Journal of Molecular Pathophysiology
www.scopemed.org
DOI: 10.5455/jmp.20141204081445

Original Research

Determining the risk of coronary artery disease using genetic markers in Asian Indians

Jayashree Shanker[1], Prathima Arvind[1], Jiny Nair[1], Vandana Ravindran[2], Srikarthika Jambunathan[1], Vijay V. Kakkar[3]

[1]Mary and Garry Weston Functional Genomics Unit, Thrombosis Research Institute, Bangalore, Karnataka, India, [2]Elizabeth and Emmanuel Kaye Bioinformatics and Statistics unit, Thrombosis Research Institute, Bangalore, Karnataka, India, [3] Scientific Chairman, Thrombosis Research Institute, Bangalore, Karnataka, India

Address for correspondence:
Dr. Jayashree Shanker, Section Head, Mary and Garry Weston Functional Genomics Unit, Thrombosis Research Institute, India; Narayana Health, 258/A, Bommasandra Industrial Area, Anekal Taluk, Bangalore - 560 099, Karnataka, India.
Phone: +91-80-27835303,
Fax: +91-80-27835302,
E-mail: jayashreeshanker@triindia.org.in

Received: November 09, 2014
Accepted: December 04, 2014
Published: December 27, 2014

ABSTRACT

Objective: Recent discoveries of single-nucleotide polymorphisms (SNPs) have spurred the development of risk prediction models for coronary artery disease (CAD). We sought to generate a genetic risk score (GRS) for CAD in a representative cohort of Asian Indians. **Methods and Results:** In Stage 1, 88 variants belonging to 65 genes and the 9p21.3 locus, identified from genome-wide association studies and internal findings were genotyped by Taqman assay in 500 CAD patients (cases) and 500 controls. Twelve SNPs showed significant independent association with CAD. In Stage 2, seven out of the twelve SNPs, analyzed in 534 cases and 534 controls, showed persistent association and belonged to two loci - 9p21.3 (rs10757278, rs2383206, rs10757274, rs1333049, rs4977574) and CELSR2-PSRC1-SORT1 (rs646776, rs599839). Two representative variants, rs10757274 (odds ratio [OR] 1.28, 95% confidence interval [CI] 1.08-1.52) and rs599839 (OR 1.33, 95% CI 1.1-1.62) constituted our pilot GRS. Subjects in the 4th GRS quartile showed a higher risk of CAD as compared to the 1st quartile after adjusting for the classical risk factors (RFs) (OR 2.51, 95% CI 1.82 - 3.45, $P < 0.001$). In receiver operating characteristic analysis, the two GRS SNPs along with all the conventional RFs (Model 3, C = 0.837) showed better discrimination of CAD than either conventional RFs (Model 1, C = 0.766) or SNPs (Model 2, C = 0.576) alone ($Z = -6.6046$, $P = 3.98 \times 10^{-11}$), with 24% net reclassification of subjects in the intermediate risk group. **Conclusion:** A pilot 2-SNP GRS showed 2.5-fold higher risk of CAD in Asian Indians with a modest discrimination, offering scope for further improvement with the addition of newer genetic variants.

KEY WORDS: Asian Indians, coronary artery disease, genetic risk score, single nucleotide polymorphism

INTRODUCTION

Coronary artery disease (CAD) has a complex etiology involving both genetic and environmental risk factors (RFs) and their interactions thereof. There has been a keen interest to develop risk prediction models that can influence, improve and streamline clinical management decisions and possibly even lead to prevention of CAD. To this effect, several risk prediction tools have been developed that fundamentally rely on the conventional RFs [1-3], with appropriate drugs targeting them. Although such efforts have nominally increased the average life span of CAD patients by three years in the United States of America [4], nearly 15% of the subjects who are classified as low risk eventually develop CAD [5], fuelling the need for better predictive algorithms that incorporate recent discoveries in the genetics of cardiovascular disease. Unparalleled advances in genomic technologies have fostered the discovery of novel genetic variants associated with CAD, myocardial infarction or the RFs through genome-wide association studies (GWAS) [6] with subsequent validation of the interesting novel variants

across global populations [7,8]. Thus, the concept of a genetic risk score (GRS) that incorporates a number of genetic variants encompassing a wide spectrum of biological pathways are now established.

The specific variants and the aggregate number of markers incorporated in the GRS models differ across the studies, yielding inconsistent results. While some show modest improvement in the c index by virtue of the addition of genetic variants to the conventional RFs [9-11], others have failed to show a significant incremental change [12,13]. A few studies have reported clinically meaningful risk stratification based on net reclassification and integrated discrimination improvement (IDI) scores [10,14]. At best, the findings have been modest, with individual variants carrying a low power of discrimination. Similar studies have not been published in Asian Indians, a population that has high rates of CAD incidence, prevalence and mortality when compared to the other world populations, and is underscored by the presence of a strong family history and early onset of the disease. In this context, the aim of the present study was to identify and evaluate the performance of a number of genetic variants selected from published GWAS and internal discoveries of candidate gene studies, followed by validation of the results in independent datasets selected from the Indian Atherosclerosis Research Study (IARS).

Materials and Methods

The participants were selected from the IARS cohort, enrolled from May 2004 to December 2011. An overview of the IARS study design has been published [15]. Briefly, the IARS is an ongoing epidemiological study with an objective to investigate the genetic factors and biomarkers against a backdrop of the conventional RFs associated with CAD in Asian Indians living in India. The IARS cohort comprise of proband (index cases) and their family members having a strong history of cardiovascular disease as well as healthy community-based controls, matched for age and gender to the proband and without a positive family history of cardiovascular disease. Recruitment of cases and controls is based on predefined inclusion/exclusion criteria, with age at onset ≤60 years for men and ≤65 years for women. CAD patients (cases) showed clinical evidence of stable angina or myocardial infarction diagnosed through coronary angiography and electrocardiogram (ECG) and treated with standard medication or invasive procedures such as percutaneous coronary intervention or bypass surgery. Control subjects were clinically asymptomatic and showed normal ECG readings. All participants provided a voluntary informed signed consent. The IARS protocol was approved by the institutional ethics committee and follows the guidelines of the Indian Council of Medical Research on bioethics [16].

All study participants provided a fasting sample of blood and urine. Details of demographics, anthropometrics, medical history and pedigree were recorded for each participant during a face-to-face interview. Prevalence of Type 2 diabetes and hypertension was ascertained based on self-report of physician's diagnosis and/or use of prescription medication. Lipid markers, namely serum triglycerides (TG) and total cholesterol (TC) were estimated using standard enzymatic analysis in a Cobas-Fara II Clinical Chemistry Auto analyzer (F. Hoffman La Roche Ltd., Switzerland). High-density lipoprotein cholesterol (HDL-C) concentrations were estimated after precipitating the non-HDL-C fractions with a mixture of 43.24 mg/dl (2.4 mmol/l) phosphotungstic acid and 312.70 mg/dL (39 mmol/l) of magnesium chloride (Bayer Diagnostics, Gujarat, India). Plasma low-density lipoprotein cholesterol (LDL-C) concentrations were calculated using the Friedewald's equation [17]. Inter-assay coefficient of variation for the commercial controls and normal serum pool ranged from 4.9% to 7.0% for TC, 6.1-7.7% for TG, and 7.1-12.2% for HDL-C.

Study Design and Cohort Selection

An overview of the study design, conducted in two Stages (training set and test set), is shown in [Figure 1]. In Stage 1, 1000 subjects including 500 cases (IARS proband) and 500 controls, matched for age, gender and mother tongue/state of origin to the case, were selected (training set). In Stage 2, additional 1068 subjects that included 534 cases and 534 controls, matched for age and gender were selected for validation of the Stage 1 findings (test set).

Single-nucleotide Polymorphisms (SNP) Selection

We initially selected 88 SNPs from 65 putative candidate genes and from the 9p21.3 locus based on published reports of GWAS ($P = 1 \times 10^{-6-8}$) [18,19] and from internal findings [20-22]. Selected SNPs were predominantly located in genes associated with lipids, inflammation and immune mechanism, thrombosis, cell proliferation and novel genomic loci including those with unknown function. Supplementary Table 1 provides a summary of the 88 SNPs and the associated genes selected for the initial analysis.

DNA was extracted by a modified salting out procedure [23] and quantified using Nanodrop spectrophotometer (Thermo Fisher Scientific, Delaware, USA). The polymorphic status of the 52 SNPs was established by targeted sequencing of around 500 bp of genomic region encompassing the SNP of interest on pooled DNA samples, constructed with equimolar

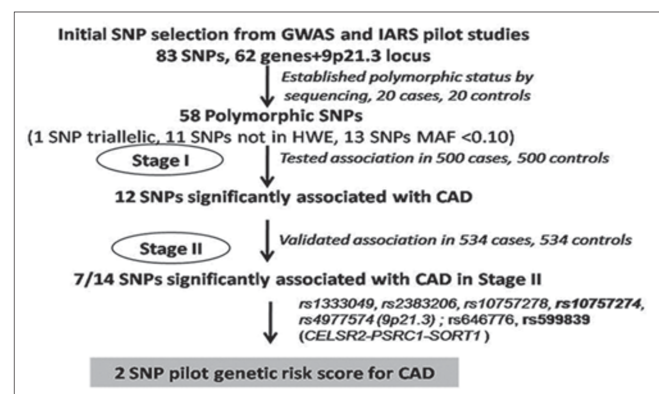

Figure 1: Overview of study design, single nucleotide polymorphism (SNPs) marked in bold represent the two genetic risk score SNP

Table 1: Clinical profile of Stage 1 and Stage 2 cohorts

Variables	Stage 1 (*n*=1000)		*P* value	Stage 2 (*n*=1068)		*P* value
	Cases	Controls		Cases	Controls	
Age, years	48.5±0.3	48.6±0.3	0.71	51.6±0.4	51.5±0.4	0.97
Hypertension *n* (%)	266 (53.2)	76 (15.2)	<0.0001	293 (54.9)	86 (16.1)	<0.0001
Diabetes *n* (%)	228 (45.6)	67 (13.4)	<0.0001	225 (42.1)	102 (19.1)	<0.0001
Smoking *n* (%)	190 (38.1)	106 (21.3)	<0.0001	216 (40.4)	126 (23.6)	<0.0001
Systolic BP, mmHg	122.8±0.8	124.9±0.8	0.067	123.5±0.8	127.2±0.8	0.001
Diastolic BP, mmHg	80.3±0.5	82.3±0.5	0.001	80.1±0.4	81.7±0.4	0.008
FBS, mg/dL (mmol/L)	120.5±2.7 (6.68±0.14)	97.2±2.4 (5.39±0.13)	<0.0001	127.5±2.5 (7.07±0.13)	106.2±2.0 (5.89±0.11)	<0.0001
BMI, kg/m^2	26.0±0.2	25.5±0.2	0.027	25.9±0.2	25.5±0.2	0.10
WHR	0.95±0.004	0.94±0.003	0.16	0.95±0.004	0.96±0.003	0.003
WC, cm	90.2±0.5	89.4±0.5	0.26	91.4±0.5	89.7±0.5	0.014
TC, mg/dL (mmol/L)	151.8±1.9 (8.42±0.10)	176.9±1.9 (9.81±0.10)	<0.0001	152.8±1.7 (8.48±0.09)	176.7±1.7 (9.80±0.09)	<0.0001
TG, mg/dL (mmol/L)	167.5±4.4 (9.29±0.24)	162.0±4.4 (8.99±0.24)	0.37	164.0±4.7 (9.10±0.26)	171.7±4.8 (9.52±0.26)	0.25
HDL-C, mg/dL (mmol/L)	37.4±0.4 (2.07±0.02)	42.0±0.4 (2.33±0.02)	<0.0001	37.5±0.4 (2.08±0.02)	38.7±0.4 (2.14±0.02)	0.034
LDL-C, mg/dL (mmol/L)	81.0±1.6 (4.49±0.08)	103.7±1.6 (5.75±0.08)	<0.0001	82.4±1.4 (4,57±0.07)	105.7±1.5 (5.86±0.08)	<0.0001
Statins *n* (%)	368 (73.6)	–	–	404 (75.7)	–	–

Data are mean±SE or n (%), BMI: Body mass index; BP: Blood pressure; HDL-C: High-density lipoprotein cholesterol, LDL-C: Low-density lipoprotein cholesterol, SE: Standard error, WC: Waist circumference, WHR: Waist hip ratio, FBS: Fasting blood sugar

concentrations of DNA sample in 20 cases and 20 controls. Oligonucleotide primers were designed using 'RExPrimer', a web-based tool and the regions of interest was amplified on a 9700 polymerase chain reaction (PCR) instrument. The PCR products were purified by ExoSAPit digestion (Amersham Biosciences, Piscataway, USA), sequenced bi-directionally using Big Dye Terminator v3.1 sequencing chemistry and analyzed on a 3130XL automated genetic analyzer, with SeqScape v2.5 software (Applied Biosystems, Foster City, USA). Details of the sequencing primers and the size of PCR product for the 88 SNPs is shown in Supplementary Table 1. Around 31 SNPs were directly taken up for genotyping by Taqman assay without prior sequencing in Stage 1, while five SNPs were prioritized based on our own internal discoveries. After excluding five SNPs that were invariant during the sequencing stage, and one triallelic SNP, 82 SNPs were selected in Stage 1 for genotyping by Taq Man allelic discrimination assay on HT 7900 Real Time PCR instrument (Applied Biosystems, Foster city, USA). Eight wells with positive in-house control samples having known genotypes and four 'no template control' wells were run with each 384-well experiment for quantity and quality assessment. Genotypes were confirmed by sequencing in a random selection of 48 samples for each SNP, ensuring representation of all the possible genotypes.

Statistical Methods

Routine statistical analysis was carried out using SPSS v17.0 software (SPSS Inc, Chicago, USA). Results are expressed as mean ± standard error for the continuous variables except age that was expressed as mean ± standard error of the mean. Chi-square test and binary logistic regression were used for testing the association of SNPs with CAD and estimating the odds ratios (ORs) and 95% confidence intervals (CIs), while Student's *t*-test, univariate and multivariate analysis were used to test for the mean differences in quantitative traits between the cases and controls. Age, gender, diabetes, hypertension, smoking, TC and HDL-C were treated as covariates and appropriately adjusted for during analysis. SNP Stats, an online software, was also used to estimate the allele frequency, genotype frequency, hardy weinberg equilibrium (HWE) and the association of genotypes and haplotypes with CAD [24].

The 'PredictABEL' package version 1.2 in 'R' statistical software (http://www.genabel.org) was used to assess the performance and utility of the various risk-prediction models that included the conventional RFs and SNPs [25]. GRS was generated using either the additive allele count method where the number of risk alleles carried by an individual for each SNP (0, 1 or 2) was simply counted (unweighted GRS - GRS_{UW}) or using the β-coefficient values derived from binary logistic regression analysis (weighted GRS - GRS_W). Correlations among the SNPs within a gene cluster (linkage disequilibrium [LD]) was estimated by Haploview v3.32 software (http://www.broad.mit.edu/mpg/haploview/) [26].Bonferroni correction was done using SNPassoc program in "R" package [27].

Accuracy of the discrimination was assessed based on area under curve the receiver operating characteristic curve (AUC or c index) for the three different models: Model 1, conventional RFs alone; Model 2, GRS alone; and Model 3, conventional RFs (Model 1) and GRS (Model 2) combined. Age, gender, diabetes, hypertension, smoking and log values of TC, TG, HDL-C and LDL-C were treated as conventional RFs in Model 1. Body mass index, waist circumference, hip circumference and waist-hip ratio did not show a significant association in regression analysis and were, therefore, not included for further analysis. Three types of analysis were performed based on the combination of the conventional RFs used in Model 1. In analysis 1, all of the above RFs were included; in analysis 2, lipids were excluded whereas, in Model 3, only diabetes, hypertension and smoking were considered. Significant difference in the AUCs between Model 1 and Model 3 was calculated using the De Long method [28]. The net reclassification index (NRI) and IDI scores were calculated by comparing the AUC generated under the different risk prediction models [11]. Calibration of the prediction models was tested using Hosmer–Lemeshow test [29].

Results

Clinical Characteristics of the Study Population

The clinical profile of study participants in Stage 1 and Stage 2 is shown in [Table 1]. The mean age of the cases and controls were comparable. Cases showed a higher frequency of diabetes, hypertension and smoking when compared to the controls. Statin usage was recorded only among the cases while data were unavailable for 1 case and 30 control subjects

Building a GRS for CAD

Of the 88 SNPs in the initial list, one was tri-allelic (rs3091244). Five SNPs were found to be non-polymorphic at the sequencing stage. Of the remaining 82 SNPs that were investigated in Stage I, 11 SNPs were not in HWE in the controls ($P < 0.05$), while 13 SNPs showed a low minor allele frequency (MAF) (8 SNPs with MAF 0.01 to 0.05 and 5 SNPs with MAF 0.06 to 0.09) and were therefore not considered for further analysis. A total of 58 genetic variants (MAF >0.10) were finally analyzed for association with CAD in Stage 1. An overview of the study design and the significant SNPs in the two stages is depicted in [Figure 1].

Stage 1 Analysis

Out of 58 SNPs, 12 SNPs showed a significant association with CAD. Five of them belonged to the well-established CAD risk locus on the 9p21.3 region (rs1333049, rs2383206, rs10757278, rs10757274, rs4977574), two belonged to the CELSR2-PSRC1-SORT1 gene cluster on 1p13.1 locus (rs646776, rs599839), while there was one SNP each of CXCL12 (rs501120), SLC22A3-LpPLA2-LPA gene cluster (rs3127599), CNNM2 (rs12413409), SMAD3 (rs17228212) and BCAP29 (rs10953541). Table 2 provides the details of the 12 SNPs, their associated genes/loci, MAF, allele and genotype frequencies and the associated ORs.

Stage 2 Analysis

In Stage 2, seven out of the twelve significant SNPs from Stage 1, 5 from the 9p21.3 locus (rs1333049, rs2383206, rs10757278, rs10757274, rs4977574) and two from CELSR2-PSRC1-SORT1 cluster (rs646776, rs599839) retained significant association with CAD. Since there was a strong correlation (LD) among the variants in the 9p21.3 and CELSR2-PSRC1-SORT1 loci, one best representative variant from each cluster i.e. rs10757274 (OR 1.7, 95% CI 1.19 -2.35) from the former and rs599839 (OR 1.51, 95% CI 0.94 - 2.43) from the latter were selected and constituted the pilot GRS. Mean GRS_{UW} (2.63 ± 0.04 vs. 2.39 ± 0.043) and weighted GRS_{W} (0.845 ± 0.012 vs. 0.770 ± 0.01) were significantly higher in cases than in the controls ($P < 0.0001$).

Table 3 shows a summary of the AUC values, the NRI and the IDI scores for the different combination of conventional RFs in Model 1, Model 2 and Model 3 along with the corresponding NRI and AUC values while [Figure 2] depicts the ROC plot for Model 1 (all RFs), Model 2 (2 GRS SNPs) and Model 3 (conventional RFs + SNPs), respectively. There was 24% net re-classification of subjects in analysis 1 where the Model 1 included age, gender, diabetes, hypertension, smoking and the log-transformed lipid levels, namely TC, TG, HDL-C and LDL-C. However, the increment in the AUC from Model 1 to Model 3 was relatively better for analysis 2, which included age, gender, diabetes, hypertension and smoking (excluding lipids) (0.112), as compared to analysis 1, which included all the common RFs (0.071) or analysis 3, which included only diabetes, hypertension and smoking (0.105). Overall, the net reclassification of subjects was better for analysis 1 (NRI = 0.2397) when compared to analysis 2 (NRI = 0.1442) or analysis 3 (NRI = 0.0974), respectively. The predictive probabilities generated using Model 3 showed better distribution when compared to either Model 1 or Model 2 across all the three types of analysis and the data for analysis 1 is shown [Figure 3a-c].

LD Analysis of 9p21.3 Common Variants and CELSR2-PSRC1-SORT1 SNPs

We have previously reported the presence of a strong pair-wise LD among the five variants in the 9p21.3 locus (r^2 0.93-0.99) [30] and the two SNPs in the CELSR2-PSRC1-SORT1 cluster (r^2 0.98) [31].

Table 2: Allelic and genotype frequency of SNPs significantly associated with CAD in Stage 1

SNP ID	Gene name	Locus	SNP*	Allele frequency		Genotype†			OR (95% CI)
				1	2	11	12	22	
rs1330049	ANRIL	9p21.3	C>G	0.54	0.46	0.28	0.52	0.20	1.44 (1.20-1.72)
rs2383206	ANRIL	9p21.3	G>A	0.56	0.44	0.30	0.52	0.18	1.51 (1.26-1.80)
rs10757278	ANRIL	9p21.3	G>A	0.56	0.44	0.29	0.53	0.18	1.36 (1.14-1.62)
rs10757274	ANRIL	9p21.3	G>A	0.54	0.46	0.28	0.52	0.20	1.45 (1.22-1.73)
rs4977574	CDKN2A/2B, ANRIL	9p21.3	G>A	0.54	0.46	0.28	0.52	0.20	1.43 (1.20-1.70)
rs646776	CELSR2-PSRC1-SORT1	1p13.1	T>C	0.74	0.26	0.54	0.38	0.08	1.29 (1.06-1.58)
rs599839	CELSR2-PSRC1-SORT1	1p13.1	A>G	0.73	0.27	0.55	0.37	0.08	1.33 (1.09-1.63)
rs501120	CXCL12	10q11.21	T>C	0.64	0.36	0.43	0.43	0.14	1.21 (1.00-1.45)
rs3127599	SLC22A3-LpPLA2-LPA	6q21-26	C>T	0.79	0.21	0.62	0.33	0.05	1.25 (1.01-1.55)
rs12413409	CNNM2	10q24.32	G>A	0.81	0.19	0.66	0.30	0.04	0.46 (0.24-0.90)
rs17228212	SMAD3	1q41	T>C	0.88	0.12	0.78	0.21	0.02	1.33 (1.03-1.72)
rs10953541	BCAP29	15q22.33	C>T	0.84	0.16	0.71	0.26	0.03	0.73 (0.55-0.97)

*Underlined allele denotes minor allele. †All SNPs were in hardy-weinberg equilibrium, CAD: Coronary artery disease, CI: Confidence interval, ID: Identification, OR: Odds ratio; SNP: Single-nucleotide polymorphism

Table 3: AUC, NRI and IDI for the various risk prediction models

Model	AUC, 95% CI	Difference in AUC between Model 1 and Model 3	DeLong test (*P* value)	NRI, 95% CI (*P* value)	IDI, 95% CI (*P* value)
Analysis 1					
Model 1 (age, gender, diabetes, hypertension, smoking, lipids)	0.766 (0.738-0.794)	0.071	Z=−6.605 ($P<0.0001$)	0.2397, 0.1881-0.2913 ($P<0.0001$)	0.1347, 0.1143-0.1550 ($P<0.0001$)
Model 2 (SNPs alone)	0.576 (0.542-0.609)				
Model 3 (Model 1+Model 2)	0.837 (0.813-0.861)				
Analysis 2					
Model 1 (age, gender, diabetes, hypertension, smoking)	0.678 (0.646-0.710)	0.112	Z=−7.7889 ($P=6.76\times10^{-15}$)	0.1442, 0.1032-0.1852 ($P<0.0001$)	0.1588, 0.1370-0.1805 ($P=0.0001$)
Model 2 (SNPs alone	0.576 (0.542-0.609)				
Model 3 (Model 1+Model2)	0.790 (0.762-0.817)				
Analysis 3					
Model 1 (diabetes, hypertension, smoking)	0.681 (0.650-0.713)	0.105	Z=−7.4221 ($P=1.15\times10^{-13}$)	0.0974, 0.0637-0.1310 ($P<0.0001$)	0.154, 0.1325-0.1756 ($P<0.00001$)
Model 2 (SNPs alone)	0.576 (0.542-0.609)				
Model 3 (Model 1+Model3)	0.786 (0.759-0.813)				

AUC: Area under curve, NRI: Net reclassification index, IDI: Integrated discrimination improvement, CI: Confidence interval, SNP: Single-nucleotide polymorphism

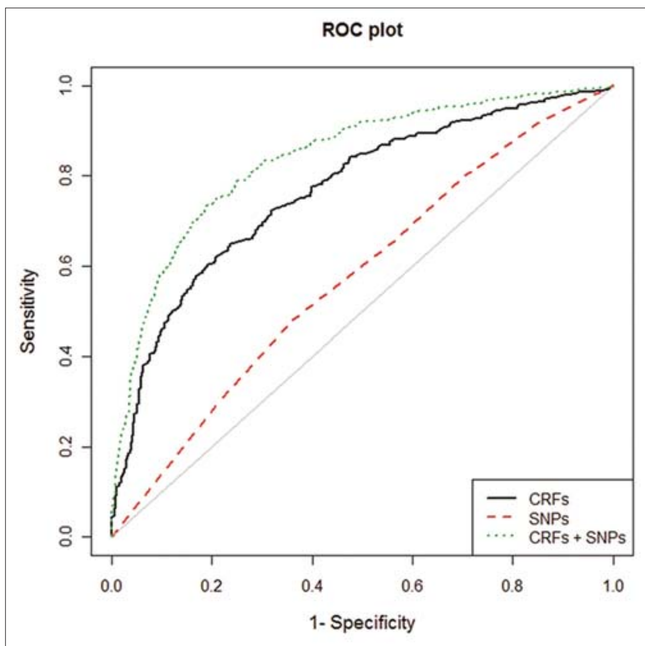

Figure 2: Depiction of receiver operating characteristic plot for the three risk prediction models Model 1: Conventional risk factors (CRFs) alone (AUC = 0.766); Model 2: Two single nucleotide polymorphisms (SNPs) alone (area under curve [AUC] = 0.576); Model 3 conventional risk factors + SNPs combined (AUC = 0.837)

CELSR2-PSRC1-SORT1 Gene Cluster and Lipid Levels

There was a significant association of SNPs rs599839 (A>G) and rs646776 (T>C) with TC and LDL-C levels in the cases in Stage 1 and in both cases and controls in Stage 2. Here, the AA or TT homozygote was associated with significantly higher lipid levels than either GG or CC, while AG or TC heterozygote showed intermediary levels [Table 4]. Age, gender and statins were used as covariates in the multivariate analysis.

Effect of Statins on Plasma Lipid Levels

About statin usage, more than 70% of CAD patients were prescribed statins while none of the controls either in Stage 1 or in Stage 2 were on the drug. As expected, the mean levels of TC and LDL-C was higher in the non-statin group as compared to those on statin medication as shown in the supplementary [Table 2].

Discussion

Genetic risk models can augment the power of the conventional RFs to identify suitable 'high-risk' candidates for aggressive therapy. Using a two-stage study design, from the initial panel of 88 SNPs, we identified seven SNPs to be significantly associated with CAD. Further, two representative SNPs, rs10757274 and rs599839, one each from the 9p21.3 locus and the cholesterol locus, respectively, constituted our pilot GRS. The use of two independent datasets, one for training and the other for validating the initial findings, is a recommended method for assessing the performance of a new biomarker/algorithms for risk stratification [32].

The genetic variants used in the construction of a risk model can play a critical role in risk stratification. The initial panel of 88 SNPs was obtained from published GWAS that showed a genome-wide significance ($P < 1 \times 10^{-6-8}$), was validated across independent populations and performed well in meta-analysis [8,19,33]. A few SNPs (rs6046, rs5128 and rs10757278) were also included from internal discovery [20-22]. Interestingly, these 88 SNPs belonged to predominantly four biological pathways, namely lipids and lipoprotein metabolism, inflammation and immune response, cell proliferation and thrombosis. Other pathways were endothelial integrity, oxidation-reduction state and apoptosis, while the exact function of a few novel genes (KIAA1462, MIA3, PHACTR1 etc.) is yet to be elucidated. Nonetheless, only seven of these

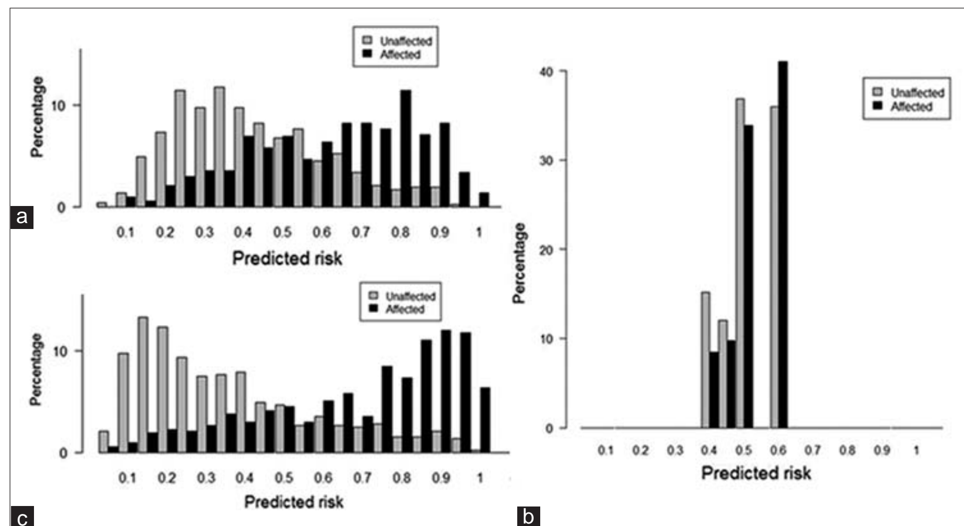

Figure 3: Distribution of the predicted risk probabilities in cases and controls (a) Model 1 includes the six most common conventional risk factors (age, gender, diabetes, hypertension, smoking, lipids) alone; (b) Model 2 includes the genetic risk score single nucleotide polymorphism (SNPs) alone; (c) Model 3 includes both the conventional risk factors and SNPs

Table 4: Mean difference in plasma lipid levels across rs599839 and rs646776 genotypes

Stage 1 (*N*=500) (cases only)				
rs599839 – A>G	AA (*n*=290)	AG (*n*=177)	GG (*n*=31)	*P* value
TC (mg/dl)	156.06±2.93	147.94±2.74	137.10±6.12	0.027
LDL-C (mg/dl)	83.74±2.57	78.32±2.43	73.16±5.11	NS
rs646776 – T>C	TT (*n*=287)	TC (*n*=182)	CC (*n*=29)	*P* value
TC (mg/dl)	155.56±2.88	148.47±2.91	138.79±6.43	0.07
LDL-C (mg/dl)	84.01±2.58	77.70±2.39	74.14±5.42	NS
Stage 2 (N=1068) (cases and controls)				
rs599839 – A>G	AA (*N*=553)	AG (*N*=422)	GG (*N*=78)	*P* value
TC (mg/dl)	167.97±1.76	162.18±2.02	154.90±3.94	0.009
LDL-C (mg/dl)	97.17±1.50	90.63±1.76	86.54±3.30	0.003
rs646776 – T>C	TT (*N*=561)	TC (*N*=416)	CC (*N*=76)	*P* value
TC (mg/dl)	167.96±1.75	162.49±2.05	152.49±3.84	0.003
LDL-C (mg/dl)	97.29±1.49	90.61±1.79	84.92±3.28	0.001

All lipid values are shown as mean±SE, SE: Standard error, LDL-c: Low density lipoprotein-cholesterol, TC: Total cholesterol

variants emerged as significant after Stage 2 analysis - five variants in the 9p21.3 and two SNPs in the CELSR2-PSRC1-SORT1 locus.

The 9p21.3 locus is one of the best-replicated regions for CAD, showing consistent association across different geographical boundaries [8]. Growing evidence from *in vitro* and *in vivo* studies indicate that ANRIL, a non-coding RNA located within this region, closely interacts with the neighboring tumor suppressor genes, CDKN2A and CDKN2B, and regulate important cellular processes such as proliferation and cell cycle regulation through epigenetic mechanisms [30,34]. Furthermore, studies including the present one have shown that the association of the 9p21.3 common variants with CAD is independent of the well-known RFs [35]. The robust association between rs10757274 and CAD has been previously demonstrated in other populations [36-38]. The CELSR2-PSRC1-SORT1 gene cluster on 1p13.1 chromosomal region is also a highly replicated locus for CAD. GWAS on lipid traits [39] and CAD [40] provide strong evidence on the importance of this locus in CAD development, wherein the genetic variants appear to modulate risk through a well-established cardiovascular risk factor, namely high cholesterol [41]. We have previously reported that the common alleles, 'T' in rs646776 or 'A' in rs599839, are associated with high TC and LDL-C levels [31].

Apart from the above-mentioned variants, five other variants, CXCL12 (rs501120), SLC22A3-LpPLA2-LPA gene cluster (rs3127599), CNNM2 (rs12413409), SMAD3 (rs17228212) and BCAP29 (rs10953541) showed significant association with CAD in Stage 1 analysis. These variants were identified through GWAS [35,42-45]. However, none of them retained significance in Stage 2 analysis. Furthermore, the variants that had previously demonstrated significant association with CAD in our own pilot studies in subsets of the IARS cohort [21,22], could not be replicated beyond Stage 1. The probable reasons could be either the presence of ethnic specific association patterns since most of the selected variants were initially reported on Caucasians, modest cohort size or due to low effect size of the risk alleles.

The type and number of SNPs used to build a GRS have varied across different studies: <10 [14-47], between 10 and 20 [13,11], or even >100 [12] SNPs, with no particular correlation between the type of SNPs included in the model and their performance. A GRS defined by 48 high-risk alleles was shown to predict Major Adverse Cardiac Events [48]. In the present study, although seven SNPs were independently replicated in Stage I and II, they were in strong LD. Therefore, two of the best-associated SNPs, one from each cluster, were included in the pilot GRS.

Although a simple addition of the risk alleles is the most popular method for constructing a GRS (GRS_{UW}) [10,49], where equal weightage is given to each risk allele, the weighted

GRS model uses β-coefficient values generated from the logistic regression function and gives different weightage to the individual SNPs [50,51]. Some studies have considered both the methods [14]. In the present study too we have assessed the performance of GRS using the unweighted and weighted methods and found both of them to perform well with higher mean values seen in the cases than in the controls. Further, subjects in the top GRS quartile showed > 2-fold risk of CAD when compared to the bottom quartile after adjusting for the known RFs. This is comparable to the other reports that have shown 1.6-2.2 times increased risk of CAD [13,46].

Arguably, statistical parameters such as c index (measure of discrimination of predictive model), NRI (measure of clinical utility) and IDI (percentage of total reclassification) have been widely used to evaluate the robustness of a risk-prediction model. We used the above metrics to estimate significant increase in AUCs between the standard and updated risk prediction models. Conventionally, the base model (Model 1) includes the conventional RFs. Therefore, we used different combinations of the six factors namely age, gender, diabetes, hypertension, smoking and lipids in Model 1. While the addition of all of the above-mentioned RFs yielded a significant net reclassification of around 24%, exclusion of lipids from Model 1 (analysis 2) showed a modest increment in AUC of 0.112 between Model 1 and Model 3. Based on the NRI and IDI values, Model 1 that included all the conventional RFs was considered as the best predictive model in this study. Some studies have shown no incremental difference [13], while others have shown modest improvement in c statistics, particularly when using the weighted GRS Model [52]. In the present study, both NRI and IDI were statistically significant in Stage 2 across the different combinations of the classical RFs. There was 24% net reclassification of subjects having intermediate risk in the updated Model 3 that included the two GRS SNPs and all the six conventional RFs. Such a modest reclassification implies that the SNPs selected for the GRS carry a low power of discrimination and association, which is an important consideration in a small cohort such as in the present study. In fact, very few studies have shown substantial improvement in the net reclassification of subjects [11]. A study involving a 5-SNP GRS showed marginal improvement in the c index and a 28% net reclassification of subjects in the intermediate-risk group [14], while the REGICOR (Registre Gironi del cor) and Framingham Heart Study have shown up to 17.4% net reclassification [46]. Studies involving a prospective cohort could hold the key toward understanding the robustness of reclassification based on prevailing knowledge on the genetic variants associated with CAD and the traditional RFs in a primary clinical setting [53].

While factors such as stringent phenotypic classification can influence the association of SNPs, other factors such as sample size, risk allele frequency and the relative importance of the selected variants located within genes governing critical biological functions can influence the performance of a GRS. A classic example is the present study where the 9p21.3 common variants and the SNPs in the cholesterol locus were validated. Simulation studies have shown that in order to achieve a c statistic between 0.80 and 0.85, around 100 uncorrelated genetic variants with relative risk of 1.5 and MAF of 10% that explain ~20% heritability of cardiovascular disease would be required for attaining a good discrimination [54]. Nevertheless, it is said that a higher increment in AUC score can be achieved with a liberal inclusion of SNPs rather than the stringent inclusion of only top performing SNPs [55]. There is an ongoing debate as to whether the available repertoire of genetic variants carries sufficient power of discrimination or whether the available statistical tools are insensitive to the effects of the small contributions made by these individual alleles towards the classification of the risk groups. The era of next generation sequencing technology holds great promise to unearth novel markers with low MAF that may escape detection in routine genetic analysis and thus fill the missing gaps in the heritability of CAD.

The strength of this study include a two-stage study design for the discovery and validation of the significant SNPs in a hitherto untested young population, (average age ~50 years), a matching case–control group, selection of SNPs from diverse biological pathways and a robust statistical analysis with appropriate adjustment for the potential confounders. However, we acknowledge certain shortcomings, such as the selection of discovery and validation cohort from the same IARS genetic pool and the construction of a pilot GRS based only on 2 SNPs. Although in an ideal scenario, the initial panel of 88 SNPs should have been genotyped on the entire cohort of 1034 cases and 1034 controls to enhance the likelihood of discovery, a 2-stage study design was adopted keeping the cost considerations in mind. In fact, the potential SNPs that showed borderline significance (rs20455, rs3127599) were eliminated in Stage 1 itself. Population stratification is another critical issue that can have confounding effect on disease association patterns, particularly in the people of Indian origin. Matters such as early human migration combined with distinct social boundaries, rigid endogamy practices, and evolutionary forces have played a critical role in building the diverse and complex genetic architecture of the present-day Indian population [56]. Such factors may pose some difficulties towards understanding the genetic susceptibility to complex diseases in Indians.

CONCLUSION

The 9p21.3 common variants and SNPs near the CELSR2-PSRC1-SORT1 cluster emerged as the best genetic markers for CAD in our study. It is just a question of time before other similar robust markers are unearthed, which could lead to a more sensitive genetic risk prediction model. The foundation for building the genetic architecture of cardiovascular disease has now been laid with discoveries through GWAS, while the next generation sequencing technologies are helping to erect a scaffold and functional underpinnings is soon to follow. Hand in hand with the reduction in cost of operating these sophisticated platforms, accessibility to query larger and more diverse populations will become possible, leading to enhanced power to detect novel genetic elements and thereby, ushering in an era of personalized medical diagnosis and therapy.

Acknowledgments

We gratefully acknowledge the support extended by the trustees of the Thrombosis Research Institute, London and Bangalore and the financial support received from the Weston Foundation, UK, Foundation Bay, Switzerland, the Tata Social Welfare Trust, India (TSWT/IG/SNB/JP/Sdm) and Department of Biotechnology, Ministry of Science and Technology, Government of India (BT/01/CDE/08/07), Emmanuel Kaye foundation, UK and Bharti foundation, India. We thank the study participants and the doctors from Narayana Health, Bangalore and the Asian Heart Institute and Research Centre, Mumbai for their valuable cooperation. The sponsors did not participate in the design, conduct, sample collection, analysis and interpretation of the data or in the preparation, review or approval of the manuscript. A special thanks to Ankit Sharma (Thrombosis Research Institute, India) for his assistance in running the 'R' statistical program.

References

1. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study. BMJ 2007;335:136.
2. Ridker PM, Paynter NP, Rifai N, Gaziano JM, Cook NR. C-reactive protein and parental history improve global cardiovascular risk prediction: The Reynolds Risk Score for men. Circulation 2008;118:2243-51, 4p following 2251.
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;97:1837-47.
4. Lenfant C. Shattuck lecture – clinical research to clinical practice – lost in translation? N Engl J Med 2003;349:868-74.
5. Khot UN, Khot MB, Bajzer CT, Sapp SK, Ohman EM, Brener SJ, *et al.* Prevalence of conventional risk factors in patients with coronary heart disease. JAMA 2003;290:898-904.
6. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 2009;106:9362-7.
7. Holdt LM, Teupser D. Recent studies of the human chromosome 9p21 locus, which is associated with atherosclerosis in human populations. Arterioscler Thromb Vasc Biol 2012;32:196-206.
8. Palomaki GE, Melillo S, Bradley LA. Association between 9p21 genomic markers and heart disease: A meta-analysis. JAMA 2010;303:648-56.
9. Davies RW, Dandona S, Stewart AF, Chen L, Ellis SG, Tang WH, *et al.* Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies. Circ Cardiovasc Genet 2010;3:468-74.
10. Lluís-Ganella C, Lucas G, Subirana I, Sentí M, Jimenez-Conde J, Marrugat J, *et al.* Additive effect of multiple genetic variants on the risk of coronary artery disease. Rev Esp Cardiol 2010;63:925-33.
11. Thanassoulis G, Peloso GM, Pencina MJ, Hoffmann U, Fox CS, Cupples LA, *et al.* A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: The Framingham Heart Study. Circ Cardiovasc Genet 2012;5:113-21.
12. Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, Miletich JP, *et al.* Association between a literature-based genetic risk score and cardiovascular events in women. JAMA 2010;303:631-7.
13. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, *et al.* A multilocus genetic risk score for coronary heart disease: Case-control and prospective cohort analyses. Lancet 2010;376:1393-400.
14. Anderson JL, Horne BD, Camp NJ, Muhlestein JB, Hopkins PN, Cannon-Albright LA, *et al.* Joint effects of common genetic variants from multiple genes and pathways on the risk of premature coronary artery disease. Am Heart J 2010;160:250-256.e3.
15. Shanker J, Maitra A, Rao VS, Mundkur L, Dhanalakshmi B, Hebbagodi S, *et al.* Rationale, design & preliminary findings of the Indian Atherosclerosis Research Study. Indian Heart J 2010;62:286-95.
16. Kumar NK. Bioethics activities in India. East Mediterr Health J 2006;12 Suppl 1:S56-65.
17. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972;18:499-502.
18. Ogawa N, Imai Y, Morita H, Nagai R. Genome-wide association study of coronary artery disease. Int J Hypertens 2010;2010:790539.
19. Peden JF, Farrall M. Thirty-five common variants for coronary artery disease: The fruits of much collaborative labour. Hum Mol Genet 2011;20:R198-205.
20. Maitra A, Dash D, John S, Sannappa PR, Das AP, Shanker J, *et al.* A common variant in chromosome 9p21 associated with coronary artery disease in Asian Indians. J Genet 2009;88:113-8.
21. Shanker J, Perumal G, Maitra A, Rao VS, Natesha BK, John S, *et al.* Genotype-phenotype relationship of F7 R353Q polymorphism and plasma factor VII coagulant activity in Asian Indian families predisposed to coronary artery disease. J Genet 2009;88:291-7.
22. Shanker J, Perumal G, Rao VS, Khadrinarasimhiah NB, John S, Hebbagodi S, *et al.* Genetic studies on the APOA1-C3-A5 gene cluster in Asian Indians with premature coronary artery disease. Lipids Health Dis 2008;7:33.
23. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.
24. Solé X, Guinó E, Valls J, Iniesta R, Moreno V. SNPStats: A web tool for the analysis of association studies. Bioinformatics 2006;22:1928-9.
25. Kundu S, Aulchenko YS, van Duijn CM, Janssens AC. PredictABEL: An R package for the assessment of risk prediction models. Eur J Epidemiol 2011;26:261-4.
26. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.
27. González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, *et al.* SNPassoc: An R package to perform whole genome association studies. Bioinformatics 2007;23:644-5.
28. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 1988;44:837-45.
29. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 1997;16:965-80.
30. Shanker J, Arvind P, Jambunathan S, Nair J, Kakkar V. Genetic analysis of the 9p21.3 CAD risk locus in Asian Indians. Thromb Haemost 2014;111:960-9.
31. Arvind P, Nair J, Jambunathan S, Kakkar VV, Shanker J. CELSR2-PSRC1-SORT1 gene expression and association with coronary artery disease and plasma lipid levels in an Asian Indian cohort. J Cardiol 2014;64:339-46.
32. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MS, *et al.* Criteria for evaluation of novel markers of cardiovascular risk: A scientific statement from the American Heart Association. Circulation 2009;119:2408-16.
33. Preuss M, König IR, Thompson JR, Erdmann J, Absher D, Assimes TL, *et al.* Design of the Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Study: A Genome-wide association meta-analysis involving more than 22 000 cases and 60 000 controls. Circ Cardiovasc Genet 2010;3:475-83.
34. Congrains A, Kamide K, Oguro R, Yasuda O, Miyata K, Yamamoto E, *et al.* Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. Atherosclerosis 2012;220:449-55.
35. Schunkert H, Götz A, Braund P, McGinnis R, Tregouet DA, Mangino M, *et al.* Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. Circulation 2008;117:1675-84.
36. Esparragón FR, Companioni O, Bello MG, Ríos NB, Pérez JC. Replication of relevant SNPs associated with cardiovascular disease susceptibility obtained from GWAs in a case-control study in a Canarian population. Dis Markers 2012;32:231-9.
37. Shen GQ, Li L, Rao S, Abdullah KG, Ban JM, Lee BS, *et al.* Four

SNPs on chromosome 9p21 in a South Korean population implicate a genetic locus that confers high cross-race risk for development of coronary artery disease. Arterioscler Thromb Vasc Biol 2008;28:360-5.
38. Yamagishi K, Folsom AR, Rosamond WD, Boerwinkle E, ARIC Investigators. A genetic variant on chromosome 9p21 and incident heart failure in the ARIC study. Eur Heart J 2009;30:1222-8.
39. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, *et al*. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet 2008;40:189-97.
40. Zeller T, Blankenberg S, Diemert P. Genomewide association studies in cardiovascular disease – an update 2011. Clin Chem 2012;58:92-103.
41. Samani NJ, Braund PS, Erdmann J, Götz A, Tomaszewski M, Linsel-Nitschke P, *et al.* The novel genetic variant predisposing to coronary artery disease in the region of the PSRC1 and CELSR2 genes on chromosome 1 associates with serum cholesterol. J Mol Med (Berl) 2008;86:1233-41.
42. IBC K CAD Consortium. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. PLoS Genet 2011;7:e1002260.
43. Myocardial Infarction Genetics Consortium, Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 2009;41:334-41.
44. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, *et al*. Genomewide association analysis of coronary artery disease. N Engl J Med 2007;357:443-53.
45. Trégouët DA, König IR, Erdmann J, Munteanu A, Braund PS, Hall AS, *et al.* Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. Nat Genet 2009;41:283-5.
46. Lluis-Ganella C, Subirana I, Lucas G, Tomás M, Muñoz D, Sentí M, *et al.* Assessment of the value of a genetic risk score in improving the estimation of coronary risk. Atherosclerosis 2012;222:456-63.
47. Qi L, Ma J, Qi Q, Hartiala J, Allayee H, Campos H. Genetic risk score and risk of myocardial infarction in Hispanics. Circulation 2011;123:374-80.
48. Andreassi MG, Adlerstein D, Carpeggiani C, Shehi E, Fantinato S, Ghezzi E, *et al.* Individual and summed effects of high-risk genetic polymorphisms on recurrent cardiovascular events following ischemic heart disease. Atherosclerosis 2012;223:409-15.
49. Yiannakouris N, Katsoulis M, Trichopoulou A, Ordovas JM, Trichopoulos D. Additive influence of genetic predisposition and conventional risk factors in the incidence of coronary heart disease: A population-based study in Greece. BMJ Open 2014;4:e004387.
50. Hernesniemi JA, Seppälä I, Lyytikäinen LP, Mononen N, Oksala N, Hutri-Kähönen N, *et al*. 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 2012;7:e28931.
51. Shah S, Casas JP, Gaunt TR, Cooper J, Drenos F, Zabaneh D, *et al.* Influence of common genetic variation on blood lipid levels, cardiovascular risk, and coronary events in two British prospective cohort studies. Eur Heart J 2013;34:972-81.
52. Lv X, Zhang Y, Rao S, Qiu J, Wang M, Luo X, *et al.* Joint effects of genetic variants in multiple loci on the risk of coronary artery disease in Chinese Han subjects. Circ J 2012;76:1987-92.
53. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, *et al*. Finding the missing heritability of complex diseases. Nature 2009;461:747-53.
54. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM. Predictive testing for complex diseases using multiple genes: Fact or fiction? Genet Med 2006;8:395-400.
55. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. Hum Mol Genet 2009;18:3525-31.
56. Tamang R, Singh L, Thangaraj K. Complex genetic origin of Indian populations and its implications. J Biosci 2012;37:911-9.

© SAGEYA. This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.

Source of Support: Nil, Conflict of Interest: None declared.