Genetic risk analysis of coronary artery disease in the Pakistani subjectsusing a genetic risk score of 21 variants

SaleemUllah Shahid[1]*, Shabana[1], Jackie A Cooper[2], Katherine E Beaney[2], Kawah Li[2], Abdul Rehman[1], Steve E Humphries[2]

1 Department of Microbiology and Molecular Genetics, University of the Punjab, Lahore, Pakistan 54590.

2Centre for Cardiovascular Genetics, British Heart Foundation Laboratories, University College London, London, WC1E6JF, UK.

*Corresponding Author

SaleemUllahShahid

00923007592547

saleemullahshahid@hotmail.com

Running title: Genetic risk score of coronary artery disease in Pakistan

Number of Tables: 4

Number of Figures: 1

Number of Supplementary Tables: 8

Number of Supplementary Figures: 2

Abstract

Back ground and Aims: Conventional risk factors like age, gender, blood lipids, hypertension and smoking have been the basis of coronary artery disease (CAD) risk prediction algorithms, but provide only modest discrimination. A genetic risk score (GRS) may provide improved discrimination over and above conventional risk factors alone. The current study analysed the genetic risk of CAD in Pakistani subjects using a GRS of 21 loci in 18 genes and examined whether its association with blood lipids in this cohort.

Methods: 625 subjects were genotyped for the variants, *NOS3* rs1799983, *SMAD3* rs17228212, *APOB*rs1042031, *LPA*rs3798220, *LPA* rs10455872, *SORT1*rs646776, *APOE* rs429358, *GLUL* rs10911021 and *FTO* rs9939609 (by TaqMan) and *MIA3* rs17465637,*CDKN2A* rs10757274, *DAB2IP* rs7025486, *CXCL12* rs1746048, *ACE* rs4341, *APOA5* rs662799, *CETP* rs708272, *MRAS* rs9818870, *LPL* rs328,*LPL* rs1801177, *PCSK9* rs11591147and *APOE* rs7412 (by KASPar technique).

Results: Individually, risk allele frequencies were not significantly higher in cases than controls (p>0.05) except for *APOB* rs1042031 and *FTO* rs9939609 (p=0.007 and 0.003 respectively), and did not associate with CAD except rs1042031 and rs993969 (p=0.01 and 0.009 respectively). However, the GRS of 21 SNPs was significantly higher in cases than controls (17.53±2.52 vs16.64±2.44, p<0.001) and was associated with CAD risk. CAD risk in the top quintile of GRS was 2.96 (95% CI 1.71-5.13). Atherogenic blood lipid levels showed significant positive association with GRS.

Conclusion: The GRS was quantitatively associated with d CAD risk and showed association with blood lipid levels, suggesting that the mechanism of these variants is likely to be in part at least through creating an atherogenic lipid profile in subjects carrying high numbers of risk alleles.

Key words: Coronary artery disease, Genetic risk score, conventional risk factors

Introduction

Coronary artery disease (CAD) is a chronic disorder progressing silently and usually has established to an advance stage by the time symptoms start appearing. Despite all measures, CAD remains the single largest killer worldwide. In high income countries, the CAD mortality rate has declined since 1980 and has shifted to an older age group, whereas, middle and low income countries bear three quarters of the global CAD burden. South Asians are at a greater risk and the prevalence is 50% to 300% higher than rest of the world (Enas & Kannan, 2005). The prevalence of CAD is even higher in Pakistan (Jafar, Jafary, Jessani, & Chaturvedi, 2005) with more than 30% of the population above 45 years of age being affected by the disease (Gaziano, Bitton, Anand, Abrahams-Gessel, & Murphy, 2010). The disease burden has almost doubled in urban Karachi since 1970 (Aziz, Uddin, Faruqui, Patel, & Jaffery, 2012). According to latest WHO reports, cardiovascular diseases (CVD) are among the biggest non-communicable killers in Pakistan and CAD represents a major type of CVD (http://www.who.int/countries/pak/en/).

CAD is a multifactorial disorder and arises from an interaction between environmental and genetic factors. The identifiable environmental risk factors have been identified in about 80% of CAD cases (Alwan, 2011). Most of the CAD risk factors are modifiable therefore to target life style changes or for drug intervention, those who are at most risk of developing disease should be properly identified. The conventional CAD risk factors (CRFs) like age, gender, blood lipids, smoking, blood pressure and diabetes have been the basis of CAD risk prediction algorithms developed by many consortia. These risk prediction algorithms include the Framingham risk score (Wilson et al., 1998), the Prospective Cardiovascular Munster Heart Study (PROCAM) (Assmann, Cullen, & Schulte, 2002), the Systematic Coronary Risk Evaluation (SCORE) system (Conroy et al., 2003), the Reynolds risk score (Ridker, Buring, Rifai, & Cook, 2007) and QRISK2 (Hippisley-Cox et al., 2008). These CRF algorithms calculate 10 years CAD risk and the individuals are then classified according to their risk. The high risk category individuals qualify for the preventive treatment (statin), and until recently, the cut-off for statin treatment has been set at 20% 10 years CAD risk (Wood et al., 2005). Lower cut-off value has been proposed in both UK and USA (10% and 7.5% respectively). The use of CAD risk prediction scores has increased the average life time of CAD patients by three years in USA (Lenfant, 2003). However, the risk assessment using CRFs provide only modest discrimination and do not fully explain the underlying risk (Wang et al., 2006). These scores lack accuracy and may overestimate risk in low risk subjects or underestimate risk in subjects at high risk (Brindle, Beswick, Fahey, & Ebrahim, 2006; Wilson et al., 1998).

Almost 15-20% cases who developed the disease in their later life were underestimated (Thanassoulis & Vasan, 2010) with most of the cases occurring in intermediate and low risk subjects (Collins & Altman, 2009; Cooper, Miller, & Humphries, 2005).

The variability in disease susceptibility in individuals exposed to similar environmental factors and having almost same CRFs can be attributed to the genetic variations (Stranger et al., 2007).Genetic testing may improve discrimination over and above the CRFs. Family history of early heart disease has long been a known CAD risk factor and heritability of CAD has been estimated to be more than 40% (Peden & Farrall, 2011). Historically, the genetic risk of a disease was assessed through the presence of the disease in the proband's relatives and the genetic component was described as heritability estimate. Then the 'candidate gene' approach was used, where common variants were determined in the genes regulating biochemical pathways of disease pathogenesis (Wray & Goddard, 2010). Since 2007, additional genes associated with CAD have been identified by Genome Wide Association studies (GWAS) (I. K. C. Consortium, 2011). Single nucleotide polymorphisms (SNPs) can be used as markers of genetic variability. The CAD associated SNPs are common in the general population with a minimal to moderate relative risk. Most of them are located in non-coding DNA region implying that they influence by regulating the expression of upstream or downstream genes. Another striking feature of CAD risk SNPs is that most of them operate independently of known CAD risk factors. This indicates that many unknown pathways involved in development of CAD still need to be explored (Folkersen et al., 2010). However, the risk associated with single SNP is modest, because of the low effect sizes of common variants, and therefore a large number of SNPs need to be genotyped for the genetic analysis of CAD like complex disease.

A genetic risk score (GRS) of a disease is calculated by summing up the number of risk alleles at all the loci included in the genetic risk of that disease. The GRS is a multi-locus profile used to transpose the discoveries from candidate gene studies and GWASs into population health tools (Belsky et al., 2013; Fava et al., 2014). A GRS summarizes the effect of multiple variants in a quantitative manner and hence is superior over the predictive power of a single SNP. The use of GRS information into risk prediction of CAD can bridge up the genomic research with more applied clinical practice. Different researchers have used varying number and types of loci for inclusion in CAD genetic risk scoring, the number ranging from less than 10 to more than 100 (Anderson et al., 2010; Paynter et al., 2010; Qi et al., 2011; Ripatti et al., 2010; Thanassoulis et al., 2012).

The majority of genetic studies and GWAS have been conducted on European/Caucasian people. It remained a routine practice to transpose the results obtained from such studies conducted in developed countries to the rest of the world, but there remains an immense requirement to extend genetic studies to other ethnicities also. The allele frequencies of many common variants vary widely between ethnicities. For example, the association of the 9p21 region with CAD has not been replicated in African Americans (Assimes et al., 2008; McPherson et al., 2007). Similarly, the linkage disequilibrium and effect size of common variants may vary across ethnicities. Moreover, a genetic marker may not be associated with a trait in all ethnicities and in such cases the applicability is limited to only those populations where the genotype to phenotype association is clearly seen (Ioannidis, 2009). The Pakistani population, like other Asian countries is under represented in international genetic studies like HAP MAP or 1000 genomes project. To date the genetic architecture of CAD has not been evaluated properly for this population. A preliminary report of the use of a 19 SNPs GRS in CAD risk analysis in the Pakistani subjects has been published (Beaney et al., 2015), but the study was underpowered to detect the same effect as observed in Europeans. In the current study, we included two additional SNPs to construct a CAD GRS and increased the sample size. We hypothesized that to predict CAD risk; the GRS of 21 SNPs will be superior over single SNPs having small effect size and modest association.

Materials and methods

The study comprised of 405 diagnosed cases of CAD and 220 healthy controls. The selection criteria for the subject recruitment has been described previously (Shahid, Cooper, Rehman, & Humphries, 2016). The CAD cases were recruited from tertiary care hospitals in Lahore during February 2012 to June 2013. These selected subjects had suffered from a non-fatal myocardial infarction, with diagnosis of myocardial infarction made by the consultant cardiologist based on the reports of ECG, cardiac echo, angiography, troponine T/I and clinical history. Only those CAD cases were selected which were recently diagnosed and had not started lipid lowering or antihypertensive drugs therapy. The controls were apparently healthy subjects, not having any family history of CAD. It was taken care that cases and controls represented all the socioeconomic groups. Subjects with obesity (BMI> XXX) were also excluded from the study but not those with Type 2 diabetes because the number of CAD subjects with type 2 diabetes was high and the sample size would have become too small to have adequate power. All participants gave a written informed consent. The study was

approved by the ethics committee, University of the Punjab, Lahore and all the procedures were in compliance with the Helsinki declaration.

Genotyping

The DNA was extracted from whole blood leucocytes using Wizard genomic DNA purification Kit (Promega, USA). The DNA samples were quantified using nanodrop (ND-8000, USA). The concentration of DNA samples was standardized to 1.25ng/ul. The genotyping was carried out in specially designed 384 well plates (Micro Amp). The DNA samples were arrayed by an automated robotic liquid handling system (Biomerk-FX, Beckman Couter). Two high throughput florescent based genotyping techniques, TaqMan and KASPar, were used for genotyping the SNPs. The details of genotyping techniques have been given somewhere else (Shahid, Cooper, Beaney, et al., 2016). The information on SNPs included is provided in supplementary table 1.

The SNPs *NOS3* (rs1799983), *SMAD3* (rs17228212), *APOB* (rs1042031), *LPA* (rs3798220), *LPA* (rs10455872), *SORT1* (rs646776), *APOE* (rs429358), *GLUL* (rs10911021) and *FTO* (rs9939609) were genotyped by TaqMan technique using qPCR master mix (KAPA Biosystems, USA). The SNPs *MIA3* (rs17465637), *CDKN2A* (rs10757274), *DAB2IP* (rs7025486), *CXCL12* (rs1746048), *ACE* (rs4341), *APOA5* (rs662799), *CETP* (rs708272), *MRAS* (rs9818870) *LPL* (rs328), *LPL* (rs1801177), *PCSK9* (rs11591147) and *APOE* (rs7412) were genotyped by KASPar technique with touchdown thermal cycler programme. The SNP *LPL* (rs1801177) was monomorphic in this population but data from this SNP is shown for completeness. The list of primers and probes used for TaqMan and KASPar are given in supplementary tables 2 and 3 respectively. After amplification, the results were analysed on ABI Prism 7900HT (Applied Biosystems/Life Technologies) and the genotypes were called using sequence detection software (SDS), version 2.0. The quality check of genotyping techniques was maintained by the inclusion of non-template controls (NTCs). There were 16 NTCs included in each plate of 384 wells. Only those runs were included in the analysis where none of the NTCs crossed the amplification cut-off line. Only the samples which were clearly clustered were included in the study. While genotyping the variants with very low risk allele frequency like *APOE* rs7412, *LPA* rs3798220, *LPA* rs10455872 and *LPL* rs1801177, known heterozygotes were added to avoid false negative calls. The genotypes were also randomly confirmed by the conventional direct DNA sequencing, and 10-15% of samples from each run were outsourced (source biosciences, UK) for direct sequencing and the results

were always similar to that of TaqMan/KASPar. The list of primers used for direct DNA sequencing is given in supplementary table 4.

Statistical analysis

The results were statistically analysed using statistical package for social sciences (SPSS) IBM, version 22. The continuous variables were compared between cases and controls using independent sample student *t* test. Hardy Weinberg equilibrium was accessed by a χ^2 goodness of fit test. The categorical variables such as risk allele frequencies (RAFs) were compared between cases and controls by χ^2 test. All the analyses were adjusted for age, gender, BMI, hypertensive and diabetic status. Since CAD is a binary variable, the association of the SNPs with CAD was examined using binary logistic regression. The effect of increasing GRS values on CAD was calculated through GRS quintile analysis. The distribution of GRS in cases and controls was compared visually by histograms. The power of the GRS to discriminate between CAD cases and controls was examined by receiver operative curve (ROC) analysis. Blood lipid levels across different number of risk allele in GRS were calculated by one way analysis of variance (ANOVA). The effect size/beta effect which is per risk allele effect of of GRS on lipid levels was calculated by linear regression.

Constructing a GRS

The un-weighted GRS was calculated by simply summing up the number of risk alleles at all the loci included in the study. The risk alleles were considered to be acting in additive manner *i.e.*, each risk allele had equal contribution to the outcome and each risk allele was coded as 1. So the protective homozygous genotype with no risk allele was coded 0, heterozygous individual carrying one risk and one normal allele was coded as 1 and the risk homozygous individual having both risk alleles was coded as 2. In this way the GRS of an individual can range from 0 (no risk allele) to 42 (with all the alleles being risk alleles for 21 loci).

Results

The baseline biochemical and anthropometric parameters of the subjects under study are given in supplementary table 5. The cases were more diabetic and hypertensive, smoking rate was also high in cases than controls. Total cholesterol (TC), triglycerides (TG) and LDL-C were significantly higher whereas, HDL-C was lower in cases than controls. Individually, the RAFs of the studied SNPs were higher in CAD cases compared to controls but the difference

was only statistically significant (p<0.05) for*APOB* rs1042031 and *FTO* rs9939609. The RAFs in cases versus controls, the *p*-values for statistical difference between frequencies and confidence interval (C.I) are shown (supplementary Table 6).

The association of SNPs with CAD was assessed by deriving their odds ratios (OR). The CAD odds of all the SNPs were greater than 1 (except *APOE* rs7412) but were not statistically significant in the studied sample size. Only,*APOB* rs1042031 and *FTO* rs9939609 were significantly associated with CAD (Table 1). However, the GRS of 21 SNPs was significantly higher in CAD cases than controls and was also significantlyassociated with CAD risk (supplementary Table 7).The GRS quintile analysis showed that the increase in GRS was significantly associated with CAD as shown by inter-quintile *p* value for CAD association. Compared to those in the bottom quintile of the score, CAD risk in the top quintile of the GRS was 2.96 (95% CI 1.71-5.13) (Table 2).

The GRS in whole sample set including cases and controls was normally distributed (Supplementary Fig. 1). The GRS histogram for cases exhibited a shift to the right with higher GRS values more prevalent. Comparatively, a left shift of GRS was observed in controls with the lower GRS being more prevalent. The most prevalent GRS in controls was 17 and 40% individuals had this value. In cases, 18 GRS was the most prevalent, present in 65% subjects followed by GRS value of 19 which was present in 60% subjects. Similarly, the upper GRS quintiles were more prevalent in cases and lower quintiles were more prevalent in controls (Fig. 1). A ROC analysis was conducted to estimate whether the gene score had potential to discriminate between cases and controls. The ROC was discriminating between cases and controls and the area under ROC was 0.602 (0.56-0.65) which was statistically significant (p<0.001) (Supplementary Fig. 2).

The mean lipid levels along different GRS values are shown in table 3. There is a significant increase in atherogenic lipids and a decrease in atheroprotective lipids with increase in gene score. The mean TC, LDL-C and TG increased and HDL-C decreased with gene score. The effect size of GRS on TC was 3.7±0.7mg/dl i.e. addition of each risk allele in GRS increased TC by 3.7±0.7mg/dl.The effect size of GRS on LDL-C was 4±0.5mg/dl which was statistically highly significant. Similarly, lower HDL-C values were observed towards higher GRS levels and the effect size of GRS on HDL-C was 2±0.3mg/dl which was a decrease in HDL-C per risk allele held by that individual. Similarly, TG levels increased with increase in GRS and the effect size of each risk allele in GRS on TG was 4.2±1mg/dl (Table 4).

Discussion

The genetics of complex diseases like CAD is an interplay of different factors because the outcome is probabilistic by definition. The statistical parameters which have been used include risk prediction (relative risk, odds ratio, hazard ratio), family analysis (liability, threshold models) and regression (linear/logistic) (Cordell, 2009). The potential validity of a GRS can be examined on the merits of discrimination, risk reclassification and its clinical utility. However, due to modest risk associated with individual variants, low power of discrimination and lack of replication in different ethnicities, the genetic analysis could explain only a small part of heritability, leaving their clinical utility questionable (Carreras-Torres et al., 2013). In the present study, we have studied SNPs at 21 loci to examine their combined effect and utility in genetic risk analysis in the Pakistani population and the combined GRS was significantly higher in cases than controls and was associated with CAD. There was a graded and continual increase in CAD risk with increasing number of CAD risk SNP alleles carried and individuals in the top quintile of the GRS had a CAD risk of 2.96 (95% CI 1.71-5.13). Even though the score distribution overlaps between cases and controls, the GRS is significantly associated with CAD risk and as such can be used as a tool to identify subjects at highest risk for lifestyle or therapeutic interventions.

The approach of using a GRS in CAD risk analysis is relatively new in Pakistani subjects; however, their use is well established in western countries. In the developed countries where CAD CRFs of people are well documented and monitored, CAD risk prediction algorithms based on CRFs are available. The GRS in these subjects can then be examined to check whether the inclusion of genetic risk information is able to improve the risk prediction over and above CRFs. In Pakistani subjects, data on routine CRF monitoring was not available and we used the GRS to examine whether it improves the discrimination power over the use of single SNPs.

These SNPs except *GLUL* rs10911021 and *FTO* rs9939609 were previously genotyped as a group in NPHSII and the 19 SNPs GRS is available for use in CAD risk prediction along with 10 years Framingham risk score in UK (Beaney et al., 2015). The SNPs included in this study were taken from meta-analysis of candidate gene studies (mostly belonging to lipid metabolism genes) or were CAD GWAS hits. All the included SNPs were not in LD even if present in the same gene.

The rationale behind selecting rs9939609 is that it has been reported that the presence of two alleles at the rs9939609 site of the FTO gene increased BMI by about 1 kg/m^2, body mass by 2.3Kg and 1.3-fold higher risk of overweight and obesity in both adults and children. It has been estimated that per unit increase in BMI increase cardiovascular disease morbidity by 8% (Li et al., 2006). However, we found significant association of the risk allele of rs9939609 with CAD in Pakistani subjects independent of BMI (Shahid, Shabana, Rehman, & Hasnain, 2016). The variant rs10911021 is a new locus identified to be associated with diabetes in subjects with coronary heart disease (Qi et al., 2013). Since its identification, only a few studies have investigated its role in different diseases. One study identified this variant to be a predictor of all cause mortality in diabetic subjects (Prudente et al., 2015). This intergenic SNP is approximately 270 Kb from the gene encoding glutamate ammonia ligase (GLUL) enzyme belonging to the glutamine synthase family. It has been found that individuals homozygous for the risk allele (C) have a lower plasma pyroglutamic acid/glutamic acid ratio resulting in impairment of the γ-glutamyl cycle which consequently increases oxidative predisposing diabetic individuals to CHD (Qi et al., 2013). We have confirmed the association of this SNP with CHD in subjects with T2D but not non-diabetic subjects (Beaney et al 2016) but have no further information on the possible risk mechanism of this SNP. .

When tested directly for CAD risk prediction , different GRSs have shown varying results in different studies. In a study using 24 variants in a sample of European ancestry, the authors failed to prove an association between GRS and subclinical atherosclerosis (Hernesniemi et al., 2012). A GRS including 101 variants failed to improve the prediction over and above family history (Paynter et al., 2010). Recently, a GRS including 13 SNPs was reported to be associated with CAD (Mehta & N, 2011) and in another study GRS of 6 lipid metabolic genes improved the discrimination of angiographically proven coronary disease (Anderson et al., 2010). Similarly in another study, a 13 SNPs score was associated with the first MI event (Ripatti et al., 2010). While some researchers were able to improve the net reclassification by the inclusion of GRS, the improvement remained modest (Davies et al., 2010; Lluís-Ganella et al., 2010), and even some failed to show a significant change in net reclassification index(Paynter et al., 2010; Ripatti et al., 2010).

In this study, the GRS was calculated assuming that all the SNPs had equal effect on the outcome and worked additively. However, this may not always be the case because the effect size of some SNPs is relatively high while some have more modest effects. This problem may be solved by the use of an externally weighted GRS, where the coded genotype is first

multiplied by an already estimated effect size of that SNP from a large study, such as a meta-analysis of GWASstudies, with the effect size being the log natural of odds ratio (OR). The effect sizes calculated from studies on such large number of samples are yet not available in Pakistan, and the only available effect sizes of the SNPs are from studies on Caucasians. The effect sizes of the SNPs may vary among ethnicities as linkage disequilibrium and allele frequencies vary (Wang & Tao Elston, 2007). We therefore,used the unweighted GRS which is also the most commonly used one (Lluís-Ganella et al., 2010; Yiannakouris, Katsoulis, Trichopoulou, Ordovas, & Trichopoulos, 2014). The SNP coding (0,1,2) was adjusted in such a way that all the SNPs were positively associated with the outcome.

We previously described the 19 SNPs score and a 13 SNPs score derived from it using only those SNPs present in genes/loci more robustly associated with CAD in the CARDIoGRAMplusC4D using 308 cases and 130 controls (Deloukas et al., 2013). The weighted GRS did not significantly differ between cases and controls and we found that the study may be adequately powered by increasing the samples to 340 cases and 340 controls (Beaney et al., 2015). Therefore, by increasing the number of samples (405 cases and 220 controls) and genotyping two new SNPs, *FTO* rs9939609 and *GLUL* 10911021, the GRS became higher than previously reported and was significantly associated with CAD. We already have reported the association of *FTO* rs9939609 with CAD in Pakistani people which was independent of blood lipid levels (Shahid, Rehman, & Hasnain, 2016) and the SNP *GLUL* rs10911021 was reported to be v associated with CAD risk in type 2 diabetes mellitus (Prudente et al., 2015).

In order to examine whether we have correctly genotyped the SNPs and to confirm the allele frequencies for these SNPs, we compared the allele frequencies of our subjects with PJL, which is a Pakistani Punjabi population from Lahore, in which 96 subjects were genotyped for many SNPs in the 1000 genomes project phase III (G. P. Consortium, 2012). The allele frequencies in our subjects did not significantly differ from those observed in PJL (Supplementary Table 8).

In conclusion, the 21 SNPs risk score can be used for genetic risk analysis in the Pakistani people but the results need to be replicated with bigger sample sizes and meta-analysis of individual SNPs for CAD association in the Pakistani population. The GRS of these 21 loci is also strongly associated with lipid profile, suggesting that the mechanism of these risk SNPs is likely to be in part at least through creating a more atherogenic lipid profile in subjects carrying high numbers of risk alleles.

Conflict of Interest: The authors declare that they have no competing interests.

Financial support

SS is supported by HEC Pakistan through grant number IRSIP 24 BMS 41.KB is supported by an MRC case award (1270920) with Randox Laboratories. SEH is a British Heart Foundation Professor, he and JC are supported by the British Heart Foundation (RG008/08) and by the National Institute for Health Research, University College London Hospitals Biomedical Research Centre.

Author contribution:

Saleem Ullah Shahid, Designed the study, performed the experiments, analysed the results and wrote the manuscript. Shabana, Katherine Beaney and Kawah Li helped in performing experiments, result analysis and manuscript writing. Jackie A Cooper helped in statistical analysis. Abdul Rehman, provided technical support and supervised the study. Steve E. Humphries, Provided logistic support, designed and supervised the project.

Table 1:Observed coronary artery disease odds ratio of the studied SNPs.

Gene	SNP	OR	C.I	*p*-value
MIA3	rs17465637	1.14	0.89-1.5	0.29
CDKN2A	rs10757274	1.18	0.93-1.5	0.17
DAB2IP	rs7025486	1.01	0.79-1.3	0.91
CXCL12	rs1746048	1.22	0.95-1.6	0.12
ACE	rs4341	1.22	0.97-1.5	0.09
NOS3	rs1799983	1.15	0.86-1.5	0.33
APOA5	rs662799	1.02	0.75-1.4	0.9
SMAD3	rs17228212	1.22	0.91-1.6	0.19
APOB	rs1042031	1.62	1.1-2.4	*0.01
CETP	rs708272	1.03	0.82-1.3	0.81
LPA	rs3798220	2.2	0.24-19.63	0.49
LPA	rs10455872	1.25	0.49-3.2	0.64
MRAS	rs9818870	1.09	0.73-1.6	0.68
LPL	rs328	1.5	0.98-2.3	0.06
LPL	rs1801177	-	-	-
SORT1	rs646776	1.2	0.92-1.5	0.19
PCSK9	rs11591147	3.71	0.34-41.2	0.29
APOE	rs429358	1.14	0.79-1.65	0.48
APOE	rs7412	1	0.56-1.77	0.98
GLUL	rs10911021	1.3	1-1.6	0.053
FTO	rs9939609	1.43	1-2.1	*0.009

OR: Odds ratio,*Statistically significant association of the SNP with CAD.

Table 2: Association of genetic risk score quintiles with CAD risk.

Genetic risk ScoreQuintiles	Allele ranges	OR (95% CI)	*p-value
1	<13	1	
2	13 to 15	1.54 (0.99-2.38)	0.052
3	15 to 16	2.19 (1.28-3.76)	0.004
4	16 to 17	2.81 (1.56-5.05)	0.001
5	>17	2.96 (1.71-5.13)	<0.001

*p:Inter quintile p-value. Analyses were adjusted for age, gender, BMI, hypertensive and diabetic status.

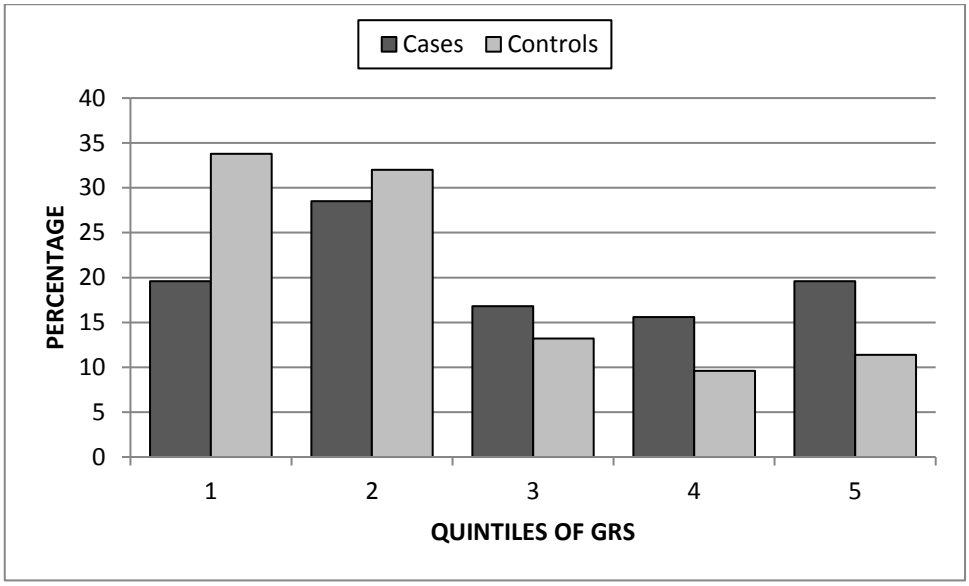
Table 3:Lipid levels in subjects with increasing number of genetic risk score alleles.

Gene score	Number	TC±SD	LDL-C±SD	HDL-C±SD	TG±SD
9	1	135.6	79.5	72.2	155.7
11	8	150.3±10.3	82.3±21	71.3±16.3	156.1±37
12	11	170.3±28.6	88.6±12.8	66.1±19.5	175.8±44.1
13	20	185.1±33.5	90.6±21.4	62.8±11.8	196.1±74.2
14	48	190.2±40.3	94.1±26	58.1±12.6	203.7±65.2
15	65	195.8±48.8	95.8±25.8	60.9±17.1	205.9±62.4
16	87	196.3±43.8	98.4±25	53.1±15.8	205±71.9
17	98	196.3±35.4	98.7±26.7	55.3±17.9	205.7±76.8
18	97	191.6±40.1	101.4±29.8	49.7±13.7	207.6±54.4
19	84	196±44.6	101.8±27.5	48.9±17.3	210.3±71.8
20	40	207.9±55	114.4±28.3	48.7±16.4	215.2±49.4
21	32	214.8±50	124.6±25.4	48.3±15.6	219.8±71.5
22	23	218.2±55.5	134.3±23.3	46.5±14.7	231±45.1
23	6	239.2±39.3	134.5±7.3	43.5±9.5	268.8±63.1
24	1	243	143	39	280
25	2	255.3±7.1	146.5±9.2	30±4.2	302.5±24.7

Table 4: Effect size of genetic risk score on blood lipid levels.

Effect size (β)± standard error and *p*-values							
TC	*p*-value	LDL-C	*p*-value	HDL-C	*p*-value	TG	*p*-value
3.7±0.7	$1.37x10^{-7}$	4±0.5	$1.9x10^{-20}$	2±.3	$3.6x10^{-15}$	4.2±1	$7.1x10^{-5}$

β is the increase/decrease in lipid levels perallele increase in genetic risk score .



HAVE ADDED AXIS TITLES AND LINES FOR GRID

Figure 1: Analysis of FREQUENCY OF SUBJECTS BY genetic risk score ~~quintiles by the frequency of subjects.~~

PLEASE ADD QUINTILE CUT-OFFS

Supplementary Table 1: Basic features of the SNPs used for gene score

Gene	Chromosome	SNP number	Base change	Risk allele	SNP type	References
MIA3	1q41	rs17465637	C>A	C	GWAS	(Kathiresan et al., 2009)
CDKN2A	9p21,3	rs10757274	A>G	G	GWAS	(Kathiresan et al., 2009)
DAB2IP	9q33,2	rs7025486	G>A	A	GWAS	(Harrison et al., 2012)
CXCL12	10q11,21	rs1746048	C>T	C	GWAS	(Samani et al., 2007)
ACE	17q23,3	rs4341	C>G	G	Candidate	(Casas, Cooper, Miller, Hingorani, & Humphries, 2006)
NOS3	7q36,1	rs1799983	G>T	T	Candidate	(Casas et al., 2006)
APOA5	11q23,3	rs662799	A>G	G	Candidate	(Sarwar & N Sandhu M.S, 2010)
SMAD3	15q22,33	rs17228212	T>C	C	GWAS	(Samani et al., 2007)
APOB	2p24,1	rs1042031	G>A	G	Candidate	(J. Casas, P et al., 2006)
CETP	16q13	rs708272	C>T	C	Candidate	(J. Casas, P et al., 2006)
LPA	6q25,3	rs3798220	T>C	C	Candidate	(Clarke et al., 2009)
LPA	6q26	rs10455872	A>G	G	Candidate	(Clarke et al., 2009)
MRAS	3q22,3	rs9818870	C>T	T	GWAS	(Erdmann et al., 2009)
LPL	8p21,3	rs328	C>G	C	Candidate	(J. Casas, P et al., 2006)
LPL	8p21,3	rs1801177	G>A	A	Candidate	(Sagoo et al., 2008)
SORT1	1p13,3	rs646776	A>G	A	Candidate	(Kathiresan et al., 2009)
PCSK9	1p32,3	rs11591147	G>T	G	Candidate	(Benn, Nordestgaard, Grande, Schnohr, & Tybjærg-Hansen, 2010)
APOE	19q13,32	rs429358	T>C	C	Candidate	(Bennet et al., 2007)
APOE	19q13,32	rs7412	C>T	T	Candidate	(Bennet et al., 2007)
GLUL	1q25,3	rs10911021	C>T	C	Candidate	(Qi et al., 2013)
FTO	16q12,2	rs9939609	T>A	A	Candidate	(Frayling et al., 2007)

Supplementary table 2: **List of primers and probes used in TaqMan assay**

Gene	SNP	40x	Primer & probe	Sequence of primers
eNOS (E289D)	rs1799983	ENOS_G894T_F	Primers	GGCTGGACCCCAGGAAA
		ENOS_G894T_R	Primers	CACCCAGTCAATCCCTTTGGT
		ENOS_G894T_VIC	Probe VIC =T	CCCAGATGATCCCCCA
		ENOS_G894T_FAM	Probe FAM=G	CCAGATGAGCCCCCA
SMAD3	rs17228212	CT17228212_F	Primers	TCACACTGTCTTTGCCGTCATT
		CT17228212_R	Primers	AGGGACGTGTCCTCACTCA
		CT17228212_VIC	Probe VIC = C	AGTTAGGTTGCGAGTTC
		CT17228212_FAM	Probe FAM = T	TTAGGTTGCAAGTTC
APOB (E4154K)	rs1042031	GA1042031_F	Primer	GGATAACGTGTTTGATGGCTTGGTA
		GA1042031_R	Primer	ATCAATGAGTGAGTCAATCAGATGCTT
		GA1042031_V	Probe VIC =G	AGTTACTCAAGAATTCCA
		GA1042031_M	Probe FAM=A	AGTTACTCAAAAATTCCA
LPA (I1891M)	rs3798220	LPA_I1891M-205F	Primers	CACCAAGAAGTGAACCTCGAATCT
		LPA_I1891M-205R	Primers	TGTGTGGGCTCCAAGAACAG
		LPA_I1891M-205V1	Probe VIC = A	CATGTTCAGGAAATAGAAGT
		LPA_I1891M-205M1	Probe FAM =G	ATGTTCAGGAAATGGAAGT
LPA	rs10455872	TC10455872_F	Primers	GTCTTGGGTAACAAGTGAAGGATATCT
		TC10455872_R	Primers	ACACATAGCTTTTCAGACACCTTGT
		TC10455872_V	Probe VIC = A	CTCAGAACCCAATGTGTTT
		TC10455872_M	Probe FAM = G	CAGAACCCAGTGTGTTT
CELSR2/ PSRC1/ SORT1	rs646776	AG599839-143F	Primers	CTGGGTGACAGAGCAAGATTCT
		AG599839-143R	Primers	GCTTACTCTATGAGTCTTCATTTTTCTAA AATAAAGTG
		AG599839-143V1	Probe VIC = A	CAGGATCAACTTCC
		AG599839-143M1	Probe FAM =G	CAGGATCGACTTCC
APOE 112	rs429358	Pre -designed	GCTGGGCGCGGACATGGAGGACGTG[C/T] GCGGCCGCCTGGTGCAGTACCGCGG	
		C-3084793_20	Reverse chromosome 19	

Supplementary table 3:**List of primers and probes used in KASPar assay**

Gene	SNP	Sequence (50bp either side, allele label = [A/T], Other SNPs = N or IUPAC Code)	FAM	VIC
MIA3	rs17465637	GAACCAAACCATATCACTTTTTAAAACCATAATAGTTA TGCTGAGAAGTT**[C/A]**TTTTTTGTCATAGTGCAAGATA ACATGTCTTTGCTGCTGATACATTGGGT	C	A
CDKN2A	rs10757274	GGTATTACAAAAAGCTTCTCCCCCGTGGGTCAAATCTA AGCTGAGTGTTG**[A/G]**GACNTAATTGAAATTCACTAGA TAGATAGGAGATAGGGGTAGGGAATTCT	A	G
DAB2IP	rs7025486	GGGNCTTGAGTGGTGAGCAAAGAGGGGAGAACAGCC CCTGGCAGACCACT**[A/G]**GGAATCAAAGGAAGGATTTT GAAAATAACAGGAATGATAACAGTGATCTC	A	G
CETP	rs708272	TTTACCCCCTGACTCAACCCCCTAACCTGGCTCAGATC TGAACCCTAACT[**C/T]**GAACCCCANTGATTCTGGGTCT CAGACAAACACAAATCCCTATACCTGGC	C	T
APOA5	rs662799	AAGAGGCATCTGGGCCAGNGACTCTGAGCCCCAGGAA CTGGAGCGAAAGT**[A/G]**AGATTTGCCCCATGAGGAAA AGCTGAACTCCACTCGCAGGGCCTCTGAGG	A	G
LPL (S474X)	rs328	GGCACCTGCGGTATTTGTGAAATGCCATGACAAGTCTC TGAATAAGAAGT**[C/G]**AGGCTGGTGAGCATTCTGGGCT AAAGCTGACTGGGCATCCTGAGCTTGCA	C	G
MRAS	rs9818870	TCTCTTGCTGCNTTTTCACATCAGCTGTGCTGCTTGGTG CCTCTCTGATA**[C/T]**NAATACACTGACACGTCAAAGTA ACCTAATGTGGACACCATCCAGAAAAC	C	T
APOE158	rs7412	TGNNNAAGCTGNNTNAGCNGCTCCNCCNCGATGCCGA TGACCTGCAGAAG**[C/T]**GCCTGGCAGTGTACCAGGCCG GGNCCCGCGAGGGCGCCNAGCGCGGCCTC	C	T
SLCO1B1	rs4149056	NATCTACATAGGTTNTTTAAAGGAATCTGGGTCATACA TGTGGATATNTG**[T/C]**GTTCATGGGTAATATGCTTNNT GGAATAGGGGAGACTCCCATAGTACCANT	T	C
CXCL12	rs1746048	ATTTCAGGACTGAACAGAGACTGAGAAGGGTAAAGGG TGGTAGGATTGAG**[C/T]**GAGTCAGGCCAGAAACCTCTA GTTAGCTACCATGACAGAAGGGAAACATG	C	T
ACE	rs4341	TCTCTGAGCTCCCCTTACAAGCAGARGTGAGCTAAGG GCTGGARCTYAAG**[C/G]**CATTCMAMCCCCTACCAGAT STGACGAATRTGATGGCCRCRTCCCGGAAA	C	G
LPL	rs1801177	CAGTTAACCTCATATCYAATTTTTCCKTTCCAGAAAGA AGAGATTTTATY**[G/A]**ACATYGRAAGTAAATTTGCCCT AAGGAMCCCTGAAGWCACAGSTGARGAC	G	A
PCSK9	rs11591147	TGCGCAGGAGGACRAGGACGGCGACTACGAGGAGCT GGTGCTAGCCTTGC**[G/T]**TTCYGAGGAGGACGGCCTGG YCGAAGCACCCRAGCACGGAACCACASCCA	G	T

Supplementary table 4:**Sequence of primers used in PCR**

Primer	Sequence
*MIA3*_F	5'-ATCCAATCACCTTCCACCAG-3'
*MIA3*_R	5'-CCCAATGTATCAGCAGCAAA-3'
*CDKN2A*_F	5'-GTTTCTGCACATGGTGATGG-3'
*CDKN2A*_R	5'-CATTCCCCAACATTTGTCCT-3'
*DAB2IP*_F	5'-GCAGATGGTGTGACTGGAAA-3'
*DAB2IP*_R	5'-AACCCCTGGTGCTGTGTAAG-3'
*ACE*_F	5'-CCCCTTACAAGCAGAGGTGA-3'
*ACE*_R	5'-TCGGGTAAAACTGGAGGATG-3'
*CETP*_F	5'-GTGACCCCCAACACCAAATA-3'
*CETP*_R	5'-TCGCCTTCAAGGTCAAGTTC-3'
*APOA5*_F	5'-GCAGGGTGAAGATGAGATGG-3'
A*POA5*_R	5'-TAGACGGAGTGGGTGTGTCA-3'
*SMAD3*_F	5'-CTCAGATCCTTTGCGGGTAG-3'
*SMAD3*_R	5'-TCTTCTGTGCAGACCAGGTG-3'
*LPA*rs3798220_F	5'-GAAGGGGCTGGACCATATTT-3'
*LPA*rs3798220_R	5'-AAGACCACAGGTGAGCGAGT-3'
*eNOS*_F	5'-ACTCCCCACAGCTCTGCAT-3'
*eNOS*_R	5'-CAGTCAATCCCTTTGGTGCT-3'
*LPL*rs328_F	5'-CTTCCACAGGGTGATCTTCTG-3'
*LPL*rs328_R	5'-CATGAAGCTGCCTCCCTTAG-3'
*LPL*rs180_F	5'-AAATAGCATCAGCGGTGGTT-3'
*LPL*rs180_R	5'-ATGAGGTGGCAAGTGTCCTC-3'
*SLCO1B1*_F	5'-GAATCTGGGTCATACATGTGG-3'
*SLCO1B1*_R	5'-AAGGGAAAGTGATCATACAATTTAATA-3'
*PCSK9*_F	5'-GACTACGAGGAGCTGGTGCT-3'
*PCSK9*_R	5'-CCTGCACTCCACTTCCTCTC-3'
*MRAS*_F	5'-TCTTGCTGCGTTTTCACATC-3'
*MRAS*_R	5'-TTGACTCCAAGGGAAGATGG-3'
*APOB*_F	5'-GCCCAGAATCTGTACCAGGA-3'
*APOB*_R	5'-TGGAATCTGGGGAAGTTCAG-3'
*CXCL12*_F	5'-GTCCAGATGAGGCCATCAAG-3'
*CXCL12*_R	5'-TGCCAAGAAAATGACACAGC-3'
LP(a)rs10455872_F	5'-GCATAGCCAGACATGGGTTT-3'
*LP(*a)rs10455872_R	5'-TGCCATGTTTGTCTTGGGTA-3'
*CELSR2*_F	5'-TGGTGAAAAGGACACCTTCC-3'
*CELSR2*_R	5'-CTGTCCGCTTCTGTGTGGTA-3'
*APOE158*_F	5'-CTGCGTAAGCGGCTCCTC-3'
*APOE158*_R	5'-CTGCCCATCTCCTCCATC-3'
*APOE112*_F	5'-GCCTACAAATCGGAACTGGA-3'
*APOE112*_R	5'-CAGCTCCTCGGTGCTCTG-3'

Supplementary Table 5: Anthropometric and biochemical parameters of study subjects.

Variables	Cases	Controls	*p*-value
Number	405	220	
Age (years)	59.1±12.6	56 ± 10.5	0.002
Sex Males (n) Females (n)	 216 189	 120 100	0.27
Diabetes (%)	64.6	13.6	$5.1x10^{-34}$
Hypertension (%)	62.1	16.4	$8.9x10^{-28}$
Smoking (%)	29.5	10.5	$7.3x10^{-08}$
Total cholesterol	207.5±53.7	175.4±43	$8.8x10^{-14}$
Triglycerides	212.4±70	188±66.3	$2.6x10^{-5}$
LDL-C	106±28.9	84.7±17	$6.3x10^{-22}$
HDL-C	45.2±11.9	67.4±16.3	$1.8x10^{-66}$

supplementary **Table 6: Comparison of RAFs between cases and controls**

Gene	SNP	RAFs (C.I)		
		Cases	Controls	*p*-value
SORT1	rs646776	0.75 (0.72-0.78)	0.72 (0.67-0.76)	0.19
APOB	rs1042031	0.92 (0.90-0.94)	0.87 (0.84-0.91)	*0.007
APOE	rs429358	0.12 (0.10-0.14)	0.11 (0.08-0.13)	0.46
APOE	rs7412	0.04 (0.03-0.05)	0.04 (0.02-0.06)	0.98
LPL	rs328	0.94 (0.92-0.95)	0.91 (0.88-0.93)	0.06
LPL	rs1801177	-	-	-
APOA5	rs662799	0.17 (0.14-0.20)	0.167 (0.13-0.20)	0.89
CETP	rs708272	0.551 (0.52-0.59)	0.543 (0.50-0.59)	0.8
LPA	rs3798220	0.005 (0.0-0.01)	0.002 (0.00-0.01)	0.66
LPA	rs10455872	0.017 (0.01-0.03)	0.014 (0.00-0.02)	0.81
CDKN2A	rs10757274	0.505 (0.47-0.54)	0.463 (0.42-0.51)	0.162
MIA3	rs17465637	0.651 (0.62-0.68)	0.621 (0.58-.67)	0.292
DAB2IP	rs7025486	0.318 (0.29-0.35)	0.315 (0.27-0.36)	0.913
SMAD3	rs17228212	0.207 (0.18-0.23)	0.176 (0.14-0.21)	0.19
MRAS	rs9818870	0.094 (0.07-0.11)	0.087 (0.06-0.11)	0.67
CXCL12	Rs1746048	0.675 (0.64-0.71)	0.63 (0.58-0.68)	0.114
ACE	rs4341	0.577 (0.54-0.61)	0.525 (0.48-0.57)	0.079
NOS3	rs1799983	0.202 (0.17-0.23)	0.178 (0.14-0.21)	0.312
PCSK9	rs11591147	0.999 (1-1)	0.995 (0.99-1)	0.252
GLUL	rs10911021	0.68 (0.64-0.70)	0.62 (0.61-0.70)	0.055
FTO	rs9939609	0.35 (0.32-0.39)	0.28 (0.25-32)	*0.003

C.I: Confidence interval, RAFs: Risk allele frequencies, * significantly high RAF in cases than controls.

Supplementary Table 7: Genetic risk score in cases and controls and its association with CAD.

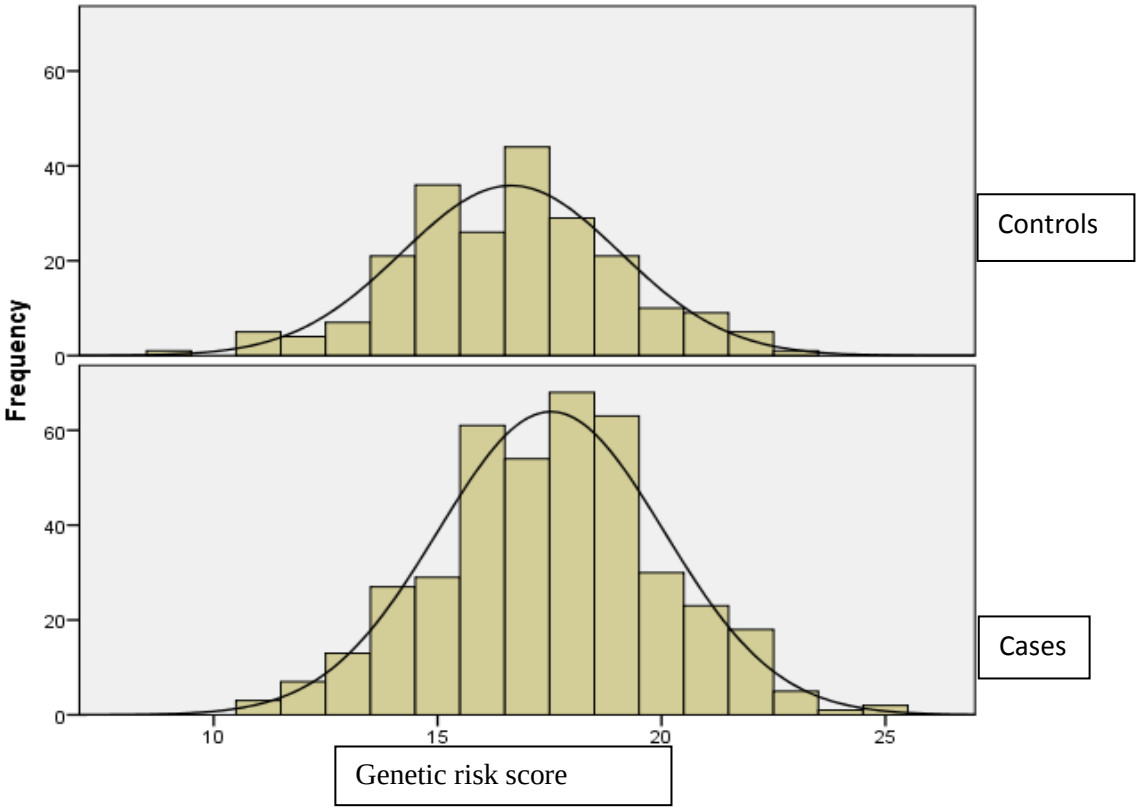
Mean gene score	Cases	Controls	*p*-value	OR	95% C.I	*p*-value
Un-weighted	17.53±2.52	16.64±2.44	$2.4x10^{-5}$	1.16	1.08-1.23	<0.0001

All the analyses were adjusted for age, gender, BMI, hypertensive and diabetic status,

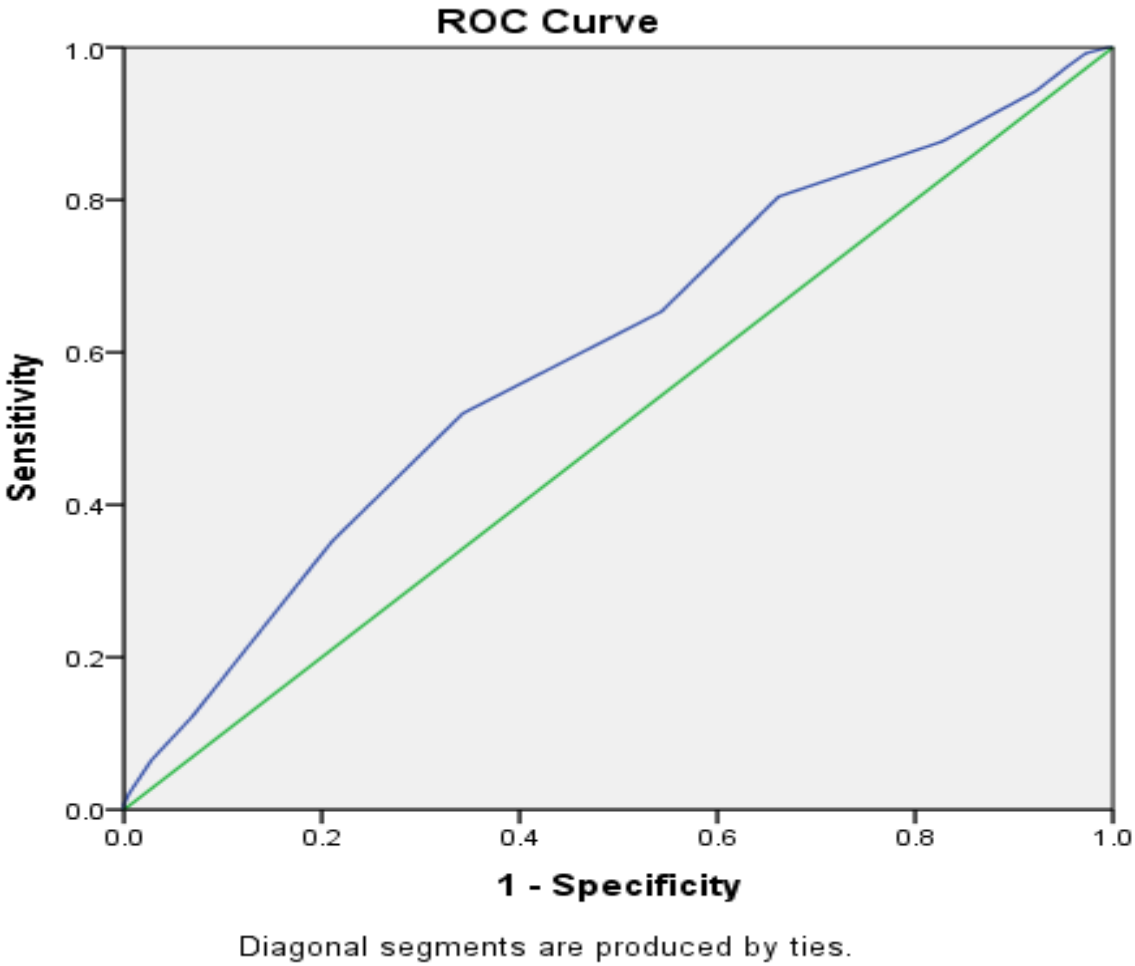
supplementary Table 8: Comparison of risk allele frequencies in studied samples with Pakistani Punjabi population from Lahore (PJL) in 1000 genomes project phase III.

Gene	SNP	Risk allele frequencies		
		Current study	PJL	**p*-value
MIA3	rs17465637	0.64	0.62	0.5
CDKN2A	rs10757274	0.49	0.495	0.9
DAB2IP	rs7025486	0.317	0.26	0.1
CXCL12	rs1746048	0.659	0.635	0.5
ACE	rs4341	0.559	0.63	0.06
NOS3	rs1799983	0.193	0.151	0.2
APOA5	rs662799	0.169	0.151	0.5
SMAD3	rs17228212	0.196	0.13	0.03
APOB	rs1042031	0.905	0.906	0.9
CETP	rs708272	0.548	0.547	0.97
LPA	rs3798220	0.004	0.005	0.81
LPA	rs10455872	0.016	0.005	0.25
MRAS	rs9818870	0.091	0.115	0.31
LPL	rs328	0.925	0.948	0.26
LPL	rs1801177	0	0	
CELSR2	rs646776	0.739	0.745	0.87
PCSK9	rs11591147	0.998	0	0.5
APOE	rs429358	0.114	0.083	0.2
APOE	rs7412	0.04	0.036	0.77

* is p value between risk allele frequencies in subjects from this study and that of from PJL.



Supplementary Figure 1: Histogram showing comparison of genetic risk score in cases and controls.



Supplementary Figure 2: Receiver operator curve, area under ROC=0.602 (C.I=0.56-0.65, *P*<0.001).

Alwan, A. (2011). *Global status report on noncommunicable diseases 2010*: World Health Organization.

Anderson, J. L., Horne, B. D., Camp, N. J., Muhlestein, J. B., Hopkins, P. N., Cannon-Albright, L. A., . . . Nicholas, Z. P. (2010). Joint effects of common genetic variants from multiple genes and pathways on the risk of premature coronary artery disease. *American Heart Journal, 160*(2), 250-256. e253.

Assimes, T. L., Knowles, J. W., Basu, A., Iribarren, C., Southwick, A., Tang, H., . . . Rubin, G. D. (2008). Susceptibility locus for clinical and subclinical coronary artery disease at chromosome 9p21 in the multi-ethnic ADVANCE study. *Human Mol. Genet., 17*(15), 2320-2328.

Assmann, G., Cullen, P., & Schulte, H. (2002). Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Münster (PROCAM) study. *Circulation, 105*(3), 310-315.

Aziz, Uddin, K., Faruqui, A., Patel, N., & Jaffery, H. (2012). Prevalence and awareness of cardiovascular disease including life styles in a lower middle class urban community in an Asian country. *Pak. Heart J., 41*(3-4).

Beaney, K. E., Cooper, J. A., Shahid, S. U., Ahmed, W., Qamar, R., Drenos, F., . . . Humphries, S. E. (2015). Clinical utility of a coronary heart disease risk prediction gene score in UK healthy middle aged men and in the Pakistani population. *PloS one, 10*(7), e0130754.

Belsky, D. W., Moffitt, T. E., Sugden, K., Williams, B., Houts, R., McCarthy, J., & Caspi, A. (2013). Development and evaluation of a genetic risk score for obesity. *Biodemograp. Soc. Biol., 59*(1), 85-100.

Benn, M., Nordestgaard, B. G., Grande, P., Schnohr, P., & Tybjærg-Hansen, A. (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.

Bennet, A. M., Di Angelantonio, E., Ye, Z., Wensley, F., Dahlin, A., Ahlbom, A., . . . de Faire, U. (2007). Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA, 298*(11), 1300-1311.

Brindle, P., Beswick, A., Fahey, T., & Ebrahim, S. (2006). Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. *Heart, 92*(12), 1752-1759.

Carreras-Torres, R., Kundu, S., Zanetti, D., Esteban, E., Via, M., & Moral, P. (2013). Genetic risk score of NOS gene variants associated with myocardial infarction correlates with coronary incidence across Europe. *PloS one, 9*(5), e96504-e96504.

Casas, P, J., Cavalleri, G. L., Bautista, L. E., Smeeth, L., Humphries, S. E., & Hingorani, A. D. (2006). Endothelial nitric oxide synthase gene polymorphisms and cardiovascular disease: a HuGE review. *American Journal of Epidemiology, 164*(10), 921-935.

Casas, J., P, Cooper, J., Miller, G. J., Hingorani, A. D., & Humphries, S. E. (2006). Investigating the genetic determinants of cardiovascular disease using candidate genes and meta-analysis of association studies. *Annals of Human Genetics, 70*(2), 145-169.

Clarke, R., Peden, J. F., Hopewell, J. C., Kyriakou, T., Goel, A., Heath, S. C., . . . Rust, S. (2009). Genetic variants associated with Lp (a) lipoprotein level and coronary disease. *New England Journal of Medicine, 361*(26), 2518-2528.

Collins, G., & Altman, D. (2009). An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. *BMJ (Clinical research ed.), 340*, c2442-c2442.

Conroy, R., Pyörälä, K., Fitzgerald, A. E., Sans, S., Menotti, A., De Backer, G., . . . Keil, U. (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal, 24*(11), 987-1003.

Consortium, G. P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature, 491*(7422), 56-65.

Consortium, I. K. C. (2011). Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet, 7*(9), e1002260.

Cooper, J. A., Miller, G. J., & Humphries, S. E. (2005). A comparison of the PROCAM and Framingham point-scoring systems for estimation of individual risk of coronary heart disease in the Second Northwick Park Heart Study. *Atherosclerosis, 181*(1), 93-100.

Cordell, H. J. (2009). Detecting gene–gene interactions that underlie human diseases. *Nat. Rev. Genet., 10*(6), 392-404.

Davies, R. W., Dandona, S., Stewart, A. F., Chen, L., Ellis, S. G., Tang, W. W., . . . Wells, G. A. (2010). Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies. *Circulation: Cardiovascular Genetics, 3*(5), 468-474.

Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T. L., Thompson, J. R., . . . Goldstein, B. A. (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetics, 45*(1), 25-33.

Enas, & Kannan, S. (2005). *How to beat the heart disease epidemic among South Asians: A prevention and management guide for Asian Indians and their doctors*: Advanced Heart Lipid Clinic.

Erdmann, J., Großhennig, A., Braund, P. S., König, I. R., Hengstenberg, C., Hall, A. S., . . . Trégouët, D.-A. (2009). New susceptibility locus for coronary artery disease on chromosome 3q22. 3. *Nature genetics, 41*(3), 280-282.

Fava, C., Sjögren, M., Olsson, S., Lövkvist, H., Jood, K., Engström, G., . . . Lindgren, A. (2014). A genetic risk score for hypertension associates with the risk of ischemic stroke in a Swedish case–control study. *European Journal of Human Genetics*.

Folkersen, L. van't Hooft, F., Chernogubova, E., Agardh, H. E., Hansson, G. K., Hedin, U., . . . Eriksson, P. (2010). Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. [Research Support, Non-U.S. Gov't]. *Circ. Cardiovasc. Genet., 3*(4), 365-373.

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., . . . Rayner, N. W. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science, 316*(5826), 889-894.

Gaziano, T. A., Bitton, A., Anand, S., Abrahams-Gessel, S., & Murphy, A. (2010). Growing epidemic of coronary heart disease in low-and middle-income countries. *Current problems in cardiology, 35*(2), 72-115.

Harrison, C, S., Cooper, J. A., Li, K., Talmud, P. J., Sofat, R., . . . Neil, A. (2012). Association of a sequence variant in *DAB2IP* with coronary heart disease. *European Heart Journal, 33*(7), 881-888.

Hernesniemi, J. A., Seppälä, I., Lyytikäinen, L.-P., Mononen, N., Oksala, N., Hutri-Kähönen, N., . . . Schork, N. J. (2012). Genetic profiling using genome-wide significant coronary artery disease risk variants does not improve the prediction of subclinical atherosclerosis: the cardiovascular risk in young finns study, the Bogalusa heart study and the health 2000 Survey: a meta-analysis of three independent studies. *PloS one, 7*(1), e28931.

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., Minhas, R., Sheikh, A., & Brindle, P. (2008). Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ, 336*(7659), 1475-1482.

Ioannidis, J. P. (2009). Prediction of cardiovascular disease outcomes and established cardiovascular risk factors by genome-wide association markers. *Circulation: Cardiovascular Genetics, 2*(1), 7-15.

Jafar, T. H., Jafary, F. H., Jessani, S., & Chaturvedi, N. (2005). Heart disease epidemic in Pakistan: women and men at equal risk. *American Heart Journal, 150*(2), 221-226.

Kathiresan, S., Voight, B. F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P. M., . . . Schunkert, H. (2009). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nature Genetics, 41*(3), 334-341.

Lenfant, C. (2003). Clinical research to clinical practice—lost in translation? *New England Journal of Medicine, 349*(9), 868-874.

Li, T. Y., Rana, J. S., Manson, J. E., Willett, W. C., Stampfer, M. J., Colditz, G. A., . . . Hu, F. B. (2006). Obesity as compared with physical activity in predicting risk of coronary heart disease in women. *Circulation, 113*(4), 499-506.

Lluís-Ganella, C., Lucas, G., Subirana, I., Sentí, M., Jimenez-Conde, J., Marrugat, J., . . . Elosua, R. (2010). Additive effects of multiple genetic variants on the risk of coronary artery disease. *Revista Espanola de Cardiologia, 63*(8), 925-933.

McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R., . . . Folsom, A. R. (2007). A common allele on chromosome 9 associated with coronary heart disease. *Science, 316*(5830), 1488-1491.

Mehta, & N, N. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Circulation: Cardiovascular Genetics, 4*(3), 327-329.

Paynter, N. P., Chasman, D. I., Paré, G., Buring, J. E., Cook, N. R., Miletich, J. P., & Ridker, P. M. (2010). Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA, 303*(7), 631-637.

Peden, J. F., & Farrall, M. (2011). Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. *Human Molecular Genetics, 20*(R2), R198-R205.

Prudente, S., Shah, H., Bailetti, D., Pezzolesi, M., Buranasupkajorn, P., Mercuri, L., . . . Trischitta, V. (2015). Genetic variant at the GLUL locus predicts all-cause mortality in patients with type 2 diabetes. *Diabetes*, db141653.

Qi, L., Ma, J., Qi, Q., Hartiala, J., Allayee, H., & Campos, H. (2011). Genetic risk score and risk of myocardial infarction in Hispanics. *Circulation, 123*(4), 374-380.

Qi, L., Qi, Q., Prudente, S., Mendonca, C., Andreozzi, F., di Pietro, N., . . . Formoso, G. (2013). Association Between a Genetic Variant Related to Glutamic Acid Metabolism and Coronary Heart Disease in Individuals With Type 2 Diabetes. *JAMA, 310*(8), 821-828.

Ridker, P. M., Buring, J. E., Rifai, N., & Cook, N. R. (2007). Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA, 297*(6), 611-619.

Ripatti, S., Tikkanen, E., Orho-Melander, M., Havulinna, A. S., Silander, K., Sharma, A., . . . Sinisalo, J. (2010). A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet, 376*(9750), 1393-1400.

Sagoo, G. S., Tatt, I., Salanti, G., Butterworth, A. S., Sarwar, N., van Maarle, M., . . . Higgins, J. P. (2008). Seven lipoprotein lipase gene polymorphisms, lipid fractions, and coronary disease: a HuGE association review and meta-analysis. [Meta-Analysis

Research Support, Non-U.S. Gov't

Review]. *American Journal of Epidemiology, 168*(11), 1233-1246.

Samani, N. J., Erdmann, J., Hall, A. S., Hengstenberg, C., Mangino, M., Mayer, B., . . . Wichmann, H.-E. (2007). Genomewide association analysis of coronary artery disease. *New England Journal of Medicine, 357*(5), 443-453.

Sarwar, & N Sandhu M.S, R. S. L. (2010). Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. *Lancet, 101*, 1634-1639.

Shahid, S. U., Cooper, J. A., Beaney, K. E., Li, K., Rehman, A., & Humphries, S. E. (2016). Effect of SORT1, APOB and APOE polymorphisms on LDL-C and coronary heart disease in Pakistani subjects and their comparison with Northwick Park Heart Study II. *Lipids in health and disease, 15*(1), 1.

Shahid, S. U., Cooper, J. A., Rehman, A., & Humphries, S. E. (2016). Association of ACE and NOS3 Gene Polymorphism with Blood Pressure in a Case Control Study of Coronary Artery Disease in Punjab, Pakistan. *Pakistan Journal of Zoology, 48*(4).

Shahid, S. U., Rehman, A., & Hasnain, S. (2016). Role of a common variant of Fat Mass and Obesity associated (FTO) gene in obesity and coronary artery disease in subjects from Punjab, Pakistan: a case control study. *Lipids in health and disease, 15*(1), 1.

Shahid, S. U., Shabana, Rehman, A., & Hasnain, S. (2016). Role of a common variant of Fat Mass and Obesity associated (FTO) gene in obesity and coronary artery disease in subjects from Punjab, Pakistan: a case control study. *Lipids in Health and Disease, 15*(1), 1.

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., . . . Lee, C. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science, 315*(5813), 848-853.

Thanassoulis, G., Peloso, G. M., Pencina, M. J., Hoffmann, U., Fox, C. S., Cupples, L. A., . . . O'Donnell, C. J. (2012). A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: The Framingham Heart Study. *Circulation: Cardiovascular Genetics, 5*(1), 113-121.

Thanassoulis, G., & Vasan, R. S. (2010). Genetic cardiovascular risk prediction will we get there? *Circulation, 122*(22), 2323-2334.

Wang, J, T., Gona, P., Larson, M. G., Tofler, G. H., Levy, D., . . . Robins, S. J. (2006). Multiple biomarkers for the prediction of first major cardiovascular events and death. *New England Journal of Medicine, 355*(25), 2631-2639.

Wang, & Tao Elston, R. C. (2007). Improved power by use of a weighted score test for linkage disequilibrium mapping. *The American Journal of Human Genetics, 80*(2), 353-360.

Wilson, P. W., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., & Kannel, W. B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation, 97*(18), 1837-1847.

Wood, D., Wray, R., Poulter, N., Williams, B., Kirby, M., Patel, V., . . . Potter, J. (2005). JBS 2: Joint British Societies' guidelines on prevention of cardiovascular disease in clinical practice. *Heart, 91*.

Wray, N. R., & Goddard, M. E. (2010). Multi-locus models of genetic risk of disease. *Genome Med., 2*(2), 10.11-10.13.

Yiannakouris, N., Katsoulis, M., Trichopoulou, A., Ordovas, J. M., & Trichopoulos, D. (2014). Additive influence of genetic predisposition and conventional risk factors in the incidence of coronary heart disease: a population-based study in Greece. *BMJ, 4*(2), e004387.