reduced 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. In addition, replacing those SNPs which have non-significant CHD risk effects with SNPs with larger risk effects should also improve clinical utility.

Thus, the aim of the current study was firstly to assess if the performance of the 19 SNP GS could be improved by updating the weightings, using the more accurate effect sizes derived from the CARDIoGRAMplusC4D GWAS meta-analysis (16). Secondly, we sought to investigate the relationship between the updated gene score and CHD CRFs. Thirdly, we assessed the use of the combined CRF plus GS scores in CHD risk prediction to determine if including the 19 SNP GS could provide any additional clinical utility. Finally, we assessed if removing SNPs not found to be associated with CHD in the CARDIoGRAMplusC4D meta-analysis and replacing them in the GS with those that were, improved its performance.

Methods:

The Second Northwick Park Heart Study (NPHSII)

NPHSII is a prospective CHD study of approximately 3000 men which has been described previously (24). Briefly, middle-aged men (50-64) were recruited from 9 general practices in the UK. Anyone with a history of CHD was excluded. There was a median of 13.5 years follow-up. CHD was defined as acute myocardial infarction (MI), silent MI or undergoing coronary surgery. Family history of early CHD was collected as reported previously (25). All subjects gave written informed consent and the study had ethical approval from the national research ethics service (NRES) Committee London-Central. The baseline characteristics have been published previously by CHD outcome in follow-up (23). The 2012 version of QRISK2 was calculated.

Genotyping:

For the 19 SNPs previously genotyped and published (23), some had been genotyped using restriction length polymorphism methods (21) but the majority of SNPs, as well as the seven SNPs added in this study, were genotyped using Taqman (Applied Biosystems, Life Technologies, Carlsbad California, USA) assays and KASPar (LGC. Teddington, Middlesex, UK) assays. The call rate for each genotype is listed in Table 1.

Gene Score

For the 19 SNP GS (23), we updated the weightings to the beta-coefficients determined in the CARDIoGRAMplusC4D meta-analysis as previously described (26). The values used are given in Table 1. Three SNPs (rs7412 and rs429358 in *APOE* and rs11591147 in *PCSK9*) were not included in this analysis and therefore the effect size from the most recent meta-analysis concerning the relationship between these SNPs and CHD was used as indicated. The SNPs included in the 19 SNP GS (and the 21 SNP GS) are shown in Table 1.

Statistical analysis

Use of an earlier version of the 19 SNP GS in NPHSII has been published (26). We updated the weightings to the beta-coefficients determined in the CARDIoGRAMplusC4D meta-analysis as previously described (26). Briefly, unweighted gene scores were determined by simply counting the number of risk alleles present for each SNP for each individual, while the weighted scores were calculated for each individual by multiplying the number of risk alleles carried for each SNP by the published effect size for that SNP and summing the values for all SNPs to obtain the individual's gene score. Numeric variables were compared using t-tests and categorical variables were compared using chi-squared tests. Regression was used to assess the relationship between the GS and CHD and CRFs. Calibration of the combined (CRF plus GS) risk scores was assessed with the Hosmer-Lemeshow test (using ten

degrees of freedom). The ability of the risk scores to discriminate between those who did and did not have an event was assessed the area under the receiver operator characteristic curve (AUC). values were compared using DeLong's test. Reclassification of individuals in different risk categories with the addition of the GS to a CRF risk score was assessed using the net reclassification index (NRI). Where imputation was used, this was performed by assigning the missing genotyping as the mean number of alleles for that particular SNP, only in those NPHSII participants with a single missing genotype. A p-value<0.05 was considered to be statistically significant.

Results:

Baseline Characteristics

The baseline characteristics of the NPHSII cohort are presented in Table 2. As expected, the men who developed CHD during ten-year follow-up were older, had higher BMI, higher systolic blood pressure, higher total cholesterol, low density lipoprotein (LDL) cholesterol, C-reactive protein (CRP) and fibrinogen and a higher proportion were smokers at baseline. Furthermore, those who developed CHD had a higher ten-year CHD risk as calculated using the Framingham risk score and those who developed CVD had a higher ten-year CVD risk as calculated using the QRISK2 score.

Association of GS with CHD

The mean GS was higher in those who went to develop CHD and CVD during the ten-year follow-up, (as shown in Table 3). Both the un-weighted and weighted GSs were associated with CHD after adjustment for age (OR=1.39 per standard deviation (sd) of GS (95% confidence intervals (Cis) 1.14-1.70 $p=1.37 \times 10^{-3}$) and OR=1.46 per sd of GS (95% CIs 1.04-1.80 $p=2.39 \times 10^{-4}$ respectively), with the weighted score showing a stronger association. This association remained only for the weighted score when the model was adjusted for the CRFs age, cholesterol, high density lipoprotein (HDL)-cholesterol, systolic blood pressure, smoking and family history, OR=1.31 per sd of GS (95% 1.03-1.68), p=0.03.

Association of GS with CRFs for CHD

The 19 SNP GS was associated with four lipid traits, cholesterol, LDL-cholesterol, apolipoprotein-B, and lipoprotein (a) (all $p<0.05$) as shown in Table 3A. It was also associated with family history of CHD (p=0.03), with 39% of these of the in top quartile of gene score having a family history of CHD compared to 31% in the bottom quartile.

The GS in CHD Risk Prediction

We assessed two CRF scores for CHD in NPHSII – the Framingham score and QRISK2. While QRISK2 was well calibrated (p=0.10, Figure 1B), the Framingham score generally underestimated risk resulting in poor calibration ($p=4.21 \times 10^{-5}$, Figure 1A). Therefore, further analysis was performed with QRISK2 only. Complete data (ten-year CVD outcome, genotyping and QRISK2 score) was available for 1213 NPHSII participants for QRISK2 plus 19 SNP GS. This combined QRISK2 plus 19 SNP GS score remained well calibrated (p=0.17, Figure 1B) and showed an improvement in discrimination compared to QRISK2 alone (AUC 0.68 v 0.70 p=0.02, Figure 3). The addition of the 19 SNP GS to QRISK2 also resulted in improved risk classification in the group who developed CHD, giving a positive NRI (NRI=0.07, p=0.04, Table 4).

Addition of extra SNPs

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 (p>0.01) (Table 1). We then selected seven SNPs from those robustly associated with CHD in the CARDIoGRAMplusC4D. The SNPs were ranked according to risk allele frequency multiplied by the beta-coefficient (Supplementary Table 1). The top seven SNPs, discounting those in loci already included in the GS (Table 1), were then genotyped in NPHSII and added into the GS to create a 21 SNP GS.

The weighted 21 SNP GS was higher in those who went on develop CHD (3.04 (n=862) v 3.16 (n=83), $p=2.6x10^{-4}$) and CVD (3.05 (n=855) v 3.15 (n=99), $p=4.41x10^{-4}$). The gene score was associated with CHD OR=1.40 per sd of GS (1.15-1.69) $p=7.31x10^{-4}$ and remained so after adjustment for age, cholesterol, HDL-C, smoking, systolic blood pressure and family history of CHD, OR=1.59 per sd of GS (1.22-2.10) $p=8.40x10^{-4}$. As with the 19 SNP GS, the 21 SNP GS was associated with lipid traits, cholesterol, LDL-cholesterol, and lipoprotein (a) (all p<0.05) as shown in Table 3B.

When the 21 SNP GS was combined with QRISK2, this combined score remained well-calibrated (p=0.11) and showed improved discrimination compared to QRISK2 alone (AUC 0.66 v 0.69, p=0.04, Figure 2). However, while the NRI was positive, it was not statistically significant (NRI=0.08, p=0.10). One possibility for this could be loss of power due to sample drop-out, since only 954 participants had genotype data for all 21 SNPs as well as QRISK2 and follow-up data. To increase the number of participants that could be included in the analysis and as the data was considered to be missing at random, the missing genotype for those with only one missing genotype were imputed and the 21 SNP GS calculated.

In the imputed data set, the gene score was similarly associated with CHD (OR=1.38 per sd of GS (1.16-1.63) $p=1.31x10^{-4}$) and remained so after adjustment for age, cholesterol, HDL-C, smoking, systolic blood pressure and family history of CHD, OR=1.39 per sd of GS (1.14-1.71) $p=1.42x10^{-3}$. There was a significant association between the imputed 21 SNP GS and QRISK2 (p=0.03), which was not the case with the un-imputed data set, (again probably reflecting the lower power in the unimputed data set). However, when the combined QRISK2 plus imputed 21 SNP GS score was calculated using this data (n=1736) the score was poorly calibrated (p=0.03) and did not show an improvement in discrimination (AUC 0.69 v 0.71, p=0.55) or reclassification (NRI=0.05, p=0.10).

Discussion

In this study we assessed the relationship between a 19 SNP GS and CHD in a cohort of middle-aged men from the UK. The SNPs were originally been chosen from meta-analyses of candidate gene studies and early GWASs (23). In 2013 the CARDIoGRAMplusC4D consortium published meta-analyses of GWAS results based on >130,000 controls and >60,000 cases (16) which should provide a much more robust estimate of the effect size pertaining to each CHD risk loci than was previously available. Therefore the weightings used in the 19 SNP GS were updated to those effect sizes determined in the CARDIoGRAMplusC4D meta-analysis. A combined risk score of QRISK2 plus the updated 19 SNP GS was found to have good calibration and to show improved discrimination between those who did and did not develop CVD in follow-up and improved reclassification in those who developed CVD, compared to the QRISK2 score alone. Therefore, our results including this GS in an estimation of CHD risk (along with QRISK2) could have clinical utility in the UK population. This result is likely to be relevant to subjects of similar ethnic origin such as white Caucasian populations in North America, Austalasia, and Europe, but extension to other ethnic groups requires additional study.

Having improved the performance of the 19 SNP GS by updating the weightings used in the calculation of the GS, we then assessed a 21 SNP GS created from the 19 SNP GS. The five SNPs with the weakest evidence of an association with CHD in the CARDIoGRAMplusC4D meta-analysis removed and replaced with the top seven ranked variants in loci not already included. Of the seven added SNPs three have been found to be associated with blood pressure traits (GUCY1A3 (enzyme catalysing conversion of GTP to GMP, activated by nitric oxide) rs7692387, CYP17A1 (enzyme of the cytochrome P450 superfamily, involved in steroid biosynthesis) rs12413409, ZC3CH1 (protein involved in the regulation of mitosis) rs11556924) and one with lipid traits (LDLR (encodes the LDL receptor) rs1122608)(16).It was expected that the addition of these robustly associated loci would more accurately reflect an individual's genetic risk of CHD and that further improve the performance of the combined QRISK2 plus GS risk score compared to the QRISK2 score alone. This appeared to be the case, although, possibly because of the reduced numbers available with complete genotype data, the NRI, which increased from 0.07 to 0.08, was not statistically significant. We expected that with increased numbers achieved by imputation of missing SNP data we would detect a statistically significant positive NRI and maintain good calibration while showing improved discrimination compared to QRISK2 alone. However, using the imputed data set, the combined QRISK2 plus imputed 21 SNP GS score showed poor calibration and no improvement in discrimination or NRI compared to QRISK2 alone. Further analysis found a statistically significant association between the 21 SNP GS and QRISK2 in the imputed data set, which was not observed in smaller un-imputed data set.

The results also show that QRISK2 was better at predicating cardiovascular outcome in NPHSII compared to the Framingham score, with the Framingham score overestimating risk in NPHSII. 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 (12). 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 (3, 12, 27).

The improved performance of the 19 SNP GS with the updated weightings (as detailed in (23)) demonstrates that the effect sizes derived from the CARDIoGRAMplusC4D analysis more accurately reflect the impact of the SNPs CHD risk. All of the updated weightings were lower, indicating that the original effect sizes were inflated. This is a common problem in genetic studies (28). However, 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 Rotterdam study and UCLEB consortium, where the gene score was more strongly associated with prevalent rather than incident disease (17), (18)). 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 GS (19). This issue is likely to be more pertinent for the CARDIoGRAMplusC4D SNPs whereas the majority of SNPs included in the 19 SNP GS have a clear mechanism of action to impact CHD and rather than purely CHD survival. This may partly explain the relatively strong performance of the updated 19 SNP GSs in NPHSII compared to the relatively poor performance of the CARDIoGRAMplusC4D GSs in much larger studies. Ultimately, a large-scale well-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 provide the weights for the GS SNPs and this should improve its performance.

It not surprising that the 19 SNP GS is associated with a number of lipid traits (cholesterol measures, apolipoprotein-B, lipoprotein (a)) given that a number of SNPs in the GS are located in genes encoded proteins involved in lipid metabolism. The 19 SNP GS was associated with CHD even after adjustment for total cholesterol indicating that a higher 19 SNP GS value can reflect lifelong genetically raised total cholesterol, which will confer a higher risk of CHD than can be reflected by a single point measurement of total cholesterol in later-life. Furthermore, the maintenance of the association between the 19 SNP GS and CHD despite the inclusion of the family history of CHD in the model suggests that these two related measures can provide information concerning different aspects of CHD risk.

This work has a number of limitations. One aim of our work is to identify a minimum SNP data set that will have clinical utility in CHD risk stratification. With more than 53 CHD SNPs now identified we chose to rank SNPs by the product of their European risk allele frequency and reported odds ratio, and examined the improvement achieved using only the top seven. 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, as demonstrated in the study using all 53 SNPs (18). A second limitation is that all of the participants of NPHSII are male and with a mean age of 56 years represent a group where CHD is highly prevalent and where intervention to reduce future risk is recommended. It is known that the pathogenesis of atherosclerosis is different between the sexes (29), but there is no evidence to suggest the risk variant effect sizes differ between men and women, and a subgroup analysis performed as part of the CARDIoGRAMplusC4D meta-analysis, found no trend for different odds ratios in either sex, but it would be ideal to test the GSs in a data set with both sexes. Another issue is that QRISK2 is updated annually and we did not have access to the most recent version, although the CRFs included are the same. Moreover, our cohort was recruited in 1989 and therefore the cardiovascular risk profile of this group may differ from that of the UK population now. Indeed, one of the reasons suggested for the overestimation of CHD risk by Framingham score is that the data it was derived from was collected at the time of peak cardiovascular risk in that community. However, this time-lag is inevitable in a prospective study and ultimately the benefits (particularly in minimising bias) outweigh the disadvantages.

The clinical utility of the GS described here depends on the context, as pointed out recently for T2D (30). 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 (31). 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.

References:

1. Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. Primary prevention of coronary heart disease in women through diet and lifestyle. New Engl J Med. 2000;343(1):16-22.
2. Stamler J, Dyer AR, Shekelle RB, Neaton J, Stamler R. Relationship of baseline major risk factors to coronary and all-cause mortality, and to longevity: findings from long-term follow-up of Chicago cohorts. Cardiology. 1993;82(2-3):191-222.
3. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. 1998;97(18):1837-47.
4. 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.
5. Simons LA, Simons J, Friedlander Y, McCallum J, Palaniappan L. Risk functions for prediction of cardiovascular disease in elderly Australians: the Dubbo Study. Med J Aust. 2003;178(3):113-6.
6. 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.
7. 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.
8. 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.
9. 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.
10. Lipid Modification: Cardiovascular Risk Assessment and the Modification of Blood Lipids for the Primary and Secondary Prevention of Cardiovascular Disease. National Institute for Health and Clinical Excellence: Guidance. London2014.
11. Rose G. Strategy of prevention: lessons from cardiovascular disease. Br Med J (Clin Res Ed). 1981;282(6279):1847-51.
12. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ. 2010;340:c2442.
13. Madjid M, Willerson JT. Inflammatory markers in coronary heart disease. Br Med Bull. 2011;100:23-38.
14. Kamstrup PR, Tybjaerg-Hansen A, Nordestgaard BG. Extreme lipoprotein(a) levels and improved cardiovascular risk prediction. J Am Coll Cardiol. 2013;61(11):1146-56.
15. Joint British Societies' consensus recommendations for the prevention of cardiovascular disease (JBS3). Heart. 2014;100 Suppl 2:ii1-ii67.
16. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. 2013;45(1):25-U52.
17. de Vries PS, Kavousi M, Ligthart S, Uitterlinden AG, Hofman A, Franco OH, et al. 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;44(2):682-8.
18. Morris RW, Cooper JA, Shah T, Wong A, Drenos F, Engmann J, et al. Marginal role for 53 common genetic variants in cardiovascular disease prediction. Heart. 2016;102(20):1640.
19. Ganna A, Magnusson PK, Pedersen NL, de Faire U, Reilly M, Arnlov J, et al. Multilocus genetic risk scores for coronary heart disease prediction. Atertio Thromb Vasc Biol. 2013;33(9):2267-72.
20. 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.

21. Humphries SE, Cooper JA, Talmud PJ, Miller GJ. Candidate gene genotypes, along with conventional risk factor assessment, improve estimation of coronary heart disease risk in healthy UK men. Clin Chem. 2007;53(1):8-16.
22. Talmud PJ, Cooper JA, Palmen J, Lovering R, Drenos F, Hingorani AD, et al. Chromosome 9p21.3 coronary heart disease locus genotype and prospective risk of CHD in healthy middle-aged men. Clin Chem. 2008;54(3):467-74.
23. Beaney KE, Cooper JA, Ullah Shahid S, Ahmed W, Qamar R, Drenos F, et al. 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;10(7):e0130754.
24. Miller GJ, Bauer KA, Barzegar S, Foley AJ, Mitchell JP, Cooper JA, et al. The effects of quality and timing of venepuncture on markers of blood coagulation in healthy middle-aged men. Thromb Haemost. 1995;73(1):82-6.
25. Hawe E, Talmud PJ, Miller GJ, Humphries SE. Family history is a coronary heart disease risk factor in the Second Northwick Park Heart Study. Ann Hum Genet. 2003;67(Pt 2):97-106.
26. Larifla L, Beaney KE, Foucan L, Bangou J, Michel CT, Martino J, et al. Influence of Genetic Risk Factors on Coronary Heart Disease Occurrence in Afro-Caribbeans. Can J Cardiol. 2016.
27. 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.
28. Kraft P. Curses--winner's and otherwise--in genetic epidemiology. Epidemiology. 2008;19(5):649-51; discussion 57-8.
29. Appelman Y, van Rijn BB, Ten Haaf ME, Boersma E, Peters SA. Sex differences in cardiovascular risk factors and disease prevention. Atherosclerosis. 2015;241(1):211-8.
30. Drenos F. The use of genetic information in the prediction of Type 2 diabetes. Pers Med. 2015;12(5):483-96.
31. 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.
32. Bennet AM, Di Angelantonio E, Ye Z, Wensley F, Dahlin A, Ahlbom A, et al. Association of apolipoprotein E genotypes with lipid levels and coronary risk. JAMA. 2007;298(11):1300-11.
33. Benn M, Nordestgaard BG, Grande P, Schnohr P, Tybjaerg-Hansen A. PCSK9 R46L, low-density lipoprotein cholesterol levels, and risk of ischemic heart disease: 3 independent studies and meta-analyses. J Am Coll Cardiol. 2010;55(25):2833-42.

Funding:

KB was funded by a MRC CASE studentship with Randox Laboratories (1270920) and JC and SEH were supported by the BHF (BHF PG08/008). FD was supported by the UK Medical Research Council (MC_UU_12013/1-9).

Conflict of Interest:

KB was funded by a MRC CASE studentship with Randox Laboratories (1270920). SEH is the medical director and minority shareholder of the University College London start-up coronary heart disease risk genetic testing company StoreGene, and has received honoraria for speaking at educational meetings with a pharmaceutical sponsor, but has donated all personal payments to various medical charities.

Tables:

Table 1: SNPs included in the 19 SNP GS and the 21 SNP GS

SNPs included in the 19 SNP GS								
Gene/Locus	SNP	Risk Allele	OR	OR in original score	Frequency	p-value*	Source	Call Rate
*APOE**	rs7412	C	1.25	0.80**	0.87	-	(32)	97 %
*APOE**	rs429358	C	1.06	1.06	0.26	-	(32)	97 %
*PCSK9**	rs11591147	G	1.39	1.43	0.99	-	(33)	85 %
CDKN2A/9p21*	rs10757274[a]	G	1.23	1.29	0.47	$1.39x10^{-52}$	Cplus4D	97 %
*SORT1**	rs599839	A	1.11	1.19	0.77	$3.8x10^{-15}$	Cplus4D	96 %
*LPA**	rs10455872	G	1.32	1.70	0.06	$3.80x10^{-13}$	CG GWAS	93 %
*MIA3**	rs17465637	C	1.14	1.14	0.74	$1.36x10^{-8}$	CG GWAS	92 %
*MRAS**	rs9818870	T	1.07	1.15	0.14	$2.62x10^{-9}$	Cplus4D	94 %
*CXCL12**	rs1746048[b]	C	1.07	1.17	0.83	$1.79x10^{-8}$	Cplus4D	96 %
*LPL**	rs328	C	1.09	1.25	0.91	$2.34x10^{-4}$	CG GWAS	95 %
*DAB2IP**	rs7025486	A	1.04	1.16	0.29	$2.14x10^{-3}$	Cplus4D	97 %
*LPL**	rs1801177[e]	A	1.10	1.33	0.06	$4.04x10^{-4}$	Cplus4D	87 %
*LPA**	rs3798220	C	1.28	1.92	0.01	$4.90x10^{-5}$	Cplus4D	96 %
*APOA5**	rs662799	G	1.05	1.19	0.06	0.01	Cplus4D	89 %
CETP	rs708272[d]	C	1.04	1.28	0.56	0.04	CG GWAS	90 %
ACE	rs4341[c]	G	1.01	1.22	0.52	0.43	CG GWAS	98 %
APOB	rs1042031	A	1.01	1.73	0.18	0.80	Cplus4D	96 %
NOS3	rs1799983	G	1.00	1.31	0.67	0.90	CG GWAS	85 %
SMAD3	rs17228212	C	1.01	1.21	0.31	0.94	Cplus4D	96 %

Additional SNPs included in the 21 SNP GS						
Gene/Locus	SNP	Risk Allele	OR	Frequency	Source	Call Rate
GUCY1A3	rs7692387	G	1.13	0.81	Cplus4D	91 %
PPAP2B	rs17114036	T	1.11	0.91	Cplus4D	91 %
CYP17A1-CNNM2-NT5C2	rs12413409	G	1.10	0.89	Cplus4D	91 %
LDLR	rs1122608	G	1.10	0.76	Cplus4D	90 %
COL4A1-COL4A2	rs9515203	T	1.08	0.74	Cplus4D	92 %

PHACTR1	rs9369640[f]	A	1.09	0.65	Cplus4D	98 %
ZC3HC1	Rs11556924	C	1.09	0.65	Cplus4D	89 %

* Included in both the 19 SNP GS and the 21 SNP GS. [a]Weighting for rs1333049 (r^2=0.88). [b]Weighting for rs501120 used (r^2=0.97). [c]Weighting for rs4343 (r^2=0.96). [d]Weighting for rs711752 (r^2=1). [e]Weighting for rs7016529 (r^2=1).[f] tagging rs12526453 (r^2=0.90). All r^2 values calculated from 1000 Genomes phase 1 EUR data. Data on coronary artery disease / myocardial infarction have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG. **In the original score, the protective allele was included rather than the risk allele. OR= odds ratio. Cplus4D = CARDIoGRAMplusC4D meta-analysis. CG GWAS=CARDIoGRAM GWAS.

Table 2 : Baseline characteristics in NPHSII for those who did and did not go on to develop CHD during ten-year follow-up

Trait	NPHSII No CHD (n=2491)	NPHSII CHD (n=284)	p-value
Age (years)	55.91 (3.42)	56.64 (3.60)	4.12×10^{-3}
Sex (% Male)	100 %	100 %	-
Smoking % (n)	25 % (n=567)	38 % (n=50)	1.90×10^{-3}
Family history of CHD % (n)	34 % (n=731)	53 % (n=66)	5.68×10^{-3}
BMI (kg/m^2)	26.38 (3.42)	27.19 (3.44)	9.61×10^{-4}
Systolic Blood Pressure (mmHg)	137.00 (18.59)	144.09 (20.10)	9.68×10^{-7}
Total Cholesterol (mmol/l)	5.71 (1.01)	6.13 (1.05)	4.79×10^{-8}
LDL-cholesterol (mmol/l)	3.07 (1.00)	3.48 (0.97)	2.66×10^{-7}
HDL-cholesterol (mmol/l)	1.72 (0.59)	1.57 (0.53)	2.60×10^{-4}
Triglycerides (mmol/l)*	1.78 (0.93)	2.02 (1.02)	3.13×10^{-3}
Apolipoprotein-B (g/l)	0.89 (0.26)	0.97 (0.24)	2.06×10^{-4}
Apolipoprotein-A1 (g/l)	1.63 (0.32)	1.61 (0.28)	0.22
Lipoprotein (a)* (g/l)	0.09 (0.03-0.30)	0.13 (0.03-0.30)	0.12
C-reactive Protein* (mg/l)	267.83 (50.16)	284.65 (49.21)	9.13×10^{-5}
Fibrinogen *(g/l)	2.75 (3.19)	4.47 (5.18)	3.21×10^{-5}
Framingham ten-year CHD risk	0.12 (0.07-0.15)	0.17 (0.09-0.21)	4.33×1011^{-4}
QRISK2 ten-year CVD risk**	0.09 (0.07-0.13)	0.13 (0.09-0.17)	1.93×10^{-14}

All variables are presented as the mean plus standard deviation except where indicated. Categorical variables were compared using chi-squared tests and continuous variables were compared using Welch's t-tests, apart from the Lipoprotein (a), Framingham and QRISK2 risk scores (shown as proportions) which were compared using Mann Whitney tests (the median and interquartile range are given). *Values were log transformed and geometric mean and approximate standard deviation are shown.**QRISK2 values shown are for those who did and did not go on to develop CVD.

Table 3A: Association between 19 SNP weighted gene score and CRFs for CHD

Trait	Tertile 1 (n=454)	Tertile 2 (n=453)	Tertile 3 (n=453)	p-value
Age (years)	56.5 (3.48)	56.21 (3.51)	56.43 (3.59)	0.73
Smoking % (n)	26 % (n=120)	28 % (n=129)	30 % (n=138)	0.18
Family history of CHD % (n)	31 % (n=133)	37 % (n=155)	39 % (n=165)	0.03
BMI (kg/m^2)*	26.30 (3.43)	26.78 (3.49)	26.45 (3.41)	0.51
Systolic blood pressure (mmHg)	138.85 (19.70)	139.97 (19.25)	139.33 (190.34)	0.71
Total Cholesterol (mmol/l)	5.64 (1.00)	5.81 (0.95)	5.86 (1.01)	$5.74x10^{-4}$
LDL-cholesterol (mmol/l)	3.02 (1.02)	3.22 (0.94)	3.28 (0.99)	$3.88x10^{-4}$
HDL-cholesterol (mmol/l)	1.69 (0.59)	1.65 (0.54)	1.63 (0.61)	0.15
Triglyceride *(mmol/l)	1.78 (0.93)	1.83 (0.95)	1.84 (1.00)	0.10
Apolipoprotein-B (g/l)	0.89 (0.28)	0.91 (0.26)	0.93 (0.26)	0.01
Apolipoprotein-A1 g/l)	1.61 (0.33)	1.60 (0.29)	1.59 (0.33)	0.50
Lipoprotein (a)** (g/l)	0.08 (0.02-0.18)	0.08 (0.03-0.20)	0.19 (0.04-0.42)	$3.68x10^{-9}$
C-reactive Protein*(mg/l)	3.41 (3.94)	3.28 (3.71)	3.16 (3.67)	0.34
Fibrinogen *(g/l)	2.73 (5.12)	2.73 (5.07)	2.72 (5.18)	0.69
QRISK*	0.09 (0.07-0.14)	0.10 (0.08-0.14)	0.11 (0.08-0.15)	0.05

Gene score was divided into tertiles and regression analysis performed. The mean and standard deviation are shown for each tertile. *Variable was log transformed and the geometric mean and approximate standard deviation are shown, expect for QRISK2 where the median and interquartile range are shown (proportions given). **Median and interquartile range shown and tobit regression on ln(lpa+1). CHD = coronary heart disease. CRFs= conventional risk factors.

Table 3B: Association between 21 SNP weighted gene score and CRFs for CHD

Trait	Tertile 1 (n=353)	Tertile 2 (n=343)	Tertile 3 (n=342)	p-value
Age (years)	56.20 (3.53)	56.18 (3.43)	56.47 (3.62)	0.32
Smoking % (n)	25 % (n=89)	27 % (n=92)	30 % (n=102)	0.17
Family history of CHD % (n)	33% (n=107)	36 % (n=114)	37 % (n=118)	0.27
BMI (kg/m^2)*	26.12 (3.38)	26.35 (3.32)	26.16 (3.35)	0.87
Systolic blood pressure (mmHg)	138.60 (19.64)	138.50 (18.69)	139.95 (19.22)	0.36
Total Cholesterol (mmol/l)	5.57 (0.97)	5.75 (0.97)	5.83 (1.02)	$4.66x10^{-4}$
LDL-cholesterol (mmol/l)	2.93 (0.95)	3.13 (0.94)	3.18 (1.00)	$1.28x10^{-3}$
HDL-cholesterol (mmol/l)	1.70 (0.58)	1.69 (0.61)	1.69 (0.61)	0.81
Triglyceride *(mmol/l)	1.58 (0.63)	1.57 (0.64)	1.55 (0.76)	0.72
Apolipoprotein-B (g/l)	0.88 (0.27)	0.89 (0.25)	0.91 (0.25)	0.13
Apolipoprotein-A1 g/l)	1.62 (0.33)	1.62 (0.32)	1.63 (0.32)	0.65
Lipoprotein (a)** (g/l)	0.07 (0.02-0.20)	0.07 (0.02-0.24)	0.15 (0.04-0.36)	$6.38x10^{-6}$
C-reactive Protein*(mg/l)	3.11 (3.54)	3.15 (3.61)	2.78 (3.07)	0.23
Fibrinogen *(mg/dl)	2.73 (5.14)	2.68 (4.82)	2.72 (4.90)	0.72
QRISK*	0.10 (0.07-0.14)	0.10 (0.07-0.14)	0.11 (0.08-0.14)	0.06

Gene score was divided into tertiles and regression analysis performed. The mean and standard deviation are shown for each tertile. *Variable was log transformed and the geometric mean and approximate standard deviation are shown, expect for QRISK2 where the median and interquartile range are shown (proportions given).**Median and interquartile range shown and tobit regression on ln(lpa+1). CHD = coronary heart disease. CRFs= conventional risk factors.

Table 4: Reclassification of NPHII participants with the addition of the gene scores to QRISK2

Risk Score	Reclassified at lower risk	No change in risk classification	Reclassified at higher risk	NRI (95 % CIs)	p-value
QRISK2 + 19 SNP GS					
No CHD	51	945	84	0.07 (0.002-0.13)	0.04
CHD	3	114	16		
Event rate	5.56 %	10.76 %	16.00 %		
QRISK2 + 21 SNP GS					
No CHD	41	738	76	0.08 (-0.01-0.17)	0.10
CHD	4	79	16		
Event rate	8.89 %	9.67 %	17.39 %		

CRFs= conventional risk factors. NRI=net reclassification index.

Figure Legends: Figure 1A: Calibration plot of observed probabilities and predicted probabilites of CHD, by decile of risk score for the Framingham score alone and the Framingam score plus the 19 SNP GS
CHD=Coronary heart disease. GS=Gene score. SNP= single nucleotide polymorphism.

Figure 1B: Calibration plot of observed probabilities and predicted probabilites of CHD, by decile of risk score for the QRISK2 score alone and the QRISK2 score plus the19 SNP GS
CHD=Coronary heart disease. GS=Gene score. SNP= single nucleotide polymorphism.

Figure 2: ROC curves for CRF score alone and with the addition of the 19 SNP GS and 21 SNP GS
CRF=conventional risk factor. GS=gene score.

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

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

OR=odds ratio. RAF=risk allele frequency. *Additional SNPs selected for inclusion in the 21 SNP GS. [+]SNPs already included in the GS (or in linkage disequilibrium with SNPs that were).