Institutionen för Medicinsk Epidemiologi och Biostatistik

Lipid-related Genes and Their Associations with Coronary Heart Disease

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Atrium, Nobels väg 12 B

Fredagen den 16 maj, 2014, kl 09.00

av

Ci Song

Huvudhandledare:
Professor Erik Ingelsson
Uppsala Universitet
Institutionen för medicinska vetenskaper

Bihandledare:
Professor Nancy Pedersen
Karolinska Institutet
Institutionen för Medicinsk Epidemiologi och Biostatistik

Doktor Sara Hägg
Karolinska Institutet
Institutionen för Medicinsk Epidemiologi och Biostatistik

Fakultetsopponent:
Professor Marju Orho-Melander
Lunds Universitet
Institutionen för kliniska vetenskaper, Malmö

Betygsnämnd:
Professor Bertil Lindahl
Uppsala universitet
Institutionen för medicinska vetenskaper

Professor Kamila Czene
Karolinska Institutet
Institutionen för Medicinsk Epidemiologi och Biostatistik

Docent Nicola Orsini
Karolinska Institutet
Institutet för miljömedicin

Stockholm 2014

ABSTRACT

Coronary heart disease (CHD) is a complex disease caused by the long-term progression of atherosclerosis. Lipids, including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglycerides (TG) are well established biomarkers for CHD, and are used to predict individual 10-year risk for CHD both in the Framingham risk score and in the System for Cardiac Operative Risk Evaluation (SCORE) project. Exploring the role and relationship of lipid-related genes and lipids in CHD progression can enhance our understanding of CHD etiology. This PhD thesis was based on the following studies:

Study I explored the associations between lipid-related genes and myocardial infarction incidence. We found that a missense variant in the *ABCA1* gene was consistently associated with myocardial infarction in two Swedish cohorts, as well as being weakly associated with CHD in a large meta-analysis that included 173,975 individuals of European descent.

Study II investigated large-scale genetic associations between lipid fractions, including data from 188,577 individuals. In this study, there were 157 independent loci associated with at least one lipid trait at the genome-wide significant level (P-value $< 5 \times 10^{-8}$), of which 62 loci were not previously reported to be associated with lipids in humans.

Study III examined the modifying effect of lifestyles on the genetic variance of CHD using twin modeling. We found that a decrease in genetic variance of CHD was dependent on increasing body mass index (BMI). This finding indicates that more genetic variants or larger effect sizes of genetic variants influence CHD in the lean group compared to the obese group.

Study IV investigated the causal relations between lipids and chronic low-degree inflammation, measured as high sensitivity C reaction protein (CRP) using a Mendelian randomization study design. We found strong observational associations between dyslipidemia (decreased HDL-C and/or increased TG) and elevated CRP, but no evidence supporting causal relations between dyslipidemia and CRP. The observational associations are likely driven by other common causes such as adiposity.

In conclusion, the findings in the current thesis highlight genetic determinants of blood lipid levels and their associations with CHD. Our results suggest a BMI-gene interaction effect on CHD. Further, our results suggest that the observational associations between dyslipidemia and chronic low-degree inflammation are more likely to be driven by other common causes such as adiposity. All these findings shed light on the complex mechanisms underlying CHD, particularly those that act through lipid metabolism pathways.

From the Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

LIPID-RELATED GENES AND THEIR ASSOCIATIONS WITH CORONARY HEART DISEASE

Ci Song

宋 辞

Stockholm 2014

Published by Karolinska Institutet. Printed by Universitetsservice US-AB

ISBN 978-91-7549-576-7

ABSTRACT

Coronary heart disease (CHD) is a complex disease caused by the long-term progression of atherosclerosis. Lipids, including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglycerides (TG) are well established biomarkers for CHD, and are used to predict individual 10-year risk for CHD both in the Framingham risk score and in the System for Cardiac Operative Risk Evaluation (SCORE) project. Exploring the role and relationship of lipid-related genes and lipids in CHD progression can enhance our understanding of CHD etiology. This PhD thesis was based on the following studies:

Study I explored the associations between lipid-related genes and myocardial infarction incidence. We found that a missense variant in the *ABCA1* gene was consistently associated with myocardial infarction in two Swedish cohorts, as well as being weakly associated with CHD in a large meta-analysis that included 173,975 individuals of European descent.

Study II investigated large-scale genetic associations between lipid fractions, including data from 188,577 individuals. In this study, there were 157 independent loci associated with at least one lipid trait at the genome-wide significant level (P-value $< 5 \times 10^{-8}$), of which 62 loci were not previously reported to be associated with lipids in humans.

Study III examined the modifying effect of lifestyles on the genetic variance of CHD using twin modeling. We found that a decrease in genetic variance of CHD was dependent on increasing body mass index (BMI). This finding indicates that more genetic variants or larger effect sizes of genetic variants influence CHD in the lean group compared to the obese group.

Study IV investigated the causal relations between lipids and chronic low-degree inflammation, measured as high sensitivity C reaction protein (CRP) using a Mendelian randomization study design. We found strong observational associations between dyslipidemia (decreased HDL-C and/or increased TG) and elevated CRP, but no evidence supporting causal relations between dyslipidemia and CRP. The observational associations are likely driven by other common causes such as adiposity.

In conclusion, the findings in the current thesis highlight genetic determinants of blood lipid levels and their associations with CHD. Our results suggest a BMI-gene interaction effect on CHD. Further, our results suggest that the observational associations between dyslipidemia and chronic low-degree inflammation are more likely to be driven by other common causes such as adiposity. All these findings shed light on the complex mechanisms underlying CHD, particularly those that act through lipid metabolism pathways.

LIST OF PUBLICATIONS

This thesis is based on the following original articles, which will be referred to in the text by their Roman numerals (I-IV)

I. **Song C**, Pedersen NL, Reynolds CA, Sabater-Lleal M, Kanoni S, Willenborg C; CARDIoGRAMplusC4D Consortium, Syvänen AC, Watkins H, Hamsten A, Prince JA, Ingelsson E. Genetic variants from lipid-related pathways and risk for incident myocardial infarction. *PLoS One*. 2013;8(3):e60454.

II. Global Lipids Genetics Consortium. Discovery and Refinement of Loci Associated with Lipid Levels. *Nat Genet*. 2013 Nov;45(11):1274-83. doi: 10.1038/ng.2797.

III. **Song C**, Chang Z, Magnusson PK, Ingelsson E, Pedersen NL. The role of lifestyle in moderating heritable factors in coronary heart disease. *J Intern Med*. 2013 Dec 12. doi: 10.1111/joim.12177.

IV. **Song C**, Fall T, Ploner A, Lind L, Lannfelt L, Pedersen NL, Hägg S, Ingelsson E. The interplay between fasting blood lipids and inflammation: a Mendelian randomization study. *Manuscript.*

RELATED PUBLICATIONS

(not included in thesis)

I. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert MF, Benazzo A, Rybin D, Jackson AU, Stringham HM, **Song C**, … , Ingelsson E, McCarthy MI, Prokopenko I; on behalf of the MAGIC investigators. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes*. 2013 Dec 2.

II. Do R, Willer CJ, Schmidt EM, …, **Song C**,…, Daly MJ, Neale BM, Kathiresan S. Polymorphisms associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet.* 2013 Oct 6. doi: 10.1038/ng.2795.

III. den Hoed M, Eijgelsheim M, Esko T, …, **Song C**, …, Milan DJ, Snieder H, Samani NJ, Loos RJ. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet*. 2013 Apr 14;45(6):621-31.

IV. CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013 Jan;45(1):25-33.

V. Debette S, Visvikis-Siest S, Chen MH, Ndiaye NC, **Song C**, Destefano A, Safa R, Azimi Nezhad M, Sawyer D, Marteau JB, Xanthakis V, Siest G, Sullivan L, Pfister M, Smith H, Choi SH, Lamont J, Lind L, Yang Q, Fitzgerald P, Ingelsson E, Vasan RS, Seshadri S. Identification of cis- and trans-Acting Genetic Variants Explaining Up to Half the Variation in Circulating Vascular Endothelial Growth Factor Levels. *Circ Res*. 2011 Aug 19;109(5):554-63.

VI. Kilpeläinen TO, Zillikens MC, Stančákova A, …, **Song C** … Fox CS, Laakso M, Loos RJ. Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nature Genetics*. 2011 Jun 26;43(8):753-60.

CONTENTS

1 Introduction 1
1.1 Coronary heart disease 1
1.1.1 Current burden and distribution of coronary heart disease 1
1.1.2 Genetic determinants of coronary heart disease 1
1.1.3 Risk factors for coronary heart disease 2
1.2 Lipids and coronary heart disease 3
1.2.1 Lipid metabolism 3
1.2.2 Utility of lipid measurements in coronary heart disease 5
1.2.3 Lipids and inflammation 5
1.3 Lifestyle intervention for coronary heart disease 6
2 Aims 8
3 Study samples 9
3.1 Swedish Twin Registry 9
3.1.1 Cohort I and Cohort II 9
3.1.2 GOSH 9
3.1.3 TwinGene 9
3.2 ULSAM 10
3.3 PIVUS 10
3.4 Follow-up and coronary heart disease definition 10
3.4.1 National Patient Register 10
3.4.2 Cause of Death Register 10
3.4.3 Coronary heart disease definition 10
3.5 Blood biomarker measurements 11
3.6 Genetic markers 11
3.6.1 GOLD chip 11
3.6.2 Metabochip 11
3.6.3 Genome-wide arrays 12
3.6.4 Quality control for genotype data 12
4 Study designs and methods 13
4.1 Study I 13
4.2 Study II 13
4.3 Study III 15
4.4 Study IV 16
5 Study summaries 18
5.1 Study I 18
5.2 Study II 19
5.3 Study III 20
5.4 Study IV 22
6 Discussion 25
6.1 Findings and implications 25
6.1.1 *ABCA1* gene and incident MI 25
6.1.2 Genetic determinants of lipid metabolism 25
6.1.3 BMI-gene interaction effect on CHD 26
6.1.4 Causal relationships between blood lipids and CRP 26
6.2 Methodological considerations 27
6.2.1 Left truncation when using genetic data in survival models 27

6.2.2 Twin modeling addressing gene-environment interaction 27
6.2.3 Causal inference using Mendelian randomization 28
7 Conclusions 30
8 Future perspectives 31
8.1 Missing heritability 31
8.1.1 More common SNPs with very modest effects 31
8.1.2 Rare variants 31
8.1.3 Interaction effects 32
8.2 From association to function in the post-GWAS era 32
8.2.1 Fine-mapping of known loci 32
8.2.2 Functional annotation 32
8.2.3 Experimental models to understand gene function 33
9 Acknowledgements 34
10 References 36

LIST OF ABBREVIATIONS

A	Additive genetic variance
ABI	Ankle-brachial index
ApoB	Apolipoprotein B
BMI	Body mass index
C	Shared environmental variance
CABG	Coronary artery bypass graft
CAC	Coronary artery calcification
CARDIoGRAMplusC4D	Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) genetic consortium
CDR	Cause of Death Register
CETP	Cholesteryl ester transfer protein
CHD	Coronary heart disease
CI	Confidence interval
CNV	Copy number variants
CRP	C-reactive protein
CVD	Cardiovascular disease
DALY	Disability-adjusted life year
DBP	Diastolic blood pressure
DNA	Deoxyribonucleic acid
E	Non-shared environmental variance
FMD	Flow-mediated dilation
1000G	1000 Human Genome Project
GPR146	G protein–coupled receptor 146
GWA	Genome-wide association
GWAS	Genome-wide association study
GRS	Genetic risk score
HDL	High-density lipoprotein
HDL-C	High-density lipoprotein cholesterol
HMG-CoA	Hydroxy-3-methyglutaryl coenzyme A
HR	Hazard ratio
ICD	International Classification of Diseases
IDL	Intermediate-density lipoprotein
IL-6R	Interleukin-6 receptor
IMT	Intima-media thickness
IV	Instrumental variable
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LDL-C	Low-density lipoprotein cholesterol
LPL	Lipoprotein lipase
MAF	Minor allele frequency
MI	Myocardial infarction
MMP	Metalloproteinaseses
MR	Mendelian randomization
NIH	National Institutes of Health
NPR	National Patient Register

OR	Odds ratio
PTCA	Percutaneous transluminal coronary angioplasty
SALT	Screening Across the Lifespan Twin Study
SBP	Systolic blood pressure
SCORE	System for Cardiac Operative Risk Evaluation
SD	Standard deviation
2SLS	Two-stage least squares
SNP	Single-nucleotide polymorphism
STR	Swedish Twin Registry
T2D	Type 2 diabetes
TG	Triglycerides
VLDL	Very low-density lipoprotein

1 INTRODUCTION

1.1 CORONARY HEART DISEASE

1.1.1 Current burden and distribution of coronary heart disease

Coronary heart disease (CHD), also known as coronary artery disease (CAD) or ischemic heart disease (IHD), is caused by reduced blood flow in narrowing coronary arteries. CHD is a long-term consequence of the atherosclerotic process.

CHD is the leading cause of morbidity and mortality worldwide.[1] In 2011, the total number of deaths due to CHD was estimated at 7 million.[2] The Global Burden of Disease Study projections indicate that CHD will continue to be the global leading cause of death up to at least the year 2030.[3] However, in recent years, the age-standardized mortality rate for CHD has decreased in many developed countries, while it has increased in developing countries. It is estimated that developing countries will contribute 71% (7.8 of 11.1 million) of total mortality due to CHD in 2020, compared to 56% (3.5 of the 6.2 million) in 1990.[4] In addition, non-fatal CHD events contribute significantly to overall disability, estimated by disability-adjusted life years (DALYs). By 2020, the global burden of CHD is projected to cause 82 million DALYs.[4] Population-specific projections have indicated that the numbers of DALYs are expected to increase in India and China.[4] By contrast, the number of DALYs is expected to decrease in countries with established market economies. The increased mortality rates and DALYs in developing countries have been suggested to primarily be due to changes in demography, such as urbanization, and lifestyle changes.[4]

1.1.2 Genetic determinants of coronary heart disease

Familial aggregation of CHD suggests evidence of a genetic predisposition. Individuals with parental cardiovascular disease (CVD) have approximately a 2-fold greater risk of CHD,[5, 6] while even greater risk is exhibited in those with affected siblings.[7] Such families represent 72% of early onset CHD cases and 48% of cases at all ages.[8] Twin studies further elucidate the familiar aggregation of CHD into genetic influences and shared environmental influences, such as social economic status and dietary habits. Previous twin studies reported 40% heritability of CHD mortality in Swedish populations,[9] and similar heritability of CHD in other Caucasian European ancestry populations.[7] In addition, twin studies also suggest a very low shared environmental influence on CHD. In agreement with these studies, adoptee investigations show that adoptees have a similar risk of CHD mortality as their biological parents, but not as their adoptive parents.[10]

Candidate gene association studies for CHD have focused mainly on variants in known pathological pathways, such as hyperlipidemia or inflammation. Although over 5,000 publications have identified candidate genes implicated in CHD, very few findings have been robust, mainly due to inadequate sample sizes, between-study heterogeneity and lack of proper replication. Candidate gene studies for CHD have been previously comprehensively reviewed elsewhere.[11]

Over the past 7 years, genome-wide association studies (GWASs) have identified genetic variants related to CHD using a hypothesis-free design. Recently, the total number of susceptibility loci for CHD has updated to 46, based on joined analyses from

the Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) genetics consortium (CARDIoGRAMplusC4D).[12] This study included 34 independent samples with patients of European or South Asian descent and comprised 41,513 cases and 65,919 controls. Among the leading 46 CHD-related loci, 12 were associated with at least one lipid trait reaching GWA significance level, 5 with blood pressure, but none with type 2 diabetes (T2D).[12] These 46 loci, together with another 104 loci at a 5% false discovery rate, explain 10.6% of CHD heritability. Network analysis based on candidate loci using a 10% false discovery rate mapped the most significant pathways to lipid metabolism and inflammation. In addition to the loci found in Europeans, a GWAS in Han Chinese subjects revealed four novel loci previously not observed in Europeans,[13] two of which were reported to be associated with hypertension in Europeans.[13]

All CHD-related loci based on GWASs are described in detail in their respective original studies and a statement from the American Heart Association.[12-22]

1.1.3 Risk factors for coronary heart disease

The 10-year individual risk for CHD can be estimated by Framingham risk score[23] and System for Cardiac Operative Risk Evaluation (SCORE) risk charts[24] based on multiple traditional risk factors, including age, sex, ethnicity, smoking, blood pressure, blood cholesterol, and diabetes.[25] In addition, obesity and lifestyle factors, including physical activity, harmful use of alcohol and unhealthy diet (e.g., high cholesterol, high saturate fatty acids, high salt, or high sugar), are also well established traditional risk factors for CHD. Population attributable risk calculations show that traditional risk factors can explain 75–85% of CHD occurrence in the United States.[25]

Widespread concerns over CHD have directed focus towards the identification of novel risk factors for this disease. These novel risk factors can enhance our understanding of atherosclerosis pathology, providing intervention targets for CHD treatment and acquiring more accurate and reliable CHD risk assessments, particularly for the intermediate risk group that carry a 10–20% CHD risk.[26] Epidemiological studies have identified novel blood and urine biomarkers associated with CHD. For example, increased circulating inflammatory marker C-reactive protein (CRP) is associated with CHD risk.[27-30] Furthermore, CRP improves the risk reclassification for intermediate risk individuals (10–20% CHD risk).[29, 30] However, recent studies suggest that CRP is a biomarker rather than a causal risk factor for CHD.[31] In addition, circulating lipoprotein (a) (Lp[a]), which is a protein that is attached to low-density lipoprotein (LDL) particles, is positively associated with CHD in perspective studies[32] and it was proposed as a causal factor based on genetic studies.[33-35]

Furthermore, structural and functional changes in subclinical atherosclerosis are associated with increased CHD risk, including carotid intima-media thickness (IMT), carotid plaques, brachial artery flow-mediated dilation (FMD), and coronary artery calcification (CAC) score.[36-39] A systematic review reported that carotid IMT, carotid plaques, and CAC can also improve risk reclassification of the intermediate risk group.[40]

The American College of Cardiology and the American Heart Association jointly updated the risk assessment guidelines for atherosclerotic CVD in 2013.[41] After reviewing publications from 2008 to 2013 for various novel risk factors, this report recommends that one or more of the following risk factors can be considered in the treatment decision making process when there is no definitive decision based on quantitative risk assessment: family history, CRP, CAC score, and ankle-brachial index (ABI).[41]

1.2 LIPIDS AND CORONARY HEART DISEASE

1.2.1 Lipid metabolism

The term "blood lipids" mainly refers to the levels of cholesterol and triglycerides (TG) in the blood. There are three pathways for maintaining circulating blood lipids levels: 1) the exogenous pathway for delivering dietary lipids; 2) the endogenous pathway for delivering lipids synthesized in the liver; and 3) the reverse cholesterol transport pathway for transferring cholesterol in peripheral tissues back to the liver.

1.2.1.1 Exogenous pathway

In the exogenous pathway, dietary fats and cholesterol are absorbed and packaged together with apolipoprotein B-48 (shown as B-48 in **Figure 1**) to form nascent chylomicrons in intestines. Nascent chylomicrons are secreted into the blood and become mature chylomicrons by combining with apolipoprotein C-II (C-II in **Figure 1**) and apolipoprotein E (E in **Figure 1**) donated from high-density lipoprotein (HDL) particles. Circulating chylomicrons are hydrolyzed by lipoprotein lipase (LPL), which release fatty acids and glycerol into peripheral tissues to be used for energy (e.g., in muscle tissue) and storage (e.g., in adipose tissue). Chylomicron remnants are removed from the blood within 8–10 hours after food intake, via chylomicron remnant receptors expressed in the liver (**Figure 1**).

1.2.1.2 Endogenous pathway

In the endogenous pathway, TG and cholesterol esters are synthesized in the liver and packaged to form very low-density lipoproteins (VLDLs), which are secreted into the blood. Circulating VLDLs release fatty acids and glycerol into peripheral tissues in a manner similar to chylomicrons in the exogenous pathway. Next, some of the VLDL remnants are removed from the blood via LDL receptors in the liver. The remaining VLDL remnant particles become intermediate-density lipoproteins (IDLs), which are smaller and denser than VLDLs. The IDL particles can either be taken up in liver by LDL receptors or hydrolyzed in the liver by hepatic-triglyceride lipase (HL in **Figure 1**) to form LDLs.

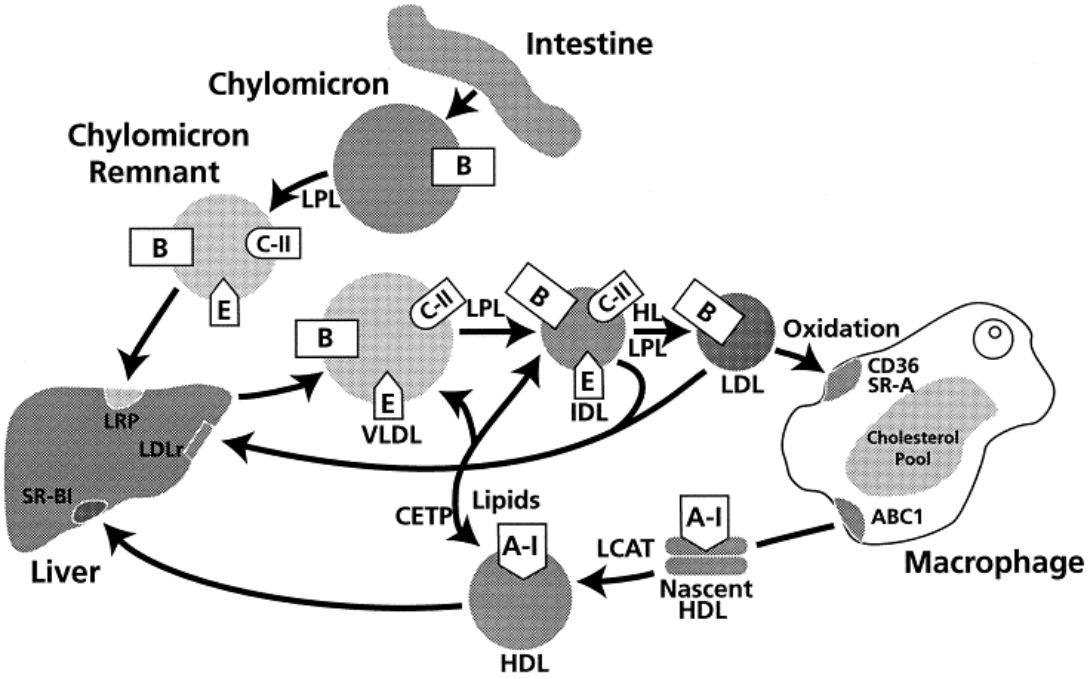
LDLs are the main vehicle for circulating cholesterol to peripheral tissues, forming LDL cholesterol (LDL-C). Most of the LDL particles are taken up by LDL receptors in the liver; the remaining LDLs are removed by way of scavenger pathways at the cellular level. As LDL is taken up by receptors, free cholesterol is released and accumulates within the cells. LDL receptor activity and LDL uptake regulate plasma LDL concentration by several mechanisms, including decreasing the synthesis of hydroxy-3-methyglutaryl coenzyme A (HMG-CoA) reductase (which controls the rate of cholesterol synthesis), suppressing the synthesis of new LDL receptors in cells, and

activating the enzyme acyl-coenzyme A cholesterol acyltransferase, which esterifies free cholesterol into cholesterol esters, thereby storing cholesterol in the cell.

1.2.1.3 Reverse cholesterol transport pathway

HDL particles carry cholesterol from atherosclerotic lesions, which is an important process of preventing atherosclerosis, and HDL cholesterol (HDL-C) is therefore considered the "good" cholesterol. Nascent HDL is primarily synthesized in the liver and intestine, and is composed of phospholipids, apolipoprotein A-1, and very little cholesterol. Nascent HDL captures free cholesterol in peripheral tissues, carries it back to the liver, and becomes a mature HDL particle by activating lecithin-cholesterol acyl transferase (LCAT) and apolipoprotein A-1 (A-1 in **Figure 1**).

Figure 1. *Lipid metabolism and lipoproteins involved in this process (adapted from Kwiterovich et al, 2000).*



Multiple proteins and enzymes are involved in regulating lipid metabolism and maintaining blood lipid levels within a physiological range. Mutations in these regulators can disrupt lipid metabolism and cause cholesterol accumulation. Previous GWASs for lipid traits reported 95 loci associated with at least one of total cholesterol, HDL-C, LDL-C, and TG during fasting status.[42] Multiple loci had been observed in the aforementioned study, which were nearby/in genes known to be involved in lipid regulation, such as *APOB, APOE, CEPT, LCAT, LDLR, LPL*. The targets of lipid-lowering drug targets were seen as well, such as *HMGCR* for statins.[42] Of the 95 reported loci, 59 had not previously been associated with the level of circulating lipids. In total, these 95 loci explain 10%–20% of total variance for each lipid fraction, which represents 25%–30% of the genetic variance. Genetic studies using larger sample sizes are required to reveal more loci associated with lipids and to uncover functional single-nucleotide polymorphisms (SNPs) in known loci.

1.2.2 Utility of lipid measurements in coronary heart disease

Lipid measurements are essential both for risk prediction in the primary prevention of CVD and to identify treatment targets in the primary and secondary prevention of CHD. Elevated total cholesterol, LDL-C, and reduced HDL-C levels have all been shown to be strong independent risk factors for CHD in numerous epidemiological studies.

Statins are the first line of treatment for most people with high cholesterol, and they have been shown in multiple research studies to reduce both CHD incidence and mortality.[43-45] They are HMG-CoA reductase inhibitors that suppress cholesterol biosynthesis, and are currently the most commonly prescribed lipid-lowering medication. Statins primarily operate by lowering LDL-C levels, but they also cause a slight raise in HDL-C and a slight decrease in TG levels. In recent updates from ACC/AHA Blood Cholesterol Guidelines, extended use of statins for both primary and secondary prevention of CHD is recommended. In addition, four major patient groups that potentially benefit from statins were identified: 1) patients with clinical CVD (recommended high-intensity statin therapy); 2) individuals with familiar hyperlipidemia who have LDL-C ≥190 mg/dL (4.9 mmol/L); 3) diabetic individuals with LDL-C levels between 70 and 189 mg/dL (1.8-4.9 mmol/L) and non-clinical CVD; and 4) non-diabetic individuals with a 10-year risk for CVD ≥7.5% and non-clinical CVD.[46]

Elevating HDL-C level has been considered a secondary target of lipid modifying pharmaceuticals for CHD prevention in the past decade. However, several drugs that aimed to elevate HDL-C level subsequently failed to prevent CHD.[47] Furthermore, recent genetic studies indicate that HDL-C is not a causal factor for CHD.[48, 49] Some studies have shown that the HDL particle function is a better biomarker for HDL than HDL-C. Further studies are required to identify the causal effect of HDL particles on CHD.

Although overwhelming evidence already supports the theory that abnormal lipid metabolism can affect the atherosclerotic process, much previous research regarding lipid metabolism primarily focused on blood total cholesterol, HDL-C, LDL-C, and TG levels. More recently, mass spectrometry-based methods that comprehensively annotate lipid species (via lipidomics) and other metabolites (via metabolomics) provide a powerful platform for identifying novel lipid markers of CHD risk. Furthermore, these technique will increase our understanding of lipid metabolic pathways involved in atherosclerosis pathogenesis, which will facilitate the development of new treatments.[50]

1.2.3 Lipids and inflammation

Both blood lipids and inflammation are important risk factors for CHD. The relations between lipids and inflammation have been discussed in recent years. There is some evidence that suggests that lipids play a role in inflammation. Previous studies have shown that high cholesterol diets, as well as saturated fatty acid consumption, are associated with increased CRP levels, which is known as metabolic inflammation.[51] In addition, transcription factors sensing circulating lipids can regulate the expression of genes involved in inflammatory pathways, including peroxisome proliferator-activated receptors,[52, 53] liver X receptors,[54] and farnesoid X receptors.[55] On the other hand, some

studies suggest that inflammation alters lipid metabolism. For instance, lipid metabolism is modified during acute phase inflammation. Chronic inflammatory diseases like dermatomyositis,[56] rheumatoid arthritis,[57] and systemic lupus erythematosus[58] are often accompanied by reduced HDL-C and elevated TG levels. In addition, inflammation has been suggested to decrease toxicity through modifying lipid metabolism, which represents a highly conserved response aimed at tissue repair in evolution.[59] There are also reasons to believe the third explanation, i.e. common causes for chronic inflammation and dyslipidemia. Mendelian randomization (MR) study has shown body mass index (BMI) causally affect both lipids and CRP.[60]

1.3 LIFESTYLE INTERVENTION FOR CORONARY HEART DISEASE

The National Institutes of Health (NIH) in the United States suggests that CHD can be prevented through a healthy lifestyle, including the following items: following a healthy diet, being physically active, maintaining a healthy weight, quitting smoking, and managing stress.[55] Lifestyle factors, including smoking, physical inactivity, harmful use of alcohol, unhealthy diet, and high BMI are essential risk factors for CHD as well as stroke.

Prospective studies consistently show lower CHD mortality in former smokers than in current smokers. Furthermore, CHD risk is reduced relatively quickly after smoking cessation.[61, 62] Patients who quit smoking after initial CHD diagnosis experienced a 50% reduced risk of CHD occurrence and total mortality,[63, 64] even in elderly patients.[65] Studies in both animals and humans have shown that smoking induces endothelium injury, which causes increased platelet activation and enhanced thrombus growth and propagation.[2, 66-68]

Physical activity has an independent protective effect on CHD.[69-75] When using category/quantitative measurements of physical activity, a dose effect of physical activity on CHD protection was also reported.[75] Any amount of activity for people with a sedentary lifestyle is encouraged and additional health benefits are acquired with more active exercise.

In a recent systematic review, it was reported that all levels of alcohol consumption above 2.5 g per day reduced the risk of incident CHD, but a greater risk of cardiovascular mortality and stroke incidence amongst those who consumed more than 60 g per day was also observed.[76]

Individuals who were overweight or obese had an increased risk of CHD compared to those of healthy bodyweight.[77-80] A pooled analysis based on 97 prospective cohorts reported that each 5 kg/m^2 increase in BMI was associated with a 27% increased hazard for CHD.[80] Although weight loss is potentially beneficial for CHD prevention, medication for lowering weight over the long-term was not successful,[81-83] and weight loss surgery is only recommended for extremely obese people.[84, 85] A recent clinical trial reported that physical activity intervention aimed at weight loss did not reduce CHD risk in overweight/obese patients with T2D.[86]

Taking together, lifestyle intervention studies show that CHD is partially preventable. There are a few studies have reported single genotype–environment interaction effects

on CHD, i.e., *APOE*- smoking,[87] *FTO*- physical inactivity,[88] *ADH1C*- alcohol consumption[89] interaction effects on CHD risk. However, few studies have addressed the overall effects of various lifestyle factors in modifying the role of genetic effects on CHD. Understanding gene-lifestyle interactions and CHD is critical to provide appropriate lifestyle intervention guidelines for specific groups.

2 AIMS

The overall aim of this thesis is to explore the role of lipid-related genes in CHD incidence. The specific aims include:

I. To investigate associations between lipid-related genetic variants and incident myocardial infarction (MI) using a candidate gene study design;

II. To investigate genetic variants associated with lipids using a genome-wide association design;

III. To investigate gene-environmental interaction effects on CHD using a twin study design;

IV. To investigate causal relationships between different lipid fractions and inflammation using a bi-directional MR study design.

3 STUDY SAMPLES

3.1 SWEDISH TWIN REGISTRY

The Swedish Twin Register (STR) is a nationwide twin register in Sweden. At present, it contains information on more than 194,000 twins born between 1886 and 2008. Over the past 10 years, DNA information was obtained from 45,000 twins, 15,000 of which also provided blood serum measurements. Thus, the STR is a powerful resource for molecular epidemiological research.[90] In the current thesis, we used the oldest cohorts of the STR to investigate CHD. Detailed information regarding each cohort within the STR is described below.

3.1.1 Cohort I and Cohort II

Cohort I is a birth cohort including all same-sex twin pairs born between 1886 and 1925. Several questionnaires were sent to these twins between 1960 and 1970,[91] which contained questions regarding the following information: twin similarity, marital status, urban or rural living, smoking, height and weight, education, dental status, family, diseases status, and medication.

Cohort II is a birth cohort including all same-sex twin pairs born between 1926 and 1958. A mailed questionnaire was sent to those individuals in 1972 to 1973;[91] similar information to Cohort I was collected, and questions regarding personality were also included. Cohort I and II were used in Study III.

3.1.2 GOSH

GENDER, OCTO-twin, SATSA, HARMONY, collectively known as GOSH, is a combination of four STR sub-cohorts. GENDER (Sex differences in health and aging) is a study of unlike-sexed twins born during 1906 to 1925.[92] Participant information was collected using questionnaires, health assessments, cognitive testing, and blood sampling in six cycles. OCTO-Twin (Individual differences among the oldest-old) is a study of the etiology of individual differences among the oldest-old.[93] All 80-year-old and older twin pairs who were both alive at study initiation in Sweden were recruited to participate. The study was completed with five waves of data collected at 2-year intervals. SATSA (Swedish Adoption and Twin Study of Aging) was initiated in 1984 and included approximately 150 twin pairs reared apart and 150 twin pairs reared together.[94] The first three waves of the study were conducted at 3-year intervals with the fourth in 1999, the fifth in 2002, and the sixth in 2005. HARMONY (the Study of Dementia in Swedish Twins) was initiated to evaluate the role of genetic and environmental factors on dementia.[95] All twins from the STR aged 65 and older were invited to participate in a computer-assisted telephone screening interview in 1998, from which 1,488 twins not identified in earlier studies updated their clinical diagnoses in the HARMONY study. There were a total of 2,618 individuals from GOSH with blood samples, which we used in Study I and II.

3.1.3 TwinGene

Between 2004 and 2008, a total of 12,614 twins from the STR participated in the TwinGene project. All individuals answered a mailed health questionnaire. In addition, they also underwent a health checkup and gave blood samples in a local healthcare facility. More extensive phenotypic information from these individuals can also be

obtained from questionnaires implemented in the Screening Across the Lifespan Twin Study (SALT). In addition, most same-sex twins in this project were from Cohort II, which provides rich phenotypic information obtained during their youth and during middle age. TwinGene data was used in Study II and IV.

3.2 ULSAM

Uppsala Longitudinal Study of Adult Men (ULSAM) is a community-based longitudinal study of unrelated individuals.[96] All men born between 1920 and 1924 who were residents of Uppsala County, Sweden were invited to a health examination. All men participated in the investigation at 50 years of age, and they have been re-examined five times to date. Blood samples were collected during re-examination at age 71. The ULSAM data was used in Study I, II, and IV.

3.3 PIVUS

Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) is another community-based longitudinal study of unrelated individuals.[97] A random sample of 1,016 subjects aged 70 years living in the community of Uppsala, Sweden participated in this study between 2001 and 2004. Basic information, blood samples, and comprehensive vasculature measurements were recorded at ages 70, 75, and 80. The PIVUS data was used in Study II and IV.

3.4 FOLLOW-UP AND CORONARY HEART DISEASE DEFINITION

3.4.1 National Patient Register

The Swedish National Patient Register (NPR) is a nationwide register that collects information from in-hospital treatments in Sweden, which is maintained by the National Board of Health and Welfare. The disease status of participants from all STR cohorts and ULSAM can be linked in the NPR via a Swedish personal number. The register contains information on the cause of hospitalization with discharge diagnosis codes, including the principle cause and several contributing causes. The first CHD case was recorded in 1969 and the registry was completed with nationwide coverage in 1987. Information from the NPR was used to identify the first hospitalization information resulting from MI (the outcome in Study I) and resulting from CHD (the outcome in Study III).

3.4.2 Cause of Death Register

The Cause of Death Register (CDR), also kept by the National Board of Health and Welfare, includes mandatory reported information from death certificates for all deceased individuals in Sweden. It contains data on the principal and contributing causes of death, coded according to the International Classification of Diseases (ICD). The register was established in 1952, and it contains complete information on causes of death since 1961. Information from the CDR was used to identify death resulting from MI (the outcome in Study I) and death resulting from CHD (the outcome in Study III).

3.4.3 Coronary heart disease definition

MI occurrence was defined by the following ICD codes: ICD-7 codes before 1968, 420.10, 420.17; ICD-8 code from 1968 to 1986, 410; ICD-9 code from 1987 to 1996, 410; and ICD-10 codes from 1997 to present, I21 and I22. The first occurrence of MI was the endpoint in Study I.

The occurrence of CHD was defined as occurrence of MI or occurrence of unstable angina. ICD codes for CHD are ICD-7 codes before 1968, 420; ICD-8 code from 1968 to 1986, 410, 411; ICD-9 code from 1987 to 1996, 410, 411B; and ICD-10 codes from 1997 to present, I20, I21 and I22. In addition, individuals who had undergone surgery for percutaneous transluminal coronary angioplasty (PTCA) and coronary artery bypass graft (CABG) were considered to be a CHD case. Surgical codes for PTCA and CABG considered were 3080, 3127, 3158 before year 1997 and FNG02, FNG05, FNC, FND, FNE from year 1997 onward. The first occurrence of CHD was the endpoint in Study III.

Only the main diagnoses (i.e., principal cause of hospitalization or underlying cause of death) were included in the outcome definition to increase the validity of the outcome (positive predictive value of ~95%).[98]

3.5 BLOOD BIOMARKER MEASUREMENTS

Serum total cholesterol, HDL-C, and TG were measured in GOSH, TwinGene, PIVUS, and ULSAM using standard laboratory techniques. LDL-C was calculated based on Friedewald's formula (LDL-C = total cholesterol – HDL-C – 0.45*TG). Lipid fractions were measured after overnight fasting, except for individuals from OCTO-Twin in GOSH and 3% of participants in TwinGene. Lipid measures were used as outcomes in Study II and Study IV. High sensitivity CRP was measured in TwinGene, PIVUS, and ULSAM and used in Study IV.

3.6 GENETIC MARKERS

3.6.1 GOLD chip

The GOLDchip was a custom chip designed to identify lipid-related gene regions. All candidate genes selected were related to at least one serum HDL-C, LDL-C, or TG trait through a literature search based on publications between 2003 and 2008. All known SNPs were taken into consideration if they are within or in linkage disequilibrium (LD) with potential gene regions based on HapMap2 CEU panel, NCBI Build36. A region approximately 20 kb upstream and 10 kb downstream of each examined gene was selected using the UCSC Genome Browser. After selection, Illumina scores for all markers were calculated; those that did not satisfy the criteria for Illumina probe chemistry were then replaced with an SNP in perfect LD ($r^2 = 1$) if available, or other tagging SNPs.

GOSH samples were genotyped using GOLDchip, which were used in Study I. Genotyping was performed using the Illumina GoldenGate assay system (San Diego, CA, USA) on the Illumina BeadStation 500GX at Uppsala University SNP Technology Platform.

3.6.2 Metabochip

Metabochip is a custom chip that prioritizes loci based on previous GWASs for 23 cardiovascular and metabolic traits.[99] SNPs were selected if they were correlated with at least one lipid fraction with a $P < 0.0005$ in the original GWASs. Based on the study by Tesclovich et al,[42] there were 5,023 independent SNPs for HDL-C, 5,055 SNPs for LDL-C, 5,056 SNPs for TG, and 938 SNPs for total cholesterol. In addition, 28,923

additional SNPs were selected for fine mapping 65 lipid-related regions. There were 50,459 and 93,308 SNPs selected for prioritizing and fine mapping for non-lipid traits, respectively, based on previous GWASs.

GOSH, PIVUS, and ULSAM samples were genotyped using Metabochip, which were used in Study II. Genotyping was performed at Uppsala University SNP Technology Platform using the Illumina iSelect Metabochip genotyping array.

3.6.3 Genome-wide arrays

Illumina OmniExpress and Omni2.5M are two genome-wide genotype microarrays designed by Illumina.[60] TwinGene and PIVUS individuals were genotyped using Illumina OmniExpress genotyping array and ULSAM individuals were genotyped using Illumina Omni2.5M genotyping array. Imputation for all studies was performed after quality control based on haplotypes from both HapMap 2 and 1000 genome project using IMPUTE version 2.

Selected SNPs from PIVUS, ULSAM, and TwinGene data based on the HapMap 2 panel in NCBI Build36 were used in Study IV. Genotyping was performed at Wellcome Trust Sanger Institute, Hinxton, UK (PIVUS) and Uppsala University SNP Technology Platform (ULSAM and TwinGene).

3.6.4 Quality control for genotype data

Similar exclusion criteria were implemented for all studies with respect to quality control.

All samples with bad genotyping quality were excluded: 1) call rate < 95%; heterozygosity > ±3 standard deviation (SD); 3) duplicates; 4) gender discordance; 5) ethnic outliner based on HapMap2 multi-ethnic panel; 6) relatedness in PIVUS and ULSAM (pi-hat > 0.2);

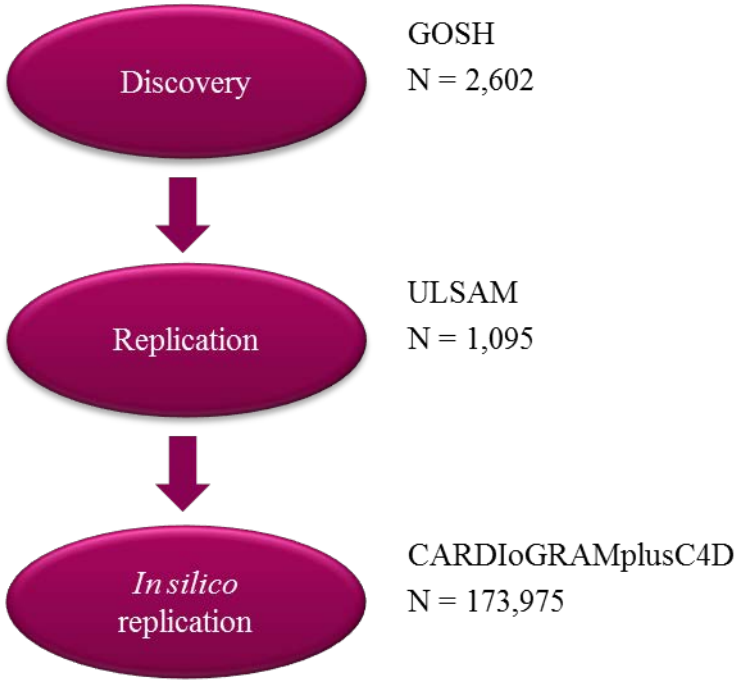
SNPs meeting the following conditions were excluded: 1) all monomorphisms; 2) minor allele frequency (MAF) < 0.01; 3) *P*-value for Hardy-Weinberg equilibrium (HWE) < 10^{-6}; 4) missing genotype rate > 0.01 (if MAF < 0.05) or missing genotype rate > 0.05 (if MAF > 0.05). In addition, SNPs were also excluded if they exhibited large position disagreements, did not map at all in the genome, if they mapped more than once in the genome, and if they have bad probe assays.

4 STUDY DESIGNS AND METHODS

4.1 STUDY I

In Study I, we used a candidate gene study design to investigate the association between lipid-related genes and MI incidence, as shown in **Figure 2**.

Figure 2. *Candidate gene design for investigating associations between lipid-related genes and myocardial infarction incidence.*



First, we used the population-based cohort GOSH within the STR as the discovery sample. In total, 554 SNPs from 41 lipid-related gene regions were selected for GOLDchip. A sex-adjusted Cox regression was performed for each SNP-MI incidence association. Considering that life-time disease information from NPR and CDR was available for most individuals in GOSH, we allowed each individual to enter the study at age 18 and followed them until MI onset, death, or the end of 2008.

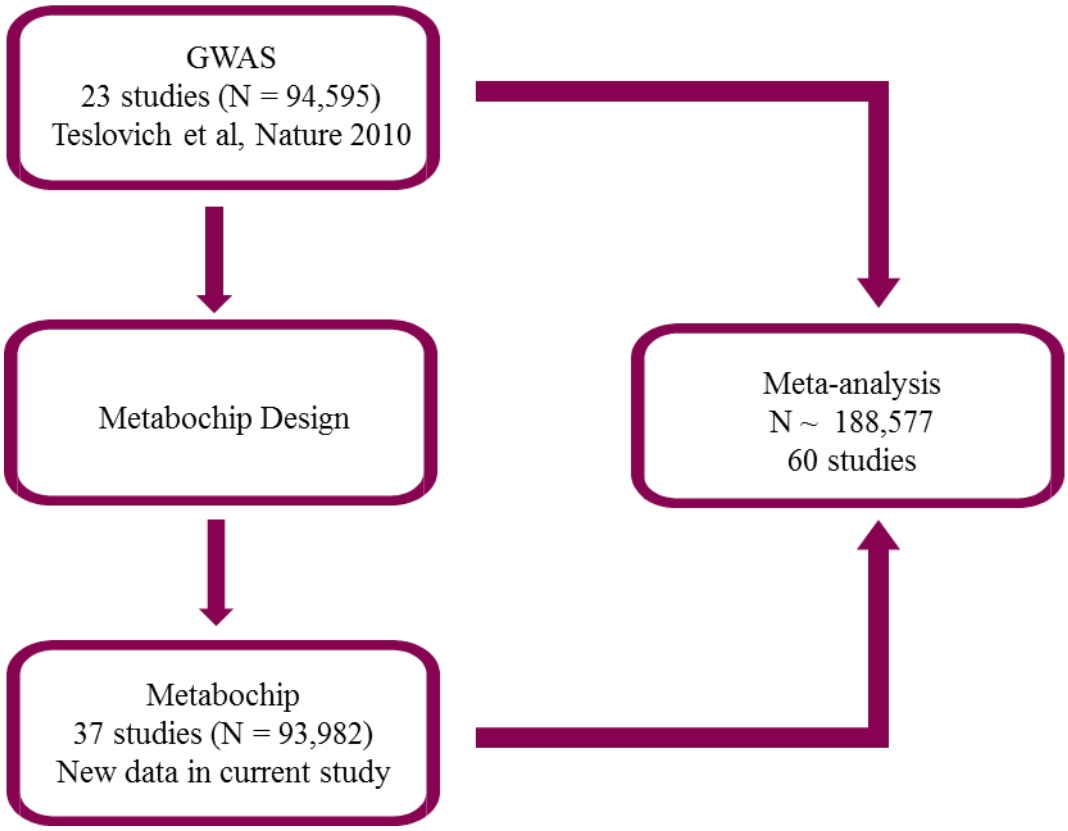
Second, SNPs that were significantly associated with MI incidence in the discovery sample in ULSAM were directly genotyped; a similar Cox regression model was performed for SNP-MI incidence.

Third, for SNPs that were consistently associated with MI incidence in both discovery and replication samples, we performed a fixed-effect meta-analysis using all studies with European ancestry from CARIDoGRAMplusC4D consortium as an *in silico* replication.

4.2 STUDY II

In Study II, we combined data from genome-wide genotyping arrays and data from Metabochip to explore genetic determinations for total cholesterol, HDL-C, LDL-C, and TG, as shown in **Figure 3**.

Figure 3. *Study design for joint GWAS and Metabochip meta-analysis.*



Teslovich et al reported 95 independent loci associated with lipids based on GWASs in 94,595 individuals from 23 studies.[42] The results from that study were used to select prioritized SNPs and fine mapping regions for lipids on the Metabochip, as described above.

There were 37 studies genotyped using Metabochip, which included 93,982 total individuals. In each participating study, individuals were excluded if they did not fast overnight (> 8 hours) or if they were taking lipid-lowering medication. The residuals for each lipid fraction were obtained after adjusting for age, square of age and sex, and then inverse-normal transformed. SNP-lipid association tests were performed using linear regression with additive genetic models and using normalized residuals as the dependent variables. The population structure was further adjusted for studies with independent participants (i.e. PIVUS and ULSAM), while mixed effect models were performed in studies with family-clustered participants (i.e. GOSH). Large-scale associations between SNPs and lipids were performed in each participant study using statistical package including PLINK (i.e. PIVUS and ULSAM), MERLIN (i.e. GOSH), SNPTest, MMAP, EMMAX and GenABEL. A meta-analysis was performed with the Stouffer Z-score method using METAL.[100]

A locus was considered to be a novel independent signal if it reached the genome-wide significant level (P-value < $5*10^{-8}$) in the combined results and was at least 1 Mb away from the nearest, previously lipid-related loci reported. We used only European samples for the discovery of new genome-wide significant loci.

4.3 STUDY III

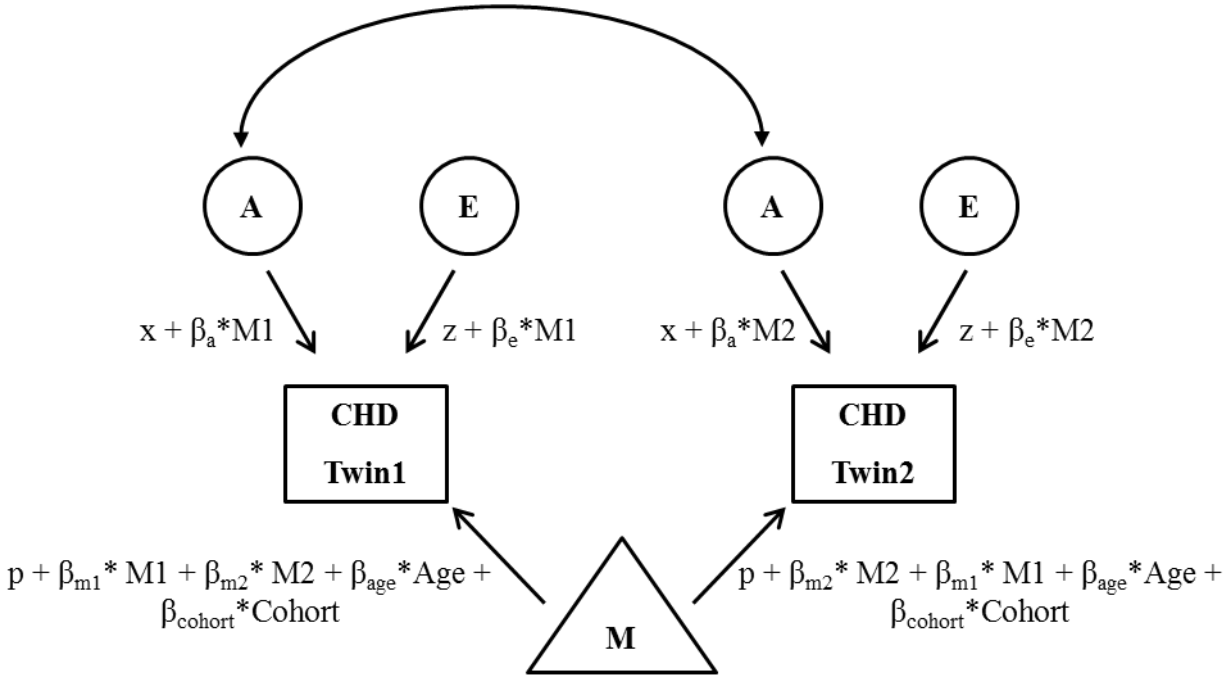
In Study III, we used a moderator twin model[101] to investigate the gene-lifestyle interaction effects on CHD.

Lifestyle information was obtained from Cohort I and Cohort II within the STR. In total, 51,065 same-sex twins from 25,715 twin pairs were eligible for this study. During the 40-year follow-up, 7,264 incident CHD events were recorded.

Classic twin modeling relies on the differences in genetic similarity between monozygotic (MZ) and dizygotic (DZ) twins.[102] Using this information, we decomposed the phenotypic variance into additive genetic variance (A), shared environmental variance (C), and non-shared environmental variance (E). Because shared environmental factors have little influence on CHD,[103, 104] we used an AE model to study the gene-lifestyle interaction effect on CHD.

In a moderator twin model (**Figure 4**), we first adjusted for age (β_{age}*age), cohort effect (β_{cohort}*cohort), and measure of the moderator (M) in both twins within a twin pair (β_{m1}* M1 and β_{m2}* M2) for the risk of CHD (p). Next, the total variance in CHD risk was decomposed into A and E. We hypothesized that both A and E could differ depending on the moderator (β_a*M and β_e*M, respectively). If A is significantly different depending on the moderator, i.e., β_a significantly differs from zero, then we conclude that a gene-moderator interaction effect is evident. We investigated the role of age and different lifestyle factors (e.g., smoking, alcohol consumption, sedentary lifestyle and BMI) as moderators in our primary analyses. The twin modelling was performed using the structural equation modeling package Mx 1.703.[105]

Figure 4. *Moderator twin model for investigating moderator effect on genetic influence of coronary heart disease.*



When a gene-lifestyle interaction effect was evident, we further allowed A and E to differ depending on age and lifestyle at the same time in order to explore the gene-lifestyle interaction while removing the age effect.

4.4 STUDY IV

In Study IV, we used a Mendelian randomization (MR) study design to investigate the bi-direction causal relationships between fasting lipids and CRP, as shown in **Figure 5**.

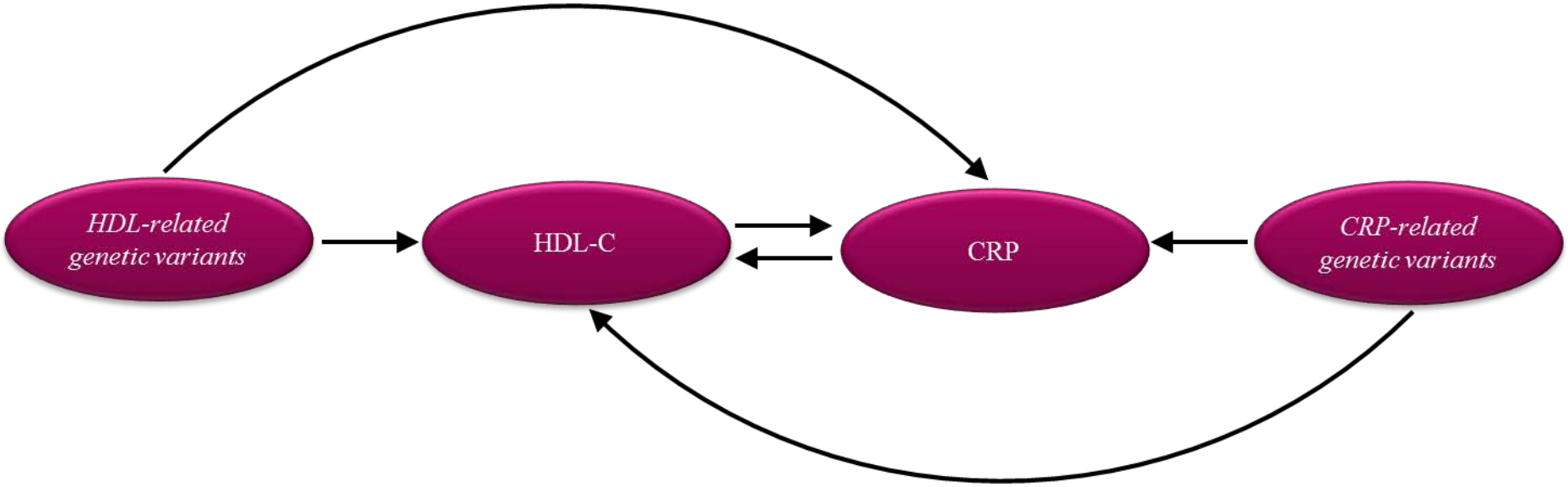
Fasting lipids and CRP levels, as well as genetic information, were obtained from PIVUS, ULSAM, and TwinGene. There were 10,723 individuals eligible for this study after excluding individuals who lacked fasting lipid and/or CRP measurements, were using lipid-lowering medications, were in acute phase inflammation (CRP > 10 mg/L), and those that lacked genetic data or that did not pass sample genotyping quality control.

We investigated the associations between lipids and CRP in three steps. In the first step, we estimated the effects of lipid fractions on CRP using linear regression models adjusted for age and sex. We further adjusted for the following CHD risk factors: smoking, BMI, waist-hip ratio, diabetes, and systolic blood pressure (SBP) and use of anti-hypertensive medication. To understand which factors explained the observed associations, we used exploratory analyses by adding each CHD risk factor one by one to the age- and sex-adjusted model. If the effect size changed with more than 10% when adding a specific factor, it indicated that the observed association was partly confounded by this factor. The same approaches were applied to investigate the observational effects of CRP on each lipid fraction.

In the second step, we explored the causal associations between lipids and CRP using a MR design, as shown in **Figure 5**. The causal effects of lipid fractions on CRP were estimated using genetic risk scores (GRSs) or single genetic loci related with each lipid fraction as instrumental variables (IVs) in a two-stage least squares (2SLS) linear regression.[106] The causal effects of CRP on lipid fractions were estimated using CRP-related GRS or the *CRP* locus as IVs. Genetic variants as IV estimators for lipids were selected based on the results from Study II of the current thesis, while for CRP they were selected based on results from Dehghan et al.[16] STATA 12.1 (StataCorp, College Station, TX, USA) package *ivreg2* was used for the IV analyses. All results from the first and second steps were based on three study samples using fixed-effect meta-analyses.

In the third step, we applied the approach by Do et al, which addresses causal inference using summarized data,[49] to test the robustness of our conclusions from IV analyses. In this step, we investigated whether the effects of lipid-related SNPs on each lipid fraction were correlated with their effects on CRP, after adjusting for their effects on other lipid fractions, such as BMI, waist-hip ratio, fasting glucose, and fasting insulin. A significant correlation suggests a causal effect of the lipid fraction on CRP. The same approach was applied to investigate causal effects of CRP on lipid fractions based on correlations between the effects of CRP-related SNPs on CRP and lipids.

Figure 5. *Bi-direction Mendelian randomization for causal relations between lipids and C-reactive protein.*



The bi-directional Mendelian randomization design exemplified for the association of high-density lipoprotein cholesterol (HDL-C) and C-reactive protein (CRP). The same design was also applied for the other lipid fractions.

5 STUDY SUMMARIES

5.1 STUDY I

The missense variant rs4149313 in the ATP-binding cassette, sub-family A, member 1 gene (*ABCA1*) was consistently related to MI incidence in both the discovery study GOSH and replication study ULSAM. Each copy of the effect allele of rs4149313 increased the hazard of MI incidence in GOSH by 36%, while this increase was 40% in ULSAM (**Figure 6**). In the *in silico* replication, the association between rs4149313 and CHD among all studies with European ancestry was of borderline significance (odds ratio [OR] = 1.03; 95% confidence interval [CI] 1.00, 1.06), as shown in **Figure 7**.

Figure 6. *Kaplan-Meier survival functions for rs4149313 and survival of myocardial infarct in GOSH (a) and ULSAM (b).*

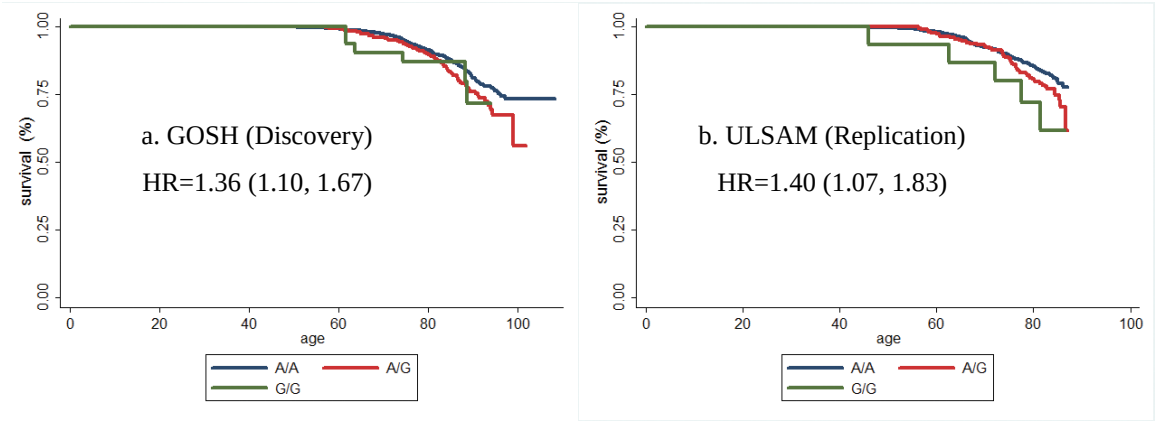
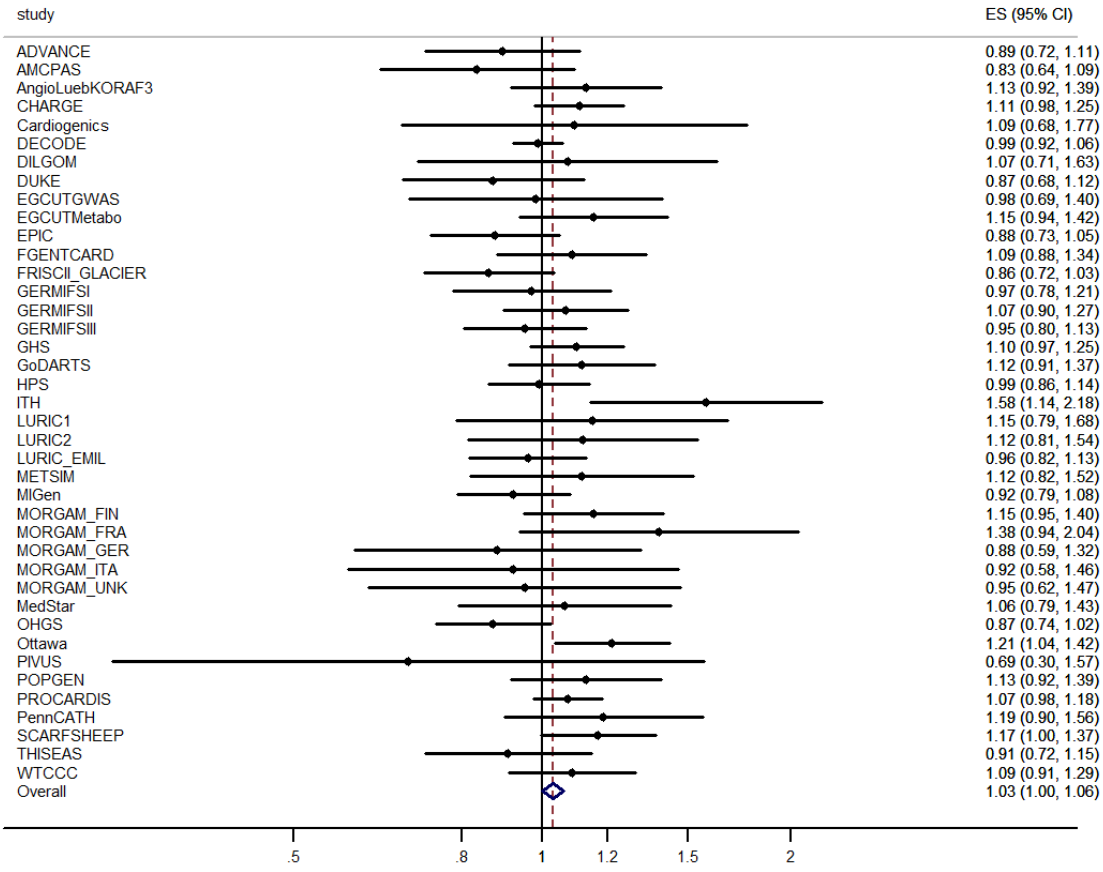


Figure 7. *Forest plot for independent replication of association between rs4149313 and coronary artery disease in CARDIoGRAMplusC4D.*

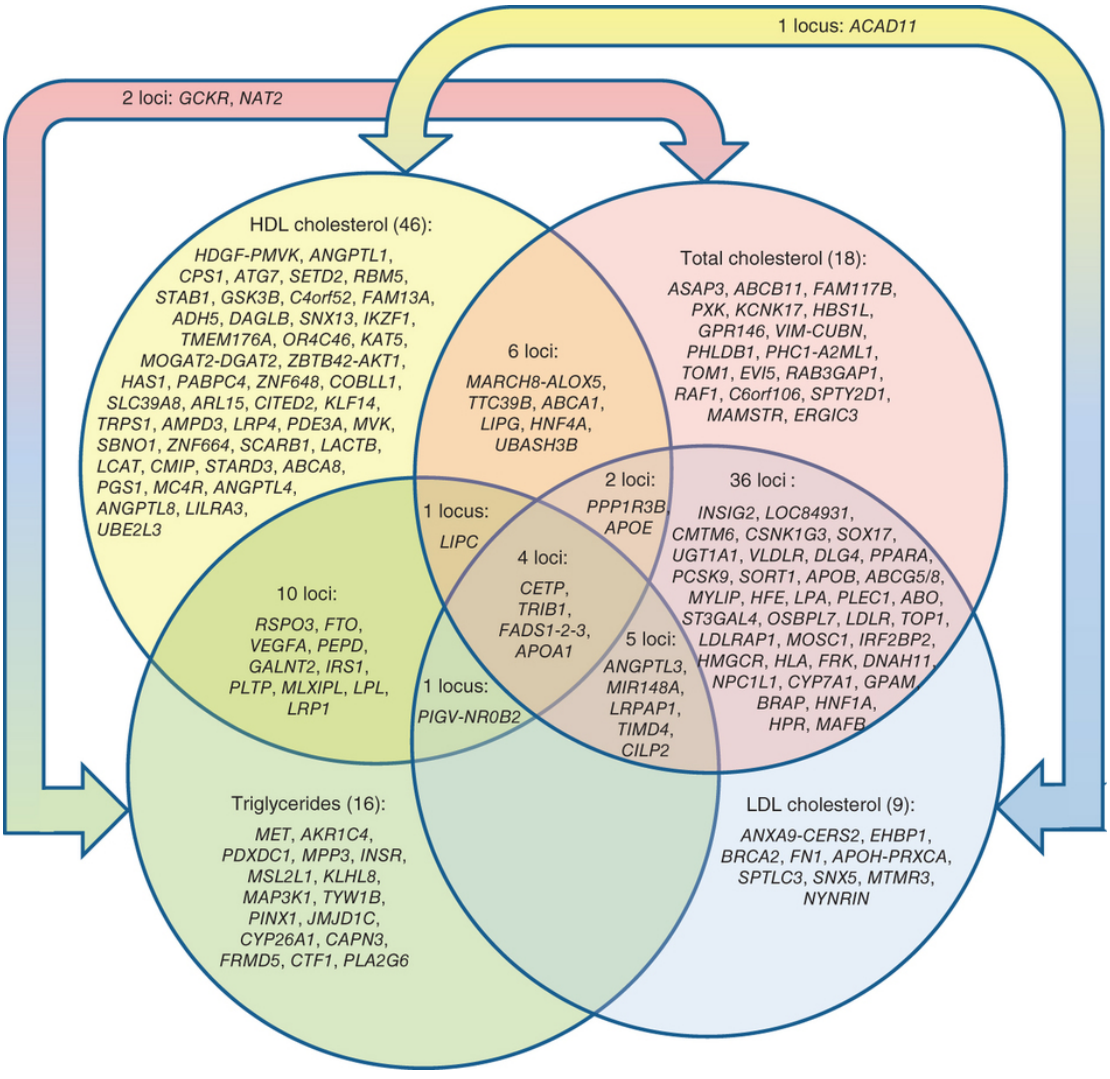


5.2 STUDY II

In this study, we uncovered 157 loci related with at least one lipid fraction based on a meta-analysis of up to 188,577 individuals (**Figure 8**). Each locus was separated by at least 1 Mb apart and had a low LD ($r^2 < 0.10$). Only the annotated genes for lead SNPs, which showed the strongest association with respective lipid traits, were reported.

For 62 of the 157 loci, this was the first time they have been reported to be associated with lipids; 24 demonstrated the strongest evidence of association with HDL-C, 15 with LDL cholesterol, 15 with total cholesterol, and eight with TG level (**Figure 8**). The effect sizes associated with newly identified loci were generally smaller than loci identified in previous GWASs. For the 62 newly identified variants, the proportion of trait variance explained in the Framingham offspring ranged from 1.6% (HDL cholesterol) to 2.6% (total cholesterol; 2.1% for triglycerides, 2.4% for LDL cholesterol). This brings the total trait variation explained by both primary and secondary association signals at previously known loci and these 62 new loci to ~13.8% for HDL-C, ~14.4% for LDL-C, ~11.7% for TG, and ~14.5 % for total cholesterol.

Figure 8. *Overlap of loci associated with different lipid traits.*



5.3 STUDY III

Table 1 presents the modifying effect of age, smoking, sedentary lifestyle, alcohol consumption, and BMI on genetic variance and non-shared environmental variance of CHD (β_a and β_e shown in **Figure 4** and **Table 1**). Both increasing age and increasing BMI were significantly related to increasing CHD genetic variance, as well as related to even larger increasing environmental CHD variance.

Considering that there was a positive correlation between age and BMI, we further investigated the genetic variance and non-shared environmental variance of CHD dependent on both age and BMI in the same model. We found that after removing the age effect, genetic variance of CHD decreased when BMI increased. BMI-moderating effect on genetic variance of CHD was significant in men (β_a = -0.04 ; 95% CI -0.07, 0.00; **Figure 9**), and showed a similar trend in women (β_a = -0.06; 95% CI -0.11, 0.02). This suggests that genetic factors play a more important role in CHD in lean individuals compared to obese individuals.

There was no evidence that smoking, sedentary lifestyle, or alcohol consumption modified the genetic variance of CHD.

Figure 9. *Genetic and environmental variance of coronary heart disease dependent on body mass index after removing the age effect on coronary heart disease variance.*

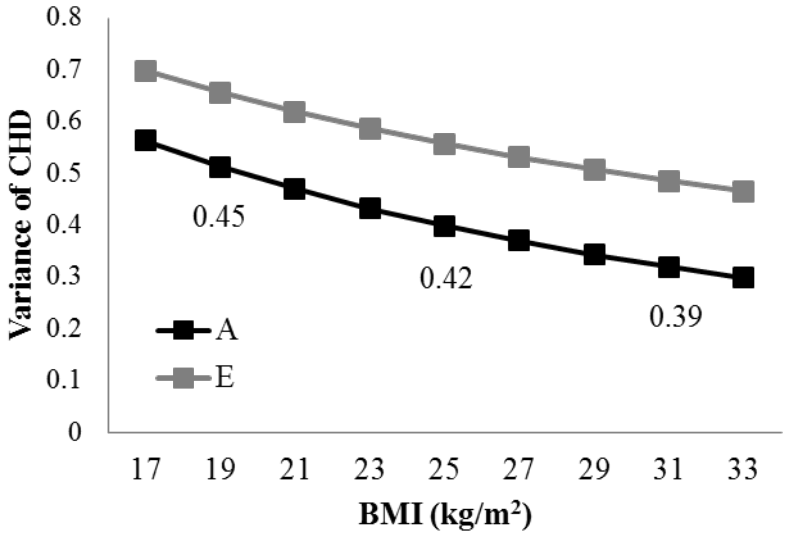


Table 1. Moderating effects of various life-style factors on genetic variance and non-shared environmental variance in coronary heart disease.

	Men		Women	
	Moderating effect on genetic component (95% CI)	**Moderating effect on environmental component (95% CI)**	**Moderating effect on genetic component (95% CI)**	**Moderating effect on environmental component (95% CI)**
Age	0.31 (0.27, 0.31)	0.45 (0.41, 0.46)	0.23 (0.19, 0.24)	0.45 (0.37, 0.45)
Smoking	0.06 (–0.06, 0.19)	0.26 (0.12, 0.40)	–0.12 (–0.24, 0.01)	–0.22 (–0.32, –0.10)
Sedentary lifestyle	0.03 (–0.15, 0.25)	–0.10 (–0.29, 0.12)	0.24 (–0.01, 0.53)	0.01 (–0.24, 0.30)
Alcohol consumption	–0.15 (–0.27 –0.01)	–0.08 (–0.22, 0.07)	0.12 (–0.32, 0.36)	–0.34 (–0.55, 0.06)
BMI	0.07 (0.03, 0.11)	0.26 (0.20, 0.33)	0.05 (0.00, 0.09)	0.33 (0.24, 0.38)

Moderating effect on genetic component is β_a, and moderating effect on environmental component is β_e

5.4 STUDY IV

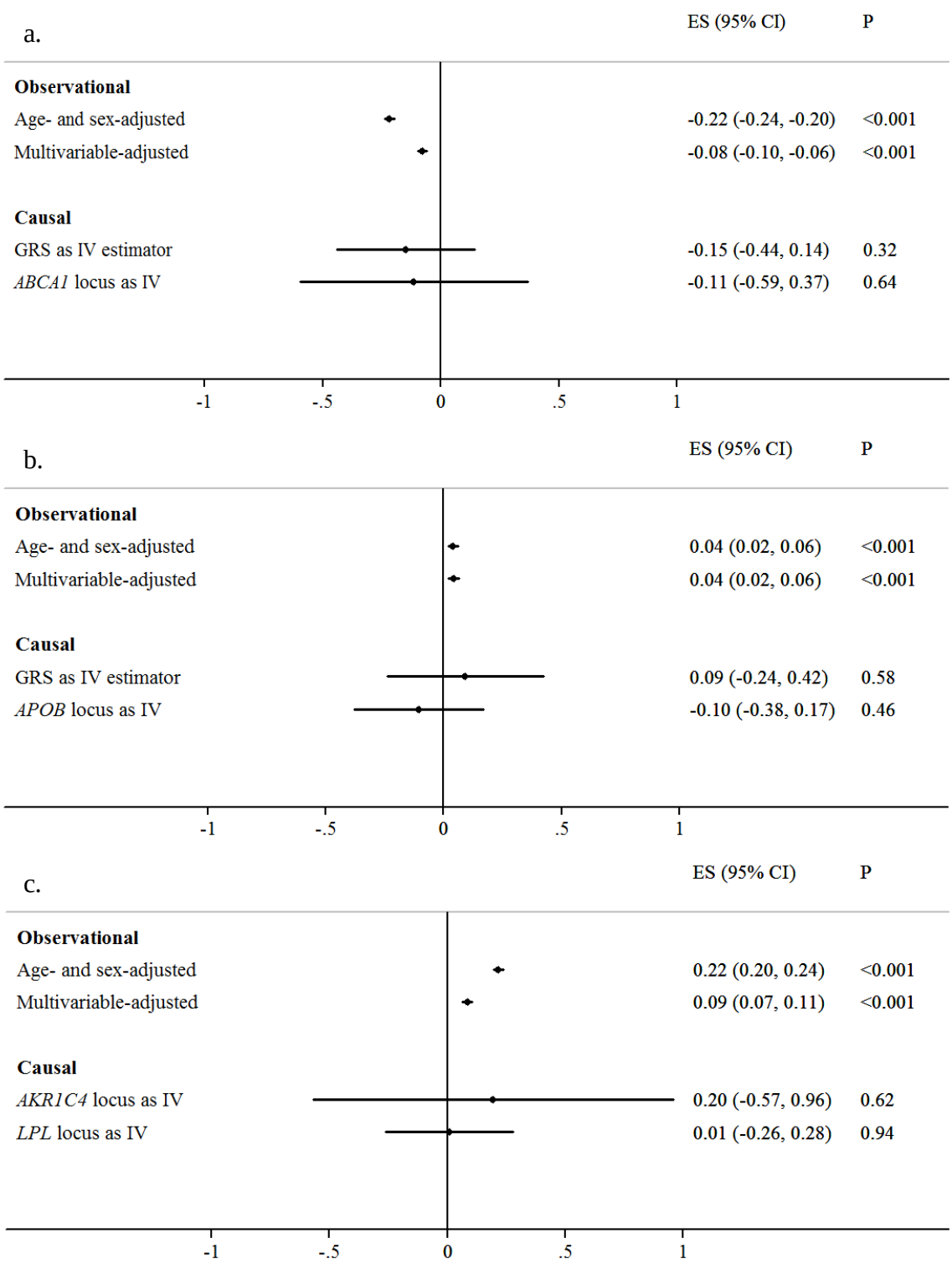
As shown in **Figure 10a**, HDL-C and CRP was negatively associated after adjusting for age and sex (β = -0.22, *P*-value < 0.001; **Figure 10a**), and this association was attenuated when additionally adjusting for other CHD risk factors (smoking, BMI, waist-hip ratio, diabetes, SBP and anti-hypertension medication; β = -0.08, *P*-value < 0.001; **Figure 10a**). Exploratory analyses suggested that the observed association between HDL-C and CRP were partly confounded by adiposity, as both BMI and waist-hip ratio attenuated the association 50% and 28%, respectively. Using the HDL-C-related GRS as an IV estimator, we did not observe a causal effect of HDL-C on CRP (β = -0.15, *P*-value = 0.32; **Figure 10a**). Similar results was shown when using a variant from the *ABCA1* locus as the IV (β = -0.11, *P*-value = 0.64; **Figure 10a**), as well as when applying Do et al's method. Taken together, our results found no evidence of a causal effect of HDL-C on CRP.

Figure 10b presents a weak and positive association between LDL-C and CRP when adjusting for age and sex (β = 0.04, *P*-value < 0.001), and the association was not attenuated when additionally adjusting for CHD risk factors as described above (β = 0.04, *P*-value < 0.001). We did not observe a causal effect of LDL-C on CRP when using either the LDL-C-related GRS (β = 0.09, *P*-value = 0.58) or the *APOB* locus (β = –0.10, *P*-value = 0.46) as IVs. Do et al's method presented the similar results. Taken together, our results did not provide evidence for a causal effect of LDL-C on CRP.

Figure 10c shows a strong positive association between TG and CRP when adjusting for age and sex (β = 0.20, *P*-value < 0.001). The association was attenuated when adjusting for CHD risk factors (β = 0.09, *P*-value < 0.001), which was mainly due to confounding effects of BMI and waist-hip ratio (reduced the association between TG and CRP with 47% and 31%, respectively). We did not observed a causal effect of TG on CRP in the IV analyses using either the *AKR1C4* locus (β = 0.20, *P*-value = 0.62) or the *LPL* locus (β = 0.01, *P*-value = 0.94) as IVs. Do et al's method produced similar results. Taken together, our results did not provide evidence for a causal effect of TG on CRP.

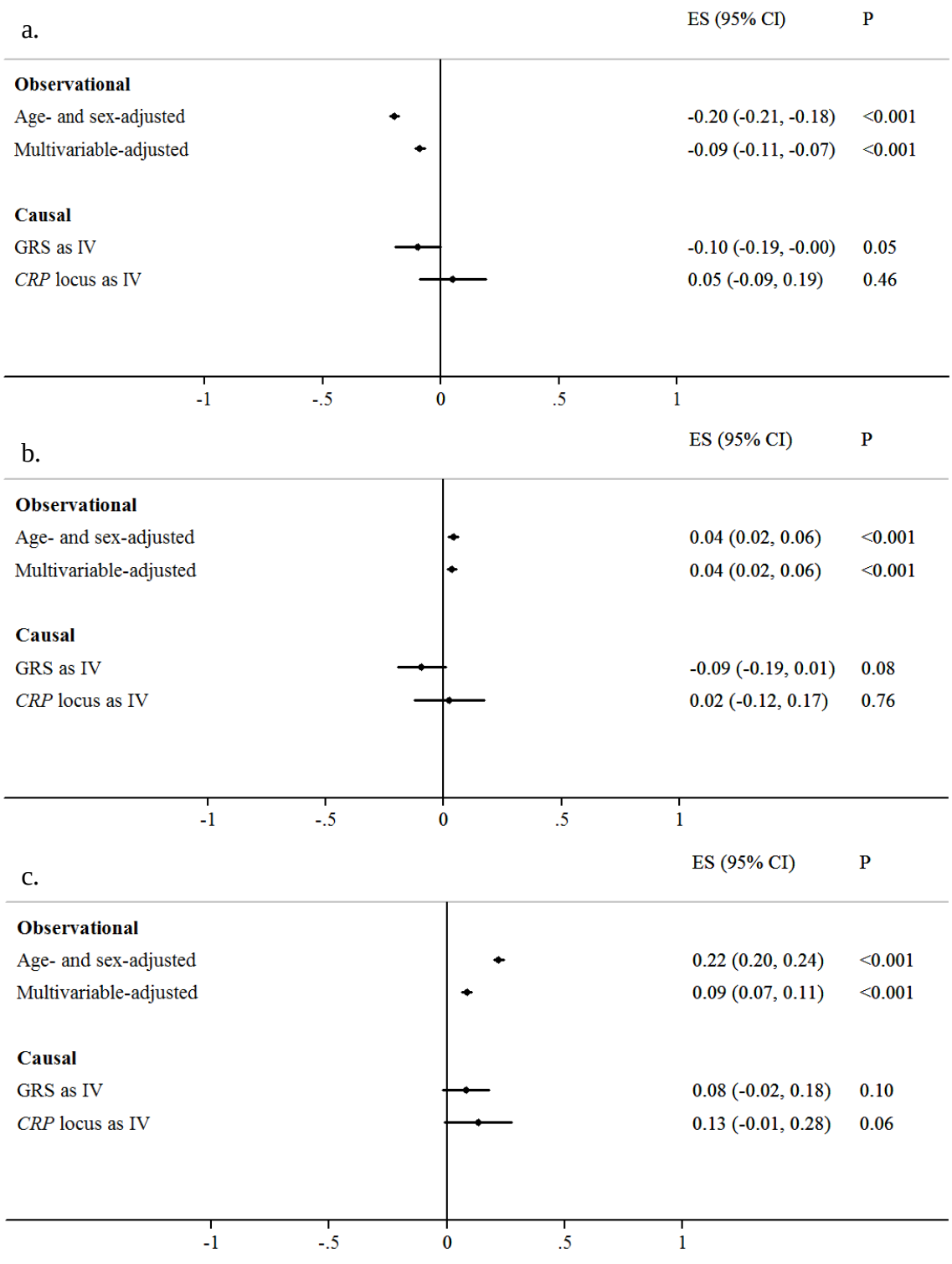
As shown in **Figure 11**, the observational effects of CRP on HDL-C, LDL-C, and TG were similar to the observational effects of HDL-C, LDL-C, and TG on CRP when adjusting for age and sex and when further adjusting for CHD risk factors. Exploratory analyses consistently suggested that the associations between CRP and HDL-C/TG were mainly confounded by BMI and waist-hip ratio. Using the CRP-related GRS or the *CRP* locus as IVs, we did not observe any causal effect of CRP on HDL-C (β = –0.10, *P*-value = 0.05; β = 0.05, *P*-value = 0.46, respectively), LDL-C (β = –0.09, *P*-value = 0.08; β = 0.02, *P*-value = 0.76, respectively), or TG (β = 0.08, *P*-value = 0.10; β = 0.13, *P*-value = 0.06, respectively). Do et al's method also suggested no causal effects of CRP on lipids. Taken together, our results did not provide evidence for causal effects of CRP on HDL-C, LDL-C, or TG.

Figure 10. *The observational and causal effects of HDL-C (a), LDL-C (b) and TG(c) on CRP.*



IV refers to instrumental variable; GRS refers to genetic risk score; ES (95% CI) refers beta coefficient with 95% confidence interval; and P refers to *P*-value.

Figure 11. *The observational and causal effects of CRP on HDL-C (a), LDL-C (b) and TG(c).*



IV refers to instrumental variable; GRS refers to genetic risk score; and ES (95% CI) refers beta coefficient with 95% confidence interval; and P refers to *P*-value.

6 DISCUSSION

6.1 FINDINGS AND IMPLICATIONS

6.1.1 *ABCA1* gene and incident MI

In study I, we investigated associations between SNPs in 41 lipid-related regions and MI incidence using a candidate gene approach. We found one missense SNP (rs4149313) in the *ABCA1* gene that was consistently associated with MI incidence in two Swedish cohorts. Furthermore, the association between rs4149313 and CHD was borderline significant in a meta-analysis including 173,975 individuals of European ancestry. Our results are suggestive of a weak association between this variant and the development of atherosclerosis and MI.

The *ABCA1* gene encodes a cellular transporter that mediates phospholipids and cholesterol transport from inside cells into the blood to from nascent HDL. Certain mutations in *ABCA1* lead to Tangier disease, which is characterized by a severe reduction in HDL.[107] rs4149313 is a missense SNP in exon 18 of *ABCA1*, which encodes the amino acid substitution isoleucine to methionine (I883M). This amino acid substitution is located at the N-terminal of the first ATP-binding motif. The functional prediction of this amino acid substitution is controversial. It was reported to be benign by Polymorphism Phenotype 2 (PolyPhen2), while, conversely, it was indicated to be functional in PANTHER. However, *in vitro* experiments demonstrate that cholesterol efflux in HEK-293 cells transfected with the rs4149313 variant significantly differed from wild-type cells,[108] which supports the biological functionality of this variant. In addition, this variant is conserved in different species, including zebra fish, western clawed frogs, chickens, common turkey, mice, rats, guinea pigs, dogs, bovines, and lowland gorillas. The conservation of this variant also indicates its importance, as well as offers opportunities for functional studies in different model animals.

6.1.2 Genetic determinants of lipid metabolism

In Study II, we reported 62 independent novel loci associated with blood lipids. Combining this data with previous findings,[42] a total of 157 loci have been associated with blood lipid levels to date. Through literature review and pathway analyses, there is evidence supporting 38 new loci in LD with genes that functionally regulate blood lipid levels.

Functional studies are required to identify causal variants as well as to understand the molecular mechanisms of relevant genes. For example, we found variants near the G protein-coupled receptor 146 gene *(GPR146*) that were associated with total cholesterol. Knockout of *Gpr146* in mice modifies blood cholesterol levels suggesting that GPR146 is a potential target for hyperlipidemia medication. For the 62 newly identified loci, there are a total of 240 genes within 100 kb of each locus. These genes provide attractive, albeit challenging due to the extensive work needed to explore the large number of genes, candidates for future functional studies.

When comparing lipid-related loci with GWASs for other metabolic and cardiovascular traits, many loci were also associated with CHD, BMI, SBP, diastolic blood pressure (DBP) and T2D. Furthermore, the effects of SNPs on LDL-C and TG, but not HDL-C, were highly correlated with the effect of the same SNPs on CHD, which indicates

causal effects of LDL-C and TG, but not HDL-C, on CHD. More detailed analyses addressing the causal effects of lipids on CHD based on the results of the current study were published in a companion article by Do et al.[49]

6.1.3 BMI-gene interaction effect on CHD

In Study III, we found that after removing the age effect, genetic variance of CHD decreases when BMI increases. This indicates that genetic factors predisposing to CHD play a more important role in lean individuals compared to obese individuals. Genetic variance of CHD change dependent on BMI can be explained via two possible underlying mechanisms: 1) additional genes may influence CHD variance in those with a low BMI; and 2) the magnitude of the effect of some CHD-related genes may differ according to BMI level.

Several studies investigating gene-environment effects on metabolic traits support the aforementioned second mechanism. For instance, Perry et al reported that 29 out of 36 known T2D related loci had a larger effect size on T2D in the lean group compared to obese group.[109] In addition, Lamina et al reported that a GRS associated with increased HDL-C had a smaller effect in the obese group compared to the lean group.[110] Together with our results, this suggests higher expression of HDL-, T2D-, or CHD-related genes in lean individuals compared to obese individuals. A GWAS for CHD stratified by BMI with adequate study power should be performed to confirm this hypothesis. Based on the results of our study, such a GWAS would be especially beneficial in the normal-weight population as these susceptibility variants may be masked in GWASs that include individuals with higher BMIs.

6.1.4 Causal relationships between blood lipids and CRP

In Study IV, we investigated causal relationships between fasting blood lipids and CRP in 10,732 individuals from three Swedish cohorts. We did not find robust evidence for causal relationships between CRP and any of the lipid fractions.

In line with our results, randomized trials for cholesteryl ester transfer protein (CETP) inhibitors, which target and elevate HDL-C, have achieved their primary aim but have not been shown to reduce CRP level or CHD risk.[47, 111] Furthermore, intervention trials with statin therapy have shown that reduced CRP levels are independent of changes in LDL-C levels.[112] Taking together, this evidence suggests that there are no causal effects of HDL-C or LDL-C on CRP. Currently, there are no intervention trials that primarily target lowering blood TG levels, and more studies are required to explore the causal effect of TG on CRP. Furthermore, intervention trials using monoclonal antibodies against IL-6R have been shown to reduce CRP while moderately elevating HDL-C, LDL-C, and TG in rheumatoid arthritis patients.[113, 114] One possible interpretation of this result is that low-grade inflammation can lead to changes in both CRP and lipid fractions, and that the effects on lipid levels are not mediated by CRP, but by inflammatory molecules higher up the inflammation cascade. This hypothesis requires further investigation using other biomarkers than CRP.

The observational associations between CRP and HDL-C and between CRP and TG were attenuated over 50% when adjusting for CHD risk factors, particularly BMI and waist-hip ratio, which suggests that adiposity confounds the relation of CRP and

dyslipidemia. This result is in line with a recent large MR study which reported that BMI is a causal factor for both CRP and dyslipidemia.[60]

In the *post-hoc* power calculation for IV analyses, the statistical power for using GRSs as IV estimators to investigate causal relations between lipids and CRP were limited, especially when investigating causal effects of lipids on CRP, which is the main limitation for the current study. Therefore, we applied Do et al's method, which is less sensitive to pleiotropy and provides better power; this method produced results that were consistent with our main analyses. In addition, we acknowledge that low-grade inflammation can be a causal factor for regulating lipid fractions and that its effect on lipids is mediated by inflammatory molecules higher up in the inflammation cascade instead of by CRP.

6.2 METHODOLOGICAL CONSIDERATIONS

6.2.1 Left truncation when using genetic data in survival models

In Sweden, all medical care is publically funded and available to all citizens, and all in-patient care, medication and death information is registered in national registries. Therefore, we were able to identify the time at the first onset of CHD for those who participated in each cohort before or after answering the questionnaire and when they gave blood samples. However, we were not able to include those who had fatal CHD before the cohort was initiated.

In Study I, we investigated the first MI incident, including cases that had yet to participate in GOSH or ULSAM. As DNA sequence is stable over the course of one's life, this approach increased the power of the study and avoided recall bias from disease self-reporting. However, with this approach, individuals experiencing fatal MI before baseline measurements could be obtained did not enter our study. The potential survival bias introduced by such early onset fatal cases could potentially decrease the effect size of our findings, and subsequently cause us to accept the null hypothesis. However, because MI is a disease that develops over time, occurs repeatedly, and mainly affects the elderly, we consider using age 18 as the entry age as an appropriate method for dealing with our data to maximize statistical power. In additional sensitivity analyses, the association of rs4149313 with incident MI was weaker when using blood draw as the time of study entry. This could be due to a potential survival bias caused by those whose first MI incidence occurred before 1987, or due to decreased power by losing 37% of the cases compared to our main analysis.

6.2.2 Twin modeling addressing gene-environment interaction

The moderate twin model for gene-environment interaction was first developed by Purcell et al.[101] Using this model, the authors explored how the variance components (e.g., genetic variance, shared environmental variance, and non-shared environmental variance) of a continuous trait can vary depending on a measured moderator. A binary trait like CHD cannot be measured using this method directly as the total variance of CHD cannot be estimated using a specific scale. Medland et al modified the method for binary traits. Using this modified method, they constrained the total variance of the outcome to be one at a certain value of a moderator, and allowed variance components to change dependent on the moderator. [115] Furthermore, in their simulation, the

narrowest confidence interval for heritability could be obtained if total variance was constrained to one when the measured moderator equaled its mean level.[115]

Purcell et al pointed out that in twin modeling, gene-environment interaction effects could be flagged by gene-environment correlations, which refers to the same genes regulating both the moderator and phenotype investigated.[101] In Study III of the current thesis, we removed the CHD variation explained by the moderator via adjusting for the measure of moderator for the risk of CHD. Using the method, the "false positive" gene-environment interaction effects due to gene-environment correlation was excluded. In addition, van der Sluis et al reported that a cross twin correlation between phenotype in twin 1 and moderator in twin 2 could also cause false positive results using Purcell et al's method[116]. Therefore, in Study III, we followed Medland's Mx script of moderator twin model for binary traits[115] and additionally added the measure of moderator for both twins in a pair to adjust CHD risk.

6.2.3 Causal inference using Mendelian randomization

Identifying causal relationships is always an intriguing topic in medical research. It is very difficult to explore causal effects using observational epidemiological research due to unmeasured confounders and reverse causality. Randomized controlled trials (RCTs) randomize the study subjects before intervention to avoid confounders of the intervention effect on outcome and different types of biases; hence, RCT is the gold standard to study causal effects. However, there are several disadvantages in RCT design, such as high cost, being time consuming, and not always being generalizable or ethical.

MR design is based on principles of Mendelian inheritance: genetic polymorphisms segregate during gamete formation and are randomly assorted at fertilization. This process conducts a natural randomization; that is, all covariates between individuals with different polymorphisms can theoretically be balanced.[117] Therefore, if certain measured genetic polymorphisms directly affect exposure (e.g., *CRP* gene effect on serum CRP levels in Study IV), then the non-confounded association between the exposure and outcome can be evaluated using an IV analysis approach.[106] Genetic variants used as IV estimators need to: 1) be associated with the exposure; 2) not be associated with confounders for the exposure-outcome association; and 3) be independent of the outcome conditioned on the exposure and confounders.[118]

With individual-level data, the causal effect of the exposure on the outcome can be quantified as the ratio of the effect of each allele change of the IV on outcome divided by effect of the IV on exposure.[106, 119] Genetic variants can usually only explain a small proportion of the phenotypic variation, as was the case in Study I and Study II in the current thesis. Therefore, to obtain adequate statistical power, large sample sizes are required to study causal associations.[120] Using GRS can partially boost statistical power, as including multiple genetic variants explains a larger proportion of phenotypic variance.[121, 122] However, large study samples are still essential, especially when effects in observational and causal associations between exposure and outcome are not large, as shown in Study IV of the current thesis.[123]

Meta-analyses for GWASs in large consortia have reported regression coefficients summarizing the GWA with various traits based on large sample sizes, like Study II in the current thesis. They are a powerful source for MR investigations, and an alternative to individual-level data. Statistical methods have been developed to calculate causal estimates based on summarized data. For instance, simulation data shows causal estimates are similar using GRSs based on summarized data to using individual-level data if the sample size is the same.[124] Furthermore, Do et al reports that the effects of lipid-related SNPs on TG correlate with their effects on CHD even after adjusting for their effects on HDL-C and LDL-C, which indicates a causal effect of TG on CHD.[49] Using the method described by Do et al, the effects of SNPs on both lipid fractions and CHD are accurately estimated based on meta-analyses of GWASs. Moreover, potential pleiotropic effects of TG-related SNPs on HDL-C and LDL-C are corrected for by adjusting for SNP effects on these lipid fractions. Therefore, we applied Do et al's method in Study IV to test the robustness of our conclusions from IV analyses.

7 CONCLUSIONS

Study I reported associations between lipid-related SNPs and incident MI in two community-based longitudinal studies with *in silico* replication in a large meta-analysis of GWASs. We consistently detected a weak association between rs4149313 and the development of atherosclerosis and MI across studies. rs4149313 is one of the few amino acid change variants in the *ABCA1* gene known to be associated with reduced cholesterol efflux.

Study II annotated 157 independent loci associated with at least one lipid fraction, including total cholesterol, HDL-C, LDL-C and TG at the GWA significance level. Moreover, we identified 62 novel loci that have not previously been reported to be associated with lipid levels in humans.

Study III showed that differences in the genetic component of CHD are a function of BMI after correcting for the effect of age. Our results suggest that genetic factors may play a more prominent role in disease development in the absence of obesity.

Study IV showed strong observational correlations between dyslipidemia and CRP. However, no evidence supported causal effects of any one lipid fraction on CRP, or supports causal effects of CRP on lipids. Our results might be limited due to insufficient statistical power of the present study or they may indicate that the relationship between blood lipid levels and CRP are driven by other common causes, such as adiposity.

8 FUTURE PERSPECTIVES

8.1 MISSING HERITABILITY

GWASs are initiated to identify the genetic effect of common variants on common diseases.[125, 126] Although these studies have provided valuable insights into the genetics of CHD and its risk factors, all of these findings taken together explain a relatively small proportion of heritability of CHD and other complex traits. This has led to much discussion about the "missing heritability" in CHD, as well as other traits. Although several hypotheses as to why this heritability remains "missing" have been proposed, no single answer has yet been agreed upon.

8.1.1 More common SNPs with very modest effects

Population genetic theory states that decreased reproductive fitness reduces frequency of or entirely erases high-risk genotypes in the population. This theory suggests that genetic influence on common diseases should be largely composed of many common variants with small to modest effects.[127] The effect sizes of SNPs found in GWASs nearly fit Poisson distribution, i.e. very few SNPs have a large effect size.[128] Furthermore, when increasing sample sizes, more common variants with even smaller effect sizes can be found. For example, we found 62 novel loci in Study II of the current thesis based on data from 188,577 individuals, which generally have smaller effects than the 95 loci reported in a previous GWAS based on 94,595 individuals.[42] Even larger meta-analyses of GWASs are likely to identify additional common variants with modest effects. Further, the incomplete tagging of causal variants by currently known lead variants within loci is another common explanation to the missing heritability.[129] That said, arguments against common variants explaining the whole missing heritability have also been raised. Park et al estimated the expected number of genetic loci that should be possible to find in GWASs with adequate power.[128] Their results suggested that even though more common variants with modest effect sizes in future GWASs will be uncovered as the sample size increases, these variants together will explain a relative small proportion of missing heritability.[128]

8.1.2 Rare variants

Since long-term evolutionary selection minimizes individuals at the extremes of traits affecting fitness, high-risk genotypes with large effects are rare (usually MAF < 1%). McCarthy et al suggested that risk allele frequencies of genetic variants are negatively correlated with their effect size.[130] Mendelian diseases, which are caused by mutations in a single gene, are extreme examples of the rare variant model. Rare variants cannot be captured using standard GWAS chips due to insufficient density of genotyping arrays. Furthermore, structural variants, including copy number variants (CNV; i.e., insertions and deletions) as well as copy neutral variants (i.e., inversions and translocations) although rare, can contribute to the "missing heritability". For example, several rare CNV have been reported to be associated with psychological diseases. [131-133]

Fine-mapping approaches can be applied to identify rare variants. Targeted re-sequencing, exome sequencing or whole-genome sequencing with adequate sample sizes is an ideal way for fine-mapping studies.[134] Custom-designed genotype arrays, such as the Metabochip used in Study II, enrich genotypes in regions involving SNPs

associated with cardiovascular and metabolic traits in previous GWASs, which is an alternative platform for fine-mapping studies. In addition, imputation based on known haplotype information from the 1000 Human Genome Project (1000G) can fill in missing genotype data between tag SNPs from GWA genotyping arrays.[135] Such 1000G-imputed data can also be used in fine-mapping studies without *de novo* sequencing. All these fine-mapping designs will be useful to identify rare variants, but adequate sample sizes are needed.

8.1.3 Interaction effects

Heritability estimates based on twin modeling are known to be narrow-sense heritability. This method decomposes phenotypic variance into ACE components, described in Study III of the current thesis. As narrow-sense heritability refers to the proportion of the A component in total phenotypic variance, this method assumes that there is no interaction effect between the different components. Therefore, it is possible that interaction effects can mask heritability estimates. Previous twin study reports that gene-gene interactions primarily modelling into genetic variance can contribute to the missing heritability. In addition, if there is a gene-environment interaction, an A-C interaction will be modelled into A, while an A-E interaction will be modelled into E. For example, in Study III of the current thesis, we found CHD heritability to be higher in the lean group compared to the obese group. Further gene-gene interaction and gene-environment interaction studies with adequate sample size will explain part of the missing heritability.

Taken together, it is likely that there is more than one explanation to the missing heritability. The missing heritability reflects the complexity of genetic architecture. Future studies using different design will complete the missing pieces with regards to understanding heritability.

8.2 FROM ASSOCIATION TO FUNCTION IN THE POST-GWAS ERA

GWASs for metabolic and cardiovascular traits are fruitful and have highlighted multiple loci associated with CHD and its risk factors. A key question in the next step is to identify the causal variants and to understand their function.

8.2.1 Fine-mapping of known loci

Although genetic variants identified in GWASs may not be the causal variants, they may be in LD with casual variants in the nearby region. Fine-mapping studies can identify a complete catalog of all genetic variants in the GWAS loci, which is a good starting point for identifying causal variants.[136]

8.2.2 Functional annotation

Multiple studies have reported that variants revealed from GWASs are enriched in genetic regions influencing gene expression.[137, 138] Integrating intermediate functional phenotypes such as transcriptomics, proteomics and metabolomics with classic genetic analysis can further address the functional consequences of GWAS variants.

Multiple computational approaches have been developed to assign potential function to SNPs based on publically available data. For the SNPs in the coding regions, there are tools to predict their possible effects on proteins. Such tools include PolyPhen2 and

PANTHER, which have been used in Study I and II. SNPs in the non-coding regions may affect gene expression at transcriptional, post-transcriptional and post-translational level. There is a range of computational tools that have been developed to assign potential function of non-coding SNPs, and these have been previously reviewed.[136] In addition, pathway analyses link the top loci in the same biological pathway on a trait from GWAS findings.[139]

8.2.3 Experimental models to understand gene function

Once candidate causal genes for diseases have been identified based on approaches described above, *in vitro* and *in vivo* experiments can be applied to validate these genes and to understand their underlying biological mechanisms. *In vitro* studies based on clinical samples are good for translating the biological mechanisms of candidate genes in humans. However, it can prove difficult to obtain access to clinical samples with relevant genetic variants to culture and manipulate primary cells. Established cell lines then provide a good alternative to overcoming these difficulties, but may lead to identifying false functional effects of candidate genes due to cellular genomic instability and karyotypic abnormalities.[136] *In vivo* studies in animal models, e.g., mice and zebrafish, provide opportunities to monitor novel candidate gene function involved in disease development. For example, the functional validation of *GALNT2*, *PPP1R3B*, *TTC39B* and *SORT1*, which are lipid-related loci highlighted from GWASs, have been performed in mouse models, and other investigations are ongoing.[42, 140]

Furthermore, several recently developed tools enable genomic modification at single points via engineered nucleases, e.g., zinc finger nucleases[141] and transcription-activator-like effector nucleases.[142] In addition, the engineered genomic modifying system clustered regularly interspaced short palindromic repeats (CRISPR) can target multiple independent loci and investigate joint effects of them.[143] These methods will largely enhance functional research in the post-GWAS era.

GWASs have been criticized for lacking clinical utility due to small effects of genetic variants on phenotypes. Lipid-lowering drugs statins target *HMGCR*, which is a gene within a locus associated with LDL-C with small effect size in GWAS. This is a good example that the drug efficacy is not necessarily related with effect size of the single variant in a GWAS. GWAS-derived functional studies will disentangle the complex biological mechanisms of disease and develop suitable treatment strategy.

9 ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere gratitude and appreciation to all my colleagues and friends during my PhD study in Sweden, who have supported me to complete this thesis.

Erik Ingelsson, my main supervisor. Thank you for introducing me to the field of epidemiology, for your endless support in my studies, and for being a fantastic professional role-model with great work efficiency and true passionate about science. I could never have had a better supervisor. You are my academic idol!

Nancy Pedersen, my co-supervisor. Thank you for being so supportive in all my studies and for your day-to-day tutoring on twin modelling. I would also like to thank you for letting me participate in your group events after Erik's group move to Uppsala, which allow to be connected with "a group environment" at MEB. It means a lot to me.

Sara Hägg, my co-supervisor. You have been taking care of me like an elder sister. Thanks a lot for all the discussion about both my project work and my life in Sweden.

Alexander Ploner, my mentor and biostatistician. You are always so nice to me. Thanks a lot for all the fika chatting, discussion regarding statistical questions in my studies, and for always encouraging me. It is so lucky to have you around during my PhD.

All my co-authors: **Anders Hamsten**, **Ann-Christine Syvänen**, **Chandra Reynolds**, **Christina Willenborg**, **Hugh Watkins**, **Jonathan Prince**, **Lars Lannfelt**, **Lars Lind**, **Maria Sabater-Lleal**, **Patrik Magnusson**, **Stavroula Kanoni**, **Tove Fall**. Many thanks for your important contributions and suggestions when preparing the manuscripts in the current thesis.

Current and past members in the Ingelsson research group: **Andrea Ganna**, **Åsa Hedman**, **Åsa Michelgård Palmquist**, **He Gao**, **Jitender Kumar**, **Katherine Kasiman**, **Marcel den Hoed**, **Stefan Gustafsson**, **Tove Fall**, **Anastasia Emmanouilidou**, **Manoj Bandaru**, **Samira Salihovic**, and **Tiffany Klingström**. You are such a group of brilliant and vigorous people. You are an excellent group for both scientific and social interactions. It is great pleasure to work with you.

Current and past members in the Pedersen research group and the past Hear and Mind group: **Adina Feldman**, **Anna Dahl**, **Fei Yang**, **Kaavya Narasimhalu**, **Ida Karlsson**, **Iffat Rahman**, **Johanna Sieurin**, **Karin Wirdefeldt**, **Kathleen Bokenberger**, **Lotte Gerritsen**, **Marie Krushammar** and **Yiqiang Zhang**. Thanks a lot for sharing your interesting research opinions with me and all the social activities.

Hans-Olov Adami, **Nancy Pedersen**, **Henrik Grönberg**, **Paul Lichtenstein** and all members of the executive group, thank you for making MEB such a wonderful place for research.

The TA-staff and IT/DATA-staff at MEB. I would like to thank you for all the help in these years and for the nice and friendly smile☺ In particular, **Camilla Ahlquist**, thanks a lot for helping me to get through the process of defense preparation. I would be so lost without your help!

Dan Grander, **Agneta Mode**, **Klas Kärre** and **Maria Johansson.** Thank you for the nice Biomedicine master program, where I started my journey in Sweden.

Current and past Chinese friends in Sweden: from MEB: **Chen Suo**, **Fang Fang**, **Haiyun Wang**, **Huan Song**, **Jiangrong Wang** , **Jianwei Zhu**, **Jiaqi Huang**, **Jie Song**, **Ruoqing Chen**, **Shuyang Yao**, **Tong Gong**, **Xu Chen**, **Zhiwei Liu**, and **Zongli Zheng**; From external: **Bin Zhao, Heng Wang, Ningsha Wang, Qinzi Yan, Qiuling Huang, Shuo Liu, Xinming Wang, Yang Xu , Zhili Song, Chan Shan, Danfang Zhang, Douhua Lou, Han Yan, Jie Yan, Juan Du, Junwei Zhang, Liu Xi, Ning Wang, Qin Li, Xiaoda Wang, Xiaoyuan Ren, Xue Yuan, Yao Shi, Yunlong Yang, Zhengguo Qiao, Ziming Du, Tianwei Gu, Dong Yang, Jianren Song, Kai Du , Meiqiongzi Zhang, Menghan Gao, Na Guan, Qiang Zhang, Qin Xiao , Yang Xuan , Yixin Wang,** and **Yuning Zhang**. It is my great pleasure to know all of you during my study in Sweden!

Ranjita DuttaRoy, thanks for being such a nice friend ever since master study!

Zheng, thanks for being such a great companion, friend, co-author, tutor and so on.

最后， 我想感谢我的父母及所有家人对我无条件的支持和包容。 父母在，不远游，游必有方。谨以此文表示出门在外这些年, 并非完全虚度年华。

10 REFERENCES

1. WHO. Global burden of disease. Http://www.Who.Int/healthinfo/global_burden_disease/estimates_regional/en/index1.Html.2013
2. Jonas MA, Oates JA, Ockene JK, Hennekens CH. Statement on smoking and cardiovascular disease for health care professionals. American heart association. *Circulation*. 1992;86:1664-1669
3. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS medicine*. 2006;3:e442
4. Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global burden of disease study. *Lancet*. 1997;349:1498-1504
5. Lloyd-Jones DM, Nam BH, D'Agostino RB, Sr., Levy D, Murabito JM, Wang TJ, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: A prospective study of parents and offspring. *JAMA : the journal of the American Medical Association*. 2004;291:2204-2211
6. Chow CK, Islam S, Bautista L, Rumboldt Z, Yusufali A, Xie C, et al. Parental history and myocardial infarction risk across the world: The interheart study. *Journal of the American College of Cardiology*. 2011;57:619-627
7. Murabito JM, Pencina MJ, Nam BH, D'Agostino RB, Sr., Wang TJ, Lloyd-Jones D, et al. Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. *JAMA : the journal of the American Medical Association*. 2005;294:3117-3123
8. Williams RR, Hunt SC, Heiss G, Province MA, Bensen JT, Higgins M, et al. Usefulness of cardiovascular family history data for population-based preventive medicine and medical research (the health family tree study and the nhlbi family heart study). *The American journal of cardiology*. 2001;87:129-135
9. 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*. 2002;252:247-254
10. Sorensen TI, Nielsen GG, Andersen PK, Teasdale TW. Genetic and environmental influences on premature death in adult adoptees. *The New England journal of medicine*. 1988;318:727-732
11. Mayer B, Erdmann J, Schunkert H. Genetics and heritability of coronary artery disease and myocardial infarction. *Clinical research in cardiology : official journal of the German Cardiac Society*. 2007;96:1-7
12. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics*. 2013;45:25-33
13. Lu X, Wang L, Chen S, He L, Yang X, Shi Y, et al. Genome-wide association study in han chinese identifies four new susceptibility loci for coronary artery disease. *Nature genetics*. 2012;44:890-894
14. Ganesh SK, Arnett DK, Assimes TL, Basson CT, Chakravarti A, Ellinor PT, et al. Genetics and genomics for the prevention and treatment of cardiovascular disease: Update: A scientific statement from the american heart association. *Circulation*. 2013;128:2813-2851
15. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nature genetics*. 2009;41:334-341
16. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for c-reactive protein levels. *Circulation*. 2011;123:731-738
17. Erdmann J, Grosshennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nature genetics*. 2009;41:280-282
18. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics*. 2011;43:333-338

19. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488-1491
20. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. *The New England journal of medicine*. 2007;357:443-453
21. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491-1493
22. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nature genetics*. 2009;41:342-347
23. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837-1847
24. Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in europe: The score project. *European heart journal*. 2003;24:987-1003
25. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: The american heart association's strategic impact goal through 2020 and beyond. *Circulation*. 2010;121:586-613
26. Greenland P, Smith SC, Jr., Grundy SM. Improving coronary heart disease risk assessment in asymptomatic people: Role of traditional risk factors and noninvasive cardiovascular tests. *Circulation*. 2001;104:1863-1867
27. Kaptoge S, Di Angelantonio E, Lowe G, Pepys MB, Thompson SG, Collins R, et al. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: An individual participant meta-analysis. *Lancet*. 2010;375:132-140
28. Arroyo-Espliguero R, Avanzas P, Cosin-Sales J, Aldama G, Pizzi C, Kaski JC. C-reactive protein elevation and disease activity in patients with coronary artery disease. *European heart journal*. 2004;25:401-408
29. Helfand M, Buckley DI, Freeman M, Fu R, Rogers K, Fleming C, et al. Emerging risk factors for coronary heart disease: A summary of systematic reviews conducted for the u.S. Preventive services task force. *Annals of internal medicine*. 2009;151:496-507
30. Buckley DI, Fu R, Freeman M, Rogers K, Helfand M. C-reactive protein as a risk factor for coronary heart disease: A systematic review and meta-analyses for the u.S. Preventive services task force. *Annals of internal medicine*. 2009;151:483-495
31. Wensley F, Gao P, Burgess S, Kaptoge S, Di Angelantonio E, Shah T, et al. Association between c reactive protein and coronary heart disease: Mendelian randomisation analysis based on individual participant data. *Bmj*. 2011;342:d548
32. Danesh J, Collins R, Peto R. Lipoprotein(a) and coronary heart disease. Meta-analysis of prospective studies. *Circulation*. 2000;102:1082-1085
33. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, et al. Genetic variants associated with lp(a) lipoprotein level and coronary disease. *The New England journal of medicine*. 2009;361:2518-2528
34. Kamstrup PR, Tybjaerg-Hansen A, Steffensen R, Nordestgaard BG. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA : the journal of the American Medical Association*. 2009;301:2331-2339
35. Tregouet DA, Konig IR, Erdmann J, Munteanu A, Braund PS, Hall AS, et al. Genome-wide haplotype association study identifies the slc22a3-lpal2-lpa gene cluster as a risk locus for coronary artery disease. *Nature genetics*. 2009;41:283-285
36. Greenland P, Bonow RO, Brundage BH, Budoff MJ, Eisenberg MJ, Grundy SM, et al. Accf/aha 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: A report of the

american college of cardiology foundation clinical expert consensus task force (accf/aha writing committee to update the 2000 expert consensus document on electron beam computed tomography) developed in collaboration with the society of atherosclerosis imaging and prevention and the society of cardiovascular computed tomography. *Journal of the American College of Cardiology*. 2007;49:378-402
37. Witte DR, Westerink J, de Koning EJ, van der Graaf Y, Grobbee DE, Bots ML. Is the association between flow-mediated dilation and cardiovascular risk limited to low-risk populations? *Journal of the American College of Cardiology*. 2005;45:1987-1993
38. Lorenz MW, Markus HS, Bots ML, Rosvall M, Sitzer M. Prediction of clinical cardiovascular events with carotid intima-media thickness: A systematic review and meta-analysis. *Circulation*. 2007;115:459-467
39. Wyman RA, Mays ME, McBride PE, Stein JH. Ultrasound-detected carotid plaque as a predictor of cardiovascular events. *Vascular medicine*. 2006;11:123-130
40. Peters SA, den Ruijter HM, Bots ML, Moons KG. Improvements in risk stratification for the occurrence of cardiovascular disease by imaging subclinical atherosclerosis: A systematic review. *Heart*. 2012;98:177-184
41. Goff DC, Jr., Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Sr., Gibbons R, 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. *Circulation*. 2013
42. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-713
43. Pignone M, Phillips C, Mulrow C. Use of lipid lowering drugs for primary prevention of coronary heart disease: Meta-analysis of randomised trials. *Bmj*. 2000;321:983-986
44. Mihaylova B, Emberson J, Blackwell L, Keech A, Simes J, Barnes EH, et al. The effects of lowering ldl cholesterol with statin therapy in people at low risk of vascular disease: Meta-analysis of individual data from 27 randomised trials. *Lancet*. 2012;380:581-590
45. Taylor F, Huffman MD, Macedo AF, Moore TH, Burke M, Davey Smith G, et al. Statins for the primary prevention of cardiovascular disease. *The Cochrane database of systematic reviews*. 2013;1:CD004816
46. Stone NJ, Robinson J, Lichtenstein AH, Merz CN, Blum CB, Eckel RH, 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*. 2013
47. Nissen SE, Tardif JC, Nicholls SJ, Revkin JH, Shear CL, Duggan WT, et al. Effect of torcetrapib on the progression of coronary atherosclerosis. *The New England journal of medicine*. 2007;356:1304-1316
48. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma hdl cholesterol and risk of myocardial infarction: A mendelian randomisation study. *Lancet*. 2012;380:572-580
49. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nature genetics*. 2013;45:1345-1352
50. Brown JM, Hazen SL. Seeking a unique lipid signature predicting cardiovascular disease risk. *Circulation*. 2014
51. Tannock LR, O'Brien KD, Knopp RH, Retzlaff B, Fish B, Wener MH, et al. Cholesterol feeding increases c-reactive protein and serum amyloid a levels in lean insulin-sensitive subjects. *Circulation*. 2005;111:3058-3062
52. Chawla A. Control of macrophage activation and function by ppars. *Circulation research*. 2010;106:1559-1569
53. Choi JM, Bothwell AL. The nuclear receptor ppars as important regulators of t-cell functions and autoimmune diseases. *Molecules and cells*. 2012;33:217-222

54. Joseph SB, Castrillo A, Laffitte BA, Mangelsdorf DJ, Tontonoz P. Reciprocal regulation of inflammation and lipid metabolism by liver x receptors. *Nature medicine*. 2003;9:213-219
55. Hollman DA, Milona A, van Erpecum KJ, van Mil SW. Anti-inflammatory and metabolic actions of fxr: Insights into molecular mechanisms. *Biochimica et biophysica acta*. 2012;1821:1443-1452
56. Wang H, Tang J, Chen X, Li F, Luo J. Lipid profiles in untreated patients with dermatomyositis. *Journal of the European Academy of Dermatology and Venereology : JEADV*. 2013;27:175-179
57. Park YB, Lee SK, Lee WK, Suh CH, Lee CW, Lee CH, et al. Lipid profiles in untreated patients with rheumatoid arthritis. *The Journal of rheumatology*. 1999;26:1701-1704
58. de Carvalho JF, Bonfa E, Borba EF. Systemic lupus erythematosus and "lupus dyslipoproteinemia". *Autoimmunity reviews*. 2008;7:246-250
59. Esteve E, Ricart W, Fernandez-Real JM. Dyslipidemia and inflammation: An evolutionary conserved mechanism. *Clinical nutrition*. 2005;24:16-31
60. Fall T, Hagg S, Magi R, Ploner A, Fischer K, Horikoshi M, et al. The role of adiposity in cardiometabolic traits: A mendelian randomization analysis. *PLoS medicine*. 2013;10:e1001474
61. Gordon T, Kannel WB, McGee D, Dawber TR. Death and coronary attacks in men after giving up cigarette smoking. A report from the framingham study. *Lancet*. 1974;2:1345-1348
62. Ockene JK, Kuller LH, Svendsen KH, Meilahn E. The relationship of smoking cessation to coronary heart disease and lung cancer in the multiple risk factor intervention trial (mrfit). *American journal of public health*. 1990;80:954-958
63. Sparrow D, Dawber TR. The influence of cigarette smoking on prognosis after a first myocardial infarction. A report from the framingham study. *Journal of chronic diseases*. 1978;31:425-432
64. Salonen JT. Stopping smoking and long-term mortality after acute myocardial infarction. *British heart journal*. 1980;43:463-469
65. LaCroix AZ, Lang J, Scherr P, Wallace RB, Cornoni-Huntley J, Berkman L, et al. Smoking and mortality among older men and women in three communities. *The New England journal of medicine*. 1991;324:1619-1625
66. Boutet M, Bazin M, Turcotte H, Lagace R. Effects of cigarette smoke on rat thoracic aorta. *Artery*. 1980;7:56-72
67. Dadak C, Leithner C, Sinzinger H, Silberbauer K. Diminished prostacyclin formation in umbilical arteries of babies born to women who smoke. *Lancet*. 1981;1:94
68. Nowak J, Murray JJ, Oates JA, FitzGerald GA. Biochemical evidence of a chronic abnormality in platelet and vascular function in healthy individuals who smoke cigarettes. *Circulation*. 1987;76:6-14
69. Berlin JA, Colditz GA. A meta-analysis of physical activity in the prevention of coronary heart disease. *American journal of epidemiology*. 1990;132:612-628
70. Eaton CB. Relation of physical activity and cardiovascular fitness to coronary heart disease, part i: A meta-analysis of the independent relation of physical activity and coronary heart disease. *The Journal of the American Board of Family Practice / American Board of Family Practice*. 1992;5:31-42
71. Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, et al. Physical activity and public health. A recommendation from the centers for disease control and prevention and the american college of sports medicine. *JAMA : the journal of the American Medical Association*. 1995;273:402-407
72. Williams PT. Physical fitness and activity as separate heart disease risk factors: A meta-analysis. *Medicine and science in sports and exercise*. 2001;33:754-761
73. Oguma Y, Shinoda-Tagawa T. Physical activity decreases cardiovascular disease risk in women: Review and meta-analysis. *American journal of preventive medicine*. 2004;26:407-418
74. Sofi F, Capalbo A, Cesari F, Abbate R, Gensini GF. Physical activity during leisure time and primary prevention of coronary heart disease: An updated meta-analysis of cohort studies. *European journal of cardiovascular prevention and rehabilitation : official journal of the European Society of Cardiology,*

Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology. 2008;15:247-257
75. Sattelmair J, Pertman J, Ding EL, Kohl HW, 3rd, Haskell W, Lee IM. Dose response between physical activity and risk of coronary heart disease: A meta-analysis. *Circulation*. 2011;124:789-795
76. Ronksley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA. Association of alcohol consumption with selected cardiovascular disease outcomes: A systematic review and meta-analysis. *Bmj*. 2011;342:d671
77. Prospective Studies C, Whitlock G, Lewington S, Sherliker P, Clarke R, Emberson J, et al. Body-mass index and cause-specific mortality in 900 000 adults: Collaborative analyses of 57 prospective studies. *Lancet*. 2009;373:1083-1096
78. Wormser D, Kaptoge S, Di Angelantonio E, Wood AM, Pennells L, Thompson A, et al. Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: Collaborative analysis of 58 prospective studies. *Lancet*. 2011;377:1085-1095
79. Ni Mhurchu C, Rodgers A, Pan WH, Gu DF, Woodward M, Asia Pacific Cohort Studies C. Body mass index and cardiovascular disease in the asia-pacific region: An overview of 33 cohorts involving 310 000 participants. *International journal of epidemiology*. 2004;33:751-758
80. The global burden of metabolic risk factors for chronic diseases, collaboration. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: A pooled analysis of 97 prospective cohorts with 1.8 million participants. *Lancet*. 2013
81. Gray LJ, Cooper N, Dunkley A, Warren FC, Ara R, Abrams K, et al. A systematic review and mixed treatment comparison of pharmacological interventions for the treatment of obesity. *Obesity reviews : an official journal of the International Association for the Study of Obesity*. 2012;13:483-498
82. Rucker D, Padwal R, Li SK, Curioni C, Lau DC. Long term pharmacotherapy for obesity and overweight: Updated meta-analysis. *Bmj*. 2007;335:1194-1199
83. Finucane MM, Stevens GA, Cowan MJ, Danaei G, Lin JK, Paciorek CJ, et al. National, regional, and global trends in body-mass index since 1980: Systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *Lancet*. 2011;377:557-567
84. Sjostrom L, Narbro K, Sjostrom CD, Karason K, Larsson B, Wedel H, et al. Effects of bariatric surgery on mortality in swedish obese subjects. *The New England journal of medicine*. 2007;357:741-752
85. Buchwald H, Avidor Y, Braunwald E, Jensen MD, Pories W, Fahrbach K, et al. Bariatric surgery: A systematic review and meta-analysis. *JAMA : the journal of the American Medical Association*. 2004;292:1724-1737
86. Look ARG, Wing RR, Bolin P, Brancati FL, Bray GA, Clark JM, et al. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *The New England journal of medicine*. 2013;369:145-154
87. Humphries SE, Talmud PJ, Hawe E, Bolla M, Day IN, Miller GJ. Apolipoprotein e4 and coronary heart disease in middle-aged men who smoke: A prospective study. *Lancet*. 2001;358:115-119
88. Ahmad T, Chasman DI, Mora S, Pare G, Cook NR, Buring JE, et al. The fat-mass and obesity-associated (fto) gene, physical activity, and risk of incident cardiovascular events in white women. *American heart journal*. 2010;160:1163-1169
89. Younis J, Cooper JA, Miller GJ, Humphries SE, Talmud PJ. Genetic variation in alcohol dehydrogenase 1c and the beneficial effect of alcohol intake on coronary heart disease risk in the second northwick park heart study. *Atherosclerosis*. 2005;180:225-232
90. Magnusson PK, Almqvist C, Rahman I, Ganna A, Viktorin A, Walum H, et al. The swedish twin registry: Establishment of a biobank and other recent developments. *Twin research and human genetics : the official journal of the International Society for Twin Studies*. 2013;16:317-329

91. Lichtenstein P, De Faire U, Floderus B, Svartengren M, Svedberg P, Pedersen NL. The swedish twin registry: A unique resource for clinical, epidemiological and genetic studies. *Journal of internal medicine*. 2002;252:184-205
92. Gold CH, Malmberg B, McClearn GE, Pedersen NL, Berg S. Gender and health: A study of older unlike-sex twins. *The journals of gerontology. Series B, Psychological sciences and social sciences*. 2002;57:S168-176
93. McClearn GE, Johansson B, Berg S, Pedersen NL, Ahern F, Petrill SA, et al. Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science*. 1997;276:1560-1563
94. Pedersen NL, Friberg L, Floderus-Myrhed B, McClearn GE, Plomin R. Swedish early separated twins: Identification and characterization. *Acta geneticae medicae et gemellologiae*. 1984;33:243-250
95. Gatz M, Fratiglioni L, Johansson B, Berg S, Mortimer JA, Reynolds CA, et al. Complete ascertainment of dementia in the swedish twin registry: The harmony study. *Neurobiology of aging*. 2005;26:439-447
96. Lind L, Andren B, Arnlov J. The doppler-derived myocardial performance index is determined by both left ventricular systolic and diastolic function as well as by afterload and left ventricular mass. *Echocardiography*. 2005;22:211-216
97. Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: The prospective investigation of the vasculature in uppsala seniors (pivus) study. *Arteriosclerosis, thrombosis, and vascular biology*. 2005;25:2368-2375
98. Hammar N, Alfredsson L, Rosen M, Spetz CL, Kahan T, Ysberg AS. A national record linkage to study acute myocardial infarction incidence and case fatality in sweden. *International journal of epidemiology*. 2001;30 Suppl 1:S30-34
99. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS genetics*. 2012;8:e1002793
100. Willer CJ, Li Y, Abecasis GR. Metal: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190-2191
101. Purcell S. Variance components models for gene-environment interaction in twin analysis. *Twin research : the official journal of the International Society for Twin Studies*. 2002;5:554-571
102. Plomin R, DeFries JC, McClearn GE, McGuffin P. *Behvioral genetics, 5th edn.* New York: Worth Publishers; 2008.
103. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, de Faire U. Genetic influences on chd-death and the impact of known risk factors: Comparison of two frailty models. *Behavior genetics*. 2004;34:585-592
104. Zdravkovic S, Wienke A, Pedersen NL, de Faire U. Genetic influences on angina pectoris and its impact on coronary heart disease. *European journal of human genetics : EJHG*. 2007;15:872-877
105. 11463413, MC N, SM B, G X, HH M. *Mx: Statistical modeling, 6th edn*. Richmond, VA: Department of Psychiatry; 2003.
106. Greenland S. An introduction to instrumental variables for epidemiologists. *International journal of epidemiology*. 2000;29:722-729
107. Brooks-Wilson A, Marcil M, Clee SM, Zhang LH, Roomp K, van Dam M, et al. Mutations in abc1 in tangier disease and familial high-density lipoprotein deficiency. *Nature genetics*. 1999;22:336-345
108. Brunham LR, Singaraja RR, Pape TD, Kejariwal A, Thomas PD, Hayden MR. Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the abca1 gene. *PLoS genetics*. 2005;1:e83
109. Perry JR, Voight BF, Yengo L, Amin N, Dupuis J, Ganser M, et al. Stratifying type 2 diabetes cases by bmi identifies genetic risk variants in lama1 and enrichment for risk variants in lean compared to obese cases. *PLoS genetics*. 2012;8:e1002741
110. Lamina C, Forer L, Schonherr S, Kollerits B, Ried JS, Gieger C, et al. Evaluation of gene-obesity interaction effects on cholesterol levels: A genetic

predisposition score on hdl-cholesterol is modified by obesity. *Atherosclerosis*. 2012;225:363-369
111. Han S, Levoci L, Fischer P, Wang SP, Gagen K, Chen Y, et al. Inhibition of cholesteryl ester transfer protein by anacetrapib does not impair the anti-inflammatory properties of high density lipoprotein. *Biochimica et biophysica acta*. 2013;1831:825-833
112. Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM, Jr., Kastelein JJ, et al. Rosuvastatin to prevent vascular events in men and women with elevated c-reactive protein. *The New England journal of medicine*. 2008;359:2195-2207
113. Maini RN, Taylor PC, Szechinski J, Pavelka K, Broll J, Balint G, et al. Double-blind randomized controlled clinical trial of the interleukin-6 receptor antagonist, tocilizumab, in european patients with rheumatoid arthritis who had an incomplete response to methotrexate. *Arthritis and rheumatism*. 2006;54:2817-2829
114. Kawashiri SY, Kawakami A, Yamasaki S, Imazato T, Iwamoto N, Fujikawa K, et al. Effects of the anti-interleukin-6 receptor antibody, tocilizumab, on serum lipid levels in patients with rheumatoid arthritis. *Rheumatology international*. 2011;31:451-456
115. Medland SE, Neale MC, Eaves LJ, Neale BM. A note on the parameterization of purcell's g x e model for ordinal and binary data. *Behavior genetics*. 2009;39:220-229
116. van der Sluis S, Posthuma D, Dolan CV. A note on false positives and power in g x e modelling of twin data. *Behavior genetics*. 2012;42:170-186
117. Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *Bmj*. 2005;330:1076-1079
118. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Statistical methods in medical research*. 2007;16:309-330
119. Thomas DC, Conti DV. Commentary: The concept of 'mendelian randomization'. *International journal of epidemiology*. 2004;33:21-25
120. Schatzkin A, Abnet CC, Cross AJ, Gunter M, Pfeiffer R, Gail M, et al. Mendelian randomization: How it can--and cannot--help confirm causal relations between nutrition and cancer. *Cancer prevention research*. 2009;2:104-113
121. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical methods in medical research*. 2012;21:223-242
122. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for mendelian randomization studies using multiple genetic variants. *International journal of epidemiology*. 2011;40:740-752
123. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in mendelian randomization studies. *International journal of epidemiology*. 2013;42:1497-1501
124. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology*. 2013;37:658-665
125. Lander ES. The new genomics: Global views of biology. *Science*. 1996;274:536-539
126. 20562874Pritchard JK, Cox NJ. The allelic architecture of human disease genes: Common disease-common variant...Or not? *Human molecular genetics*. 2002;11:2417-2423
127. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: Common disease-common variant...Or not? *Human molecular genetics*. 2002;11:2417-2423
128. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature genetics*. 2010;42:570-575
129. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747-753

130. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature reviews. Genetics*. 2008;9:356-369
131. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *The New England journal of medicine*. 2008;358:667-675
132. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008;455:232-236
133. International schizophrenia consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008;455:237-241
134. Johnson AD, Hwang SJ, Voorman A, Morrison A, Peloso GM, Hsu YH, et al. Resequencing and clinical associations of the 9p21.3 region: A comprehensive investigation in the framingham heart study. *Circulation*. 2013;127:799-810
135. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56-65
136. Edwards SL, Beesley J, French JD, Dunning AM. Beyond gwass: Illuminating the dark road from association to function. *American journal of human genetics*. 2013;93:779-797
137. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*. 2009;325:1246-1250
138. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated snps are more likely to be eqtls: Annotation to enhance discovery from gwas. *PLoS genetics*. 2010;6:e1000888
139. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nature reviews. Genetics*. 2010;11:843-854
140. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via sort1 at the 1p13 cholesterol locus. *Nature*. 2010;466:714-719
141. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*. 2010;11:636-646
142. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, et al. A tale nuclease architecture for efficient genome editing. *Nature biotechnology*. 2011;29:143-148
143. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using crispr/cas systems. *Science*. 2013;339:819-823