Tanzania Journal of Health Research *https://doi.org/10.4314/thrb.v21i1.3*
Volume 21, Number 1, April 2019

Intronic variants in the long non-coding RNA CDKN2B-AS1 are strongly associated with the risk of coronary artery disease in the Northern Tribes of Tanzania

Gokce Akan[1], Peter Kısenge[2], Tulızo Shemu Sanga[2], Erasto Mbugı[1], Mehmet Kerem Turkcan[3], Mohammed Janabı[2] and 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*

Abstract
Introduction: Sub-Saharan Africa (SSA) is facing a rising epidemic of non-communicable diseases including the coronary artery disease (CAD) ranking at the top of the list. Chromosome locus 9p21.3 containing *CDKN2B* antisense RNA 1 (*CDKN2B*-AS1), identified in many genome-wide association studies for coronary artery disease (CAD), encompasses multiple single nucleotide polymorphisms (SNPs). This study aimed to conduct the first genetic study evaluating the common polymorphisms in 9p21.3 locus in Tanzanian CAD patients from different regions of Tanzania and their associations with CAD risk factors.
Material and Methods: A total of 90 patients from Northern region (N-CAD) of Tanzania and 65 patients from other regions (South, East, West and Central) (R-CAD) were included in the study. Further the biochemical analysis the genotyping of common variants was performed with the LightSNiP typing assay using qRT-PCR method.
Results: Our analyses revealed that both genotype and allele frequencies of rs10757274, rs10757278 and rs10811656 were significantly different between the groups ($p<0.05$, respectively). We identified that one previously undescribed three-marker haplotype (rs1333049, rs10757274 and rs10757278) encompassing *CDKN2B-AS1* was overrepresented (G-G-G, the risk haplotype, $p<0.05$) in N-CAD group compared to R-CAD group. The AUC of a risk model based on non-genetic factors was 0.730 (0.654-0.797) and the combination with genetic risk factors improved the AUC to 0.784 (95%CI=0.713-0.844, $p<0.0001$).
Conclusion: Our results identified the presence of a novel three-marker haplotype having a significant association with CAD in Northern Tanzania. Moreover, combination of the nongenetic and genetic risk models were demonstrated to indicate good diagnostic accuracy for CAD in Northern Tanzania.

Keywords: Coronary Artery Disease, 9p21.3, Single Nucleotide Polymorphism, Tribe, Tanzanian Population

Introduction

The contribution of non-communicable diseases (NCDs) to the health challenges is growing globally and becoming major health problem facing the world today and a global leading cause of death and disability (Lee et al., 2012). Sub-Saharan Africa (SSA) is characterized by the greatest burden of communicable diseases (CDs) among all regions in the world (Boutayeb, 2006). But a disturbing trend in the last decade has been the consistent rise of NCDs in SSA (Dalal et al., 2011). The prevalence of some NCDs in SSA is beginning to match with that in high-income countries and are projected to increase at least 5-fold by 2100 (Adebamowo et al., 2017). Coronary artery disease (CAD) as a main NCDs once considered rare in SSA, its incidence is now becoming worrisome in this region. The number of deaths due to CAD increased by 87% from 1990 to 2013 in SSA, likely due to aging and growth of the SSA population (Adebamowo et al., 2017). Moreover, the number of deaths caused by NCDs will be increased by 2020 with CAD

*Corresponding e-mail: fatmahan.atalar@gmail.com

expected to be the most leading causes of death and disability among adult populations (World Health Organization, 2002). The spectrum and pattern of CAD along with their risk factors are changing in urban areas as a result of progressive urbanization and westernization of lifestyle (Tantchou Tchoumi et al., 2011; Mocumbi, 2012). The result of migration from rural areas to urban areas and living in urban areas compared to rural areas is strongly associated with a higher prevalence of hypertension, glucose intolerance, obesity dyslipidemia, and also CAD (Unwin et al., 2010).

CAD is caused by a combination of genetic, physiological, environmental and behavioral factors. Modifiable risk factors such as diabetes, cholesterol, hypertension, and smoking has repeatedly shown that 30% to 40% of deaths from CAD can be prevented and the fact that a major proportion of the susceptibility to CAD is due to genetic risk factors has been recognized for more than five decades. For reduced or eliminated of the CAD, comprehensive prevention will require knowledge of the genetic risk factors (Roberts, 2014).

Genome-wide association studies (GWASs) after development of DNA sequencing technology have facilitated to increase our knowledge for understanding the genetic basis of complex diseases, for preventing the disease and translation toward new therapeutics (Visscher et al., 2017).

In 2007, GWAS identified a new susceptibility locus for CAD mapped at chromosome 9p21.3 (Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007). This locus contains the coding sequences of genes for two cyclin-dependent kinase inhibitors; 2 cyclin kinases inhibitors (*CDKN2A/B*) *CDKN2A* ($p16^{INK4a}$, $p14^{ARF}$), *CDKN2B* ($p15^{INK4b}$), methylthioadenosine phosphorylase (*MTAP*) and large nonprotein coding RNA in *INK4* locus, termed *CDKN2B-AS1* (*ANRIL*). Moreover, CDKN2*B-AS1* has been proposed to regulate their neighbour adjacent protein coding genes which include *CDKN2A/B* through transcription factors (Congrains et al., 2012). Also, several single nucleotide polymorphisms (SNP) associated with CAD, are located on the 9p21.3 locus. These SNPs show a strong linkage disequilibrium forming a risk haplotype 58 kilobase in length (Helgadottir et al., 2007; McPherson et al., 2007). Associations of the SNPs with CAD pathogenesis have also been replicated in many case-control studies in non-African populations (Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 20017; Congrains et al., 2012; Hinohara et al., 2008; Zhou et al., 2012; Yan et al., 2016).

Africa is the ancestral homeland of all modern humans in line with knowledge of human evolutionary history (Teo et al., 2010). Though the genetic studies of populations on mitochondrial (mt) DNA and nuclear DNA markers consistently indicate that knowledge of genetic diversity and population structure in Africa has high levels of haplotype diversity and low levels of linkage disequilibrium (LD) (González-Santos et al., 2015). Because of the important role of African populations in human history, characterizing their patterns of genetic diversity is crucial for reconstructing for understanding the genetic basis of complex diseases. GWASs in Africa could shed light on understanding the genetic background of complex diseases (Tishkoff & Williams, 2002).

In the present study, we selected from GWAS six SNPs (rs1333049, rs2383206, rs2383207, rs10757274, rs10757278 and rs10811656) located on 9p21.3 locus. These SNPs were previously reported to be strongly associated with CAD in non-African populations (Helgadottir et al., 2007; McPherson et al., 2007). Associations of the SNPs with CAD pathogenesis have also been replicated in many case-control studies in non-African populations (Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 20017; Congrains et al., 2012; Hinohara et al., 2008; Zhou et al., 2012; Yan et al., 2016). And then we investigated their associations with CAD in different regions of Tanzania along with the CAD risk factors.

Materials and Methods

Study Population

The study was performed at Muhimbili National Hospital (MNH) Jakaya Kikwete Cardiac Institute (JKCI) from January 2016 through February 2017. Individuals enrolled in the study were selected among patients admitted to the cardiology outpatient clinic. A total of 155 CAD patients; 90 CAD patients from Northern region (mean age 63.10±11.25) and 65 CAD patients from other regions (Southern region, Western region, Central region and Eastern region of Tanzania) (mean age 59.67±9.21) were enrolled in this study. CAD was defined as ≥ 50 % luminal narrowing in at least one coronary artery. All subjects enrolled in this study were African Tanzanians belonging to different tribes.

Detailed information on demographics, past history, lifestyle factors and coronary risk factors (diabetes, hypertension, obesity and etc.) record were completed through personal interviews. Also, informed consent was obtained from each participant. Body mass index (BMI) was calculated with regard to the formula BMI = kg/m^2. Obesity was defined as a BMI ≥30. For diagnosis of dyslipidemia, triglyceride, Low-density lipoprotein-cholesterol (LDL-C) and high-density lipoprotein (HDL-C) level were used as parameters according to National Cholesterol Education Program Adult Treatment Panel III (NCEP-ATP III) (NCEP-ATP III, 2001). Hypertension was defined as a systolic blood pressure ≥140 mmHg and diastolic blood pressure ≥90 mmHg or on the basis that patients were already being treated with anti-hypertensive drugs. Diabetes was diagnosed either by the 1999 World Health Organization (WHO) criteria 11 or self-report of being previously diagnosed as diabetic. All participants gave written consent after receiving a full explanation of the study. Ethics approval was obtained from the Ethics Committee of the Muhimbili University of Health and Allied Sciences (MUHAS).

Biochemical parameters

Serum total cholesterol (TC) and HDL-C, triglycerides (TG) were measured by routine enzymatic endpoint methods and using an automated autoanalyzer (Analyzer A15 Biosystems, Philippines). Fasting glucose was determined using the enzymatic reference method with glucose oxidase. LDL-C and Very Low-density lipoprotein-cholesterol (VLDL-C) were calculated using the Friedewald Formula.

DNA Isolation

Genomic DNA was obtained from peripheral blood leukocytes with MagnaPure Compact (Roche, Germany). DNA qualities and quantities were determined by NanoDrop™ 1000 Spectrophotometer (Thermo Scientific, Wilmington, Delaware USA). The extracted DNA was stored at -20°C before subsequent processes.

Genotyping

Genotyping of the 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). 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. The associations between variants in six SNPs and CAD risk factors were estimated by computing odds ratios (ORs) and 95% confidence

intervals (CIs). A multiple logistic regression model was used to adjust for multiple CAD risk factors. 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. The ROC curves were calculated using leave-one-out cross-validation and random forests with 1000 trees (DeLong et al., 1988; Pedregosa et al., 2011). For classification, differences were considered significant at $p<0.05$.

Results

General Characteristics of the Subjects

A total of 155 CAD patients; 90 CAD patients from Northern region and 65 CAD patients from other regions were enrolled in this study. All individuals were African Tanzanians belonging to different tribes. Most of the patients were from the Northen region of Tanzania (n=90), and the rest from South, West, East and Central regions of Tanzania (n=20, n=15, n=15 and n=15 respectively). The majority of the Northern participants were living in an urban location and were ethnically Chagga tribe.

Of all patients, 105 (67.7%) had obesity, 104 (67.4%) had type 2 diabetes mellitus, 134 (86.4%) had hypertension and 85 (54.8%) had hyperlipidemia. Moreover 38 (24.5%) participants were smokers and 66 (42.5%) participants declared a positive family history of CAD. The prevalence of obesity, diabetes, hypertension, hyperlipidemia and family history did differ between N-CAD and R-CAD (***p<0.05***, respectively). There was no significant difference in prevalence of smoking between N-CAD and R-CAD group ($p>0.05$). According to biochemical analysis the CAD patients from Northern region (N-CAD) and other regions (R-CAD), there were no significant differences observed between CAD patients from Northern region and other regions in terms of age, height, VLDL, systolic BP, and diastolic BP ($p>0.05$, respectively). However, N-CAD patients had significantly greater weight, BMI, fasting blood glucose, serum TC, TG, LDL-C and HDL-C (***p<0.05***, respectively) than R-CAD patients. There were no significant differences observed between male CAD patients from Northern region and other regions in terms of systolic BP, diastolic BP and VLDL ($p>0.05$, respectively). However, N-CAD male patients had significantly higher differences BMI, fasting blood glucose, serum TC, TG, LDL-C and significantly lower HDL-C level (***p<0.05***, respectively) than male R-CAD patients. Moreover, there were significant differences detected between female N-CAD patients and female R-CAD patients with regard to level of BMI, fasting blood glucose, serum TC, LDL-C HDL-C and VLDL (***p<0.05***, respectively).

Genotypes and Allele Frequencies in N-CAD and R-CAD patients and Their Associations with CAD

The genotype and allelic distributions of rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656 SNPs in N-CAD patients and R-CAD patients are presented in Table 1. Significant differences were observed in genotype frequencies of rs10757274, rs10757278 and rs10811656 between N-CAD patients and R-CAD patients (***p<0.05***). The risk genotypes of rs10757274, rs10757278 (GG genotypes of the rs10757274 and rs10757278) and rs10811656 (TT genotype for the rs10811656) were associated with an increased risk to CAD in northern region of Tanzania (***p<0.005,*** respectively) The risk alleles frequencies of rs10757274, rs10757278 and rs10811656 (G alleles of rs10757274, rs10757278, and T allele of rs10811656) were found significantly higher (OR=3.072 CI 95% 1.899-4.970, OR=2.611 95% CI:1.606-4.246 and OR=2.115 95% CI:1.334-3.351, respectively) in N-CAD patients compared to R-CAD patients (Table 1).

On the other hand we have not noted significant difference in genotype and allele frequency distributions of rs1333049, rs2383206 and rs2383207 SNPs between N-CAD and R-CAD patients. We also investigated gender and genotype distributions of SNPs, a significant difference was also observed between gender and genotype distributions of rs10757274, rs10757278 and rs10811656 SNPs. GG genotypes of rs10757274 and rs10757278 and TT genotype of rs10811656

were found to be more frequent in the female and male N-CAD patients than in female and male R-CAD (***p<0.05***, respectively).

We also examined the associations of rs10757274, rs10757278 and rs10811656 with the risk of CAD in female and male N-CAD patients. In female N-CAD patients, G alleles of rs10757274, rs10757278 and T allele of rs10811656 increased the risk of CAD 3.2 times (95% CI:1.33-7.70) (***p=0.007***), 4.2 times (95% CI:1.74-10.13) (***p=0.001***) and 4.12 times (95% CI:1.75-9.68) (***p=0.008***) respectively in comparison to female R-CAD patients. In male patients, G alleles of rs10757274, rs10757278 and T allele of rs10811656 increased the risk of CAD in male N-CAD patients; 3.7 times (95% CI:1.74-7.84) (***p=0.000***); 2.65 times (95% CI:1.15-6.11) (***p=0.020***) and 7 times (95%1.42-34.28) (***p=0.008***), respectively compared to male R-CAD patients. 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 rs10757274, a significant association between this polymorphism and increased risk of CAD was found in all three models, dominant model (OR:8.25, 95% CI:3.64-18.68, ***p<0.0001***), recessive model (OR:0.31, 95% CI:0.11-0.88, ***p=0.022***) and additive model (OR:3.07, 95% CI:1.89-4.97, ***p<0.0001***). Moreover, significant positive correlations between rs10757278 and CAD risk were also identified in dominant (OR:5.10, CI:2.29-11.36, ***p<0.0001***) and additive model (OR:1.95,95% CI:1.22-3.10, ***p=0.0044***). Similarly, an increased risk of CAD was alo found with rs10811656 in dominant model (OR:9.31, 95% CI:3.30-26.21, ***p<0.0001*** and additive model (OR:2.11, 95% CI:1.33-3.35, ***p=0.0013***).

Haplotype analysis

The most common haplotypes of the six polymorphisms, calculated by Haploview 4.2 are summarized in Table 2. Any common haplotypes associated with the disease and rare haplotypes (with frequency<5%) were excluded from the association analysis. The haplotypes were generated using the three SNPs (rs1333049, rs10757274 and rs10757278) encompassing the long non-coding RNA RNA *CDKN2B-AS1* between CAD patients and controls, and eight different haplotypes were generated (with frequency<5%) (Figure 1). GGG haplotype was found to be a highrisk haplotype for coronary artery disease risk (***p<0.05***). The significancy remained after applying Bonferroni correction.

Table 1: The genotypic and allelic frequency distributions of SNPs on chromosome 9p21.3 in study cohorts

SNP	Genotypic Frequencies n (%)		P-Value	Allelic Frequencies			X^2	OR/CI(95%)	P-Value
Genotype	N-CAD(n=90)	R-CAD(n=65)		Allele	N-CAD(n=90)	R-CAD(n=65)			
rs1333049 GG	37(41.1)	27(41.53)							
GC	45(50)	25(38.47)	0.116	G/C	0.66/0.34	0.61/0.39	0.93	0.79/0.49-1.26	0.333
CC	8(8.9)	13(20)							
rs2383206 AA	18(18.9)	18(27.69)							
AG	58(65.6)	36(55.38)	0.819	A/G	0.52/0.48	0.55/0.45	0.42	1.16/0.73-1.82	0.517
GG	14(15.6)	11(16.93)							
rs2383207 AA	17(20)	24(36.93)							
AG	59(64.4)	31(47.69)	0.226	A/G	0.52/0.48	0.61/0.39	2.24	1.41/0.89-2.24	0.134
GG	14(15.6)	10(15.38)							
rs10757274 AA	10(11.1)	33(50.77)							
AG	61(67.8)	27(41.53)	*0.000*	A/G	0.45/0.55	0.72/0.28	21.59	3.07/1.89-4.97	*0.000*
GG	19(21.1)	5(7.7)							
rs10757278 AA	11(12.9)	27(41.54)							
AG	65(72.2)	30(46.15)	*0.000*	A/G	0.46/0.54	0.69/0.31	15.33	2.61/1.60-4.24	*0.000*
GG	14(15.6)	8(12.3)							
rs10811656 CC	5(5.6)	23(35.39)							
CT	54(60)	24(36.92)	*0.000*	C/T	0.36/0.64	0.54/0.46	10.29	2.11/1.33-3.35	*0.001*
TT	31(34.4)	18(27.69)							

OR: Odd Ratio, CI: Confidence Interval *The genotypic and allelic frequency distributions of polymorphisms between the groups were compared using HWE test.

In all cases differences were considered significant at $p< 0.05$.

https://doi.org/10.4314/thrb.v21i1.3

Table 2: 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
AAG	0.459	83.5 : 96.5, 40.4 : 49.6	0.464, 0.449	0.057	0.8112
GGC	0.235	37.0 : 143.0, 26.5 : 63.5	0.205, 0.295	2.666	0.1025
AGC	0.072	11.7 : 168.3, 7.7 : 82.3	0.065, 0.085	0.375	0.5404
AAC	0.065	12.1 : 167.9, 5.4 : 84.6	0.067, 0.060	0.042	0.8367
AGG	0.060	10.7 : 169.3, 5.5 : 84.5	0.059, 0.061	0.002	0.9604
GAG	0.055	12.1 : 167.9, 2.8 : 87.2	0.067, 0.031	1.483	0.2234
GGG	0.033	8.6 : 171.4, 0.3 : 89.7	0.048, 0.003	3.742	***<0.05***
GAC	0.021	4.3 : 175.7, 1.3 : 88.7	0.024, 0.015	0.228	0.633

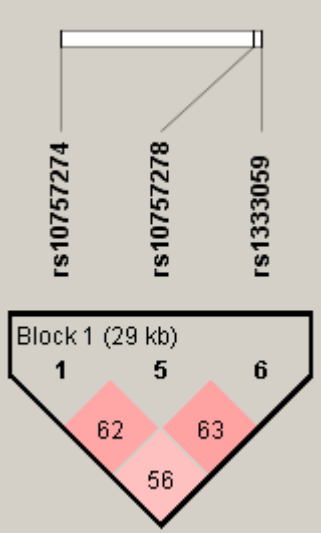

Figure 1. Linkage disequilibrium pattern of the SNPs along the 9p21.3 region. The graphic illustrates the distinct haplotypes defined using the Haploview programme. The linkage disequilibrium (LD) between any two SNPs is shown in the cross cell. LD is presented with standard color schemes, bright red 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 r^2-value.

ROC analysis combining the non-genetic and genetic risk factors

We presented two ROC analysis in our study. The first ROC analysis evaluated that, whether the SNPs identified to be associated with CAD enhance the predictive value of known non-genetic risk factors such as BMI, cholesterol, triglyceride, glucose levels. Therefore two models for CAD prediction based on clinical features and SNPs were built. The first model (the clinical only model) consisted of known CAD risk factors collected in this study: BMI, cholesterol, triglyceride, glucose levels. For the second model (clinical+genetic model), the six associated SNPs (rs1333049, rs2383207, rs2383206, rs10757274, rs10757278 and rs10811656) were entered into the model assuming an additive model of inheritance, in addition to those included in the clinical-only model. SNPs rs10757274, rs10757278 and rs10811656 were significantly associated with CAD (***p=0.0004***), but the other three SNPs rs1333049, rs2383207, rs2383206 were not associated with CAD ($p > 0.05$) after adjusting for other variables in the model and therefore were removed from the clinical+genetic model. The final clinical+genetic model included BMI, cholesterol, triglyceride, glucose, rs10757274, rs10757278 and rs10811656. As shown in Table 3,

both models were significantly predictive of CAD, with area under curve (AUC) of 0.730 (95% CI:0.654-0.797) for clinical only model (***p=0.0001***), and AUC of 0.784 (95% CI:0.713-0.844) (***p<0.0001***) for clinical+genetic model. Most importantly, all three CAD-associated SNPs rs10757274, rs10757278 and rs10811656, improved the predictive power for CAD over the model composing of only non-genetic known risk factors, with an improvement in AUC of 0.0388 (95% CI: 0.0246-0.102, ***p=0.02307***).

Table 3: Comparison between two multivariate models with and without CAD associated SNPs. AUC: area under curve.

Model	Variables	AUC (95%CI)	P value	Difference in AUCs (95%CI) Model 2-Model 1	P value
Non-Genetic Risk Factors	BMI, cholesterol, triglycerides, glucose	0.730 (0.654-0.797)	**0.0001**		
Non-Genetic + Genetic Risk Factors	BMI, cholesterol, triglycerides, glucose, rs10757274, rs10757278 and rs10811656	0.784 (0.713-0.844)	**<0.0001**	0.0388 (0.0246-0.102)	***0.02307***

DeLong et al ., 1988

We performed another ROC analysis to evaluate whether the distribution of SNPs identified to be associated with CAD differs between regions (N-CAD and R-CAD) significantly, two multivariate models were built. The first model (the clinical only model) consisted of known general features including age, BMI, weight and height. For the second model (clinical+genetic model), the six associated SNPs (rs1333049, rs2383206, rs2383207, rs10757274, rs10757278 and rs10811656) were entered into the model assuming an additive model of inheritance, in addition to those included in the clinical-only model. The Second clinical+genetic model included age, BMI, weight, height, rs1333049, rs2383206, rs2383207, rs10757274, rs10757278 and rs10811656. The clinical + genetic model had a significantly higher AUC of 0.818 (***p=0.002***) (Figure 2A) compared to the clinical-only model with an AUC of 0.662, suggesting significant differences between different regions. Moreover, the models used were also used for the estimation of features important to prediction (Figure 2B).

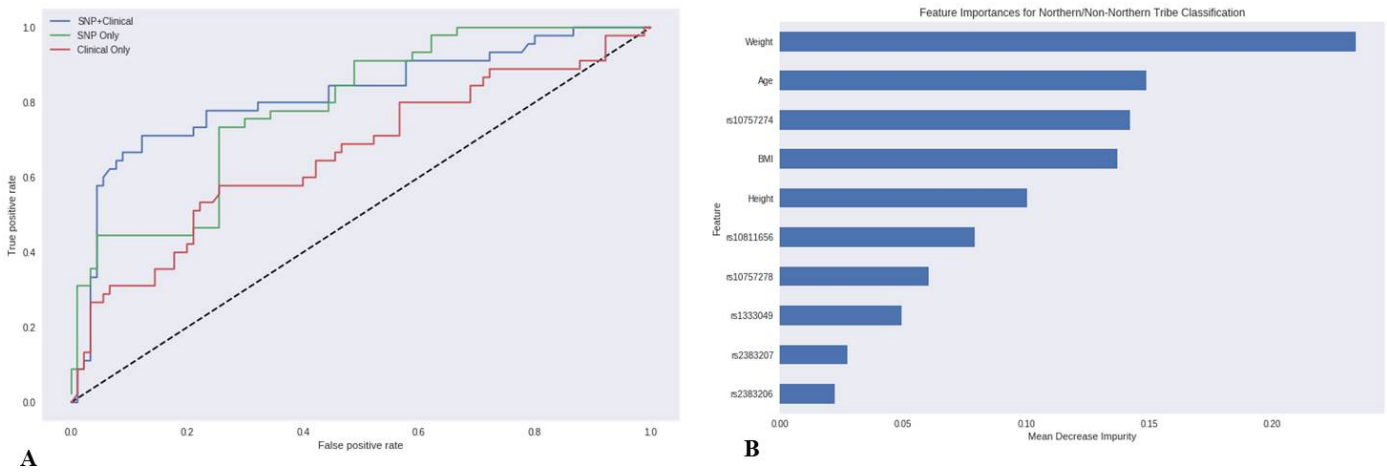

Figure 2: (A) ROC Curve Comparison between models with and without CAD associated SNPs. AUC: area under curve. **(B)** Future importance analysis in Tanzanian CAD patients.

Discussion

For over the centuries, CDs were the main cause of death worldwide. As the results of medical research, the discovery of vaccinations and antibiotics, and improvement of living conditions CDs are declined and NCDs began to cause enormous problems in industrialized countries. SSA is today facing a double burden of NCDs and they still are increasing rapidly. The main NCD is CAD and is the leading cause of disability and death in SSA. And the incidence of CAD is increasing dramatically as a result of urbanization and migrations from rural areas to urban areas in Tanzania. The current population of Tanzania is 59 million and 32.6 % of the population is living in urban areas but the urban population is increasing at a rate of 4.2% per year, compared to 1.9% for the rural population (Unvin et al., 2010). The epidemiological transition is associated with parallel increases in CAD and their risk factors (Mocumbi, 2012). The preventable factors are tobacco smoking, obesity, hypertension, diabetes, raised blood cholesterol levels but the unpreventable factor is the genetic predisposition. Recently, GWAS studies reveal that a significant association with CAD was detected for multiple SNPs across a genomic region of approximately 58 kilobase pairs on the chromosome 9p21.3 and substantial linkage disequilibrium were also determined between the SNPs (Helgadottir et al., 2007; McPherson et al., 2007). This association was demonstrated in Caucasian populations but not in African populations (Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007). The present study was set out to investigate the association of the common SNPs in the 9p21.3 locus with CAD in CAD patients from different regions of Tanzania and CAD risk factors. The majority of our patients was from Northern region of Tanzania which is known as an urban location and were ethnically from Chagga tribe. There are more than 120 distinct ethnic groups and tribes in Tanzania and Chagga tribe is the third largest (total population of Chagga is more than 2 million) ethnic group in Tanzania (Levinson, 1998) They traditionally live on the southern and eastern slopes of Mount Kilimanjaro and Mount Meru (John, 2011) and near Moshi. Many Chaggas works as clerks, teachers, and administrators, and many engage in business activities.

We observed that obesity and BMI were significantly different between N-CAD and R-CAD patients. Our anthropometric data is in agreement with a number of studies stating that obesity and BMI are generally higher in urban compared to rural areas (Unvin et al., 2010; Asprey et al., 2000; Njelekelaet al., 2003). We have also found a high prevalence of dyslipidemia in N-CAD patients compared to R-CAD patients. Moreover, we found significant differences in TC, LDL, TG and HDL levels between N-CAD and R-CAD patients. The prevalence of diabetes and glucose levels were also notably higher in N-CAD patients than in R-CAD patients. In addition, we found that a high prevalence of hypertension in N-CAD patients compare to R-CAD patients. But there

were no significant differences in blood pressure between N-CAD and R- CAD patients which might well be due to the antihypertensive medication use. A number of population based studies in Tanzania on hypertension involving the general population have shown the high population prevalence of hypertension ranging from 19% in rural areas to 35% in urban areas, with the highest prevalence of 70% found among individuals aged 70 years and above (Edwards et al., 2000).

In the present study, we evaluated the common six SNPs on the 9p21.3 locus, rs1333049, rs2383206, rs2383207, rs10757274, rs10757278, and rs10811656, extracted from previous GWAS studies of European populations. They were shown to be associated with the risk of CAD in populations of European ancestry (Helgadottir et al., 2007; Samani et al., 2007). We could only determine significant associations of rs10757274, rs10757278 and rs10811656 with the risk of CAD patients from the northern region of Tanzania which have been replicated in multiple different populations (Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007; Yan et al., 2016; Maitra et a., 2009; Zhou et al., 2008; Shen et al., 2008; Aleyasin et al., 2017). In our study cohort, we could not demonstrate a significant association between rs1333049, rs2383207 and rs2383206 and N-CAD. This is in agreement with previous studies that reported the lack of association between rs1333049, rs2383207 and rs2383206 SNPs and CAD in American African and North African populations (Zanetti et al., 2016; Gong et al., 2011; Silander et al., 2009).

First of all, our results confirm the previous GWAS results indicating the association of 9p21.3 locus with CAD and secondly contribute to the literature by reporting the presence of this association with Tanzanian ethic groups (Yan et al., 2016; Maitra et al., 2009; Zanetti et al., 2016; Foroughmand et l., 2015; Lemmens et al., 2009). The SNPs that we found to be significantly associated with N-CAD, are encompassing the long non-coding RNA *CDKN2B-AS1* located in 9p21.3 locus; rs10757274 is located about 100000 base pairs upper parts of the 9p21.3 locus and it is located in the intronic region of *CDKN2B-AS1* gene. Moreover, other SNPs rs10757274 and rs10811656 are located in one of *CDKN2B-AS1* gene enhancers and disrupt a binding site for STAT1 which would in turn inhibit *CDKN2B-AS1* expression and modulate inflammatory signaling (Harismendy et al., 2011). In the contrary, Rotterdam study based on an elderly population, reported that the rs10757274 and rs10757278 are not associated with CAD (Pilbrow et al., 2012). Also, the rs10757274 was shown not to be associated with CAD in North Indian population (18). Other studies in Africans ancestry (American Africans and North Africans) CAD patients did not address the association between the risk allele of rs10757274, rs10757278 SNPs and phenotypic back-ground of CAD (Zhou et al., 2008; Shen et al., 2008; Assimes et al., 2008; Beckie et al., 2011). The absence of association may be due to the presence of different ethnicity and possible epigenetic factors. Furthermore, we observed in the present study a significant difference in genotype distributions of rs10757274, rs10757278 and rs10811656 variants between genders. Risk genotypes were found to be more frequent in the female and male N-CAD patients than in R-CAD patients. In contrast several investigations done with white and black women, demonstrated a low frequency of rs10757274 and rs10757278 in black women compare to white women (Beckie et al., 2011; Beckie et al., 2011). Moreover, the same study reported that black women have a low frequency of rs1333049, rs2383207 and rs2383206 compare to white women. In our study cohort, we have also confirmed the absence of associations between rs1333049, rs2383207, and rs2383206 and female CAD patients. Besides the similarities, in some point, our data do also differ from the previously reported African data. The differences with those studies might be due to genetic diversity within Africa itself. Besides, the International HapMap Project also showed that Africa is genetically diverse continent than Europe and Asia. Our results also showed the genetic heterogeneity within Tanzania and in Africa. Even though there are few studies which focused on characterizing the genome-wide genetic diversity in African populations (Jones et al., 2015; Disotell, 2012). The International HapMap (Baudot et al., 2009), 1000 Genomes (Abecasis et al., 2010), and especially the African Genome Variation Projects (AGVP) have provided remarkable insight regarding the vast genetic variations existing across African ancestry populations (Gurdasani et al., 2015). Moreover, the gene-gene and gene-environment interactions could also

be different in various ethnic setups, thereby making the magnitude of the effect of the SNPs to be different within Tanzania and Africa. Moreover, the frequencies of SNPs that were common in our population were similar to European population but different than Western and Northern African populations. One might well even explain this genetic proximity of Tanzanian and European populations by a possible common gene pool formed during the migration time around 2.700–3.300 years ago (Pickrell et al., 2014). Furthermore, G-G-G haplotype (rs1333049, rs10757274, and rs10757278) encompassing the long non-coding RNA *CDKN2B-AS1*, was found linked with a significant increase (high-risk haplotypes) to CAD risk in the Northern region of Tanzania. It suggests that SNPs of this locus, rs1333049, rs10757274, and rs10757278, may have a cooperative effect on the rise of the CAD in the Northern region of Tanzania. Two multivariate models of potential prognostic markers were built; the first one included only clinical model and the second one included clinical+genetic model containing CAD-associated SNPs (rs10757274, rs10757278, and rs10811656). According to ROC analysis, the comparison of these two models indicated, clinical+genetic model including to three SNPs rs10757274, rs10757278, and rs10811656 together with CAD risk factors such as age, BMI, total cholesterol, triglyceride, and glucose levels could possibly serve as prognostic biomarker for CAD in Northern region of Tanzania.

As a conclusion, high prevalence of CAD clinical and genetic risk factors imply a continuing burden of CAD morbidity and mortality in Africa. To the best of our knowledge, the present work is the first study demonstrating the genetic association of 9p21.3 locus with CAD in East Africa, and local populations such as Chagga tribe from Northern Tanzania. And the first findings from our study show that the prevalence of CAD is higher in the Northern region of Tanzania which implies the people from the Northern region of Tanzania are prone to CAD more than other regions of Tanzania. rs10757274, rs10757278 and rs10811656 SNPs located on the 9p21.3 locus, are associated with the CAD risk in the Northern region of Tanzania. High CAD risk in the Northern region of Tanzania might well be increased due to the rapid cultural and social changes, aging population, increasing urbanization, dietary changes, reduced physical activity and unhealthy behaviors compared to other regions in Tanzania. Further large cohorts are required to investigate the current genetic associations and to check for their interactions with various genetic as well as environmental factors for CAD in Tanzania.

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

Funding: This research did not receive any specific Grant from funding agencies in the public, commercial, or not-for-profit sectors.

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, Mr. Ismael Adolf and Mr. Ally Luhaga from Muhimbili University of Health and Allied Sciences Genetics Laboratory.

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

References:

Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. (2010) Genomes Project, A map of human genome variation from population-scale Nature 467: 1061-1073

Adebamowo SN, Tekola-Ayele F, Adeyemo AA, Rotimi CN. (2017) Genomics of Cardiometabolic Disorders in Sub-Saharan Africa. Public Health Genomics. 20: 9-26.

Aleyasin SA, Navidi T, Davoudi S. (2017) Association between rs10757274 and rs2383206 SNPs as Genetic Risk Factors in Iranian Patients with Coronary Artery Disease. J Tehran Heart Cent. 12: 114-118.

Aspray J, Mugusi F, Rashid S, Whiting D, Edwards R, Albert G, Unwin N. (2000) Non-Communicable Disease Health Intervention Rural and urban differences in diabetes prevalence in Tanzania: the role of obesity, physical inactivity and urban living. Transactions of the Royal Society of Tropical Medicine & Hygiene 94: 637- 644.

Assimes TL, Knowles JW, Basu A, Iribarren C, Southwick A, Tang H, Absher D, Li J, Fair JM, Rubin GD. (2008) Susceptibility locus for clinical and subclinical coronary artery disease at chromosome 9p21 in the multi-ethnic ADVANCE study. Hum Mol Genet 17: 320-2328.

Baudot A, Real FX, Izarzugaza JM, Valencia A. (2009) From cancer genomes to cancer models: bridging the gaps. EMBO Rep 10: 359–366.

Beckie TM, Beckstead JW, Groer MW. (2011) The association between variants on chromosome 9p21 and inflammatory biomarkers in ethnically diverse women with coronary heart disease: a pilot study. Biol Res Nurs. 13: 306-319.

Beckie TM, Groër MW, Beckstead JW. (2011) The relationship between polymorphisms on chromosome 9p21 and age of onset of coronary heart disease in black and white women. Genet Test Mol Biomarkers.15: 435-442.

Boutayeb A. (2006) The double burden of communicable and non-communicable diseases in developing countries. Trans R Soc Trop Med Hyg. 100: 1919.

Congrains A, Kamide K, Oguro R, Yasuda O, Miyata K, Yamamoto E, Kawai T, Kusunoki H, Yamamoto H, Takeya Y, Yamamoto K, Onishi M, Sugimoto K, Katsuya T, Awata N, Ikebe K, Gondo Y, Oike Y, Ohishi M, Rakugi H. (2012) Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. Atherosclerosis. 220: 449-455.

Dalal S, Beunza JJ, Volmink J, Adebamowo C, Bajunirwe F, Njelekela M, Mozaffarian D, Fawzi W, Willett W, Adami HO, Holmes MD. (2011) Non-communicable diseases in sub-Saharan Africa: what we know now. Int J Epidemiol. 40: 885-901.

DeLong ER, DeLong DM, Clarke-Pearson DL. (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44: 837-845.

Disotell TR. (2012) Archaic human genomics. Am J Phys Anthropol.149: 24–39

Edwards R, Unwin N, Mugusi. (2000) Hypertension prevalence and care in an urban and. J Hypertens. 18:145–152.

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

Gong Y, Beitelshees AL, Cooper-DeHoff RM, Lobmeyer MT, Langaee TY, Wu J, Cresci S, Province MA, Spertus JA, Pepine CJ, Johnson JA. (2011) Chromosome 9p21 haplotypes and prognosis in white and black patients with coronary artery disease. Circ Cardiovasc Genet. 42: 169-178.

González-Santos M, Montinaro F, Oosthuizen O, Oosthuizen E, Busby GB, Anagnostou P, Destro-Bisol G, Pascali V, Capelli C. (2015) Genome-Wide SNP Analysis of Southern African Populations Provides New Insights into the Dispersal of Bantu-Speaking Groups. Genome Biol Evol.7: 2560-2568.

Gurdasani D, Carstensen T, Tekola-Ayele F. (2015) The African Genome Variation Project shapes medical genetics in Africa. Nature 517: 327–332.

Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA. (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. Nature. 470: 264-268.

Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Jonasdottir A, Sigurdsson A, Baker A, Palsson A. et al. (2007) A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science 316: 1491-1493.

Hinohara K, Nakajima T, Takahashi M, Hohda S, Sasaoka T, Nakahara K, Chida K, Sawabe M, Arimura T, Sato A. et al. (2008) Replication of the association between a chromosome 9p21 polymorphism and coronary artery disease in Japanese and Korean populations. J Hum Genet 53: 357-359.

John A. (2011) Ethnic Groups of Africa and the Middle East: An Encyclopedia, Shoup, ABC-CLIO.

Jones B. (2015) Population genetics: the African Genome Variation Project. Nat Rev Genet 16: 68–69.

Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT; Lancet Physical Activity Series Working Group. (2012) Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet. 380: 219-229.

Lemmens R, Abboud S, Robberecht W, Vanhees L, Pandolfo M, Thijs V, Goris A. (2009) Variant on 9p21 strongly associates with coronary heart disease, but lacks association with common stroke. Eur J Hum Genet. 17: 1287-1293.

Levinson David. (1998) Ethnic Groups Worldwide: A Ready Reference Handbook. Oryx Press.

Maitra A, Dash D, John S, Sannappa PR, Das AP, Shanker J, Rao VS, Sridhare H, Kakkar VV. (2009) A common variant in chromosome 9p21 associated with coronary artery disease in Asian Indians. J Genet 88: 113-118.

McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR et al. (2007) A common allele on chromosome 9 associated with coronary heart disease. Science 316: 1488-1491

Mocumbi AO. (2012 Lack of focus on cardiovascular disease in sub-Saharan Africa. Cardiovasc Diagn Ther. 2: 74-77.

National Cholesterol Education Program (NCEP) Expert Panel on Detection. (2001) Evaluation and treatment of high blood cholesterol in adults (Adult Treatment Panel III) executive summary of the third report. JAMA 285: 2486e97

Njelekela M, Sato T, Nara Y, Miki T, Kuga S, Noguchi T, Kanda T, Yamori M, Ntogwisangu J, Masesa Z, Mashalla Y, Mtabaji J, Yamori Y. (2003) Nutritional variation and cardiovascular risk factors in Tanzania--rural-urban difference. South African Medical Journal 93: 295-299.

Pedregosa, Fabian, Gaël VG, Alexandre GA, Vincent MV, Bertrand TB, Olivier GO, Mathieu BM.(2011) Scikit-learn: Machine learning in Python. Journal of machine learning research 12: 2825-2830.

Pickrell JK, Patterson N, Loh PR, Lipson M, Berger B, Stoneking M, Pakendorf B, Reich D. (2014) Ancient west Eurasian ancestry in southern and eastern Africa. Proc Natl Acad Sci U S A. 111: 2632-2637.

Pilbrow AP, Folkersen L, Pearson JF, Brown CM, McNoe L, Wang NM, et al.(2012) The chromosome 9p21.3 coronary heart disease risk allele is associated with altered gene expression in normal heart and vascular tissues. PLoS One 7: e39574.

Roberts R. (2014) Genetics of coronary artery disease. Circ Res. 114: 1890-903.

Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Richard D, Meitinger T, Baraud P, Wichmann E. et al. (2007) WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. N Engl J Med. 357: 443-453.

Shen GQ, Rao S, Martinelli N, L, L, Olivieri O, Corrocher R, Abdullah KG, Hazen SL, Smith j, Barnard J, Plow EF. et al. (2008) Association between four SNPs on chromosome 9p21 and myocardial in¬farction is replicated in an Italian population. J Hum Genet 53: 144-150.

Silander K, Tang H, Myles S, Jakkula E, Timpson NJ, Cavalli-Sforza L, Peltonen L. (2009) Worldwide patterns of haplotype diversity at 9p21.3, a locus associated with type 2 diabetes and coronary heart disease. Genome Med. 1: 51.

Tantchou Tchoumi JC, Ambassa JC, Kingue S, Giamberti A, Cirri S, Frigiola A, Butera G. (2011) Occurrence, aetiology and challenges in the management of congestive heart failure in sub-Saharan Africa: experience of the Cardiac Centre in Shisong, Cameroon. Pan Afr Med J 8: 11.

Teo YY, Small KS, Kwiatkowski DP. (2010) Methodological challenges of genome-wide association analysis in Africa. Nat Rev Genet. 11: 149-60.

Tishkoff SA, Williams SM. (2002) Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet. 3: 611-21.

Unwin N, James P, McLarty D, Machybia H, Nkulila P, Tamin B, Nguluma M, McNally R. (2010) Rural to urban migration and changes in cardiovascular risk factors in Tanzania: a prospective cohort study. BMC Public Health 10: 272.

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 101: 5-22.

World Health Organization The world health report 2002. Geneva: World Health Organization; 2002

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

Zanetti D, Via M, Carreras-Torres R, Esteban E, Chaabani H, Anaibar F, Harich N, Habbal R, Ghalim N, Moral P. (2016) Analysis of Genomic Regions Associated With Coronary Artery Disease Reveals Continent-Specific Single Nucleotide Polymorphisms in North African Populations. J Epidemiol. 26: 264-271.

Zhou L, Zhang,X. He M. Cheng L, Chen Y, Hu FB, Wu T. (2008) Associations between single nucleotide polymorphisms on chromosome 9p21 and risk of coronary heart disease in Chinese Han population. Arterioscler Thromb Vasc Biol 28: 2085-2089.

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.