M.Sc. Thesis – P. Joseph; McMaster University – Health Research Methodology

Title Heading:

Genetic Risk Score and Myocardial Infarction Risk

THE IMPACT OF A CORONARY ARTERY DISEASE GENETIC RISK SCORE ON MYOCARDIAL INFARCTION RISK IN A MULTI-ETHNIC POPULATION: AN INTERHEART STUDY

By PHLIP JOSEPH, M.D., B.A.SC

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

Master of Science (2013); Health Research Methodology; McMaster

University; Hamilton, Ontario

Title: The impact of a coronary artery disease genetic risk score on

myocardial infarction risk in a multi-ethnic population: an INTERHEART study.

AUTHOR: Philip Joseph, M.D., B.A.Sc

SUPERVISOR: Sonia Anand, M.D., Ph.D.

PAGES: 160

Abstract:

Background: Genome wide association studies (GWAS) performed in Caucasian populations have identified several single nucleotide polymorphisms (SNPs) associated with coronary artery disease (CAD), although their cumulative impact in other ethnicities is unknown. Using a genetic risk score (GRS), we examined the impact of CAD related SNPs on myocardial infarction (MI) in a multi-ethnic population.

Methods: We included 4083 MI cases and 4473 controls from the INTERHEART case: control study, stratified by six ethnic groups: European, South Asian, other Asian, Arab, Latin American, and African. We created a GRS comprised of 25 SNPS, and tested its association with MI in individual ethnicities using logistic regression, and across ethnic groups through meta-analyses. Results were adjusted for age, sex, and modifiable risk factors.

Results: The GRS was significantly associated with MI in Europeans (odds ratio [OR] = 1.08, 95% confidence interval [CI] 1.04-1.12 per risk allele), South Asians (OR = 1.09, 95% CI 1.05-1.14), other Asians (OR = 1.09, 95% CI 1.04-1.15), and Arabs (OR = 1.07, 95% CI 1.03-1.12). In Latin Americans and Africans the GRS was not significant. Meta-analysis of ethnic groups demonstrated a 1.06 (95% CI 1.03-1.09) increase in the odds of MI with the GRS per risk allele. Significant heterogeneity was observed, which was reduced by exclusion of Latin Americans (I2=63% to

0%). Above clinical risk factors, the GRS modestly increased population attributable risk (PAR) (0.92 to 0.94), concordance statistic (0.73 to 0.74), net reclassification improvement (0.14), and integrated discriminatory improvement (0.007).

Conclusions: The GRS was associated with a significant increase in the odds of MI in multiple ethnic groups. Improvements in PAR, discrimination and reclassification were modest above clinical factors.

Acknowledgements:

I would like to thank Dr. Sonia Anand for her support and direction as my primary research supervisor for this thesis and my master's degree.

I would also like to thank Dr. Salim Yusuf and Dr. Guillaume Pare for their direction and support both with my thesis and master's degree.

I would like to thanks Dr. Jamie Engert for his support as an external examiner for my thesis.

I would also like to thank Michelle Zhang and Senay Asma for their statistical support for this thesis.

Table of Contents:

List of Figures and Tables

TABLES:

List of Abbreviations and Symbols

ApoA1	Apolipoprotein A1
ApoB	Apolipoprotein B
BMI	Body mass index
CAD	Coronary Artery Disease
C-statistic	Concordance statistic
CVD	Cardiovascular Disease
DALY	Disability Adjusted Life Years
DBP	Diastolic Blood Pressure
dsDNA	Double stranded deoxyribonucleic acid
DNA	Deoxyribonucleic acid
GRS	Genetic Risk Score
GWAS	Genome wide association study
HIC	High income country
IDI	Integrated discriminatory improvement
IDL	Intermediate density lipoprotein
IHD	Ischemic heart disease
IP	Integral ‘1-specificity’
IS	Integral sensitivity
LIC	Low income country
LMIC	Low middle income country
LDL	Low density lipoprotein
MDS	Multi-dimensional scaling
MI	Myocardial Infarction

MIC	Middle income country
NRI	Net reclassification improvement
OR	Odds ratio
PAR	Population attributable risk
PC	Principle component
PCA	Principle components analysis
PURE	Prospective Urban Rural Epidemiology
RR	Relative risk
SNP	Single nucleotide polymorphism
SBP	Systolic blood pressure
VIF	Variance inflation factor
VLDL	Very low density lipoprotein

Declaration of Academic Achievement:

Dr. Philip Joseph was involved in the conception and design of the study, performed the primary statistical analyses, and was the primary author of this manuscript.

Dr. Sonia Anand was the primary supervisor of the thesis. She was involved in the conception and design of the study, review of the statistical analysis and review of the manuscript.

Drs. Salim Yusuf and Guillaume Pare were involved in the conception and design of the study, review of the statistical analysis and review of the manuscript.

Dr. Jamie Engert was involved with review of the statistical analysis and review of the manuscript.

Michelle Zhang and Senay Asma provided statistical support for the analysis conducted by Dr. Joseph.

Introduction:

Although cardiovascular disease (CVD) remains the leading cause of mortality in most regions of the world, the burden of CVD has dramatically shifted from high-income countries (HIC), with 80% of cases now occurring in middle-income countries (MIC) and low-income countries (LIC) (Gaziano, 2005). The INTERHEART study demonstrated that globally, acute myocardial infarction burden is significantly influenced by potentially modifiable risk factors, which account for 90% of the population attributable risk (PAR) for myocardial infarction (MI), with their effects being consistent across several regions of the world (Yusuf et al., 2004).

While the relationship between modifiable factors and coronary artery disease (CAD) risk are well established, the impact of genetic factors and their ability to improve CAD risk stratification requires additional study. Genome wide association studies (GWAS) have now identified single nucleotide polymorphisms (SNPs) in over 30 loci associated with CAD risk, although individually these variants only modestly increase CAD risk (Coronary Artery Disease Genetics Consortium, 2011; Schunkert et al., 2011). To better understand the cumulative CAD risk associated with these genetic variants, several studies have evaluated their impact using genetic risk scores (GRS). Higher scores have been associated with increased CAD risk, and may improve risk prediction above known modifiable risk factors (Brautbar et al., 2012; Davies et al., 2010; Ripatti et al., 2010; Thanassoulis et al., 2012). However, these polymorphisms and genetic scores have primarily

been studied in Caucasian populations, and their effects may significantly differ in other ethnic groups.

Characterizing the impact of known CAD related genetic polymorphisms in a multi-ethnic population is needed in order to understand how these polymorphisms influence genetic risk globally, and whether they can be used to improve CAD risk prediction in multiple ethnicities. This thesis will evaluate the impact of a CAD related genetic risk score in INTERHEART, a multi-ethnic, case-control study examining risk factors for incident MI across 52 countries. First, I will summarize the current evidence surrounding the impact of both modifiable and genetic factors from a global perspective. Secondly, I will examine the association between the genetic risk score and MI across six ethnic groups in INTERHEART. Thirdly, I will determine the added impact of the genetic risk score in addition to modifiable risk factors.

Chapter 1: Modifiable Risk Factors for Coronary Artery Disease Development

Cardiovascular disease, characterized by coronary artery disease, cerebrovascular disease, heart failure and peripheral arterial disease, is the leading cause of death and disease burden globally (Lozano et al., 2012; Murray et al., 2012). Over the past few decades, there has been a dramatic shift of CVD burden from HICs to MICs and LICs. Consequently, while age adjusted CVD related mortality is now declining in many HICs, CVD is the leading cause of mortality in LICs and MICs, where eighty percent of new CVD cases now occur (Gaziano, 2005).

The increased burden of CVD occurring in LICs and MICs is closely related to lifestyle transitions occurring with economic development and urbanization. Low- and middle-income countries are experiencing large increases in cardiovascular risk factors, such as increasing cholesterol levels, blood pressure and body mass index, increasing cardiovascular risk in entire populations.

In the INTERHEART study, nine risk factors accounted for a large proportion of the population attributable risk (PAR) for MI. Smoking, apolipoprotein B/apolipoprotein A1 (ApoB/A1) ratio, hypertension, diabetes, abdominal obesity, psychosocial factors, consumption of fruit and vegetables, alcohol consumption patterns, and physical activity patterns accounted for approximately 90% of the PAR for MI across multiple regions in the world (Yusuf et al., 2004). Since the risk associated with these factors are at least partially modifiable through lifestyle or

pharmacologic interventions, they have been previously classified as modifiable risk factors, which we will also refer to them as throughout this thesis. This chapter will describe global trends in risk factors that their impacts on CAD risk with an emphasis on the nine INTERHEART modifiable risk factors.

1.1 Modifiable Risk Factors Influencing Cardiovascular Disease Development

1.1.1 Blood Pressure and Hypertension

Global Trends in Population Blood Pressure and Hypertension Prevalence

It is estimated that high blood pressure accounts for 9.4 million deaths per year, and 7% of disability adjusted life years (DALYs) globally, making it the leading risk factor for disability worldwide (Lim et al., 2012). Daneai et al. evaluated blood pressure trends in 135 countries and 5.4 million participants from 1980 to 2008. Across the study, there was a 1mmHg reduction in systolic blood pressure (SBP) per decade during the period, corresponding to an estimated reduction in global uncontrolled hypertension prevalence from 33% to 29% in men, and 29% to 25% in women. However, significant regional variability has been observed, with reductions in population SBP in HICs, and increases in SBP in many MICs and LICs. From 1980 to 2008, age standardized SBP has declined in men and women in North America, Western Europe and Australia. During the same period, South Asian, Southeast Asian, Oceanic and East African populations, as well as West African females, experienced significant

rises, ranging from 0.8–1.6 mm Hg per decade in men and 1.0–2.7 mm Hg per decade in women (Danaei, Finucane, Lin, et al., 2011)

Association Between Blood Pressure/Hypertension, and Coronary Artery Disease

Hypertension is associated with an increase in the risk of CAD. In INTERHEART, a history of hypertension was associated with a 1.91 increased odds for MI, accounting for 17.9% of the PAR (Yusuf et al., 2004). Although hypertension is defined condition (e.g. a blood pressure of greater than 140/90 mmHg) the relationship between blood pressure and CAD risk is continuous. In a meta-analysis of observational studies including 1 million adults without vascular disease, a 20 mmHg increase in SBP was associated with a two-fold increase in ischemic heart disease (IHD) related mortality. This relationship was log-linear, down to a SBP of 115 mmHg. A similar effect on IHD related mortality was observed with a 10 mmHg difference in diastolic blood pressure (DBP) in the studied ranges of 75-110 mmHg, with no lower threshold of benefit (Lewington et al., 2002) .

Reduction of Coronary Artery Disease with Blood Pressure Reduction

Several studies have demonstrated that CAD risk can be reduced with lowering blood pressure. In the Blood Pressure Lowering Treatment Trialists' Collaboration meta-analysis of randomized control trials, studies comparing treatment with one hypertensive agent to placebo (with a

corresponding reduction of SBP by 3-10 mmHg) demonstrated a 20% relative risk reduction in the risk of CAD. When trails of aggressive versus conservative blood pressure reduction were compared (corresponding to a mean reduction in blood pressure of 3/3 mmHg) more aggressive strategies was associated with a 19% relative risk reduction in CAD (Neal et al., 2000). Current evidence suggests that while a diagnosis of hypertension is associated with CAD, the risk associated with increasing blood pressure levels and CAD is continuous, even at blood pressure levels considered to be 'normal'. The impact of blood pressure lowering on CAD related outcomes parallels what has been observed in observational studies, with incremental reductions in the risk of CAD proportional to the degree of blood pressure reduction achieved.

Sub-optimal Treatment of Hypertension Globally

Despite clinical trial evidence that pharmacologic agents lower blood pressure and the risk of CAD in those with hypertension, the management of hypertension globally is relatively poor. In the United States, improved recognition and treatment of hypertension has been observed over the last decade, although 20% of individuals with hypertension are still unaware of the diagnosis, and 30% of individuals have inadequate hypertension control (Egan et al., 2010). Global hypertension awareness, treatment and control were shown to substantially worse in the multi-national Prospective Urban Rural Epidemiology (PURE) study in 142 042 participants across 17 countries. In PURE, 41% of participants had hypertension, although only

47% were aware of the diagnosis. Of those aware, 88% received pharmacologic therapy, but treatment with more than one agent occurred in only 31%, and only 33% achieved adequate control. In LICs and Low middle-income countries (LMICs), rates of hypertension awareness and treatment are significantly lower, despite the mortality related to hypertension greatest in these regions (Chow et al., 2013). Given the impact of hypertension on CVD risk, reducing cardiovascular related morbidity and mortality in the future will greatly depend on the ability to overcome barriers that limit hypertension awareness and control, particularly in LIC and MICs.

1.1.2 Atherogenic Lipid Burden

Global Trends in Total Cholesterol

From a population perspective, small changes in serum cholesterol values can have a significant impact on cardiovascular risk. From 1980 to 2008, there have been opposing trends observed in population mean cholesterol levels in several regions of the world. North America, Australia, Europe and central Asia have experienced significant declines in population serum cholesterol during this period, while levels have increased in East Asian, Southeast Asian and Pacific regions (Farzadfar et al., 2011). The increases in serum cholesterol level observed in these regions are mostly the result of dietary transitions that have occurred with economic development and urbanization.

Association Between Atherogenic Lipid Burden and Coronary Artery Disease

The principle components of cholesterol that promote atherosclerosis development are atherogenic lipid particles, such as low-density lipoproteins (LDL), intermediate density lipoproteins (IDL) and very low density lipoproteins (VLDL). Apolipoprotein B is a principle component of LDL, IDL, and VLDL, facilitates VLDL production, and has been shown to mediate interactions between atherogenic lipid particles and the vascular wall. Apolipoprotein A1 is a principle component of high-density lipoproteins.

The relationship between atherogenic lipid burden (measured by serum cholesterol, LDL-cholesterol, ApoB or ApoB/A1 ratio) and CAD risk is well established. In the Prospective Studies Collaboration meta-analysis of 900 000 adults across Europe and North America, there was a graded relationship between total cholesterol and IHD related mortality across several age ranges, with no lower threshold observed. A 1 mmol/L lower total cholesterol was associated with a decrease in the relative risk of IHD mortality by 50% in those ages 40-49 years age, 33% in those 50-69 years, and 17% in those 70-89 years (Prospective Studies Collaboration, 2007). In INTERHEART, the ApoB/A1 ratio was the most significant risk factor for MI, accounting for 54% of the PAR. Those at the highest quintile of ApoB/A1 ratio were at a 3.87 odds increased MI risk when compared to the lowest quintile. In both men and women, the highest ApoB/A1 ratios were observed in South Asians, Arabs, Persians and Latin Americans,

with the lowest ratios in Chinese and black-African populations (Yusuf et al., 2004).

The Impact of Atherogenic Lipid Reduction on Coronary Artery Disease Risk

Atherogenic lipid reduction using pharmacologic agents can significantly reduce CAD, with the proportion of risk reduction directly related to the degree of lipid lowering achieved. In a meta-analysis of 90 056 individuals from randomized control trails comparing statins to placebo, a 1.0 mmol/L reduction in LDL concentration was associated with a 19% reduction in IHD mortality. Similar effects on cardiovascular event risk reduction have been observed for reductions in ApoB related to statin use (Boekholdt et al., 2012). Consistent with observational data suggesting that there no threshold effect to reducing atherogenic lipid burden, further reductions in IHD mortality have been observed with intensive lipid lowering therapy using higher dose statins targeting a greater reduction in LDL (Cholesterol Treatment Trialists Collaboration, 2010).

1.1.3 Overweight and Obesity

Global Trends in Body Mass Index and the Prevalence of Being Overweight or Obese

It is estimated that 1.5 billion individuals worldwide are either overweight or obese. Globally, mean body mass index (BMI) has steadily increased between 1980 and 2008, reaching an age standardized mean of 24.1

kg/m2 in men and 23.8 kg/m2 in women. In men and women, the age standardized mean BMI has increased by 0.4 kg/m^2 and 0.5 kg/m^2 per decade, and increases were observed in most regions of the world, with the largest changes observed in Oceanic countries. It is estimated that these changes have translated into a doubling of obesity prevalence globally from 1980 to 2008, occurring in 9.8% of men and 13.8 of women (Finucane et al., 2011).

Association Between Overweight/Obesity and Coronary Artery Disease

The World Health Organization thresholds for overweight and obesity are >25kg/m^2 and >30kg/m^2 respectively. Abdominal obesity, characterized by increased waist circumference or the ratio of waist circumference to hip circumference (waist-to-hip ratio), is associated with increased visceral fat accumulation, diabetes, metabolic syndrome and cardiovascular disease. General guidelines for increased cardiometabolic disease risk associated with abdominal obesity are a threshold waist circumference of greater than 94 cm in men and 80 cm in women, and a waist-to-hip ratio of 0.90 in men and 0.85 in women (Genest et al., 2009). However, South and East Asian populations demonstrate equal CVD risks to European populations at lower thresholds, suggesting that cut-off values for abdominal obesity should be reduced in these populations to portray equal cardiovascular risk.

Irrespective of the measurement method utilized, overweight and obese individuals are at an increased risk of developing coronary artery

disease. Wormser et al. examined the impact of BMI, waist circumference and waist-to-hip ratio on CAD risk in a meta-analysis of 203 338 individuals across 17 countries. In individuals greater than a BMI of 20 kg/m^2, one standard deviation (SD) increase in BMI (4.5 kg/m^2) was associated with a 1.26 and 1.24 increase in the hazard ratio for men and women respectively. The hazard for CAD was 1.24 in men, and 1.31 in women per 1 SD (12cm) increase in waist circumference; and 1.24 and 1.30 in men and women per 1 SD (0.083) increase in waist-to-hip ratio (Emerging Risk Factors Collaboration, 2011). Similar findings were observed in INTERHEART, where the waist-to-hip ratio was associated with a 1.12 and 1.62 increase in the odds of MI for the second and upper tertile respectively, when compared to the lowest tertile; and accounted for 20.1% of PAR (Yusuf, et al., 2004). In INTERHEART, both waist-to-hip ratio and waist circumference had better predictive ability for MI risk when compared to BMI, although other studies have reported no improvement in prediction or reclassification with these measures (Emerging Risk Factors Collaboration, 2011; Yusuf et al., 2005).

Reduction in Cardiovascular Risk Associated with Weight Loss

Reduction of weight in obese patients is a significant challenge, although both lifestyle (consisting of dietary and physical activity strategies) and surgical interventions have demonstrated consistent weight loss effects (Appel et al., 2011). While these interventions have been shown to improve some measures associated with CVD risk (e.g.

blood pressure, hemoglobin A1c [HbA1c]), there is limited data thus far that they improve cardiovascular outcomes. Wing et al. evaluated the impact of an intensive lifestyle intervention in obese patients with type 2 diabetes, and did not observe a difference in cardiovascular outcomes when compared to usual care after a mean follow up of 9.6 years, although the percent mean weight loss was not drastically greater in the intensive group (6.0%) when compared to the control (3.5%) by study end (Look et al., 2013). Larger reductions in weight loss have been observed with bariatric surgery when compared to intensive medical therapy strategies, and observational data suggests that those undergoing bariatric surgery have a lower long-term mortality (Sjostrom et al., 2004). Although randomized trials have demonstrated improvements in glycemic control, reduction in triglycerides, and increased HDL with surgery compared to medical therapy, there is insufficient clinical trial data evaluating the impact of surgery on CVD outcomes (Schauer et al., 2012).

Uniform strategies to globally reduce obesity may be challenging as obesity is closely related to socioeconomic factors, which differ considerably between regions. For example, in HICs obesity risk is greatest in low socioeconomic groups, while in MIC and LIC obesity risk is associated with higher socioeconomic status (Dinsa, Goryakin, Fumagalli, & Suhrcke, 2012). In addition to broad health promotion strategies, complex strategies will be required to address the variation in behavioral, lifestyle and socioeconomic factors that promote obesity in different regions of the world.

1.1.4 Smoking and Tobacco Use

Smoking and Tobacco Use in Different Regions of the World

Using household surveys between 2008-2010, the Global Adult Tobacco Survey evaluated smoking and tobacco use in 16 countries to characterize and compare smoking prevalence in different regions of the world. Overall, it was estimated that there were 852 million tobacco users worldwide, of which 661 million were smokers, and 247 million used smokeless tobacco. In men, smoking prevalence was highest in Russia (61.5%), China (55.2%) and the Ukraine (50%); and lowest in Brazil (22%) the United Kingdom (23%), the United States (24%), India (24%) and Mexico (25%). Less women smoked compared to men in all countries, particularly in South Asian, most East Asian and Arabic countries, where the prevalence in women was less than 4%. The highest prevalence of female smokers was in Poland (24%), Russia (22%) and the United Kingdom (21%). The highest proportions of young male smokers (ages 15-24) were observed in Russia (53%) and the Ukraine (45%), while the proportion of young female smokers was highest in Russia (33%) and the United Kingdom (29%) (Giovino et al., 2012). Manufactured or hand rolled cigarettes were the predominant method of tobacco in most regions of the world. The prevalence of smokeless tobacco use was highest in India (26%) and the Bangladesh (29%), Thailand (4%) and Egypt (2%). In the remaining 12 countries, the prevalence of smokeless tobacco use was less than 1% (Giovino, et al., 2012). Globally, significant regional

differences exist in tobacco use among men and women, even among countries in the same economic strata, likely reflecting cultural differences that impact patterns of tobacco use.

Association Between Smoking and Coronary Artery Disease Risk

Cross section and prospective observational studies have observed a 1.4-3 fold increase in CAD risk with smoking. The risk of CAD associated with smoking increases both with cumulative and current exposure (D. G. Cook, Shaper, Pocock, & Kussick, 1986; Prescott, Hippe, Schnohr, Hein, & Vestbo, 1998). In a prospective study of 10 427 Canadians enrolled in the 1994 National Population Health Survey (NPHS), current smokers who smoked less than 25 cigarettes per day had a 1.4 fold increase in CAD risk compared to non-smokers, while those smoking greater than 25 cigarettes per day were at two fold increased risk (M. W. K. Shield, 2013). Adjusting for other cardiovascular risk factors, studies suggest that female smokers have a 25% relative risk increase of CAD with smoking when compared to males, although it is unclear whether this difference is related to biological mechanisms or smoking behaviors (R. R. Huxley & Woodward, 2011). Consistent with other observational studies, current smokers in INTERHEART were at a 2.95 odds for MI compared to those who never smoked. Each additional cigarette smoked per day was associated with a 6% increase in the odds of MI. The PAR for MI in individuals who ever smoked was 35.7% (Yusuf et al., 2004).

Although fewer studies have evaluated the association between smokeless tobacco and CAD risk, similar effects have been observed. Meta-analysis of 11 studies evaluating the risks associated with smokeless tobacco report a 1.13 increase in the odds of fatal MI among current users of smokeless tobacco compared to those who never used tobacco (Boffetta & Straif, 2009). In INTERHEART, chewing tobacco was associated with a 2.36 increased odds of MI (Teo et al., 2006). These data suggest that all forms of tobacco consumption are strongly associated with CAD risk.

Reduction in Coronary Artery Disease Risk Associated with Smoking Cessation

Observational studies demonstrate that the risk of CAD decreases with quitting smoking, although large variability has been observed in the rate at which this risk decreases. Some studies suggest that CAD risk returns to that of the general population as little as 5 years after quitting smoking, although other studies demonstrate that an increased risk persists past 20 years (Doll & Peto, 1976; Prescott, et al., 1998; Tverdal, Thelle, Stensvold, Leren, & Bjartveit, 1993). In the NPHS, the relative risk of CAD continued to decline with number of years quitting smoking, although the risk reduction became similar to what was observed in non-smokers after 20 years of continuous cessation (M. W. K. Shield, 2013).

1.1.5 Alcohol Use

Global Patterns in Alcohol Consumption

Sheild et al. examined alcohol consumption patterns were evaluated in 2005 across 241 countries and territories. Overall, 41% of participants consumed alcohol, and 14% were former drinkers. The proportion of current drinkers was higher in men (50%) when compared to women (32%). Three percent of women and 9% of men were considered heavy drinkers (defined as greater than 40g/day in women and 60g/day in men) (K. D. Shield et al., 2013).

Regionally, Australia and Southern Latin America had the highest proportion of individuals who ever consumed alcohol (91% of the population in each country). Conversely, lifetime abstinence form alcohol was highest in North Africa/Middle East and South Asia (82% of the population in each country). Abstinence rates ranged from 13-28% in Europe, and 22% in North America. The proportion of heavy drinkers was greatest in Eastern Europe (12% of women and 25% of men), and Central Europe (9% of women and 22% of men); and lowest in the Oceanic countries (less than 1% of the population). Current data suggest that a large proportion of the population consume alcohol, with 1 in 10 female drinkers and 1 in 5 male drinkers reporting heavy drinking. The proportion of both men and women reporting heavy drinking is particularly high in Eastern and Central Europe (K. D. Shield, et al., 2013).

Association between Alcohol Consumption and CAD Risk

Alcohol consumption has a complex relationship with overall cardiovascular risk, with observational studies demonstrating an overall benefit in CAD related outcomes, but increased stroke risk at higher levels of consumption. In a meta-analysis of 84 prospective observational studies, Ronksley et al. observed that any alcohol consumption was associated with a 0.71 relative risk reduction in incident CAD and a 0.75 relative risk reduction in CAD related mortality when compared to abstinence (Ronksley, Brien, Turner, Mukamal, & Ghali, 2011). Similar degrees of risk reduction were observed for these CAD outcomes at all levels of alcohol consumption. Any alcohol consumption was associated with a 0.75 relative risk reduction in mortality secondary to cardiovascular causes, although the relationship between amount of alcohol consumption and cardiovascular outcomes was ‘U shaped’, due to a higher stroke risk associated with heavy alcohol consumption. A similar relationship between alcohol consumption and incident MI was observed in INTERHEART, where regular alcohol use (greater than 3-4 drinks per week) was associated with a 0.91 odds of MI compared to no alcohol use (Yusuf et al., 2004).

1.6 Diabetes

Trends in Fasting Plasma Glucose and Diabetes Prevalence Globally

The prevalence of diabetes continues to increase globally, with the burden of disease transitioning to MICs and LICs, where two-thirds of cases now

occur. The Global Burden of Disease study characterized changes in glycaemia and diabetes prevalence across 199 countries from 1980 to 2008. In this period, mean fasting blood glucose (FBG) increased by 0.09 mmol/L per decade in women to 5.50 mmol/L, and 0.07 mmol/L per decade to 5.43 mmol/L in men. In parallel, diabetes prevalence increased from 8.3% to 9.8% in men and 7.5% to 9.2% in women during the same period (Danaei, Finucane, Lu, et al., 2011).

Across most regions of the world, population mean FBG either increased or remained unchanged. Regions with the highest prevalence of diabetes were Oceanic regions, North Africa, the Middle East and Caribbean, occurring in 21-25% of men and 21-32% of women. In men, largest increases in FBG were observed in Oceanic countries, Latin America, and South Asia; while in women the largest increases were observed in Oceanic countries, Central Asian, North Africa and the Middle East. Interestingly, an increase in FBG was correlated with increasing BMI in most regions (0.71 in men and 0.51 in women), characterizing the consistent association between weight gain and dysglycaemia across the world (Danaei, Finucane, Lu, et al., 2011).

Association Between Diabetes and Coronary Artery Disease

Vascular disease is the most common cause of death among diabetics. Patients with diabetes are at a similar risk of death from MI as patients with established CAD. Haffner et al. followed patients without diabetes, with diabetes, and with CAD over a mean follow up of seven

years, and observed similar a incidence of MI in those with diabetes (2.7%/year) compared to those with CAD and no diabetes (2.9%/year), which was approximately six times the annual incidence of MI in subjects without diabetes or CAD (0.5%/year) (Haffner, Lehto, Ronnemaa, Pyorala, & Laakso, 1998). A history of diabetes was associated with a 2.07 increase in the odds of incident MI in INTERHEART. Similarly, a large increase in the risk of fatal coronary heart disease has been observed in prospective studies. In a meta-analysis of 37 prospective studies including 447 064 participants, participants with diabetes had a 3.5 increase in the relative risk of fatal coronary heart disease compared to non-diabetics. In women, diabetes was associated with a 1.46 relative risk increase in fatal coronary heart disease when compared to men (R. Huxley, Barzi, & Woodward, 2006). Consistent with other studies, there was an attenuation of risk after adjusting for other cardiovascular risk factors, suggesting that at least a proportion of the excess risk in women can be attributed to clustering of risk factors (Kanaya, Grady, & Barrett-Connor, 2002).

The risk associated with developing diabetes is largely modifiable. Hu et al. demonstrated that female participants without lifestyle factors (defined as a healthy diet, normal BMI, 30 min/day of at least moderate activity, non-smoking, and consuming a half to one drink of alcohol per day) were at a 90% reduced risk of developing diabetes when compared to those with all risk factors (Hu et al., 2001). Given that urbanization and development have resulted in major lifestyle changes affecting diet and physical activity, in addition to increased obesity prevalence, it is not

surprising that diabetes prevalence has significantly increased throughout LICs and MICs. In those at increased risk of developing diabetes, strategies aimed at lifestyle modification have been shown to reduce the incidence of disease development. Given the strong association between diabetes and cardiovascular risk globally, additional strategies are needed which reduce diabetes risk in populations, identify and manage those at increased risk for developing diabetes, and optimize diabetes control in those with the condition.

1.1.7 Dietary Factors and the Prudent Diet

The Nutritional Transition and Dietary Patterns

Vandevijvere et al. defines the nutrition transition as, “changes in dietary patterns and nutrient intakes when populations adopt modern lifestyles during economic and social development, urbanization and acculturation.” (Vandevijvere et al., 2013). Major nutritional transitions have already occurred in HICs over the past several decades, resulting in overconsumption of energy dense foods. In the National Health and Nutrition Examination Survey in the United States between 2003 and 2006, the greatest food categories contributing to total dietary energy intake were breads and rolls (7.2%), and processed pastries (7.2%). Poultry (14.4%) and beef (14.0%) were the largest contributors to protein intake. Fat intake was primarily from oils and miscellaneous fats (9.8%), cheese (8.8%), beef (7.9%) and pastries (7.7%). Primary sources of

carbohydrate consumption were beverages (11.4%), and breads and rolls (10.9%) (O'Neil, Keast, Fulgoni, & Nicklas, 2012).

With rapid access to food sources and urbanization in many MICs and LICs, equally rapid nutritional transitions are occurring in these regions. From 1989-2004 dietary transitions in China have resulted in a reduction in the consumption of rice, wheat, cereals, tubers; and an increased consumption of energy dense foods such as animal products, processed meats (particularly in Urban regions), oils, and dairy products. Interestingly, fresh fruits and vegetables consumption has also increased in both rural and urban regions (Zhai et al., 2009). A similar dietary shift to a higher consumption of meats, prepared food, dairy and eggs, with a reduction in the consumption of cereal products has been observed in South East Asia (Lipoeto, Wattanapenpaiboon, Malik, & Wahlqvist, 2004). In India, consumption of cereals has decreased over the past two decades, although refined carbohydrate consumption has increased, and the largest proportion of energy consumption remains carbohydrates (73% in rural areas and 68% in urban areas). Percent dietary intake from fat has also increased in India from 14% in 1980 to 19% in 2000, with higher consumption reported in urban areas. Protein intake varies widely in India based on region, but consumption of protein, including animal protein, has increased in urban areas. While fresh vegetable intake has increased in both urban and rural areas, average intake is only approximately 50% of the recommended daily allowance (Misra et al., 2011).

In INTERHEART, food consumption was recorded from individuals across 52 countries and grouped into three major dietary patterns by factor analysis: oriental (based on a high consumption of tofu, soy and other sauces), western (based on a high consumption of fried foods, salty snacks and meat) and prudent (based on a high consumption of fruits and vegetables) (Iqbal et al., 2008; Yusuf, et al., 2004). Individuals were evaluated in all three dietary groups, allowing for overlap in dietary practices, and each diet was grouped into quartiles representing increasing consumption of the specific dietary pattern. The highest rates of oriental diet consumption were in South Asia and China; while prudent diet rates were highest in North America, Europe and Australia. Similar rates of western diet consumption were seen in all regions of the world, with approximately one quarter to one third of individuals reporting high consumption of the western diet pattern.

Associations between Dietary Components, Prudent Diets and Coronary Artery Disease

Since diet is comprised of several food types, each of which may impact CAD risk differently, the relationship is complex. Specific dietary components have been shown to impact CAD risk. In a meta-analysis of 13 prospective studies, including 278 459 participants, He et al. examined the impact of fruit and vegetable intake on the incidence of CAD. Servings of greater than 5/day was associated with a 0.83 relative risk reduction in CAD compared to less than 3 servings/day (He, Nowson, Lucas, &

MacGregor, 2007). In a meta-analysis of 20 cohort studies of 1 218 380 individuals, processed meat consumption was associated with a 1.47 relative risk increase in CAD per 50g serving/day. While red meat consumption did not impact incident CAD in this study, other studies have demonstrated an association between both processed and unprocessed red meat consumption and CVD related mortality (Pan et al., 2012). From epidemiologic data, it is difficult to make firm conclusions about the impact of dietary saturated fat intake on CAD risk. A recent meta-analysis of 21 studies including 347,747 participants did not demonstrate an association between incident CAD and saturated fat intake. However, replacement saturated fats with poly-unsaturated fats has been shown to reduce MI or cardiac death, suggesting that the ratio of polyunsaturated to saturated fat intake may be of greater importance (Mozaffarian, Micha, & Wallace, 2010). Increasing consumption of trans-saturated fats is associated with a higher CAD risk (Oh, Hu, Manson, Stampfer, & Willett, 2005).

In INTERHEART, there was an inverse relationship between prudent diet and MI risk. Compared to the lowest quartile of prudent diet consumption, individuals in the highest quartile were at a 0.67 (95% CI 0.59-0.76) odds of MI. The western dietary pattern demonstrated a U shaped relationship with MI risk, where, compared to the lowest quartile, a lower risk was observed in the second quartile (OR 0.87, 95% CI 0.78-0.98), and a higher risk was observed in the third (OR 1.12, 95% CI 1.00-1.25) and fourth (OR 1.35, 95% CI 1.210-1.51) quartile. There was no

relationship between increasing consumption of the oriental diet pattern and MI risk (Iqbal, et al., 2008).

1.1.8 Physical Activity and Sedentary Behavior

Global Prevalence of Physical Inactivity and Sedentariness

A second lifestyle transition occurring as a consequence of economic development and urbanization is decreasing physical activity levels, and increasing sedentary behavior. The prevalence of physical inactivity and sedentariness in 122 countries was reported by Hallal et al. Being physically active was defined as at least 20 min of vigorous activity 3 times a week; at least 30 minutes of moderate activity 5 times a week; or at least 600 metabolic equivalents per week; and encompassed both work and leisure. Worldwide in adults, 33.9% of women and 27.9% of men were inactive, with greater physical inactivity occurring with increasing age. Physical inactivity was highest in HICs and lowest in LICs. The proportion of the population who were physically inactive was highest in the Eastern Mediterranean (43%) and the Americas (43%) and lowest in South East Asia (17%) (Hallal et al., 2012).

Sedentariness was defined as sitting for greater than 4 hours per day, and was reported in 41.5% of adult individuals, also increasing with age. The proportion of individuals reporting sedentariness was highest in Europe (64%) and the Americas (55%), and lowest in South East Asian (23%) (Hallal, et al., 2012).

Association Between Physical Activity, Sedentary Behavior and Coronary Artery Disease

There is a graded relationship between physical activity and CAD risk. In a meta-analysis of 9 prospective cohort studies, increasing leisure time physical activity levels were associated with a reduction in incident coronary heart disease. Compared to those who reported no physical activity, the relative risk of incident coronary heart disease was 0.86 in those engaging in 150 min/week of at least moderate intensity activity, and 0.80 with 300 min/week of activity (Sattelmair et al., 2011). Similarly, sedentary time has been associated with an increased risk of CVD (Wilmot et al., 2012). In INTERHEART, associations between both occupational sedentary behavior and leisure activity were examined. Compared to sedentary occupation levels, mild (e.g. walking) to moderate physical activity at work (e.g. walking, climbing, lifting) reduced MI risk, while the risk of MI associated with heavy physical labor did not differ. The benefit of mild occupational activity was predominantly seen in MICs and HICs, while the benefits of moderate physical activity were primarily observed in MICs (Held et al., 2012). Those who engaged in moderate or strenuous leisure physical activities were also at a lower risk for MI compared to those who did no engage in leisure physical activity. A lower leisure physical activity level was also associated with diabetes (Held, et al., 2012).

1.1.9 Psychosocial Stressors

Association Between Psychosocial Stressors and Coronary Artery Disease Risk

Across regions of the world, the perception of psychosocial stress differs due to cultural, economic and social factors. Still, common patterns of psychosocial stressors have emerged as risk factors for CAD. Among psychosocial factors, the associations between job strain and depression, and CAD have been the most widely examined. Meta-analysis of 13 cohort studies reported job strain in 15% of participants, and observed an associated 1.23 relative risk of incident CAD. Separate meta-analysis of 14 studies reported a 1.40 relative risk of MI related to the presence of depressive symptoms (Van der Kooy et al., 2007).

In INTERHEART, several psychosocial stressors have been identified which impact MI risk globally, including work stress, home stress, financial stress, and specific stressful events. Both frequent work stress (OR 1.38 [99% CI 1·19–1·61]) and permanent work stress (OR 2.14 [1·73–2·64]) were associated with increased MI risk. Similar risks were observed with frequent (OR = 1.52 [1.34–1.72]) and permanent (OR = 2.12 [1.68–2.65]) stress at home. Moreover, depression was associated with an increased risk of MI (OR 1.55 [1.42–1.69]), while a high locus of control (control over one's life circumstances) was associated with reduced risk. Together, these stressors, in addition to exposure to financial stress and stressful events, accounted for 33% of the PAR for MI (Rosengren et al., 2004; Yusuf, et al., 2004).

1.2 Cumulative Risk of Myocardial Infarction Associated with Modifiable Risk Factors

In INTERHEART, smoking and an abnormal ApoB/A1 ratio (top vs. lowest quintile) were the largest risk factors for incident MI. Diabetes, hypertension, abdominal obesity, and psychosocial stressors were also significantly associated with MI risk, but to a lesser extent. Physical activity, moderate alcohol consumption (when compared to no or low alcohol consumption), and daily fruit and vegetable intake were associated with reduced odds of MI.

Results from INTERHEART demonstrate that cumulative burden of modifiable risk factors on MI risk. Across the study, these nine risk factors accounted for 90.4% of the PAR for MI, with the PAR for just smoking, diabetes, hypertension and abnormal ApoB/A1 ratio was 75.8%. Therefore, a majority of MI cases are associated with four risk factors, and all nine risk factors overwhelmingly contribute to MI risk globally. Although those suffering an MI earlier in life are thought to be more genetically susceptible, the PAR for MI in men <=55 years of age and women <65 years of age was greater than that of older individuals (93.8% vs. 87.9%), demonstrating the significant role that modifiable risk factors play in early MI. In all regions studies, modifiable risk factors significantly influenced the development of incident MI (Yusuf, et al., 2004).

Similarly, the cumulative burden of risk factors significantly increases the lifetime risk of developing CVD. In the Framingham study,

Lloyd-Jones et al. estimated the contribution of traditional risk factors to the lifetime risk of CVD in middle age individuals. In a 50-year old male with optimal risk factors, the lifetime CVD risk to 95 years of age was 5%. At the same age, the lifetime risk of CVD with at least two risk factors was 69%. In a 50 year old woman, lifetime risk of CVD with no risk factors compared to at least two risk factors was 8% and 50% respectively (Lloyd-Jones et al., 2006).

1.3 Risk Stratification Tools Utilizing Modifiable Risk Factors

Several risk stratification tools have been developed which use demographic and modifiable risk factors to predict cardiovascular events in individuals. The Framingham risk score is the most widely used risk stratification tool to assess the risk of development cardiovascular events. The score predicts the 10-year risk of CVD using age, sex, smoking history, SBP, LDL, HDL and total cholesterol. Studies demonstrate good risk discriminatory ability in multi-ethnic populations. D'agostino et al. evaluated the Framingham score in seven cohorts comprising five distinct ethnic groups: White, Black, Hispanic, and Native American and Japanese American. The score demonstrated good discriminatory ability across ethnic groups, with concordance statistics (c-statistics) ranging from 0.63-0.79 in men and 0.66-0.83 in women (D'Agostino, Grundy, Sullivan, Wilson, & Group, 2001). In separate studies of Chinese American participants, Italian, and European Mediterranean participants, the Framingham risk score was found to overestimate cardiovascular events,

although re-calibrated scores performed similar to what had been observed in other populations (Liu et al., 2004; Marrugat et al., 2003; Menotti, Puddu, & Lanti, 2000).

The Reynolds risk score estimates 10-year CVD event risk age, sex, total cholesterol, HDL, SBP, family history and high sensitivity C-reactive protein (as a measure of inflammation). In a cohort of 24 558 women in which the score was derived and validated the risk prediction tool demonstrated good discriminatory ability, with a c-statistic of 0.81 (Ridker, Buring, Rifai, & Cook, 2007). In a separate cohort of women, net reclassification improvement increased by 12.9% when compared to the Framingham Risk Score (N. R. Cook et al., 2012). In 10 724 men, the Reynolds risks score demonstrated moderate discriminatory ability (c-statistic = 0.71), although lower than what was observed in women (Ridker, et al., 2007; Ridker, Paynter, Rifai, Gaziano, & Cook, 2008).

Given that coronary artery disease rates vary widely in different regions, the EURO-SCORE was created as a risk stratification tool for estimating cardiovascular disease risk in the European population. The score estimates the 10 year risk of a fatal CVD event using age, sex, total cholesterol, and SBP. Using participant data from 12 studies in different European countries, a modifiable risk factor based prediction tool was developed that demonstrated good discriminatory ability to predict cardiovascular disease among middle age adults (45-64 years), with c-statistics ranging between 0.71-0.84 in the participating regions (Conroy et al., 2003).

One limitation of the Framingham score is that it does not account for several modifiable risk factors that are known to influence MI risk. Using the INTERHEART study, McGorrian et al. derived several multi-ethnic modifiable risk factor scores to predict cardiovascular events in individuals. Of these scores, the 'short' INTERHEART risk score provided the best discriminatory ability while maintaining model parsimony, with an observed c-statistic of 0.71 (0.70-0.73) across ethnic groups (Conroy, et al., 2003). Overall, various modifiable risk factor based risk prediction scores have been shown to have good discriminatory ability for cardiovascular events, with both Framingham and INTERHEART risk score generalizable to multiple ethnic groups. Still, based on observed c-statistics, it appears that most of these scores have only moderate discriminatory ability, suggesting that cardiovascular disease risk prediction can be further improved with additional risk stratification tools.

1.4 Summary:

The global impact of modifiable risk factors on MI risk has been well studied. An overwhelming proportion of MI cases globally are associated with modifiable risk factors, with the nine INTERHEART modifiable risk factors accounting for 90% of the PAR (Yusuf, et al., 2004). Moreover, these risk factors are common across the world, and in many regions their prevalence is increasing, particularly in LIC and MICs. Several risk prediction scores have been developed to determine the individual risk of developing cardiovascular disease, with some validated in multiple ethnic

groups, and most demonstrating moderate discriminatory ability for predicting cardiovascular events. It is of great interest whether novel risk stratification tools can improve upon our current ability to predict cardiovascular outcomes.

Chapter 2: The Impact of Genetic Factors on the Risk of Coronary Artery Disease

Familial genetic studies suggest that approximately 40-60% of CAD is related to heritable factors (Marenberg, Risch, Berkman, Floderus, & de Faire, 1994; Zdravkovic et al., 2002). However, GWAS demonstrate a more modest association between individual genetic variants and CAD risk. Currently, SNPs in over 40 loci have been associated with CAD risk utilizing GWAS, although most are associated with a modest increase MI risk (CARDIOGRAMplusC4D Consortium, 2013; Coronary Artery Disease Genetics Consortium, 2011; Schunkert, et al., 2011). The cumulative risk attributable to CAD related polymorphisms has been evaluated through 'genetic risk scores' (GRS). Most studies examining the impact of GRS have been limited to Caucasian populations, and it is not known whether these observations can be generalized to other ethnic populations. This chapter will review polymorphisms associated with CAD identified in GWAS, the impact of GRS on MI risk, and current limitations in the literature.

2.1 Studies Examining Family History and Heritability as Indicators of Genetic Risk

Family History and Coronary Artery Disease Risk:

Several studies have attempted to crudely estimate the impact of genetic factors on CVD or CAD development. In a prospective study of 117,156 women, Colditz et al. observed a 2.8 relative risk of non-fatal MI and a 5.0 relative risk of fatal MI in those with a parental history of MI less than 60 years of age, when compared to no premature parental history of MI (Colditz et al., 1986). In the Physicians' Health Study of 22 071 men, a family history of MI maternally, paternally or in both parents was associated with a 1.71, 1.40, and 1.85 relative risk of CVD respectively, when compared to no parental history. Similar results were reported in 39 876 women from the Women's Health Study, where a family history of coronary artery disease maternally, paternally or in both parents was associated with a 1.46, 1.15, and 2.05 relative risk of CVD respectively, when compared to no family history (Sesso et al., 2001).

In the Framingham offspring study, Lloyd-Jones et al. evaluated the relationship between parental cardiovascular disease and 8-year CAD risk in middle age offspring without a history of CAD disease. Premature CAD in a parent (<55 years of age in men and <65 years of age in women) was associated with a 2.0 (95% CI, 1.2-3.1) fold increased risk of CAD in men and 1.7 (95% CI, 0.9-3.1) fold increased risk in women. (Lloyd-Jones et al., 2004).

In the INTERHEART study, where a graded relationship was seen between parental history of CAD and the odds of MI. In INTERHEART, one parent with a history of MI after the age of 50 was associated with a 1.67 (95% CI 1.55 - 1.81) increased odds of MI, while a history of MI in one parent less than 50 years of age was associated with a 2.36 (95% CI 1.89 - 2.95) increased odds of MI. A history of both parents with MI less than 50 years of age was associated with the greatest MI risk (6.56, 95% CI 1.39-30.95) (Chow et al., 2011).

Heritability Studies in Twins

Studies of monozygotic and dizygotic twins provide additional support that a moderate proportion of CAD risk is heritable. Studies from a large registry over 20 000 twins in Sweden have demonstrated a significant association between premature CAD related death in one twin and risk of death of the other twin, with the risk greater in monozygotic twins compared to dizygotic twins. In male monozygotic twins, the relative hazard of death was 8.1 (95% CI 2.7-24.5) when a twin had a sibling who died prematurely of CAD (prior to 55 years of age) when compared to those with siblings who did not die prematurely of CAD. Among dizygotic twins, the relative hazard of death was 3.8 (95% CI 1.4-10.5) in those with a twin with who died of premature CAD. Among monozygotic and dizygotic female twins, the relative hazard of death in those with a twin that died prematurely of CAD (prior to 65 years of age) was 15.0 (95% CI 7.1- 31.9) and 2.6 (95% CI 1.0-7.1) respectively (Marenberg, et al., 1994). Estimated

heritability was 0.57 and 0.38 in male and female twins respectively (Zdravkovic, et al., 2002).

Both twin studies and family history studies suggest that a moderate proportion of CAD related risk is attributable to genetic factors. However, these studies do not directly measure genetic factors, but rely on surrogate measures such as family history of premature CAD. Families often share similar lifestyle behaviors and environments, which may persist into adulthood, and also influence CAD risk. Due to this confounding, it is difficult to ascertain whether the observed risks are fully attributable to genetic factors, or multiple factors, both genetic and environmental, which are shared within a family. Still, knowledge of family history of CAD remains a simple tool that adds prognostic information in CVD risk stratification.

2.2 Genome Wide Association Studies

Small differences in deoxyribonucleic acid (DNA) structure or expression can result in marked differences in phenotype. The most common variation in DNA structure between individuals is the SNP, characterized by a single base pair substitution at a given position within the DNA. Common SNPs occur at a frequency of greater than 1-5% in the population (Cirulli & Goldstein, 2010). It is estimated that approximately 11 million common SNPs exist (Hirschhorn & Daly, 2005). A group of alleles may associate with each other in a non-random method, known as linkage disequilibrium. Polymorphisms that are highly associated with one

another, or in high linkage disequilibrium are commonly referred to as haplotype blocks (Slatkin, 2008). Subsequently, a group of SNPs in high linkage disequilibrium with one another can be identified by genotyping a single ‘tag SNP’.

Most GWAS have focused on common diseases, which are hypothesized to have polygenic influences. The most promising ability of the current GWAS study design is the ability to identify multiple common polymorphisms associated with a particular disease. Thus far, case: control designs have been the most efficient method to detect novel associations. GWAS employ a subset of tag SNPs to represent polymorphisms in high linkage disequilibrium. Several methods have been utilized in GWAS to identify tag SNPs for genotyping, including linkage disequilibrium mapping, and with current genotype platforms over 1 000 000 SNPs can be genotyped to represent the genome. However, these techniques are limited to mapping common polymorphisms, and are not robust for the detection of rare polymorphisms (occurring at an allele frequency of <1%) (Hirschhorn & Daly, 2005).

Although GWAS is a robust method to identify multiple genetic associations, some methodological issues warrant consideration. Rather than pre-specifying genes associated with a given disease based on previous literature, GWAS utilize a ‘hypothesis free’ approach where associations between millions of SNPs and the phenotype of interest can be tested. With multiple testing, the risk of a finding a “significant” association by chance increases as the number of false positive

associations increase with the number of association tests performed. Several methods have been developed to reduce the chances of identifying false positive associations. Most GWAS use a conservative Bonferroni corrected p-value of $5x10^{-8}$ (correcting for 1 000 000 independent tests) as the threshold for achieving genome wide significance (Hirschhorn & Daly, 2005). Moreover, significant associations require replication in additional populations. Meta-analyses of GWAS studies can improve the statistical power to detect true associations through larger sample sizes, and provide more generalizable results and effects sizes for significant SNPs.

A second methodological issue in genetic studies is population stratification. Geographically separated populations can have differences in any given allele frequency due to natural selection of favorable genotypes within a population, or genetic drift. The effect of population stratification is greatest when ethnicities from different continents are combined, although even smaller differences within a self-reported ethnic group, or sub-stratification, may also occur (Tian, Gregersen, & Seldin, 2008). If significant variability in the frequency of a polymorphisms exists either between or within a population, false positive associations may be observed, which are due to systematic differences within the population, rather than a true association between the SNP and phenotype of interest (Hirschhorn & Daly, 2005).

Several methods have been employed to minimize the impact of population stratification, including structured association tests, principal

component analysis (PCA) and multidimensional scaling (MDS). Structured association tests use model- or distance-based clustering algorithms to fit ancestral information into ethnic groups. An example of one such analysis software is STRUCTURE (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000). Principle component analysis and MDS represent similar methods in which multiple inter-correlated variables are expressed as a small number of variables representing patterns of variability (Abdi, 2010). Both methods use related alleles to quantify the genetic relationships between individuals through similarity matrices, and then derive a set of principle component (PC) or co-ordinate variables that reflect these relationships. Eigenvectors representing the largest degree of variability are chosen so that their representative principle components also represent the greatest degree of variation (Tian, et al., 2008). Principal components or co-ordinates, derived from PCA and MDS respectively, can both be used to adjust for ethnic variability in a population, and cluster individuals in ancestral groups in order to identify outliers. A further discussion of the methods used in this study will be provided in Chapter 4.

Genome Wide Associations in Coronary Artery Disease

Coronary artery disease is a complex disease with both modifiable and polygenic influences. Several GWAS have evaluated the impact of genetic polymorphisms on CAD risk. Currently, SNPs from approximately 30 loci

have been consistently associated with CAD in large meta-analyses of GWAS.

9p21 Risk Allele and Coronary Artery Disease Risk

Single nucleotide polymorphisms in the 9p21 chromosomal region have demonstrated the most robust association with CAD risk. In 2007, McPherson et al. and Helgadottir et al. simultaneously demonstrated that SNPs in the 9p21 region were associated with an increased risk of MI in "white" Caucasians (Helgadottir et al., 2007; McPherson et al., 2007). Since then, several GWAS in CAD have replicated these findings both in white Caucasians and other ethnicities. In INTERHEART, multiple SNPs within the 9p21 loci were associated with increased MI risk in white Caucasians, South Asians, Chinese, and Latin Americans (Do et al., 2011). In the largest meta-analysis of 14 GWAS studies of European populations (22 233 cases and 64 762 controls), the Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) consortium demonstrated that each risk allele (rs4977574-G) of the 9p21 polymorphisms was associated with 1.29 (95% CI 1.23-1.36) increase in the odds of CAD, far exceeding a level of genome wide significance ($p=1.35x10^{-22}$) (Schunkert, et al., 2011). Similar results were seen in the Coronary Artery Disease (C4D) Genetics consortium meta-analysis of while Caucasian and South Asian populations, with meta-analysis of ethnic specific data demonstrating a 1.20 (95% CI 1.16-1.25) odds

increase in CAD per rs4977574-G risk allele (Coronary Artery Disease Genetics Consortium, 2011).

Despite the consistent association between the 9p21 polymorphisms and CAD risk demonstrated in several studies, its mechanisms of action remains unknown. Polymorphisms within the region largely reside in non-coding regions of DNA, and likely influence nearby gene expression (McPherson, et al., 2007). The closest cluster of genes to the 9p21 region are *CDKN2A* and *CDKN2B*, which are involved in regulation of the cell cycle through the expression of cyclin-dependent kinase inhibitors (McPherson, 2010). In animal studies, polymorphisms within 9p21 have been shown to occupy enhancer regions that affect STAT1 protein binding, which in turn regulate the expression of the nearby *CDKN2B* gene, and *CDKN2BAS* (also known as ANRIL) (Visel et al., 2010). However, the mechanisms by which *CDKN2B* modulate CAD development remain unclear, and further mechanistic studies are necessary to clarify this pathway.

The effect of 9p21 on CAD risk may be modified by diet. In 8,114 participants from the INTERHEART study, Anand et al. examined whether physical activity, smoking and diet modified the effects of 9p21 polymorphisms on MI risk. While no interactions were observed with smoking and physical activity, MI risk associated with the 9p21 risk allele was significantly reduced in participants who consumed a ‘prudent’ diet. Each risk allele was associated with a 1.32 (95% CI 1.18–1.48) increased odds of MI in participants with a low prudent diet score, but not associated

with MI risk (OR 1.02, 95% CI 0.92–1.14) in participants with a high prudent diet score. This interaction was significantly reduced after adjusting for raw vegetable intake, suggesting that the effect of 9p21 was strongly influenced by this component of the prudent diet. Results were replicated in 19,129 European participants, demonstrating that the interaction between fruits and vegetable intake and 9p21 risk was consistent across independent populations and studies (Do, et al., 2011). It is postulated that dietary factors influence risk associated with the 9p21 region through regulation of nearby genes. In a mouse model, deletion of the 9p21 region was shown to altered expression of nearby *CDKN2A* and *CDKN2B*; and a higher mortality was observed in mice exposed to a high fat diet. Still, our current understanding of 9p21 is limited, and further research is needed to clarify the mechanism of this gene-environment interaction.

Additional Genes Associated with Coronary Artery Disease Risk in GWAS

In addition to the 9p21 region, several loci have been shown to influence CAD risk in GWAS. Meta-analyses of GWAS have been instrumental in confirming many initial GWAS associations, identifying false positive effects, and providing additional power to detect new polymorphisms association with CAD risk. The CARDIoGRAM and C4D meta-analyses have provided the most robust results thus far (Coronary Artery Disease Genetics Consortium, 2011; Schunkert, et al., 2011). From these meta-analyses of GWAS studies, SNPs from 29 loci have been shown to

influence CAD risk at or near levels of genome wide significance. A summary of loci associated with CAD and included in this thesis, in addition to their potential functional significances, is shown in Table 1. Recently, meta-analysis of these consortiums has increased the number of identified CAD related SNPs through GWAS to greater than 40 (CARDIOGRAMplusC4D Consortium, 2013).

CARDIoGRAM Meta-Analysis

The CARDIoGRAM meta-analysis included 22 233 cases and 64 762 controls of European ancestry from 14 GWAS studies, with significant loci replicated in an additional 60 738 participants. Ten loci previously associated with CAD were replicated in the meta-analysis (see Table 2.1). With the exception of the rare *LPA* polymorphism (rs3798220-C), the frequencies of the risk alleles were relatively common, ranging from 15-87% in the population. Risk alleles from the previously identified loci were associated with 1.07-1.54 increased odds of CAD. In addition, 13 novel loci were identified and successfully replicated in the study (see Table 2.1). Risk allele frequencies representing these novel loci ranged from 13-91% in the population, and were associated with 1.06-1.17 increased odds of CAD (Schunkert, et al., 2011). As expected, odds ratios for novel loci were lower than that of previously known loci, as smaller GWAS studies were unable to detect these smaller effects. Interestingly, only three of the identified novel loci were also associated with at least one modifiable risk factor.

C4D Meta-Analysis

Discovery phase of the C4D meta-analysis was comprised of four studies, which included 15 420 CAD cases and 15 062 controls from Caucasian (European and South Asian) populations. To prevent false positive results due to population stratification, significant SNPs were identified separately in each ethnicity, and results were meta-analyzed to determine the combined effect across ethnicities. Significant SNPs underwent replication in a population consisting of 21 408 cases and 19 185 controls of European or South Asian ancestry. Eleven previously identified loci were associated with CAD in the discovery population, and five novel loci were identified at a threshold meeting genome wide significance in the combined (discovery and replication) population. Risk allele frequencies in novel loci ranged from 32-80%, and were associated with modest effects (1.05-1.10 odds of CAD per risk allele) (Coronary Artery Disease Genetics Consortium, 2011).

Several observations can be drawn from the novel loci identified through the CARDIoGRAM and C4D meta-analyses. Firstly, most loci associated with CAD through GWAS do not appear to act through modifiable risk factors, but regulate independent pathways affecting atherosclerosis or thrombosis. Therefore, the potential exists for these genes to account for additional CAD risk not attributable to modifiable risk factors. As expected, most SNPs identified are common in the population. Individually most polymorphisms are associated with a less than 10% increase in CAD risk per risk allele. While failing to identify loci that have

large impacts on CAD risk may be disappointing, the results of GWAS studies are expected, as these studies are most fruitful for detecting common polymorphisms, most of which will demonstrate modest effects, rather than rare polymorphisms with large effects. Finally, most polymorphisms have been identified in white Caucasian populations, and their association with CAD risk in other ethnicities is poorly understood.

2.3 Cumulative Coronary Artery Disease Burden Associated with Genetic Variants from Genome Wide Association Studies

With the identification of multiple polymorphisms associated with CAD through GWAS, determining their cumulative impact on CAD risk, and their ability to improve CAD risk prediction in addition to modifiable risk factors, has been of considerable interest. Several studies have attempted to characterize the cumulative CAD risk associated with these polymorphisms through the development of 'genetic risk scores' (GRS).

2.3.1 Creation of Genetic Risk Scores

Genetic risk scores derived from GWAS use risk alleles from multiple loci to characterize the overall genetic risk burden in a given individual. Several methods have been used to create these scores.

The most common method used is the allele counting method, in which a score is computed by adding the total number of risk alleles an individual possesses from a number of candidate SNPs. For each SNP, the risk allele is determined from the previous literature, with bi-allelic

polymorphisms having 0, 1 or 2 copies of the risk allele. The GRS is calculated by adding the total number of risk alleles in all SNPs of interest (Brautbar, et al., 2012; Meigs et al., 2008). Both weighted and unweighted genetic risk scores can be derived using the allele counting method. In the unweighted method, each risk allele is given the same weight and the cumulative score a sum of the number of risk alleles present. Weighted methods give each risk allele a specific weighting in the genetic risk score based on their effect size in the previous literature (Meigs, et al., 2008; Ripatti, et al., 2010).

Alternatively, weighted scores can be derived using logistic regression methods. In a derivation population, optimal weights for each risk allele are determined by optimizing the log-likelihood for the overall regression model. The impact of the score is then tested on a second population. Some studies have demonstrated greater CAD risk prediction with the use of a weighted gene scores derived from regression methods (Davies, et al., 2010).

2.3.2 Studies Examining Coronary Artery Disease Risk and Predictive Ability Associated with Genetic Risk Scores

Using GRS, several studies have evaluated the association between CAD related polymorphisms identified in GWAS and CAD. Ripatti et al. evaluated the cumulative impact of 13 SNPs associated with CAD in GWAS and CAD in both case: control and prospective cohort populations of European ancestry. The GRS was constructed using an allele counting

method, with weighting each allele based on its effects described in previous literature. The highest quintile of GRS was associated with a 1.66 (95% CI 1·35–2·04) risk of CAD compared to the lowest quintile, after adjusting for age, sex and traditional risk factors. Compared to other modifiable risk factors, this risk was lower than that of elevated lipids, and similar to the effects of hypertension. Despite an independent association with CAD risk, the GRS did not improve risk prediction above modifiable risk factors based on the area under the curve c-statistic, or NRI (Ripatti, et al., 2010).

With the identification of additional SNPs, Vaarhost et al. evaluated the impact of 29 CAD related SNPs on CAD risk in 2221 participants followed prospectively for 12.1 years. Genetic risk scores were created using both unweighted and weighted allele counting. After adjusting for traditional risk factors, no effect was observed with the non-weighted score (HR 1.03 per risk allele, 95% CI, 0.95–1.12), while a significant effect was observed with the weighted score (HR 1.12, 95% CI 1.04–1.21 per risk allele). The difference in observed effect sizes suggests that the non-weighted score underestimates risk when compared the weighted method. Neither score improved CAD risk prediction above traditional risk factors by c-statistic, although a marginal improvement in NRI was observed with the weighted score (2.8%, p=0.031) (Vaarhorst et al., 2012).

A similar association with cardiovascular outcomes was observed using a GRS comprised of 13 SNPs in the Framingham cohort of 3,014 participants prospectively followed for a median of 11 years. The score

was constructed using a non-weighted allele counting method. After adjusting for traditional risk factors and family history, the score was associated with a 1.07 (95% CI 1.00–1.15, P=0.04) per allele increase in CVD risk. Addition of the GRS to age, sex and traditional risk factors resulted in marginal improvement of risk prediction when assessed by change in c-statistic (0.819 to 0.822) and NRI (0.17, 95% CI, 0.01– 0.33). In a secondary analysis, a genetic risk score comprised of 29 GWAS identified SNPs did not provide additional risk prediction above the 13 SNP score. Finally, the addition of 89 SNPs associated with modifiable risk factors reduced the impact of the GRS (HR 1.01, 95% CI 0.99 –1.03, P=0.48) (Thanassoulis, et al., 2012). While adding SNPs associated with CAD risk factors may have been expected to result in a more significant association between GRS and CAD, most polymorphisms associated with modifiable risk factors have not been shown to directly impact CAD risk, and including these genes would result in a smaller observed effect by 'diluting' the effects of more strongly associated polymorphisms.

Additional studies in white Caucasian men and women have demonstrated similar results (Ganna et al., 2013). Overall, in multiple studies of white Caucasian populations, a higher number of CAD related genetic polymorphisms is associated with increased CAD risk. Consistent with the effects of individual SNPS, the per risk allele association with CAD was modest, although in one study individuals in the highest genetic risk quintile were at 66% increased risk of CAD compared to those in the

lowest quintile of risk. Genetic risk scores appear to modestly improve CAD risk prediction above what is ascertained from traditional risk factors.

2.4 Limitations of the Current Literature

The majority of studies evaluating GRS have been performed in European populations. Given that genetic risk factors can vary widely between regions and ethnicities, and that the majority of CAD cases now occur in MICs and LICs regions, which include multiple ethnicities, results of GRS studies primarily done in European populations may not be generalizable to the regions suffering the greatest CAD burden. Therefore, it is of considerable interest to determine whether known CAD related polymorphisms are associated with and improve CAD risk prediction in a multi-ethnic population.

2.5 Summary

Genome wide association studies have been fruitful in identifying common polymorphisms in over 30 loci, which individually are associated with modest increases in CAD risk. Moreover, genetic risk is cumulative, as demonstrated in several studies examining GRS, where higher scores are associated with increasing CAD risk. Although independently associated with CAD, this genetic information appears to only modestly improve CAD risk prediction above what is ascertained through traditional risk factors.

Chapter 3: Study Objectives and Methods

3.1 Research Question and Objectives

Research Question:

In a multi-ethnic population, what is the association between a GRS comprised of previously known CAD related SNPs and MI.

Secondary Question: Does the GRS improve the prediction of MI over and above modifiable risk factors as assessed by the AUC C statistic, and does the GRS lead to net reclassification improvement over modifiable risk factors alone

Hypothesis:

In a multi-ethnic population, the burden of CAD related genetic variants is associated with myocardial infarction risk.

Objectives:

The primary objective is to determine MI risk associated with a GRS, comprised of previously known CAD related SNPs, in a multiethnic population.

The secondary objective is to determine the impact of the CAD related GRS above modifiable risk factors in a multi-ethnic population

3.2 Methods

3.2.1 Population:

Study Population:

The INTERHEART study consisted of 29 281 participants (12 461 cases and 14 820 controls) in a case-control design. The objective of the study was to examine risk factors for incident acute MI in 262 centers from 52 countries in Europe, Australia, North America, South America, Asia and the Middle East (Yusuf et al., 2004).

Definition of MI in INTERHEART:

Myocardial infarction was defined by characteristic symptoms and electrocardiographic changes consistent with acute MI.

Selection of Cases and Controls:

Cases were screened from coronary care or cardiology units within 24 hours of symptom onset. For each case, at least one age (+/- 5 years) and sex matched control subject was recruited with no prior history of heart disease, and no symptoms of exertional chest pain. The control group included participants who were hospitalized for non-cardiac reasons, or individuals from the community.

3.2.2 Selection of Single Nucleotide Polymorphisms:

The ability of a gene score to accurately characterize genetic risk is heavily influenced by the selection of appropriate SNPs. Previous studies

have demonstrated that GRS which have included alleles not directly associated with CAD (such as polymorphisms only associated with CAD related risk factors) are less effective at identifying an association with CAD risk compared to risk scores only including robust polymorphisms directly associated with CAD. For this reason we included only polymorphisms which had robust associations with CAD in meta-analyses of GWAS.

I conducted a literature review of GWAS studies Single nucleotide polymorphisms were selected for this study based on their previous associations with CAD risk in the CARDIOGRAM and C4D meta-analyses of GWAS studies. For each CAD related locus identified in the meta-analyses, one polymorphism which met the threshold for genome-wide significance (p-value < 0.5 x 10-8) was selected for genotyping. Although the *PCSK9* gene was slightly under the threshold for genome wide significance in the CARDIOGRAM meta-analysis (9x10-8), we included it in the GRS since it has been replicated in several previous studies both related to CAD and lipid traits (Cohen, Boerwinkle, Mosley, & Hobbs, 2006; Huang et al., 2009; Kathiresan & Myocardial Infarction Genetics, 2008; Schunkert, et al., 2011; Zhang et al., 2013). In total, genotyping was performed on 29 SNPs from independent loci. Twenty-two of the selected SNPs were from the CARDIOGRAM study and evaluated exclusively in a European cohort, while 7 were from the C4D study, representing both European and South Asian populations (Coronary Artery Disease Genetics Consortium, 2011; Schunkert, et al., 2011).

3.2.3 Collection of Blood samples for Genetic Analysis:

Each participant provided 20 ml of non-fasting blood, which was collected and centrifuged within 2 hours of their admission, then immediately frozen at -20 C or -70 C. Samples were subsequently shipped from each enrollment site to a central blood storage facility and stored in liquid nitrogen. Extraction of DNA from buffy coats and genotyping were performed at the Population Health Research Institute, McMaster University.

3.2.4 Genotyping of Samples:

Genotyping was performed in Dr. Guillaume Pare's Genetic and Molecular Epidemiology Laboratory, located in Hamilton, Ontario. A custom GoldenGate 384-plex Assay panel included the 29 SNPs of interest was developed using Illumina's Assay Design Tool. Extraction of DNA from white blood cells was performed on 9438 individuals. Samples were quantitated using Quant-iT™ PicoGreen ® dsDNA Reagent (LifeTechnologies) according to the manufacturer's recommended protocol. Samples were then inputted into a high-throughput, 96-well plate based Veracode® GoldenGate assay (Illumina). Loci were identified using allele specific primer extension and ligation, followed by polymerase chain reaction to amplify regions of interest on Veracode® holographic microbeads. Samples were then analyzed from microbeads using the Illumina BeadXpress Reader, using "IllumiCode" addresses.

3.2.5 Quality Control Analysis of Genotyped Samples:

All genotyped samples underwent a series of quality control evaluations prior to inclusion in the analysis. This process involved an assessment for missing data, sex check, ethnicity check and departure from Hardy-Weinberg equilibrium. Assessment of missing data, sex check and ethnicity checks were performed prior to obtaining the dataset through Dr. Guillaume Pare's laboratory. Evaluation for deviation from Hardy-Weinberg equilibrium was performed using P-LINK software.

Evaluation of missing data:

Abnormality high proportions of missing data suggest genotyping errors occurring at either the participant or SNP level. Participants were excluded from the analysis if greater than 5% of the total number of SNPs were unsuccessfully genotyped. In addition, exclusion of SNPs occurred if greater than 5% of participants were unsuccessfully genotyped for any given polymorphism.

Sex Verification:

As a quality control measure, self-reported sex was cross-referenced with sex specific SNPs that were genotyped in the array. Identified discrepancies between self-reported sex and the genotype data were subsequently excluded from the analysis.

Ethnicity verification and Principal Component Analysis

As previously discussed, population stratification, the presence different ancestral groups in a population, can result in false positive associations driven by differences in allelic frequencies rather than true associations with the phenotype of interest. This is most evidence when populations from different continents are combined. Self-reporting of ethnicity can reduce the impact of population stratification, although misclassification, admixture and sub-population stratification can still result in false positive associations.

Principle component analysis can be used to reduce the impact of population stratification in a study population. Using PCA, multiple inter-correlated variables can be expressed as a small number of orthogonal variables representing patterns of similarity (Abdi, 2010). Moreover, examining patterns between variables, which would not otherwise be readily apparent by examining individual variables, can be utilized to identify outliers. In INTERHEART, ethnicity was self-reported. For the genetic analysis, ethnic groups were categorized into six groups based on common ancestries: European, South Asian, Other Asian, Arab, Latin American and African. For the overall cohort and for each ethnic group, principle component analysis was used to extract information from ancestral related SNPs into principle component variables representing ethnic variability in the population. Supplementary figure 1 demonstrates the clustering of participants into each ethnic group based on PC1 and PC2 variables for the full cohort. Supplementary figure 2a shows the

distributions of participants based on ethnic specific PC1 and PC2 variables. There is no clustering apparent in the Arab, Latin American, Other Asian and African ethnic groups. Although it appears that clustering is present in the European and South Asian groups, this was explained by our thresholds to remove missing alleles and participants with missing genotype data. When these thresholds were reduced from 5% to 1%, clustering in these ethnic groups was no longer apparent (Supplementary Figure 2b). Outliers that deviated from ethnic clusters based on PCA were subsequently excluded from the analysis, reducing the potential impact of population stratification.

Even among participants within an ethnic cluster, variability in genotype frequencies due to subtle ancestral differences can result in false positive associations. This 'sub-stratification' occurring within ethnic groups can also be adjusted for using PCA (Tian, et al., 2008). Because PCs are orthogonal, the majority of linear patterns between data are explained by the first few PCs, although subsequent PC may explain additional non-linear patterns within the data (Tian, et al., 2008). In our analysis, the first five PC variables created for each ethnicity were also used as covariates when individual ethnicities were analyzed in order to adjust for effects due to sub-stratification.

Hardy-Weinberg Equilibrium:

The Hardy-Weinberg (HW) equation allows for the calculation of bi-allelic frequencies in a population. The equation estimates the frequencies of two alleles occurring at random, and is defined as:

$$p^2 + 2pq + q^2 = 1 \quad (1)$$

where p is the major allele frequency and q is the minor allele frequency (Moonesinghe et al., 2010). Deviation from HW equilibrium may suggest a systematic error with genotyping which may result in erroneous results. Several methods have been utilized to assess deviation from HW equilibrium. For this analysis, we assessed for deviation from HW equilibrium using a Chi-square goodness of fit test. Since allele frequencies may vary in different ethnicities and between cases and controls, each SNP was assessed for deviation from HW equilibrium in the control population of each ethnic group. A p-value of less than < 0.0001 was considered statistically significant. SNPs deviating from HW equilibrium in at least one ethnic group were removed from the analysis.

3.2.6 Creation of the Genetic Risk Score:

Twenty-five SNPs that passed quality control measures were included in our GRS. The score was created using a non-weighted allele counting method, with one risk allele for any given SNP contributing to one point in the score (resulting in a maximum score of 50). For each SNP, the risk

allele was defined as the allele associated with increased CAD risk based on the previous literature.

Consideration was given to weighting each risk allele based on effects identified in the previous literature. However, a non-weighted approach was preferred for two reasons. Firstly, in CAD studies weighted gene risk scores have demonstrated only marginal benefits in increasing CAD prediction when compared to non-weighted scores. Secondly, effect sizes have been identified in European or South Asian populations, and given genetic risks vary between ethnicities, weights applied to SNPs may not be generalizable to a multi-ethnic cohort. To reduce this potential source of error, we chose a non-weighted risk score for our analysis, which is consistent with previous studies examining GRS in multi-ethnic populations (Anand et al., 2013).

3.2.7 Collection of modifiable risk factor variables

History and Physical Measurements:

All participants were administered a standardized questionnaire and physical measurements. Information was collected on baseline demographics (age, sex, self-reported ethnicity), cardiovascular risk factors (hypertension, diabetes) and behaviors (smoking history, alcohol consumption, fruit and vegetable intake, and physical activity) and psychosocial factors (depression, financial stress, locus of control, perceived stress and stressful life events). Waist and hip circumferences were measured using standardized non-stretchable tape measures at the

narrowest point between the costal margin and iliac crest, and at the widest diameter around the buttocks respectively.

Collection of Lipoproteins:

Blood samples for lipoprotein analysis were collected together with genetic samples. Since ApoB and ApoA1 levels do not significantly differ from the fasting to non-fasting state, samples could be non-fasting, and were obtained at different post-prandial periods, ideally within 24 hours from the onset of symptoms in MI cases.

Samples were centrifuged and immediately frozen at -20 C to -70 C, then subsequently transferred to the PHRI for storage at -160 C. ApoA1 and Apo B were measured with the Roche Hitachi 917 analyzer, using standardized methods and laboratory kits.

3.2.8 Definition of modifiable risk factor variables

Smoking history was stratified into current smokers, former smokers, and those who have never smoked. Current smoking was defined as smoking any tobacco product within the past 12 months, while former smokers had quit smoking at least one year prior to enrollment. We defined physically active as engaging in at least moderate exercise for greater than 4 hours per week. Moderate alcohol consumption was defined as consuming 3 or more alcoholic drinks per week. We characterized fruit and vegetable intake as daily consumption or fruits and vegetables, daily consumption of either fruits or vegetables, or the lack of daily fruit or vegetable intake.

Waist to hip and ApoB to ApoA1 ratios were divided into groups by tertiles and quintiles respectively, using the control group as the reference population. In the original INTERHEART study, five variables were used to characterize psychosocial factors. For this analysis, we used the global stress measure as the representative psychosocial factor.

3.2.9 Statistical Analyses:

Analysis of Individual Single Nucleotide Polymorphisms

Effects were determined for each of the 25 candidate SNPs included in the score. For each SNP, an estimate of the effect was determined stratified by ethnicity using logistic regression. For each model, the dependent variable was MI, and the independent variable was the candidate SNP. An additive model was used for each SNP, where and each SNP treated as a continuous variable with and given a number ranging from 0-2 based on the number of risk alleles present. Therefore, resulting odds ratios represent the per risk allele odds of MI associated with a given SNP. To adjust for confounding, additional co-variables included in each model were age, sex and five additional variables generated from our principle components analysis, which was added to adjust for sub-stratification.

After obtaining odds ratios and variance estimates for each ethnic group, results for each SNP were meta-analyzed to provide a final estimate of the effect. Meta-analyses were performed using the R-META package in the R statistical program. Each ethnic group was weighted according to the inverse of the variance estimate observed. Chi-square

and I^2 methods were used to test for heterogeneity between ethnic groups. A fixed effects model was used to determine effect sizes and 95% confidence intervals, unless moderate heterogeneity was observed, defined as Chi-square test for heterogeneity <0.10 or an I2 >50%, in which case a random effects model was used ("Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0," 2008). In total, 25 meta-analyses were performed to determine the effect sizes of each SNP included across our multi-ethnic cohort.

Analysis of the Genetic Risk Score

The association between GRS and MI was determined in each ethnic group using logistic regression. In the univariate analyses, incident MI was the dependent variable, and the continuous GRS was the independent variable. Multivariable analysis included age, sex, and variables representing the nine INTERHEART modifiable risk factors as covariates. Results for each ethnic group were reported as odds ratios with 95% confidence intervals, representing the per allele odds of MI. Logistic regression analyses for the genetic risk score was performed in SAS version 9.2 for Windows.

To determine an overall effect of the GRS in the entire cohort, results of the multivariable analysis for each ethnicity were meta-analyzed, using the R-META package in the R statistical program. The same weighting methods and methods to determine heterogeneity were used as described in the previous section (2.9.1).

Finally, to determine whether substratification within ethnic groups influenced MI risk associated with the GRS, sensitivity analysis within each ethnic group was performed by adjusting for PC variables in the multivariable models, and results were meta-analyzed to determine the overall effect across ethnic groups. A p-value of less than 0.05 was considered to be statistically significant.

Assessment for Multi-Collinearity in the Adjusted Logistic Regression Models

The inclusion of highly correlated variables in a regression model can result in inflated standard deviations and confidence intervals. Simple correlation coefficient testing can assess for continuous variables that are strongly associated, signaling the need to remove one variable from the model. However, correlation coefficient testing does not adequately characterize relationships between categorical variables, and whether high levels of correlation exist between multiple variables (multi-collinearity). Variance inflation factor (VIF) testing uses linear regression to test for high levels of correlation between multiple variables included in a regression model. We used VIF to test for multi-collinearity among the variables included in our gene risk score adjusted logistic regression model. The presence of significant multi-collinearity associated with a given set of variables was suspected if the VIF was greater than 10 (Kutner, 2004). Testing for multi-collinearity was performed in SAS version 9.2 for Windows.

Analysis of Myocardial Infarction Risk Discrimination with the Genetic Risk Score Using the Concordance Statistic:

A commonly used method of determining the overall discriminatory ability of a logistic regression model can be determined by evaluating how it discriminates those having an event from those who do not. The concordance statistic (C-statistic) is a rank based measure that can be used to measure the discriminatory ability of a logistic regression model. In all possible pairings of individuals who have and have not experienced an event, the C-statistic represents the proportion of pairs where the model based predicted probability of an event was higher in the individual who actually experienced the event. (Royston & Altman, 2010). If all cases have a higher predicted risk than all controls the model has perfect discriminatory ability, characterized as a C-statistic of 1. Conversely, a C-statistic value of 0.5 represents a model that is unable to discriminate between those who have had the event and those who have not. A C-statistic between 0.5-0.7 suggests limited discriminatory ability, between 0.7-0.8 moderate discriminatory ability, and >0.8 good discriminatory ability between those with and without the event (Kennedy, 2010). When additional variables are added to a model, the change in the C-statistic represents the added discriminatory ability of the new model provided by the new variables.

Although the C-statistic is a good measure of the discriminatory ability of a model, there are some limitations of the measure when

characterizing risk prediction. A change is the C-statistic represents a higher proportion of individuals that have the disease being correctly classified, but it does not evaluate changes in risk among individual participants. Therefore, it is an indirect measure of risk prediction. Moreover, variables with moderate to large effects may still only result in small changes in the C-statistic if there is no significant change in the sum ranking, minimizing their perceived utility despite significantly increasing risk (Kennedy, 2010). It is necessary to acknowledge these limitations when interpreting the C-statistic as a risk prediction tool.

For this analysis, C-statistic calculations were used to compare the discriminatory ability of the GRS and modifiable risk factors. Firstly, the C-statistic for a logistic regression model containing age, sex, ethnicity and the GRS was calculated, reflecting the discriminatory ability of a model containing genetic risk predictors and demographic variables. We also calculated the C-statistic for a model containing modifiable risk factors and basic demographic variables. Finally, added discriminatory ability was determined by adding the GRS to the model containing modifiable risk factors and demographic variables, and calculating the change in c-statistic between both models. We used the Homer-Lemeshow test to assess for goodness of fit of the models to ensure that addition of the genetic risk score did not worsen model calibration. A p-value of <0.05 was considered a statistically significant change in the C-statistic. Evaluation of the C-statistic was performed in SAS version 9.2 for windows.

Net Reclassification Improvement and Integrated Discriminatory Improvement Associated with the Genetic Risk Score

The net reclassification improvement (NRI) and integrated discriminatory improvement (IDI) represent additional measures to evaluate changes in risk prediction associated with a new risk marker or test when compared to a standard measure. In general, they are more sensitive to changes the discriminatory ability of a novel predictor when compared to the C-statistic.

The NRI measures the number of individuals whose predicted risk of developing a disease changes with a new risk marker or test. Unlike the C-statistic it directly measures changes in risk prediction for each participant, then calculates the percentage of individuals that demonstrate an appropriate change in risk. The NRI is calculated using the following formula (Kennedy, 2010; Pencina, D'Agostino, & Steyerberg, 2011):

NRI = [(# events moving up/ # of events) – (# events moving down/# of events)] + [(# non-events moving down/ # of non-events) – (# non-events moving up/# of non-events)]

Both categorical and category free NRI methods have been developed to assess risk prediction. The categorical NRI uses pre-defined categories of risk for the sample population, and characterizes changes in risk as changes from one category to another. Different categories should represent clinically meaningful differences in risk. For example, using

clinical prediction tools, patients with coronary artery disease are typically classified by their 10-year risk of a major adverse cardiovascular event into low (<10%), moderate (10-20%), or high (>20%) risk groups; with different therapeutic targets Using these thresholds of risk, a categorical NRI test could be used to determine the percent of individuals that demonstrate an appropriate change in risk category (e.g. from <10% to 10-20%) based on a novel risk marker. When categories of risk cannot be established in a population, or when the prevalence the disease is not known in the study population, a categorical free NRI may be calculated, which represent the percent of individuals that demonstrate any change in predicted risk in the appropriate direction (Pencina, et al., 2011).

While the NRI evaluates the proportion of participants in which the test improves risk prediction, it does not test the degree of improvement that occurs, particularly using the category free model in which small changes in improvements in predictive ability result in changes in the index.

The IDI is a measure of the average improvement in risk prediction that occurs with a new risk marker or test. The test measures the average improvement in sensitivity, in addition to improvement in '1-specificity', by comparing models without (old model) and with (new model) the novel predictor of interest. Therefore the test characterizes the average improvement in sensitivity of a new predictor without a reduction in specificity (Kerr, McClelland, Brown, & Lumley, 2011). The IDI can be

represented as follows (Kennedy, 2010; Pencina, D'Agostino, D'Agostino, & Vasan, 2008):

IDI = (IS new model – IS old model) − (IP new model – IP old model)

Where IS represents the integral (or average) sensitivity and IP represents the integral (or average) '1-specificity' of each model. The IDI can also be estimated using the predicted probabilities of events between models (Pencina, et al., 2008):

IDI = (probability of an event in new model – probability of an event in old model) + (probability of a nonevent model old model – probability of a nonevent in new model)

We compared logistic regression models, which included modifiable risk factors and demographic variables with and without the genetic risk score, and calculated the NRI and IDI between both models. Since an accurate measure of disease incidence is required to correctly interpret the categorical NRI (which is not possible in a case: control design), we did not employ this method. Instead the category free NRI was used to characterize the proportion of individuals in our study that experienced any appropriate change in risk. Moreover, we calculated the IDI in order to determine the average improvement in the predictive ability of the model associated with the genetic risk score. Calculation of the NRI and IDI was

performed using the statistical software 'PredictABEL' program R, for Windows.

Population Attributable Risk Associated with the Genetic Risk Score:

Population attributable risk (PAR) is useful to characterize the risk associated with a risk factor within a population, and represents the percent of cases that would not occur if this factors were absent in the population. The PAR is dependent both on the effect of the risk factor, and its prevalence in the population (Natarajan, Lipsitz, & Rimm, 2007):

PAR = [Prevalence of exposure (Relative Risk of Exposure-1)]

/ [1 + (Prevalence of exposure (Relative Risk of Exposure-1)]

In a case: control study, the relative risk of the exposure may be approximated by the odds ratio. In addition to characterizing the PAR associated with a single risk factor, the PAR associated with multiple risk factors can also be determined, with or without adjusting for important covariates.

In this analysis, we determined PARs associated with the GRS; with modifiable and demographic factors; and the increase in PAR associated with the GRS above what is attributable to modifiable and demographic factors.

For calculation of the PAR, continuous risk factors were categorized into categorical variables in order to provide a measure of the prevalence of each factor. The genetic risk score was categorized into quartiles, and since the PAR is meant to determine the effect associated with absence of a risk factor, we compared the top three quartiles of the genetic risk score to the lowest quartile (representing the 'absence' of this risk factor). Similarly, we compared the top four quintiles of ApoB/A1 ratio to the lowest quintile, and the top two tertiles of waist-to-hip ratio to the lowest tertile. Calculation of the PAR was performed using the software program, IRAP for windows.

Exploratory Analyses:

Subgroup Analysis:

It is hypothesized that genetic polymorphisms exert a greater effect in those who present with CAD at an earlier age. Some studies support this hypothesis, which demonstrated an increased risk of MI associated with having a first degree relative with early coronary artery disease (Lloyd-Jones, et al., 2004). Moreover, the effects of some CAD related SNPs have been shown to be slightly larger in cohorts examining MI at an early age (Ardissino et al., 2011). However, in INTERHEART, the PAR related to modifiable risk factors was higher in those who presented with MI before the age of 50, accounting for the majority of MI cases even at an early age (Yusuf, et al., 2004).

To further evaluate whether the impact of the genetic risk score is modified by age, we performed subgroup analysis stratified by quintiles of age and tested for interactions between subgroups. Although stratifying into smaller subgroups of age would reduce our statistical power to identify significant interactions, if age were a true effect modifier, we would expect to observe large effect sizes at very young ages.

We also evaluated whether sex modified the association between the genetic risk score and the odds of MI. Interaction testing was performed using Wald's test, with a p-value < 0.05 considered statistically significant.

Impact of the Genetic Risk Score after Adjusting for Family History

In INTERHEART, cases and controls were asked about a maternal or paternal history of CAD. Given that a family history of CAD is associated with increased MI risk, and that this risk is thought to be mediated by genetic factors, we performed an exploratory analysis evaluating whether the effect of the GRS changed after adjusted for family history. For this analysis, family history was defined as a mother or father with CAD. In each ethnic group, family history of CAD was added to the multivariable analysis which previously included demographic variables, modifiable risk factors, and the GRS. Odds ratios were then compared to the models which did not include a family history of CAD.

3.2.10 Statistical Power

Primary outcome:

We calculated the power to detect a significant association between the continuous genetic risks score and MI. For statistical power calculation, I used an alpha of 0.05, a disease prevalence of 0.07, and a standard deviation of the genetic risk score of 3. From previous genetic risk score studies, effect sizes of scores ranged from 1.03-1.12. Assuming a 5% drop out rate of participants due to errors in genotyping, our study would have 4500 participants in each group. This would result in 80% power to detect a significant effect at an odds ratio of 1.02, and over 90% power to detect a significant effect at an odds ratio of 1.04 (see Supplementary Table 1)

Individual Risk Alleles

Although the primary outcome was the association between the genetic risk score and MI risk, we also calculated the statistical power to identify significant associations in individual risk alleles. We assumed an alpha error of 0.05. We used allele frequencies and effects from the literature in white Caucasians to determine the sample size needed to achieve 80% and 90% power. A summary table of ranges of sample sizes is provided in Supplementary Table 2.

Chapter 4: Results

4.1 Quality Control Measures

In total 882 participants and 4 SNPs were excluded due to failure of a quality control measure (*PHACTR1* [rs1332844], *SH2B3*, [rs3184504], *PDGFD*, [rs974819], *TCF21* [rs12190287]). A summary of excluded participants and polymorphisms due to failure to meet specific quality control tests is summarized in Figure 1. After quality control testing, 8556 participants (4083 cases and 4473 controls) were selected for analysis. Twenty-five SNPs were selected for inclusion in the genetic risk score.

In the remaining 8556 participants, the GRS was missing in 431 (5%) of individuals due to at least one missing allele. There was no significant difference in the missing number of scores between cases and controls (p=0.14). With regards to the modifiable risk factors the following number of individuals with missing data were observed: 234 (2.7%) for fruit and vegetable intake, 72 (0.8%) for hypertension, 246 (2.8%) for smoking, 303 (3.5%) for global stress, 307 (3.6%) for waist to hip ratio, 79 (0.9%) for diabetes, 159 (1.8%) for alcohol, 125 (1.5%) for physical activity, and 19 (0.2%) for ApoB/A1 ratio.

4.2 Baseline Characteristics of the Study Population

A summary of baseline characteristics of the study population is summarized in Table 2a. The mean age of the study population was 55.7 years old, and 79% of participants were male. The mean GRS was 26.27(SD 2.96) in the study population. The GRS was distributed normally

in both cases and controls (Supplementary Figure 3). Median GRS was 26 (interquartile range, 24-28). Non-communicable diseases associated with CAD were common, with 29.5% of individuals reporting a history of hypertension, and 14.6% having diabetes. Behavioral factors associated with MI were also common in the population. 37.9% of study participants were active smokers, and 22.4% reporting a history of prior smoking. The majority of individuals (87.7%) reported consuming less than 3 alcoholic beverages in a week. Only 16.3% of individuals engaged in regular physical activity. 17.9% of participants did not consume fruits or vegetables daily. 19.8% of participants reported several period of stress at home or at work over the past 12 months, and 7.43% of individuals reported permanent stress.

A summary of baseline characteristics of the study population stratified by ethnicity is provided in Table 2b. Mean age of participants was highest in Europeans (59.8 [12.1]) and Latinos (58.5 [12.4]), and lowest in Africans (49.5 SD [10.3]). Mean genetic risk score was highest in South Asians (26.58 (2.85)), and lowest in Arabs (25.95 [SD 3.10]). There were significant differences in the burden of modifiable risk factors between ethnic groups. The prevalence of hypertension was highest in Europeans (37.9%) and Latinos (37.3%), which also represented the oldest ethnic groups in our cohort. The number of participants with a waist-to-hip ratio >0.95 (the upper tertile value in control participants) was highest in Latinos (45.2%) and lowest in South Asians (36.0%). A significantly elevated APOB/APOPA1 ratio was also common in the population, with 27.7% of

the study population having a ratio greater than 1.02. The prevalence of this abnormal value was greatest in Africans (37.9%) and lowest in Europeans (20.9%). Diabetes was highest in Africans (17.1%) despite being the youngest ethnic group in our study.

Cardiovascular risk related behaviors also varied significantly between ethnic groups. Smoking history (either current or former smoking) was highest in Latinos (71.5%) and lowest in Arabs (55.9%). Alcohol consumption was highest in Europeans (24.3%) and lowest in Arabs (3.3%) and Africans (3.5%). Daily fruit or vegetable consumption was lowest in Africans (78.8%) and highest in South Asians (85.1%). Moderate physical activity was lowest in Arabs (7.0%), Africans (7.3%) and South Asians (10.2%), and highest in Europeans (28.2%). Permanent stress was highest in Latin Americans (11.1%) and lowest in Africans (5.1%).

4.3 Individual Single Nucleotide Polymorphisms and their Association with Myocardial Infarction Risk

4.3.1 Frequencies of Individual Risk Alleles

In general, the frequencies of the literature derived risk alleles were similar to what has been observed in the prior literature. In 10 of the 25 SNPs, the literature derived risk allele was the major allele across ethnicities. In the remaining 15 SNPs, the literature derived risk allele was the minor allele, or varied between the major and minor allele according to ethnicity. A summary of allele frequencies for candidate SNP included in the genetic risk score is summarized in Table 3.

4.3.2 Odds of Myocardial Infarction Associated with Individual Risk Alleles

Effects in Europeans Compared to Previous Studies in European Populations:

To determine whether effects of each risk allele were similar in our population when compared to previous studies, we examined whether effects in Europeans were similar to what has been observed in CARDIOGRAM or C4D. In 8 SNPs, the effect of the literature based risk allele was similar to what was observed in these reference studies. Fifteen SNPs did not demonstrate any significant effect in our European study population. Only one SNP had a significant effect in the opposite direction when compared to the white European population in our reference studies.

We also compared the effect of the risk allele in our European group, to what had been observed in Europeans populations in our reference studies by meta-analyzing effects and examining heterogeneity. Odds ratios and confidence intervals were available for all but four SNPs from our reference studies (*ADAMST7* [rs4380028], *KIA1462* [rs2505083], and *7q22* [rs10953541]). In 17 of the remaining 21 SNPs, there was no significant heterogeneity observed between the effect of the risk allele in our study population, and what was observed in Europeans in our reference studies (Supplementary table 1).

Effects in the Overall Study Population, and Differences between Ethnicities:

To determine the overall effect of each SNP on the odds of MI in the study population, separate logistic regression analyses was performed in each ethnicity and these results were meta-analyzed (see Supplementary Appendix, Figure 1). For each logistic regression model, age, sex and five PC variables (from within ethnic group) were included as covariates.

For 9 of the 25 candidate SNPs, the literature derived risk allele was associated with a significant increase in the odds of MI in the entire study population. The largest effect observed was associated with the 9p21 polymorphism, which increased the odds of MI by 1.21 (95% CI 1.10-1.32) per risk allele. Other significant SNPs were associated with smaller effects, increasing the odds of MI by 1.07-1.14. For 16 SNPs, the literature derived risk allele was not significantly associated with MI.

To explore whether results were consistent across ethnic groups, we tested for heterogeneity between ethnic groups in each meta-analysis. Heterogeneity was calculated using the chi-square test for heterogeneity and the I^2 value. Consistent with previous literature, a p-value of <0.10 on Chi-square analysis or an $I^2 > 50\%$ was our threshold for heterogeneity in meta-analysis (Higgins, 2008). For 18 SNPs, the effects were consistent across ethnicities. For 7 SNPs, significant heterogeneity was observed, suggesting that the effects of these risk alleles differed significantly between the ethnic groups. For six of these SNPs (the exception being 9p21) overall results were non-significant.

4.4 Association between Modifiable Risk Factors and Myocardial Infarction Risk

Associations of modifiable risk factors and odds of MI are summarized in Table 4. The modifiable risk factor with the largest impact on the odds of MI was ApoA/B1 ratio, with the highest quintile associated with a 3.50 (95% CI 2.95-4.16) odds of MI compared to the lowest (reference) quintile. Current Smoking (OR 2.83, 95% CI 2.50-3.20), permanent stress (OR 2.00, 95% CI 1.61-2.49), and diabetes (OR 2.24, 95% CI 1.93-2.60) were each associated with a greater than two fold increase in MI risk. Regular physical activity (OR 0.86, 95% CI 0.75-1.00), daily fruit or vegetable intake (OR 0.83, 95% CI 0.72-0.95 when compared to neither daily) and daily fruit and vegetable intake (OR 0.79, 95% CI 0.68-0.91 when compared to neither daily) were each associated with a lower odds of MI.

4.5 Association between the Genetic Risk Score and Myocardial Infarction Risk, and Differences between Ethnicities

Individual Ethnic Groups

In the univariate analysis, the GRS was significantly associated with MI in Europeans, South Asians, Other Asians, and Arabs. In Latin Americans and Africans, the GRS was not significantly associated with MI (Table 5). Multivariable Analysis demonstrated similar results. Per risk allele, the associated between GRS and MI was largest in Europeans (OR 1.09, 95% CI 1.04-1.12), South Asians (OR 1.09, 95% CI 1.05-1.14), and Other

Asians (OR 1.09, 95% CI 1.04-1.15). The effect of the GRS was lower but still significant in Arabs (OR, 95% CI 1.07 (1.03-1.12). There was no significant effect observed in Latin Americans (OR 0.99, 95% CI 0.95-1.03) or Africans (OR 1.04, 95% CI 0.98-1.10). There was no multi-collinearity observed after VIF testing.

Meta-Analysis of Genetic Risk Score Effects across Ethnic Groups

Adjusted results of the GRS in each ethnic group were meta-analyzed. In the entire cohort, the GRS was associated with a 1.06 (95% CI 1.03-1.09) increase in the odds of MI per risk allele (figure 3). Significant heterogeneity was observed (I^2 for heterogeneity = 63%), which improved after exclusion of the Latin American subgroup (I^2 for heterogeneity = 0%)(Figure 4). Results were similar following the addition of PC covariates to our third multi-variable model (Figure 5), suggesting that subgroup stratification was not a significant confounder in the association between the genetic risk score and MI.

4.6 Subgroup Analyses

Age

After stratifying by quintiles of age, we did not identify a significant interaction between age and the GRS (p-value for interaction = 0.22) (see Table 6a). Furthermore, I performed a sensitivity analysis excluding Latin Americans given that the effect of the gene score differed significantly in this group compared to the other ethnic groups (Table 6b). In this analysis,

results were similar, with no significant interaction between quintiles of age and the GRS (0.13).

To determine whether the number of risk alleles was greater in younger participants (less than 55 years of age) compared to older participants, the mean GRS in cases for each age group were compared. In the full cohort, there was no significant difference in the mean number of risk alleles in cases less than 55 years of age compared to cases greater than or equal to 55 years of age (p=0.17).

Sex

After stratifying by sex, we did not identify a significant interaction between sex and the GRS (p value for interaction = 0.08) (Table 6a). Furthermore, we performed a sensitivity analysis excluding Latin Americans given that the effect of the score differed significantly in this group compared to the other ethnic groups. In this sensitivity analysis, results were similar, with no significant interaction between sex and the GRS (p value for interaction = 0.18) (Table 6b).

4.7 Impact of the Genetic Risk Score after Adjusting for Family History

After adjusting for a family history of CAD in either parent, the effect of the GRS remained significant in Europeans (OR 1.08 [1.04-1.12]), South Asians (OR 1.10 [1.05- 1.14]), other Asians (OR 1.10 [1.04-1.15]), and Arabs (OR 1.07 [1.03- 1.12]). There was no significant effect demonstrated in Latin Americans (OR 0.99 [0.95-1.03]) or Africans (OR

1.04 [0.98- 1.11]). Overall, effect sizes associated with the GRS remained similar after inclusion of a family history of CAD.

4.8 Population Attributable Risk Associated with Modifiable Risk Factors and the Genetic Risk Score

A GRS >23, corresponding to the highest four quintiles of risk, was associated with a PAR of 0.32 (0.25-39). The PAR associated with the nine modifiable risk factors in our analysis was 0.90 (0.87-0.92). Addition of the GRS to modifiable risk factors increased the PAR associated with these factors to 0.92 (0.90-0.94). There was no difference in estimates of the PAR after exclusion of the Latin American subgroup (tables 7a and 7b).

4.9 Discriminatory Ability and Reclassification Improvement with the Genetic Risk Score

C-Statistic

Using the C-statistic, we evaluated the discriminatory ability of the GRS (in addition to baseline demographics) in our population. In addition, we examined the discriminatory ability of the GRS above modifiable risk factors. Our logistic regression model which included the GRS, age, sex and the remaining five ethnic groups as covariates resulted in a C-statistic of 0.57 (95% CI 0.56-0.59). Comparatively, the C-statistic of our logistic regression model including age, sex, ethnicity and modifiable risk factors was 0.73 (95% CI 0.72-0.74). The addition of the genetic risk score to our

model that included demographic variables and modifiable risk factors resulted in significant but modest improvement in the C-statistic (0.74, 95% CI 0.72-0.75, p-value for chance < 0.0001). Sensitivity analysis excluding Latin Americans demonstrated similar results. C-statistic results are summarized in Tables 8a and 8b.

NRI and IDI

We further examined the added predictive ability of the genetic risk score above modifiable and demographic factors using the NRI and IDI. Using the non-categorical approach, the addition of the genetic risk score resulted in a NRI of 0.137 (95% CI 0.087-0.188). The calculated IDI after the addition of the genetic risk score was 0.007 (95% CI 0.005-0.010)

Chapter 5: Discussion

In our multiethnic population a higher GRS, was associated with increased odds of MI in Europeans, South Asians, other Asians and Arabs. There was no significant association observed in Latin Americans and Africans. In the entire cohort, the GRS was significantly associated with MI, although the effects of the score significantly differed in Latin Americans compared to the other ethnic groups. Despite a significant association with MI, the cumulative impact of these polymorphisms was relatively small compared to the effects of modifiable risk factors, and provided modest additional ability to predict MI.

5.1 Associations with Modifiable Risk Factors and MI Risk

We performed multivariate logistic regression analysis to determine whether the associations between modifiable risk factors and MI were similar to the larger INTERHEART study. Overall effect sizes were smaller when compared to the larger INTERHEART study, although the directions of associations were the same (see table 4). Similar to the larger study, the largest effects on MI risk were seen with an elevated ApoA1/ApoB ratio and current smoking. Both physical activity and consumption of fruits and vegetables were protective. While moderate alcohol intake was associated with a reduction in MI risk in the full study (OR 0.79, 95% CI 0·73–0·86), the effect was not significant in our substudy (0.96, 95% CI 0.82-1.13). Since we selected a subgroup of participants from INTERHEART, we expect that effect sizes would differ from the larger

study. However, the observed effects were similar for most modifiable risk factors, supporting their use as covariates in our study sample.

5.2 Association Between Individual SNPs and Myocardial Infarction Risk in a Multi-Ethnic Cohort

Based on our meta-analyses of European groups in our study with reference GWAS studies, most polymorphisms demonstrated similar effects with what was observed in the literature, with only 4 of 21 SNPs demonstrating at least moderate heterogeneity (supplementary table 1). Across our six ethnic groups, 9 of the 25 polymorphisms studied were associated with MI. The 9p21 risk allele conferred the largest risk, with each risk allele associated with a 1.21 (95% CI 1.10-1.32) increase in the odds of MI. The observed effect in our multi-ethnic cohort was similar to what has been demonstrated in previous studies in Europeans and South Asians (Coronary Artery Disease Genetics Consortium, 2011; Schunkert, et al., 2011). Interestingly, there was moderate heterogeneity observed with the effect of the 9p21 risk allele between ethnic groups, with the largest effect observed in the South Asian group (OR 1.38, 95% CI 1.21-1.56), and the smallest effect observed in the Latin group (OR 1.06, 95% CI 0.91-1.23) However, consistent with the previous literature, the direction of the observed effect of the 9p21 risk allele was similar across multiple ethnicities (Ding et al., 2009; Dong, Wang, Wang, & Ding, 2013).

Eight additional risk alleles were significantly associated with MI in our study, although their effects were smaller, with odds ratios ranging

from 1.07-1.14 per risk allele. Given that risk alleles and genetic effects can differ significantly between ethnic groups, we expected some heterogeneity to occur between ethnic groups. For 7 SNPs, significant heterogeneity was observed.

Although only 9 of our candidate polymorphisms were significantly associated with an increased odds of MI, these results were expected. For individual SNPs, statistical power is affected by effect size and allele frequency. Based on the literature derived effects and risk allele frequencies of our 29 candidate genes in white Caucasians, our study would have 80% power for 23/29 SNPs, and 90% power for 15/29 alleles at a two sided alpha of 0.05 (see Supplementary Table 2). Therefore, the study was underpowered to detect significant effects in a large proportion of SNPs.

Moreover, significant heterogeneity was observed for 7 SNPs, which is expected when evaluating the impact of genetic markers in a multi-ethnic cohort. Polymorphisms identified through GWAS are usually not causal, but are alleles in high linkage disequilibrium with the causal allele. Since the linkage disequilibrium between SNPs is known to vary significantly between ethnicities, SNPs associated with MI in one ethnicity may not be in linkage with the causal mutation in another. For this reason, the effects of some risk alleles varied considerably between ethnic groups, and in some cases were in opposite directions, reducing their overall effect. In our analysis, only 1 of the 7 alleles (9p21) that demonstrated significant heterogeneity was significant in the overall population.

5.3 Association between Genetic Risk Score and Myocardial Infarction Risk in our Multi-Ethnic Cohort

In our meta-analysis of ethnic groups, the GRS was associated with a modest per allele increase in the odds of MI (OR 1.06, 95% CI 1.04-1.08). Similar results within ethnic groups were observed after adjusting for demographic and modifiable risk factors, demonstrating that genetic risk burden in our overall cohort was independent of traditional cardiovascular risk factors. We did not expect that the association between the GRS and MI risk would significantly change after adjusting for modifiable risk factors, since the dispersion of these alleles in the population would most likely occur at random, and would not be influenced by the development of cardiovascular risk factors later in life. Still, some SNPs share pathways with modifiable risk factors, therefore it was important to demonstrate that the association between the GRS and MI was independent of these factors (Schunkert, et al., 2011; Teslovich et al., 2010).

The association between the GRS and MI in this multi-ethnic cohort was similar to what has been observed in CAD studies of white Caucasian populations (Coronary Artery Disease Genetics Consortium, 2011; Schunkert, et al., 2011). In previous studies, the per allele risk of CAD observed in white Caucasian populations ranged from 1.03-1.12 (Davies, et al., 2010; Ripatti, et al., 2010; Thanassoulis, et al., 2012; Vaarhorst, et al., 2012). In our study, largest effects were observed in Europeans, South Asians, other Asians and Arabs. In these ethnic groups, the score was significantly associated with the odds of MI. In both Africans and Latin

Americans, there was no significant effect observed with the GRS. Moreover, meta-analysis of ethnic groups demonstrated significant heterogeneity, which was reduced by exclusion of the Latin American group, suggesting that impact of the GRS significantly different in Latin Americans compared to other groups.

The observed differences with the GRS in Latin Americans may be explained by differences in linkage disequilibrium between candidate alleles and causative polymorphisms. Many studies have demonstrated differences in the linkage disequilibrium of groups of SNPs among different ethnicities (Hirunsatit et al., 2007; Lohmueller et al., 2006; Nakajima et al., 2002). Since patterns of linkage disequilibrium can vary significantly between ethnic groups, associations between candidate polymorphisms and the phenotypic trait of interest can also vary significantly. Seven of our 25 candidate SNPs demonstrated at least moderate heterogeneity when individual SNPs were evaluated, which may have contributed to the overall heterogeneity observed in the effects of the GRS. Latin American groups have varying and complicated ancestral patterns based on ancestral informative markers. For example, Latin groups residing in Brazil demonstrate significant and variable admixture comprised of European, African and Aboriginal populations (Manta et al., 2013). It is possible that haplotype blocks related to the SNPs of interest differed significantly when compared to our other ethnicities, resulting in the observed heterogeneity. Unfortunately, ethnic variability in haplotype

structure associated with our CAD related candidate SNPs has not been well studied, but remains an important area of future research.

In Africans, the GRS was also non-significant (OR 1.04, 95% CI 0.98-1.16), although the direction of effect was consistent with the four significant ethnic groups in which the score was significant, and no significant heterogeneity in the effects of the GRS was observed between these five groups. A smaller observed effect is possible given that haplotype blocks are much smaller in those of African ancestry, and resulting differences in linkage disequilibrium patterns may result in weaker associations between the candidate SNPs and causal mutations. Also, at the observed effect size, our study would have less than 60% power to detect a significant effect (supplementary table 1). Therefore, we may have been underpowered in this ethnic group.

5.4 Effect of Age and Sex on the Genetic Risk Score

We performed pre-specified subgroup analyses to evaluate whether specific factors modified the effect of the genetic risk score. Some studies suggest that individuals who present with a MI at an early age may be at an increased genetic risk compared to those who present later in life. In the Framingham offspring study, of middle age men and women (mean age of 44 years), a premature history of parental CAD was associated with a 1.7-2.0 increase in the relative risk of CAD. Similar results were observed in INTERHEART, where a graded relationship was observed between the number of parents with early MI, and the odds of MI in the

study. In fact, a history of both parents with MI less than 50 years of age was associated with a 6.56 (95% CI 1.39-30.95) fold increase in the odds of MI. In a GWAS by the Myocardial Infarction Genetics Consortium of 2,967 cases of early-onset MI and 3,075 age-sex matched controls, eight SNPs were associated with early MI, with per allele odds ratios that were slightly higher than what has been observed in meta-analyses of CAD GWAS. However, as observed in INTERHEART, many those presenting early with MI also have a greater burden of modifiable risk factors. In our cohort, we did not observe a significant change in the effect of the genetic risk score when stratified by age. When participants were stratified by quintiles of age, there was no significant interaction observed with the effect of the genetic risk score between subgroups, suggesting that the cumulative risk of these polymorphisms is not influenced by age. Our analysis suggests that the cumulative impact of known polymorphisms associated with CAD through GWAS is similar in those who present with CAD at an early age or later in life. However, it is important to note that our study was not powered to detect interactions with age. Moreover, our analysis focused on common SNPs identified in GWAS, while the association between family history and MI may be mediated through rare, as of yet unidentified alleles with larger effect sizes.

5.5 Population Attributable Risk Associated with the Genetic Risk Score

In our study, a genetic risk score >23 resulted in a PAR of 32%. Comparatively, the PAR associated with modifiable risk factors was much greater (90%). Importantly, there was only a modest increase in the PAR with the addition of the genetic risk score to modifiable risk factors (92%). This is likely due to the high prevalence of modifiable risk factors in the population. In our study, over 90% of the control population had at least one modifiable CVD risk factor. Therefore, the added risk attributable to the genetic risk score would be low given that most individuals had concomitant risk factors.

5.6 Genetic Risk Score and Risk Prediction

Using the c-statistic, we estimated the ability of the genetic risk score to discriminate between MI cases and controls, and the added discriminatory ability of the genetic risk score after accounting for modifiable factors. Our logistic regression model including age, sex, ethnicity and the genetic risk score had a modest predictive ability in discriminating MI risk, with an observed c-statistic of 0.57 (0.56-0.59). Our logistic regression model, which including age, sex, ethnicity and modifiable risk factors, had better predictive ability, with an observed c-statistic of 0.73 (0.72-0.74), suggesting that the discriminatory ability of demographic and modifiable risk factors is much greater than that of demographic and genetic factors. Moreover, addition of the genetic risk

score to our model which included age, sex, ethnicity and modifiable risk factors resulted in a modest improvement in the c-statistic, from 0.73 (95% CI 0.72-0.74) to 0.74 (95% CI 0.72-0.75).

We further examined the added predictive ability of genetic risk score to modifiable and demographic risk factors with the NRI and IDI. Using the uncategorized NRI, the genetic risk score resulted in a 13.7% increase in reclassification when compared to modifiable and demographic risk factors alone. Although the change in NRI suggests a significant improvement in risk prediction, using a non-categorical approach, even small increases in MI risk would result in the reclassification of individuals. The change in IDI was 0.007, and together with a small change in the C-statistic, suggests that the degree of improvement in risk prediction was modest.

5.7 Summary of Findings

This study is the first to demonstrate that a genetic risk score comprised of CAD related polymorphisms identified in white Caucasian populations, is also associated with MI risk in a multi-ethnic population. The genetic risk score was significantly associated with MI risk in While Caucasians, South Asians, other Asians and Arabs. Although the genetic risk score was not significant in our African subgroup, our study was underpowered to detect a significant difference. Moreover, heterogeneity testing suggests that the effect in Africans was similar to the other significant ethnicities. The impact of the genetic risk score in Latin

Americans differed when compared to the other ethnicities, with no significant effect observed in this ethnic group. Even though the genetic risk score was significantly associated with MI risk, the impact was modest, with each risk allele associated with a 1.06 (95% CI 1.03-1.09) increase in the risk of MI in the entire cohort, and its' effect only marginally increased (OR 1.08, 95% CI 1.06-1.10) with the exclusion of the Latin American group. Although the PAR associated with the genetic risk score was 0.32, it was smaller than that of modifiable risk factors, and only modestly increased risk above what was due to modifiable factors. In keeping with this, the genetic risk score was also shown to improve risk prediction above modifiable and demographic factors, although the added predictive ability was modest, and similar to what has been observed in previous studies of European populations. Several factors likely account for these findings. Firstly, the candidate polymorphisms that have been identified through GWAS studies have modest effects both in previous analysis and in our multi-ethnic cohort. Therefore, it would be unexpected to see large effect sizes associated with a score comprised of these risk alleles.

Secondly, modifiable risk factors associated with MI are also common in the population. In our study, over 90% of control subjects had at least one modifiable risk factor present. Given the high prevalence of these factors, the addition of genetic factors may only result in modest improvements in the attributable risk, and risk prediction, compared to modifiable factors alone. Our finding support this premise, as

demonstrated by small improvements in the PAR, c-statistic, NRI and IDI when the genetic risk score was added to demographic and modifiable factors.

5.8 Limitations

There are some limitations of this study that warrant consideration. Firstly, our genetic risk score only included polymorphisms that were directly associated with MI risk in GWAS. Many modifiable risk factors (e.g hypertension, diabetes, and obesity) and behaviors (e.g. smoking) associated with CAD also have genetic influences. Therefore, additional genetic effects occurring through these modifiable risk factors, but not directly influencing MI risk, would not be captured in the gene score. However, previous studies have evaluated the impact of genetic risk scores comprised of SNPs associated with both CAD and modifiable risk factors, and have observed either no or marginal improvement in their impact when polymorphisms associated with modifiable risk factors and CAD were incorporated. Therefore, based on previous studies, we would not expect an increase in the impact of the genetic risk score with the inclusion of additional SNPs related to modifiable risk factors.

Secondly, our genetic risk score was derived from polymorphisms that were identified primarily in Caucasian populations (either European, or South Asian and European). As previously discussed, candidate polymorphisms identified through GWAS are usually not the causative mutation for a disease, but are in linkage disequilibrium with causative

mutation(s). In different ethnicities, linkage disequilibrium among groups of SNPs can vary significantly. Subsequently, a candidate SNP may be associated with the causal trait in one ethnicity, but not in another. This important difference between ethnicities may explain the lack of any association observed in Latin Americans, and the heterogeneous effect of the genetic risk score in this ethnic group. Still, in our other ethnic groups, results were fairly homogenous.

Finally, a large proportion of individuals in our population had modifiable risk factors. In this context, the added predictive value of the gene score above already known modifiable risk factors would be expected to be small. Our results may not be generalizable to individuals without risk factors for cardiovascular disease. It is possible that in this patient population, a genetic risk score may result in greater MI risk prediction compared to our observations. Further research is needed to evaluate the impact of genetic risk scores in participants with otherwise no risk factors, and to determine whether such an approach is a cost effective method to improve risk stratification.

5.9 Future Directions in Research

Our study adds to the current evidence suggesting that known CAD related genetic risk polymorphisms identified from GWAS have a moderate cumulative impact on MI risk in a multi-ethnic population, and marginally improve risk prediction above what can be ascertained from modifiable risk factors. Based on the current literature, there appears to be

modest gain from genetic based risk prediction tools that use currently identified polymorphism in the general population. However, additional steps are needed to further understand the impact of genetic variability on CAD and improving risk prediction, particularly in multi-ethnic populations.

Given that candidate polymorphisms identified in Caucasians through GWAS are most likely not causal, and may not have the same relationship in other ethnicities, further work is needed to identify the causative polymorphisms for CAD development. Identifying these polymorphisms requires fine mapping of regions in which candidate alleles lie, and may uncover polymorphisms with significantly greater effects than what has currently been observed in GWAS. Since the burden of CAD now resides in regions spanning multiple ethnic groups, identifying causative polymorphisms with consistent effects across ethnicities may be most helpful in developing risk prediction tools that can be used globally. Moreover, by identifying the causative polymorphisms through fine mapping techniques, a greater proportion of genetic variability associated with MI may be explained.

Although GWAS have been fruitful in identifying common SNPs that have moderate associations with CAD risk, the overall genetic impact explained by these polymorphisms has been modest. It is estimated that currently known polymorphisms associated with CAD explain approximately only 10-15% of CAD heritability (Schunkert, et al., 2011). With common alleles explaining a relatively small amount of CAD related risk, there is increasing interest in the ‘rare polymorphism hypothesis’,

where additional genetic CAD risk may be explained by rare polymorphisms (occurring at a frequency of less than 1% in the population) that have much larger effect sizes and cannot be identified by current GWAS techniques (Hirschhorn & Daly, 2005). Whole exome or genome sequencing involves the sequencing of all base pairs residing within exomes, or the entire genome respectively. These methods can detect rare polymorphisms that are not identifiable using current GWAS methods. Currently, technical and cost limitations have prohibited the use of these methods in large population studies, with most studies focusing on familial populations, or studies examining the extremes of specific traits (i.e. blood pressure) to facilitate the identification of candidate genes. With improvements in technology, and reductions in the cost of genotyping, it may become feasible to genotype large cohorts through exome or whole genome sequencing techniques to potentially identify rare polymorphisms with larger effects that may further improve genetic based risk prediction.

5.10 Conclusions

Our results suggest that the overall burden of disease associated with known CAD related genetic polymorphisms is modest, and smaller than that of modifiable risk factors. Our findings are consistent with the results of previous studies in Caucasian populations, and provide a more generalizable estimation of the impact of these polymorphisms on CAD genetic burden globally by examining their effects in a multi-ethnic cohort. Based on the current evidence, it appears that currently identified genetic

polymorphisms associated with CAD through GWAS have a modest impact on CAD burden, and only marginally improve CAD risk prediction.

Bibliography:

Abdi, H. W., L. (2010). Principle Component Analysis. *WIREs Comp Stat, 2*, 433-459.

Anand, S. S., Meyre, D., Pare, G., Bailey, S. D., Xie, C., Zhang, X., et al. (2013). Genetic Information and the Prediction of Incident Type 2 Diabetes in a High-Risk Multiethnic Population: The EpiDREAM genetic study. *Diabetes Care, 36*(9), 2836-2842.

Appel, L. J., Clark, J. M., Yeh, H. C., Wang, N. Y., Coughlin, J. W., Daumit, G., et al. (2011). Comparative effectiveness of weight-loss interventions in clinical practice. *N Engl J Med, 365*(21), 1959-1968.

Ardissino, D., Berzuini, C., Merlini, P. A., Mannuccio Mannucci, P., Surti, A., Burtt, N., et al. (2011). Influence of 9p21.3 genetic variants on clinical and angiographic outcomes in early-onset myocardial infarction. *J Am Coll Cardiol, 58*(4), 426-434.

Boekholdt, S. M., Arsenault, B. J., Mora, S., Pedersen, T. R., LaRosa, J. C., Nestel, P. J., et al. (2012). Association of LDL cholesterol, non-HDL cholesterol, and apolipoprotein B levels with risk of cardiovascular events among patients treated with statins: a meta-analysis. *JAMA, 307*(12), 1302-1309.

Boffetta, P., & Straif, K. (2009). Use of smokeless tobacco and risk of myocardial infarction and stroke: systematic review with meta-analysis. *BMJ, 339*, b3060.

Brautbar, A., Pompeii, L. A., Dehghan, A., Ngwa, J. S., Nambi, V., Virani, S. S., et al. (2012). A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. *Atherosclerosis, 223*(2), 421-426.

Cholesterol Treatment Trialists Collaboration, C., Baigent, C., Blackwell, L., Emberson, J., Holland, L. E., Reith, C., et al. (2010). Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet, 376*(9753), 1670-1681.

Chow, C. K., Islam, S., Bautista, L., Rumboldt, Z., Yusufali, A., Xie, C., et al. (2011). Parental history and myocardial infarction risk across the world: the INTERHEART Study. *J Am Coll Cardiol, 57*(5), 619-627.

Chow, C. K., Teo, K. K., Rangarajan, S., Islam, S., Gupta, R., Avezum, A., et al. (2013). Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high-, middle-, and low-income countries. *JAMA, 310*(9), 959-968.

Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet, 11*(6), 415-425.

Cohen, J. C., Boerwinkle, E., Mosley, T. H., Jr., & Hobbs, H. H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med, 354*(12), 1264-1272.

Colditz, G. A., Stampfer, M. J., Willett, W. C., Rosner, B., Speizer, F. E., & Hennekens, C. H. (1986). A prospective study of parental history of myocardial infarction and coronary heart disease in women. *Am J Epidemiol, 123*(1), 48-58.

Conroy, R. M., Pyorala, K., Fitzgerald, A. P., Sans, S., Menotti, A., De Backer, G., et al. (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J, 24*(11), 987-1003.

Cook, D. G., Shaper, A. G., Pocock, S. J., & Kussick, S. J. (1986). Giving up smoking and the risk of heart attacks. A report from The British Regional Heart Study. *Lancet, 2*(8520), 1376-1380.

Cook, N. R., Paynter, N. P., Eaton, C. B., Manson, J. E., Martin, L. W., Robinson, J. G., et al. (2012). Comparison of the Framingham and Reynolds Risk scores for global cardiovascular risk prediction in the

multiethnic Women's Health Initiative. *Circulation, 125*(14), 1748-1756, S1741-1711.

Coronary Artery Disease Genetics Consortium. (2011). A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet, 43*(4), 339-344.

D'Agostino, R. B., Sr., Grundy, S., Sullivan, L. M., Wilson, P., & Group, C. H. D. R. P. (2001). Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA, 286*(2), 180-187.

Danaei, G., Finucane, M. M., Lin, J. K., Singh, G. M., Paciorek, C. J., Cowan, M. J., et al. (2011). National, regional, and global trends in systolic blood pressure since 1980: systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 5.4 million participants. *Lancet, 377*(9765), 568-577.

Danaei, G., Finucane, M. M., Lu, Y., Singh, G. M., Cowan, M. J., Paciorek, C. J., et al. (2011). National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet, 378*(9785), 31-40.

Davies, R. W., Dandona, S., Stewart, A. F., Chen, L., Ellis, S. G., Tang, W. H., et al. (2010). Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies. *Circ Cardiovasc Genet, 3*(5), 468-474.

Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T. L., et al. (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet, 45*(1), 25-33.

Ding, H., Xu, Y., Wang, X., Wang, Q., Zhang, L., Tu, Y., et al. (2009). 9p21 is a shared susceptibility locus strongly for coronary artery disease and weakly for ischemic stroke in Chinese Han population. *Circ Cardiovasc Genet, 2*(4), 338-346.

Dinsa, G. D., Goryakin, Y., Fumagalli, E., & Suhrcke, M. (2012). Obesity and socioeconomic status in developing countries: a systematic review. *Obes Rev, 13*(11), 1067-1079.

Do, R., Xie, C., Zhang, X., Mannisto, S., Harald, K., Islam, S., et al. (2011). The effect of chromosome 9p21 variants on cardiovascular disease may be modified by dietary intake: evidence from a case/control and a prospective study. *PLoS Med, 8*(10), e1001106.

Doll, R., & Peto, R. (1976). Mortality in relation to smoking: 20 years' observations on male British doctors. *Br Med J, 2*(6051), 1525-1536.

Dong, L., Wang, H., Wang, D. W., & Ding, H. (2013). Association of chromosome 9p21 genetic variants with risk of coronary heart disease in the East Asian population: a meta-analysis. *Ann Hum Genet, 77*(3), 183-190.

Egan, B. M., Zhao, Y., & Axon, R. N. (2010). US trends in prevalence, awareness, treatment, and control of hypertension, 1988-2008. *JAMA, 303*(20), 2043-2050.

Emerging Risk Factors Collaboration, C., Seshasai, S. R., Kaptoge, S., Thompson, A., Di Angelantonio, E., Gao, P., et al. (2011). Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med, 364*(9), 829-841.

Emerging Risk Factors Collaboration, C., Wormser, D., Kaptoge, S., Di Angelantonio, E., Wood, A. M., Pennells, L., et al. (2011). Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. *Lancet, 377*(9771), 1085-1095.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics, 164*(4), 1567-1587.

Farzadfar, F., Finucane, M. M., Danaei, G., Pelizzari, P. M., Cowan, M. J., Paciorek, C. J., et al. (2011). National, regional, and global trends in serum total cholesterol since 1980: systematic analysis of health examination surveys and epidemiological studies with 321 country-years and 3.0 million participants. *Lancet, 377*(9765), 578-586.

Finucane, M. M., Stevens, G. A., Cowan, M. J., Danaei, G., Lin, J. K., Paciorek, C. J., et al. (2011). National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *Lancet, 377*(9765), 557-567.

Ganna, A., Magnusson, P. K., Pedersen, N. L., de Faire, U., Reilly, M., Arnlov, J., et al. (2013). Multilocus Genetic Risk Scores for Coronary Heart Disease Prediction. *Arterioscler Thromb Vasc Biol*.

Gaziano, T. A. (2005). Cardiovascular disease in the developing world and its cost-effective management. *Circulation, 112*(23), 3547-3553.

Genest, J., McPherson, R., Frohlich, J., Anderson, T., Campbell, N., Carpentier, A., et al. (2009). 2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of

dyslipidemia and prevention of cardiovascular disease in the adult - 2009 recommendations. *Can J Cardiol, 25*(10), 567-579.

Giovino, G. A., Mirza, S. A., Samet, J. M., Gupta, P. C., Jarvis, M. J., Bhala, N., et al. (2012). Tobacco use in 3 billion individuals from 16 countries: an analysis of nationally representative cross-sectional household surveys. *Lancet, 380*(9842), 668-679.

Haffner, S. M., Lehto, S., Ronnemaa, T., Pyorala, K., & Laakso, M. (1998). Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *N Engl J Med, 339*(4), 229-234.

Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., Haskell, W., Ekelund, U., et al. (2012). Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet, 380*(9838), 247-257.

He, F. J., Nowson, C. A., Lucas, M., & MacGregor, G. A. (2007). Increased consumption of fruit and vegetables is related to a reduced risk of coronary heart disease: meta-analysis of cohort studies. *J Hum Hypertens, 21*(9), 717-728.

Held, C., Iqbal, R., Lear, S. A., Rosengren, A., Islam, S., Mathew, J., et al. (2012). Physical activity levels, ownership of goods promoting

sedentary behaviour and risk of myocardial infarction: results of the INTERHEART study. *Eur Heart J, 33*(4), 452-466.

Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., et al. (2007). A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science, 316*(5830), 1491-1493.

Higgins JPT, Green S (Eds.). Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0. (2008). Available from http://www.cochrane-handbook.org

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet, 6*(2), 95-108.

Hirunsatit, R., Ilomaki, R., Malison, R., Rasanen, P., Ilomaki, E., Kranzler, H. R., et al. (2007). Sequence variation and linkage disequilibrium in the GABA transporter-1 gene (SLC6A1) in five populations: implications for pharmacogenetic research. *BMC Genet, 8*, 71.

Hu, F. B., Manson, J. E., Stampfer, M. J., Colditz, G., Liu, S., Solomon, C. G., et al. (2001). Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med, 345*(11), 790-797.

Huang, C. C., Fornage, M., Lloyd-Jones, D. M., Wei, G. S., Boerwinkle, E., & Liu, K. (2009). Longitudinal association of PCSK9 sequence variations with low-density lipoprotein cholesterol levels: the Coronary Artery Risk Development in Young Adults Study. *Circ Cardiovasc Genet, 2*(4), 354-361.

Huxley, R., Barzi, F., & Woodward, M. (2006). Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies. *BMJ, 332*(7533), 73-78.

Huxley, R. R., & Woodward, M. (2011). Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. *Lancet, 378*(9799), 1297-1305.

Iqbal, R., Anand, S., Ounpuu, S., Islam, S., Zhang, X., Rangarajan, S., et al. (2008). Dietary patterns and the risk of acute myocardial infarction in 52 countries: results of the INTERHEART study. *Circulation, 118*(19), 1929-1937.

Kanaya, A. M., Grady, D., & Barrett-Connor, E. (2002). Explaining the sex difference in coronary heart disease mortality among patients with type 2 diabetes mellitus: a meta-analysis. *Arch Intern Med, 162*(15), 1737-1745.

Kathiresan, S., & Myocardial Infarction Genetics, C. (2008). A PCSK9 missense variant associated with a reduced risk of early-onset myocardial infarction. *N Engl J Med, 358*(21), 2299-2300.

Kennedy, K. F. P., M.J. (2010). A SAS macro to compute added predictive ability of new markers predicting a dichotomous outcome. Retrieved July 23, 2013, 2013, from http://analytics.ncsu.edu/sesug/2010/SDA07.Kennedy.pdf.

Kerr, K. F., McClelland, R. L., Brown, E. R., & Lumley, T. (2011). Evaluating the incremental value of new biomarkers with integrated discrimination improvement. *Am J Epidemiol, 174*(3), 364-374.

Kutner, M. N., CJ; Neter, J; . (2004). *Applied Linear Regression Models, 4th Edition* (Fourth ed.): MgGraw-Hill Irwin.

Lewington, S., Clarke, R., Qizilbash, N., Peto, R., Collins, R., & Prospective Studies, C. (2002). Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet, 360*(9349), 1903-1913.

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., et al. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the

Global Burden of Disease Study 2010. *Lancet, 380*(9859), 2224-2260.

Lipoeto, N. I., Wattanapenpaiboon, N., Malik, A., & Wahlqvist, M. L. (2004). Nutrition transition in west Sumatra, Indonesia. *Asia Pac J Clin Nutr, 13*(3), 312-316.

Liu, J., Hong, Y., D'Agostino, R. B., Sr., Wu, Z., Wang, W., Sun, J., et al. (2004). Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study. *JAMA, 291*(21), 2591-2599.

Lloyd-Jones, D. M., Leip, E. P., Larson, M. G., D'Agostino, R. B., Beiser, A., Wilson, P. W., et al. (2006). Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. *Circulation, 113*(6), 791-798.

Lloyd-Jones, D. M., Nam, B. H., D'Agostino, R. B., Sr., Levy, D., Murabito, J. M., Wang, T. J., et al. (2004). Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA, 291*(18), 2204-2211.

Lohmueller, K. E., Wong, L. J., Mauney, M. M., Jiang, L., Felder, R. A., Jose, P. A., et al. (2006). Patterns of genetic variation in the

hypertension candidate gene GRK4: ethnic variation and haplotype structure. *Ann Hum Genet, 70*(Pt 1), 27-41.

Look, A. R. G., Wing, R. R., Bolin, P., Brancati, F. L., Bray, G. A., Clark, J. M., et al. (2013). Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *N Engl J Med, 369*(2), 145-154.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., et al. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet, 380*(9859), 2095-2128.

Manta, F. S., Pereira, R., Caiafa, A., Silva, D. A., Gusmao, L., & Carvalho, E. F. (2013). Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. *Ann Hum Biol, 40*(1), 94-98.

Marenberg, M. E., Risch, N., Berkman, L. F., Floderus, B., & de Faire, U. (1994). Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med, 330*(15), 1041-1046.

Marrugat, J., D'Agostino, R., Sullivan, L., Elosua, R., Wilson, P., Ordovas, J., et al. (2003). An adaptation of the Framingham coronary heart disease risk function to European Mediterranean areas. *J Epidemiol Community Health, 57*(8), 634-638.

McPherson, R. (2010). Chromosome 9p21 and coronary artery disease. *N Engl J Med, 362*(18), 1736-1737.

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

Meigs, J. B., Shrader, P., Sullivan, L. M., McAteer, J. B., Fox, C. S., Dupuis, J., et al. (2008). Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med, 359*(21), 2208-2219.

Menotti, A., Puddu, P. E., & Lanti, M. (2000). Comparison of the Framingham risk function-based coronary chart with risk function from an Italian population study. *Eur Heart J, 21*(5), 365-370.

Misra, A., Singhal, N., Sivakumar, B., Bhagat, N., Jaiswal, A., & Khurana, L. (2011). Nutrition transition in India: secular trends in dietary intake and their relationship to diet-related non-communicable diseases. *J Diabetes, 3*(4), 278-292.

Moonesinghe, R., Yesupriya, A., Chang, M. H., Dowling, N. F., Khoury, M. J., Scott, A. J., et al. (2010). A Hardy-Weinberg equilibrium test for analyzing population genetic surveys with complex sample designs. *Am J Epidemiol, 171*(8), 932-941.

Mozaffarian, D., Micha, R., & Wallace, S. (2010). Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: a systematic review and meta-analysis of randomized controlled trials. *PLoS Med, 7*(3), e1000252.

Murray, C. J., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., et al. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet, 380*(9859), 2197-2223.

Nakajima, T., Jorde, L. B., Ishigami, T., Umemura, S., Emi, M., Lalouel, J. M., et al. (2002). Nucleotide diversity and haplotype structure of the human angiotensinogen gene in two populations. *Am J Hum Genet, 70*(1), 108-123.

Natarajan, S., Lipsitz, S. R., & Rimm, E. (2007). A simple method of determining confidence intervals for population attributable risk from complex surveys. *Stat Med, 26*(17), 3229-3239.

Neal, B., MacMahon, S., Chapman, N., & Blood Pressure Lowering Treatment Trialists, C. (2000). Effects of ACE inhibitors, calcium antagonists, and other blood-pressure-lowering drugs: results of prospectively designed overviews of randomised trials. Blood Pressure Lowering Treatment Trialists' Collaboration. *Lancet, 356*(9246), 1955-1964.

O'Neil, C. E., Keast, D. R., Fulgoni, V. L., & Nicklas, T. A. (2012). Food sources of energy and nutrients among adults in the US: NHANES 2003-2006. *Nutrients, 4*(12), 2097-2120.

Oh, K., Hu, F. B., Manson, J. E., Stampfer, M. J., & Willett, W. C. (2005). Dietary fat intake and risk of coronary heart disease in women: 20 years of follow-up of the nurses' health study. *Am J Epidemiol, 161*(7), 672-679.

Pan, A., Sun, Q., Bernstein, A. M., Schulze, M. B., Manson, J. E., Stampfer, M. J., et al. (2012). Red meat consumption and mortality: results from 2 prospective cohort studies. *Arch Intern Med, 172*(7), 555-563.

Pencina, M. J., D'Agostino, R. B., Sr., D'Agostino, R. B., Jr., & Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med, 27*(2), 157-172; discussion 207-112.

Pencina, M. J., D'Agostino, R. B., Sr., & Steyerberg, E. W. (2011). Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med, 30*(1), 11-21.

Prescott, E., Hippe, M., Schnohr, P., Hein, H. O., & Vestbo, J. (1998). Smoking and risk of myocardial infarction in women and men: longitudinal population study. *BMJ, 316*(7137), 1043-1047.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Prospective Studies, C., Lewington, S., Whitlock, G., Clarke, R., Sherliker, P., Emberson, J., et al. (2007). Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet, 370*(9602), 1829-1839.

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

Ridker, P. M., Paynter, N. P., Rifai, N., Gaziano, J. M., & Cook, N. R. (2008). C-reactive protein and parental history improve global cardiovascular risk prediction: the Reynolds Risk Score for men. *Circulation, 118*(22), 2243-2251, 2244p following 2251.

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

Ronksley, P. E., Brien, S. E., Turner, B. J., Mukamal, K. J., & Ghali, W. A. (2011). Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *BMJ, 342*, d671.

Rosengren, A., Hawken, S., Ounpuu, S., Sliwa, K., Zubaid, M., Almahmeed, W. A., et al. (2004). Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. *Lancet, 364*(9438), 953-962.

Royston, P., & Altman, D. G. (2010). Visualizing and assessing discrimination in the logistic regression model. *Stat Med, 29*(24), 2508-2520.

Sattelmair, J., Pertman, J., Ding, E. L., Kohl, H. W., 3rd, Haskell, W., & Lee, I. M. (2011). Dose response between physical activity and risk of coronary heart disease: a meta-analysis. *Circulation, 124*(7), 789-795.

Schauer, P. R., Kashyap, S. R., Wolski, K., Brethauer, S. A., Kirwan, J. P., Pothier, C. E., et al. (2012). Bariatric surgery versus intensive medical therapy in obese patients with diabetes. *N Engl J Med, 366*(17), 1567-1576.

Schunkert, H., Konig, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet, 43*(4), 333-338.

Sesso, H. D., Lee, I. M., Gaziano, J. M., Rexrode, K. M., Glynn, R. J., & Buring, J. E. (2001). Maternal and paternal history of myocardial infarction and risk of cardiovascular disease in men and women. *Circulation, 104*(4), 393-398.

Shield, K. D., Rylett, M., Gmel, G., Gmel, G., Kehoe-Chan, T. A., & Rehm, J. (2013). Global alcohol exposure estimates by country, territory and region for 2005--a contribution to the Comparative Risk Assessment for the 2010 Global Burden of Disease Study. *Addiction, 108*(5), 912-922.

Shield, M. W., K. (2013). Smoking, smoking cessation and heart disease risk: A 16 year follow-up study. *Health Rep, 24*(2), 12-22.

Sjostrom, L., Lindroos, A. K., Peltonen, M., Torgerson, J., Bouchard, C., Carlsson, B., et al. (2004). Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery. *N Engl J Med, 351*(26), 2683-2693.

Slatkin, M. (2008). Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet, 9*(6), 477-485.

Teo, K. K., Ounpuu, S., Hawken, S., Pandey, M. R., Valentin, V., Hunt, D., et al. (2006). Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study. *Lancet, 368*(9536), 647-658.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature, 466*(7307), 707-713.

Thanassoulis, G., Peloso, G. M., Pencina, M. J., Hoffmann, U., Fox, C. S., Cupples, L. A., et al. (2012). A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham Heart Study. *Circ Cardiovasc Genet, 5*(1), 113-121.

Tian, C., Gregersen, P. K., & Seldin, M. F. (2008). Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet, 17*(R2), R143-150.

Tverdal, A., Thelle, D., Stensvold, I., Leren, P., & Bjartveit, K. (1993). Mortality in relation to smoking history: 13 years' follow-up of 68,000 Norwegian men and women 35-49 years. *J Clin Epidemiol, 46*(5), 475-487.

Vaarhorst, A. A., Lu, Y., Heijmans, B. T., Dolle, M. E., Bohringer, S., Putter, H., et al. (2012). Literature-based genetic risk scores for coronary heart disease: the Cardiovascular Registry Maastricht (CAREMA) prospective cohort study. *Circ Cardiovasc Genet, 5*(2), 202-209.

Van der Kooy, K., van Hout, H., Marwijk, H., Marten, H., Stehouwer, C., & Beekman, A. (2007). Depression and the risk for cardiovascular diseases: systematic review and meta analysis. *Int J Geriatr Psychiatry, 22*(7), 613-626.

Vandevijvere, S., Monteiro, C., Krebs-Smith, S. M., Lee, A., Swinburn, B., Kelly, B., et al. (2013). Monitoring and benchmarking population diet quality globally: a step-wise approach. *Obes Rev, 14 Suppl 1*, 135-149.

Visel, A., Zhu, Y., May, D., Afzal, V., Gong, E., Attanasio, C., et al. (2010). Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature, 464*(7287), 409-412.

Wilmot, E. G., Edwardson, C. L., Achana, F. A., Davies, M. J., Gorely, T., Gray, L. J., et al. (2012). Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. *Diabetologia, 55*(11), 2895-2905.

Yusuf, S., Hawken, S., Ounpuu, S., Bautista, L., Franzosi, M. G., Commerford, P., et al. (2005). Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet, 366*(9497), 1640-1649.

Yusuf, S., Hawken, S., Ounpuu, S., Dans, T., Avezum, A., Lanas, F., et al. (2004). Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet, 364*(9438), 937-952.

Zdravkovic, S., Wienke, A., Pedersen, N. L., Marenberg, M. E., Yashin, A. I., & De Faire, U. (2002). Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med, 252*(3), 247-254.

Zhai, F., Wang, H., Du, S., He, Y., Wang, Z., Ge, K., et al. (2009). Prospective study on nutrition transition in China. *Nutr Rev, 67 Suppl 1*, S56-61.

Zhang, L., Yuan, F., Liu, P., Fei, L., Huang, Y., Xu, L., et al. (2013). Association between PCSK9 and LDLR gene polymorphisms with coronary heart disease: case-control study and meta-analysis. *Clin Biochem, 46*(9), 727-732.

Appendix 1: Tables and Figures

Table 1: Single Nucleotide Polymorphisms Associated with Coronary Artery Disease in C4D Meta-Analysis and CARDIOGRAM Study

SNP	Loci	OR	RAF	P-value	Possible Function(s) of Gene Related to CVD
rs4977574-G	*9p21 – CDKN2A/B ANRIL*	1.20	0.46	1.62x 10-25	CDKN2A/B encodes protein that is involved in regulation of the cell cycle ANRIL is a non-coding region but may influence expression of nearby genes
rs646776-T	*CELSR2P SRC1SOR T1*	1.14	0.78*	6.05x 10-10	Encoded protein mediates LDL metabolism and polymorphisms are associated with plasma LDL concentration
rs1412444-T	*LIPA*	1.09	0.34-0.51	2.76x 10-19	Encodes lipase A, which mediates the hydrolysis of cholesteryl esters and triglycerides
rs4380028-C	*ADAMTS7 MORF 4*	1.07	0.6-0.7	2.41x 10-9	Encodes metalloproteinase that accumulates preferentially in carotid artery neo-intima
rs974819-T	*PDGFD*	1.08	0.29-0.35	3.71x 10-9	Encodes PDGF-D, which is expressed in vascular tissue and may mediate atherosclerosis

rs2505083-C	*KIAA1462*	1.08	0.34-0.42	3.12x 10-8	The function of the protein product encoded by this gene is largely unknown
rs10953541-C	*7q22 (GPR22/COG5/DUS4L/BCAP29/SLC26A4 cluster)*	1.1	0.75-0.84	3.86x 10-8	Cluster of genes involved in encoding G-protein couple receptors expressed in coronary arteries
rs1746048-C	*CXCL12 10q11.21*	1.05	0.87	2.12x 10-10	Gene for stromal-cell–derived factor 1 precursor, association with CVD unknown
rs1122608-G	*LDLR 19p13.2*	1.08	0.77	9.73x 10-10	Encodes LDL receptor which is regulated LDL transport
rs3798220-C	*LPA 6q25.3*	1.54	0.02	9.62x 10-12	Gene for Lipoprotein(a), a lipoprotein that is associated with CVD development
rs17465637-C	*MIA3 1q41*	1.08	0.74	1.36x 10-8	Regulates protein involved in the export of proteins from the endoplasmic reticulum, although role in atherosclerosis remains unclear
rs9818870-C	*MRAS 3q22.3*	1.07	0.15	3.34x 10-8	M-Ras protein belongs to Ras superfamily of GTP-binding proteins. It is widely expressed in tissues and may be involved in TNF-alpha lymphocyte expression
rs9982601-T	*SLC5A3 MRPS6 KCNE2*	1.08	0.15	4.22x 10-10	Unclear relationship between this gene cluster and CAD risk

	21q22.11				
rs11206510-T	*PCSK9 1p32.3*	1.05	0.82	9.1x 10-8	Has been shown to influence LDL cholesterol and consequently may affect risk of myocardial infarction
rs12526453-C	*PHACTR1 6p24.1*	1.10	0.67	1.15x 10-9	Inhibitor of protein phosphatase 1, an enzyme that dephosphorylates serine and threonine residues on a range of proteins, unclear relationship to CVD development
rs6725887-C	*WDR12 2q33.1*	1.11	0.15	1.12x 10-9	Encoded protein complex interacts with several proteins to enable cell proliferation
rs579459-C	*ABO 9q34.2*	1.10	0.21	4.08x 10-14	Polymorphisms associated with variability in Von Willebrand Factor plasma concentration, which may influence thrombosis
rs17609940-G	*ANKSIA 6p21.31*	1.07	0.75	1.3x 10-8	Unclear role in CVD development
rs4773144-G	*COL4A1 COL4A2 13q34*	1.07	0.44	3.84x 10-9	Encoded proteins involved in collagen synthesis which may influence atherosclerosis development
rs12413409-G	*CYP17A1 CNNM2 NT5C2 10q24.32*	1.12	0.89	9.51x 10-10	Gene cluster associated with hypertension in the CARDIoGRAM study
rs2895811-C	*HHIPL1 14q32.2*	1.07	0.43	1.14x 10-10	Unclear relationship to CVD development

rs12936587-G	*RASD1 SMCR3 PEMT 17p11.2*	1.07	0.56	4.45x 10-10	Unclear relationship between gene cluster and CAD development
rs216172-C	*SMG6 SRR 17p13.3*	1.07	0.37	1.15x 10-9	Unclear relationship between gene cluster and CAD development
rs12190287-C	*TCF21 6q23.2*	1.08	0.62	1.07x 10-12	Shown to influence differentiation of developing coronary vessels
rs46522-T	*UBE2Z GIP ATP5G1 SNF8 17q21.32*	1.06	0.53	1.81x 10-8	Possible gene transcription effects, although relationship with CAD development is unclear
rs11556924-C	*ZC3HC1 7q32.2*	1.09	0.62	9.18x 10-18	Function variant affecting protein structure of R363H, unclear relation to CAD development
rs964184-G	*ZNF259 APOA5-A4-C3-A1 11q23.3*	1.13	0.13	1.02x 10-17	Associated with plasma lipids in the CARDIOGRAM study
rs3184504-T	*SH2B3*	1.07	0.44	6.35 E-6	Encoded protein is involved in the regulation of endothelial cell adherence and migration
rs17114036-A	*PPAP2B*	1.17	0.91	1.43x 10-8	Gene product degrades bioactive lysophospholipids and mediates smooth muscle cell modulation and inflammation following vascular injury

SNP = single nucleotide polymorphism, OR = odds ratio, RAF = risk

allele frequency

Table 2A: Baseline Characteristics of INTERHEART Participants with and without Myocardial Infarction

Baseline Characteristic	Total (8556)	Myocardial Infarction Cases (4083)	Controls (4473)	p-value
Mean Age (SD)	55.73 (12.1)	56.31 (12.03)	55.21 (12.09)	<0.0001
Male	6763 (79.04)	3232 (79.16)	3531 (78.94)	0.80
Mean Gene Score (SD)	26.27 (2.96)	26.59 (2.87)	25.88 (3.01)	<0.0001
Median Gene Score (IQR)	26 (24-28)	27 (25-29)	26 (24-28)	<0.0001
Hypertension	2509 (29.57)	1542 (38.35)	967 (21.67)	<0.0001
Diabetes	1200 (14.16)	807 (20.09)	393 (8.81)	<0.0001
ApoB/A1 Ratio				
<0.61	1257 (14.72)	365 (8.95)	892 (20.01)	<0.0001
0.61-0.73	1427 (16.72)	535 (13.12)	892 (20.01)	
0.73-0.86	1620 (18.98)	729 (17.87)	891 (19.99)	
0.86-1.02	1872 (21.93)	980 (24.03)	892 (20.01)	
>1.02	2361 (27.66)	1470 (36.04)	891 (19.99)	
Waist to Hip Ratio				
<0.90	2313 (28.04)	848 (22.00)	1465 (33.33)	<0.0001
0.90-0.95	2609 (31.63)	1142 (29.63)	1467 (33.38)	
>0.95	3327 (40.33)	1864 (48.37)	1463 (33.29)	
Smoking				
Never	3344 (40.24)	1256 (31.95)	2088 (47.68)	<0.0001
Former	1859 (22.37)	840 (21.37)	1019 (23.27)	
Current	3107 (37.39)	1835 (46.68)	1272 (29.05)	

Alcohol Consumption				
< 3x/wk	7361 (87.66)	3446 (87.15)	3915 (88.12)	0.1801
>/= 3x/wk	1036 (12.34)	508 (12.85)	528 (11.88)	
Fruit and Vegetable Intake				<0.0001
Neither daily	1490 (17.90)	772 (19.74)	718 (16.27)	
Fruits or veg daily	3617 (43.46)	1707 (43.66)	1910 (43.29)	
Fruits and veg daily	3215 (38.63)	1431 (36.60)	1784 (40.44)	
Physically Active				
Yes	1374 (16.30)	571 (14.40)	803 (17.98)	<0.0001
Stress at Home or Work				
None	1987 (24.08)	883 (22.69)	1104 (25.31)	<0.0001
Some periods	4017 (48.67)	1795 (46.13)	2222 (50.94)	
Several periods	1636 (19.82)	836 (21.49)	800 (18.34)	
Permanent stress	613 (7.43)	377 (9.69)	236 (5.41)	

Table 2B: Baseline Characteristics of INTERHEART Participants by Ethnic Group

Baseline Characteristic	Total (8556)	European (1886)	South Asian (1883)	Other Asian (1223)
Mean Age (SD)	55.73 (12.1)	59.8 (12.1)	54.7 (11.7)	55.7 (11.8)
Male	6763 (79.04)	1323 (70.15)	1568 (85.68)	1025 (83.81)
Mean Gene Score (SD)	26.27 (2.96)	26.26 (2.90)	26.58 (2.85)	26.42 (2.83)
Median Gene Score (IQR)	26 (24-28)	26 (24-28)	27 (25-28)	26 (25-28)
Hypertension	2509 (29.57)	711 (37.92)	461 (25.51)	345 (28.63)
Diabetes	1200 (14.16)	208 (11.10)	269 (14.89)	173 (14.37)
ApoB/A1 Ratio				
<0.61	1257 (14.72)	371 (19.68)	220 (12.08)	203 (16.63)
0.61-0.73	1427 (16.72)	368 (19.52)	298 (16.36)	218 (17.85)
0.73-0.86	1620 (18.98)	383 (20.32)	330 (18.12)	253 (20.72)
0.86-1.02	1872 (21.93)	370 (19.63)	428 (23.50)	249 (20.39)
>1.02	2361 (27.66)	393 (20.85)	545 (29.93)	298 (24.41)
Waist to Hip Ratio				
<0.90	2313 (28.04)	599 (33.00)	532 (30.09)	350 (29.36)
0.90-0.95	2609 (31.63)	455 (25.07)	599 (33.88)	365 (30.62)
>0.95	3327 (40.33)	761 (41.93)	637 (36.03)	477 (40.02)
Smoking				
Current	3344 (40.24)	760 (41.01)	660 (37.35)	480 (40.47)
Former	1859 (22.37)	438 (23.64)	399 (22.58)	270 (22.77)
Never	3107 (37.39)	655 (35.35)	708 (40.07)	436 (36.76)
Alcohol				

Consumption				
< 3 times/week	7361 (87.66)	1399 (75.66)	1622 (90.36)	985 (82.15)
>/= 3times/week	1036 (12.34)	450 (24.34)	173 (9.64)	214 (17.85)
Fruit and Vegetable Intake				
Neither daily	1490 (17.90)	373 (20.21)	264 (14.91)	241 (20.22)
Fruits or veg daily	3617 (43.46)	770 (41.71)	914 (51.61)	458 (38.42)
Fruits and veg daily	3215 (38.63)	703 (38.08)	593 (33.48)	493 (41.36)
Physical Activity				
No	7057 (83.70)	1338 (71.78)	1612 (89.81)	985 (81.74)
Yes	1374 (16.30)	526 (28.22)	183 (10.19)	220 (18.26)
Stress at Home or Work				
None	1987 (24.08)	438 (23.96)	394 (22.70)	326 (27.72)
Some periods	4017 (48.67)	936 (51.20)	861 (49.60)	529 (44.98)
Several periods	1636 (19.82)	316 (17.29)	375 (21.60)	239 (20.32)
Permanent stress	613 (7.43)	138 (7.55)	106 (6.11)	82 (6.97)

SD = standard deviation, IQR = interquartile range, veg = vegetables

Table 2B (Continued): Baseline Characteristics of INTERHEART Participants by Ethnic Group

Baseline Characteristic	Arab (1371)	Latino (1470)	African (776)	P-value
Mean Age (SD)	52.3 (10.5)	58.5 (12.4)	49.5 (10.3)	<0.0001
Male	1063 (77.53)	1082 (73.61)	702 (90.46)	<0.0001
Mean Gene Score (SD)	25.95 (3.10)	25.96 (3.16)	26.55 (2.81)	<0.0001
Median Gene Score (IQR)	26 (24-28)	26 (24-28)	27 (25-28)	<0.0001
Hypertension	310 (22.74)	545 (37.30)	137 (17.72)	<0.0001
Diabetes	201 (14.75)	217 (14.90)	132 (17.08)	0.0006
ApoB/A1 Ratio				
<0.61	199 (14.55)	206 (14.03)	58 (7.49)	<0.0001
0.61-0.73	201 (14.69)	238 (16.21)	104 (13.44)	
0.73-0.86	252 (18.42)	275 (18.73)	127 (16.41)	
0.86-1.02	300 (21.93)	334 (22.75)	191 (24.68)	
>1.02	416 (30.41)	415 (28.27)	294 (37.98)	
Waist to Hip Ratio				<0.0001
<0.90	320 (24.02)	349 (25.18)	163 (21.56)	
0.90-0.95	519 (38.96)	411 (29.65)	260 (34.39)	
>0.95	493 (37.01)	626 (45.17)	333 (44.05)	
Smoking				
Current	526 (39.40)	594 (41.36)	324 (44.20)	<0.0001
Former	220 (16.48)	432 (30.08)	100 (13.64)	
Never	589 (44.12)	410 (28.55)	309 (42.16)	
Alcohol Consumption				

<3 times/week	1311 (96.68)	1306 (91.14)	738 (96.47)	<0.0001
>/= 3 times/week	45 (3.32)	127 (8.86)	27 (3.53)	
Fruit and Vegetable Intake				
Neither daily	228 (17.00)	224 (15.81)	160 (21.19)	<0.0001
Fruits or veg daily	557 (41.54)	584 (41.21)	334 (44.24)	
Fruits and veg daily	556 (41.46)	609 (42.98)	261 (34.57)	
Physical Activity				
No	1262 (93.00)	1148 (79.61)	712 (92.71)	<0.0001
Yes	95 (7.00)	294 (20.39)	56 (7.29)	
Stress at Home or Work				
None	329 (24.87)	342 (23.95)	158 (20.73)	<0.0001
Some periods	653 (49.36)	650 (45.52)	388 (50.92)	
Several periods	251 (18.97)	278 (19.47)	177 (23.23)	
Permanent stress	90 (6.80)	158 (11.06)	39 (5.12)	

SD = standard deviation, IQR = interquartile range, veg = vegetables

Table 3: Allele Frequencies of Single Polymorphisms

Previously Associated with Coronary Artery Disease

Included in the Genetic Risk Score

LOCI	SNP	Literature Risk Allele	Total RAF
9p21	RS4977574	G	0.47-0.54
CELSR2-PSRC1-SORT1	RS646776	T	0.77-0.83
WDR12	RS6725887	C	0.03-0.12
SCL5A3-MRPS6-KCNE2	RS9982601	T	0.04-0.16
MRAS	RS2306374	C	0.07-0.14
LDLR	RS1122608	G	0.15-0.25
CXCL12	RS1746048	C	0.67-0.82
MIA3	RS17465637	C	0.54-0.70
PCSK9	RS11206510	T	0.82-0.95
LIPA	RS1412444	T	0.36-0.49
ADAMSTS7-MORFL1	RS4380028	C	0.59-0.61
7Q22	RS10953541	C	0.80-0.90
KIAA1462	RS2505083	C	0.33-0.45
LPA	RS3798220	C	0.006-0.15
PPAP2B	RS17114036	A	0.88-0.91
ANKS1A	RS17609940	G	0.04-0.16
ZC3HC1	RS11556924	C	0.67-0.85
ABO	RS579459	C	0.16-0.23
CYP17A1,CNNM	RS12413409	G	0.79-

2, NT5C2			0.91
ZNF259,APOA5-A4-C3-A1	RS964184	G	0.17-0.28
COL4A1,COL4A2	RS4773144	G	0.40-0.51
HHIPL1	RS2895811	C	0.33-0.43
RASD1,SMCR3, PEMT	RS12936587	G	0.62-0.79
SMG6,SRR	RS216172	C	0.33-0.39
UBE2Z,GIP,ATP5G1,SNF8	RS46522	T	0.38-0.57

SNP = single nucleotide polymorphism, RAF = risk allele frequency

Table 3 (Continued): Allele Frequencies of Single Polymorphisms Previously Associated with Coronary Artery Disease Included in the Genetic Risk Score

LOCI	European RAF	South Asian RAF	Other Asian RAF	Arab RAF	Latin RAF	African RAF
9p21	0.47	0.52	0.54	0.51	0.47	0.54
CELSR2-PSRC1-SORT1	0.77	0.85	0.83	0.75	0.76	0.77
WDR12	0.03	0.12	0.04	0.08	0.08	0.06
SCL5A3-MRPS6-KCNE2	0.16	0.05	0.04	0.11	0.11	0.08
MRAS	0.14	0.08	0.07	0.10	0.09	0.10
LDLR	0.21	0.19	0.18	0.19	0.15	0.25
CXCL12	0.82	0.74	0.69	0.67	0.71	0.68
MIA3	0.70	0.61	0.62	0.58	0.54	0.61
PCSK9	0.82	0.92	0.95	0.85	0.88	0.91
LIPA	0.36	0.41	0.47	0.43	0.44	0.51
ADAMSTS7-MORFL1	0.63	0.64	0.66	0.63	0.59	0.61
7Q22	0.80	0.83	0.85	0.90	0.85	0.85
KIAA1462	0.45	0.28	0.29	0.31	0.40	0.33
LPA	0.02	0.04	0.04	0.01	0.15	0.006
PPAP2B	0.91	0.94	0.94	0.88	0.91	0.92
ANKS1A	0.15	0.07	0.05	0.04	0.16	0.06
ZC3HC1	0.67	0.83	0.85	0.81	0.77	0.80
ABO	0.23	0.20	0.17	0.17	0.18	0.16
CYP17A1,CNNM2, NT5C2	0.91	0.79	0.81	0.91	0.87	0.83
ZNF259, APOA5-A4-C3-A1	0.17	0.23	0.24	0.22	0.28	0.23
COL4A1, COL4A2	0.44	0.43	0.40	0.51	0.45	0.45
HHIPL1	0.42	0.37	0.37	0.36	0.33	0.43

RASD1,SMCR3,PEMT	0.62	0.76	0.79	0.65	0.66	0.77
SMG6,SRR	0.35	0.33	0.35	0.38	0.33	0.39
UBE2Z,GIP,ATP5G1,SNF8	0.51	0.57	0.53	0.38	0.43	0.46

SNP = single nucleotide polymorphism, RAF = risk allele frequency

Table 4: Impact of Modifiable Risk Factors on Myocardial Infarction Risk in Multivariable Logistic Regression Analysis

Baseline Characteristic	Beta (SE)	Odds Ratio (95% CI)	p-value
Hypertension	0.668 (0.059)	1.95 (1.74-2.19)	<0.0001
Diabetes	0.806 (0.077)	2.24 (1.93-2.60)	<0.0001
ApoB/A1 Ratio			
<0.61	Reference		
0.61-0.73	0.360 (0.095)	1.43 (1.19-1.73)	0.0001
0.73-0.86	0.600 (0.091)	1.82 (1.52-2.17)	<0.0001
0.86-1.02	0.948 (0.088)	2.41 (2.01-2.90)	<0.0001
>1.02	1.253 (0.088)	3.50 (2.95-4.16)	<0.0001
Waist to Hip Ratio			
<0.90			
0.90-0.95	Reference		
>0.95	0.210 (0.070)	1.23 (1.08-1.41)	0.0024
	0.618 (0.067)	1.86 (1.63-2.12)	<0.0001
Smoking			
Never	Reference		
Former	0.259 (0.070)	1.29 (1.13-1.49)	0.0002
Current	1.04 (0.063)	2.83 (2.50-3.20)	<0.0001
Alcohol Consumption			
< 3 times/week	Reference		
>/= 3 times/week	-0.039 (0.081)	0.96 (0.82-1.13)	0.630
Fruit and Vegetable Intake			
Neither daily	Reference		
Fruits or veg daily	-0.1867 (0.072)	0.83 (0.72-0.95)	0.009
Fruits and veg daily	-0.2399 (0.074)	0.79 (0.68-0.91)	0.0011
Physically Active	-0.141 (0.072)	0.86 (0.75-1.00)	0.049
Stress at Home or Work			
None	Reference		
Some periods	-0.031 (0.064)	0.97 (0.86-1.10)	0.629
Several periods	0.256 (0.078)	1.29 (1.13-1.49)	0.001
Permanent stress	0.694 (0.111)	2.00 (1.61-2.49)	<0.0001

Model adjusted for age, sex, ethnicity, and modifiable risk factors

Table 5: Univariate and Multivariable Analysis of Associations Between Genetic Risk Score and Myocardial Infarction in Individual Ethnic Groups

Ethnicity	**Odds Ratio (95% CI)**	**p-value**	**Odds Ratio (95% CI)**	**p-value**
European	1.10 (1.06-1.13)	<0.0001	1.08 (1.04-1.12)	<0.0001
South Asian	1.10 (1.06-1.13)	<0.0001	1.09 (1.05-1.14)	<0.0001
Other Asian	1.08 (1.04-1.13)	0.0003	1.09 (1.04-1.15)	0.001
Arab	1.06 (1.03-1.10)	0.002	1.07 (1.03-1.12)	0.002
Latin American	1.03 (1.00-1.07)	0.06	0.99 (0.95-1.03)	0.75
African	1.03 (0.98-1.09)	0.25	1.04 (0.98-1.10)	0.24

Table 6a: Subgroup Analysis of Genetic Risk Score

Category	Number of participants	Effect (95% confidence interval)	P-value for interaction
Age			
< 45	1352	1.03 (0.99,1.08)	0.22
45-51	1587	1.10 (1.05,1.14)	
53-58	1555	1.08 (1.04,1.12)	
59-67	1440	1.06 (1.02,1.10)	
>67	1443	1.05 (1.01,1.09)	
Sex			
Male	6763	1.05 (1.03,1.07)	0.08
Female	1793	1.09 (1.05, 1.14)	

Models adjusted for age, and sex when not stratified by these variables, in addition to ethnicity, modifiable risk factors and genetic risk score

Table 6b: Subgroup Analysis of Genetic Risk Score

Excluding Latin group

Category	Number of participants	Effect (95% confidence interval)	P-value for interaction
Age*			
< 45	1169	1.03 (0.99,1.08)	0.13
45-51	1392	1.10 (1.06,1.15)	
53-58	1271	1.10 (1.06,1.15)	
59-67	1179	1.09 (1.04-1.14)	
>67	1117	1.07 (1.02-1.11)	
Sex			
Male	6128	1.07 (1.05, 1.09)	0.18
Female	1194	1.10 (1.05, 1.15)	

Models adjusted for age, and sex when not stratified by these variables, in addition to ethnicity, modifiable risk factors and genetic risk score

Table 7a: Population Attributable Risk Estimates for Genetic Risk Score and Modifiable Risk Factors

	Population Attributable Risk
Genetic Risk Score > 23	0.32 (0.25-0.39)
Modifiable Risk Factors	0.90 (0.87-0.92)
Modifiable Risk Factors + Genetic Risk Score > 23	0.92 (0.90-0.94)

Table 7b: Population Attributable Risk Estimates for Genetic Risk Score and Modifiable Risk Factors Excluding Latin American Subgroup

	Population Attributable Risk
Genetic Risk Score >23	0.32 (0.25-0.39)
Modifiable Risk Factors	0.90 (0.87-0.92)
Modifiable Risk Factors + Genetic Risk Score >23	0.92 (0.90-0.94)

Table 8a: Concordance Statistic for Modifiable Risk Factors and Genetic Risk Score

	C-Statistic (95% CI)	p-value for difference
Model 1:		
Modifiable Risk Factors	0.731 (0.719,0.742)	
Modifiable Risk Factors + Genetic Risk Score	0.735 (0.723,0.746)	0.001
Model 2:		
Genotype Score	0.573 (0.560,0.586)	
Modifiable Risk Factors + Genetic Risk Score	0.735 (0.723,0.746)	<0.0001

Table 8b: Concordance Statistic for Modifiable Risk Factors and Genetic Risk Score Excluding Latin group

	C-Statistic (95% CI)	p-value for difference
Model 1:		
Modifiable Risk Factors	0.729 (0.717,0.742)	
Modifiable Risk Factors + Genetic Risk Score	0.735 (0.722,0.747)	0.0004
Model 2:		
Genotype Score	0.583 (0.569,0.597)	
Modifiable Risk Factors + Genetic Risk Score	0.735 (0.722,0.747)	<0.0001

Figure 1: Exclusion of Participants and Single Nucleotide Polymorphisms

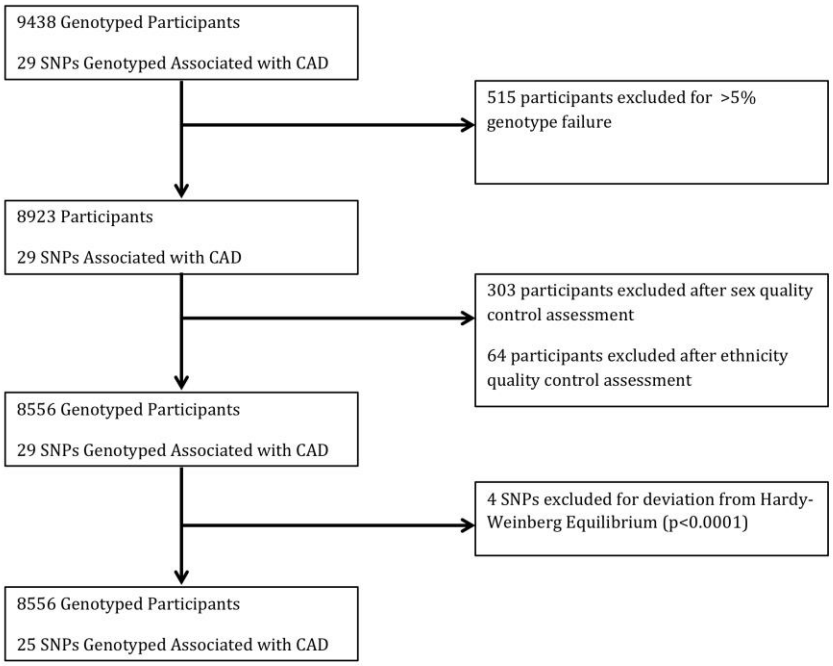


Figure 2: Meta-Analysis of Univariate Association Between Genetic Risk Score and Myocardial Infarction

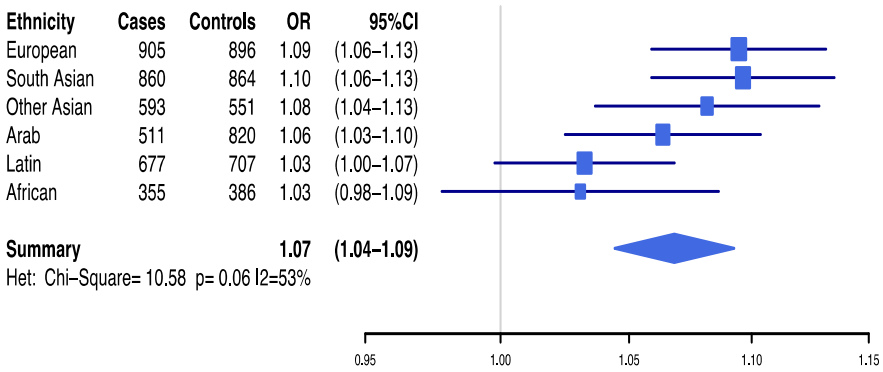


Figure 3: Meta-Analysis of Full Multi-variable Association Between Genetic Risk Score and Myocardial Infarction

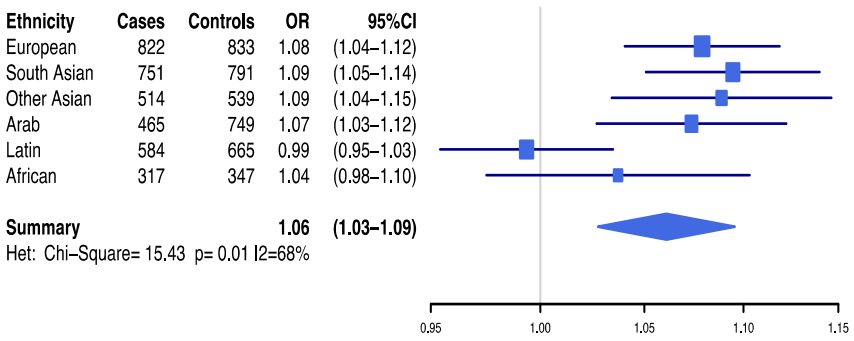


Figure 4: Meta-Analysis of Full Multi-variable Association

Between Genetic Risk Score and Myocardial Infarction

After Exclusion of Latin American Group

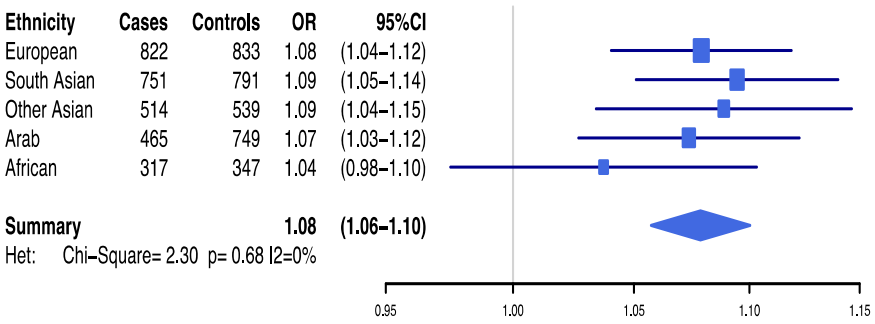
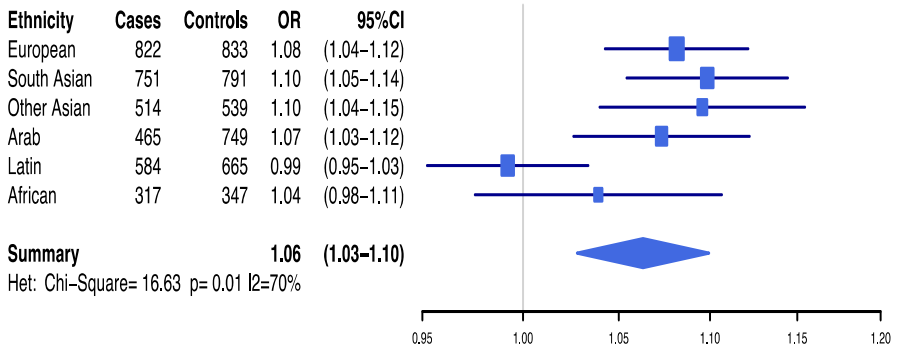


Figure 5: Meta-Analysis of Multi-variable (with PCA) Association Between Genetic Risk Score and Myocardial Infarction



Supplementary Appendix 1: Supplementary Tables and Figures

Supplementary Table 1: Comparison of Effects of Each Candidate Single Nucleotide Polymorphisms in Europeans in INTERHEART compared to Previous Literature

LOCI	OR 95% CI from GWAS MA in Europeans*	Effect in Europeans in IH Study	Final Estimate
9p21	1.27 (1.19, 1.35)	1.30 (1.14, 1.48)	1.28 (1.20, 1.35)
CELSR2-PSRC1-SORT1	1.32 (1.22, 1.42)	1.11 (1.02,1.21)	1.29 (1.20, 1.38)
WDR12	1.14 (1.09, 1.19)	1.25 (1.03, 1.52)	1.15 (1.10, 1.20)
SCL5A3-MRPS6-KCNE2	1.18 (1.12, 1.24)	1.29 (1.08, 1.54)	1.19 (1.13, 1.25)
MRAS	1.14 (1.09, 1.20)	1.05 (0.88, 1.26)	1.14 (1.09, 1.19)
*LDLR***	1.14 (1.09, 1.18)	0.97 (0.90, 1.04)	1.13 (1.08, 1.18)
CXCL12	1.09 (1.07, 1.13)	1.22 (1.03, 1.43)	1.09 (1.07, 1.11)
MIA3	1.14 (1.09, 1.20)	1.14 (0.99, 1.31)	1.14 (1.09, 1.19)
*PCSK9***	1.08 (1.05, 1.11)	0.95 (0.80, 1.13)	1.08 (1.05, 1.11)
LPA	1.54 (1.36, 1.74)	1.93 (1.19, 3.14)	1.56 (1.38, 1.76)
PPAP2B	1.17 (1.13, 1.22)	1.24 (0.98, 1.57)	1.17 (1.13, 1.21)
ANKS1A	1.07 (1.05, 1.10)	0.97 (0.82, 1.16)	1.07 (1.05, 1.09)
ZC3HC1	1.09 (1.07, 1.12)	1.04 (0.91, 1.19)	1.09 (1.07,1.11)
ABO	1.10 (1.07, 1.13)	1.18 (1.01, 1.39)	1.10 (1.07, 1.13)
CYP17A1,CNNM2,NT5C2	1.12 (1.08, 1.16)	1.22 (0.98, 1.53)	1.12 (1.08, 1.16)
ZNF259,APOA5-A4-C3-A1	1.13 (1.10, 1.16)	1.11 (0.94, 1.32)	1.12 (1.09, 1.15)
*COL4A1,COL4A2***	1.07 (1.05, 1.09)	0.87 (0.77, 0.99)	1.07 (1.05, 1.09)

HHIPL1	1.07 (1.05, 1.10)	1.07 (0.94, 1.22)	1.07 (1.05, 1.09)
RASD1,SMCR3, PEMT	1.07 (1.05, 1.09)	1.10 (0.97, 1.25)	1.07 (1.05, 1.09)
SMG6,SRR	1.07 (1.05, 1.09)	1.00 (0.85-1.14)	1.07 (1.05, 1.09)
*UBE2Z,GIP,ATP 5G1,SNF8***	1.06 (1.04, 1.08)	1.31 (1.16,1.49)	1.07 (1.04, 1.09)

*Effects observed in European Cohorts in CARDIOGRAM and

CD4 Meta-Analyses

** Genes associated with significant heterogeneity

(I^2 >50% or p-value for heterogeneity <0.10)

No estimates were available for ADAMST7, KIA1462, and

7q22 from GWAS Meta-Analyses

Supplementary Table 1 (Continued): Comparison of Effects of Each Candidate Single Nucleotide Polymorphisms in Europeans in INTERHEART compared to Previous Literature

LOCI	I^2 for Heterogeneity	P for Heterogeneity
9p21	0%	0.74
CELSR2-PSRC1-SORT1	48%	0.10
WDR12	0%	0.36
SCL5A3-MRPS6-KCNE2	0%	0.35
MRAS	0%	0.68
*LDLR***	68%	0.08
CXCL12	42%	0.19
MIA3	0%	0.99
*PCSK9***	51%	0.15
LPA	0%	0.37
PPAP2B	0%	0.63
ANKS1A	6%	0.30
ZC3HC1	0%	0.51
ABO	0%	0.36
CYP17A1,CNNM2,NT5C2	0%	0.44
ZNF259,APOA5-A4-C3-A1	0%	0.86
*COL4A1,COL4A2***	89%	0.002
HHIPL1	0%	0.97
RASD1,SMCR3, PEMT	0%	0.67
SMG6,SRR	1%	0.31
*UBE2Z,GIP,ATP5G1,SNF8***	91%	<0.001

*Effects observed in European Cohorts in CARDIOGRAM and CD4 Meta-Analyses

** Genes associated with significant heterogeneity (I^2 >50% or p-value for heterogeneity <0.10)

No estimates were available for ADAMST7, KIA1462, and

7q22 from GWAS Meta-Analyses

Supplementary Table 2: Power Calculation

for Genetic Risk Score

Odds Ratio	Power	N
1.02	0.6	2778
	0.7	3500
	0.8	4451
	0.9	5958
1.04	0.6	710
	0.7	894
	0.8	1137
	0.9	1522
1.06	0.6	322
	0.7	406
	0.8	517
	0.9	692
1.08	0.6	186
	0.7	234
	0.8	297
	0.9	398
1.10	0.6	122
	0.7	153
	0.8	195
	0.9	261

Standard deviation of 3, two-sided alpha 0.05, prevalence of CAD = 0.07

Supplementary Table 3: Sample Size Estimates for Individual SNPs based on Effect Sizes and Frequencies Derived in Literature in White Caucasians

SNP	Associated Loci	OR	Risk Allele Frequency
rs4977574-G	9p21	1.20	0.46
rs646776-T	*CELSR2/PSRC1 /SORT1*	1.14	0.78*
rs1412444-T	*LIPA*	1.09	0.34
rs4380028-C	*ADAMTS7*	1.07	0.6
rs974819-T	*PDGFD*	1.08	0.29
rs2505083-C	*KIAA1462*	1.08	0.42
rs10953541-C	*7q22*	1.1	0.75
rs1746048-C	CXCL12	1.05	0.87
rs1122608-G	LDLR	1.08	0.77
rs3798220-C	LPA	1.54	0.02
rs17465637-C	MIA3	1.08	0.74
rs9818870-C	MRAS	1.07	0.15
rs9982601-T	MRPS6/SLC5A3 /KCNE2	1.08	0.15
rs11206510-T	PCSK9	1.05	0.82
rs12526453-C	PHACTR1	1.10	0.67
rs6725887-C	WDR12	1.11	0.15
rs579459-C	ABO	1.10	0.21
rs17609940-G	ANKSIA	1.07	0.75
rs4773144-G	COL4A1/4A2	1.07	0.44
rs12413409-G	CYP17A1 CNNM2	1.12	0.89
rs2895811-C	HHIPL1	1.07	0.43
rs12936587-G	RASD1/SMCR3/ PEMT	1.07	0.56
rs216172-C	SMG6/SRR	1.07	0.37
rs12190287-C	TCF21	1.08	0.62
rs46522-T	UBE2Z/GIP/ATP 5G1/SNF8	1.06	0.53
rs11556924-C	ZC3HC1	1.09	0.62
rs964184-G	ZNF259/APOA5-A4-C3-A1	1.13	0.13
rs3184504-T	SH2B3	1.07	0.44
rs17114036-A	PPAP2B	1.17	0.91

Estimated using a two-sided alpha = 0.05

*Bolded numbers represent values that are above the

sample size of the study (N=8556)

Supplementary Table 3 (Continued) : Sample Size Estimates for Individual SNPs based on Effect Sizes and Frequencies Derived in Literature in White Caucasians

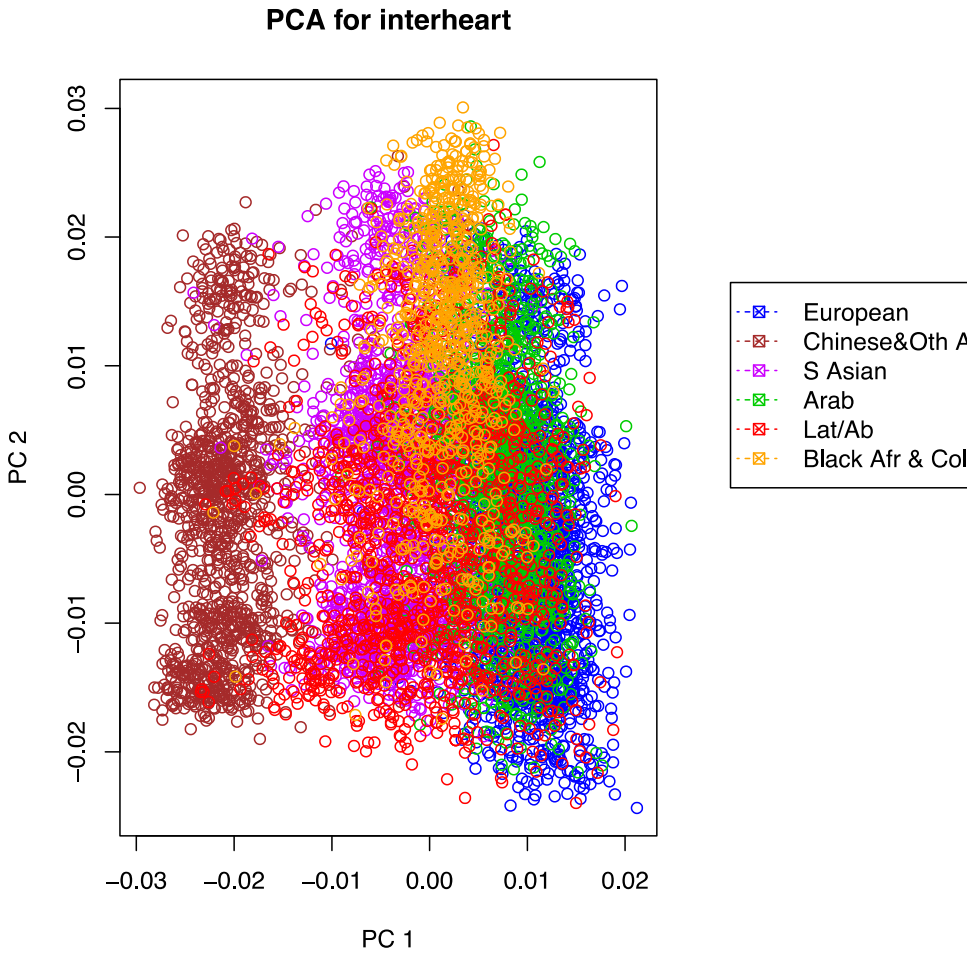
SNP	80% Power Sample Size*	90% Power Sample Size*
rs4977574-G	949	1270
rs646776-T	2751	3683
rs1412444-T	4659	6237
rs4380028-C	7189	**9624**
rs974819-T	6352	8505
rs2505083-C	5415	7250
rs10953541-C	4706	6300
rs1746048-C	**29608**	**39637**
rs1122608-G	7619	**10200**
rs3798220-C	1814	2428
rs17465637-C	7000	**9371**
rs9818870-C	**13182**	**17648**
rs9982601-T	**10161**	**13603**
rs11206510-T	**22641**	**30311**
rs12526453-C	3966	5309
rs6725887-C	5483	7341
rs579459-C	5091	6815
rs17609940-G	**9281**	**12424**
rs4773144-G	6938	**9289**
rs12413409-G	6482	**8677**
rs2895811-C	6971	**9333**
rs12936587-G	6987	**9353**
rs216172-C	7305	**9779**
rs12190287-C	5673	7594
rs46522-T	**9298**	**12447**
rs11556924-C	4529	6063
rs964184-G	4475	5991
rs3184504-T	6938	**9289**
rs17114036-A	4105	5495

Estimated using a two-sided alpha = 0.05

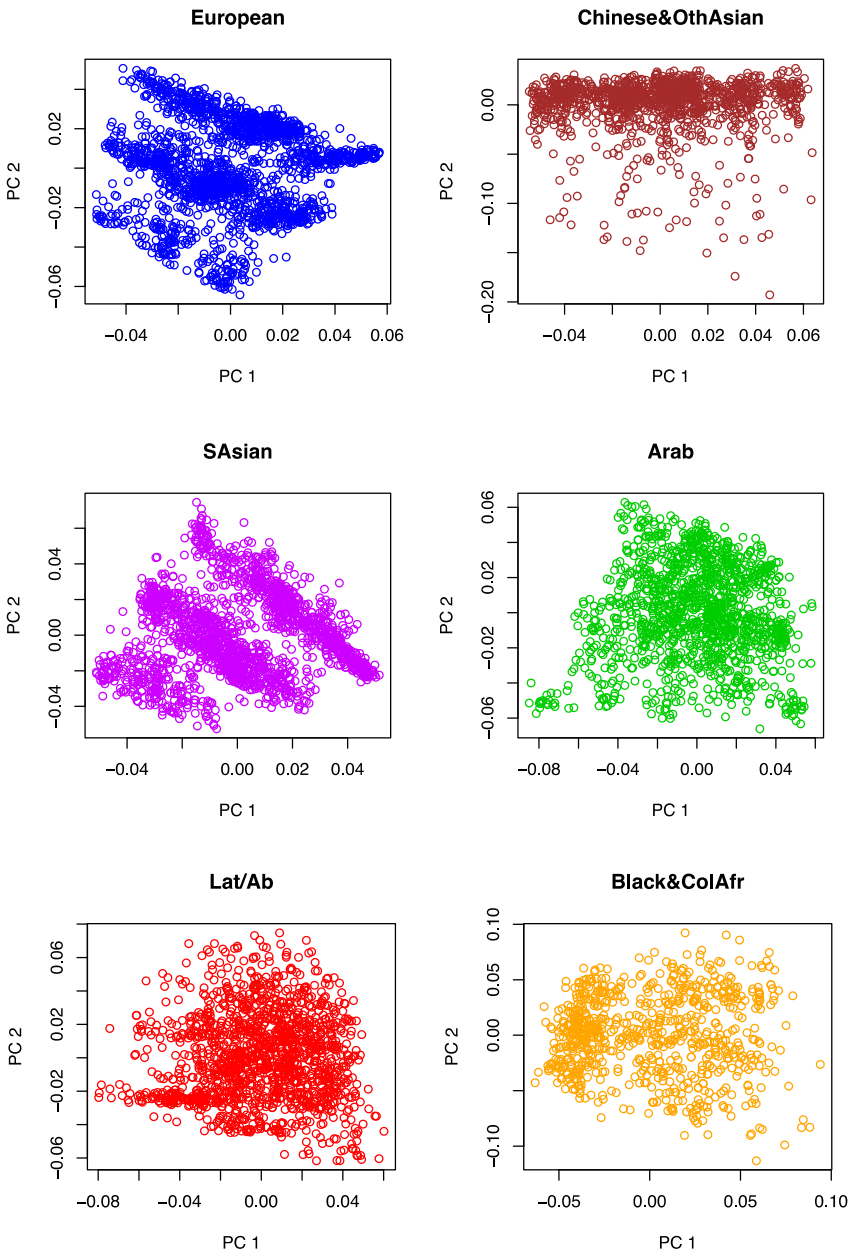
*Bolded numbers represent values that are above the sample size of the study (N=85)

Supplementary Figure 1: Principle Component Analysis

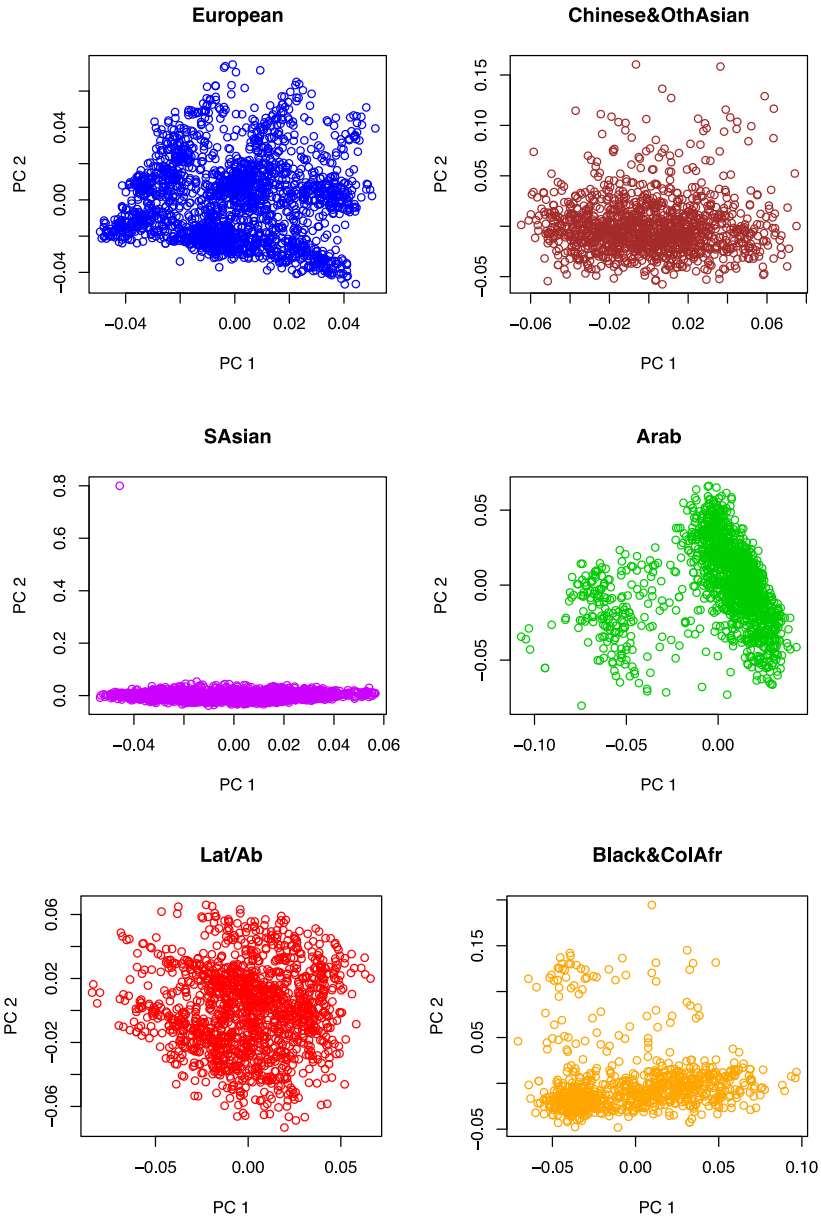
Resulting in Clustering of Individual Ethnic Groups



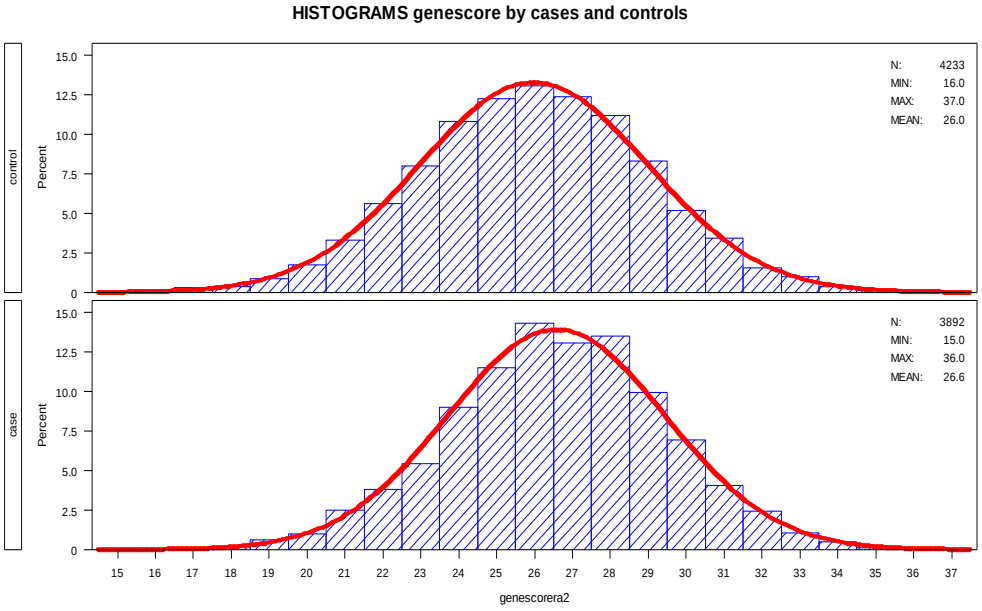
Supplementary Figure 2a: Ethnic Specific Principle Component Analyses for Individual Ethnic Groups



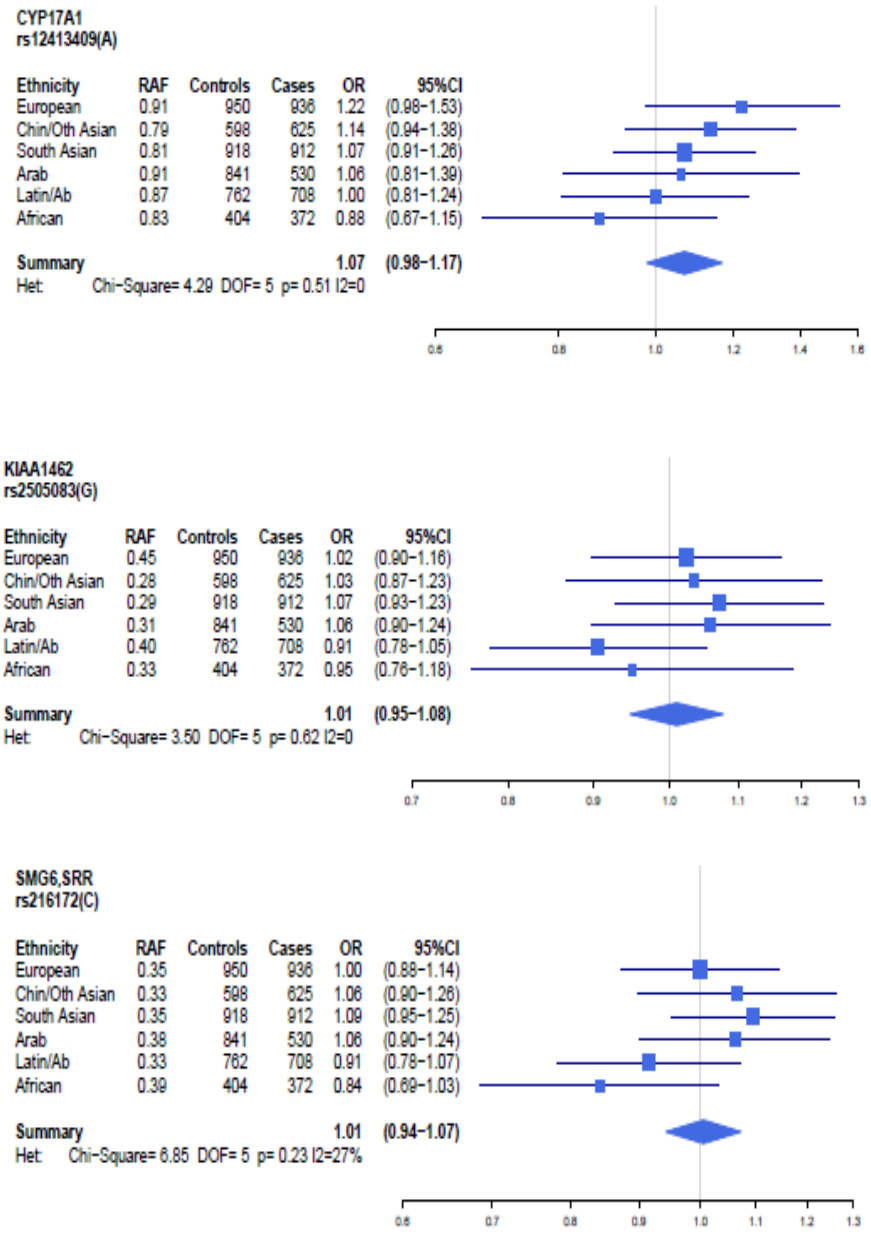
Supplementary Figure 2b: Ethnic Specific Principle Component Analyses for Individual Ethnic Groups After Changing Missing Data Threshold from 0.05 to 0.01

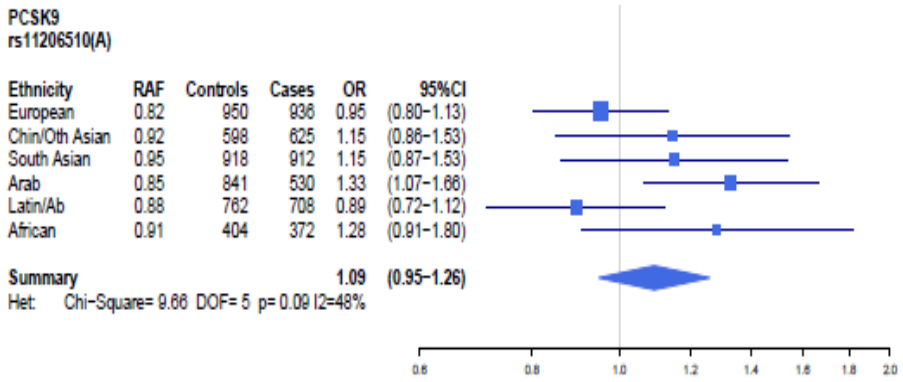


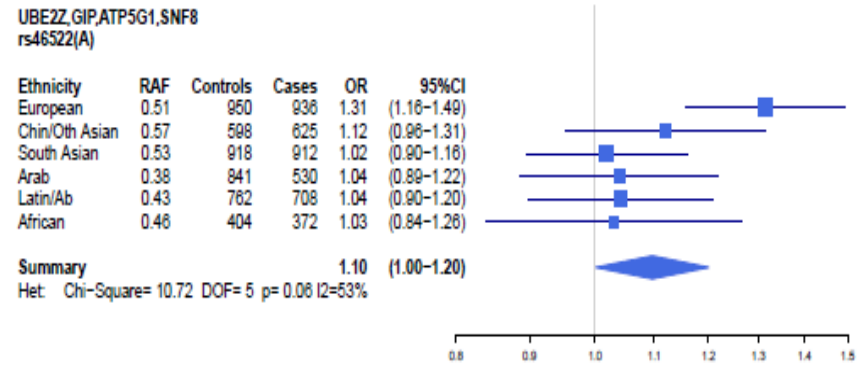
Supplementary Figure 3: Distribution of Genetic Risk Score in Cases and Controls

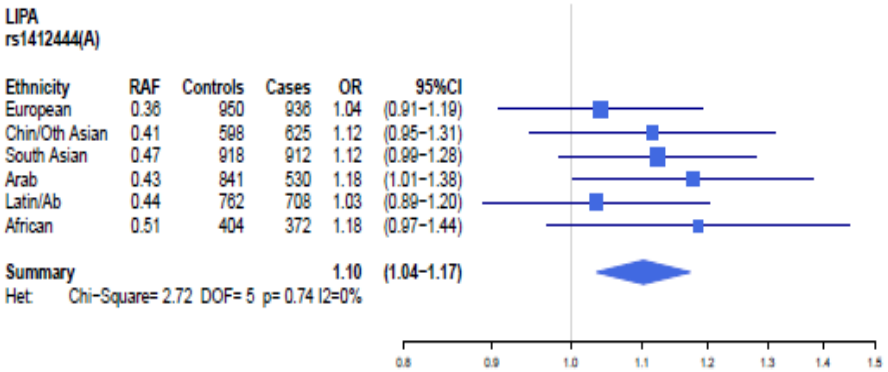


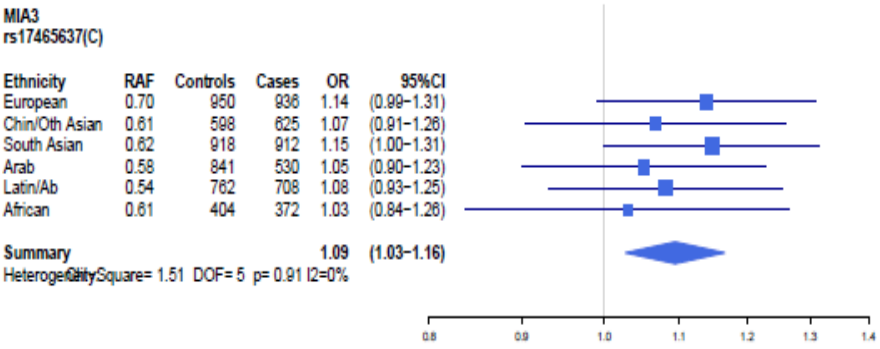
Supplementary Figure 4: Meta-Analysis of Individual Single Nucleotide Polymorphisms and their Association with Myocardial Infarction Risk across Ethnic Groups




PCSK9
rs11206510(A)
Ethnicity RAF Controls Cases OR 95%CI
European 0.82 950 936 0.95 (0.80−1.13)
Chin/Oth Asian 0.92 598 625 1.15 (0.86−1.53)
South Asian 0.95 918 912 1.15 (0.87−1.53)
Arab 0.85 841 530 1.33 (1.07−1.66)
Latin/Ab 0.88 762 708 0.89 (0.72−1.12)
African 0.91 404 372 1.28 (0.91−1.80)
Summary 1.09 (0.95−1.26)
Het: Chi−Square= 9.66 DOF= 5 p= 0.09 I2=48%
0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0


UBE2Z,GIP,ATP5G1,SNF8
rs46522(A)
Ethnicity RAF Controls Cases OR 95%CI
European 0.51 950 936 1.31 (1.16−1.49)
Chin/Oth Asian 0.57 598 625 1.12 (0.96−1.31)
South Asian 0.53 918 912 1.02 (0.90−1.16)
Arab 0.38 841 530 1.04 (0.89−1.22)
Latin/Ab 0.43 762 708 1.04 (0.90−1.20)
African 0.46 404 372 1.03 (0.84−1.26)
Summary 1.10 (1.00−1.20)
Het: Chi−Square= 10.72 DOF= 5 p= 0.06 I2=53%
0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5

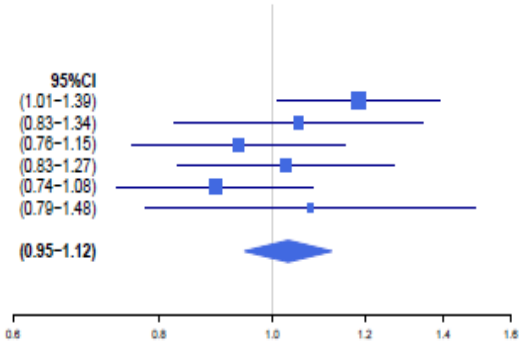

LIPA
rs1412444(A)
Ethnicity RAF Controls Cases OR 95%CI
European 0.36 950 936 1.04 (0.91−1.19)
Chin/Oth Asian 0.41 598 625 1.12 (0.95−1.31)
South Asian 0.47 918 912 1.12 (0.99−1.28)
Arab 0.43 841 530 1.18 (1.01−1.38)
Latin/Ab 0.44 762 708 1.03 (0.89−1.20)
African 0.51 404 372 1.18 (0.97−1.44)
Summary 1.10 (1.04−1.17)
Het: Chi−Square= 2.72 DOF= 5 p= 0.74 I2=0%
0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5


MIA3
rs17465637(C)
Ethnicity RAF Controls Cases OR 95%CI
European 0.70 950 936 1.14 (0.99−1.31)
Chin/Oth Asian 0.61 598 625 1.07 (0.91−1.26)
South Asian 0.62 918 912 1.15 (1.00−1.31)
Arab 0.58 841 530 1.05 (0.90−1.23)
Latin/Ab 0.54 762 708 1.08 (0.93−1.25)
African 0.61 404 372 1.03 (0.84−1.26)
Summary 1.09 (1.03−1.16)
Heterogeneity Square= 1.51 DOF= 5 p= 0.91 I2=0%
0.8 0.9 1.0 1.1 1.2 1.3 1.4

ABO
rs579459(G)

Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.23	950	936	1.18	(1.01–1.39)
Chin/Oth Asian	0.20	598	625	1.05	(0.83–1.34)
South Asian	0.17	918	912	0.94	(0.76–1.15)
Arab	0.17	841	530	1.03	(0.83–1.27)
Latin/Ab	0.18	762	708	0.89	(0.74–1.08)
African	0.16	404	372	1.08	(0.79–1.48)
Summary				**1.03**	**(0.95–1.12)**

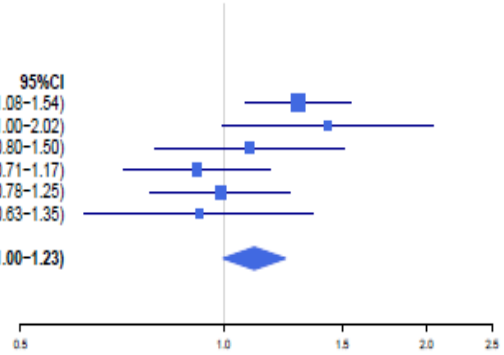
Het: Chi-Square= 6.15 DOF= 5 p= 0.29 I2=19%



SCL5A3-MRPS6-KCNE2
rs9982601(A)

Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.16	950	936	1.29	(1.08–1.54)
Chin/Oth Asian	0.05	598	625	1.42	(1.00–2.02)
South Asian	0.04	918	912	1.09	(0.80–1.50)
Arab	0.11	841	530	0.91	(0.71–1.17)
Latin/Ab	0.11	762	708	0.99	(0.78–1.25)
African	0.08	404	372	0.92	(0.63–1.35)
Summary				**1.11**	**(1.00–1.23)**

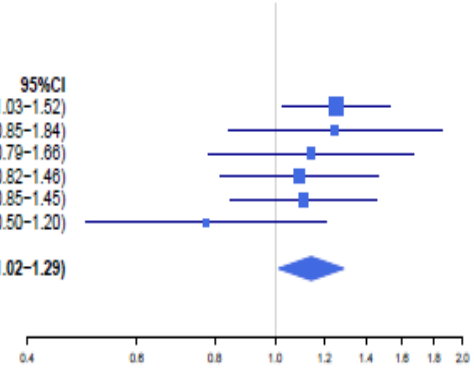
Het: Chi-Square= 8.96 DOF= 5 p= 0.11 I2=44%



WDR12
rs6725887(G)

Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.12	950	936	1.25	(1.03–1.52)
Chin/Oth Asian	0.04	598	625	1.25	(0.85–1.84)
South Asian	0.03	918	912	1.14	(0.79–1.66)
Arab	0.08	841	530	1.10	(0.82–1.46)
Latin/Ab	0.08	762	708	1.11	(0.85–1.45)
African	0.06	404	372	0.78	(0.50–1.20)
Summary				**1.14**	**(1.02–1.29)**

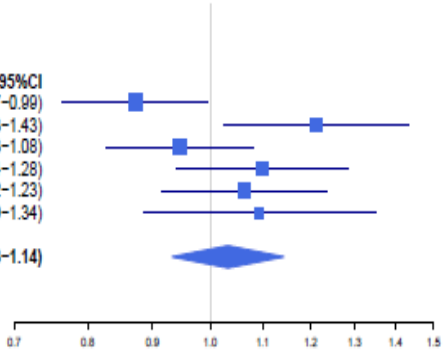
Het: Chi-Square= 4.22 DOF= 5 p= 0.52 I2=0%

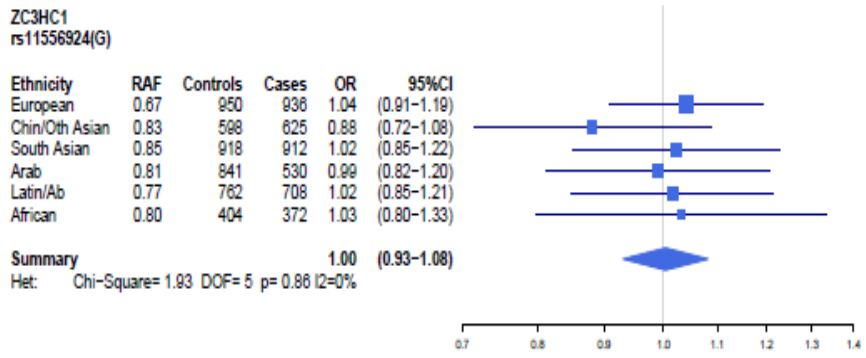


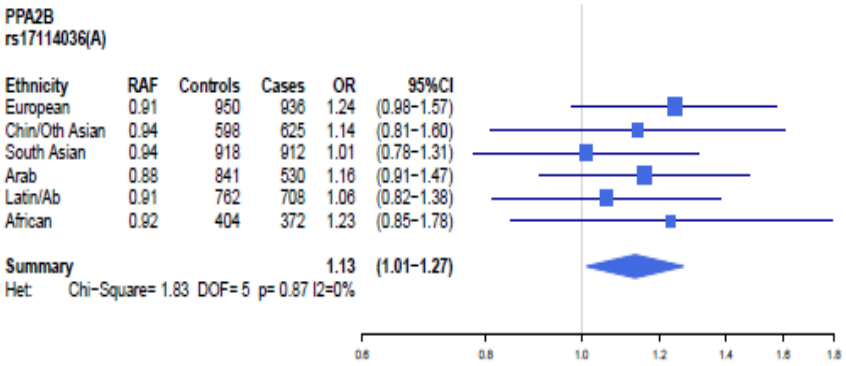
COL4A1,COL4A2
rs4773144(G)

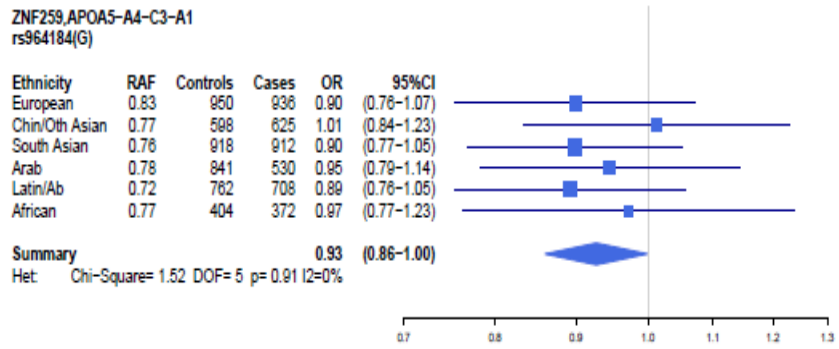
Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.44	950	936	0.87	(0.77–0.99)
Chin/Oth Asian	0.43	598	625	1.21	(1.03–1.43)
South Asian	0.4	918	912	0.95	(0.83–1.08)
Arab	0.51	841	530	1.10	(0.94–1.28)
Latin/Ab	0.45	762	708	1.06	(0.92–1.23)
African	0.45	404	372	1.09	(0.89–1.34)
Summary				**1.03**	**(0.93–1.14)**

Het: Chi-Square= 12.79 DOF= 5 p= 0.03 I2=61%




ZC3HC1
rs11556924(G)
Ethnicity RAF Controls Cases OR 95%CI
European 0.67 950 936 1.04 (0.91−1.19)
Chin/Oth Asian 0.83 598 625 0.88 (0.72−1.08)
South Asian 0.85 918 912 1.02 (0.85−1.22)
Arab 0.81 841 530 0.99 (0.82−1.20)
Latin/Ab 0.77 762 708 1.02 (0.85−1.21)
African 0.80 404 372 1.03 (0.80−1.33)
Summary 1.00 (0.93−1.08)
Het: Chi−Square= 1.93 DOF= 5 p= 0.86 I2=0%
0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4

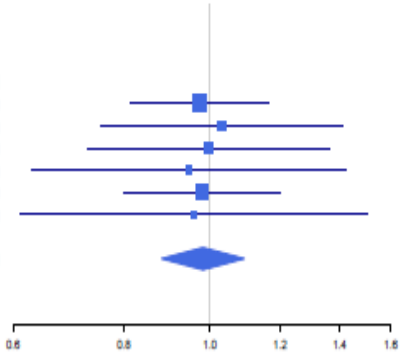

PPA2B
rs17114036(A)
Ethnicity RAF Controls Cases OR 95%CI
European 0.91 950 936 1.24 (0.98−1.57)
Chin/Oth Asian 0.94 598 625 1.14 (0.81−1.60)
South Asian 0.94 918 912 1.01 (0.78−1.31)
Arab 0.88 841 530 1.16 (0.91−1.47)
Latin/Ab 0.91 762 708 1.06 (0.82−1.38)
African 0.92 404 372 1.23 (0.85−1.78)
Summary 1.13 (1.01−1.27)
Het: Chi−Square= 1.83 DOF= 5 p= 0.87 I2=0%
0.6 0.8 1.0 1.2 1.4 1.6 1.8


ZNF259,APOA5−A4−C3−A1
rs964184(G)
Ethnicity RAF Controls Cases OR 95%CI
European 0.83 950 936 0.90 (0.76−1.07)
Chin/Oth Asian 0.77 598 625 1.01 (0.84−1.23)
South Asian 0.76 918 912 0.90 (0.77−1.05)
Arab 0.78 841 530 0.95 (0.79−1.14)
Latin/Ab 0.72 762 708 0.89 (0.76−1.05)
African 0.77 404 372 0.97 (0.77−1.23)
Summary 0.93 (0.86−1.00)
Het: Chi−Square= 1.52 DOF= 5 p= 0.91 I2=0%
0.7 0.8 0.9 1.0 1.1 1.2 1.3

ANKS1A
rs17609940(G)

Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.85	950	936	0.97	(0.82–1.16)
Chin/Oth Asian	0.93	598	625	1.03	(0.76–1.40)
South Asian	0.95	918	912	1.00	(0.73–1.36)
Arab	0.96	841	530	0.95	(0.64–1.41)
Latin/Ab	0.84	762	708	0.98	(0.80–1.19)
African	0.94	404	372	0.96	(0.62–1.49)
Summary				**0.98**	**(0.88–1.09)**

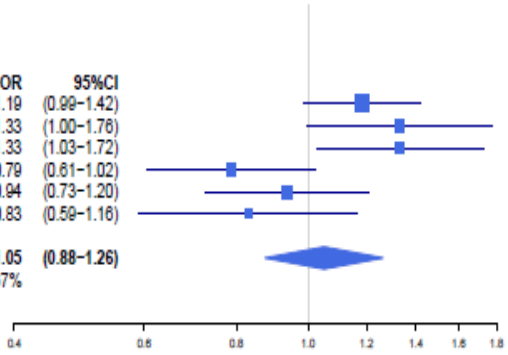
Het: Chi-Square= 0.16 DOF= 5 p= 1.00 I2=0%



MRAS
rs2306374(G)

Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.14	950	936	1.19	(0.99–1.42)
Chin/Oth Asian	0.08	598	625	1.33	(1.00–1.76)
South Asian	0.07	918	912	1.33	(1.03–1.72)
Arab	0.10	841	530	0.79	(0.61–1.02)
Latin/Ab	0.09	762	708	0.94	(0.73–1.20)
African	0.10	404	372	0.83	(0.59–1.16)
Summary				**1.05**	**(0.88–1.26)**

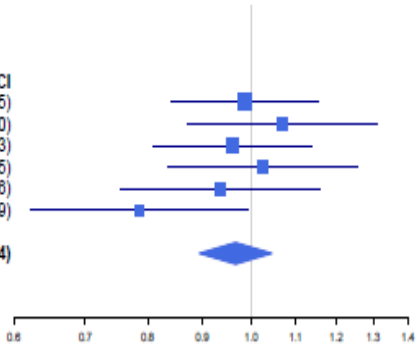
Het: Chi-Square= 15.21 DOF= 5 p= 0.01 I2=67%



LDLR
rs1122608(C)

Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.79	950	936	0.99	(0.85–1.15)
Chin/Oth Asian	0.81	598	625	1.07	(0.88–1.30)
South Asian	0.82	918	912	0.96	(0.81–1.13)
Arab	0.81	841	530	1.02	(0.84–1.25)
Latin/Ab	0.85	762	708	0.94	(0.76–1.16)
African	0.75	404	372	0.79	(0.63–0.99)
Summary				**0.97**	**(0.90–1.04)**

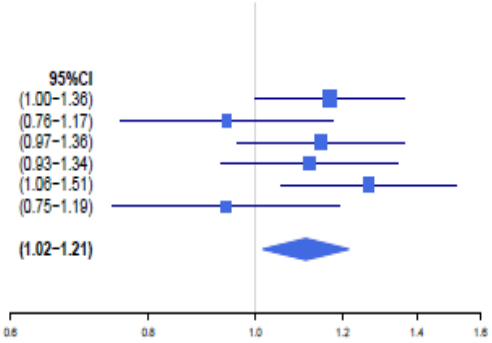
Het: Chi-Square= 4.58 DOF= 5 p= 0.47 I2=0%

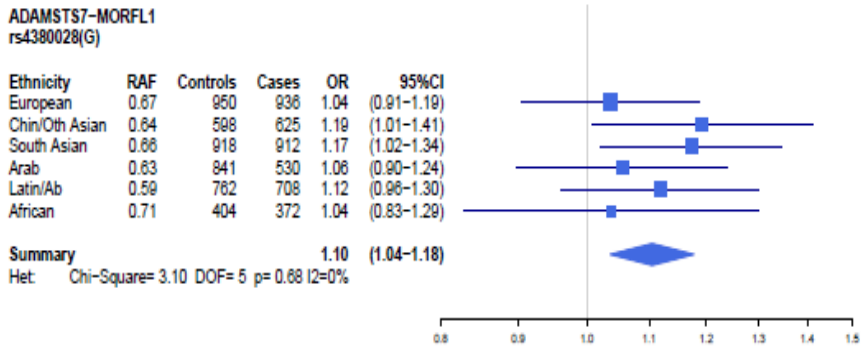


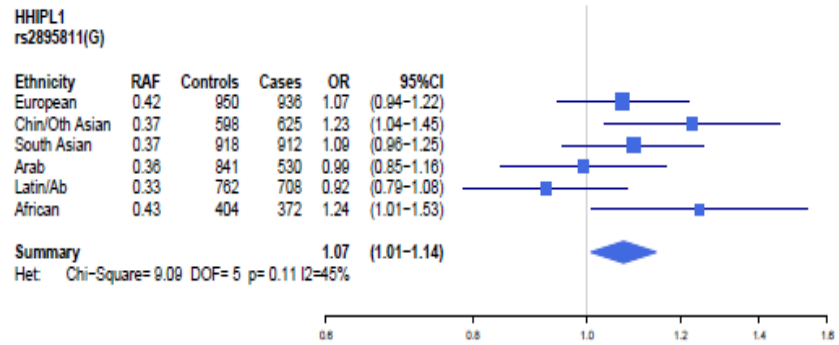
CELSR2-PSRC1-SORT1
rs646776(A)

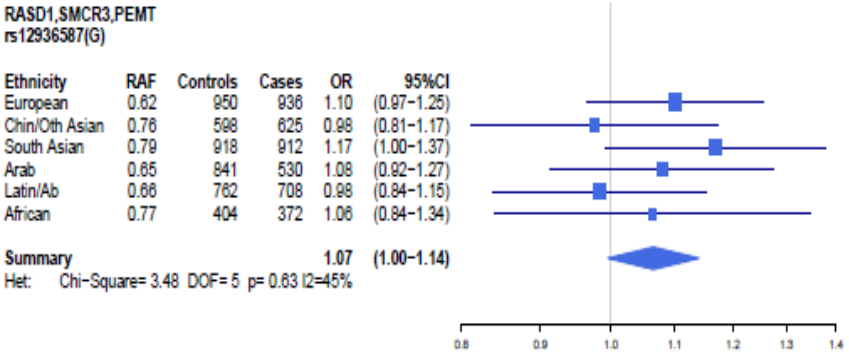
Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.77	950	936	1.17	(1.00–1.36)
Chin/Oth Asian	0.85	598	625	0.94	(0.76–1.17)
South Asian	0.83	918	912	1.15	(0.97–1.36)
Arab	0.75	841	530	1.12	(0.93–1.34)
Latin/Ab	0.76	762	708	1.27	(1.06–1.51)
African	0.77	404	372	0.94	(0.75–1.19)
Summary				**1.11**	**(1.02–1.21)**

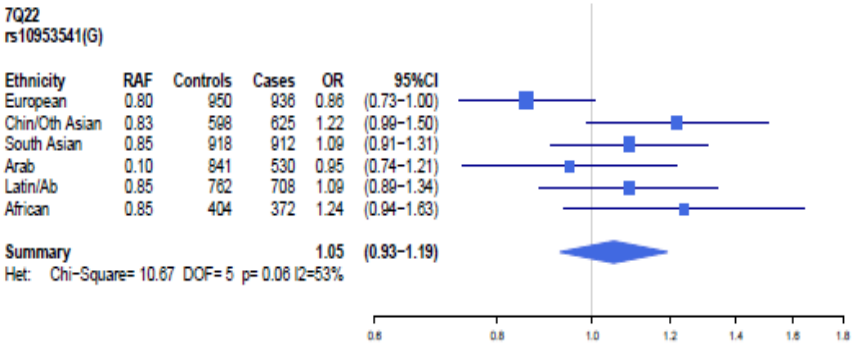
Het: Chi-Square= 6.80 DOF= 5 p= 0.24 I2=26%




ADAMSTS7−MORFL1
rs4380028(G)
Ethnicity RAF Controls Cases OR 95%CI
European 0.67 950 936 1.04 (0.91−1.19)
Chin/Oth Asian 0.64 598 625 1.19 (1.01−1.41)
South Asian 0.66 918 912 1.17 (1.02−1.34)
Arab 0.63 841 530 1.06 (0.90−1.24)
Latin/Ab 0.59 762 708 1.12 (0.96−1.30)
African 0.71 404 372 1.04 (0.83−1.29)
Summary 1.10 (1.04−1.18)
Het: Chi−Square= 3.10 DOF= 5 p= 0.68 I2=0%
0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5


HHIPL1
rs2895811(G)
Ethnicity RAF Controls Cases OR 95%CI
European 0.42 950 936 1.07 (0.94−1.22)
Chin/Oth Asian 0.37 598 625 1.23 (1.04−1.45)
South Asian 0.37 918 912 1.09 (0.96−1.25)
Arab 0.36 841 530 0.99 (0.85−1.16)
Latin/Ab 0.33 762 708 0.92 (0.79−1.08)
African 0.43 404 372 1.24 (1.01−1.53)
Summary 1.07 (1.01−1.14)
Het: Chi−Square= 9.09 DOF= 5 p= 0.11 I2=45%
0.6 0.8 1.0 1.2 1.4 1.6

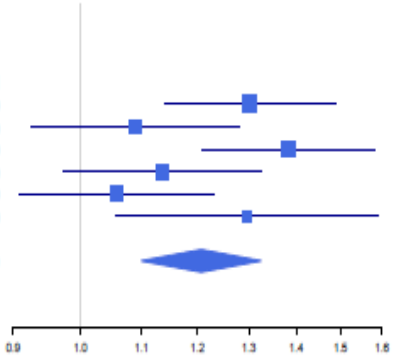

RASD1,SMCR3,PEMT
rs12936587(G)
Ethnicity RAF Controls Cases OR 95%CI
European 0.62 950 936 1.10 (0.97−1.25)
Chin/Oth Asian 0.76 598 625 0.98 (0.81−1.17)
South Asian 0.79 918 912 1.17 (1.00−1.37)
Arab 0.65 841 530 1.08 (0.92−1.27)
Latin/Ab 0.66 762 708 0.98 (0.84−1.15)
African 0.77 404 372 1.06 (0.84−1.34)
Summary 1.07 (1.00−1.14)
Het: Chi−Square= 3.48 DOF= 5 p= 0.63 I2=45%
0.8 0.9 1.0 1.1 1.2 1.3 1.4


7Q22
rs10953541(G)
Ethnicity RAF Controls Cases OR 95%CI
European 0.80 950 936 0.86 (0.73−1.00)
Chin/Oth Asian 0.83 598 625 1.22 (0.99−1.50)
South Asian 0.85 918 912 1.09 (0.91−1.31)
Arab 0.10 841 530 0.95 (0.74−1.21)
Latin/Ab 0.85 762 708 1.09 (0.89−1.34)
African 0.85 404 372 1.24 (0.94−1.63)
Summary 1.05 (0.93−1.19)
Het: Chi−Square= 10.67 DOF= 5 p= 0.06 I2=53%
0.6 0.8 1.0 1.2 1.4 1.6 1.8

9p21
rs4977574(G)

Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.47	950	936	1.30	(1.14–1.48)
Chin/Oth Asian	0.52	598	625	1.09	(0.93–1.28)
South Asian	0.54	918	912	1.38	(1.21–1.58)
Arab	0.51	841	530	1.14	(0.98–1.32)
Latin/Ab	0.47	762	708	1.06	(0.91–1.23)
African	0.54	404	372	1.30	(1.06–1.58)
Summary				**1.21**	**(1.10–1.32)**

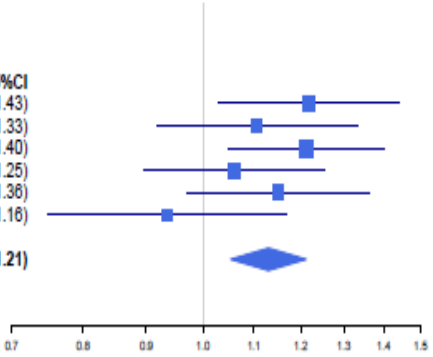
Het: Chi-Square= 10.95 DOF= 5 p= 0.05 I2=54%



CXCL12
rs1746048(G)

Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.82	950	936	1.22	(1.03–1.43)
Chin/Oth Asian	0.74	598	625	1.11	(0.92–1.33)
South Asian	0.69	918	912	1.21	(1.05–1.40)
Arab	0.67	841	530	1.06	(0.90–1.25)
Latin/Ab	0.71	762	708	1.15	(0.97–1.36)
African	0.68	404	372	0.94	(0.75–1.16)
Summary				**1.13**	**(1.05–1.21)**

Het: Chi-Square= 5.29 DOF= 5 p= 0.38 I2=6%



LPA
rs3798220(G)

Ethnicity	RAF	Controls	Cases	OR	95%CI
European	0.02	950	936	1.93	(1.19–3.14)
Chin/Oth Asian	0.04	598	625	0.83	(0.56–1.22)
South Asian	0.04	918	912	0.85	(0.62–1.17)
Arab	0.01	841	530	1.58	(0.75–3.31)
Latin/Ab	0.15	762	708	1.16	(0.88–1.54)
African	0.006	404	372	0.32	(0.07–1.46)
Summary				**1.09**	**(0.80–1.48)**

Het: Chi-Square= 13.20 DOF= 5 p= 0.02 I2=62%