Interactions between Cardiovascular Risk Factors and a Coronary Artery Disease Genetic Risk Score

Justine Desrochers

Department of Human Genetics, Faculty of Medicine and Health Sciences,

McGill University, Montreal, Quebec, Canada

December 2023

A thesis submitted to the McGill University in partial fulfillment of the requirements of the degree of Master of Science.

Abstract

Objective : Genome-wide association studies of coronary heart disease (CAD) have identified many variants that contribute to its etiology, frequently with small effect sizes. When combined into a genetic risk score (GRS), the sum of these variants demonstrates larger effects, allowing for a personalized prediction of risk. However, a GRS may interact with known risk factors and GRS associations with diseases may not be consistent across all sub-populations. My study aimed to identify potential GRS interactions with age, sex, hypertension, dyslipidemia, obesity, lipoprotein(a), smoking and diabetes.

Approach and Results : Using Cox proportional hazard models for incident CAD in 344,130 unrelated individuals of European ancestry in the UK Biobank, I analyzed a CAD GRS containing 204 single nucleotide polymorphisms (SNPs) (denoted as GRS_{204}). I also examined GRS subsets by partitioning variants based on their effects on four atherosclerotic risk factors: apolipoprotein B (apoB), lipoprotein a (Lp(a)), diabetes mellitus (DM) and hypertension (HTN). The GRS_{204} was significantly associated with incident CAD (HR per standard deviation (95% CI), 1.37 (1.35, 1.40); $P < 2 \times 10^{-16}$). The effect of the GRS_{204} on incident CAD decreased with age (HRs of 1.47 (1.43, 1.52); and 1.33 (1.31, 1.36) for individuals <55 and ≥55 respectively (interaction $P = 3.60 \times 10^{-8}$)). The GRS_{204} demonstrated a significantly stronger association in men (HRs of 1.40 (1.38, 1.43); and 1.32 (1.29, 1.36) for men and women respectively; both $P < 2 \times 10^{-16}$; interaction $P = 1.09 \times 10^{-4}$). The GRS_{204} also significantly interacted with diabetes and dyslipidemia, with a stronger association observed in non-diabetic individuals compared to diabetics (HRs of 1.39 (1.37, 1.41), and 1.26 (1.21, 1.32), respectively; interaction $P = 7.28 \times 10^{-7}$) and in individuals with dyslipidemia (HR of 1.40 (1.36, 1.45), compared to those without

dyslipidemia (HR (95% CI), 1.34 (1.32, 1.37), interaction P = 4.11 x 10^{-3}). These results were generally consistent across all of the GRS subsets.

Conclusion : The GRS_{204} demonstrated a stronger association in men, younger individuals, those without diabetes, and those with dyslipidemia. GRS interactions may identify subgroups of individuals at higher genetic risk and improve risk prediction.

Résumé

Objectif : Des études d'associations pangénomiques sur la maladie coronarienne (MC) ont identifié des variants génétiques qui contribuent à son étiologie, typiquement avec de faibles effets individuels. Quand ces variants sont combinés dans un score de risque génétique (GRS), leur somme démontre des effets plus importants, permettant une prédiction plus précise du risque. Un GRS peut toutefois être influencé par des facteurs de risque connus et les associations entre GRS et MC peuvent ne pas être constants dans toutes les sous-populations. La présente étude visait à identifier les interactions potentielles du GRS avec l'âge, le sexe, l'hypertension, la dyslipidémie, l'obésité, la lipoprotéine(a), le tabagisme et le diabète.

Approches et résultats : En utilisant des modèles de risques proportionnels de Cox pour la MC incidente chez 344 130 individus d'ascendance européenne non apparentés de la UK Biobank, j'ai analysé un GRS de la maladie coronarienne contenant 204 polymorphismes de nucléotides simples (SNP). J'ai également examiné certains sous-ensembles de ce GRS en séparant les variants en fonction de leurs effets sur les facteurs de risque de l'athérosclérose: l'apolipoprotéine B (apoB), la lipoprotéine a (Lp(a)), le diabète (DM) et l'hypertension artérielle (HTN). Le GRS_{204} a été associé de manière significative à la MC incidente (ratio de risque (RR) par écart-type (intervalle de confiance à 95%), 1.37 (1.35, 1.40); $P < 2 \times 10^{-16}$). L'effet du GRS204 sur l'incidence de la MC diminue avec l'âge (RR de 1.47 (1.43, 1.52) et 1.33 (1.31, 1.36) pour les individus âgés <55 et ≥55 ans respectivement). Le GRS_{204} a aussi démontré une association significativement plus forte chez les hommes (RR de 1.40 (1.38, 1.43); et 1.32 (1.29, 1.36) pour les hommes et les femmes respectivement; $P < 2 \times 10^{-16}$ dans les deux cas; interaction $P = 1.09 \times 10^{-4}$). Le GRS_{204} interagit également de manière significative avec le diabète et la dyslipidémie, avec une association plus forte observée chez les non-diabétiques que chez les diabétiques (RR

de 1.39 (1.37, 1.41), et 1.26 (1.21, 1.32), respectivement; interaction P = 7.28 x 10^{-7}) et chez les personnes souffrant de dyslipidémie (RR de 1.40 (1.36, 1.45), par rapport aux personnes sans dyslipidémie (RR de 1.34 (1.32, 1.37), interaction P = 4.11 x 10^{-3}). Ces résultats sont généralement constants pour tous les sous-ensembles du GRS analysés.

Conclusion : Le GRS_{204} a démontré une association plus forte chez les hommes, les individus plus jeunes, ceux qui n'ont pas de diabète et ceux qui souffrent de dyslipidémie. Les facteurs d'interaction avec le GRS peuvent permettre d'identifier des sous-groupes d'individus présentant un risque génétique plus élevé et d'améliorer la prédiction du risque.

Table of Contents

List of Abbreviations

95%CI	95 percent confidence interval
ApoB	apolipoprotein B
BMI	body mass index
C4D	the coronary artery disease genetics
CAD	coronary artery disease genetics
CARDIOGRAM	coronary artery disease genome wide replication and meta-analysis study
CARDIoGRAMplusC4D	coronary artery disease genome wide replication and meta-analysis plus the coronary artery disease genetics
CLSA	canadian longitudinal study on aging
CNV	copy number variant
DM	diabetes mellitus
DNA	deoxyribonucleic acid
GRS	genetic risk score
GWAS	genome-wide association study
HDL	high-density lipoprotein
HR	hazard ratio
HTN	hypertension
LD	linkage disequilibrium
LDL-C	low-density lipoprotein cholesterol
Lp(a)	lipoprotein(a)
MESA	multi-ethnic study of atherosclerosis
MI-GENES	myocardial infarction genes
MVP	million veteran program
NRI	net reclassification index
OR	odds ratio
P+T	pruning plus thresholding
PCE	pooled cohort equation
PCE+CAD GRS	pooled cohort equation and coronary artery disease genetic risk score
PCSK9	proprotein convertase subtilisin/kexin type 9
RF+CAD GRS	risk factors and coronary artery disease genetic risk score
ROS	reactive oxygen species
SNP	single nucleotide polymorphism
UKB	united kingdom biobank
X-chr	x-chromosome

List of Figures

List of tables

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my co-supervisors, Dr. Jamie Engert and Dr. George Thanassoulis, for taking me on as an undergraduate and graduate student. Their invaluable feedback and mentorship helped me succeed and flourish throughout this degree.

Additionally, this endeavour would not have been possible without the guidance of my supervisory committee members, Dr. Patrick Dion and Dr. Jacques Genest. They generously provided insights and expertise which greatly elevated the quality and methodology of my project.

I would also like to extend my sincere gratitude to the indispensable members of the Engert-Thanassoulis laboratories. In addition to creating a warm and welcoming environment, they provided thoughtful suggestions and encouragement when needed. Special thanks to Ms. Line Dufresne, without whom none of this project would have been possible. Her patience, statistical knowledge and precious feedback were invaluable for my project. I am also grateful to the students from this laboratory for troubleshooting certain things with me and for brainstorming project ideas with me.

Lastly, I would like to thank my family and close friends for their unwavering support and encouragement over the past two and a half years. Special thanks to my dear parents for always believing in me along the way.

Chapter 1: Introduction

Coronary artery disease (CAD), the most common type of heart disease, is the leading cause of death worldwide[1]. In the last few decades, genome-wide association studies (GWAS) have enabled the large-scale detection of single nucleotide polymorphisms (SNPs) associated with CAD[2-6]. CAD risk variants can be integrated into a CAD genetic risk score (GRS), capturing and quantifying part of an individual's genetic susceptibility to CAD[7-9], and several CAD GRS studies based on previously identified genome-wide significant variants have been performed[8,10,11]. Recently, age and sex[12,13] have both been shown to interact with a CAD GRS, but fewer studies[8,14,15] have investigated interactions with other risk factors or specific pleiotropy between CAD GRS and other atherosclerotic risk factors such as diabetes, hypertension, dyslipidemia, smoking and body mass index (BMI). In this study, I aimed to investigate the relationship between a CAD GRS, including specific SNP subsets based on their association with risk factors, and several prominent CAD risk factors.

1.1 Coronary Artery Disease

Coronary artery disease (CAD) accounts for one third of all deaths worldwide[1,16,17]. According to the 2022 Heart Disease and Stroke Statistics update from the American Heart Association, CAD prevalence is 7.2% in American adults above the age of 20 years old. Notably, American men have a higher CAD prevalence than women (8.3% and 6.2%, respectively)[17]. In Canada in 2022, 2.6 million people aged 20 and older were living with CAD[18]. In addition, Zhu *et al.* detected opposing trends of CAD prevalence in developed and developing countries, noting that the prevalence is rising in developing countries, while decreasing in developed countries[19]. In terms of socioeconomic status, individuals with a low socioeconomic status have a higher

prevalence than individuals with a high socioeconomic status[20]. They are also more likely to have poorer outcomes[20].

The number of CAD cases has increased by over five million in the United States in the past decade in part due to the rise in the number of individuals affected by metabolic CAD risk factors such as obesity, diabetes, hypertension and dyslipidemia[16,17,21]. This has created a large economic burden on the healthcare system. From 1996 to 2016, total spending on cardiovascular care in US adults increased from 212 billion dollars to 320 billion dollars (public, private payers and out-of-pocket spending included)[22]. This budget, which represents 15% of all United States health care spending, is needed to treat and manage CAD and CAD risk factors[22]. Likewise, CAD has also generated a substantial societal burden whereby affected individuals prematurely exit the labor market following dire disability or death[23,24]. In addition, medical leave of absences, long-term leave of absence due to hospitalization, and reduction of working hours due to disability all contribute to the societal burden of CAD[25].

1.1.1 CAD Pathogenesis

Atherosclerosis is the main driver of CAD pathogenesis[26] (Figure 1). Characterized by a deregulation of lipid homeostasis, this process develops over a long period of time, up to 50 years in some cases[27]. The initial stage of atherosclerosis is endothelial cell dysfunction. This occurs when the lining of the arterial intima becomes unstable, allowing for the accumulation and retention of monocytes and various lipids, including LDL-C and Lp(a), through leaky junctions[28,29]. The injured endothelial cells involved in endothelial activation release numerous chemokines, inflammatory cytokines, and mediators which leads to increased reactive oxygen species (ROS) levels[30-32]. High ROS levels can, in turn, stimulate an inflammatory response[33].

This creates a cycle where high ROS levels and high levels of inflammatory markers positively influence one another[34].

Monocytes retained in the arterial intima mature into macrophages through specific chemokines and cytokines such as interleukin-8[29,35]. Elevated ROS levels establish an ideal environment where LDL-C and Lp(a) are oxidized and subsequently phagocytosed by macrophages[28]. As oxidized LDL-C accumulates in macrophages, these cells convert into foam cells[28,29], which cluster together to form fatty streaks[30]. With time, the latter evolve into fibrous plaques which are characterized as stable or unstable depending on the amount of inflammation still present[36]. An inflammation-rich plaque is susceptible to thinning of its fibrous cap and rupture of the plaque[30]. This results in thrombus formation and adverse cardiac events such as myocardial infarction, stroke, and death[27,30].

1.1.2 Risk Factors

Risk factors for CAD can, for the most part, be classified into two categories: non-modifiable (age, sex, and family history) and modifiable (e.g. obesity, smoking, dyslipidemia, diabetes, and hypertension)[37]. Age is an independent risk factor for CAD despite it being associated with other risk factors such as diabetes and hypertension[38,39]. Ageing acts through various pathways including increased endothelial cell dysfunction and greater ROS production levels[38,40,41]. Another non-modifiable risk factor is sex: women tend to develop CAD 7 to 10 years later than men[42,43]. Additionally, women have poorer clinical outcomes than men following

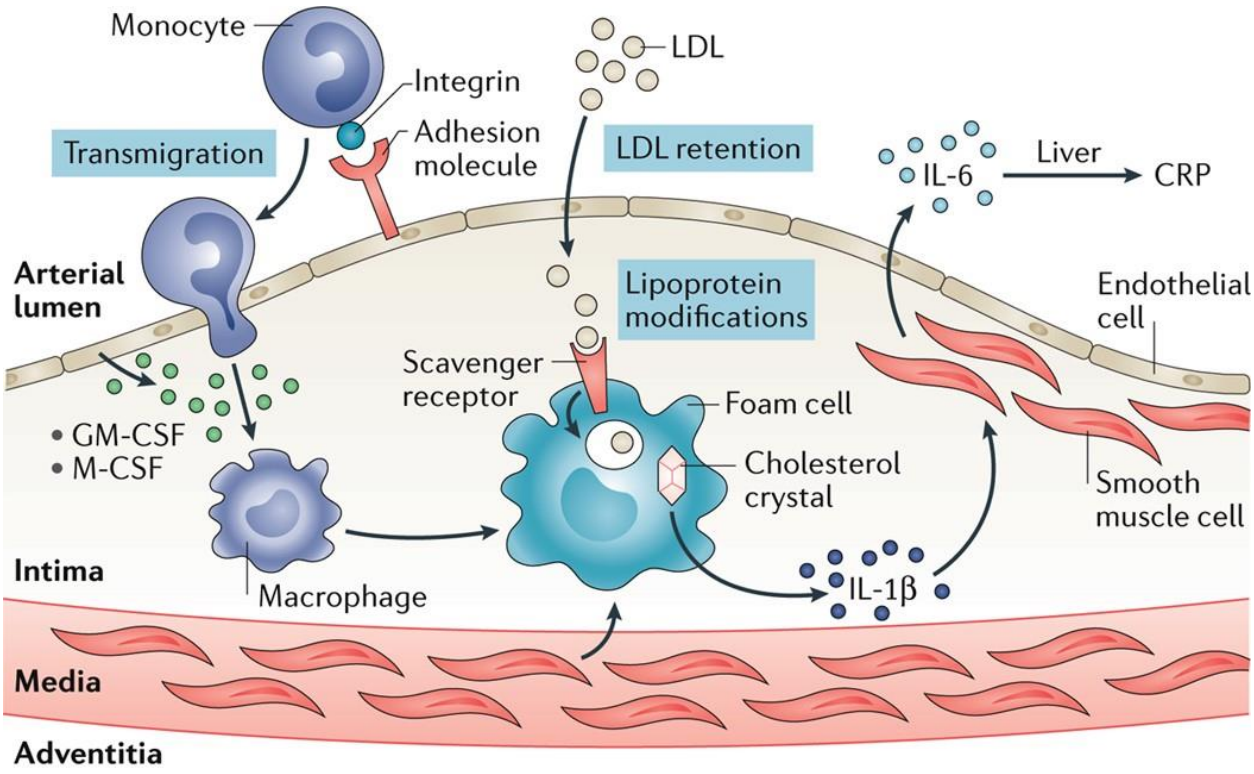


Figure 1: Steps involved in atherosclerosis

Stepwise sequences of atherosclerosis development in the arterial lumen and intima. LDL, low-density lipoprotein; GM-CSF, granulocyte-macrophage colony-stimulating factor; M-CSF, macrophage colony-stimulating factor; IL-1B, interleukin 1 beta; CRP, C-reactive protein. (Gistera A, Hansson GK. The immunology of atherosclerosis. *Nat Rev Nephrol*. Jun 2017;13(6):368-380) (Reproduced with permission of the publisher)

CAD diagnosis[44]. As for family history of CAD, it was found by the Framingham study to increase CAD by 2.6 and 2.3-fold in men and women, respectively, when defined as a parental CAD event[45].

Many modifiable risk factors are also independent and affect different stages of the atherosclerosis process. First, obesity, defined as a BMI greater or equal to 30kg/m^2 [46], is an independent risk factor for CAD[47]. It primarily affects early stages of atherosclerosis. For example, obese individuals overexpress pro-inflammatory cytokines which leads to an upregulation of LDL-C and Lp(a) oxidation in the arterial intima[48,49]. Additionally, obesity is a contributor to diabetes as it predisposes individuals to insulin resistance and beta-cell dysfunction, among other processes[50,51]. A second independent risk factor for CAD is diabetes mellitus (type I and type II). Diabetes-induced inflammation triggers a pro-inflammatory cytokine response, thus accelerating endothelial dysfunction and ROS production[52,53] which in turn hastens the accumulation and oxidation of lipids in the arterial intima and the formation of foam cells. The third modifiable CAD risk factor to highlight is hypertension, defined as a systolic blood pressure above 130 mmHg or a diastolic blood pressure above 80 mmHg[54]. High blood pressure increases the permeability of the endothelium in the arterial intima, thus allowing more monocytes and lipids to migrate into the arterial wall[55]. In addition, hypertension also increases the likelihood of plaques becoming unstable or rupturing[56]. Dyslipidemia, defined as deregulation of lipid levels in the blood (i.e., high total cholesterol or LDL-C or triglyceride levels or low HDL cholesterol levels), is another independent risk factor for CAD which increases the accumulation of various lipids in the arterial wall, and is a major contributor to atherosclerosis[57,58]. Lastly, smoking, a preventable risk factor for CAD, aggravates many pathways involved in atherosclerosis by promoting vascular and endothelial dysfunction,

increasing the build up of lipids in the arterial intima and their oxidization, and creating a procoagulant state susceptible to thrombosis[59,60].

One exception is Lp(a), an independent risk factor for CAD. High Lp(a) is noted clinically when the concentration is greater than 50mg/dL[61]. High Lp(a) levels accelerate atherosclerosis by, in part, promoting endothelial activation and the formation of foam cells through increased Lp(a) phagocytosis[62,63]. Notably, elevated Lp(a) plasma levels are largely independent of LDL-C plasma levels even though both lipid particles are atherogenic[30].

1.1.3 CAD Assessment and Treatment

Clinicians often use assessment tools to estimate the risk of CAD in individuals. The American Heart Association advocates for the use of the Pooled Cohort Equations (PCE) if the individuals have no pre-existing cardiovascular disease and they are between the ages of 40 and 79[64]. The PCE estimates an individual's 10-year risk of atherosclerotic cardiovascular disease. It relies on age, sex, race and atherosclerotic risk factors including diabetes, smoking status and systolic blood pressure[64,65]. Individuals with an elevated 10-year risk (≥ 7.5%) can either be recommended a high-intensity or moderate-intensity treatment plan in addition to diet and lifestyle recommendations for primary prevention of cardiovascular disease[66].

CAD treatment can vary between individuals as any treatment and recommendations are tailored to an individual's disease severity and comorbidities. One of the initial treatment options for CAD is diet and lifestyle modifications aiming to mitigate certain risk factors[67]. Diet changes can include following dietary guidelines to help maintain an appropriate body weight and lower cholesterol and blood pressure levels. Lifestyle modifications including limiting alcohol consumption, weight management, regular and consistent exercise and cigarette cessation can

mitigate atherosclerosis progression and the risk of thrombosis. Additionally, medical therapy ranges from anti-ischemic drugs like beta-blockers and antiplatelet drugs to antithrombotic and cholesterol-lowering medications such as statins[36]. Finally, heart surgery can be a critical component of CAD treatment depending on disease progression[68]. Surgeries such as minimally invasive heart surgery, percutaneous coronary intervention and coronary artery bypass grafting are routinely used to treat CAD.

1.2 Genetics

Since the late 20th century, scientists and epidemiologists noted familial clustering of CAD: individuals with a positive family history of CAD are far more likely to be diagnosed with the disease. Family aggregation studies from the 1990s onwards estimated that the presence of a family history of CAD increased an individual's risk of CAD by 2.5 to 4-fold compared to no family history[69-71]. As these studies did not account for environmental factors such as smoking and diet, later studies focused on twins who shared the same common environment. Well-known twin studies from Nordic countries have evaluated CAD heritability to be around 40 to 60%[72-74]. Although twin studies hinted at a significant genetic component, it was the completion of the Human Genome Project in 2003, which provided the first sequence of the human genome. This initial sequence covered around 92% of the total human genome[75] and paved the way for tremendous progress in CAD genetics.

While identifying genetic variants was not the main goal of the Human Genome Project, it provided an ideal opportunity to annotate and analyze around 3 million SNPs found in the human genome[76]. Shortly thereafter another global research project, the International HapMap Consortium, characterized SNPs genotyped in 270 individuals[77]. Notably, these individuals come from four diverse populations (or genetic ancestries) to acquire and compare the allele frequency

and linkage disequilibrium (LD) differences across populations[77]. In total, the International HapMap Consortium eventually annotated approximately 3.1 million SNPs[78].

1.2.1 Genome-Wide Association Studies

The publication of human sequences and polymorphism databases by the Human Genome Project and the International HapMap Consortium paved the way for the first genome-wide association studies (GWAS) to be performed[5]. The purpose of a GWAS is to detect genetic variants statistically associated with a disease or trait by comparing allele or genotype frequencies of variants in cases versus controls. A GWAS for a particular disease or trait usually consists of three main steps: recruitment or identification of cases and controls, genotyping individuals to identify genetic variants and performing association analyses[79] (Figure 2).

In 2007, the first GWAS for CAD was conducted in individuals of European ancestry and published by three independent research groups[80-82]. They discovered the first locus predisposing to CAD at chromosome 9p21. The 9p21 locus contains around 60 SNPs in high linkage disequilibrium with each other. In addition, many of these SNPs are very common in individuals of European ancestry[2]. Approximately 75% of individuals of European ancestry carry at least one risk allele. These research groups also emphasized that the risk conferred by variants in the 9p21 locus was independent of known risk factors for CAD[80-82]. Notably, these findings from the first GWAS were replicated in many studies of European individuals[83] as well as other ancestries[84,85].

In the years following these GWAS publications, databanks accelerated recruitment of cases and controls for a myriad of diseases and disorders, facilitating the production and publication of dozens of cardiovascular GWAS. Larger sample sizes enabled the discovery of more associated novel loci with smaller effect sizes. Research groups from around the world

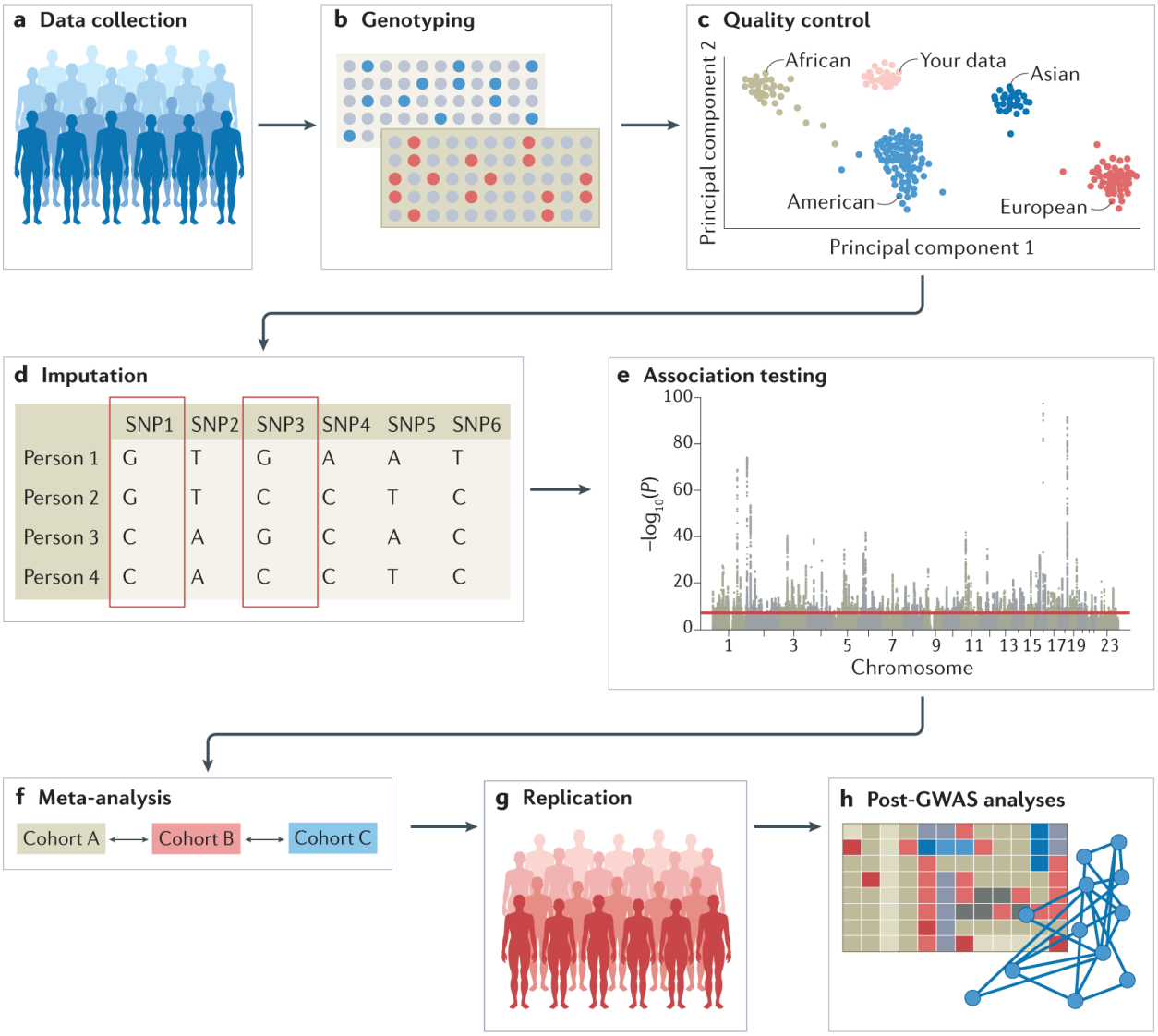


Figure 2: Synopsis of steps necessary to conduct a GWAS

Step-wise visual representation of the different steps involved in conducting a GWAS. (Uffelmann E, Huang Q, Munung N, et al. Genome-wide association studies. *Nature Reviews Methods Primers*. 2021;1(1):1-21) (Reproduced with permission of the publisher)

came together to create international consortia such as CARDIoGRAM and C4D which were extremely successful in discovering novel CAD polymorphisms[7]. In the past decade, these two consortia as well as their combined consortium (CARDIoGRAMplusC4D) have identified over 80 novel CAD SNPs, mostly in individuals of European ancestry. Moreover, extremely large databanks have started releasing their genotypic data to researchers. For example, since 2017, the UK Biobank (UKB), a large prospective study that recruited over 502,000 participants and collected their genotypic and phenotypic information, has become a rich resource for the entire scientific community[86]. Subsequent CAD GWAS publications using UKB data alone or in a meta-analysis discovered over 150 novel CAD variants. For example, van der Harst and Verweij utilized the UKB to perform a GWAS in the UKB which identified 64 novel CAD loci. Likewise, Tcheandjieu *et al.* discovered 95 novel CAD loci through a GWAS using the Million Veteran Program (MVP) and a meta-GWAS which included the UKB[3]. To date, at least 321 genome-wide independent significant loci for CAD have been identified[87].

These CAD loci constitute potential causal variants and genes for CAD, and thus provide potential drug targets for CAD management and treatment[88-90]. In 2003, scientists identified gain-of-function mutations in the PCSK9 gene which led to extremely high LDL-C levels[91]. In contrast, further research from Cohen *et al.* highlighted the protective effect of loss-of-function variants in the PCSK9 gene that disrupt its function; effectively, individuals carrying these variants have lower LDL-C levels and, thus, a lower risk of CAD diagnosis[92]. Ultimately, this observation led to the creation of two monoclonal antibodies that mimic the protective variants by inhibiting the PCSK9 protein in blood[93,94]. Indeed, randomized clinical trials for these two drugs demonstrated that they significantly lowered LDL-C levels and reduced the risk of cardiovascular events[93,94]. This provided robust proof of concept that genetics can contribute to

the understanding and development of therapeutics. Additional studies have investigated the clinical translation and druggability of other CAD loci[87,95].

1.2.2 Genetic Risk Scores

On their own, the discovered common variants associated with CAD typically have small effect sizes[87,96,97], limiting their individual use as predictors of CAD risk. However, identified variants can be integrated into a genetic risk score (GRS) in which the summation shows a stronger effect. A GRS can be calculated for any disease or disorder as long as that phenotype has sufficient GWAS summary statistics and relevant available data. GRS are based on 3 key information from GWAS: SNP IDs, risk alleles and effect sizes (odds ratio (OR)) (Figure 3). The traditional way of calculating a GRS (also known as a polygenic risk score or a genome-wide risk score) is by summing the risk alleles multiplied by their effect size (also known as a weighted sum)[98]. As GRS are based on germline variants, they are not age-dependent and can be calculated at birth[99]. Importantly, a CAD GRS captures and quantifies only a part of an individual's genetic susceptibility and predisposition to CAD.

Recently, there has been a rise in direct-to-consumer genetic testing (i.e., at-home DNA kits). Valued at 1.56 billion USD in 2022, this market has capitalized on GRS, which companies can perform using a DNA sample sent in by consumers sometimes in conjunction with ancestry testing. Indeed, companies such as 23andMe and Color Genomics are offering a wide array of commercialized GRS ranging from breast and hereditary prostate cancer to heart disease and type 2 diabetes [100]. While some companies clearly state that their commercialized GRS are "for people who are of mainly European ancestry"[101], 23andMe claim they

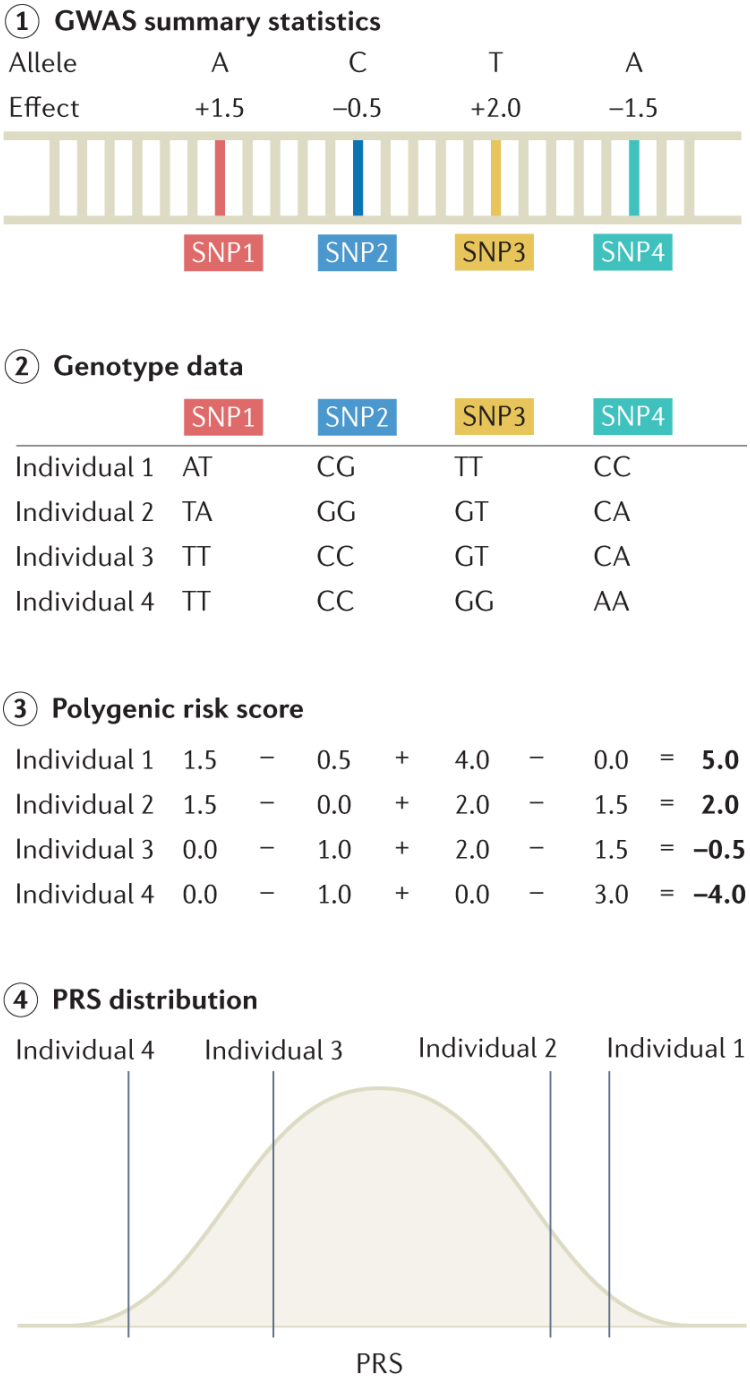


Figure 3: Steps to calculate a GRS

Four essential steps required to calculate a GRS and assess its distribution. PRS, polygenic risk score. (Uffelmann E, Huang Q, Munung N, et al. Genome-wide association studies. *Nature Reviews Methods Primers*. 2021;1(1):1-21) (Reproduced with permission of the publisher)

"can adjust and test our computer models on people from many backgrounds to make sure they also work well for non-Europeans"[102] despite recent research demonstrating a clear difficulty of GRS transferability between European and non-European ancestries[103-105]. Although many concerns persist concerning this business, especially surrounding the validity of a client's interpretation of the GRS and the variability in the GRS methods used, direct-to-consumer GRS appear to be here to stay. To be certain that customers benefit the most from access to commercialized GRS, it will be crucial for further research to focus on GRS construction methods, possible interpretation, and the role of physicians[100,101,106].

Based on summary results data from GWAS, dozens of GRS construction methods have been developed and optimized including pruning and thresholding (P+T) and LDpred2[107-109]. The former creates clusters of SNPs in high LD that also meet a p-value threshold and then selects the most significant variant from each cluster[110]. The latter is a more recent Bayesian technique which incorporates GWAS summary statistics and a linkage disequilibrium (LD) matrix[107,111]. A recent study by Wang *et al.* noted that a CAD LDpred2-based GRS had a stronger predictive ability for CAD than others built using more traditional methods like P+T[112]. One main difference between the two methods that could partially explain this result lies in the number of SNPs included in the GRS[107,108]. As the LDpred2 method does not filter out any SNPs, the GRS can contain hundreds of thousands or more than a million SNPs whereas a P+T-based GRS usually contains a few hundred SNPs.

Many CAD GRS studies have been performed based on previously identified genome-wide significant variants. Researchers such as Tada *et al.* and Inouye *et al.* have emphasized that individuals with a high GRS had a significantly higher risk of CAD than those with a low GRS (2.4-fold and 4.17-fold risk, respectively)[10,113], highlighting the reliability of GRS as a predictive

tool. Importantly, Mega *et al.* concluded that individuals with a high genetic risk for CAD (top 20% of GRS) displayed the greatest risk reductions of CAD events with the use of statins[11]. Ference *et al.* observed a 3-fold reduction in risk of CAD in individuals with long-term exposure to low LDL-C due to genetic variants compared to LDL-C lowering medications later in life[114]. This result suggests that targeted therapeutic interventions earlier in life can have a significant impact on CAD risk. A possible weakness in CAD GRS published to date is that the overwhelming majority are exclusively autosomal and do not include variants from the X and Y chromosome. Only one CAD GRS that included X chromosome (X-chr) variants has been published[3]. This study concluded that their trans-ethnic GRS which also included X-chr variants outperformed other existing autosomal GRS for risk prediction.

Another important characteristic of GRS is that they are independent of family history. Tada *et al.* observed that the significant association between a CAD GRS and CAD events did not vary according to self-reported family history[113]. Similarly, Hindieh *et al.* reported a similar trend with their 30-SNP CAD GRS while highlighting that common variants are unlikely to give rise to a family history of CAD[115]. Thus, GRS have the potential to be an effective additional tool for risk assessment and primary prevention.

1.3 Hypothesis and Objectives

In recent years, a few studies have shown that CAD GRS are significantly associated with some traditional CAD risk factors (e.g., hypertension, diabetes, age and more). In 2019, a CAD GRS containing 300 polymorphisms was significantly associated with hypertension, type 2 diabetes, and hypercholesterolemia in the UKB[8]. Additionally, Inouye *et al.* developed a "metaGRS" consisting of 1.7 million genetic variants and analyzed its association with many risk factors in the UKB[10]. They observed that the "metaGRS" was significantly associated with BMI,

diabetes, hypertension, smoking, high cholesterol and family history of heart disease. This last result is notable because family history and a CAD GRS are usually viewed as independent[113,115,116], and it could be partially explained by the increased power of the UKB.

Recently, a small number of studies have demonstrated that two main risk factors for CAD, age and sex, can interact with a CAD GRS. Using a 161-loci CAD GRS, Huang *et al.* identified a significant interaction between the GRS and sex and a CAD GRS with over a million variants was able to detect an interaction with sex in the UKB[12,13]. Moreover, Marston *et al.* recently illustrated a significant interaction between a 241-variant CAD GRS and age for incident cases in the UKB[117]. Further, in 2023, a group led by Cristen Willer used the "metaGRS" developed by Inouye *et al.* to investigate CAD GRS interaction with age and sex[118]. They confirmed a significant interaction with age and sex in both HUNT2, a Norwegian-based data bank, and the UKB. However, only a few studies have investigated interactions between CAD GRS SNPs and other atherosclerotic risk factors including diabetes, hypertension, dyslipidemia, smoking, and BMI[9,15]. While one study from Cole *et al.* reported pleiotropy between BMI and CAD[14], little research has been done to directly explore the possible pleiotropic effects between CAD and atherosclerotic risk factors.

As part of this thesis, I hypothesized that a CAD GRS would significantly interact with various atherosclerotic risk factors. I also hypothesized that a CAD GRS will not have the same strength in different atherosclerotic risk factor subgroups. I tested these through the three following objectives:

1.3.1 Objective 1: Determine if the known association between a CAD GRS and CAD interacts with specific risk factors.

I performed cox proportional hazard analyses to test for interactions between a CAD GRS and atherosclerotic risk factors (age, sex, dyslipidemia, hypertension, diabetes, BMI, smoking behavior, Lp(a) levels).

1.3.2 Objective 2: Determine if the known association between a CAD GRS and CAD varies in risk factor specific subsets of the GRS.

I performed cox proportional hazard analyses for each subset of the CAD GRS based on atherosclerotic risk factors in individuals of European ancestry in UKB. I then compared these results to each other and to the complete CAD GRS.

1.3.3 Objective 3: Determine if GRS construction methods influence the association between GRS and CAD and interactions between the GRS and atherosclerotic risk factors.

I contrasted two different construction methods, pruning and thresholding and LDpred2, by performing generalized linear models to test for cross-sectional association between each CAD GRS and CAD and interactions between each CAD GRS and atherosclerotic risk factors.

Chapter 2: Methods

2.1 Study Design and Participants

The UKB recruited approximately 502,000 British participants aged between 38 and 73 years from the general population, through one of 22 assessment centers, between 2006 and 2010. Participants completed a standardized questionnaire and provided blood samples, as previously described[86]. Additionally, the UKB sample data was linked with data from the UK's National Health Service which allowed diagnoses to be identified in participant's medical records[119]. UKB received ethics approval from the Northwest Multi-Centre Research Ethics committee. All participants gave written informed consent. I excluded participants with missing genetic sex information or genetic sex that differed from the self-reported as well as samples that were identified as outliers for heterozygosity and missing rates. Participants taking cholesterol-lowering medication at baseline were also excluded. Among 1st degree relatives, only one (selected randomly) from each family was kept. This yielded 344,130 individuals of European ancestry, 5,207 individuals of South Asian ancestry, 6,104 of African ancestry and 1,288 Chinese ancestry individuals for this study. Individuals of European ancestry include 321,403 White British individuals, 9,533 White Irish individuals and 13,194 White Other individuals (Ethnic terms defined by the UKB).

2.2 GRS and Subdivisions

I used a weighted CAD GRS composed of 204 autosomal SNPs (denoted as GRS_{204}) as previously described[9]. All included SNPs had an imputation quality score > 0.3. The P + T method, which involves creating clusters of SNPs in high LD, that also meet a p-value threshold

and selecting the most significant variant from each cluster[110], was used to compile this GRS and linkage disequilibrium in Europeans between any pair of SNPs was $r^2 < 0.2$[9].

The GRS_{204} was partitioned according to significant associations in the UKB of included variants with Lp(a), apoB, DM or HTN. SNPs with a nominally significant p value ($p < 0.05$) for a given risk factor were included in the risk-factor-specific GRS. Thus, the risk factor-specific GRS contained 41, 90, 66 and 121 SNPs for Lp(a), apoB, DM and HTN, respectively. The GRS without these risk factor SNPs contained 163, 114, 138 and 83 SNPs for Lp(a), apoB, DM and HTN, respectively (Table 1). This subdivision method was validated by comparing each GRS subset pair ($GRS_{rf(+)}$ and $GRS_{rf(-)}$) for its association with risk factors (Table 2).

While my work with the GRS_{204} was in progress, the Million Veteran Program (MVP) CAD GWAS was published in 2022[3]. Specifically, it identified 95 novel CAD loci including nine X chromosome loci. Notably, 33 novel loci were identified in a European-ancestry GWAS while 62 novel loci were identified in a multi-ancestry GWAS. I created a weighted CAD GRS with established and novel SNPs from that study (denoted as GRS_{MVP}). The GRS_{MVP} contained 258 SNPs (249 autosomal SNPs and nine X chromosome SNPs), and the effect sizes for my GRS calculations are from Tcheandjieu et al.'s summary statistics.

In additional sensitivity analysis, I used LDpred2 (grid model) to build a CAD GRS based on the 2015 CARDIoGRAMplusC4D GWAS summary statistics (denoted as $GRS_{LDpred2}$)[97,107]. This model allows tuning of two hyperparameters: SNP heritability (h^2) and the proportion of causal variants (p)[107]. As the LDpred2-grid model requires testing and validation

cohorts, the UKB European-ancestry cohort was split into two cohorts. The LD correlation matrix was computed in the testing cohort among 1,316,447 SNPs. In total, 102 grid models were generated from *p*, h^2 and sparsity combinations (17 *p* values, 3 h^2 values and presence/absence of sparsity). I used the GRS model determined by the best AUC among the 102 grid models for the association with CAD, adjusting for age and sex. Statistical analyses were performed in the validation samples with the best LDpred2-grid model and 556 552 SNPs.

2.3 Outcome Definitions

Briefly, I included myocardial infarction, acute and chronic ischemic heart disease, coronary artery disease and replacement of a coronary artery, but not angina. Specifically, CAD in UKB was defined as the presence of one or more of the following ICD9, ICD10 and OPCS4 codes: 410, 411, 412, 413, 414, I21, I22, I23, I24, I25, K40, K41, K42, K43, K44, K45, K46 and K49.

Table 1. SNP classification according to GRS204 subdivisions

Locus#	Locus	CHR	Position (hg19)	rsID	Classification
1	MORN1	1	2252205	rs36096196	b, d, e, h
2	PRDM16	1	3325912	rs2493298	b, d, e, h
3	FHL3	1	38461319	rs61776719	b, d, e, h
4	PCSK9	1	55496039	rs11206510	a, c, f, h
4	PCSK9	1	55505647	rs11591147	a, c, f, h
5	PLPP3(PPAP2B)	1	56966350	rs17114046	b, d, e, h
5	PLPP3(PPAP2B)	1	57016950	rs112470402	b, d, e, h
5	PLPP3(PPAP2B)	1	56986303	rs147055617	a, d, e, g
6	PSRC1(SORT1)	1	109821511	rs602633	a, c, f, h
7	NGF	1	115753482	rs11806316	b, d, f, h
8	TDRKH	1	151762308	rs11810571	a, c, e, h
9	IL6R	1	154422067	rs4845625	b, c, f, h
10	ATP1B1	1	169094459	rs1892094	b, d, e, g
11	DDX59,CAMSAP2	1	200646073	rs6700559	b, d, f, g
12	LMOD1	1	201872264	rs2820315	b, d, e, g
13	HHAT	1	210468999	rs60154123	b, d, e, h
14	MIA3	1	222823529	rs17465637	b, d, f, h
15	AGT	1	230845794	rs699	a, c, e, h
16	OSR1(AK097927)	2	19942473	rs16986953	a, d, e, h
17	APOB	2	21291529	rs668948	a, c, f, g
18	ABCG8,ABCG5	2	44081627	rs4076834	a, d, f, h
18	ABCG8,ABCG5	2	44073881	rs6544713	a, d, f, h
19	PRKCE	2	45896437	rs582384	b, d, e, h
20	VAMP8,VAMP5	2	85809989	rs1561198	a, d, f, h
21	ZEB2,TEX41	2	145801461	rs2252641	b, d, e, h
21	ZEB2,TEX41	2	145270592	rs6740731	b, d, f, h
21	ZEB2,TEX41	2	145286559	rs17678683	b, d, f, h

Legend:

a: apoB (+) subset

b: apoB (-) subset

c: Lpa (+) subset

d: Lpa (-) subset

e: HTN (+) subset

f: HTN (-) subset

g: DM (+) subset

h: DM (-) subset

22	FIGN	2	164957251	rs12999907	a, d, e, h
23	CALCRL	2	188196469	rs840616	b, d, e, h
24	WDR12,NBEAL1	2	203893999	rs115654617	a, d, f, g
25	FN1	2	216304384	rs1250229	a, d, e, h
25	FN1	2	216291359	rs17517928	a, d, e, h
26	TNS1	2	218683154	rs2571445	a, d, e, h
26	TNS1	2	218669225	rs61741262	a, d, e, h
27	LOC646736	2	227100698	rs2972146	a, d, e, g
28	KCNJ13,GIGYF2	2	233633460	rs1801251	b, d, f, g
29	COL6A3	2	238223955	rs11677932	b, d, f, h
30	FGD5	3	14901525	rs13079221	b, d, e, g
31	SNORD77,ALS2CL	3	46688562	rs7633770	a, d, e, g
32	CDC25A	3	48193515	rs7617773	b, d, e, h
33	RHOA	3	49448566	rs7623687	b, d, e, h
34	UMPS,ITGB5	3	124450081	rs4678145	b, d, f, h
35	DNAJC13	3	132257961	rs10512861	a, d, f, h
36	STAG1	3	136069472	rs667920	a, c, e, g
37	MRAS	3	138092889	rs185244	b, d, e, g
38	ARHGEF26	3	153839866	rs12493885	a, d, e, g
39	CCNL1	3	156852592	rs4266144	b, d, e, h
40	FNDC3B	3	172115902	rs12897	b, d, f, h
41	HGFAC,RGS12	4	3449652	rs16844401	a, d, e, h
42	REST,NOA1	4	57838583	rs17087335	a, d, f, h
43	SHROOM3	4	77416627	rs12500824	a, d, e, h
44	FGF5	4	81181072	rs10857147	a, d, e, h
45	HNRNPD	4	82587050	rs11099493	a, d, f, h
46	UNC5C	4	96117371	rs3775058	b, c, e, g
47	MAD2L1	4	120909501	rs7678555	a, d, e, g
48	ZNF827	4	146782837	rs35879803	b, d, e, h
49	EDNRA	4	148281001	rs4593108	b, d, f, h
49	EDNRA	4	148400819	rs6842241	b, d, e, h

Legend:

a: apoB (+) subset

b: apoB (-) subset

c: Lpa (+) subset

d: Lpa (-) subset

e: HTN (+) subset

f: HTN (-) subset

g: DM (+) subset

h: DM (-) subset

50	GUCY1A3,MAP9	4	156635309	rs7692387	b, d, e, g
50	GUCY1A3,MAP9	4	156436517	rs13118820	a, d, e, g
51	PALLD	4	169687725	rs7696431	b, d, e, h
52	SEMA5A	5	9556694	rs1508798	b, d, f, h
53	LOC101928448	5	55860781	rs3936511	a, d, e, g
54	LOX	5	121413208	rs1800449	b, d, f, h
55	SLC22A4-SLC22A5	5	131667353	rs273909	a, d, e, h
56	ARHGAP26	5	142516897	rs246600	b, d, e, h
57	FOXC1	6	1617143	rs9501744	b, d, e, h
58	PHACTR1	6	12756658	rs1412748	b, d, f, h
58	PHACTR1	6	12903957	rs9349379	b, d, e, h
59	HDGFL1	6	22598259	rs7766436	a, d, f, h
60	C2	6	31919578	rs2072633	a, c, e, g
61	ANKS1A,C6orf16	6	34618893	rs2814993	a, c, e, h
61	ANKS1A,C6orf16	6	35034800	rs17609940	b, d, e, h
62	CDKN1A,PANDAR	6	36638636	rs1321309	b, d, e, h
63	KCNK5	6	39174922	rs10947789	b, d, e, h
64	VEGFA	6	43758873	rs6905288	a, d, e, g
65	PRIM2	6	57160572	rs9367716	b, d, f, h
66	RP11-379B8.1	6	82612271	rs4613862	a, d, f, h
67	CENPW	6	126717064	rs1591805	a, d, f, g
68	TCF21	6	134209837	rs2327429	b, d, e, h
68	TCF21	6	134214227	rs2327433	b, d, e, h
69	PLEKHG1	6	150997401	rs17080091	a, d, e, h
70	LPA,PLG,LPAL2,SLC22A3	6	160679400	rs624249	a, c, f, g
70	IGF2R	6	160465291	rs688359	b, c, e, h
70	LPA,PLG,LPAL2,SLC22A3	6	160863532	rs2048327	a, c, e, g
70	LPA,PLG,LPAL2,SLC22A3	6	161143608	rs4252120	b, c, f, g
70	LPA,PLG,LPAL2,SLC22A3	6	161056112	rs9365196	a, c, f, h
70	LPA,PLG,LPAL2,SLC22A3	6	161102643	rs9457995	a, c, e, h
70	LPA,PLG,LPAL2,SLC22A3	6	161005610	rs55730499	a, c, e, g

Legend:

a: apoB (+) subset

b: apoB (-) subset

c: Lpa (+) subset

d: Lpa (-) subset

e: HTN (+) subset

f: HTN (-) subset

g: DM (+) subset

h: DM (-) subset

70	LPA,PLG,LPAL2,SLC22A3	6	160911596	rs147555597	a, c, e, h
70	LPA,PLG,LPAL2,SLC22A3	6	161111700	rs186696265	a, c, e, g
71	MAD1L1	7	1937261	rs10267593	b, d, e, g
72	DAGLB*	7	6446027	rs10951983	b, d, f, h
73	TMEM106B	7	12261911	rs11509880	b, d, f, g
74	HDAC9	7	19049388	rs2107595	b, d, e, h
75	CCM2	7	45077978	rs2107732	b, d, e, g
76	7q22(BCAP29)	7	107244545	rs10953541	b, d, e, h
77	CFTR,CCTNBP2	7	117332914	rs975722	b, d, e, h
78	ZC3HC1	7	129663496	rs11556924	b, d, e, h
79	PARP12	7	139757136	rs10237377	b, d, e, h
80	NOS3	7	150690176	rs3918226	a, d, e, g
81	NAT2	8	18286997	rs6997340	a, d, e, g
82	LPL	8	19824667	rs15285	a, c, e, g
82	LPL	8	19800529	rs6997330	a, c, e, g
83	BMP1	8	22033615	rs6984210	b, d, f, g
84	ZFPM2	8	106565414	rs10093110	b, d, f, h
85	TRIB1	8	126490972	rs2954029	a, d, e, h
86	CDKN2B,CDKN2A	9	21706571	rs896655	b, d, f, h
86	CDKN2B,CDKN2A	9	22073996	rs1855185	b, d, f, g
86	CDKN2B,CDKN2A	9	22098619	rs2891168	a, d, e, g
86	CDKN2B,CDKN2A	9	21970916	rs3731249	b, c, f, h
86	CDKN2B,CDKN2A	9	22062012	rs4977754	b, c, e, h
86	CDKN2B,CDKN2A	9	22113324	rs13301964	a, d, f, g
87	KLF4	9	110517794	rs944172	a, d, f, h
88	SVEP1	9	113169775	rs111245230	b, d, e, g
89	DAB2IP	9	124420173	rs885150	b, d, f, h
90	ABO	9	136149399	rs507666	a, d, e, g
91	CDC123	10	12303813	rs61848342	b, d, e, g
92	KIAA1462	10	30317073	rs9337951	b, c, e, h
93	CXCL12	10	44777560	rs1657346	b, d, f, h

Legend:

a: apoB (+) subset

b: apoB (-) subset

c: Lpa (+) subset

d: Lpa (-) subset

e: HTN (+) subset

f: HTN (-) subset

g: DM (+) subset

h: DM (-) subset

93	CXCL12	10	44480811	rs1870634	b, d, f, h
94	TSPAN14	10	82251514	rs17680741	b, d, e, h
95	LIPA	10	91004886	rs2246942	a, c, f, h
96	AS3MT,CYP17A1,CNNM2	10	104638480	rs3740390	b, d, e, h
97	STN1	10	105693644	rs4918072	b, d, e, g
98	HTRA1	10	124237612	rs4752700	b, d, f, h
99	TRIM5,TRIM22	11	5701074	rs11601507	a, d, f, g
100	SWAP70	11	9751196	rs10840293	a, d, e, h
100	MRVI1,CTR9	11	10745394	rs11042937	b, c, e, h
101	ARNTL	11	13301548	rs1351525	b, d, e, g
102	HSD17B12	11	43696917	rs7116641	a, d, e, g
103	PCNX3	11	65391317	rs12801636	a, d, e, g
104	SERPINH1	11	75274150	rs590121	b, d, f, h
104	SERPINH1	11	75284334	rs659418	b, d, f, g
105	ARHGAP42	11	100624599	rs7947761	b, d, e, h
106	PDGFD,DYNC2H1	11	103660567	rs974819	b, d, f, h
107	APOA1-A5-A4-C3,ZNF259	11	116648917	rs964184	a, d, f, g
108	C1S	12	7175872	rs11838267	b, d, f, h
109	LOC156393	12	20220033	rs10841443	b, d, e, h
110	HOXC4	12	54513915	rs11170820	b, d, f, h
111	LRP1	12	57527283	rs11172113	a, d, f, h
112	ATP2B1	12	90013089	rs2681492	b, d, e, h
113	NDUFA12	12	95355541	rs7306455	b, d, e, h
114	SH2B3,ATXN2,HNF1A	12	111884608	rs3184504	a, d, e, g
115	KSR2	12	118265441	rs11830157	b, d, f, h
116	HNF1A	12	121416988	rs2244608	a, c, e, g
117	SCARB1,CCDC92	12	124427306	rs11057401	a, c, f, g
117	SCARB1,CCDC92	12	125307053	rs11057830	a, d, e, h
118	FLT1	13	28973621	rs9319428	b, d, f, h
119	N4BP2L2	13	33058333	rs9591012	b, d, e, g
120	COL4A1/A2	13	110960943	rs3809346	a, d, f, h

Legend:

a: apoB (+) subset

b: apoB (-) subset

c: Lpa (+) subset

d: Lpa (-) subset

e: HTN (+) subset

f: HTN (-) subset

g: DM (+) subset

h: DM (-) subset

120	COL4A1/A2	13	111049623	rs9515203	a, c, f, h
120	COL4A1/A2	13	110818102	rs11617955	b, d, f, g
120	COL4A1/A2	13	111040681	rs11838776	a, c, e, h
121	MCF2L	13	113631780	rs1317507	a, d, e, h
122	ARID4A	14	58794001	rs2145598	b, d, f, g
123	TMED10	14	75614504	rs3832966	a, d, e, h
124	SERPINA1,SERPINA2	14	94838142	rs112635299	b, d, e, g
125	HHIPL1,CYP46A1	14	100133942	rs2895811	a, d, e, h
125	HHIPL1,CYP46A1	14	100148961	rs8003602	b, d, f, h
126	OAZ2,RBPMS2	15	65024204	rs6494488	b, d, f, h
127	SMAD3	15	67450305	rs17228058	b, d, e, g
128	ADAMTS7	15	79017861	rs8039034	b, d, e, h
128	ADAMTS7	15	79139000	rs11637783	a, d, f, h
129	MFGE8-ABHD2	15	89574218	rs8042271	b, d, f, h
130	FURIN	15	91416550	rs17514846	a, d, e, h
131	LINC00924 (15q26.2)	15	96146414	rs17581137	b, d, f, g
132	CETP	16	56995236	rs1800775	a, c, e, h
133	DHX38,TXNL4B	16	72130815	rs1050362	a, c, e, h
134	CFDP1	16	75462055	rs12930452	a, d, e, h
135	PLCG2	16	81906423	rs7199941	b, c, f, h
136	CDH13	16	83045790	rs7500448	b, d, f, h
137	SMG6,SRR	17	2170216	rs170041	a, d, f, h
137	SMG6,SRR	17	2126504	rs216172	a, d, e, h
138	RASD1, SMCR3, PEMT	17	17543722	rs12936587	a, c, f, h
139	CORO6,ANKRD13B	17	27941886	rs13723	b, d, e, g
140	(17q11.2)	17	30033514	rs76954792	b, d, e, h
141	DHX58,KAT2A	17	40257163	rs2074158	a, d, f, h
142	GOSR2	17	45013271	rs17608766	b, d, e, h
143	UBE2Z,GIP	17	47047868	rs3895874	a, d, e, g
143	UBE2Z,ZNF652	17	47440466	rs16948048	b, d, e, h
144	BCAS3	17	59013488	rs7212798	b, d, e, h

Legend:

a: apoB (+) subset

b: apoB (-) subset

c: Lpa (+) subset

d: Lpa (-) subset

e: HTN (+) subset

f: HTN (-) subset

g: DM (+) subset

h: DM (-) subset

145	PECAM1	17	62387091	rs1867624	b, d, e, g
146	ACAA2	18	47229717	rs9964304	b, d, e, h
147	PMAIP1,MC4R	18	57838401	rs663129	b, d, f, g
148	ANGPTL4	19	8429323	rs116843064	a, c, e, g
149	LDLR	19	11277232	rs4804573	a, c, f, h
149	LDLR	19	11202306	rs6511720	a, c, e, h
150	MAP1S,FCHO1	19	17855763	rs73015714	b, d, f, h
151	ZNF507,LOC400684	19	32882020	rs12976411	b, d, f, h
152	TGFB1,CCDC97	19	41851509	rs4803455	b, d, e, h
152	TGFB1,CCDC97	19	41832231	rs12980942	b, d, f, g
152	TGFB1,CCDC97	19	41790086	rs138120077	a, d, e, h
153	APOE,APOC1,TOMM4	19	45412079	rs7412	a, c, e, h
153	SNRPD2	19	46190268	rs1964272	b, c, f, g
153	APOE,APOC1,TOMM4	19	45422946	rs4420638	a, d, f, g
154	PROCR	20	33764554	rs867186	b, d, f, h
154	NCOA6	20	33313566	rs6088590	b, d, e, g
155	ZHX3	20	39924279	rs6102343	a, d, f, h
156	PCIF1,ZNF335	20	44586023	rs3827066	a, d, f, h
157	ZNF831	20	57714025	rs260020	b, c, e, h
158	MAP3K7CL	21	30533076	rs2832227	b, d, f, g
159	MRPS6	21	35593827	rs28451064	a, d, e, h
160	POM121L9P,ADORA2A	22	24658858	rs180803	b, d, f, h

Legend:

a: apoB (+) subset

b: apoB (-) subset

c: Lpa (+) subset

d: Lpa (-) subset

e: HTN (+) subset

f: HTN (-) subset

g: DM (+) subset

h: DM (-) subset

Each SNP was classified to different GRS subsets depending on its association with each risk factor (apoB, Lpa, HTN, DM).

GRS indicates genetic risk score; CHR, chromosome; CAD, coronary artery disease; apoB, apolipoprotein B; HTN, hypertension; DM, diabetes mellitus; Lp(a), lipoprotein (a).

Table 2: Association of CAD GRS subsets with Lp(a), apoB, DM and HTN in UKB individuals of European ancestry

		Lp(a)		apoB		HTN	
	SNPs	Effect size (95%CI)	P value	Effect size (95%CI)	P value	Adj. OR (95%CI)	P value
GRS_{204}	204	16.07 (15.88, 16.27)	$<2.00 \times 10^{-16}$	0.031 (0.030, 0.032)	$4.38 \times E10^{-9}$	1.07 (1.07, 1.08)	$<2.00 \times 10^{-16}$
$GRS_{apoB(-)}$	114	1.12 (0.92, 1.32)	2.79×10^{-14}	**0.002 (0.001, 0.003)**	$2.03 \times E10^{-8}$	1.05 (1.05, 1.06)	$<2.00 \times 10^{-16}$
$GRS_{apoB(+)}$	90	20.29 (20.10, 20.48)	$<2.00 \times 10^{-16}$	**0.039 (0.038, 0.040)**	$<2.00 \times 10^{-16}$	1.05 (1.04, 1.06)	$<2.00 \times 10^{-16}$
$GRS_{HTN(-)}$	83	2.84 (2.64, 3.04)	5.28×10^{-8}	0.024 (0.023, 0.025)	$<2.00 \times 10^{-16}$	**0.99 (0.98, 1.00)**	4.72×10^{-3}
$GRS_{HTN(+)}$	121	18.53 (18.34, 18.72)	$<2.00 \times 10^{-16}$	0.022 (0.022, 0.023)	$<2.00 \times 10^{-16}$	**1.10 (1.10, 1.11)**	$<2.00 \times 10^{-16}$
$GRS_{Lpa(-)}$	163	**-0.07 (-0.27, 0.13)**	0.50	0.01 (0.01, 0.01)	$<2.00 \times 10^{-16}$	1.08 (1.07, 1.09)	$<2.00 \times 10^{-16}$
$GRS_{Lpa(+)}$	41	**29.36 (29.19, 29.54)**	$<2.00 \times 10^{-16}$	0.042 (0.041, 0.042)	$<2.00 \times 10^{-16}$	1.01 (1.01, 1.02)	2.14×10^{-4}
$GRS_{DM(-)}$	138	4.54 (4.34, 4.74)	$<2.00 \times 10^{-16}$	0.025 (0.025, 0.026)	$<2.00 \times 10^{-16}$	1.06 (1.05, 1.07)	$<2.00 \times 10^{-16}$
$GRS_{DM(+)}$	66	20.32 (20.13, 20.51)	$<2.00 \times 10^{-16}$	0.019 (0.018, 0.020)	$<2.00 \times 10^{-16}$	1.04 (1.04, 1.05)	$<2.00 \times 10^{-16}$

		DM	
	SNPs	Adj. OR (95%CI)	P value
GRS_{204}	204	1.02 (1.01, 1.04)	6.00×10^{-3}
$GRS_{apoB(-)}$	114	1.04 (1.02, 1.06)	3.91×10^{-6}
$GRS_{apoB(+)}$	90	1.00 (0.98, 1.02)	0.87
$GRS_{HTN(-)}$	83	0.99 (0.97, 1.01)	0.30
$GRS_{HTN(+)}$	121	1.04 (1.02, 1.05)	1.83×10^{-5}
$GRS_{Lpa(-)}$	163	1.03 (1.01, 1.05)	2.81×10^{-4}
$GRS_{Lpa(+)}$	41	1.00 (0.98, 1.01)	0.62
$GRS_{DM(-)}$	138	**0.98 (0.96, 1.00)**	1.70×10^{-2}
$GRS_{DM(+)}$	66	**1.06 (1.05, 1.08)**	1.35×10^{-13}

Cox proportional hazard analyses were performed for GRS associations with CAD, apoB, HTN and DM. All analyses are age and sex adjusted. Each GRS is weighted per standard deviation.

GRS indicates genetic risk score; HR per SD, hazard ratio per standard deviation; CI, confidence interval; P, p value; SNPs, single nucleotide polymorphisms; CAD, coronary artery disease; apoB, apolipoprotein B; HTN, hypertension; DM, diabetes mellitus.

Thresholds used to define risk factors were age ≥55 years old for older age, BMI ≥30 for obesity, high Lp(a) was considered ≥100 nmol/L, and dyslipidemia was defined as ≥1.3 g/L apoB. Diabetes mellitus was defined as the presence of any of the following: (i) use of diabetes medication or (ii) a diagnosis of diabetes mellitus or (iii) HbA1c levels ≥6.5%. Hypertension was defined as the presence of any of the following: (i) use of blood pressure medication or (ii) a diagnosis of hypertension or (iii) a systolic blood pressure level ≥140 mm Hg or a diastolic blood pressure level ≥90 mm Hg. I constructed a risk factor score for each individual based on the presence or absence of four risk factors: BMI ≥30, hypertension, diabetes mellitus, and current smoker. Thus, the score was an integer from 0 to 4.

2.4 Statistical Analysis

In the primary analysis, the CAD GRS_{204} was tested for association with incident CAD with Cox proportional hazard models adjusted for age and sex. To evaluate interaction, models included a multiplicative interaction term (GRS x risk factor) for each CAD risk factor (age, sex, dyslipidemia, hypertension, diabetes, BMI, smoking behavior, Lp(a) levels). In addition, stratified analysis for each risk factor was performed. The different GRS construction methods (GRS_{204}, GRS_{MVP} and $GRS_{LDpred2}$) were compared using generalized linear models testing the association with incident CAD adjusted for age and sex. All results are presented as HR per SD or OR per SD with 95% confidence intervals. A p-value < 0.05 was considered statistically significant. All analyses were performed using PLINK 2.0[120] and R studio version 4.2.2.

Chapter 3: Results

A total of 344,130 European ancestry individuals with complete data were included in the study. Characteristics of the study population are presented in Table 3. The median age was 57 years with an interquartile range [IQR] of [49, 62] and 145,042 (42.3%) were male. Following enrolment in UKB (baseline), 16,118 incident CAD cases occurred during a median follow-up of 10.95 years [10.06, 11.68].

3.1 Interaction Analyses

The GRS_{204} was strongly associated with CAD in individuals of European ancestry (Hazard Ratio (HR) 1.37 (95% CI, 1.35, 1.40), $P < 2.00 \times 10^{-16}$) (Table 4). The GRS_{204} also had a significant positive interaction in men ($P = 1.09 \times 10^{-4}$) and a significant negative interaction with increasing age ($P = 3.63 \times 10^{-8}$) (Figure 4). The GRS also had a significant positive interaction with dyslipidemia ($P = 4.11 \times 10^{-03}$), on incident CAD, but a negative interaction with diabetes ($P = 7.28 \times 10^{-7}$) (Table 5).

3.2 Subset Analyses

Because these interactions could be due to SNP subsets associated with specific atherosclerotic risk factors, I investigated interactions with GRS subsets. All risk factor GRS subsets showed a significant interaction with age (P values $< 8.00 \times 10^{-3}$) and, with the exception of the $GRS_{DM(-)}$. All GRS subsets demonstrated a significant interaction with sex (Table 6). Five CAD GRS subsets had significant interactions with diabetes: the $GRS_{apoB(-)}$ ($P = 0.012$), the $GRS_{apoB(+)}$ ($P = 0.013$), the $GRS_{HTN(+)}$ ($P = 1.45 \times 10\text{-}4$), the $GRS_{DM(+)}$ ($P = 2.03 \times 10\text{-}3$) and the $GRS_{Lpa(-)}$ ($P = 1.45 \times 10\text{-}4$) (Table 6). Further, five CAD GRS subsets showed a significant interaction with dyslipidemia: $GRS_{apoB(-)}$ ($P = 6.97 \times 10^{-3}$), $GRS_{HTN(+)}$ ($P = 8.18 \times 10^{-3}$) , $GRS_{Lpa(-)}$

($P = 1.66 \times 10^{-3}$), $GRS_{DM(+)}$ ($P = 3.60 \times 10^{-2}$) and $GRS_{DM(-)}$ ($P = 0.041$) (Table 6). Only the $GRS_{HTN(-)}$ demonstrated a significant interaction with hypertension ($P = 0.03$) (Table 6).

Consistent with the GRS_{204}, there were no significant interactions between the CAD GRS subsets and BMI, smoking, or Lp(a) (Table 6).

3.3 Stratified Analyses

In stratified analyses of each risk factor (Figure 4), males and those with dyslipidemia had higher HR (men, HR 1.40, (95% CI, 1.38, 1.43); women, 1.32 (1.29, 1.36); individuals with dyslipidemia, 1.40 (1.36, 1.45); those without, 1.34 (1.32, 1.37); all $P < 2x10^{-16}$). Age and diabetes had higher HRs for those without the risk factor (age <55, 1.47 (1.43, 1.52); age ≥55, 1.33 (1.31, 1.36); diabetics, 1.26 (1.21, 1.32); non-diabetics, 1.39 (1.37, 1.41); all $P < 2x10^{-16}$). Consistent directions of effect were also observed in the stratified analysis of GRS subsets (Table 7).

3.4 Other Genetic Ancestries

The GRS_{204} was also predictive of CAD in individuals of South Asian ancestry (HR 1.31, (95% CI, 1.19, 1.45), $P = 9.75 \times 10^{-08}$) (Tables 8-9) with an effect size consistent with the European ancestry sample. However, the GRS_{204} had no significant interactions with any of the risk factors in individuals of South Asian ancestry (Table 9). The GRS_{204} was not significant in individuals of African and Chinese ancestries (Table 10).

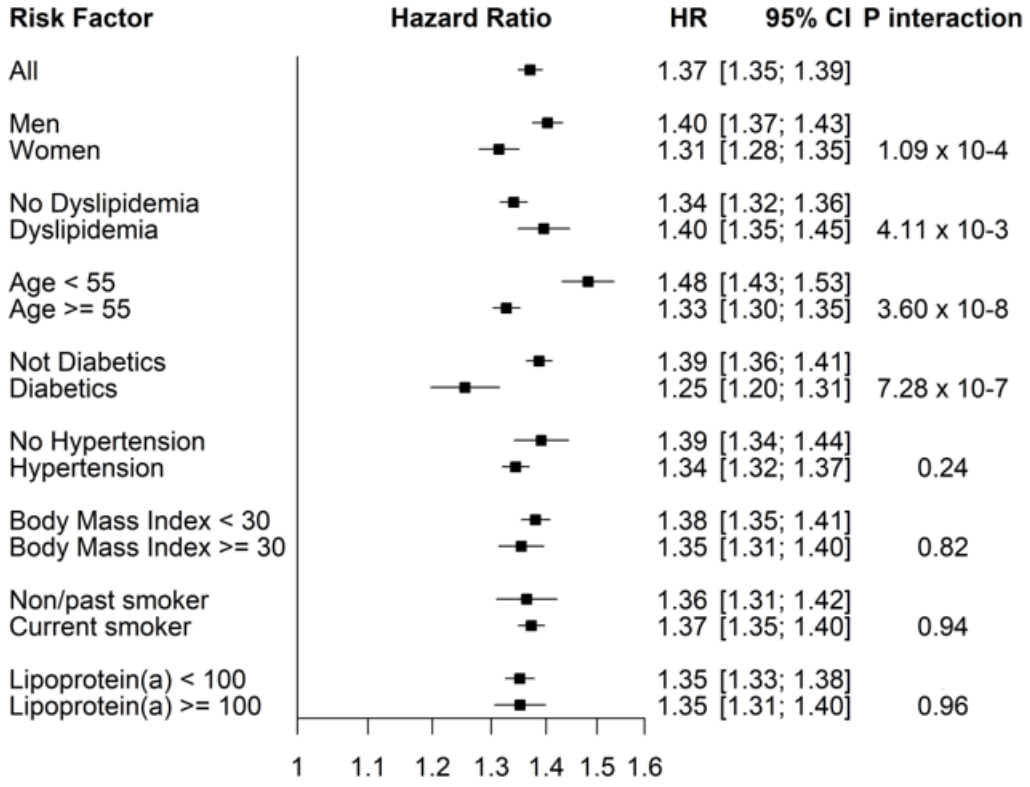
Table 3: Characteristics of European participants with genetic data from UK Biobank

Characteristic	Participants
N	344,130
Male	145,042 (42.3)
BP medication	45,362 (13.2)
Diabetes medication	2,605 (0.8)
Current smoker	35,123 (10.2)
Diabetes mellitus	13,311 (3.9)
Hypertension	170,520 (49.7)
Incident CAD cases	16,118 (4.7)
Cross-sectional CAD cases	19,336 (5.6)
Median follow-up (years)	10.95 [10.06, 11.68]
Age	57.00 [49.00, 62.00]
BMI	26.30 [23.81, 29.33]
Systolic BP (mm Hg)	135.50 [124.00, 148.50]
Diastolic BP (mm Hg)	82.00 [75.00, 89.00]
Apolipoprotein B (g/L)	1.05 [0.91, 1.21]
Total cholesterol (mmol/L)	5.86 [5.19, 6.58]
Triglycerides (mmol/L)	1.44 [1.02, 2.09]
HDL cholesterol (nmol/L)	1.44 [1.21, 1.72]
LDL cholesterol (nmol/L)	3.68 [3.16, 4.24]
Lipoprotein(a) (nmol/L)	19.90 [9.35, 58.80]
Non-HDL cholesterol (mmol/L)	4.37 [3.72, 5.08]

Data are n (%) or median [interquartile range].

BP indicates blood pressure; BMI, body mass index; CAD, coronary artery disease; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

Figure 4: Stratified analysis of CAD GRS_{204} with CAD by risk factors in the European ancestry individuals of the UK Biobank



Cox proportional hazard analyses were performed stratified by the presence or absence of risk factors among individuals of European ancestry in UKB. Age stratified analyses were adjusted for sex. Sex stratified analyses were adjusted for age. All other risk factor stratified analyses were age and sex adjusted.

GRS indicates genetic risk score; HR, hazard ratio per standard deviation; CI, confidence interval; CAD, coronary artery disease.

Table 4: Associations of CAD GRS subsets with incident CAD in UKB individuals of European ancestry

		CAD	
GRS model	SNPs	HR (95%CI)	P value
GRS_{204}	204	1.37 (1.35, 1.40)	$<2.00 \times 10^{-16}$
$GRS_{apoB(-)}$	114	1.24 (1.22, 1.26)	$<2.00 \times 10^{-16}$
$GRS_{apoB(+)}$	90	1.27 (1.26, 1.29)	$<2.00 \times 10^{-16}$
$GRS_{HTN(-)}$	83	1.20 (1.18, 1.22)	$<2.00 \times 10^{-16}$
$GRS_{HTN(+)}$	121	1.31 (1.30, 1.34)	$<2.00 \times 10^{-16}$
$GRS_{Lpa(-)}$	163	1.31 (1.29, 1.33)	$<2.00 \times 10^{-16}$
$GRS_{Lpa(+)}$	41	1.19 (1.17, 1.20)	$<2.00 \times 10^{-16}$
$GRS_{DM(-)}$	138	1.28 (1.26, 1.30)	$<2.00 \times 10^{-16}$
$GRS_{DM(+)}$	66	1.22 (1.20, 1.24)	$<2.00 \times 10^{-16}$

Cox proportional hazard analyses were performed for GRS association with CAD. All analyses were age and sex adjusted.

GRS indicates genetic risk score; HR, hazard ratio per standard deviation; CI, confidence interval; P, p value; SNPs, single nucleotide polymorphisms; CAD, coronary artery disease.

Table 5: Risk factor interactions with the GRS204 for incident CAD in UKB individuals of European ancestry

	CAD			
Risk factor	HR (95%CI)	P	HR_{int} (95%CI)	P_{int}
Age	1.41 (1.38, 1.43)	$<2.00 \times 10^{-16}$	0.95 (0.94, 0.97)	3.63×10^{-8}
Sex	1.32 (1.29, 1.35)	$<2.00 \times 10^{-16}$	1.07 (1.03, 1.10)	1.09×10^{-4}
Diabetes	1.39 (1.37, 1.41)	$<2.00 \times 10^{-16}$	0.92 (0.88, 0.97)	7.28×10^{-7}
Hypertension	1.38 (1.34, 1.43)	$<2.00 \times 10^{-16}$	0.98 (0.94, 1.02)	0.24
Obesity	1.38 (1.35, 1.40)	$<2.00 \times 10^{-16}$	1.00 (0.96, 1.03)	0.82
Current smoker	1.38 (1.35, 1.40)	$<2.00 \times 10^{-16}$	1.00 (0.96, 1.04)	0.94
Lp(a)	1.36 (1.33, 1.38)	$<2.00 \times 10^{-16}$	1.00 (0.96, 1.04)	0.96
Dyslipidemia	1.34 (1.32, 1.36)	$<2.00 \times 10^{-16}$	1.05 (1.01, 1.09)	4.11×10^{-3}
Risk factor score	1.36 (1.33, 1.40)	$<2.00 \times 10^{-16}$	0.98 (0.96, 1.00)	1.70×10^{-2}

Cox proportional hazard analyses with risk factor interactions. Risk factor interactions were adjusted for age and sex except age interaction was adjusted only for sex and sex interaction was adjusted only for age. Risk factor score had no covariates in the model.

GRS indicates genetic risk score; Lp(a), lipoprotein (a); HR, hazard ratio per standard deviation; CI, confidence interval; P, p value; CAD, coronary artery disease; HR_{int}, hazard ratio per standard deviation of the interaction; P_{int}, p value of the interaction.

Table 6: Interactions of CAD GRS subsets with risk factors on incident CAD

A		Age interaction		Sex interaction		Diabetes interaction		Current smoker interaction	
GRS model	SNPs	HR_{int} (95%CI)	P_{int}	HR_{int} (95%CI)	P_{int}	HR_{int} (95%CI)	P_{int}	HR_{int} (95%CI)	P_{int}
GRS_{204}	204	0.95 (0.94, 0.97)	3.63×10^{-8}	1.07 (1.03, 1.10)	1.09×10^{-4}	0.92 (0.88, 0.97)	7.28×10^{-4}	1.00 (0.96, 1.04)	0.49
$GRS_{apoB(-)}$	114	0.96 (0.95, 0.98)	2.51×10^{-5}	1.04 (1.00, 1.07)	4.80×10^{-2}	0.94 (0.90, 0.99)	0.012	0.98 (0.94, 1.05)	0.33
$GRS_{apoB(+)}$	90	0.96 (0.95, 0.98)	3.27×10^{-5}	1.06 (1.02, 1.09)	8.44×10^{-4}	0.94 (0.90, 0.99)	0.013	1.02 (0.97, 1.06)	0.47
$GRS_{HTN(-)}$	83	0.97 (0.95, 0.98)	9.23×10^{-5}	1.04 (1.00, 1.07)	4.20×10^{-2}	0.98 (0.93, 1.02)	0.34	0.98 (0.94, 1.02)	0.39
$GRS_{HTN(+)}$	121	0.96 (0.95, 0.98)	1.24×10^{-5}	1.06 (1.02, 1.10)	6.84×10^{-4}	0.91 (0.87, 0.96)	1.45×10^{-4}	1.02 (0.98, 1.06)	0.42
$GRS_{Lpa(-)}$	163	0.95 (0.94, 0.97)	1.61×10^{-7}	1.05 (1.01, 1.08)	8.16×10^{-3}	0.93 (0.89, 0.97)	2.14×10^{-3}	0.98 (0.94, 1.02)	0.42
$GRS_{Lpa(+)}$	41	0.98 (0.96, 0.99)	8.00×10^{-3}	1.05 (1.01, 1.08)	7.15×10^{-3}	0.96 (0.92, 1.01)	0.12	1.03 (0.99, 1.07)	0.19
$GRS_{DM(-)}$	138	0.97 (0.95, 0.98)	1.84×10^{-4}	1.03 (0.99, 1.06)	0.10	0.96 (0.91, 1.00)	0.056	0.98 (0.94, 1.02)	0.25
$GRS_{DM(+)}$	66	0.96 (0.94, 0.98)	2.68×10^{-6}	1.07 (1.03, 1.10)	7.70×10^{-5}	0.93 (0.89, 0.97)	2.03×10^{-3}	1.03 (0.99, 1.08)	0.11

B		Dyslipidemia interaction		Hypertension interaction		BMI interaction		Lp(a) interaction	
GRS model	SNPs	HR_{int} (95%CI)	P_{int}	HR_{int} (95%CI)	P_{int}	HR_{int} (95%CI)	P_{int}	HR_{int} (95%CI)	P_{int}
GRS_{204}	204	1.06 (1.02, 1.10)	4.11×10^{-3}	0.98 (0.94, 1.02)	0.24	1.00 (0.96, 1.03)	0.82	1.00 (0.96, 1.04)	0.96
$GRS_{apoB(-)}$	114	1.05 (1.01, 1.09)	6.97×10^{-3}	0.98 (0.94, 1.02)	0.24	1.00 (0.96, 1.03)	0.84	1.03 (1.00, 1.07)	7.40×10^{-2}
$GRS_{apoB(+)}$	90	1.03 (0.99, 1.07)	0.13	0.99 (0.95, 1.03)	0.52	1.00 (0.97, 1.03)	0.91	0.98 (0.94, 1.01)	0.19
$GRS_{HTN(-)}$	83	1.02 (0.98, 1.06)	0.28	0.96 (0.92, 0.99)	3.00×10^{-2}	1.02 (0.98, 1.05)	0.37	1.02 (0.98, 1.06)	0.30
$GRS_{HTN(+)}$	121	1.05 (1.01, 1.09)	8.18×10^{-3}	1.00 (0.96, 1.04)	0.93	0.98 (0.95, 1.01)	0.26	0.99 (0.95, 1.02)	0.49
$GRS_{Lpa(-)}$	163	1.06 (1.02, 1.10)	1.66×10^{-3}	0.98 (0.94, 1.02)	0.43	1.01 (0.98, 1.04)	0.60	1.03 (0.99, 1.07)	0.17
$GRS_{Lpa(+)}$	41	1.01 (0.97, 1.04)	0.72	0.98 (0.94, 1.02)	0.34	0.98 (0.95, 1.02)	0.31	0.97 (0.93, 1.01)	0.10
$GRS_{DM(-)}$	138	1.04 (1.00, 1.08)	3.60×10^{-2}	0.99 (0.96, 1.03)	0.75	1.01 (0.97, 1.04)	0.67	1.02 (0.98, 1.06)	0.37
$GRS_{DM(+)}$	66	1.04 (1.00, 1.08)	4.10×10^{-2}	0.97 (0.94, 1.01)	0.13	0.98 (0.95, 1.02)	0.35	0.97 (0.94, 1.01)	0.13

Cox proportional hazard analyses were performed including risk factor interactions among individuals of European ancestry in the UKB. Age interaction was adjusted for sex. Sex interaction was adjusted for age. All other risk factor interactions were age and sex adjusted. Each GRS is weighted per standard deviation. Panel A: interactions for age, sex, dyslipidemia and current smoker; Panel B: diabetes, hypertension, BMI and Lp(a).

GRS indicates genetic risk score; Lp(a), lipoprotein (a); BMI, body mass index; CI, confidence interval; P, p value; SNPs, single nucleotide polymorphisms; CAD, coronary artery disease; HR_{int}, hazard ratio per standard deviation of the interaction; P_{int}, p value of the interaction.

Table 7: Associations of CAD GRS subsets with incident CAD in risk factor related subsets of UKB

A		Age <55	Age ≥ 55	Men	Women	Hypertension (+)	Hypertension (-)
GRS model	SNPs	HR (95%CI)	HR (95%CI)	HR (95%CI)	HR (95%CI)	HR (95%CI)	HR (95%CI)
GRS_{204}	204	1.47 (1.43, 1.52)	1.33 (1.31, 1.36)	1.40 (1.38, 1.43)	1.32 (1.29, 1.36)	1.35 (1.33, 1.37)	1.39 (1.35, 1.44)
$GRS_{apoB(-)}$	114	1.29 (1.25, 1.34)	1.22 (1.20, 1.24)	1.25 (1.23, 1.28)	1.21 (1.18, 1.24)	1.22 (1.20, 1.24)	1.25 (1.21, 1.30)
$GRS_{apoB(+)}$	90	1.35 (1.31, 1.40)	1.24 (1.22, 1.26)	1.30 (1.27, 1.32)	1.23 (1.20, 1.27)	1.26 (1.24, 1.28)	1.28 (1.24, 1.33)
$GRS_{HTN(-)}$	83	1.26 (1.22, 1.30)	1.18 (1.16, 1.20)	1.21 (1.19, 1.24)	1.18 (1.15, 1.21)	1.19 (1.17, 1.21)	1.25 (1.21, 1.29)
$GRS_{HTN(+)}$	121	1.39 (1.35, 1.44)	1.28 (1.26, 1.30)	1.34 (1.31, 1.37)	1.27 (1.24, 1.31)	1.29 (1.27, 1.31)	1.30 (1.26, 1.34)
$GRS_{Lpa(-)}$	163	1.39 (1.34, 1.43)	1.28 (1.25, 1.30)	1.33 (1.30, 1.35)	1.27 (1.24, 1.31)	1.28 (1.26, 1.30)	1.31 (1.27, 1.36)
$GRS_{Lpa(+)}$	41	1.24 (1.20, 1.28)	1.16 (1.14, 1.18)	1.20 (1.18, 1.23)	1.16 (1.13, 1.18)	1.18 (1.16, 1.20)	1.21 (1.17, 1.25)
$GRS_{DM(-)}$	138	1.35 (1.30, 1.40)	1.26 (1.24, 1.28)	1.29 (1.27, 1.32)	1.26 (1.23, 1.30)	1.27 (1.25, 1.29)	1.29 (1.24, 1.33)
$GRS_{DM(+)}$	66	1.30 (1.26, 1.34)	1.20 (1.17, 1.22)	1.25 (1.23, 1.27)	1.18 (1.15, 1.21)	1.21 (1.19, 1.23)	1.25 (1.21, 1.29)

B		Non obese	Obese	Dyslipidemia (+)	Dyslipidemia (-)	Current smoker	Former/never smoked
GRS model	SNPs	HR (95%CI)	HR (95%CI)	HR (95%CI)	HR (95%CI)	HR (95%CI)	HR (95%CI)
GRS_{204}	204	1.38 (1.36, 1.41)	1.37 (1.33, 1.41)	1.40 (1.36, 1.45)	1.34 (1.32, 1.37)	1.36 (1.31, 1.41)	1.38 (1.36, 1.40)
$GRS_{apoB(-)}$	114	1.24 (1.22, 1.26)	1.23 (1.20, 1.27)	1.28 (1.24, 1.32)	1.22 (1.20, 1.25)	1.21 (1.16, 1.26)	1.24 (1.22, 1.26)
$GRS_{apoB(+)}$	90	1.28 (1.26, 1.30)	1.27 (1.24, 1.31)	1.27 (1.23, 1.31)	1.25 (1.23, 1.27)	1.28 (1.24, 1.33)	1.27 (1.25, 1.30)
$GRS_{HTN(-)}$	83	1.20 (1.17, 1.22)	1.21 (1.18, 1.25)	1.20 (1.16, 1.24)	1.18 (1.16, 1.20)	1.17 (1.13, 1.22)	1.20 (1.18, 1.23)
$GRS_{HTN(+)}$	121	1.33 (1.30, 1.35)	1.29 (1.26, 1.33)	1.34 (1.30, 1.39)	1.29 (1.27, 1.31)	1.32 (1.27, 1.38)	1.31 (1.29, 1.34)
$GRS_{Lpa(-)}$	163	1.31 (1.28, 1.33)	1.31 (1.28, 1.35)	1.35 (1.31, 1.40)	1.29 (1.26, 1.31)	1.28 (1.23, 1.33)	1.31 (1.29, 1.35)
$GRS_{Lpa(+)}$	41	1.19 (1.17, 1.22)	1.17 (1.14, 1.20)	1.16 (1.13, 1.20)	1.16 (1.14, 1.18)	1.21 (1.16, 1.25)	1.18 (1.16, 1.20)
$GRS_{DM(-)}$	138	1.29 (1.26, 1.31)	1.29 (1.25, 1.33)	1.30 (1.26, 1.35)	1.26 (1.24, 1.28)	1.25 (1.20, 1.30)	1.29 (1.27, 1.31)
$GRS_{DM(+)}$	66	1.23 (1.21, 1.25)	1.21 (1.17, 1.24)	1.24 (1.20, 1.28)	1.20 (1.18, 1.22)	1.25 (1.20, 1.30)	1.22 (1.20, 1.24)

C		Diabetics (+)		Diabetics (-)	Lp(a) < 100	Lp(a) ≥100	
GRS model	SNPs	HR (95%CI)	P	HR (95%CI)	HR (95%CI)	HR (95%CI)	P
GRS_{204}	204	1.26 (1.21, 1.32)	$<2.00 \times 10^{-16}$	1.39 (1.37, 1.41)	1.36 (1.33, 1.38)	1.36 (1.31, 1.40)	$<2.00 \times 10^{-16}$
$GRS_{apoB(-)}$	114	1.16 (1.11, 1.21)	2.80×10^{-11}	1.25 (1.23, 1.27)	1.23 (1.21, 1.25)	1.27 (1.23, 1.31)	$<2.00 \times 10^{-16}$
$GRS_{apoB(+)}$	90	1.20 (1.15, 1.25)	$<5.67 \times 10^{-16}$	1.29 (1.27, 1.31)	1.26 (1.24, 1.28)	1.23 (1.19, 1.27)	$<2.00 \times 10^{-16}$
$GRS_{HTN(-)}$	83	1.17 (1.12, 1.22)	1.24×10^{-11}	1.21 (1.19, 1.23)	1.19 (1.17, 1.21)	1.21 (1.17, 1.25)	$<2.00 \times 10^{-16}$
$GRS_{HTN(+)}$	121	1.20 (1.15, 1.25)	$<2.00 \times 10^{-16}$	1.33 (1.31, 1.35)	1.30 (1.28, 1.32)	1.28 (1.24, 1.32)	$<2.00 \times 10^{-16}$
$GRS_{Lpa(-)}$	163	1.21 (1.16, 1.27)	$<2.00 \times 10^{-16}$	1.32 (1.30, 1.34)	1.30 (1.28, 1.32)	1.34 (1.29, 1.38)	$<2.00 \times 10^{-16}$
$GRS_{Lpa(+)}$	41	1.14 (1.09, 1.18)	5.35×10^{-9}	1.19 (1.17, 1.21)	1.16 (1.14, 1.19)	1.13 (1.09, 1.16)	5.46×10^{-14}
$GRS_{DM(-)}$	138	1.22 (1.17, 1.28)	$<2.00 \times 10^{-16}$	1.30 (1.28, 1.32)	1.27 (1.25, 1.29)	1.29 (1.25, 1.33)	$<2.00 \times 10^{-16}$
$GRS_{DM(+)}$	66	1.13 (1.08, 1.18)	2.13×10^{-8}	1.23 (1.21, 1.25)	1.20 (1.18, 1.23)	1.17 (1.13, 1.21)	$<2.00 \times 10^{-16}$

Cox proportional hazard analyses for different risk factors among individuals of European ancestry in UKB. All stratified analyses with no p value indicated have a p value $< 2.00 \times 10^{-16}$. Age stratified analyses were adjusted for sex. Sex stratified analyses were adjusted for age. All other risk factor stratified analyses were age and sex adjusted. Each GRS is weighted per standard deviation. Panel A indicates stratified analyses for age, sex and hypertension; Panel B for obesity, dyslipidemia and current smoker; Panel C for diabetics and Lp(a).

GRS indicates genetic risk score; Lp(a), lipoprotein (a); HR, hazard ratio per standard deviation; CI, confidence interval; SNPs, single nucleotide polymorphisms; CAD, coronary artery disease; P, p value.

Table 8: Characteristics of South Asian ancestry participants with genetic data from UK Biobank

Characteristic	n (%) or median (interquartile range [IQR])
N	5,207
Male	2,581 (49.6)
BP medication	751 (14.4)
Diabetes medication	250 (4.8)
Current smoker	465 (8.9)
Diabetes mellitus	724 (13.9)
Hypertension	2,562 (49.2)
Incident CAD	399 (7.7)
Median follow-up (years)	10.61 [10.10, 11.28]
Age	51.00 [45.00, 58.00]
BMI	26.36 [24.00, 29.21]
Systolic BP (mm Hg)	132.00 [121.00, 144.50]
Diastolic BP (mm Hg)	82.50 [76.00, 89.50]
Apolipoprotein B (g/L)	1.04 [0.90, 1.18]
Total cholesterol (mmol/L)	5.52 [4.91, 6.17]
Triglycerides (mmol/L)	1.66 [1.17, 2.41]
HDL cholesterol (nmol/L)	1.23 [1.05, 1.46]
LDL cholesterol (nmol/L)	3.52 [3.05, 4.02]
Lipoprotein(a) (nmol/L)	29.20 [12.00, 66.40]
Non-HDL cholesterol (mmol/L)	4.26 [3.66, 4.89]

BP indicates blood pressure; BMI, body mass index; CAD, coronary artery disease; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

Table 9: Interaction of CAD GRS_{204} with risk factors on incident CAD in the UK Biobank South Asians

	CAD			
Risk factor	HR (95%CI)	P	HR_{int} (95%CI)	P_{int}
Age	1.31 (1.18, 1.46)	7.68×10^{-7}	1.00 (0.91, 1.10)	0.97
Sex	1.29 (1.07, 1.55)	6.23×10^{-3}	1.02 (0.82, 1.27)	0.85
Diabetes	1.30 (1.14, 1.48)	6.38×10^{-5}	1.05 (0.86, 1.29)	0.63
Hypertension	1.47 (1.17, 1.84)	9.82×10^{-4}	0.85 (0.66, 1.10)	0.22
Obesity	1.31 (1.17, 1.46)	4.11×10^{-6}	1.02 (0.81, 1.28)	0.88
Current smoker	1.33 (1.20, 1.48)	1.42×10^{-7}	0.87 (0.65, 1.17)	0.37
Lp(a)	1.32 (1.19, 1.48)	4.57×10^{-7}	0.93 (0.71, 1.20)	0.56
Dyslipidemia	1.28 (1.15 (1.43)	4.99×10^{-6}	1.09 (0.82, 1.46)	0.55
Risk factor score	1.29 (1.08, 1.54)	5.75×10^{-3}	0.98 (0.89, 1.08)	0.75

Cox proportional hazard analyses were performed with the different risk factor interactions among individuals of South Asian ancestry in UKB. Age interaction was adjusted for sex. Sex interaction was adjusted for age. All other risk factor interactions except risk factor score were age and sex adjusted. Risk factor score had no covariates in the model. Each GRS is weighted per standard deviation. GRS_{204} interactions with risk factors were not performed in individuals of African and Chinese ancestries as the GRS_{204} was not significant in those ancestries.

GRS indicates genetic risk score; Lp(a), lipoprotein (a); HR, hazard ratio per standard deviation; CI, confidence interval; CAD, coronary artery disease; P, p value; HR_{int}, hazard ratio per standard deviation of the interaction; P_{int}, p value of the interaction.

Table 10: Association of GRS204 with incident CAD in UK Biobank ancestries

	CAD	
Ancestry	HR (95%CI)	P value
European	1.37 (1.35, 1.40)	$<2.00 \times 10^{-16}$
South Asian	1.31 (1.19, 1.45)	9.75×10^{-8}
Black	1.02 (0.88, 1.19)	0.76
Chinese	0.90 (0.57, 1.42)	0.66

Cox proportional hazard analyses were performed for CAD among individuals of different ancestries in UKB. All analyses are age and sex adjusted.

GRS indicates genetic risk score; HR, hazard ratio per standard deviation; CI, confidence interval; CAD, coronary artery disease.

3.5 GRS Comparison

In a sensitivity analysis, I compared the GRS_{204} results with two other GRS. The GRS_{MVP} and the $GRS_{LDpred2}$ were also strongly associated with CAD in European ancestry individuals (Table 11) (OR 1.38, (95% CI, 1.36, 1.41), 1.38 (1.35, 1.42), all p values $P < 2.00 \times 10^{-16}$, respectively). Furthermore, all three GRS had significant interactions with age, sex, and dyslipidemia in Europeans and the GRS_{204} and $GRS_{LDpred2}$ also had a significant interaction with diabetes. Additionally, I replicated six out of the nine X-chr SNPs from Tcheandjieu *et al.*'s study in men of European ancestry in UKB (Figure A1). In women of European ancestry in UKB, I replicate two out of the nine X-chr SNPs (Figure A2). To further quantify the addition of the nine X-chr SNPS to an autosomal GRS, I compared the GRS_{204} with two versions of the GRS_{MVP} – one with the nine X-chr SNPs and one without the X-chr SNPs. I concluded that the GRS_{204} and the $GRS_{MVP(with\ X\ chr)}$ were both strongly associated with CAD in European ancestry individuals with a similar effect size whereas the $GRS_{MVP(no\ X\ chr)}$ had a slightly lower effect size (HR 1.37, (95% CI, 1.35, 1.40), 1.35 (1.33, 1.38), 1.37 (1.34, 1.39), all p values $P < 2.00 \times 10^{-16}$, respectively) (Table B2). Moreover, I contrasted both the $GRS_{MVP(with\ X\ chr)}$ and the $GRS_{MVP(no\ X\ chr)}$ in individuals of non-European ancestry in the UKB. I observed that both of these GRS were able to replicate the GRS association in South Asian (Table B3). However, both GRS were not significant among individuals of African and Chinese ancestries.

Table 11: Comparison of different GRS methods (GRSLDpred2 vs GRS204 vs GRSMVP) in European subset of UK Biobank for various associations and interactions with CAD

Method	GRS_{204}	GRS_{MVP}	$GRS_{LDpred2}$
Number of SNPs	204	258	556 552
GRS	1.39 (1.37, 1.41) $P < 2.00 \times 10^{-16}$	1.38 (1.36, 1.41) $P < 2.00 \times 10^{-16}$	1.38 (1.35, 1.42) $P < 2.00 \times 10^{-16}$
GRS*age	0.99 (0.99, 1.00) $P = 3.63 \times E10^{-08}$	0.97 (0.95, 0.99) $P = 4.00 \times E10^{-04}$	0.94 (0.92, 0.97) $P = 8.84 \times E10^{-06}$
GRS*sex	1.08 (1.05, 1.12) $P = 5.11 \times E10^{-06}$	1.09 (1.06, 1.13) $P = 3.55 \times E10^{-07}$	1.07 (1.02, 1.12) $P = 3.75 \times E10^{-03}$
GRS*HTN	0.99 (0.95, 1.03) P = 0.51	1.00 (0.96, 1.04) P = 0.88	0.99 (0.94, 1.05) P = 0.83
GRS*DM	0.94 (0.89, 0.99) $P = 1.95 \times E10^{-02}$	0.96 (0.91, 1.01) P = 0.09	0.92 (0.85, 0.99) $P = 1.86 \times E10^{-02}$
GRS*current smoker	0.98 (0.94, 1.01) P = 0.10	1.00 (0.96, 1.04) P = 0.98	0.98 (0.92, 1.04) P = 0.53
GRS*dyslipidemia	1.06 (1.02, 1.11) $P = 3.83 \times E10^{-03}$	1.07 (1.03, 1.11) $P = 1.14 \times E10^{-03}$	1.08 (1.02, 1.15) $P = 5.50 \times E10^{-03}$
GRS*BMI	0.99 (0.95, 1.03) P = 0.61	1.01 (0.97, 1.04) P = 0.66	0.96 (0.91, 1.00) P = 0.11
GRS*lpa	1.00 (0.97, 1.04) P = 0.87	1.02 (0.96, 1.07) P = 0.56	1.01 (0.97, 1.05) P = 0.78

Generalized linear models were performed with the different CAD GRS among individuals of European ancestry in UKB. Age interaction was adjusted for sex. Sex interaction was adjusted for age. All other risk factor interactions were age and sex adjusted. All results are presented as OR per SD (95%CI) with their respective p value. The GRS_{204} and the GRS_{MVP} analyses were conducted in the European cohort of UKB. The $GRS_{LDpred2}$ was conducted in the validation cohort of UKB. LDpred2 uses maximum AUC.

GRS indicates genetic risk score; HTN, hypertension; DM, diabetes mellitus; BMI, body mass index; Lp(a), lipoprotein (a); OR per SD, odds ratio per standard deviation; CI, confidence interval; P, p value; AUC, area under the curve; SNPs, single nucleotide polymorphisms; CAD, coronary artery disease.

Chapter 4: Discussion

4.1 Thesis Overview

In this thesis I sought to elucidate the impact of risk factors on the strength of a CAD GRS. In addition, I examined the impact of GRS subdivisions and GRS construction methods on CAD associations. I used Cox proportional hazard analyses to investigate CAD GRS interactions with atherosclerotic risk factors (i.e., age, sex, diabetes, dyslipidemia, hypertension, current smoking, obesity, Lp(a)) in individuals of European ancestry in the UKB. I found that subdividing a CAD GRS according to the possible role of the SNPs in specific atherosclerotic risk factors (i.e., hypertension, diabetes, Lp(a), apoB) attenuated the strength of the GRS associations with CAD but they remained significant. To determine if a similar pattern could be distinguished with CAD GRS interactions, I specifically compared interaction results between the GRS_{204} and each GRS subset for each risk factor. I found that for age, sex, diabetes, and dyslipidemia interactions, most CAD GRS subsets followed the same significant direction of effect as the GRS_{204}. In stratified analyses for each risk factor for GRS interactions. I determined that sex and dyslipidemia both had higher HRs for those with the risk factor (i.e., men and dyslipidemia) whereas age and diabetes had higher HRs for those without the risk factor. Consistent direction of effects was also identified in the stratified analysis of each risk factor for every sub-GRS. I also compared the GRS association with CAD and interactions with risk factors across different GRS construction methods. I found that all three GRS were strongly associated with CAD in European-ancestry individuals independent of the GRS construction method. Furthermore, consistent significant interactions with age, sex, diabetes, and dyslipidemia were observed among individuals of European ancestry.

4.2 Interactions

My data confirmed significant interactions between the GRS204 with sex, age and dyslipidemia and identified a novel interaction with diabetes.

My study observed that the CAD GRS_{204} has a stronger effect in men than women. In addition, the effect of the GRS_{204} and each sub-GRS, except the $GRS_{DM(-)}$, was significantly different between the sexes. My results are consistent with prior studies including one that demonstrated that a GRS composed of 161 variants had a stronger association with incident CAD in men (HR 1.38, (95% CI, 1.34, 1.41)) than in women (1.25 (1.21, 1.30))[13]. My results are also in line with those of Manikpurage *et al.* which demonstrated that a CAD GRS constructed with the LDpred software identified a significantly higher risk for CAD among men (HR 1.62, (95% CI, 1.59, 1.64)) compared to women (1.45 (1.42, 1.48)) in the UKB[12]. Although Manikpurage *et al.* reported stronger associations than I report, this is most likely due to their inclusion of prevalent and incident CAD cases (for a total of 32,694 CAD cases) whereas I included only incident cases (for a total of 23,752 CAD cases). Although some studies have proposed explanations for these sex differences, such as age differences in CAD diagnosis between men and women[13] or the greater proportion of men in the GWAS studies[3,4,121], it remains unclear why men have an increased genetic susceptibility to CAD based on a CAD GRS. Regardless of the source, the difference will need to be taken into consideration for the eventual clinical application of a CAD GRS.

In addition, my findings also demonstrated that the $GRS_{204,}$ as well as every sub-GRS, had a stronger effect size in younger individuals than older individuals. This is consistent with previously observed significant variation in the risk associated with a CAD GRS between individuals above and below the age of 57.6 years old[113]. My results are also in line with those of

Manikpurage *et al.* which demonstrated that a CAD GRS had a stronger association for individuals between 40 and 51 years old than in individuals between the ages of 63 and 73 (HR (95% CI), 1.89 (1.77, 2.02) and 1.48 (1.42, 1.53), respectively)[12]. Similarly, Marston *et al.* recently illustrated a significant interaction between a 241-variant CAD GRS and age in the UKB ($p < 0.001$)[117]. These findings of an age interaction have also been extended to individuals with diabetes. Lithovius *et al.* reported that the strength of a 158-variant CAD GRS differed significantly in diabetic individuals above and below a median age of 38.6 years. The CAD GRS had better risk discrimination in the younger age-group than the older age-group (C-index 0.637 (95% CI, 0.580, 0.695) and C-index 0.546 (95% CI, 0.516, 0.577), respectively)[122]. The stronger CAD GRS association at earlier ages is somewhat expected for genetic exposures and is consistent with other diseases as well[123-125].

My study also demonstrated an interaction of the GRS_{204} with dyslipidemia (defined as high plasma apoB levels). With the exception of $GRS_{LPA(+)}$, every sub-GRS also had a higher HR in individuals with dyslipidemia than in individuals without dyslipidemia. In a similar vein, previous work by Bolli *et al.* concluded that a CAD GRS had a stronger effect among individuals with a high LDL-C (4.71 (2.23, 9.94) for ≥190 mg/dL; 3.14 (1.52, 6.50) for 160-<190 mg/dL; 2.23 (1.08, 4.59) for 130-<160 mg/dL and 1.15 (0.54, 2.46) for 100-<130 mg/dL)[126]. Moreover, my finding is also consistent with the observations that a CAD GRS can identify individuals who demonstrate the greatest relative risk reduction with statin therapy[11,127]. Thus, a targeted approach consisting of apoB plasma levels in conjunction with a CAD GRS could lead to earlier interventions and risk reduction for individuals with dyslipidemia who also have a higher genetic predisposition for CAD, as quantified by the CAD GRS.

Our important novel result revealed that the GRS_{204} had a significantly weaker association with CAD in diabetes patients. A similar trend was also seen in every GRS risk factor subset. These results suggest that in those individuals with the prominent cardiovascular risk factor, diabetes, their CAD GRS may be less predictive. Consistent with my work, it has been shown that a CAD GRS is independent of traditional risk factors among type 2 diabetics patients but does not add to predictive performance[128]. Interestingly, Lee et al. observed an increased risk from a CAD PRS when type 2 diabetes was diagnosed at an earlier age[129]. In addition, a significantly larger effect of the GRS was observed in diabetic individuals with a higher HbA1c level[9]. Importantly, recent work points to the existence of novel genetic contributors to CAD among diabetics and more accurate CAD prediction in diabetics could come from the inclusion of such variants[130]. I would also not expect all SNPs in the GRS_{204} to have the same strength in diabetics. Future research should explore specific loci to identify those with a differing effect in diabetics.

4.3 GRS Construction Methods

My data suggest that two different GRS construction methods (pruning and thresholding and LDpred2) as well as the addition of chromosome X SNPs to an autosomal GRS do not substantially modify the strength of the association with CAD nor do they alter the significance of the observed atherosclerotic risk factor interactions with CAD GRS. Each GRS construction method demonstrated significant interactions with the same four risk factors, as previously discussed above.

My findings surrounding the two construction methods contradict those from some existing studies about LDpred2 efficiency[107,112,131]. For instance, previous work focusing on the different predictive ability of 15 GRS construction methods found that LDpred2 had the best

predictive power for CAD among individuals of European descent in the UKB[112]. Similarly, another study by Prive *et al.* noted that LDpred2 outperforms many other methods such as pruning and thresholding and lassosum for CAD in the UKB[107]. A few factors that might explain these differing results are different CAD definitions and differing statistical tools chosen to compare the methods (general linear models versus area under the receiver operating curve). In addition, fewer studies have used the LDpred2 method as it was only published and made publicly available three years ago[107]. Additional studies conducted in other large-scale cohorts such as the Million Veteran Program could further shed light on the distinctions and ideal parameters to optimize and select GRS methods.

While interest in X-chr SNPs has grown in the past decade, resulting in more publications[132], this still has not resulted in substantial X-chr-inclusive GWAS and GRS publications. To date, there has been only one published study (Tcheandjieu *et al.*) that has incorporated X-chr variants in a CAD GRS[3] and therefore limited opportunity to compare my results to previous studies. A notable difference between my results and those of Tcheandjieu *et al.* is that their GRS including the X-chr outperformed their autosomal GRS whereas my two GRS showed similar prediction abilities. However, in their publication, other parameters including a multi-population cohort and GRS construction techniques were also changed when the X-chr variants were used. Interestingly, the GRS based on the MVP GWAS results from my analyses and the GRS from Tcheandjieu *et al.*'s paper were strongly associated with CAD with similar effect size (1.37 (1.34, 1.39); 1.35 (1.31, 1.38), respectively) despite their different construction methods. Indeed, my GRS_{MVP} was created with the pruning and thresholding technique whereas Tcheandjieu *et al.*'s GRS utilised the PRSice2 method.

In addition, while the inclusion of X-chr SNPs to the $GRS_{MVP(no\ X)}$ demonstrated a stronger predictive ability for CAD, the $GRS_{MVP(X\ chr)}$ didn't outperform the GRS_{204}. It suggests that in certain cases the addition of X-chr variants can be beneficial for risk prediction. As the study led by Tcheandjieu didn't specifically look at the differential impact of adding X-chr SNPs to an autosomal GRS, these findings cannot be compared to another study. Furthermore, adding X-chr SNPs to an autosomal GRS didn't affect the CAD GRS associations in individuals of South African ancestry (i.e., both $GRS_{MVP(X\ chr)}$ and $GRS_{MVP(no\ X)}$ had significant associations with CAD with similar effect sizes). Additional research which will only be possible when more CAD GWAS including X-chr variants are published will be able to explore this potential benefit.

The lack of studies on CAD GRS including the X-chr variants can be explained by two factors. First, only 25% of all GWAS have X-chr data, of which only a small fraction pertains to CAD[132]. This severely limits the pool of CAD summary statistics to use as the basis for GRS construction. Second, not all GRS construction approaches can currently accommodate the inclusion of X-chr SNPs. For instance, the LDpred2 method is frequently used with an LD matrix provided by Prive *et al.*[107] which does not currently include any X-chr SNPs.

4.4 GRS Ancestry

My findings confirmed a significant association between GRS_{204} and CAD in individuals of South Asian and European ancestries while no significant association was identified in individuals of African and Chinese ancestries. The successful replication of a European CAD GRS in individuals of South Asian ancestry has been observed in other CAD studies as well[103,104,116,133]. These prior studies replicated their GRS association in individuals of South

Asian ancestry with a strong but attenuated effect size when compared to European results[104]. One study from Joseph *et al.* demonstrated that a 25-SNP CAD GRS, using just 25 SNPs, was significantly associated with myocardial infarction in both South Asian and European ancestries[116]. Notably, individuals of South Asian ancestry are known to be more closely related to individuals of European ancestry than individuals of African and Chinese ancestries[105,134], which may partially explain my results (Figure 5). These replication trends in non-European ancestries are consistent with those from other diseases as well[135-138]. Moreover, the lack of replication of European results among individuals of African and Chinese ancestries is most likely due to the small sample sizes[139-141] and specifically the low number of CAD cases in those subsets [10].

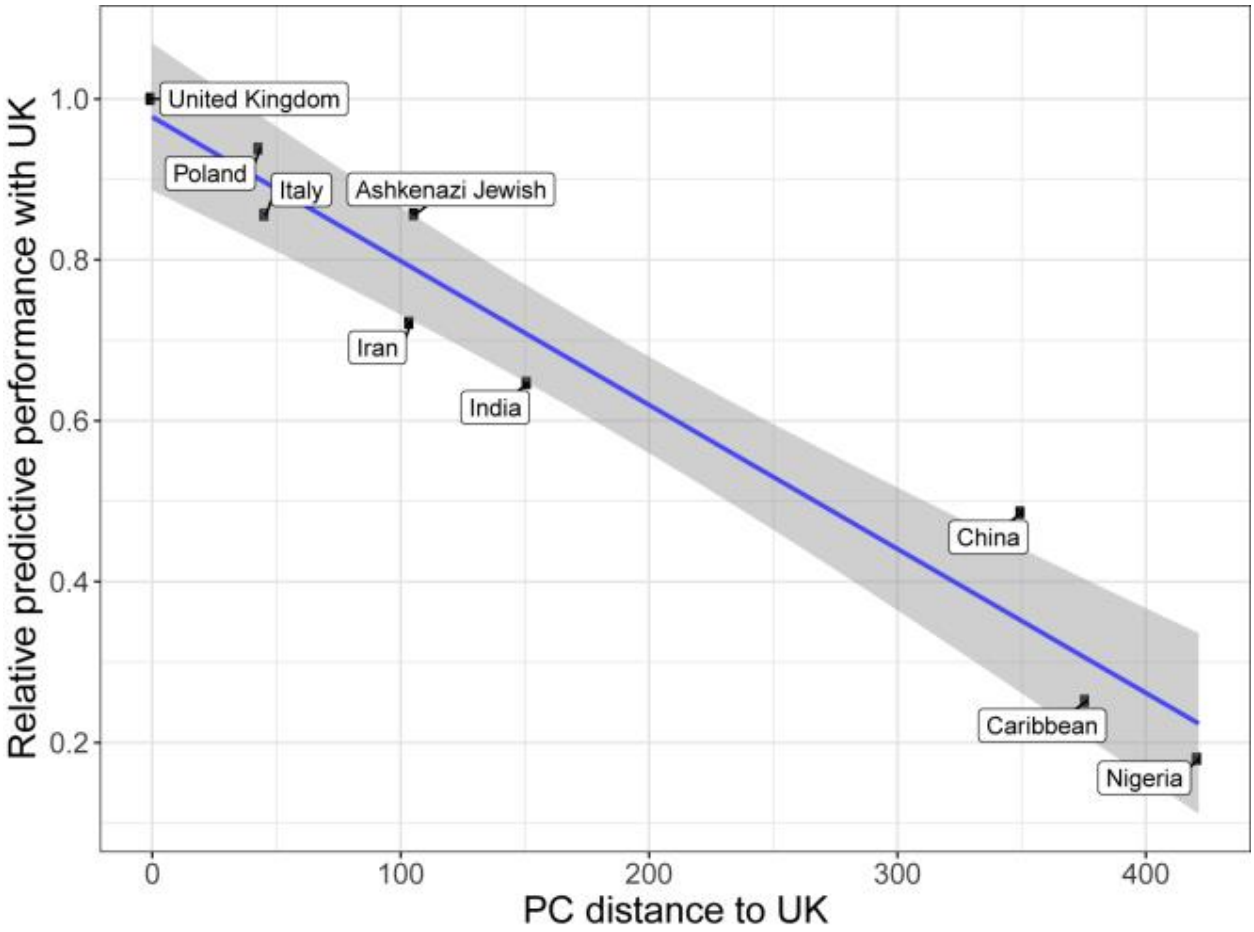


Figure 5: Comparison of relative variance and its association with PC distance

The relative predictive performance of a GRS in each ancestry is compared with each other while also factoring in the PC distance with the UK. UK, United Kingdom; PC, principal component. (Prive F, Aschard H, Carmi S, et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *Am J Hum Genet*. Feb 3 2022;109(2):373) (Reproduced with permission of the publisher)

4.5 Implementation of a CAD GRS in Clinical Settings

My results highlight interactions as an essential aspect to consider for any clinical use of a CAD GRS in the future. Specifically, the results emphasize the stronger impact of a CAD GRS at younger ages, a relevant concept for many diseases as genetics has a stronger role than environment at that age. Importantly, younger individuals are also less likely to have developed atherosclerotic risk factors which contribute directly to disease but also interact with the CAD GRS as my work has shown. This is a crucial factor to consider because results from my study suggest that an individual's susceptibility to a high CAD GRS may be less clear in the presence of certain atherosclerotic risk factors like diabetes. Younger ages may be the ideal time to discuss an individual's genetic risks including CAD[142,143]. While these are important considerations to keep in mind for any clinical use of a CAD GRS, an in-depth analysis about costs and benefits is detailed below.

4.5.1 CAD GRS Benefits

In the past decade, numerous studies have showcased the various ways in which a CAD GRS can be an effective addition to CAD management and treatment in clinical settings. Previous works have shown that CAD GRS can help predict drug efficacy in certain subgroups of individuals[11,127,144,145]. One study observed that individuals with a high CAD GRS had the greatest reduction in major cardiovascular events and death when treated with a PCSK9 inhibitor[144]. Further studies will be needed to fully elucidate the role and impact of using a CAD GRS for predicting response to cardiovascular drugs.

Many studies have shown that the addition of a CAD GRS to the PCE (PCE+CAD GRS) produced significant improvement in risk stratification as measured by net reclassification indices (NRI) and GRS risk prediction[117,143,146-150]. Another study using the QRISK3 score to

assess conventional risk factors also observed an improvement in risk stratification[151]. Thus, an assessment of PCE+CAD GRS could result in a higher proportion of individuals correctly identified for early statin intervention. Consistent with this, individuals with a high genetic risk are reclassified to statin intervention when genetic risk is considered along with the conventional risk factors[152]. Notably, a recent study by Martson *et al.* emphasized that younger individuals had the strongest NRI improvement rates among various age brackets[117]. These findings are consistent with those from Riveros-Mckay *et al.* and Saadatagah *et al.*[143,150]. These observations are consistent with my observation that every GRS evaluated in this study had a stronger effect in younger individuals. Moreover, the additive effect of PCE+CAD GRS is likely due to the fact that they capture different genetic components of CAD. PCE incorporates CAD family history which often stems from rare CAD variants whereas a CAD GRS includes common CAD variants (minor allele frequency > 1%). Indeed, multiple studies have confirmed the independence of family history and CAD GRS in various cohorts such as the Malmö Diet and Cancer study and the Gender and Sex Determinants of Cardiovascular Disease From Bench to Beyond in Premature Acute Coronary Syndrome study[103,113,115,150]. The combination of PCE+CAD GRS could be a valuable clinical tool that integrates a larger proportion of an individual's genetic background than either PCE or the CAD GRS alone. Taken altogether, these results provide further evidence that a CAD GRS can be an asset in preventing CAD, particularly when provided to younger individuals. Ultimately, risk assessments with genetic information can improve primary prevention of CAD through pharmacological therapy and lifestyle modifications. Early intervention following specific genetic risk disclosure has the potential to reshape CAD prevention and intervention by putting a larger emphasis on prevention.

Clearly, PCE+CAD GRS has major potential economic consequences, as previously described by Mujwara *et al.*[153]. Their analysis of PCE+CAD GRS concluded that the addition of the CAD GRS was cost-effective with the mean cost diminishing by 181$ per person over 10 years. Considering that 2.6 million Canadians aged 20 and older were living with CAD in 2022[154], this decline in mean cost would represent a substantial savings. The results of Mujwara *et al.*[153] also highlighted another significant ramification of the PCE+CAD GRS: there were 50 fewer CAD events compared to PCE alone over 10 years in a cohort of 10,000 individuals, which resulted in an average cost savings of 36,000$ per event averted. In total, this would save 1.8 million dollars and much more when a larger population is considered. Previous work by Hynninen *et al.* also confirmed the cost-effectiveness of traditional risk factors and CAD GRS[18]. Notably, they observed that a combination of traditional risk factors and CAD GRS had a larger net monetary benefit compared to traditional risk factors alone. Further research should help to optimize the economic impact of PCE+CAD GRS.

Disclosing CAD genetic risk to individuals has been linked to positive behavioural health changes such as weight loss, smoking cessation, and consulting with a doctor[155]. Specifically, a recent prospective study which communicated a 10-year risk based on genetic and conventional risk factors to individuals concluded that individuals with a high genetic risk were more likely to make health changes than individuals with a low genetic risk[155]. In addition, an observational study has shown that disclosure of genetic risk led to a modest increase in physical exercise and weight loss[156]. Likewise, the Myocardial Infarction Genes (MI-GENES) study demonstrated that disclosing genetic risk to CAD in addition to evaluating conventional cardiovascular risk factors prompted a greater reduction of LDL-C levels when compared to only evaluating conventional cardiovascular risk factors[157]. Indeed, the disclosure of genetic risk of CAD caused a higher

proportion of individuals, in collaboration with their doctor, to initiate statin therapy. Additional prospective large-scale studies will be needed to validate the hypothesis that positive health changes can be increased when individuals are provided with their CAD GRS.

A complete assessment of individual disease risk prior to the development of atherosclerotic risk factors may have consequential ramifications for CAD prevention, risk assessment and cardiovascular healthcare expenditure. The latter is of crucial significance as Canada faces an ageing population which exacerbates the growing burden of cardiovascular disease on the healthcare system[158].

4.5.2 CAD GRS Limitations

Although the field of CAD GRS has been the subject of a sizeable number of studies in the past decade, issues still exist regarding the implementation of a CAD GRS in clinical settings. Ancestry portability, GRS optimization, and addition to risk calculators such as the PCE are some of the considerations that need to be addressed in relation to the clinical use of a CAD GRS.

First of all, a current limitation of CAD GRS is their transferability to non-European ancestries, as observed above in chapter 3. Multiple studies confirm that performance of a CAD GRS is biased towards the ancestry which provided the summary statistics[3,104,105,133,149,159]. In the case of European ancestry, a GRS based on European summary statistics (GWAS), will perform better among individuals of European ancestry than in individuals of non-European ancestries. For instance, my findings detailed in chapter 3 are based on individuals of European ancestry and they illustrate the reduced predictive ability of the GRS_{204} in non-European ancestries, though

the issue of reduced power must also be considered. Eurocentric CAD GRS results cannot necessarily be extrapolated to other ancestry groups as the causal genetic variants of CAD and their linkage disequilibrium patterns vary between ethnic groups[160]. This trend has been seen in GRS for other diseases as well[135,161-164]. However, some improvements have been observed when using an ancestry-specific GRS for non-European ancestries. This was exemplified by Onengut-Gumuscu *et al.* when they noted that an African GWAS-based GRS provided stronger prediction for DM among individuals of African ancestry than a European GWAS-based GRS[163].These issues bring forth an ethical concern: that the implementation of a CAD GRS in clinical settings will not be equally effective for all individuals. In fact, a recent study emphasized that any clinical use of a eurocentric GRS may exacerbate health disparities between individuals of various ancestries[135]. One possible solution would be to create an ancestry-specific CAD GRS (e.g. a South Asian ancestry CAD GWAS would be used to create a South Asian GRS). However, as Martin *et al.* pointed out in their 2019 paper, 79% of GWAS participants are of European ancestry even though they only account for 16% of the global population[135] (Figure 6). This profound imbalance has led to the creation of hundreds of eurocentric GRS and very few non-European GRS. Although many recent efforts have focused on recruiting ethnically diverse participants for large data banks, it will take many years, perhaps over a decade, for GWAS parity between ancestries to be achieved. Once large-scale GWAS for various ancestries are available, additional studies should be performed to validate the efficacy of ancestry-specific GRS as well as differences in risk factor interactions.

A possible solution to this problem is to train and validate a GRS in a diverse group of individuals despite using summary statistics from a European-ancestry GWAS.

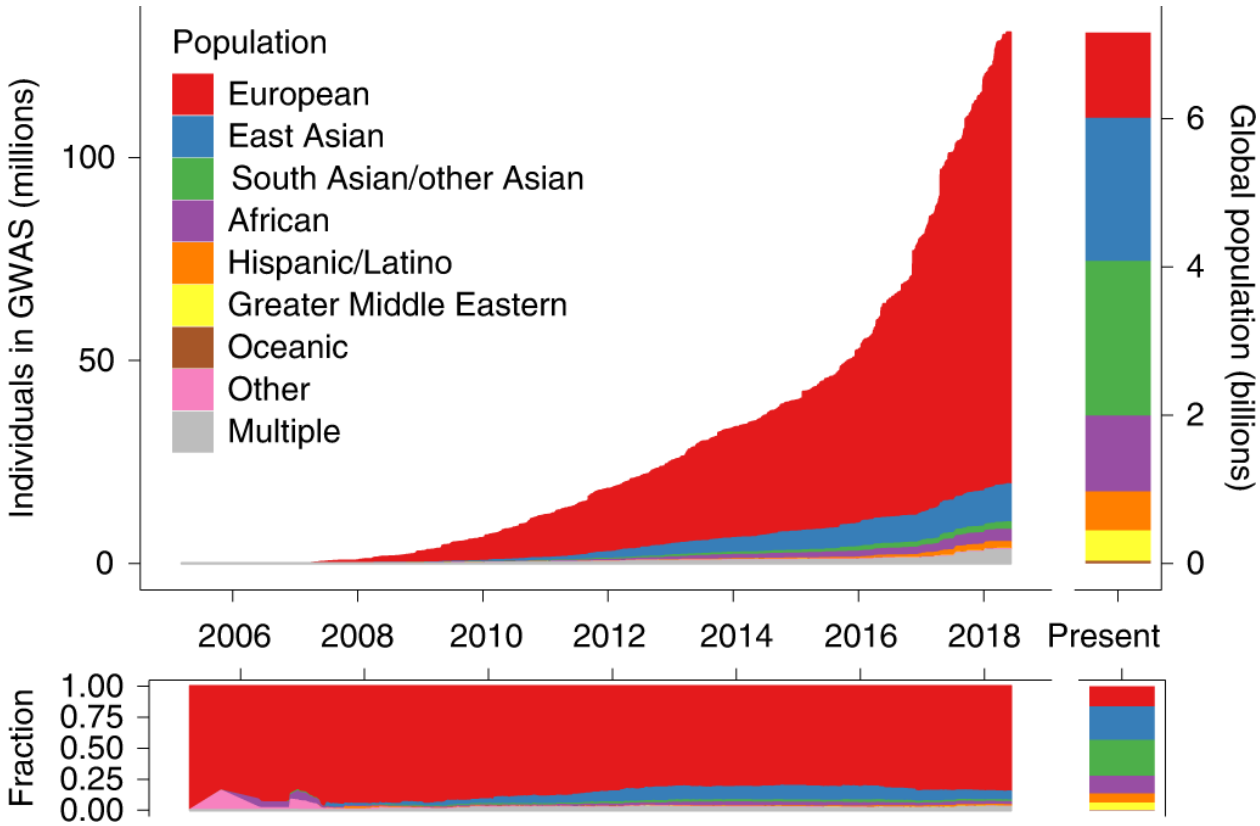


Figure 6: Ancestry of GWAS participants with reference to the global population

A comprehensive analysis of GWAS participants and their ancestries over time as well as in comparison to the global population. (Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. Apr 2019;51(4):584-591) (Reproduced with permission of the publisher)

This method has had inconclusive results in recent years: Prive *et al.* reported that in Europeans and Caribbeans a mixed-ancestry GRS had comparable predictive ability to a Eurocentric GRS whereas Márquez-Luna *et al.* and Cavazos and Witte illustrated the benefit of using a mixed-ancestry GRS instead of a Eurocentric GRS[105,164,165]. Factors such as different GRS construction methods and different sizes of non-European samples could explain these contradicting results. It is worth noting that the study from Prive *et al.* focused on CAD GRS among other diseases while the other two studies did not include CAD in their analyses. The mixed-ancestry GRS method might be more applicable for some diseases than others. Nonetheless, while some studies have investigated mixed-ancestry approaches for CAD[3,97,166], further studies are needed to get a clearer understanding of the potential advantages of this method.

Another issue that has risen in the past few years is the lack of consensus surrounding GRS construction approaches. While over 20 different construction methods for single-ancestry and multi-ancestry GRS exist[109], the scientific community has not come to a decision regarding which method is the best for clinical settings. Although it has become standard for studies to compare their own CAD GRS to existing ones in the literature, it can be very difficult to contrast these GRS comparisons studies with one another due to many factors. For example, different CAD definitions, cohorts, inclusion criteria, GRS construction methods and the number of genetic variants included in the GRS are all factors that can hinder the comparison of CAD GRS studies. Additionally, as most of these methods are mainly being evaluated in individuals of European ancestry, specific methods may perform differently for different mixtures of ancestries.

Recent work by Patel *et al.* explored combining multi-ancestry GRS for CAD and related risk factors to create a multi-ancestry multi-trait GRS including GRS for CAD, BMI, LDL, HDL, and DM[147]. They concluded that their multi-trait GRS outperformed many published CAD GRS

in an independent cohort, the MVP. Although the composition of ancestries in each trait was heavily skewed towards European ancestry, their findings illustrated that the multi-trait GRS had a stronger association with CAD in four different ancestries (European, African, East Asian, South Asian) when compared to other published CAD GRS. However, this method may only be advantageous for young individuals as the CAD risk factors themselves will be more predictive in older adults. Consequently, the GRS for those risk factors may not be useful. Additional studies are needed to determine in which situations (e.g., younger individuals) this multi-trait CAD GRS provides a better way to optimize GRS prediction.

While numerous studies have highlighted the positive impact of adding a CAD GRS to the conventional risk factors (RF+CAD GRS)[117,143,146-151], others have contradicted these findings over the past few years[167-169]. Specifically, they all noted that RF+CAD GRS did not significantly improve prediction accuracy, nor did it improve reclassification (net reclassification indices). Although there are possible explanations for the differing results, such as different CAD GRS construction method, SNP number, QC parameters, inclusion criteria, ancestry distribution, the lack of consistent evidence gives rise to some hesitation about the use of a CAD GRS in clinical settings. Furthermore, detailed guidelines about the application, interpretation and communication of CAD GRS results to individuals still need to be worked out. For example, would individuals with an extreme GRS need to consult genetic counsellors for an understanding of the genetic risk for CAD in their relatives? These queries still need to be addressed as more CAD GRS research occurs in the next few years. If future large-scale research can address some of the concerns detailed above, such as ancestry portability, then more conclusive data regarding the clinical use of a CAD GRS should be uncovered. GRS optimization, including risk factor interactions, could positively affect the accuracy of a CAD GRS.

4.6 Limitations

My study was conducted in the UKB, a large, well-characterized cohort, and across multiple ancestries. I also included 204 variants in my primary GRS, as well as evaluated different subsets and different GRS building approaches. Despite these strengths, the study has several limitations. First, as previously mentioned above, findings based on individuals of European ancestry cannot always be replicated in other ancestries. The frequency of CAD variants and their LD arrangements which vary between ancestry groups account for some of this discordance[160]. However, many studies are under way to increase the amount of available genetic data from diverse and minority populations, which should in turn improve GRS predictions in diverse cohorts[3,135,170]. In the present work, I observed that the association of the CAD GRS was replicated in South Asian individuals but not Chinese or Black individuals. This failure to replicate is most likely due at least in part to the small sample sizes[139-141] and specifically the low number of available CAD cases in these ethnic groups[10]. Second, all the observed interactions may be misestimated due to the exclusion of prevalent CAD cases. Third, participants of the UK Biobank have been shown to be healthier than the general population[171]. For example, participants were less likely to drink alcohol or smoke and self-reported fewer health conditions when compared to characteristics of the general British population and thus, participants may not be representative of the overall population. Fourth, copy number variants (CNVs) and structural variants were not considered in my CAD GRS despite certain CNVs being associated with various cardiovascular diseases[172]. Finally, it has been demonstrated that higher numbers of genetic markers in a GRS can improve prediction[113,173] even if some of these variants are not genome-wide significant[3]. Thus, partitioning a GRS, as I did for my risk factor defined subsets, and others have done[13,15] might decrease the power.

Chapter 5: Conclusions and Future directions

5.1 Conclusions

In summary, I assessed the strength of a CAD GRS in atherosclerotic risk factor subgroups. I identified four atherosclerotic risk factors that have significant interactions with a CAD GRS – age, sex, dyslipidemia, and diabetes – in individuals of European ancestry. I also replicated the GRS association with CAD in individuals of South Asian ancestry. Additionally, I investigated the impact of different GRS construction methods on the strength of a CAD GRS and the identified interactions. I concluded that each method had comparable CAD GRS associations with CAD. Approaches to include a CAD GRS in clinical prediction should consider these interaction results to optimize predictive performance.

5.2 Future Directions

5.2.1 Analyses in Other Large-Scale Cohorts of Non-European Ancestry

Future research should attempt to replicate my analyses and observations to other cohorts with a larger representation of non-European ancestries such as the MVP, All Of Us, Japan BioBank, etc. This may address the lack of replication of the CAD GRS association in Black and Chinese individuals when compared to European individuals, which is a problem that has been observed in other studies[104,105,133] . As mentioned above, it was most likely due to the small sample size available in these ancestries[135]. Performing my analyses in more diverse ancestries will allow an assessment of the transferability of European findings but also potentially highlight novel interactions between a CAD GRS and risk factors in non-European ancestries. The ability to transfer specific GRS construction methods and SNP sets (e.g. GRS_{204}, GRS_{MVP}, $GRS_{LDpred2}$, etc.) to non-European ancestries, will also be of great importance. Moreover, collaborations with

other groups in Canada and around the world including large databanks with a higher proportion of non-European ancestries could yield additional novel results, some of which may be specific to non-European ancestries. Such a meta-analysis would also provide another opportunity to analyze the application of a European-derived GRS to individuals of non-European ancestries to determine its generalizability.

5.2.2 The Responsibility of Informing Individuals About Their Risk Scores

In addition, future research must include detailed analyses on the best ways to accurately and responsibly inform individuals about their risk scores for certain diseases. Many gaps remain in the translation of risk scores from bioinformatics teams to clinicians to individuals[174]. For instance, it is crucial that clinicians make accurate interpretations and individuals have an in-depth understanding of their risk score. One study helped individuals fully comprehend their risk scores by having a genetic counsellor explain it to them in addition to meeting with their health care provider to discuss potential medical changes (e.g., statin usage)[157]. The lack of anxiety that individuals felt about their risk score in this study by Kullo *et al.* could partially be due to the genetic counselling session they all had. Moreover, other aspects of informing individuals about their risk scores to consider are consent and cost-effectiveness[174]. Can parents consent to learn about their child's risk score? Should that right be reserved for the child when they reach adulthood? Additionally, while some GRS for certain diseases might be cost-effective, additional research will need to be conducted before a blanket statement can be made about the cost-effectiveness of all GRS. Proper national or international guidelines including clinical recommendations to make based on GRS would help remedy the translation gap and set clinical standards.

Chapter 6: References

1. Khan MA, Hashim MJ, Mustafa H, et al. Global Epidemiology of Ischemic Heart Disease: Results from the Global Burden of Disease Study. *Cureus*. Jul 23 2020;12(7):e9349. doi:10.7759/cureus.9349
2. Assimes TL, Roberts R. Genetics: Implications for Prevention and Management of Coronary Artery Disease. *J Am Coll Cardiol*. Dec 27 2016;68(25):2797-2818. doi:10.1016/j.jacc.2016.10.039
3. Tcheandjieu C, Zhu X, Hilliard AT, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat Med*. Aug 2022;28(8):1679-1692. doi:10.1038/s41591-022-01891-3
4. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ Res*. Feb 2 2018;122(3):433-443. doi:10.1161/CIRCRESAHA.117.312086
5. Ozaki K, Ohnishi Y, Iida A, et al. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet*. Dec 2002;32(4):650-4. doi:10.1038/ng1047
6. Matsunaga H, Ito K, Akiyama M, et al. Transethnic Meta-Analysis of Genome-Wide Association Studies Identifies Three New Loci and Characterizes Population-Specific Differences for Coronary Artery Disease. *Circ Genom Precis Med*. Jun 2020;13(3):e002670. doi:10.1161/CIRCGEN.119.002670
7. Roberts R. Genetics in the prevention and management of coronary artery disease. *Curr Opin Cardiol*. May 2018;33(3):257-268. doi:10.1097/HCO.0000000000000501
8. Ntalla I, Kanoni S, Zeng L, et al. Genetic Risk Score for Coronary Disease Identifies Predispositions to Cardiovascular and Noncardiovascular Diseases. *J Am Coll Cardiol*. Jun 18 2019;73(23):2932-2942. doi:10.1016/j.jacc.2019.03.512
9. Morieri ML, Gao H, Pigeyre M, et al. Genetic Tools for Coronary Risk Assessment in Type 2 Diabetes: A Cohort Study From the ACCORD Clinical Trial. *Diabetes Care*. Nov 2018;41(11):2404-2413. doi:10.2337/dc18-0709
10. Inouye M, Abraham G, Nelson CP, et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *J Am Coll Cardiol*. Oct 16 2018;72(16):1883-1893. doi:10.1016/j.jacc.2018.07.079
11. Mega JL, Stitziel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet*. Jun 6 2015;385(9984):2264-2271. doi:10.1016/S0140-6736(14)61730-X
12. Manikpurage HD, Eslami A, Perrot N, et al. Polygenic Risk Score for Coronary Artery Disease Improves the Prediction of Early-Onset Myocardial Infarction and Mortality in Men. *Circ Genom Precis Med*. Dec 2021;14(6):e003452. doi:10.1161/CIRCGEN.121.003452
13. Huang Y, Hui Q, Gwinn M, et al. Sexual Differences in Genetic Predisposition of Coronary Artery Disease. *Circ Genom Precis Med*. Feb 2021;14(1):e003147. doi:10.1161/CIRCGEN.120.003147
14. Cole CB, Nikpay M, Stewart AF, McPherson R. Increased genetic risk for obesity in premature coronary artery disease. *Eur J Hum Genet*. Apr 2016;24(4):587-91. doi:10.1038/ejhg.2015.162

15. Huang Y, Hui Q, Gwinn M, et al. Interaction between genetics and smoking in determining risk of coronary artery diseases. *Genet Epidemiol*. Apr 2022;46(3-4):199-212. doi:10.1002/gepi.22446
16. Roth GA, Mensah GA, Johnson CO, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. *J Am Coll Cardiol*. Dec 22 2020;76(25):2982-3021. doi:10.1016/j.jacc.2020.11.010
17. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association. *Circulation*. Feb 22 2022;145(8):e153-e639. doi:10.1161/CIR.0000000000001052
18. Hynninen Y, Linna M, Vilkkumaa E. Value of genetic testing in the prevention of coronary heart disease events. *PLoS One*. 2019;14(1):e0210010. doi:10.1371/journal.pone.0210010
19. Zhu KF, Wang YM, Zhu JZ, Zhou QY, Wang NF. National prevalence of coronary heart disease and its relationship with human development index: A systematic review. *Eur J Prev Cardiol*. Mar 2016;23(5):530-43. doi:10.1177/2047487315587402
20. Schultz WM, Kelli HM, Lisko JC, et al. Socioeconomic Status and Cardiovascular Outcomes: Challenges and Interventions. *Circulation*. May 15 2018;137(20):2166-2178. doi:10.1161/CIRCULATIONAHA.117.029652
21. Go AS, Mozaffarian D, Roger VL, et al. Executive summary: heart disease and stroke statistics--2013 update: a report from the American Heart Association. *Circulation*. Jan 1 2013;127(1):143-52. doi:10.1161/CIR.0b013e318282ab8f
22. Birger M, Kaldjian AS, Roth GA, Moran AE, Dieleman JL, Bellows BK. Spending on Cardiovascular Disease and Cardiovascular Risk Factors in the United States: 1996 to 2016. *Circulation*. Jul 27 2021;144(4):271-282. doi:10.1161/CIRCULATIONAHA.120.053216
23. Pereira E, Pereira H. Socioeconomic impact of cardiovascular disease. *Rev Port Cardiol (Engl Ed)*. May 2020;39(5):253-254. doi:10.1016/j.repc.2020.05.002
24. Gordois AL, Toth PP, Quek RG, Proudfoot EM, Paoli CJ, Gandra SR. Productivity losses associated with cardiovascular disease: a systematic review. *Expert Rev Pharmacoecon Outcomes Res*. Dec 2016;16(6):759-769. doi:10.1080/14737167.2016.1259571
25. Kigozi J, Jowett S, Lewis M, Barton P, Coast J. The Estimation and Inclusion of Presenteeism Costs in Applied Economic Evaluation: A Systematic Review. *Value Health*. Mar 2017;20(3):496-506. doi:10.1016/j.jval.2016.12.006
26. Zhang J, Zu Y, Dhanasekara CS, et al. Detection and treatment of atherosclerosis using nanoparticles. *Wiley Interdiscip Rev Nanomed Nanobiotechnol*. Jan 2017;9(1)doi:10.1002/wnan.1412
27. Insull W, Jr. The pathology of atherosclerosis: plaque development and plaque responses to medical treatment. *Am J Med*. Jan 2009;122(1 Suppl):S3-S14. doi:10.1016/j.amjmed.2008.10.013
28. Gao S, Liu J. Association between circulating oxidized low-density lipoprotein and atherosclerotic cardiovascular disease. *Chronic Dis Transl Med*. Jun 25 2017;3(2):89-94. doi:10.1016/j.cdtm.2017.02.008
29. Malekmohammad K, Bezsonov EE, Rafieian-Kopaei M. Role of Lipid Accumulation and Inflammation in Atherosclerosis: Focus on Molecular and Cellular Mechanisms. *Front Cardiovasc Med*. 2021;8:707529. doi:10.3389/fcvm.2021.707529
30. Linton MRF, Yancey PG, Davies SS, et al. The Role of Lipids and Lipoproteins in Atherosclerosis. In: Feingold KR, Anawalt B, Boyce A, et al, eds. *Endotext*. 2000.

31. Theofilis P, Sagris M, Oikonomou E, et al. Inflammatory Mechanisms Contributing to Endothelial Dysfunction. *Biomedicines*. Jul 6 2021;9(7)doi:10.3390/biomedicines9070781
32. Burtenshaw D, Kitching M, Redmond EM, Megson IL, Cahill PA. Reactive Oxygen Species (ROS), Intimal Thickening, and Subclinical Atherosclerotic Disease. *Front Cardiovasc Med*. 2019;6:89. doi:10.3389/fcvm.2019.00089
33. Nowak WN, Deng J, Ruan XZ, Xu Q. Reactive Oxygen Species Generation and Atherosclerosis. *Arterioscler Thromb Vasc Biol*. May 2017;37(5):e41-e52. doi:10.1161/ATVBAHA.117.309228
34. Yang X, Li Y, Li Y, et al. Oxidative Stress-Mediated Atherosclerosis: Mechanisms and Therapies. *Front Physiol*. 2017;8:600. doi:10.3389/fphys.2017.00600
35. Flynn MC, Pernes G, Lee MKS, Nagareddy PR, Murphy AJ. Monocytes, Macrophages, and Metabolic Disease in Atherosclerosis. *Front Pharmacol*. 2019;10:666. doi:10.3389/fphar.2019.00666
36. Weber C, Noels H. Atherosclerosis: current pathogenesis and therapeutic options. *Nat Med*. Nov 7 2011;17(11):1410-22. doi:10.1038/nm.2538
37. Hajar R. Risk Factors for Coronary Artery Disease: Historical Perspectives. *Heart Views*. Jul-Sep 2017;18(3):109-114. doi:10.4103/HEARTVIEWS.HEARTVIEWS_106_17
38. Rodgers JL, Jones J, Bolleddu SI, et al. Cardiovascular Risks Associated with Gender and Aging. *J Cardiovasc Dev Dis*. Apr 27 2019;6(2)doi:10.3390/jcdd6020019
39. Pinto E. Blood pressure and ageing. *Postgrad Med J*. Feb 2007;83(976):109-14. doi:10.1136/pgmj.2006.048371
40. Wang JC, Bennett M. Aging and atherosclerosis: mechanisms, functional consequences, and potential therapeutics for cellular senescence. *Circ Res*. Jul 6 2012;111(2):245-59. doi:10.1161/CIRCRESAHA.111.261388
41. North BJ, Sinclair DA. The intersection between aging and cardiovascular disease. *Circ Res*. Apr 13 2012;110(8):1097-108. doi:10.1161/CIRCRESAHA.111.246876
42. Maas AH, Appelman YE. Gender differences in coronary heart disease. *Neth Heart J*. Dec 2010;18(12):598-602. doi:10.1007/s12471-010-0841-y
43. Wakabayashi I. Gender differences in cardiovascular risk factors in patients with coronary artery disease and those with type 2 diabetes. *J Thorac Dis*. May 2017;9(5):E503-E506. doi:10.21037/jtd.2017.04.30
44. Moller-Leimkuhler AM. Gender differences in cardiovascular disease and comorbid depression. *Dialogues Clin Neurosci*. 2007;9(1):71-83. doi:10.31887/DCNS.2007.9.1/ammoeller
45. Lloyd-Jones DM, Nam BH, D'Agostino RB, Sr., et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA*. May 12 2004;291(18):2204-11. doi:10.1001/jama.291.18.2204
46. Jensen MD, Ryan DH, Apovian CM, et al. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *J Am Coll Cardiol*. Jul 1 2014;63(25 Pt B):2985-3023. doi:10.1016/j.jacc.2013.11.004
47. Powell-Wiley TM, Poirier P, Burke LE, et al. Obesity and Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation*. May 25 2021;143(21):e984-e1010. doi:10.1161/CIR.0000000000000973
48. Ellulu MS, Patimah I, Khaza'ai H, Rahmat A, Abed Y. Obesity and inflammation: the linking mechanism and the complications. *Arch Med Sci*. Jun 2017;13(4):851-863. doi:10.5114/aoms.2016.58928

49. Couillard C, Ruel G, Archer WR, et al. Circulating levels of oxidative stress markers and endothelial adhesion molecules in men with abdominal obesity. *J Clin Endocrinol Metab*. Dec 2005;90(12):6454-9. doi:10.1210/jc.2004-2438
50. Al-Goblan AS, Al-Alfi MA, Khan MZ. Mechanism linking diabetes mellitus and obesity. *Diabetes Metab Syndr Obes*. 2014;7:587-91. doi:10.2147/DMSO.S67400
51. Ades PA, Savage PD. Obesity in coronary heart disease: An unaddressed behavioral risk factor. *Prev Med*. Nov 2017;104:117-119. doi:10.1016/j.ypmed.2017.04.013
52. Kaneto H, Katakami N, Matsuhisa M, Matsuoka TA. Role of reactive oxygen species in the progression of type 2 diabetes and atherosclerosis. *Mediators Inflamm*. 2010;2010:453892. doi:10.1155/2010/453892
53. Pickup JC. Inflammation and activated innate immunity in the pathogenesis of type 2 diabetes. *Diabetes Care*. Mar 2004;27(3):813-23. doi:10.2337/diacare.27.3.813
54. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. Oct 23 2018;138(17):e426-e483. doi:10.1161/CIR.0000000000000597
55. Martinez-Quinones P, McCarthy CG, Watts SW, et al. Hypertension Induced Morphological and Physiological Changes in Cells of the Arterial Wall. *Am J Hypertens*. Sep 11 2018;31(10):1067-1078. doi:10.1093/ajh/hpy083
56. Liu Y, Luo X, Jia H, Yu B. The Effect of Blood Pressure Variability on Coronary Atherosclerosis Plaques. *Front Cardiovasc Med*. 2022;9:803810. doi:10.3389/fcvm.2022.803810
57. Stein R, Ferrari F, Scolari F. Genetics, Dyslipidemia, and Cardiovascular Disease: New Insights. *Curr Cardiol Rep*. Jun 21 2019;21(8):68. doi:10.1007/s11886-019-1161-5
58. Garg R, Aggarwal S, Kumar R, Sharma G. Association of atherosclerosis with dyslipidemia and co-morbid conditions: A descriptive study. *J Nat Sci Biol Med*. Jan-Jun 2015;6(1):163-8. doi:10.4103/0976-9668.149117
59. Elkhalifa AM. Effects of cigarette smoking on coagulation screening tests and platelet counts in a Sudanese male adults population. *Saudi Med J*. Sep 2018;39(9):897-901. doi:10.15537/smj.2018.9.22630
60. Messner B, Bernhard D. Smoking and cardiovascular disease: mechanisms of endothelial dysfunction and early atherogenesis. *Arterioscler Thromb Vasc Biol*. Mar 2014;34(3):509-15. doi:10.1161/ATVBAHA.113.300156
61. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. Jun 18 2019;139(25):e1046-e1081. doi:10.1161/CIR.0000000000000624
62. Rehberger Likozar A, Zavrtanik M, Sebestjen M. Lipoprotein(a) in atherosclerosis: from pathophysiology to clinical relevance and treatment options. *Ann Med*. Aug 2020;52(5):162-177. doi:10.1080/07853890.2020.1775287
63. Wu HD, Berglund L, Dimayuga C, et al. High lipoprotein(a) levels and small apolipoprotein(a) sizes are associated with endothelial dysfunction in a multiethnic cohort. *J Am Coll Cardiol*. May 19 2004;43(10):1828-33. doi:10.1016/j.jacc.2003.08.066

64. Goff DC, Jr., Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. Jul 1 2014;63(25 Pt B):2935-2959. doi:10.1016/j.jacc.2013.11.005
65. Damen JA, Pajouheshnia R, Heus P, et al. Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. *BMC Med*. Jun 13 2019;17(1):109. doi:10.1186/s12916-019-1340-7
66. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. Jun 24 2014;129(25 Suppl 2):S1-45. doi:10.1161/01.cir.0000437738.63853.7a
67. Jia S, Liu Y, Yuan J. Evidence in Guidelines for Treatment of Coronary Artery Disease. *Adv Exp Med Biol*. 2020;1177:37-73. doi:10.1007/978-981-15-2517-9_2
68. Doenst T, Thiele H, Haasenritter J, Wahlers T, Massberg S, Haverich A. The Treatment of Coronary Artery Disease. *Dtsch Arztebl Int*. Oct 21 2022;119(42):716-723. doi:10.3238/arztebl.m2022.0277
69. Friedlander Y. Familial Clustering of Coronary Heart Disease: A Review of its Significance and Role as a Risk Factor for the Disease. *Genetic factors in coronary heart disease*. Springer; 1994:chap 37-53.
70. Chow CK, Islam S, Bautista L, et al. Parental history and myocardial infarction risk across the world: the INTERHEART Study. *J Am Coll Cardiol*. Feb 1 2011;57(5):619-27. doi:10.1016/j.jacc.2010.07.054
71. Nielsen M, Andersson C, Gerds TA, et al. Familial clustering of myocardial infarction in first-degree relatives: a nationwide study. *Eur Heart J*. Apr 2013;34(16):1198-203. doi:10.1093/eurheartj/ehs475
72. Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med*. Apr 14 1994;330(15):1041-6. doi:10.1056/NEJM199404143301503
73. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med*. Sep 2002;252(3):247-54. doi:10.1046/j.1365-2796.2002.01029.x
74. Wienke A, Holm NV, Skytthe A, Yashin AI. The heritability of mortality due to heart diseases: a correlated frailty model applied to Danish twins. *Twin Res*. Aug 2001;4(4):266-74. doi:10.1375/1369052012399
75. The International Human Genome Sequencing C. International consortium completes human genome project. *Pharmacogenomics*. May 2003;4(3):241. doi:10.1517/phgs.4.3.241.22688
76. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. Feb 15 2001;409(6822):860-921. doi:10.1038/35057062
77. International HapMap C. The International HapMap Project. *Nature*. Dec 18 2003;426(6968):789-96. doi:10.1038/nature02168
78. International HapMap C, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. Oct 18 2007;449(7164):851-61. doi:10.1038/nature06258

79. Uffelmann E, Huang Q, Munung N, et al. Genome-wide association studies. *Nature Reviews Methods Primers*. 2021;1(1):1-21.
80. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. Jun 8 2007;316(5830):1488-91. doi:10.1126/science.1142447
81. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. Jun 8 2007;316(5830):1491-3. doi:10.1126/science.1142842
82. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. Jun 7 2007;447(7145):661-78. doi:10.1038/nature05911
83. Schunkert H, Gotz A, Braund P, et al. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation*. Apr 1 2008;117(13):1675-84. doi:10.1161/CIRCULATIONAHA.107.730614
84. Shen GQ, Li L, Rao S, et al. Four SNPs on chromosome 9p21 in a South Korean population implicate a genetic locus that confers high cross-race risk for development of coronary artery disease. *Arterioscler Thromb Vasc Biol*. Feb 2008;28(2):360-5. doi:10.1161/ATVBAHA.107.157248
85. Do R, Xie C, Zhang X, et al. 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*. Oct 2011;8(10):e1001106. doi:10.1371/journal.pmed.1001106
86. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. Oct 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
87. Chen Z, Schunkert H. Genetics of coronary artery disease in the post-GWAS era. *J Intern Med*. Nov 2021;290(5):980-992. doi:10.1111/joim.13362
88. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet*. Jun 2017;18(6):331-344. doi:10.1038/nrg.2016.160
89. Finan C, Gaulton A, Kruger FA, et al. The druggable genome and support for target identification and validation in drug development. *Sci Transl Med*. Mar 29 2017;9(383)doi:10.1126/scitranslmed.aag1166
90. Trajanoska K, Bherer C, Taliun D, Zhou S, Richards JB, Mooser V. From target discovery to clinical drug development with human genetics. *Nature*. Aug 2023;620(7975):737-745. doi:10.1038/s41586-023-06388-8
91. Abifadel M, Varret M, Rabes JP, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet*. Jun 2003;34(2):154-6. doi:10.1038/ng1161
92. Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med*. Mar 23 2006;354(12):1264-72. doi:10.1056/NEJMoa054013
93. Robinson JG, Farnier M, Krempf M, et al. Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. *N Engl J Med*. Apr 16 2015;372(16):1489-99. doi:10.1056/NEJMoa1501031
94. Sabatine MS, Giugliano RP, Wiviott SD, et al. Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. *N Engl J Med*. Apr 16 2015;372(16):1500-9. doi:10.1056/NEJMoa1500858

95. Tragante V, Hemerich D, Alshabeeb M, et al. Druggability of Coronary Artery Disease Risk Loci. *Circ Genom Precis Med*. Aug 2018;11(8):e001977. doi:10.1161/CIRCGEN.117.001977
96. McPherson R, Tybjaerg-Hansen A. Genetics of Coronary Artery Disease. *Circ Res*. Feb 19 2016;118(4):564-78. doi:10.1161/CIRCRESAHA.115.306566
97. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. Oct 2015;47(10):1121-1130. doi:10.1038/ng.3396
98. Igo RP, Jr., Kinzy TG, Cooke Bailey JN. Genetic Risk Scores. *Curr Protoc Hum Genet*. Dec 2019;104(1):e95. doi:10.1002/cphg.95
99. Roberts R, Chavira J, Venner E. Genetic risk and its role in primary prevention of CAD. *Journal of Translational Genetics and Genomics*. 2022;6(4):388-402. doi:http://dx.doi.org/10.20517/jtgg.2022.07
100. Park JK, Lu CY. Polygenic Scores in the Direct-to-Consumer Setting: Challenges and Opportunities for a New Era in Consumer Genetic Testing. *J Pers Med*. Mar 23 2023;13(4)doi:10.3390/jpm13040573
101. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med*. May 18 2020;12(1):44. doi:10.1186/s13073-020-00742-5
102. 23andMe I. How 23andMe Predicted My Likelihood of Developing the "Disease of Kings.". 23andMe Blog. https://blog.23andme.com/articles/how-23andme-predicted-my-likelihood-of-developing-the-disease-of-kings
103. Hassanin E, Maj C, Klinkhammer H, Krawitz P, May P, Bobbili DR. Assessing the performance of European-derived cardiometabolic polygenic risk scores in South-Asians and their interplay with family history. *BMC Med Genomics*. Jul 12 2023;16(1):164. doi:10.1186/s12920-023-01598-5
104. Mars N, Kerminen S, Feng YA, et al. Genome-wide risk prediction of common diseases across ancestries in one million people. *Cell Genom*. Apr 13 2022;2(4):None. doi:10.1016/j.xgen.2022.100118
105. Prive F, Aschard H, Carmi S, et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *Am J Hum Genet*. Feb 3 2022;109(2):373. doi:10.1016/j.ajhg.2022.01.007
106. Oh B. Direct-to-consumer genetic testing: advantages and pitfalls. *Genomics Inform*. Sep 2019;17(3):e33. doi:10.5808/GI.2019.17.3.e33
107. Prive F, Arbel J, Vilhjalmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics*. Dec 16 2020;36(22-23):5424-31. doi:10.1093/bioinformatics/btaa1029
108. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc*. Sep 2020;15(9):2759-2772. doi:10.1038/s41596-020-0353-1
109. Khunsriraksakul C, Markus H, Olsen NJ, Carrel L, Jiang B, Liu DJ. Construction and Application of Polygenic Risk Scores in Autoimmune Diseases. *Front Immunol*. 2022;13:889296. doi:10.3389/fimmu.2022.889296
110. Lamri A, Mao S, Desai D, Gupta M, Pare G, Anand SS. Fine-tuning of Genome-Wide Polygenic Risk Scores and Prediction of Gestational Diabetes in South Asian Women. *Sci Rep*. Jun 2 2020;10(1):8941. doi:10.1038/s41598-020-65360-y
111. Baek EJ, Jung HU, Chung JY, et al. The effect of heteroscedasticity on the prediction efficiency of genome-wide polygenic score for body mass index. *Front Genet*. 2022;13:1025568. doi:10.3389/fgene.2022.1025568

112. Wang C, Zhang J, Veldsman WP, Zhou X, Zhang L. A comprehensive investigation of statistical and machine learning approaches for predicting complex human diseases on genomic variants. *Brief Bioinform*. Jan 19 2023;24(1)doi:10.1093/bib/bbac552
113. Tada H, Melander O, Louie JZ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur Heart J*. Feb 7 2016;37(6):561-7. doi:10.1093/eurheartj/ehv462
114. Ference BA, Yoo W, Alesh I, et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. *J Am Coll Cardiol*. Dec 25 2012;60(25):2631-9. doi:10.1016/j.jacc.2012.09.017
115. Hindieh W, Pilote L, Cheema A, et al. Association Between Family History, a Genetic Risk Score, and Severity of Coronary Artery Disease in Patients With Premature Acute Coronary Syndromes. *Arterioscler Thromb Vasc Biol*. Jun 2016;36(6):1286-92. doi:10.1161/ATVBAHA.115.306944
116. Joseph PG, Pare G, Asma S, et al. Impact of a Genetic Risk Score on Myocardial Infarction Risk Across Different Ethnic Populations. *Can J Cardiol*. Dec 2016;32(12):1440-1446. doi:10.1016/j.cjca.2016.05.014
117. Marston NA, Pirruccello JP, Melloni GEM, et al. Predictive Utility of a Coronary Artery Disease Polygenic Risk Score in Primary Prevention. *JAMA Cardiol*. Dec 28 2022;doi:10.1001/jamacardio.2022.4466
118. Surakka I, Wolford BN, Ritchie SC, et al. Sex-Specific Survival Bias and Interaction Modeling in Coronary Artery Disease Risk Prediction. *Circ Genom Precis Med*. Feb 2023;16(1):e003542. doi:10.1161/CIRCGEN.121.003542
119. Collins R. What makes UK Biobank special? *Lancet*. Mar 31 2012;379(9822):1173-4. doi:10.1016/S0140-6736(12)60404-8
120. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. doi:10.1186/s13742-015-0047-8
121. Vikulova DN, Grubisic M, Zhao Y, et al. Premature Atherosclerotic Cardiovascular Disease: Trends in Incidence, Risk Factors, and Sex-Related Differences, 2000 to 2016. *J Am Heart Assoc*. Jul 16 2019;8(14):e012178. doi:10.1161/JAHA.119.012178
122. Lithovius R, Antikainen AA, Mutter S, et al. Genetic Risk Score Enhances Coronary Artery Disease Risk Prediction in Individuals With Type 1 Diabetes. *Diabetes Care*. Mar 1 2022;45(3):734-741. doi:10.2337/dc21-0974
123. Jiang X, Holmes C, McVean G. The impact of age on genetic risk for common diseases. *PLoS Genet*. Aug 2021;17(8):e1009723. doi:10.1371/journal.pgen.1009723
124. Thompson DJ, Wells D, Selzam S, et al. UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. *medRxiv*. 2022:2022.06.16.22276246. doi:10.1101/2022.06.16.22276246
125. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. *Elife*. Jan 30 2020;9doi:10.7554/eLife.48376
126. Bolli A, Di Domenico P, Pastorino R, Busby GB, Botta G. Risk of Coronary Artery Disease Conferred by Low-Density Lipoprotein Cholesterol Depends on Polygenic Background. *Circulation*. Apr 6 2021;143(14):1452-1454. doi:10.1161/CIRCULATIONAHA.120.051843

127. Natarajan P, Young R, Stitziel NO, et al. Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. *Circulation*. May 30 2017;135(22):2091-2101. doi:10.1161/CIRCULATIONAHA.116.024436
128. Tsao NL, Judy R, Levin MG, et al. Evaluation of the Performance of the RECODe Equation with the Addition of Polygenic Risk Scores for Adverse Cardiovascular Outcomes in Individuals with Type II Diabetes. *medRxiv*. May 5 2023;doi:10.1101/2023.05.03.23289457
129. Lee H, Choi J, Kim NY, et al. Earlier Age at Type 2 Diabetes Diagnosis Is Associated With Increased Genetic Risk of Cardiovascular Disease. *Diabetes Care*. May 1 2023;46(5):1085-1090. doi:10.2337/dc22-2144
130. Kwak SH, Hernandez-Cancela RB, DiCorpo DA, et al. Time-to-Event Genome-Wide Association Study for Incident Cardiovascular Disease in People with Type 2 Diabetes Mellitus. *medRxiv*. Jul 28 2023;doi:10.1101/2023.07.25.23293180
131. Pain O, Glanville KP, Hagenaars SP, et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. *PLoS Genet*. May 2021;17(5):e1009021. doi:10.1371/journal.pgen.1009021
132. Sun L, Wang Z, Lu T, Manolio TA, Paterson AD. eXclusionarY: 10 years later, where are the sex chromosomes in GWASs? *Am J Hum Genet*. Jun 1 2023;110(6):903-912. doi:10.1016/j.ajhg.2023.04.009
133. Fahed AC, Aragam KG, Hindy G, et al. Transethnic Transferability of a Genome-Wide Polygenic Score for Coronary Artery Disease. *Circ Genom Precis Med*. Feb 2021;14(1):e003092. doi:10.1161/CIRCGEN.120.003092
134. Metspalu M, Romero IG, Yunusbayev B, et al. Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am J Hum Genet*. Dec 9 2011;89(6):731-44. doi:10.1016/j.ajhg.2011.11.010
135. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. Apr 2019;51(4):584-591. doi:10.1038/s41588-019-0379-x
136. Wang Y, Guo J, Ni G, Yang J, Visscher PM, Yengo L. Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat Commun*. Jul 31 2020;11(1):3865. doi:10.1038/s41467-020-17719-y
137. Breedon JR, Marshall CR, Giovannoni G, et al. Polygenic risk score prediction of multiple sclerosis in individuals of South Asian ancestry. *Brain Commun*. 2023;5(2):fcad041. doi:10.1093/braincomms/fcad041
138. Ho WK, Tan MM, Mavaddat N, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun*. Jul 31 2020;11(1):3833. doi:10.1038/s41467-020-17680-w
139. Hackshaw A. Small studies: strengths and limitations. *Eur Respir J*. Nov 2008;32(5):1141-3. doi:10.1183/09031936.00136408
140. Serdar CC, Cihan M, Yucel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochem Med (Zagreb)*. Feb 15 2021;31(1):010502. doi:10.11613/BM.2021.010502
141. Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. May 2013;14(5):365-76. doi:10.1038/nrn3475

142. Klarin D, Natarajan P. Clinical utility of polygenic risk scores for coronary artery disease. *Nat Rev Cardiol*. May 2022;19(5):291-301. doi:10.1038/s41569-021-00638-w
143. Riveros-Mckay F, Weale ME, Moore R, et al. Integrated Polygenic Tool Substantially Enhances Coronary Artery Disease Prediction. *Circ Genom Precis Med*. Apr 2021;14(2):e003304. doi:10.1161/CIRCGEN.120.003304
144. Damask A, Steg PG, Schwartz GG, et al. Patients With High Genome-Wide Polygenic Risk Scores for Coronary Artery Disease May Receive Greater Clinical Benefit From Alirocumab Treatment in the ODYSSEY OUTCOMES Trial. *Circulation*. Feb 25 2020;141(8):624-636. doi:10.1161/CIRCULATIONAHA.119.044434
145. Marston NA, Kamanu FK, Nordio F, et al. Predicting Benefit From Evolocumab Therapy in Patients With Atherosclerotic Disease Using a Genetic Risk Score: Results From the FOURIER Trial. *Circulation*. Feb 25 2020;141(8):616-623. doi:10.1161/CIRCULATIONAHA.119.043805
146. Hindy G, Aragam KG, Ng K, et al. Genome-Wide Polygenic Score, Clinical Risk Factors, and Long-Term Trajectories of Coronary Artery Disease. *Arterioscler Thromb Vasc Biol*. Nov 2020;40(11):2738-2746. doi:10.1161/ATVBAHA.120.314856
147. Patel AP, Wang M, Ruan Y, et al. A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nat Med*. Jul 2023;29(7):1793-1803. doi:10.1038/s41591-023-02429-x
148. Elliott J, Bodinier B, Bond TA, et al. Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *JAMA*. Feb 18 2020;323(7):636-645. doi:10.1001/jama.2019.22241
149. Aragam KG, Dobbyn A, Judy R, et al. Limitations of Contemporary Guidelines for Managing Patients at High Genetic Risk of Coronary Artery Disease. *J Am Coll Cardiol*. Jun 9 2020;75(22):2769-2780. doi:10.1016/j.jacc.2020.04.027
150. Saadatagah S, Naderian M, Dikilitas O, Hamed ME, Bangash H, Kullo IJ. Polygenic Risk, Rare Variants, and Family History. *JACC: Advances*. 2023;2(7):100567. doi:doi:10.1016/j.jacadv.2023.100567
151. Ramirez J, van Duijvenboden S, Young WJ, et al. Prediction of Coronary Artery Disease and Major Adverse Cardiovascular Events Using Clinical and Genetic Risk Scores for Cardiovascular Risk Factors. *Circ Genom Precis Med*. Oct 2022;15(5):e003441. doi:10.1161/CIRCGEN.121.003441
152. Gupta R. Genetics-based risk scores for prediction of premature coronary artery disease. *Indian Heart J*. Aug 24 2023;doi:10.1016/j.ihj.2023.08.003
153. Mujwara D, Henno G, Vernon ST, et al. Integrating a Polygenic Risk Score for Coronary Artery Disease as a Risk-Enhancing Factor in the Pooled Cohort Equation: A Cost-Effectiveness Analysis Study. *J Am Heart Assoc*. Jun 21 2022;11(12):e025236. doi:10.1161/JAHA.121.025236
154. Canadian Chronic Disease Surveillance System (2021).
155. Widen E, Junna N, Ruotsalainen S, et al. How Communicating Polygenic and Clinical Risk for Atherosclerotic Cardiovascular Disease Impacts Health Behavior: an Observational Follow-up Study. *Circ Genom Precis Med*. Apr 2022;15(2):e003459. doi:10.1161/CIRCGEN.121.003459
156. Knowles JW, Zarafshar S, Pavlovic A, et al. Impact of a Genetic Risk Score for Coronary Artery Disease on Reducing Cardiovascular Risk: A Pilot Randomized Controlled Study. *Front Cardiovasc Med*. 2017;4:53. doi:10.3389/fcvm.2017.00053

157. Kullo IJ, Jouni H, Austin EE, et al. Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). *Circulation*. Mar 22 2016;133(12):1181-8. doi:10.1161/CIRCULATIONAHA.115.020109
158. Age and sex, and type of dwelling data: Key results from the 2016 Census (2017).
159. Dikilitas O, Schaid DJ, Kosel ML, et al. Predictive Utility of Polygenic Risk Scores for Coronary Heart Disease in Three Major Racial and Ethnic Groups. *Am J Hum Genet*. May 7 2020;106(5):707-716. doi:10.1016/j.ajhg.2020.04.002
160. Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A. Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet*. Apr 1 2003;12(7):771-6. doi:10.1093/hmg/ddg088
161. Ding Y, Hou K, Xu Z, et al. Polygenic scoring accuracy varies across the genetic ancestry continuum. *Nature*. Jun 2023;618(7966):774-781. doi:10.1038/s41586-023-06079-4
162. Kim MS, Patel KP, Teng AK, Berens AJ, Lachance J. Genetic disease risks can be misestimated across global populations. *Genome Biol*. Nov 14 2018;19(1):179. doi:10.1186/s13059-018-1561-7
163. Onengut-Gumuscu S, Chen WM, Robertson CC, et al. Type 1 Diabetes Risk in African-Ancestry Participants and Utility of an Ancestry-Specific Genetic Risk Score. *Diabetes Care*. Mar 2019;42(3):406-415. doi:10.2337/dc18-1727
164. Cavazos TB, Witte JS. Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. *HGG Adv*. Jan 14 2021;2(1)doi:10.1016/j.xhgg.2020.100017
165. Marquez-Luna C, Loh PR, South Asian Type 2 Diabetes C, Consortium STD, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol*. Dec 2017;41(8):811-823. doi:10.1002/gepi.22083
166. Howson JMM, Zhao W, Barnes DR, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet*. Jul 2017;49(7):1113-1119. doi:10.1038/ng.3874
167. Mosley JD, Gupta DK, Tan J, et al. Predictive Accuracy of a Polygenic Risk Score Compared With a Clinical Risk Score for Incident Coronary Heart Disease. *JAMA*. Feb 18 2020;323(7):627-635. doi:10.1001/jama.2019.21782
168. Thanassoulis G, Peloso GM, Pencina MJ, et al. A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham Heart Study. *Circ Cardiovasc Genet*. Feb 1 2012;5(1):113-21. doi:10.1161/CIRCGENETICS.111.961342
169. de Vries PS, Kavousi M, Ligthart S, et al. Incremental predictive value of 152 single nucleotide polymorphisms in the 10-year risk prediction of incident coronary heart disease: the Rotterdam Study. *Int J Epidemiol*. Apr 2015;44(2):682-8. doi:10.1093/ije/dyv070
170. Khera AV, Chaffin M, Zekavat SM, et al. Whole-Genome Sequencing to Characterize Monogenic and Polygenic Contributions in Patients Hospitalized With Early-Onset Myocardial Infarction. *Circulation*. Mar 26 2019;139(13):1593-1602. doi:10.1161/CIRCULATIONAHA.118.035658
171. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. Nov 1 2017;186(9):1026-1034. doi:10.1093/aje/kwx246

172. Glessner JT, Li J, Desai A, et al. CNV Association of Diverse Clinical Phenotypes from eMERGE reveals novel disease biology underlying cardiovascular disease. *Int J Cardiol*. Jan 1 2020;298:107-113. doi:10.1016/j.ijcard.2019.07.058
173. Antiochos P, Marques-Vidal P, McDaid A, Waeber G, Vollenweider P. Association between parental history and genetic risk scores for coronary heart disease prediction: The population-based CoLaus study. *Atherosclerosis*. Jan 2016;244:59-65. doi:10.1016/j.atherosclerosis.2015.10.104
174. Adeyemo Aea. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat Med*. Nov 2021;27(11):1876-1884. doi:10.1038/s41591-021-01549-6

Appendices

Appendix A

Figure A1. MVP X-chromosome SNP replication in men of European ancestry in the UKB

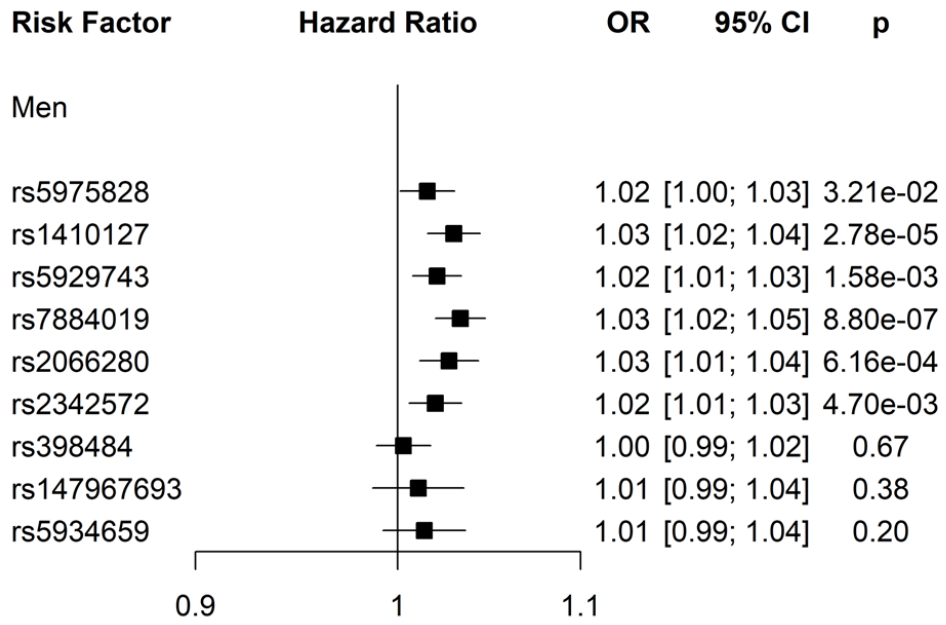
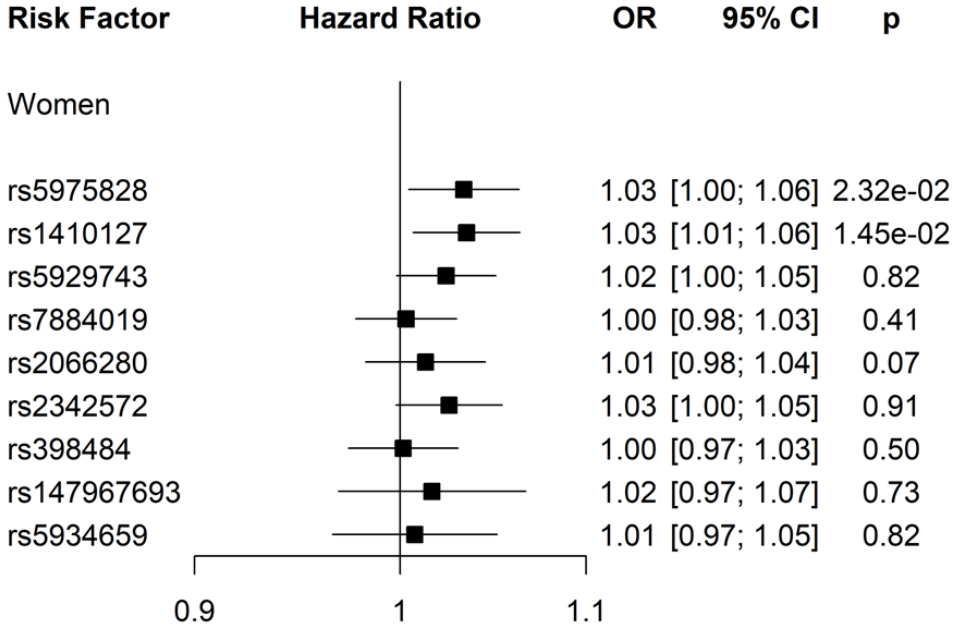


Figure A2. MVP X-chromosome SNP replication in women of European ancestry in the UKB



Appendix B

Table B1. MVP X-chromosome SNP replication among individuals of European ancestry in UKB

				MVP		UKB	
POS	ID	EA	NEA	OR	P value	OR	P value
X:135986549	rs5975828	T	C	1.03	$9.40\times E10^{-9}$	1.03	$4.48\times E10^{-04}$
X:67280381	rs1410127	C	T	1.02	$1.39\times E10^{-9}$	1.05	$2.61\times E10^{-06}$
X:135318977	rs5929743	A	G	1.02	$4.91\times E10^{-9}$	1.03	$3.03\times E10^{-04}$
X:109809489	rs7884019	A	C	1.03	$4.16\times E10^{-15}$	1.03	$2.73\times E10^{-04}$
X:80177630	rs2066280	A	T	1.03	$4.63\times E10^{-8}$	1.03	$2.36\times E10^{-03}$
X:84069371	rs2342572	C	T	1.02	$2.02\times E10^{-8}$	1.03	$1.01\times E10^{-03}$
X:77599469	rs398484	T	C	1.02	$1.59\times E10^{-8}$	1.00	0.83
X:153639255	rs147967693	T	C	1.04	$2.23\times E10^{-8}$	1.00	0.87
X:9578104	rs5934659	C	T	1.04	$5.78\times E10^{-9}$	1.02	0.24

POS, position; EA, effect allele; NEA, non-effect allele; OR, odds ratio. Cross-sectional CAD cases.

Table B2. Adjusted* associations of CAD GRS with incident CAD among individuals of European ancestry in the UK Biobank

		CAD	
WGRS (SD)	SNPs	HR (95%CI)	P value
GRS_{204}	204	1.37 (1.35, 1.40)	$<2.00 \times E10^{-16}$
$GRS_{MVP\ (no\ x\ chr)}$	249	1.36 (1.34, 1.38)	$<2.00 \times E10^{-16}$
$GRS_{MVP\ (x\ chr)}$	258	1.37 (1.35, 1.39)	$<2.00 \times E10^{-16}$

*age and sex adjusted.

Table B3. Adjusted* associations of CAD GRS with incident CAD in various ancestry subsets of the UK Biobank

			CAD	
WGRS (SD)	SNPs	subset	HR (95%CI)	P value
$GRS_{MVP(no\ x\ chr)}$	249	European	1.35 (1.33, 1.38)	$<2.00 \times E10^{-16}$
		Chinese	1.18 (0.76, 1.85)	0.47
		Black	1.03 (0.88, 1.18)	0.75
		South Asian	1.24 (1.13, 1.37)	1.73 × E10-05
$GRS_{MVP(x\ chr)}$	258	European	1.37 (1.34, 1.39)	$<2.00 \times E10^{-16}$
		Chinese	1.17 (0.74, 1.83)	0.50
		Black	1.01 (0.87, 1.18)	0.85
		South Asian	1.24 (1.12, 1.37)	2.44 × E10-05

*age and sex adjusted.

Appendix C

Permissions

ELSEVIER LICENSE
TERMS AND CONDITIONS

Oct 01, 2023

This Agreement between McGill University -- Justine Desrochers ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5591501391558
License date	Jul 17, 2023
Licensed Content Publisher	Elsevier
Licensed Content Publication	The American Journal of Human Genetics
Licensed Content Title	Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort
Licensed Content Author	Florian Privé,Hugues Aschard,Shai Carmi,Lasse Folkersen,Clive Hoggart,Paul F. O'Reilly,Bjarni J. Vilhjálmsson
Licensed Content Date	Jan 6, 2022
Licensed Content Volume	109
Licensed Content Issue	1
Licensed Content Pages	12
Start Page	12
End Page	23
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations

Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	Thesis Deposition - Desrochers
Institution name	McGill University
Expected presentation date	Sep 2023
Portions	Figure 3
Requestor Location	McGill University

3488 chemin cotes-des-neiges, apt 204

Montreal, QC H3H2M6

Canada

Attn: McGill University

Publisher Tax ID	GB 494 6272 12
Total	**0.00 CAD**

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Oct 01, 2023

This Agreement between McGill University -- Justine Desrochers ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5632570442198
License date	Sep 19, 2023
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Nephrology
Licensed Content Title	The immunology of atherosclerosis
Licensed Content Author	Anton Gisterå et al
Licensed Content Date	Apr 10, 2017
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Would you like a high resolution image with your order?	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Thesis Deposition - Desrochers
Institution name	McGill University
Expected presentation date	Sep 2023
Order reference number	Permission_thesis_2
Portions	Figure 1

Requestor Location	3488 chemin cotes-des-neiges, apt 204 Montreal, QC H3H2M6 Canada Attn: McGill University
Total	**0.00 CAD**

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Oct 01, 2023

This Agreement between McGill University -- Justine Desrochers ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5632570689271
License date	Sep 19, 2023
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Genetics
Licensed Content Title	Clinical use of current polygenic risk scores may exacerbate health disparities
Licensed Content Author	Alicia R. Martin et al
Licensed Content Date	Mar 29, 2019
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Would you like a high resolution image with your order?	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Thesis Deposition - Desrochers
Institution name	McGill University
Expected presentation date	Sep 2023
Order reference number	Permission_thesis_3
Portions	Figure 1

Requestor Location	3488 chemin cotes-des-neiges, apt 204 Montreal, QC H3H2M6 Canada Attn: McGill University
Total	**0.00 CAD**

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Oct 01, 2023

This Agreement between McGill University -- Justine Desrochers ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5640391195763
License date	Oct 01, 2023
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Methods Primers
Licensed Content Title	Genome-wide association studies
Licensed Content Author	Emil Uffelmann et al
Licensed Content Date	Aug 26, 2021
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Thesis Deposition - Desrochers
Institution name	McGill University
Expected presentation	Oct 2023

date

Order reference number	Permission 4
Portions	Figure 1, Figure 4
Requestor Location	McGill University

3488 chemin cotes-des-neiges, apt 204

Montreal, QC H3H2M6

Canada

Attn: McGill University

Total	**0.00 CAD**