Cellular and Molecular Biology
E-ISSN : 1165-158X / P-ISSN : 0145-5680
www.cellmolbiol.org

Original Research

Common SNP-based haplotype analysis of the 9p21.3 gene locus as predictor coronary artery disease in Tanzanian population

Gokce Akan[1], Peter Kisenge[2], Tulizo Shemu Sanga[2], Erasto Mbugi[1], Ismael Adolf[1], Mehmet Kerem Turkcan[3], Mohammed Janabi[2], Fatmahan Atalar[1,4*]

[1]MUHAS Genetics Laboratory, Biochemistry Department, School of Medicine, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania
[2] Jakaya Kikwete Cardiac Institute, Dar es Salaam, Tanzania
[3] Department of Electrical Engineering, Columbia University, New York, USA
[4] Istanbul University, Child Health Institute, Department of Medical Genetics, Istanbul, Turkey

[*]**Correspondence to:** fatmahan.atalar@gmail.com

Received March 10, 2019; **Accepted** July 26, 2019; **Published** July 31, 2019

Doi: http://dx.doi.org/10.14715/cmb/2019.65.6.7

Abstract: Genome-wide association studies (GWAS) have recently confirmed a strong association of the 9p21.3 locus with Coronary Artery Disease (CAD) in different populations but no data has been reported for the Tanzanian population. This study aimed to investigate the 9p21.3 locus harboring the disease-causing hotspot variations in Tanzanian CAD patients and their associations with the risk factors. 135 patients with CAD and 140 non-CAD patients were enrolled into the study. Further the biochemical analysis, the genotyping assays were performed by the use of qRT-PCR. The genotype and allele frequencies of rs1333049, rs2383207, rs2383206, rs10757274, rs10757278, and rs10811656 were significantly different between the groups ($p<0.005$). The genotype distribution of rs1333049, rs10757278 and rs10811656 polymorphisms were significantly different among patients with one, two, three stenotic vessels ($p<0.05$). For rs10757274 and rs10757278, the GG genotype indicated a significant 3-fold and 4-fold increased risk of CAD ($p<0.0001$,respectively). Additionally, haplotype analysis revealed that AAGCAG, AAACAG, GGGTGC haplotypes of 9p21.3 locus polymorphisms are associated with CAD risk. The GGGTGC haplotype was over-represented while the other two underrepresented in patients as compared to controls ($p<0.00001$,respectively) suggesting the first one a high-risk and the other two low-risk haplotypes for Tanzanian population. The AUC of a risk model based on non-genetic risk factors was 0.954 (95% CI: 0.930-0.977) and the combination with genetic risk factors improved the AUC to 0.982 (95% CI: 0.954-0.985) ($p<0.012$), indicating good diagnostic accuracy. Our results are the first data reporting statistically significant associations between 9p21.3 polymorphisms and CAD, and the very first haplotype block harboring the disease-causing variations in Tanzanian population.

Key words: Coronary artery disease; 9p21.3; Polymorphism; Haplotype; Tanzania.

Introduction

Coronary artery disease (CAD) has a higher incidence in developing countries and is known to be a major cause of death worldwide (1). In sub-Saharan Africa, CAD ranks fourth among the leading causes of death after infectious diseases (2, 3).

The hallmark of CAD is the built-up plaques in the arteries of large and medium-size through a process called atherosclerosis. The building up of the plaque is a complex process characterized by chronic inflammation (4, 5). Occlusion, thrombosis and stenosis are characteristic features of atherosclerosis that lead to its clinical outcome. The plaque can clog the microvasculature due to clotting which blocks blood flow with consequent rupture of such blood vessels. As a result, downstream tissues become devoid of oxygen and nutrients. If clotting develops in heart vessels it leads to heart attack, and when it is in brain blood vessels stroke ensues (5, 6).

CAD is a multi-factorial and heterogenic disease and many risk factors have been identified for CAD, including age, high blood pressure, dyslipidemia, diabetes, obesity, smoking and *etc*. Apart from these factors, genetic predisposition appears to be strongly involved in the pathogenesis of CAD (4, 5). It is important to understand the genetic basis of atherosclerosis to prevent CAD. A number of candidate genes have been identified to be responsible for CAD, but the exact identity of the candidate genes and the quantity of their effects on disease pathogenesis is not well known.

Recently, GWAS (7-9) and meta-analysis (10, 11) using hundreds of thousands of markers have identified several new loci which confer increased susceptibility to CAD and its progression. One prominent locus unveiled by such studies is 9p21.3. The 9p21.3 risk locus spans a >50-kb genomic locus and comprises 59 single nucleotide polymorphisms (SNPs) that are in strong linkage disequilibrium (LD). This locus is devoid of protein-coding genes, it overlaps a large nonprotein coding RNA in *INK4* locus, termed *"CDKN2B AS1 (ANRIL)"* and lies adjacent to a cluster of cell-cycle–regulating tumor-suppressor genes, including the cyclin-dependent kinase inhibitors, *CDKN2A* and *CDKN2B*. A number of these SNPs lying in 9p21.3 locus have been

studied in light of their association with CAD. Intronic polymorphisms rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656 were found to be associated with CAD (7-9, 12) These findings on possible association of aforementioned SNPs with CAD have been replicated in large scale by several studies in diverse populations (7, 9, 13-21) but so far no data from sub-Saharan Africans have been reported.

This is the first study investigating the relationship between 9p21.3 locus and CAD in Tanzanian population. The aim of the study was to investigate the relationship between the SNPs (rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656) on chromosome 9p21.3 and CAD susceptibility, along with the risk factors and severity of CAD in Tanzanian population.

Materials and Methods

Study population

The study was performed at Jakaya Kikwete Cardiac Institute located within the Muhimbili National Hospital, Dar es Salaam, Tanzania. Participants enrolled in the study were selected among patients admitted to the cardiology outpatient clinic for symptoms of angina, dyspnea and chest discomfort at the time of diagnosis. Study population included 135 patients (mean age 62.01±10.65) with CAD and 140 controls (mean age 58.21±12.62) without CAD. Control subjects, who underwent coronary angiography to rule out CAD, were also enrolled during the same time period at the same hospital. CAD was defined as ≥ 50% luminal narrowing in at least one coronary artery. Patients with CAD were divided into subgroups with one stenotic vessel, two stenotic vessels or three stenotic vessels disease according to the number of significantly affected vessels using the Coronary Artery Surgery Study classification (22). Two cardiologists who were blinded to all patient information assessed the angiograms. Subjects with CAD, HIV, viral hepatitis, cancer, collagen disease, endocrinopathies, secondary hypertension, diabetic microangiopathic complications, for female patients on birth control pills or postmenopausal were also excluded from the study. All subjects enrolled in this study were of Tanzanian origin.

Detailed information on demographics, medical history and coronary risk factors such as the presence of diabetes, hypertension, smoking, lifestyle and current medication of the participants' record were completed through personal interviews. Informed consent was obtained from each participant. Body mass index (BMI) was calculated as a ratio of mass in kilogram divided by the square of corresponding height in meters. Hypertension was defined as blood pressure ≥ 140/90 mm Hg or on the basis that patients were already being treated with antihypertensive drugs. Diabetes was defined either by 1999 World Health Organization criteria 11 or self-report of being previously diagnosed as diabetic. All subjects gave written consent after receiving a full explanation of the study. This study was approved by the Research Ethics Committee of Muhimbili University of Health and Allied Sciences (Ref. Number 2017-05-25/AEC/Vol.XII/66 Date 25.05.2017).

Biochemical parameters

Blood specimens were collected after an overnight fast of 12 hours by venipuncture using the vacutainer system in the plain bulb for serum. Blood was centrifuged and serum aliquots were preserved at −20 °C for one month prior to analysis. Serum total cholesterol (TC) and high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST) and Alkaline Phosphatase (ALP) were determined by routine enzymatic endpoint methods (Analyzer A15 Biosystems, Philippines). Fasting glucose was measured using the enzymatic reference method with glucose oxidase. Low-density lipoprotein-cholesterol (LDL-C) and Very Low-density lipoprotein-cholesterol (VLDL-C) were calculated according to Friedewald´s formula (23)

DNA extraction

From both groups, approximately 5 ml of blood were collected in falcon tubes containing 0.5 mg of ethylene diaminetetraacetic acid (EDTA) per ml as anticoagulant. Genomic DNA was obtained from 400 µl peripheral blood leukocytes with MagnaPure DNA Isolation robot (Roche, Germany). DNA quantities were determined by NanoDrop™ 1000 Spectrophotometer (Thermo Scientific, Wilmington, Delaware USA). The extracted DNA was stored at -20ºC before subsequent molecular analysis.

Genotyping

Genotyping of the rs1333049, rs2383207, rs2383206, rs10757274, rs10757278, and rs10811656 SNPs was performed by the use of Quantitative Real-Time PCR (QRT-PCR). Genotyping was carried out with the LightSNiP typing assay (TIBMolBiol, Berlin, Germany) with the LightCycler® 480 system instrument (Roche-Germany).

Statistical analysis

Statistical analysis was performed using SPSS software (Statistical Package for the Social Sciences, SPSS Inc, Chicago, IL, USA, version 25). The allelic frequency distributions of polymorphisms between the control and patient groups were compared using Chi square (x^2). Hardy-Weinberg equilibrium (HWE) was assessed by Fischer's exact test. For comparisons of differences between mean values between two groups unpaired Student *t*-test was used. To evaluate differences between groups, the data were log transformed to satisfy ANOVA criteria and then subjected to one-way ANOVA with Tukey's post hoc analysis. In all cases differences were considered significant at $p < 0.05$. Haplotypes were generated from the genotyped data. The linkage disequilibrium (LD) and haplotype analysis were performed using Haploview 4.2. Bonferroni correction was used to account for multiple testing and a two-tailed *p*-value <0.01 was considered statistically significant. For discrimination of CAD cases and healthy controls, the LightGBM library (24) was utilized for gradient boosting model fitting and the Python scientific stack (25) and scikit-learn (26) libraries were used for the analysis of the area under the curve (AUC) of the receiver operating characteristic (ROC) and to assess the difference between different areas under the curve

Table 1. Baseline characteristics of the study subjects.

Variables	CAD (n=135)	Control (n=140)	P Value*
Age (year)	62.01±10.65	58.21±12.62	0.08
Weight (kg)	89.42±11.9	73.68±13.74	***0.001***
Height (m)	1.60±0.087	1.65±0.087	***0.001***
BMI (kg/m^2)	34.78±4.43	26.18±3.72	***0.001***
Systolic BP (mmHg)	140.96±26.26	131.93±28.22	0.058
Diastolic BP (mmHg)	81.60±15.57	80.06±16.10	0.46
Glucose (mmol/L)	7.44±2.75	3.92±1.03	***0.001***
Cholesterol (mmol/L)	6.16±1.03	4.09±0.87	***0.001***
HDL (mmol/L)	0.90±0.25	1.34±0.30	***0.001***
LDL (mol/L)	4.06±0.96	2.68±0.69	***0.001***
VLDL (mmol/L)	1.20±0.46	0.36±0.19	***0.001***
TG (mmol/L)	2.35±1.48	1.10±0.48	***0.001***
ALT (U/L)	31,97±16.66	19.6±7.11	***0.001***
AST (U/L)	42.07±4.95	22.54±6.78	***0.002***
ALP (U/L)	74.6±24.2	56.53±22.6	0.078
Smoking (%) Yes/No	24% / 76%	5.2% / 94.8%	***0.001***
Sedentary Life (%) Yes/No	49% / 51%	24.1% / 75.9%	***0.001***
Family History (%) Yes/No	45.2% / 54.8%	29% / 86%	***0.002***

Data: mean ± SD *Comparisons of differences between mean values of two groups unpaired Student t-test was used.

(AUC) and to evaluate the diagnostic sensitivity and specificity. Models were trained with 5000 estimators for all experiments, and 1000 replicates were used for confidence interval estimation; missing data were imputed using the most frequent value in the data. Demographical and patient characteristic variables that were significantly associated with CAD were included into the model (clinical model). SNPs associated with CAD were then added to clinical model (genetic+clinical model). Two-sided *p*-values <0.05 were considered statistically significant.

Results

General characteristics of the subjects

Table 1 displays mean, standard deviations (SD) and *p*-values for the patients and controls relevant to the demographic and clinical characteristics as well as biochemical parameters. 135 CAD patients (age 62.01±10.65) and 140 controls without CAD (age 58.21±12.62) were enrolled in this study. There were significant differences observed between CAD patients and control groups in terms of weight, height, BMI, fasting blood glucose, serum total cholesterol, triglyceride, LDL-C, HDL-C, VLDL, ALT and AST (*p<0.005,* respectively) than controls. However, there was no difference in age, ALP and systolic blood pressure (BP) and diastolic BP (*p>0.005*, respectively).

Relationship betweeen SNPs and CAD risk

The genotype and allelic distributions of six SNPs (rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656) in CAD patients and control group are presented in Table 2 and Table 3. Significant differences were observed in genotype frequencies of rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656 variants between CAD patients and control group (p<0.005) (Table 2). The risk genotypes; rs1333049 CC genotype, rs10757274 GG genotype, rs10757278 GG genotype, and rs10811656 TT genotype, also heterozygous mutant genotype of rs2383207 AG and rs2383206 AG genotype were more frequent in CAD patients compared to non-CAD group.

Furthermore, the risk alleles, rs1333049 C allele, rs2383207 G allele, rs2383206 G allele, rs10757274 G allele, rs10757278 G allele, and rs10811656 T allele were found statistically significant (OR=2.56, 95% CI= 1.757-3.754, OR=3.31, 95% CI= 2.09-5.24, OR=2.31, 95% CI= 1.62-3.28, OR=2.78, 95% CI= 1.85-4.18,

Table 2. The genotype distributions of SNPs on chromosome 9p21.3.

SNP	Genotype Frequencies n (%)			
	Genotype	CAD (n=135)	Controls (n=140)	*P-Value
rs1333049	GG	51(37.8)	94(67.1)	
	GC	61(45.2)	35(25)	***0.002***
	CC	23(17)	11(7.9)	
rs2383207	AA	2(1.5)	37(26.4)	
	AG	26(19.3)	8(5.7)	***0.001***
	GG	107(79.3)	95(67.9)	
rs2383206	AA	28(20.7)	72(51.4)	
	AG	82(60.7)	54(38.6)	***0.004***
	GG	25(18.5)	14(10)	
rs10757274	AA	58(43)	100(71.4)	
	AG	60(44.4)	35(25)	***0.001***
	GG	17(12.6)	5(3.6)	
rs10757278	AA	47(34.8)	104(74.3)	
	AG	68(50.4)	32(22.9)	***0.001***
	GG	20(14.8)	4(2.9)	
rs10811656	CC	19(14.1)	93(66.4)	
	CT	72(53.3)	36(25.7)	***0.001***
	TT	44(32.6)	11(7.9)	

The genotype distribution of polymorphisms between the groups was compared using x^2 test.

Table 3. Allelic Frequency of six SNPs at Chromosome 9p21.3 in our study population.

SNP	Allele Frequencies				
	CAD (n=135)	Controls (n=140)	X^2	OR/CI (95%)	*P-Value
rs1333049					
G	0.6	0.8			
C	0.4	0.2	24.4	2.56/1.75-3.75	***0.001***
rs2383207					
A	0.11	0.29			
G	0.89	0.71	28	3.31/2.09-5.24	***0.001***
rs2383206					
A	0.51	0.71			
G	0.49	0.29	22.2	2.31/1.62-3.28	***0.001***
rs10757274					
A	0.65	0.84			
G	0.35	0.16	25.57	2.78/1.85-4.18	***0.001***
rs10757278					
A	0.60	0.86			
G	0.40	0.14	46.21	4/2.64-6.05	***0.001***
rs10811656					
C	0.41	0.79			
T	0.59	0.21	85.35	5.56/3.81-8.12	***0.001***

The allelic frequency of polymorphisms between the groups was compared using HWE test.

Table 4. Allelic Frequency of six SNPs at Chromosome 9p21.3 Region Stratified by Inter-African Populations.

SNP	Allele Frequencies				
	*HapMap Populations				** CAD patients
	AFR	LWK	YRI	ASW	TNZ
rs1333049					
G	0.78	0.73	0.78	0.73	0.6
C	0.21	0.26	0.21	0.27	0.4
rs2383207					
A	0.03	0.05	0.01	0.09	0.11
G	0.96	0.94	0.98	0.90	0.89
rs2383206					
A	0.58	0.60	0.55	0.50	0.51
G	0.41	0.39	0.44	0.49	0.49
rs10757274					
A	0.82	0.75	0.85	0.71	0.65
G	0.17	0.24	0.14	0.28	0.35
rs10757278					
A	0.84	0.77	0.85	0.73	0.60
G	0.15	0.22	0.14	0.27	0.40
rs10811656					
C	0.56	0.54	0.62	0.52	0.41
T	0.43	0.45	0.32	0.47	0.59

* HapMap populations (http://hapmap.ncbi.nlm.nih.gov). AFR: African, LWK: Luhya in Webuye, Kenya, YRI: Yoruba in Ibadan, Nigeria, ASW: Americans of Anfrican ancestry in SW, USA and **TNZ: Tanzanian CAD patients from our study.

OR=4, 95% CI= 2.64-6.05 and OR=5.56, 95% CI= 3.81-8.12, respectively) in CAD patients compared to controls.

Table 4 presents the interethnic allelic frequencies of the CAD patients from our study and for relevant International HapMap Project populations (http://hapmap.ncbi.nlm.gov). When compared to other black populations, a larger proportion of Tanzanian CAD patients carried the risk genotype of three SNPs; rs1333049, rs10757274 and rs1057278.

A significant difference was also observed between gender and genotype distributions of rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656 variants. The rs1333049 CC genotype, GG genotype for the rs2383207 GG genotype, rs2383206 GG genotype, rs10757274 GG genotype and rs10757278 GG genotype and rs10811656 TT genotype were found to be more frequent in the female and male CAD patients compared to control counterparts.

In female patients, the risk alleles of six SNPs (rs1333049 C allele, for rs2383207 G allele, rs2383206 G allele, rs10757274 G allele, rs10757278 G allele and for rs10811656 T allele) were found to increase the CAD risk (OR= 2.71 95% Cl= 1.57-4.66, $p<0.001$, OR=5.49 95% CI= 3.210-9.404, $p<0.001$, OR=2.28 95% CI= 1.376-3.80, $p=0.001$, OR=3.23 95%CI= 1.868-5.913, $p=0.001$, OR=3.57 95% CI= 2.037-6.273, $p<0.001$ and OR=5.49 95% Cl= 3.21-9.404, $p<0.001$, respectively) when compared to female controls.

Moreover, in male patients, the risk alleles of six SNPs (rs1333049 C allele, for rs2383207 G allele, rs2383206 G allele, rs10757274 G allele, rs10757278 G allele and for rs10811656 T allele) were found to increase the CAD risk (OR=2.43 95% Cl= 1.397-4.24, $p=0.001$, OR=1.714 95% CI= 0.792-3.710, $p<0.05$, OR= 2.27 95% CI= 1.369-3.78, $p<0.05$, OR= 2.37 95% CI= 1.315-4.284, $p<0.05$, OR= 5.25 95% CI= 2.67-10.32, $p<0.001$ and OR= 5.76 95% Cl= 3.29-10.09, $p<0.00$, respectively) when compared with male controls.

There was a difference in genotype distribution of rs10811656 between female and male CAD patients; male heterozygous mutant carriers (CT genotype) were more frequent compared to females (62.7% for males, 38.5% for females).

In genetic association studies, statistical power to detect disease susceptibility loci depended on the genetic models tested. Therefore, the genotype frequencies were further analyzed by three genetic models: additive, dominant and recessive model. For rs10757278, a significant association between this polymorphism and increased risk of CAD was found in recessive model (OR=5.409, 95%CI=3.220-9.086, $p<0.0001$) and additive model (OR=4.00, 95% CI=2.644-6.051, $p<0.0001$). Moreover, significant positive correlations between rs10811656 and CAD risk were also identified in recessive (OR=2.081, 95% CI=1.639-3.982, $p<0.0001$) and additive model (OR=5.567,95% CI=3.817-8.120, $p<0.0001$). Similarly, an increased risk of CAD was also found in recessive model (OR=3.36, 95% CI= 2.051-5.524, $p<0.0001$, OR=4.046, 95% CI=2.377-6.887, $p<0.0001$, OR=3.888, 95% CI=2.626-5.426,$p<0.0001$, OR=3.319, 95% CI=2.012-5.475, $p<0.0001$, respectively) and additive model (OR=2.568, 95% CI=1.757-3.754, $p<0.0001$, OR=2.310, 95% CI=1.626-3.281, $p<0.0001$, OR=3.313, 95% CI=2.094-5.241, $p<0.0001$, OR=2.789, 95% CI=1.859-4.184, p<0.0001, respectively) of polymorphism rs1333049, rs2383206,

Table 5. Analysis of the six selected SNPs based on the three genetic models.

9p21.3 Genotype/Allele	Model	OR/ CI(95%)	p-value
rs1333049			
CC vs GC+GG	Dominant	0.415 / 0.194-0.889	0.02
G/G vs CC+GC	Recessive	3.36 / 2.051-5.524	***<0.0001***
G vs C	Additive	2.568 / 1.757-3.754	***<0.0001***
rs2383207			
GG vs AG+A	**Dominant**	**0.552 / 0.320-0.954**	**0.03**
A/A vs GG+AG	**Recessive**	**3.888 / 2.626-5.426**	*<0.0001*
A vs G	**Additive**	**3.313 / 2.094-5.241**	*<0.0001*
rs2383206			
GG vs AG+AA	Dominant	0.489 / 0.242-0.987	0.04
A/A vs GG+AG	Recessive	4.046 / 2.377-6.887	***<0.0001***
A vs G	Additive	2.310 / 1.626-3.281	***<0.0001***
rs10757274			
GG vs AG+AA	Dominant	5.409 / 3.220-9.086	0.005
A/A vs GG+AG	Recessive	3.319 / 2.012-5.475	***<0.0001***
A vs G	Additive	2.789 / 1.859-4.184	***<0.0001***
rs10757278			
GG vs AG+AA	Dominant	0.169 / 0.056-0.509	***<0.0001***
A/A vs GG+AG	Recessive	5.409 / 3.220-9.086	***<0.0001***
A vs G	Additive	4.00 / 2.644-6.051	***<0.0001***
rs10811656			
TT vs CT+CC	Dominant	0.176 / 0.086-0.360	***<0.0001***
C/C vs TT+CT	Recessive	2.081 / 1.639-3.982	***<0.0001***
C vs T	Additive	5.567 / 3.817-8.120	***<0.0001***

Abbreviations: OR,odds ratio; CI, confidence interval; p-value≤0.05 considered as statisticlly significant. P-values in bold have still remained their significance after Bonferroni correction (0.0027). Ors were adjusted for age, sex and smoking status of the study cohort in logistic regression.

rs2383207 and rs10757274 (Table 5).

Association between SNPs, biochemical measurements, and clinical features

There was a significant association between the rs10811656 genotypes and serum TC and HDL-C levels in CAD patients ($p<0.05$, respectively). Moreover, glucose, TC, HDL-C and LDL-C levels were statistically different in rs10811656 CT and TT female carriers compared to CC (WildType) female carries ($p<0.05$). We have also conducted gene-environmental analysis to investigate the association between the SNPs and demographic characteristics of subjects for the risk of CAD. The results indicated significant interaction of the risk genotypes of rs10757274 and rs10757278 with hypertension in conferring increased risk of CAD (OR=3.36, 95% CI= 2.01-5.6, $p<0.05$; OR=5.79, 95% CI= 3.40-9.86, $p<0.001$, respectively). Although the differences were observed, no statistical significance was reached in diabetes and obesity in conferring the increased risk of CAD.

Impact of SNPs on the severity of CAD in Tanzania

The result showed the strong predictive association between genotype distribution of rs1333049, rs10757278 and rs10811656 polymorphisms and severity of CAD among patients with one, two, three stenotic vessels ($p<0.05$). A strong direct association between the proportion of patients with three stenotic vessels disease and increasing the risk variants of rs1333049, rs10757278 and rs10811656 was observed ($p<0.05$, respectively). The corresponding OR per copy of the risk allele for the SNPs were 1.739 (95% CI= 1.35-2.14), 2.1 (95% CI = 1.97-2.23), and 1.96 (95% CI= 1.627-4.82), respectively. Similarly, there was also an association between rs10811656 polymorphism among female patients with one, two, three stenotic vessels ($p<0.05$). The corresponding OR per copy of the risk allele for rs10811656 was 0.883 (95% CI= 0.583-1.192). Thus, in CAD patients, a strong association between rs10811656 polymorphism and severity of CAD was observed.

Haplotype analysis of the 9p21.3 locus polymorphisms

It was also investigated whether the six SNPs were in linkage disequilibrium. Any common haplotypes associated with the disease and rare haplotypes (with frequency<5%) were excluded from the association analysis. The most common haplotypes of the six polymorphisms, calculated by Haploview 4.2 are summarized in Table 6. The haplotypes were generated using the six 9p21.3 locus SNPs (rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656) among the CAD cases and controls, and seventeen different haplotypes were generated (with frequency<5%) (Figure 1). For commonly observed haplotypes, GGGTGC haplotype ($p<0.0001$), GGGTAG haplotype ($p=0.006$), AGGTGC haplotype ($p=0.007$), AAGTGC

Table 6. The distribution of 9p21.3 locus haplotypes in Tanzanian CAD patients and controls.

Haplotype Associations	Frequency	Case, Control Frequency	Case, Control Ratio	χ^2	P value
AAGCAG	0.273	55.1 : 214.9, 94.9 : 185.1	0.204, 0.339	12.612	***<0.0001***
AAACAG	0.120	11.9 : 258.1, 54.0 : 226.0	0.044, 0.193	28.845	***<0.0001***
GGGTGC	0.119	54.1 : 215.9, 11.4 : 268.6	0.200, 0.041	33.457	***<0.0001***
AAGTAG	0.112	33.0 : 237.0, 28.5 : 251.5	0.122, 0.102	0.577	0.4474
AGGCAG	0.063	16.1 : 253.9, 18.6 : 261.4	0.060, 0.067	0.114	0.7352
AGGTAC	0.030	11.9 : 258.1, 4.7 : 275.3	0.044, 0.017	3.495	0.0616
GGGCGC	0.027	3.6 : 266.4, 11.3 : 268.7	0.014, 0.040	3.767	0.0523
GGGTAG	0.019	9.6 : 260.4, 1.0 : 279.0	0.036, 0.003	7.536	***0.006***
AGGTGC	0.018	9.0 : 261.0, 0.9 : 279.1	0.033, 0.003	7.111	***0.0077***
GGGCAC	0.017	1.2 : 268.8, 8.2 : 271.8	0.004, 0.029	5.063	***0.0244***
AAGTGC	0.017	8.2 : 261.8, 1.1 : 278.9	0.031, 0.004	5.812	***0.0159***
AAATAG	0.017	4.5 : 265.5, 4.6 : 275.4	0.017, 0.016	0.001	0.971
AGACAG	0.013	0.5 : 269.5, 6.6 : 273.4	0.002, 0.024	5.151	***0.0232***
GAACGG	0.013	1.9 : 268.1, 5.0 : 275.0	0.007, 0.018	1.355	0.2443
AGGCAC	0.012	2.0 : 268.0, 4.5 : 275.5	0.007, 0.016	0.857	0.3547
AGGTGG	0.011	6.0 : 264.0, 0.0 : 280.0	0.022, 0.000	6.158	***0.0131***

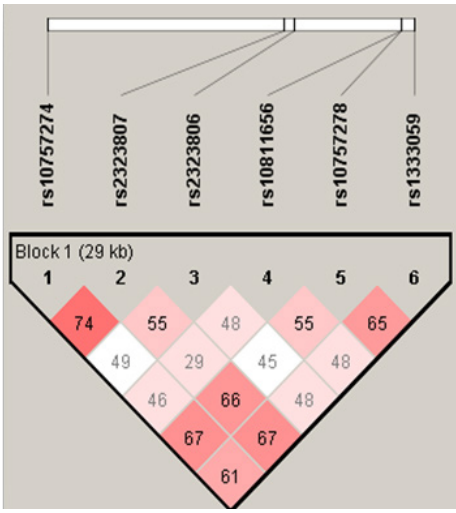

Figure 1. Linkage disequilibrium pattern of the SNPs along the 9p21.3 region.The graphic illustrates the distinct haplotypes defined using the Haploview 4.2. The linkage disequilibrium (D) between any two SNPs is shown in the cross cell. LD is presented with standard color schemes, bright red for very strong LD (LOD>2, D'=1), pink-red and blue for intermediate LD (LOD>2, D'<1 and LOD<2, D'=1, respectively) and white for no LD (LOD<2, D<1). The darker region shows higher r2-value. SNP, single nucleotide polymorphism; LD, linkage disequilibrium.

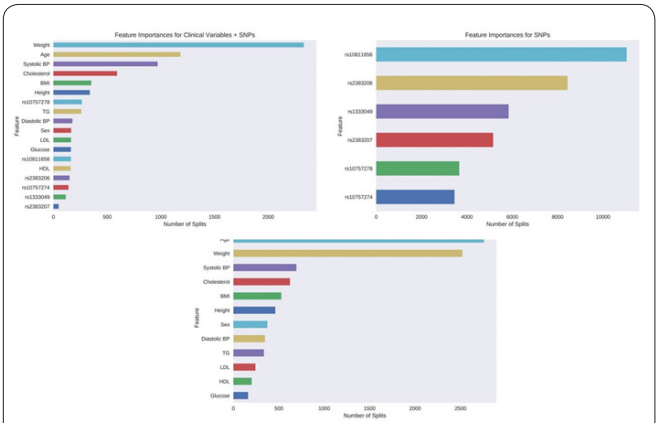

Figure 2. Feature importances for clinical variables, SNPs and the combined clinical variables+ SNPs model. We train a gradient boosting model for classification with 5000 estimators and report on the number of splits per feature for our dataset.

haplotype (*p=0.02*) and AGGTGG haplotype (*p=0.01*) were found linked with significant increase (high-risk haplotypes) in coronary artery disease risk while AAGCAG haplotype (*p=4.0E-4*), AAACAG haplotype (*p<0.0001*), GGGCAC haplotype (*p=0.0244*) and AGACAG haplotype (*p=0.0232*) were observed associated with a significant reduction (low-risk haplotypes) in coronary artery disease risk in Tanzania. The other seven haplotypes including AAGTAG, AGGCAG, AGGTAC, GGGCGC, AAATAG, GAACGG and AGGCAC were observed not associated with the risk of coronary artery disease (Table 6). Since this study was based on relatively small sample size, we applied a Bonferroni correction to decrease the type I error. Following that, the haplotypes GGGTGC, GGGTAG, AGGTGC, AAGTGC and AGGTGG still maintained their significance.

Combined genotype analysis of 9p21.3 locus SNPs

Table 7 summarizes the association studies among the combined genotypes of six SNPs and overall risk for CAD using conditional logistic regression model. The analysis revealed that SNP1 vs SNP3, rs1333049 and rs2383206, SNP3 vs SNP4, rs2383206 and rs10757274, SNP3 vs SNP6, rs2383206 and rs10811656, SNP4 vs SNP5, rs10757274 and rs10757278, SNP4 vs SNP 6, rs10757274 and rs10811656 had a positive correlations with increased risk of CAD (OR=1,759; CI=1,171-2,642; p<0.05; OR=0.462; CI=0,305-0,7*; p<0.001*; OR=1,454; CI=1,167-1,812; *p<0.001*; OR=1,377; CI=1,101-1,721; *p<0.05*; OR=1,443; CI=1,064-1,955; *p<0.05*, respectively).

Building CAD predictors and determining useful features

Models for CAD prediction based on clinical features and SNPs were built. A gradient boosting classifier was trained to obtain a measure of feature importances (Figure 2). In addition, merged output probabilities of

Table 7. Logistic regression model of SNP-SNP interactions and CAD risk.

SNP- SNP interactions	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.
SNP1vs SNP2	-0,069	0,134	0,268	1	0,605	0,933	0,717- 1,213
SNP1vs SNP3	0,565	0,208	7,411	1	***0,006***	1,759	1,171- 2,642
SNP1vs SNP4	-0,214	0,138	2,423	1	0,120	0,807	0,616- 1,057
SNP1vs SNP5	0,181	0,119	2,316	1	0,128	1,199	0,949- 1,514
SNP1vs SNP6	0,054	0,189	0,082	1	0,775	1,056	0,728- 1,530
SNP2vs SNP3	0,040	0,167	0,058	1	0,810	1,041	0,751- 1,443
SNP2vs SNP4	0,272	0,153	3,139	1	0,076	1,313	0,972- 1,773
SNP2vs SNP5	-0,026	0,126	0,041	1	0,839	0,975	0,762- 1,247
SNP2vs SNP6	0,223	0,131	2,908	1	0,088	1,250	0,967- 1,617
SNP3vs SNP4	-0,772	0,212	13,236	1	***0,001***	0,462	0,305- 0,700
SNP3vs SNP5	-0,144	0,144	0,998	1	0,318	0,866	0,653- 1,148
SNP3vs SNP6	0,374	0,112	11,139	1	***0,001***	1,454	1,167- 1,812
SNP4vs SNP5	0,320	0,114	7,879	1	***0,005***	1,377	1,101- 1,721
SNP4vs SNP6	0,366	0,155	5,579	1	***0,018***	1,443	1,064- 1,955
SNP5vs SNP6	-0,280	0,186	2,252	1	0,133	0,756	0,525- 1,089

Abbreviations: OR, Odds Ratio; CI, confidence interval; SNP1, rs1333049; SNP2, rs2383207; SNP3, rs2383206; SNP4, rs10757274; SNP5, rs10757278; SNP6, rs10811656. pvalue≤0.05 considered as statistically significant. ORs were adjusted for age and sex status of study cohort in logistic regression model.

Table 8. Comparison between two clinical models with and without SNPs, in addition to a model with SNPs only. AUC: area under curve for the ROC curve.

Model	AUC (95% CI)		AUC (95% CI)	P value
Clinical Model Model 1	0.954 (0.930-0.977)	Clinical Model and Genetic Model (Model 1+Model 2)	0.982 (0.954-0.985)	p<0.012
Genetik Model Model 2	0.815 (0.757-0.870)			

each fold for leave-one-out cross-validation was used to build ROC curves and compare models; the results are summarized in Table 8 and the ROC curves are shown in Figure 3. The first model (the clinical only model) consisted of known CAD risk factors collected in this study: BMI, age, sex, weight, height, diastolic and systolic BP, cholesterol, triglyceride, glucose, HDL, LDL levels. The second model is genetic only model that consisted of rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656 SNPs related with CAD. For the clinical+genetic model, the six associated SNPs were entered into the model assuming an additive model of inheritance, in addition to those included in the clinical-only model.

In an unrelated analysis, a multivariate logistic regression model fit was conducted on the SNP data for CAD prediction in order to obtain a predictive link between SNPs and CAD. The forest plot of the odds ratios for analysis is provided in Figure 4. As shown in Table 7 both models were significantly predictive of CAD, with area under curve (AUC) of 0.954 (95% CI:0.930-0.977) for clinical model and AUC of 0.815 (95% CI:0.757-0.870) for genetic model. Most importantly, all six CAD-associated SNPs, rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656, improved the predictive power for CAD over the model composing of only conventional known risk factors, with an improvement in AUC of 0.982 (95% CI:0.954-0.985) (*p=0.012*) (clinical model+genetic model).

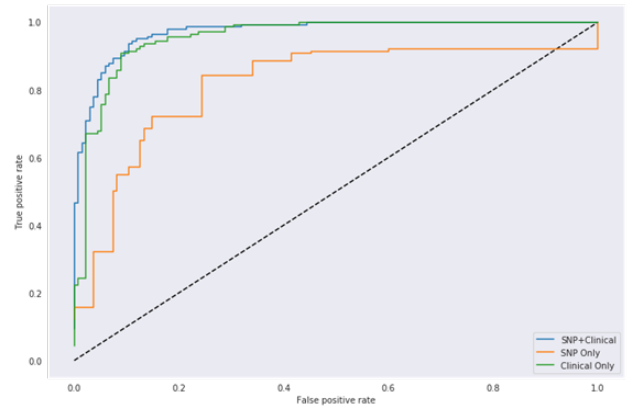

Figure 3. ROC curves of the trained models for CAD prediction. We plot the ROC curves for the three models we consider for CAD prediction.

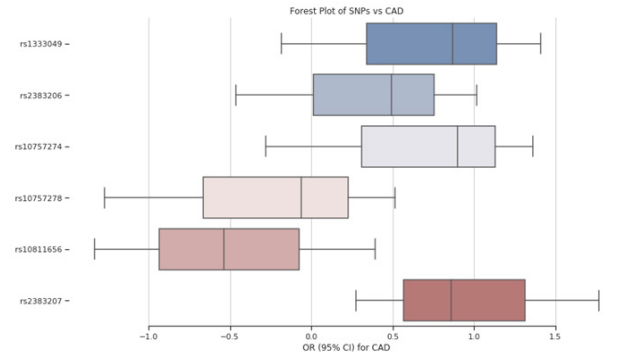

Figure 4. Forest plot of odds ratios of different SNPs for predicting CAD. We use a logistic regression model to obtain the odds ratio (OR) for the different SNPs.

Discussion

Cardiovascular disease is a global killer and it is increasing in incidence in low-income countries (1). This is probably due to change in lifestyle particularly in urban and peri-urban populations. During the last five decades, Sub-Saharan Africa has undergone enormous demographic changes coupled with an increase in the prevalence of noncommunicable diseases, such as diabetes, hypertension, coronary artery disease, obesity and *etc*. (3). There is a sufficient evidence that genetics is a contributory factor in the increased prevalence of CAD. Different approaches have been used in recent decades to discover causal genes for CAD. Recent GWAS has indicated that there is an association between the increased susceptibility to CAD and 9p21.3 locus. One such locus has been identified within chromosome 9p21.3 which codes for an antisense RNA (*CDKN2B-AS1*) and is located near the *CDKN2A-CDKN2B* gene cluster (27).

By use of GWAS, Helgadottir et al. (2007) and McPherson et al. (2007) independently and simultaneously discovered the first common CAD risk SNPs variants, located on chromosome 9p21.3 which play role in synergy with other known traditional risk factors in CAD development. The same variants within this locus were shown to be significantly associated with an increased risk to CAD (7). The risk alleles of 9p21.3 locus are carried by 75% of the European population (50% heterozygous and 25% homozygous risk) and confer risk for coronary atherosclerosis by an unknown mechanism (7, 8, 28). The risk for CAD is increased by 25% with one copy and 50% by two copies of the risk allele of the 9p21.3 (8). There are a great amount of replicated studies in different populations meant to decipher the genetic background of CAD pathophysiology. Although advancement in genomic technologies has allowed for systematic characterization of genome-wide genetic diversity in health and disease in European and Asian ancestry populations, the conduct of genetic studies in Sub Saharan Africa has been underwhelming until recently.

Thus, an attempt was undertaken in this study to determine whether SNPs on 9p21.3 locus genes are associated with CAD in Tanzania. This is the first study undertaken in Tanzanian population to replicate previous studies conducted in other populations on the association of the following SNPs variants rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656 with CAD (12- 15, 17, 29- 31). In this study, we successfully genotyped a total of six SNPs in 9p21.3 locus genes and examined their possible association with the pathogenesis of the CAD, including the first SNP rs1333049 identified through genome-wide scans by the U.K. Wellcome Trust Case-Control Consortium and the German Cardiogenics Consortium (9) to be most significantly associated with risk to CAD. The other five SNPs are also from the same locus; rs2383207, rs2383206, rs10757274, rs10757278, and rs10811656. We also investigated that whether or not these six polymorphisms are or not in linkage disequilibrium and common haplotypes of these SNPs are associated with CAD. Finally, we estimated the association among the combined genotypes of six SNPs and overall risk of CAD.

All the studied SNPs were associated with CAD patients in Tanzanian population. Genotype and allelic frequencies of rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656 were statistically significant in Tanzanian CAD patients compared to control group. Genotyping of our study population showed similar or even higher high-risk allele frequencies compared to those in previous studies which reported the association of 9p21.3 sequence variants with CAD (7, 8, 13). Our results confirm previously reported meta-analysis which deduced that the 9p21.3 risk locus was significantly associated with CAD (10).

The result of our study confirmed that the rs1333049 is significantly associated with CAD in the Tanzanian population, which is in agreement with the findings in other ethnic (7, 8, 13, 32-37), although a weak association was reported in some populations (38, 39). Our results showed not only the rs1333049 CC genotype, but also that the C allelic form is associated with CAD congruent with the results from some populations in Europe and Asia (7, 8, 13, 33-37) though G allelic form was associated with CAD in some populations (Cakmak et al. 2015). In addition, our results also confirmed the significant associations of rs2383206, rs2383207, rs10757274 and rs10757278 located on chromosome 9p21.3 with the risk of CAD as reported in multiple populations (7, 8, 15, 16, 40-42).

Helgadottir *et al.* conducted GWAS and showed that rs2383207 and rs10757278 were associated with MI and CAD (7). McPherson *et al*. found that rs2383206 and rs10757274 were associated with 30 to 40% increased risk of CHD in independent samples (n=23.000 participants) (8). Previously, case-control studies replicated the association between the SNPs on 9p21.3 (including the same SNPs) and the CAD risk in different ethnic groups and their findings are in agreement with our findings in Tanzanian population (16, 29, 39, 40). On the other hand, a few studies did not address the association between the risk allele of SNPs and phenotypic background of CAD (42, 43).

In our study, rs2383207 were more strongly associated with CAD, as confirmed to be the most common in populations from sub-Saharan Africa (44). Also our results showed that rs10811656 has a strong association with Tanzanian CAD population as demonstrated in Chinese population (41).

The "Africans" included in the various GWAS studies were from the North, West, Central and South of Africa. There is no study reported from Eastern Africa, therefore our work is the first and only study reported from East Africa. Our results differ from the previously reported African data. Although the studies of African Americans have been limited, 9p21.3 appears to be a risk factor in all ethnic groups that migrated from Africa (45). Our results are not in agreement with few studies (42, 46), stating the possible breakage of the African 9p21 risk haplotype into smaller haplotypes producing minimal or no risk for CAD. Their results might well be viable for Western, Central or Southern Africa but not for Eastern Africa.

Some of the known traditional risk factors of CAD are diabetes, dyslipidemia, obesity, and hypertension. Our result indicated the significant interaction between

the risk genotypes of rs10757274 and rs10757278 with hypertension in conferring an increased risk of CAD. It is possible that genetic factors would exert a greater influence of the traditional factors. Significant interactions were found between the traditional CAD risk factor (hypertension) and rs10757274 and rs10757278 in our study. These findings endorse the notion that many genes, each with a comparatively small effect, work in combination with other modifier genes and environmental factors (47). Other three parameters (diabetes, dyslipidemia, obesity) did not show significant associations with the SNPs in conferring an increased risk to CAD.

Available reports have discussed the association of these loci on chromosome 9p21.3 with diabetes suggesting the sharing of some pathological mechanisms with heart disease (30, 48, 49). Our study did not find a significant association with the SNPs and diabetes. A strong direct association between the proportion of patients with three stenotic vessels disease and the risk variants of rs1333049, rs10757278 and rs10811656 were seen ($p<0.05$). Similarly, several investigators have confirmed 9p21 as a predictor of the severity of CAD in an independent population (32, 45, 50, 51). In contrast, some studies did not observe any relationship between the 9p21.3 risk alleles and the number of coronary vessels involved in different populations (33, 52, 53).

In the second step of the study, we successfully established haplotypes for the 9p21.3 locus genes from different combination of six SNPs. For commonly observed haplotypes, GGGTGC, GGGTAG, AGGTGC, AAGTGC and AGGTGG haplotypes were found linked with significant increase (high-risk haplotypes) in coronary artery disease risk while AAGCAG, AAACAG, GGGCAC and AGACAG haplotypes were observed to be associated with a significant reduction (low-risk haplotypes) in coronary artery disease risk in Tanzania. High-risk haplotypes possess risky alleles which are consistently over represented in CAD patients relative to healthy controls, suggesting a leading role of selected polymorphisms in risk determination for Tanzania. To produce more information, linkage disequilibrium was calculated for six SNPs on 9p21.3 locus genes as shown in Table 5. It is possible that SNPs of this locus may have a collective effect on the disease occurrence. Furthermore, SNP-SNP interaction, rs1333049-rs2383206, rs2383206-rs10757274 rs2383206-rs10811656, rs10757274-rs10757278, and, rs10757274-rs10811656 combinations were associated with an increased CAD risk. Additionally, ROC analysis indicated that the CAD-associated SNPs (rs1333049, rs2383207, rs2383206, rs10757274, rs10757278, and rs10811656) together with the known risk factors of CAD such as BMI, age, sex, weight, height, diastolic and systolic BP, cholesterol, triglyceride, glucose, HDL, LDL levels could possibly serve as prognostic biomarkers of CAD for Eastern African Tanzanians.

At this point, it is clear that the 9p21.3 locus has been involved as a hotspot associated with CAD, but understanding the biological and functional effects of risk variants in that locus continues to be a challenge. There is growing evidence of the involvement of *CDKN2B-AS1* transcription from the 9p21.3 locus in CAD disease etiology (54). The disease-associated SNPs located on 9p21.3 would regulate the expression of *CDKN2B-AS1* rather than *CDKN2A/B* expressions (55). A recent study, it could be demonstrated that the epigenetic silencer polycomb repressive complexes 1 and 2 (*PRC1* and *PRC2*) and PRC-associated activating proteins can bind to *CDKN2B-AS1* and regulate. Moreover, the same study suggesting that *CDKN2B-AS1* may be able to modulate epigenetic regulation of *CDKN2A/B* genes expression in cis and trans-acting mechanism (56).

In conclusion, to the best of our knowledge, the present study demonstrated for the first time that SNPs (rs1333049, rs2383207, rs2383206, rs10757274, rs10757278, and rs10811656) located on the chromosome 9p21.3 locus is directly associated with increased risk of CAD in Tanzania. We confirmed that not only risk genotypes, but also risk alleles are significantly associated with CAD phenotype in Tanzanian population. Also, we have demonstrated that the risk variants of rs1333049, rs10757274, and rs10757278 SNPs confer a magnified severity to CAD. A novel finding of the present work is the possible use of the studied SNPs together with known risk factors as prognostic markers for CAD detection in Eastern African Tanzanians.

Our data do not contradict with the results of worldwide research that confirmed the association of the 9p21.3 locus with CAD. Although these data suggest that SNPs located on the chromosome 9p21.3 locus confer susceptibility to CAD across racial lines. Complex diseases such as CAD have a complicated mechanism involving gene-gene and gene-environment interactions. The potential source of variable findings is this complex mechanism that differs between populations.

Multiple and different SNPs in chromosomal locus should be further studied to elucidate the exact identity of the candidate genes and the quantity of their effect on the CAD pathogenesis. More detailed studies on 9p21.3 utilizing different approaches will give a better understanding of the underlying complex mechanism of CAD pathogenesis.

Different polymorphisms, copy number variations and epigenetic factors should be studied prospectively in Eastern African cohorts to fully understand the molecular basis underlying the association of 9p21.3 chromosome variants with the etiopathology and pathophysiology of CAD in the locus.

Acknowledgements

The authors gratefully acknowledge all the support from CAD patients and Jakaya Kikwete Cardiac Institute CATH Laboratory staff. In addition, we wish to thank individuals who gave us valuable help with the patient consent forms, sample collection and continuous support for the project: Miss Sarah Isaya Haule and Mr. Ally Luhaga from Muhimbili University of Health and Allied Sciences Genetics Laboratory (MGL).

Conflict of interest

The authors declared that there is no conflict of interest of any kind.

Author's contribution

G.A. and F.A. conceived the study. F.A., M.J., P.K., T.S.S. and G.A. designed the study methodology. M.J., P.K., and T.S.S. performed the clinical arm of the study. G.A. performed the experiments. I.A. contributed to

sample preparation. F.A. and M.K.T. contributed to the statistical analysis and interpretation of the results. G.A. and F.A.wrote the original draft of the manuscript. M.J., P.K., T.S.S., E.M., M.K.T. and I.A. edited, and reviewed the manuscript. F.A., E.M., and M.J. supervised the study.

References

1.World Health Organization. Global Action Plan for the prevention and control of noncommunicable diseases 2013–2020. 2013 http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236_eng.

2.Mirzaei M, Truswell AS, Taylor R, Leeder SR. Coronary heart disease epidemics: not all the same. Heart, 2009; 95: 740-46.

3.WHO Atlas of the African Health Statistics 2017 http://www.aho.afro.who.int/en/atlas/atlas-african-health-statistics-2017-0

4.Mannarino E. and Pirro M. Molecular biology of atherosclerosis. Clin Cases Miner Bone Metab. 2008; 5: 57-62

5.Cullen P, Rauterberg J, Lorkowski S. The pathogenesis of atherosclerosis Review. Handb Exp Pharmacol. 2005; 170: 3-70.

6.Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS. et al. Heart disease and stroke statistics–2012 update: a report from the American Heart Association. Circulation 2012; 125: 2–220.

7.Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Jonasdottir A, Sigurdsson A, Baker A, Palsson A. et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science 2007; 316: 1491-93.

8.McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR. et al. A common allele on chromosome 9 associated with coronary heart disease. Science 2007; 316: 1488-99.

9.Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Richard D, Meitinger T, Baraud P, Wichmann E. et al. WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007; 357: 443-53.

10.Palomaki GE, Melillo S, Bradley LA. Association between 9p21 genomic markers and heart disease: a meta–analysis. Journal of the American Medical Association 2010; 303: 648–56.

11.Preuss M, König IR, Thompson J R, Erdmann J, Absher D, Assimes TL, Blankenberg S, Boerwinkle E, Chen L, Cupples L A. 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. Circulation Cardiovascular Genetics 2010; 3: 475–83.

12.Yan J, Zeng J, Xie Z, Liu D, Wang L, Chen Z. Association of rs10811656 on 9P21.3 with the risk of coronary artery disease in a Chinese population. Lipids Health Dis. 2016; 15: 126.

13.Schunkert H, Gotz A, Braund P, McGinnis R, Tregouet DA, Mangino M, Linsel-Nitschke P, Cambien F, Hengsrenberg C, Stark K. et al. Cardiogenics Consortium. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. Circulation 2018; 117: 1675-84.

14.Meng W, Hughes A E, Patterson C C, Belton C, Kee F, McKeown P P. Chromosome 9p21.3 is associated with early-onset coronary heart disease in the Irish population. Disease Markers 2008; 25: 81–85.

15.Shen GQ, Li L, Rao S, Abdullah KG, Ban JM, Lee BS, Park EJ, Wang QK. 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: 360-365.

16.Zhou L, Zhang X, He M, Cheng L, Chen Y, Hu FB, Wu T. Associations between single nucleotide polymorphisms on chromosome 9p21 and risk of coronary heart disease in Chinese Han population. Arterioscler Thromb Vasc Biol. 2008; 28: 2085-89.

17.Xie F, Chu X, Wu H, Sun W, Shen M, Yang L, Wang Y, Wang Y, Shi J, Huang W. Replication of putative susceptibility loci from genome-wide association studies associated with coronary atherosclerosis in Chinese Han population. PLoS One 2011; 6: 20833.

18.Lu XF, Wang LY, Chen SF, He L, Yang X, Shi YY, Cheng J, Zhang L, Gu C, Huang J. et al. Coronary Artery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Consortium. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. Nat Genet. 2012; 44: 890-894.

19.Broadbent HM, Peden JF, Lorkowski S, Goel A, Ongen H, Green F, Clarke R, Collins R, Franzoai MG, Tognoni G. et al. PROCARDIS consortium. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Hum Mol Genet. 2008; 17: 806-14.

20.Patel RS, Su S, Neeland IJ, Ahuja A, Veledar E, Zhao J, Helgadottir A, Holm H, Gulcher JR, Stefansson K. et al. The chromosome 9p21 risk locus is associated with angiographic severity and progression of coronary artery disease. Eur Heart J. 2010; 31: 3017-23

21.Zhou LT, Qin L, Zheng DC, Song ZK, Ye L. Meta-analysis of genetic association of chromosome 9p21 with early-onset coronary artery disease. Gene 2012; 510: 185-188.

22.Weiner DA, Ryan TJ, McCabe CH, Chaitman BR, Sheffield LT, Fisher LD, Tristani F. Value of exercise testing in determining the risk classification and the response to coronary artery bypass grafting in three-vessel coronary artery disease: a report from the Coronary Artery Surgery Study (CASS) registry. Am J Cardiol. 1987; 60: 262–66.

23.Knopfholz J, Disserol CC, Pierin AJ, Schirr FL, Streisky L, Takito LL, Massucheto Ledesma P, Faria-Neto JR, Olandoski M, da Cunha CL, Bandeira AM. Validation of the friedewald formula in patients with metabolic syndrome. Cholesterol. 2014; 2014: 261878.

24.Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems 2017; 30: 3149-57.

25.Jones, Eric, Oliphant T, and Peterson. P. {SciPy}: open source scientific tools for {Python}. 2014

26.Pedregosa, Fabian, Gaël VG, Alexandre GA, Vincent MV, Bertrand TB, Olivier GO, Mathieu BM. Scikit-learn: Machine learning in Python. Journal of machine learning research 2011; 12: 2825-30.

27.Roberts R, Stewart AF. Genes and coronary artery disease: where are we? J Am Coll Cardiol. 2012; 30: 1715-21.

28.Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007; 447: 661–78.

29.Maitra A, Dash D, John S, Sannappa PR, Das AP, Shanker J, Rao VS, Sridhare H, Kakkar VV. A common variant in chromosome 9p21 associated with coronary artery disease in Asian Indians. J Genet. 2009; 88: 113-18.

30.Saxena R, Voight BF, Lyssenko V, Burtt NP, Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ. et al Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007; 316: 1331–36.

31.Helgadottir A, Thorleifsson G, Magnusson KP, Gretarsdottir S, Steinthorsdottir V, Manolescu A, Jones GT, Rinkel GJ, Blankenstejin JD, Ronkainen A. et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. Nat Genet. 2008; 217-24.

32.Ardissino D, Berzuini C, Merlini PA, Mannucci PM, Surti A, Burtt N, Voight B, Tubaro M, Peyvandi F, Spreafico M. et al.

Influence of 9p21.3 genetic variants on clinical and angiographic outcomes in early-onset myocardial infarction. J Am Coll Cardiol. 2011; 58: 426–34.

33.Hinohara K, Nakajima T, Takahashi M, Hohda S, Sasaoka T, Nakahara K, Chida K, Sawabe M, Arimura T, Sato A. et al. Replication of the association between a chromosome 9p21 polymorphism and coronary artery disease in Japanese and Korean populations. J Hum Genet 2008; 53: 357-59.

34.Bhanushali AA, Contractor A, Das BR. Variant at 9p21 rs1333049 is associated with age of onset of coronary artery disease in a Western Indian population: a case control association study. Genet Res (Camb). 2013; 95: 138-45.

35.Scheffold T, Kullmann S, Huge A, Binner P, Ochs HR, Schöls W, Thale J, Motz W, Hegge FJ, Stellbrink C. et al. Six sequence variants on chromosome 9p21.3 are associated with a positive family history myocardial infarction: a multicenter registry. BMC Cardiovasc Disord. 2011; 7: 11-9.

36.Ivanova AA, Maksimov VN, Orlov PS, Ivanoshchuk DE, S avchenko SV, Voevoda MI. Association of the genetic markers for myocardial infarction with sudden cardiac death. Indian Heart J. 2017; 69: 8-11.

37.Abid K, Mili D, Kenani A. Polymorphism on Chromosome 9p21.3 Is Associated with Severity and Early-Onset CAD in Type 2 Diabetic Tunisian Population. Dis Markers 2015; 2015:792679.

38.Shen GQ, Rao S, Martinelli N, Li L, Olivieri O, Corrocher R, Abdullah KG, Hazen SL, Smith j, Barnard J, Plow EF. et al. Association between four SNPs on chromosome 9p21 and myocardial in¬farction is replicated in an Italian population. J Hum Genet. 2008; 53: 144-50.

39.Foroughmand AM, Nikkhah E, Galehdari H, Jadbabaee MH. Association Study between Coronary Artery Disease and rs1333049 and rs10757274 Polymorphisms at 9p21 Locus in South-West Iran. Cell J. 2015; 17: 89-98.

40.Cakmak HA, Bayoğlu B, Durmaz E, Can G, Karadağ B, Cengiz M, Vural VA, Yüksel H. Evaluation of association between common genetic variants on chromosome 9p21 and coronary artery disease in Turkish population. Anatol J Cardiol. 2015; 15: 196-203.

41.Zeng Q, Yuan Y, Wang S, Sun J, Zhang T, Qi M. Polymorphisms on chromosome 9p21 confer a risk for acute coronary syndrome in a Chinese Han population. Can J Cardiol. 2013; 29: 940-44.

42.Assimes TL, Knowles JW, Basu A, Iribarren C, Southwick A, Tang H, Absher D, Li J, Fair JM, Rubin GD. Susceptibility locus for clinical and subclinical coronary artery disease at chromosome 9p21 in the multi-ethnic ADVANCE study. Hum Mol Genet. 2008; 17: 2320-28.

43.Dehghan A, Van-Hoek M, Sijbrands EJ, Oostra BA, Hofman A, Van-Duijn CM, Witteman JC. Lack of association of two common polymorphisms on 9p21 with risk of coronary heart disease and myocardial infarction; results from a prospective cohort study. BMC Med. 2008; 16: 6-30.

44.Tekola-Ayele F, Adeyemo AA, Rotimi CN. Genetic epidemiology of type 2 diabetes and cardiovascular diseases in Africa. Prog Cardiovasc Dis. 2013; 56: 251-60.

45.Gong Y, Beitelshees AL, Cooper-DeHoff RM, Lobmeyer MT, Langaee TY, Wu J, Cresci S, Province MA, Spertus JA, Pepine CJ, Johnson JA. Chromosome 9p21 haplotypes and prognosis in white and black patients with coronary artery disease. Circ Cardiovasc Genet. 2011; 4: 169-78.

46.Zanetti D, Via M, Carreras-Torres R, Esteban E, Chaabani H, Anaibar F, Harich N, Habbal R, Ghalim N, Moral P. Analysis of Genomic Regions Associated With Coronary Artery Disease Reveals Continent-Specific Single Nucleotide Polymorphisms in North African Populations. J Epidemiol. 2016;b26: 264-71.

47.Arnett DK, Baird AE, Barkley RA, Basson CT, Boerwinkle E, Ganesh SK, Herrington DM, Hong Y, Jaquish C, McDermott DA. Relevance of genetics and genomics for prevention and treatment of cardiovascular disease: a scientific statement from the American Heart Association Council on Epidemiology and Prevention, the Stroke Council, and the Functional Genomics and Translational Biology Interdisciplinary Working Group. Circulation. 2007; 115: 2878 –901.

48.Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L. et al A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007; 316: 1341–45.

49.Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry, JR, Rayner NW, Freathy RM. et. al Replication of genomewide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007; 316: 1336–41.

50.Dandona S, Stewart AF, Chen L, Williams K, So D, O'Brien E, Glover C, LeMay M, Assogba O, Vo L. et al. Gene dosage of the common variant 9p21 predicts severity of coronary artery disease. J Am Coll Cardiol. 2010; 56: 479-86.

51.Chan K, Motterle A, Laxton RC, Ye S. Common variant on chromosome 9p21 predicts severity of coronary artery disease. J Am Coll Cardiol. 2011; 57: 1497–98.

52.Chen Z, Qian Q, Ma G, Wang J, Zhang X, Feng Y, Shen C, Yao Y. A common variant on chromosome 9p21 affects the risk of early-onset coronary artery disease. Mol Biol Rep. 2009; 36: 889 –93.

53.Anderson JL, Horne BD, Kolek MJ, Muhlestein JB, Mower CP, Park JJ, May HT, Camp NJ, Carlquist JF. Genetic variation at the 9p21 locus predicts angiographic coronary artery disease prevalence but not extent and has clinical utility. Am Heart J. 2008; 156: 1155–62.

54.Jarinova O, Stewart AF, Roberts R, Wells G. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus, Arterioscler. Thromb., Vasc. Biol. 2009; 29: 1671–77.

55.Holdt LM, Beutner F, Scholz M, Gielen S. ANRIL expression is associated with atherosclerosis risk at chromosome 9p21, Arterioscler. Thromb., Vasc. Biol. 2010; 30: 620–27.

56.Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene. 2011; 30: 1956–62.