IDENTIFICATION AND CHARACTERIZATION OF GENETIC VARIANTS ASSOCIATED WITH LIPID AND LIPOPROTEIN LEVELS

By

Logan Caneel Dumitrescu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements for

the degree of

DOCTOR OF PHILOSOPHY

in

Human Genetics

May, 2011

Nashville, Tennessee

Approved:

Professor Marylyn D. Ritchie

Professor Dana C. Crawford

Professor Jonathan L. Haines

Professor Mary V. Relling

Professor Jay H. Fowke

Copyright © 2011 Logan Caneel Dumitrescu

All Rights Reserved

To my parents, Retta and Steve, for their love, strength, and humor

and

To my husband, Andy, for his patience, love, friendship, and culinary prowess

ACKNOWLEDGEMENTS

The work presented in this dissertation was supported in part by the Vanderbilt Institute for Clinical and Translational Research (VICTR), as well as a Clinical and Translational Science Award (CTSA) grant (1UL1 RR024975-01 from the National Center for Research Resources and the National Institutes of Health [NIRR/NIH]), NCI grant CA 21765, the NIH/NIGMS Pharmacogenomics Research Network (U01 GM92666), the American Lebanese Syrian Associated Charities (ALSAC), and the Human Genetics training grant (T32GM080178). The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALiCo), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004790 (WHI), and U01HG004801 (Coordinating Center). A heartfelt thank you to all the individuals who participated in these studies - this effort would not be possible without them.

The work presented here was guided and greatly improved by the input from my thesis committee: Marylyn Ritchie (my committee chair), Dana Crawford, Jonathan Haines, Mary Relling, and Jay Fowke. Special acknowledgements are due to my mentor, Dana Crawford. Dr. Crawford's expertise, patience, constructive criticism, and friendship have given me the tools I need to be an excellent scientist. She has provided the outstanding training and mentoring and for that I will always be grateful.

I also thank the members of the Crawford lab (Janina Jeff, Sarah Pendergrass, Matt Oetjens, and Sabrina Mitchell) for their creativity and friendship. They brought light into a window-less working space. Thanks to the students of the Program in Human Genetics for their insightful discussions (and arguments). I could not have

asked for better, brighter colleagues. Special thanks to Will Bush, Stephen Turner, Anna Cummings, and Kelli Ryckman, who were always available to lend an ear and a helping hand.

TABLE OF CONTENTS

Page

DEDICATION iii

ACKNOWLEDGEMENTS..... iv

LIST OF TABLES..... ix

LIST OF FIGURES..... xi

LIST OF APPENDICES xii

OVERVIEW..... xv

Chapter

I. **INTRODUCTION.....1**

Structure and Function of Lipids and Lipoproteins 2
Lipids..... 2
Lipoproteins 2
Lipid and lipoprotein metabolism 5
Risk factors for coronary heart disease..... 6
Ethnic Differences in Cardiovascular Disease and Lipid Levels 8
Strategies for Studying the Genetics of Lipids..... 10
Candidate gene approach..... 11
Genome-wide association studies 13
Gene-environment interactions 16
Summary 17

II. **THE ASSOCIATION OF COMMON VARIATION IN *LPA* AND LIPOPROTEIN(A) LEVELS DIFFERS BY RACE/ETHNICITY.....18**

Introduction 18
Methods..... 21
Study population 21
Laboratory measures..... 21
SNP selection and genotyping 22
Statistical methods 23
Genetic risk score calculation..... 27
Ethics statement 27
Results 28
Population characteristics..... 28
LPA SNP associations with Lp(a) levels 32
LPA risk allele distribution..... 38

Discussion 40
Acknowledgements 46

III. **USE OF A GWAS OF CHILDREN TO IDENTIFY NOVEL LIPID-ASSOCIATED VARIANTS 47**

Introduction 47
Methods 49
Results 52
Study population characteristics 52
Discovery GWAS results 54
Replication in NHANES III youths and adults 56
Age as a potential modifier 58
Discussion 60
Acknowledgements 64

IV. **REPLICATION AND GENERALIZATION OF LIPID GWAS-IDENTIFIED VARIANTS FROM THE LITERATURE 66**

Introduction 65
Methods 69
Study populations and phenotypes 68
SNP selection and genotyping 69
Statistical methods 70
Results 71
Study population characteristics 71
Allele frequencies 74
Replication in European-descent populations 75
Associations in non-European-descent populations 85
Generalization across non-European-descent populations 87
Power 88
Linkage disequilibrium 89
Adjustments for exposures and co-morbidities 90
Effect of including versus excluding by medication use 90
Discussion 92
Replication in European-descent populations 92
Generalization to non-European-descent populations 94
Limitations and strengths 95
Conclusions 96
Acknowledgements 97

V. **ENVIRONMENTAL MODIFIERS OF GWAS-IDENTIFIED LIPID-ASSOCIATED VARIANTS 100**

Introduction 100
Methods 102

Study population ... 102
Laboratory measurements ... 103
Determination of smoking status ... 103
SNP selection and genotyping ... 104
Statistical analysis ... 105
Results ... 105
Population characteristics ... 105
SNPxSmoking interactions ... 110
SNPxVitamin A and E interactions ... 112
Discussion ... 114
SNPxSmoking interactions ... 115
SNPxVitamin A and E interactions ... 116
Strengths and limitations ... 118
Conclusions ... 120
Acknowledgements ... 120

VI. **CONCLUSION ... 121**

Summary ... 121
Future Directions ... 125

REFERENCES ... 209

LIST OF TABLES

Table Page

1.1 Summary characteristics of lipoprotein classes 4

1.2 ATP III guidelines for lipid and lipoprotein levels 8

1.3 Overview of the differences in lipid and lipoprotein concentrations of American racial/ethnic groups in comparison to European Americans 9

2.1 SNP location and genotyping quality control metrics, stratified by race/ethnicity 25

2.2 NHANES III study population characteristics 29

2.3 Frequency of *LPA* variants in HapMap samples 31

2.4 *LPA* common haplotypes and haplotype frequencies 32

2.5 Associations between *LPA* SNPs and Lp(a) levels 35

2.6 Additive effects of *LPA* risk alleles on Lp(a) levels 37

3.1 Participant Characteristics 53

3.2 Significant Replication Results in NHANES III 57

3.3 Mean LDL-C levels in Blacks, stratified by rs2429917 genotype and age group 59

4.1 Characteristics of PAGE study populations 73

4.2 Meta-analysis of GWAS-identified HDL-C SNPs 78

4.3 Meta-analysis of GWAS-identified LDL-C SNPs 81

4.4 Meta-analysis of GWAS-identified Triglyceride SNPs 83

5.1 NHANES participant characteristics, stratified by smoking status 107

5.2 NHANES participant characteristics, including serum vitamin levels 108

5.3 Associations between lipid traits and vitamins A and E.. 109

5.4 Significant SNPxenvironment interactions in NHANES 113

LIST OF FIGURES

Figure Page

2.1 Location of genotyped *LPA* SNPs relative to the kringle repeat region and a SNP in the 5′ untranslated region 24

2.2 Overview of association results between *LPA* SNPs and Lp(a) levels 34

2.3 Distribution of *LPA* risk alleles in non-Hispanic whites, non-Hispanic blacks, and Mexican Americans 39

2.4 Pair-wise linkage disequilibrium (r^2) calculated for 19 *LPA* SNPs in non-Hispanic whites, non-Hispanic blacks, and Mexican Americans in NHANES III 41

3.1 Discovery GWAS association results 55

4.1 Meta-analysis results for GWAS-identified SNPs by population 76

5.1 *PLTP*xSmoking interaction effect on HDL-C levels in non-Hispanic blacks 111

LIST OF APPENDICES

Appendix Page

A. Associations between *LPA* SNPs and Lp(a) levels, weighted for selection and non-response biases 127

B. Additive effects of *LPA* alleles associated with increased Lp(a) levels 129

C. Differences in mean HDL-C, LDL-C, and triglycerides among St. Jude Children and NHANES III youths, stratified by race/ethnicity 130

D. Differences in mean HDL-C, LDL-C, and triglycerides among NHANES III adults and NHANES III youths, stratified by race/ethnicity 132

E. St. Jude discovery GWAS associations selected for replication 134

F. Association results of lipid candidate genes and HDL-C, LDL-C, and TG in the St. Jude discovery GWAS 136

G. Descriptions of Participants, Data Collection, and Genotyping by Study 137

H. Study characteristics by PAGE study and population 147

I. List of candidate gene and GWAS-identified SNPs targeted for genotyping in PAGE 152

J. Comparison of unadjusted, minimally adjusted, adjusted models for HDL-C, by population 155

K. Comparison of unadjusted, minimally adjusted, adjusted models for LDL-C, by population 161

L. Comparison of unadjusted, minimally adjusted, adjusted models for triglyceride concentrations, by population 167

M. HDL-C and the effects of lipid lowering medication use on genetic associations, by population 173

N. LDL-C and the effects of lipid lowering medication use on genetic associations, by population 179

O. Transformed triglycerides and the effects of lipid lowering medication use on genetic associations, by population 185

P. Coded allele frequency, by population 191

Q. Generalization of associations across non-European populations for HDL-C, LDL-C and ln(TG) 192

R. Comparison of genetic effect estimates when participants are excluded or included based on medication use with adjustments in WHI 196

S. List of 23 candidate gene and GWAS-identified SNPs genotyped in EAGLE 198

T. Gene-smoking interaction results for HDL-C levels, adjusted for age and sex 200

U. Gene-smoking interaction results for LDL-C levels, adjusted for age and sex 201

V. Gene-smoking interaction results for transformed triglyceride levels, adjusted for age and sex 202

W. Gene-vitamin A interaction results for HDL-C levels, adjusted for age and sex 203

X. Gene-vitamin A interaction results for LDL-C levels, adjusted for age and sex 204

Y. Gene-vitamin A interaction results for transformed triglyceride levels, adjusted for age and sex 205

Z. Gene-vitamin E interaction results for HDL-C levels, adjusted for age and sex 206

AA. Gene-vitamin E interaction results for LDL-C levels, adjusted for age and sex 207

BB. Gene-vitamin E interaction results for transformed triglyceride levels, adjusted for age and sex 208

OVERVIEW

> "What more powerful form of study of mankind could there be than to read our own instruction book?"
>
> — Francis S. Collins

Despite the impressive advances in our knowledge regarding the prevention and treatment of cardiovascular diseases, coronary heart disease (CHD) remains the leading killer of Americans. Dissecting the genetic architecture of this common, complex human disease is extremely challenging. However, the development of new technologies and analytical strategies for large-scale, high-throughput generation of biological data at progressively diminishing costs has greatly expanded our perspective of the genetic landscape.

The blood lipid profile, among other factors, has long been recognized as a significant predictor of CHD. Due to their central role in coronary heart disease, plasma lipids and lipoproteins have been intensively studied for many decades. The high heritability of these traits has lead to much effort in understanding the genetic factors involved. However, as with many complex traits, previous genetic studies have been only modestly successful in identifying new genes. In this work, I endeavor to identify and characterize common genetic variants that explain a proportion of the inter-individual variability in lipids levels, including low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), and lipoprotein(a) [Lp(a)] levels.

The purpose of Chapter I is to provide an overview of the biology and epidemiology of lipids and lipoproteins. I focus on their structure, metabolism, and relevance to coronary heart disease, along with how their mean concentrations differ across several racial/ethnic groups. Lastly, I summarize the strategies I have employed for studying their association with common genetic variants.

In Chapter II, I take a candidate gene approach to determine if variants in the gene that encodes for apolipoprotein(a), *LPA*, are correlated with Lp(a) levels. *LPA* makes for an especially interesting candidate gene for study because it accounts for nearly all the variation observed in Lp(a) levels, which also vary considerably among different racial/ethnic groups. Genotyped samples were drawn from participants of the Third National Health and Nutrition Examination Survey (NHANES III), a population-based, cross-sectional survey in the United States. NHANES III, along with subsequent surveys NHANES 1999-2000 and NHANES 2001-2002, are utilized repeatedly throughout this work, as they provide a large number of DNA samples linked to a myriad of health and environmental variables, including lipid and lipoprotein measurements. Furthermore, the diversity of NHANES samples (which include non-Hispanic whites, non-Hispanic blacks, and Mexican Americans) allows me to explore population-specific genetic associations with Lp(a) levels.

In Chapter III, I use a different approach – a genome-wide association study (GWAS) – to identify novel variants associated with HDL-C, LDL-C, and TG in a diverse cohort of children undergoing treatment for acute-lymphoblastic leukemia, followed by replication in an independent cohort, NHANES III. The majority of GWAS to date have been performed on adults, even though children and adolescents have different lipid distributions compared with adults. In this chapter I ask if it may be advantageous to search for novel lipid-associated genetic variants in a cohort of children, who, presumably have not been exposed to environmental modifiers as long as adults. Furthermore, I explore whether novel variants discovered in children generalize to adults and if these associations vary by age.

To date, approximately 100 lipid-associated variants have been identified through GWAS in samples of European descent. However, data for other racial/ethnic populations is just emerging. Therefore, as part of the Population Architecture using Genomics and Epidemiology (PAGE) consortium, I investigate in Chapter IV whether a select set of GWAS-identified, lipid-associated variants selected from the literature replicate and generalize in an independent cohort

of European Americans, African Americans, Mexican Americans/Hispanics, American Indians, Japanese/East Asians, and Pacific Islanders/Native Hawaiians.

Despite the ever-growing number of loci detected by GWAS, the proportion of trait variation explained is collectively small. To investigate this missing heritability, it is important to explore gene-environment interactions, which may also contribute to trait variation. In Chapter V, I test a subset of the variants explored in Chapter IV for interactions with three putative environmental modifiers (smoking status and levels of vitamin A and vitamin E) in NHANES.

Lastly, in the Chapter VI, I will summarize the work presented in Chapters II through V, emphasizing the benefits and difficulties of the different approaches utilized here. I will also discuss future directions for the field in light of rapidly advancing genotyping and sequencing technologies.

CHAPTER I

INTRODUCTION

Coronary heart disease (CHD) is the leading killer of Americans. In 2007, CHD was responsible for 1 of every 6 deaths in the United States (Roger et al., 2011). While the death rates from CHD have declined over the past decade, the disease remains a significant source of morbidity and economic burden. In 2010, the compilation of health care services, medications, and lost productivity cost the United States $316.4 billion (Lloyd-Jones et al., 2010). With such an astounding impact, new treatments and prevention strategies are desperately needed. The identification of CHD risk factors is key to understanding disease etiology and will be instrumental in advancing clinical care.

The lipid profile has long been recognized as being important in the development of CHD, and within this profile, it is known that decreased high density lipoprotein cholesterol (HDL-C) and increased low density lipoprotein cholesterol (LDL-C) and triglyceride (TG) levels are independent risk factors for CHD in the general population (Third Report of the National Cholesterol Education Program (NCEP) Expert Panel, 2002). Other lipoproteins, such as lipoprotein(a) [Lp(a)], are also emerging as potential risk factors for CHD. This introductory chapter will review (1) the structure, function, and metabolism of these lipids and lipoproteins, (2) their clinical relevance to coronary heart disease, (3) epidemiological differences between racial/ethnic populations, and (4) some of the strategies undertaken to determine genetic variants important in their metabolism.

Structure and Function of Lipids and Lipoproteins

Lipids

The two major lipids in the blood are cholesterol and triglycerides. Cholesterol is the major animal sterol and serves many vital functions, including maintaining the integrity of cell membranes and acting as a precursor for vitamin D, steroid hormones, oxysterols, and bile acids (Moffatt and Stamford, 2006). Cholesterol is found in two forms, free and esterified. The majority of circulating cholesterol is esterified and is synthesized endogenously, with the small remainder originating from our diet (Hegele, 2009).

Triglyceride (also known as triacylglycerol) is the chemical form in which most fat exists in the body. It is composed of three fatty acid chains (a key energy source) esterified to a glycerol backbone. Triglycerides are synthesized both in the liver and in the intestines, are transported through the blood, and following lipolysis at the endothelial surface, deliver free fatty acids to peripheral cells where they are stored or used for energy.

Lipoproteins

Due to their hydrophobic nature, cholesterol and triglycerides must be transported through the circulation via lipoproteins. Lipoproteins consist of a core of hydrophobic triglycerides, fat-soluble vitamins, and cholesteryl esters which is then surrounded by a layer of hydrophilic phospholipids, free cholesterol, and apolipoproteins. The two main triglyceride transporters are chylomicrons and very

low-density lipoprotein (VLDL). The two main cholesterol transporters are high-density HDL and low-density LDL.

Lipoproteins are classified by their density (Table 1.1). Density is both a result of the lipid content (i.e. percent cholesterol and triglyceride) and the lipid/protein ratio. As their name implies, high density lipoproteins are the densest lipoproteins, containing about equal amounts of lipids and proteins. As the lipid/protein ratio increases, lipoprotein size increases and density decreases, with chylomicrons being the least dense lipoprotein since they are composed almost entirely of lipids (98-99% by weight, Table 1.1).

While density determines lipoprotein classification, it is apolipoprotein content that determines lipoprotein function. The major plasma apolipoproteins are described in Table 1.1 (Moffatt and Stamford, 2006; Hegele, 2009; Fless et al., 1994). Apolipoproteins act as cofactors for enzymes and as ligands for uptake by cellular receptors. For example, apolipoprotein B-100, which is encoded by *APOB* and is the primary protein of LDL, facilitates lipid uptake through LDL receptors, which recognize and attach to certain segments of the apoB-100 molecule. In mice, homozygous *APOB* knockouts are embryonic lethal and heterozygotes have decreased LDL-C and HDL-C levels (Huang et al., 1995). Additionally, the characteristic protein component of lipoprotein(a) [Lp(a)] is apolipoprotein(a). As discussed in Chapter II, apo(a) can vary greatly in size and plays a major role in the amount of circulating Lp(a). Therefore, both lipid and apolipoprotein content play key roles in lipoprotein structure and function.

Table 1.1. Summary characteristics of lipoprotein classes. VLDL, very low density lipoprotein; LDL, low density lipoprotein; HDL, high density lipoprotein. Table adapted from Hegele 2009, Moffat & Stamford 2006, and Fless et al 1994.

Lipoprotein Class	Density (g/ml)	Triglyceride (% by weight)	Phospholipid (% by weight)	Free cholesterol (% by weight)	Esterified cholesterol (% by weight)	Protein (% by weight)	Main apolipoproteins
Chylomicrons	<0.94	80-95	3-6	1-3	2-4	1-2	A-I, A-IV, A-V, B-48, E, C-I, C-II, C-III
VLDL	0.94-1.006	45-65	15-20	4-8	16-22	6-10	B-100, E, C-I, C-II, C-III
LDL	1.019-1.063	4-8	18-24	6-8	45-50	18-22	B-100
Lp(a)	1.054-1.090	2-9	15-19	4-6	30-38	25-36	B-100, (a)
HDL	1.063-1.210	2-7	26-32	3-5	15-20	45-55	A-I, A-II, E

__Lipid and lipoprotein metabolism__

As alluded to above, lipids are obtained either exogenously (through diet) or endogenously (hepatic synthesis). Dietary fats are absorbed by the intestine through fatty acid transporters while sterols enter intestinal cells via the Niemann-Pick C1-like 1 (NPC1L1) transporter. The two lipids are then packaged together (via the microsomal TG-transfer protein, MTTP) into chylomicrons, along with apoliprotein B-48, and are secreted into the lymphatic system. Once the chylomicrons have entered the circulatory system they are quickly hydrolyzed by lipoprotein lipase (LPL), thereby releasing free fatty acids to tissues such as adipose and muscle. The remnants of this lipolysis are taken up by the liver, with the help of apo E and the LDL receptor-related protein (LRP), where they are available for catabolism (Lusis et al., 2004; Moffatt and Stamford, 2006).

The liver packages and secretes intestinally derived triglycerides and cholesterol, along with *de novo* synthesized cholesterol (produced via 3-hydoxy-3-methlyglutaryl coenzyme A reductase, HMGCR) as very low-density lipoprotein (VLDL). VLDL is lipolyzed in the circulation by LPL, giving rise to intermediate-density lipoprotein (IDL), which is then hydrolyzed by hepatic lipase (HL) and yields LDL. LDL is then removed from the circulation mainly by LDL receptor (LDLR) in conjunction with apo E. Because the kinetics of LDL uptake is slow, LDL remains in circulation for a longer period of time and is, therefore, the predominant cholesterol-carrying particle (Lusis et al., 2004; Moffatt and Stamford, 2006).

In the process known as reverse cholesterol transport, HDL is formed in circulation from chylomicrons and VLDL surface remnants, along with apo A-1 secreted by the liver and intestine. While LDL delivers cholesterol to cells, HDL removes it.

HDL precursors take up cholesterol from tissues through interactions with ABC transporters, specifically ABCA1. The cholesterol is then esterified by lecithin cholesterol acyl transferase (LCAT). Enzymes such as cholesteryl ester transfer protein (CETP) and phospholipid transfer protein (PLTP) facilitate the exchange of lipids between lipoprotein particles. Finally, HDL is taken up by the liver, via scavenger receptor class B type 1 (SR-B1), where it metabolized (Lusis et al., 2004; Moffatt and Stamford, 2006).

Risk factors for coronary heart disease

The most significant risk factors for CHD and atherosclerosis, the primary cause of CHD, are related to an imbalance of lipids and lipoprotein intake, metabolism, and catabolism. When there are an insufficient number of LDL receptors synthesized or the receptors have low affinity for their apolipoprotein ligand (both of which may be caused by genetic abnormalities), or when there is excess dietary intake of fat, plasma LDL-C may be abnormally elevated and become oxidized. This can lead to lesions and dysfunction of the vascular wall, which then can lead progressively to the attraction and migration of macrophages to the site of the lesion. In a hypercholesterolaemic environment, macrophages engulf oxidized LDL and become foam cells. Foam cells can become a problem when they accumulate along the vascular wall, form fatty streaks, and contribute to the creation of a necrotic core, which must be contained with a fibrous cap. However, rupture of the cap can lead to hemorrhage, formation of a blood clot, occlusion of coronary vessels, heart disease, or stroke.

Along with elevated LDL-C levels, elevated triglycerides and decreased HDL-C levels are well-established risk factors for CHD (Third Report of the National

Cholesterol Education Program (NCEP) Expert Panel, 2002; Gordon et al., 1977; Manninen et al., 1992). Whether or not increased levels of TG constitute an independent risk factor for CHD (due to correlation between TG and HDL-C) is debated among some investigators; however, studies have provided evidence of independence in risk prediction (Hokanson and Austin, 1996; Sarwar et al., 2007).

Lp(a) is an emerging risk factor for cardiovascular disease. Numerous clinical studies have identified high Lp(a) levels (~>30 mg/dl) as a risk factor independent from LDL for a variety of cardiovascular pathologies (Kamstrup et al., 2009; Rhoads et al., 1986; Danesh et al., 2000). Much of the focus on Lp(a) has centered on apo(a)'s strong resemblance to plasminogen, the zymogen for plasmin, which is the primary enzyme for blood clot degradation. However, the exact atherogenic mechanism of Lp(a) remains unknown and *in vivo* investigation of Lp(a) function has been impeded by the lack of availability of small animal models since Lp(a) is expressed only in humans, nonhuman primates, and the European hedgehog.

The clinical relevance of the lipid profile cannot be overstated. Studies have shown that, in men, a rise in total cholesterol (TC) from 200 to 240 mg/dl is associated with threefold increase in mortality from CHD (Stamler et al., 2000). Given the overwhelming relevance of the lipid profile to CHD, the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH) launched the National Cholesterol and Education Program (NCEP) Adult in 1985. The goal of the NCEP is to reduce the number of CHD-related deaths in the United States by reducing the number of Americans with high cholesterol. The NCEP also sponsors expert panels to develop guidelines for health professional. Table 1.2 presents the Third NCEP Adult Treatment Panel's recommendation for lipid levels for adults.

Table 1.2. NCEP ATP III guidelines for lipid and lipoprotein levels

Lipid (mg/dL)	Normal	Borderline High	High	Very High
TC	<200	200-239	≥240	
LDL-C	<129	130-159	160-189	≥190
HDL-C	>60	40-59 (Borderline-low)	≤40 (Low)	
TG	<150	150-199	200-499	≥500

Ethnic Differences in Cardiovascular Disease and Lipid Levels

The relationship between cholesterol levels and coronary heart disease (CHD) holds across all race/ethnicities. However, median levels of lipids and lipoproteins do differ across geographical ancestries and this might contribute to differences in CHD risk observed across different racial/ethnic populations. Overall differences in CHD mortality, HDL-C, LDL-C, TG, and Lp(a) levels among America's major subpopulations, in comparison to those of European Americans, are presented in Table 1.3.

Populations of African-descent (i.e. non-Hispanic blacks or African Americans) have the highest overall mortality rate from CHD of any ethnic group in the United States (Clark et al., 2001). However, African Americans have a more favorable lipid profile (lower levels of total cholesterol and triglycerides and higher levels of HDL-C) compared to European Americans as well as a lower prevalence of hypercholesterolemia (Johnson et al., 1993; Clark et al., 2001; Metcalf et al., 1998). Another consistent finding in African Americans is greater concentrations of lipoprotein(a) [Lp(a)] compared to European Americans (Howard et al., 1994; Metcalf et al., 1998; Clark et al., 2001).

Table 1.3. Overview of the differences in lipid and lipoprotein concentrations of American racial/ethnic groups in comparison to European Americans. ↑, Higher than European Americans; ↓, lower than European Americans; ↔, undetermined; CHD, coronary artery disease; LDL-C, low-density lipoprotein cholesterol; HDL-C, high density lipoprotein cholesterol; TG, triglycerides; Lp(a), lipoprotein(a). Table adapted from Moffat and Stamford 2006.

Racial/Ethnic Population	CHD mortality	LDL-C	HDL-C	TG	Lp(a)
African American	↑	↓	↑	↓	↑
Mexican American	↓	↓	↓	↑	↔
American Indian	↓	↓	↓	↓	↓

Mexican Americans (or Hispanics), have approximately 20% lower rates of mortality due to CHD and cardiovascular disease compared to European Americans, despite their less favorable cardiovascular risk profile, including lower HDL-C and higher triglycerides (Sorlie et al., 1993; Liao et al., 1997). Even though Mexican Americans appear to have lower than expected mortality from CHD, the proportion of total deaths due to CHD is similar to that of European Americans (Third Report of the National Cholesterol Education Program (NCEP) Expert Panel, 2002). In regards to Lp(a), there are inconsistencies between studies. Compared to European Americans, studies have shown Mexican Americans to have both higher (Kamboh et al., 1997) and lower (Haffner et al., 1992) mean Lp(a) levels.

Early reports of CHD mortality in American Indians suggested that this group had lower rates compared to European Americans (Nelson et al., 1990). However, more recent reports suggest that this rate appears to be increasing, possibly related to the high

and increasing prevalence of diabetes in these communities (Howard et al., 1999; Welty et al., 2002). On average, this population has lower total cholesterol, HDL-C, and LDL-C than European Americans (Howard et al., 1999; Welty et al., 2002; Robbins et al., 1996). Interestingly, American Indians have 1.5-3 times lower concentrations of Lp(a) than those of European Americans, and 5-10 times lower than those of African Americans and do not appear to associate with cardiovascular disease risk (Wang et al., 2002; Howard et al., 1994).

Observed differences in the distribution of risk factors, such as the lipid profile, and coronary heart disease between race and ethnic groups are likely a function of the frequency of specific genotypes and interactions with environmental factors. Methods to determine which genotypes and which environmental interactions are important are discussed in the next section.

Strategies for Studying the Genetics of Lipids

A central goal of human genetics is to determine relationships between DNA sequence variants and any resulting phenotypic changes. Lipid and lipoprotein distributions are attractive phenotypes for genetic study for several reasons. First and foremost, lipids are highly heritable. Twin and family studies suggest that up to 80% of lipid trait distributions can be attributed to genetics (O'Connell et al., 1988; Snieder et al., 1999; Heller et al., 1993). Furthermore, lipids, unlike most other complex traits, are not as prone to phenotypic heterogeneity. The use of standardized methods allows for accuracy and precision in trait measurement.

Genetic studies of the lipid profile are also clinically relevant as the lipid profile may be considered an intermediate clinical phenotype for CHD. Just as clinical outcome can be treated as a combination of intermediate phenotypes, so can intermediate phenotypes be treated as a combination of risk factors, both environmental and genetic (Carlson et al., 2004a). Therefore, identifying genetic variants associated with lipid levels may provide insight into the etiology of CHD. The advantage of using the lipid profile as an intermediate phenotype is the number of genetic and environmental factors influencing variation in lipid distribution is presumably smaller than the number of factors affecting CHD (Carlson et al., 2004a).

The search for specific genes involved in the regulation of lipid and lipoprotein levels in the general population has been going on for the past ~30 years. While the goal remains the same (a better understanding of lipid metabolism and its relationship with CHD), the methods of study have evolved greatly over that time. To identify common genetic variation associated with complex traits, such the lipid profile, two approaches stand out: the candidate gene approach and the genome-wide association approach. The benefits and drawbacks of both study designs are summarized below, along with a variant of the study designs: gene-environment interactions.

Candidate gene approach

Candidate gene studies focus on genes with a known or inferred biological function, which may play a role in disease or an observed phenotype. Much like traditional epidemiological approaches, these types of studies have an *a priori* hypothesis that a certain exposure (in this case, a given genotype) may be correlated with a certain disease. In the genetic study of lipids and lipoproteins, candidate gene

studies have generally centered on genes encoding apolipoproteins, lipoprotein receptors, and enzymes involved in lipid metabolism, and transport proteins (Ordovas, 2002). Typically, investigators try to identify variants in or near candidate genes that may have functional consequences (such as a change in the protein or its expression).

Some of the first key genes implicated in lipid metabolism were identified by candidate gene studies on a small subset of patients at the extremes of lipid distribution. For instance, familial hypercholesterolemia (FH), a monogenic disorder characterized by abnormally high concentrations of LDL-C in the blood, was one of the first genetic diseases of lipid metabolism to be characterized (Rader et al., 2003). Brown and Goldstein determined that rare variants in *LDLR*, the gene that encodes the LDL receptor, disrupt the normal control of lipid metabolism, thereby resulting in hypercholesterolemia (Brown and Goldstein, 1986).

There are many benefits of the candidate gene approach. For one, candidate gene studies are well suited for detecting genes with small effect sizes underlying common and complex disease (Collins et al., 1997; Risch and Merikangas, 1996). In contrast to linkage analyses and genome-wide association studies (see below), the whole genome is not scanned and, therefore, fewer markers are genotyped. Testing fewer markers result in fewer statistical tests. That is, in comparison to genome-wide association studies (discussed below), less stringent levels of significance are necessary and the need for correction for multiple testing is minimized.

One of the major difficulties with this approach is that an understanding of the mechanisms underlying the disease/phenotype must already be in place. Most candidate gene studies consider a small number of genes and variants; therefore, deciding which genes to focus on and which variants to genotype is limited by our

current biological knowledge. Another major criticism of candidate gene studies is lack of replication (Ioannidis et al., 2001). Lipid candidate gene studies have been disappointing as only a handful have been replicated consistently across studies and populations, such as the *APOA1/C3/A4/A5* gene cluster (Talmud et al., 2002; Pennacchio et al., 2002), *ABCA1* (Cohen et al., 2004), and *PCSK9* (Cohen et al., 2005). Lastly, variants discovered through candidate gene studies are often rare. For example, over 700 different variants in *LDLR* have been identified; yet, the estimated frequency of patients with familial hypercholesterolemia who are heterozygous for one of those variants in *LDLR* ranges from only 0.1 to 1.5%, depending on the population (Austin et al., 2004).

<u>Genome-wide association studies</u>

While the study of candidate genes and monogenetic disorders has been useful in detecting rare, causative variants that affect lipid levels, common variation likely also plays a major role in the distribution of lipid levels in the general population. The idea that risk for common diseases/phenotypes is influenced primarily by common variants (with frequencies > 5% in the population) is known as the common disease/common variant (CD/CV) hypothesis (Reich and Lander, 2001). This hypothesis is the underlying rationale for genome-wide association studies (GWAS) (Manolio et al., 2009).

Completion of sequencing of the human genome, the International HapMap project, and cost-effective, high-throughput genotyping technologies in the mid-2000's resulted in a new study design – GWAS. In this study paradigm, investigators, in an attempt to capture much of the common genetic variation across the genome, genotype individuals for 250,000 to over 1 million single nucleotide polymorphisms (SNPs). GWAS are indirect association studies that rely on linkage disequilibrium (LD), or the

non-random association of alleles at two or more loci. Presumably, genomic regions that contain a disease-related variant are tagged by a SNP, and are detectable by a significant association between the tagging SNP and the disease/phenotype of interest.

Genome-wide association studies allow for interrogation of the entire genome with few prior assumptions. One of the major driving forces behind GWAS was the desire for discovery of novel genes and pathways important in lipid metabolism, as only ~20 lipid-correlated loci were known. GWAS have successfully identified novel common variants which associate with lipids and lipoprotein concentrations in the general population. Beginning in 2008, a flurry of lipid-related GWAS were published (Kathiresan et al., 2008; Kooner et al., 2008; Sandhu et al., 2008; Wallace et al., 2008; Willer et al., 2008). While the majority of indicated loci were previously known, these studies identified seven new loci associated with HDL-C, LDL-C, and triglycerides (Lusis and Pajukanta, 2008). Then, one year later, three more lipid GWAS modestly expanded the number of associated loci (Aulchenko et al., 2009; Kathiresan et al., 2009; Sabatti et al., 2009). More recently, in the most comprehensive meta-analysis of GWAS to date, a total of 95 lipid-associated loci, including 59 novel associations, were indicated in greater than 100,000 individuals (Teslovich et al., 2010).

Genome-wide association studies have also been successfully applied to genetic studies of Lp(a). In 2009, two GWAS identified variants associated with Lp(a) levels. One study identified a novel SNP in the *LPA* gene that associated with high Lp(a) levels in two cohorts, and with carotid artery disease in one (Ober et al., 2009). A few months later, another study identified two additional variants, also in *LPA*, that resulted in a 2-3-fold increase in circulating Lp(a) levels (Clarke et al., 2009).

Despite the success of GWAS, it has its challenges beyond those of candidate gene studies (McCarthy et al., 2008; Shriner et al., 2007; Williams et al., 2007). For one, due to the nature of GWAS, multiple testing becomes a significant problem. Performing multiple tests can lead to an inflated type 1 error rate, thereby increasing the number of false positive associations. However, there is no consensus on when and how to correct for multiple testing. The Bonferroni corrected threshold of $\sim 5 \times 10^{-8}$ proposed by many (The International HapMap Consortium, 2005; Hoggart et al., 2008) is overly conservative because it assumes that all tests are independent, which is untrue for SNPs in strong linkage disequilibrium. Permutation testing is another option but it is computationally demanding.

A second major issue related to GWAS is that the effect sizes of common SNP associations are, with rare exception, small to modest. For example, most common variants examined by GWAS explain only ~3-5% of the variance of HDL-C, LDL-C, or triglycerides in the population. This observation has several implications. In regards to study design, large meta-analyses and collaborative consortia with very large sample sizes are required to detect these small effects. For example, the large meta-analysis by Teslovich et al examined greater than 100,000 individuals from the United States, Europe, and Australia (Teslovich et al., 2010). Furthermore, some have argued that variants with such small effect sizes are not clinically relevant, only marginally improving the accuracy of disease predication models and may not even be as predictive of traditional risk factors (Lanktree et al., 2008). However, determination of new pathways and targets important in lipid metabolism may inform new drug design and, possibly, lead to changes in clinical practice.

Gene-environment interactions

It has been argued that traditional single-SNP association studies may have reached their limit to detect common variants associated with lipids with small effect sizes (Hegele, 2010). Therefore, to illuminate more of the genetic landscape, different analytical approaches of the traditional study designs are needed. One such approach is accounting for gene-environment interactions. It is well-known that environmental factors play significant roles in shaping complex diseases and phenotypes (Ritchie et al., 2001; Moore and Williams, 2002; Hunter, 2005). It is hypothesized that inclusion of gene-environment interactions may explain more common disease than either genetics or the environment alone, thereby contributing to some of the "missing heritability" (Manolio et al., 2006; Manolio et al., 2009; Hunter, 2005).

In regard to lipid levels, many environmental modifiers have been identified, including diet, exercise, cigarette smoking, postmenopausal estrogen use, oral contraceptive use, and lipid-lowering medication (i.e. statins) use. How these environmental variables interact with genetic variants to shape lipid distributions is still largely left to be determined. For example, *APOE* is probably the most studied candidate gene; however, it is only in the last decade that it's interaction with environmental factors have been examined (Talmud and Humphries, 2002). Studies have shown that certain *APOE* genotypes interact with various cholesterol-lowering interventions (medications, hormone replacement therapy, diet, and exercise) to differentially affect lipid and lipoprotein changes.

Summary

In summary, lipids and lipoproteins are complex, clinically relevant phenotypes that are shaped by genes and the environment. Numerous genes, which play a key role in their metabolism, have been characterized; however, accounting for more genetic determinants of such an important trait is necessary. Futhermore, the modifying effect of the environment on these genetic determinants has not been fully realized. Hopefully future genetic association studies, like the ones outlined in Chapters II-V, will identify factors that impact lipid traits and could be used to further our understanding of their biology and the development of CHD risk.

CHAPTER II

THE ASSOCIATION OF COMMON VARIATION IN *LPA* AND LIPOPROTEIN(A) LEVELS DIFFERS BY RACE/ETHNICITY[1]

Introduction

Lipoprotein (a) [Lp(a)] levels have long been recognized as an independent risk factor for coronary artery disease (CAD) (Bennet et al., 2008; Berglund and Ramakrishnan, 2004; Danesh et al., 2000). However, Lp(a) concentrations and their relationship with cardiovascular disease vary across races/ethnicities. The most notable example of this discrepancy is observed between populations of European- and African-decent. While the mean Lp(a) level is two- to threefold higher in blacks relative to whites (The Emerging Risk Factors Collaboration, 2009; Guyton et al., 1985), elevated plasma Lp(a) levels have been reported to be associated with CAD in whites but have not been clearly demonstrated in blacks (Heiss et al., 1984; Moliterno et al., 1995; Sharrett et al., 2001; Sorrentino et al., 1992; Srinivasan et al., 1991).

The epidemiology of Lp(a) in other US racial/ethnic populations, such as Mexican Americans, is not as well documented and often inconsistent. For example, compared to non-Hispanic whites, studies have shown Mexican Americans to have both higher (Kamboh et al., 1997) and lower (Haffner et al., 1992) mean Lp(a) levels. The

---

[1] Adapted from: Dumitrescu L, Glenn K, Brown-Gentry K, Shephard C, Wong M, Rieder MJ, Smith JD, Nickerson DA, and Crawford DC. Variation in LPA is robustly associated with Lp(a) levels in the Third National Health and Nutrition Examination Survey. *PLoS One* (1) 6; e16604.

underlying cause(s) for these between-population differences has not been fully determined; however, there is evidence for the role of multiple, population-specific alleles in *LPA* (Chretien et al., 2006), the gene that encodes for apolipoprotein(a) [apo(a)], which when bound to apolipoprotein B-100 and a low density lipoprotein (LDL)-like particle forms Lp(a).

Lp(a) levels not only vary dramatically across populations, they also have a remarkable inter-individual variability that ranges from barely detectable to greater than 250 nmol/l (Marcovina et al., 2003). This inter-individual variability has a substantial genetic component. It has been determined that the apolipoprotein(a) gene is the major contributor to Lp(a) levels, accounting for more than 90% of the variance for that trait in European Americans (Boerwinkle et al., 1992).

Two types of genetic variants in *LPA* have been associated with Lp(a) levels: variations in the number of copies of the kringle IV-2 repeat and single nucleotide polymorphisms (SNPs). It has been estimated that the kringle IV-2 repeat alone explains 61-69% of the variability observed in Lp(a) levels in populations of European ancestry (Boerwinkle et al., 1992; Boomsma et al., 2000). In contrast, the kringle repeat appears to explain less of the variability (19-44%) in populations of African descent (Ali et al., 1998; Kraft et al., 1996; Schmidt et al., 2006) and Mexican Americans (22-48%) (Chiu et al., 2000; Rainwater et al., 1997). While the kringle IV-2 repeat polymorphism accounts for a large percentage of the variability of Lp(a) levels, the remaining variance has yet to be explained.

Recent studies have identified common SNPs in *LPA* as strongly associated with Lp(a) levels, explaining up to 36% of the trait variance in populations of European-descent (Clarke et al., 2009; Lanktree et al., 2010; Ober et al., 2009). While several studies

have indicated certain SNPs are in substantial linkage disequilibrium (LD) with the kringle IV-2 repeat polymorphism (Clarke et al., 2009; Lanktree et al., 2010), evidence also exists that some SNPs are in relatively little LD with copy number variation in *LPA* (Crawford et al., 2008) and may be independent contributors to Lp(a) levels. A recent genome-wide association study performed in a Hutterite population with kringle IV-2 repeat polymorphism data identified a SNP associated with Lp(a) levels independent of the kringle repeat, supporting the assumption that some common SNPs in *LPA* are independent of the kringle repeat polymorphisms (i.e., not in linkage disequilibrium) (Ober et al., 2009).

To date, relatively few studies have examined associations between *LPA* common SNPs and Lp(a) levels across multiple, diverse populations and no study has characterized the same panel of *LPA* common SNPs in populations of European-, African-, and Mexican-descent. To better characterize this genotype-phenotype relationship in more diverse populations, we have genotyped 19 European American and African American *LPA* tagSNPs in 7,159 participants from the Third National Health and Nutrition Examination Survey (NHANES III). NHANES III is a diverse, population-based cohort representing Americans of European-, African-, and Mexican-descent (Center for Disease Control and Prevention, 1996). We report the significant association of *LPA* SNPs and Lp(a) levels in this diverse cohort and estimate the proportion of Lp(a) variance explained by these genetic variants.

Methods

Study population

Ascertainment of the Third National Health and Nutrition Examination Survey (NHANES III) and method of DNA collection have been previously described (Crawford et al., 2006; Chang et al., 2009; Steinberg et al., 1997) and so will only be briefly described here. The National Health and Nutrition Examination Surveys are cross-sectional surveys conducted by the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). NHANES III was conducted between 1988-1990 (phase 1) and 1991-1994 (phase 2) (Centers for Disease Control and Prevention, 2004; Centers for Disease Control and Prevention, 1996). Like all the NHANES, NHANES III is a complex survey design that over-sampled minorities (non-Hispanic blacks and Mexican Americans), the young, and the elderly. All NHANES have interviews that collect demographic, socioeconomic, dietary, and health-related data. Also, all NHANES study participants undergo a detailed medical examination at a central location known as the Mobile Examination Center (MEC). The medical examination includes the collection of physiological measurements by CDC medical personnel and blood and urine samples for laboratory tests. Beginning with phase 2 of NHANES III, DNA samples were collected from study participants aged 12 years and older.

Laboratory measures

Serum total cholesterol, triglycerides, and HDL cholesterol were measured using standard enzymatic methods. LDL cholesterol was calculated using the Friedewald

equation, with missing values assigned for samples with triglyceride levels greater than 400 mg/dl. Serum Lp(a) levels were measured immunochemically by enzyme-linked immunosorbant assay (ELISA) (Strategic Diagnostics, Newark, DE), which does not have cross reactivity with plasminogen or LDL and is non-sensitive to apo(a) size heterogeneity (Center for Disease Control and Prevention, 1996). Quality control measures of the Lp(a) assay have been described elsewhere and the reliability of this Lp(a) measurement has been adequately demonstrated (Center for Disease Control and Prevention, 1996).

SNP selection and genotyping

Single nucleotide polymorphisms (SNPs) were selected from SeattleSNPs data on European Americans (n=23) and African Americans (n=24) re-sequenced for SNP discovery as previously described (Crawford et al., 2008). Briefly, tagSNPs were chosen for genotyping in both populations separately using LDSelect (Carlson et al., 2004b) at minor allele frequency (MAF) >5% and r^2>0.80. At the time of tagSNPs selection (2006), *LPA* variation data was not available for Mexican Americans or other Hispanic reference samples. Forty-nine SNPs were considered for genotyping, 35 SNPs were targeted for genotyping, and 20 were successfully genotyped. Genotyping was performed using the Illumina GoldenGate assay (as part of a custom 384 OPA) by the Center for Inherited Disease Research (CIDR) through the National Heart Lung and Blood Institute's Resequencing and Genotyping Service. A display of the chromosomal locations of all 20 *LPA* SNPs, along with their relative locations to the 5′ untranslated region (represented by rs1800769) and the kringle repeat (represented by rs9457952 and rs9457986, which flank the kringle repeat), is presented in Figure 2.1.

Genotyping call rates and tests of Hardy Weinberg Equilibrium stratified by self-reported race/ethnicity were calculated for all genotyped *LPA* SNPs (Table 2.1). The average genotyping call rate for all 20 SNPs was 95.9%. SNP rs4073498 was out of Hardy Weinberg Equilibrium (HWE; $p<0.01$) in all three racial/ethnic groups and was therefore excluded from all analyses as mandated by CDC. Five additional SNPs (rs1321195, rs1652507, rs7755463, rs7450261, and rs41265936) were found to be out HWE in one subpopulation but were carried forward in the analysis. In addition to these quality control metrics, we genotyped blinded duplicates as required by CDC, and all SNPs reported here passed quality control metrics required by CDC. All genotype data reported here were deposited into the NHANES III Genetic database and are available for secondary analysis through CDC.

Statistical methods

Analyses were performed for each self-reported race/ethnicity separately. Quality control measures were implemented in PLINK (Purcell et al., 2007). Tests of association were performed using SAS version 9.1 and were limited to participants greater than 18 years of age who had non-missing Lp(a) levels regardless of fasting status. Each genetic variant was tested for association with ln(Lp(a)+1) levels (a transformation that approximated normality) using linear regression assuming an additive genetic model. Analyses were performed adjusted for age and sex, and results were plotted using Synthesis-View (Pendergrass et al., 2010). Data were accessed remotely from the CDC's Research Data Center (RDC) in Hyattsville, Maryland using Analytic Data Research by Email (ANDRE). Statistical significance was defined as

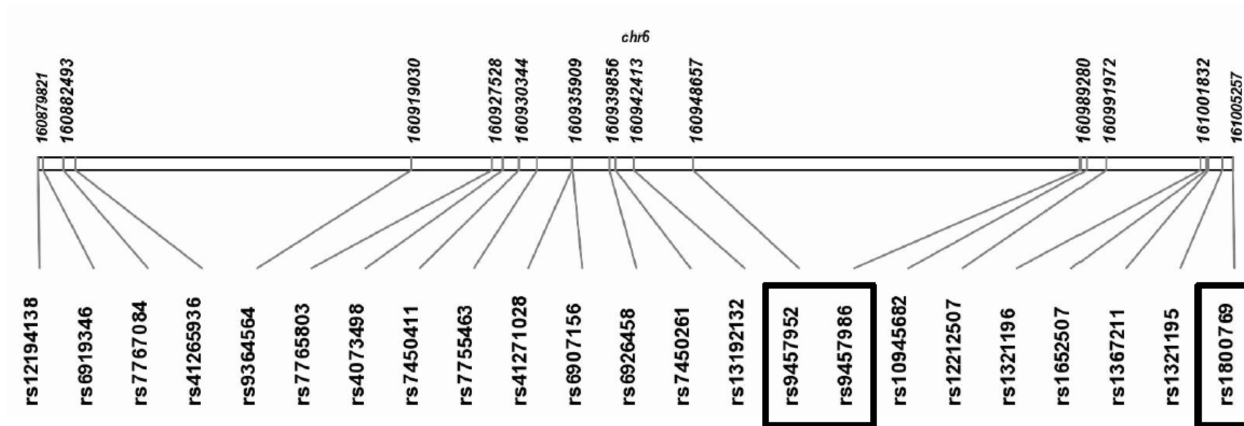


Figure 2.1. Location of genotyped *LPA* SNPs relative to the kringle repeat region and a SNP in the 5′ untranslated region. Synthesis-View (Pendergrass et al., 2010) was used to plot the 20 *LPA* SNPs genotyped in this study. Three other SNPs not genotyped in this study are also represented in this plot within the boxes: rs1800769 (which represents a 5′ UTR SNP genotyped by Rainwater et al 1997 (Rainwater et al., 1997)) and rs9457986 and rs9457952, which flank the kringle repeat. Chromosomal locations are based on genome build 36.

Table 2.1. SNP location and genotyping quality control metrics, stratified by race/ethnicity. Abbreviations: Base pair (bp), genotyping efficiency (GE), Hardy Weinberg Equilibrium (HWE), minor allele (MA), minor allele frequency (MAF). ‡Referent allele listed first

SNP	Position (bp) (Build 36)	Location	Alleles‡	Non-Hispanic Whites n=2,631				Non-Hispanic Blacks n=2,108				Mexican Americans n=2,073			
				G.E. %	HWE	MA	MAF	G.E. %	HWE	MA	MAF	G.E.%	HWE	MA	MAF
rs1321196	161001832	intron	A/G	96.0	0.488	G	0.358	96.1	0.527	G	0.436	96.0	0.910	G	0.272
rs1321195	161004146	intron	A/G	95.8	0.793	A	0.131	96.3	0.695	A	0.030	95.8	4.3E-4	A	0.089
rs1367211	161002685	intron	A/G	95.8	0.249	A	0.274	95.8	1.000	A	0.480	95.8	0.463	A	0.214
rs4073498	160928635	intron	A/G	94.2	4.0E-22	A	0.368	92.3	0.009	G	0.452	94.2	0.005	A	0.272
rs1652507	161002451	intron	A/G	95.5	0.101	G	0.168	96.1	0.760	G	0.078	95.5	0.002	G	0.428
rs6907156	160935999	intron	A/G	96.2	1.000	G	0.003	96.1	0.141	G	0.162	96.2	1.000	G	0.019
rs6919346	160880349	intron	A/G	95.4	0.038	A	0.175	96.3	1.000	A	0.035	95.4	0.895	A	0.094
rs6926458	160939856	intron	A/G	96.0	0.352	G	0.219	96.3	0.557	G	0.106	96.0	0.395	G	0.156
rs7755463	160932260	intron	A/G	96.3	0.001	A	0.005	96.1	0.623	A	0.345	96.3	0.664	A	0.028
rs7767084	160882493	intron	A/G	96.0	0.715	G	0.163	96.4	0.510	G	0.034	96.0	0.789	G	0.147
rs9364564	160919030	intron	A/G	95.8	0.683	A	0.178	96.3	0.397	A	0.086	95.8	0.403	A	0.140
rs12212507	160991972	intron	A/G	96.1	0.246	A	0.055	96.4	1.000	A	0.007	96.1	1.000	A	0.012
rs13192132	160942413	intron	A/G	95.8	0.513	G	0.353	96.3	0.858	G	0.145	95.8	0.504	G	0.245
rs10945682	160989931	intron	A/G	95.6	0.279	A	0.359	96.2	0.223	G	0.438	95.6	1.000	A	0.279
rs12194138	160879821	intron	A/T	96.0	0.776	T	0.168	96.4	1.000	T	0.027	96.0	0.486	T	0.069
rs7450261	160940495	intron	A/G	96.2	1.000	A	0.001	96.0	0.006	A	0.053	96.2	1.000	A	0.002
rs7450411	160930344	intron	A/C	95.6	0.734	A	0.179	96.1	0.637	A	0.136	95.6	0.311	A	0.142
rs7765803	160927528	L1358V	C/G	95.7	0.589	C	0.331	95.9	0.823	G	0.457	95.7	0.729	C	0.264
rs41265936	160883764	G1822A	C/G	96.2	1.000	G	0.001	96.0	4.1E-4	G	0.062	96.2	1.000	G	0.006
rs41271028	160935909	intron	A/T	96.2	1.000	T	0.003	96.4	0.150	T	0.094	96.2	1.000	T	0.013

$p<0.0001$, which represents the Bonferroni corrected p-value [$p=0.0008=0.05/(20$ SNPs x 3 populations)]. Using STATA 10.1, the frequency of risk alleles was compared between populations using Pearson's chi-squared test. Pair-wise linkage disequilibrium (r^2) was calculated using the Genome Variation Server provided by SeattleSNPs (http://gvs.gs.washington.edu/GVS/). Haplotypes were inferred by SAS/Genetics using the expectation-maximization algorithm in each subpopulation separately.

To account for oversampling and non-response in the survey, the National Center for Health Statistics provides a weighting methodology, which has been described elsewhere (Lohr, 1999). In general, weights are often applied in large-scale health surveys that have complicated sampling schemes (such as NHANES) so that one can estimate the level of a variable in the U.S. population at large, as opposed to just in one's sample population. However, since weights were not designed to produce estimates from a sample created using our specific restrictions (i.e. adults 18 years or older with *LPA* genotypes and Lp(a) measurements) and use of sampling weights when unnessary may lead to an inefficient, underpowered analysis (Korn and Graubard, 1991), unweighted results are presented in the main body of the paper, and thus point estimates are not representative of the U.S. population. Nevertheless, we performed tests of association both unweighted (using SAS version 9.1) and weighted (using SUDAAN). Unsurprisingly (Korn and Graubard, 1991), the results of the weighted analyses (Appendix A) were generally less significant than those in the unweighted analysis; however, the proportion of risk alleles when comparing the three populations remained equivalent.

Genetic risk score calculation

The Genetic Risk Score (GRS) was calculated for every participant, respective to each population separately, using SNPs that were associated with transformed Lp(a) levels at $p<0.0001$. We used a count method and assumed each SNP to be independently associated with increased levels of Lp(a). Assuming an additive genetic model for each SNP, a value of 2 was given to individuals who were homozygous for the "risk" allele (i.e. the allele associated with increased levels of transformed Lp(a) levels). Values of 1 and 0 were given to genotypes containing 1 or 0 copies of the risk allele, respectively. The GRS was calculated summing the number of risk alleles at each locus. Participants with incomplete genotype data at any SNP used in the GRS were excluded from analysis. Linear regression, with continuous GRS as the independent variable, was used to evaluate the joint effects (R^2) of associated genetic variants for Lp(a) trait variation. A weighted GRS (WGRS) was also calculated by multiplying each β-coefficient from adjusted tests of association by the number of risk alleles, and then summing the products. Compared to the GRS, the results of the WGRS do not appreciably differ (Appendix B); therefore, GRS was used for the main analyses.

Ethics statement

All procedures were approved by the CDC Ethics Review Board and written informed consent was obtained from all participants. This candidate gene association study was approved by the CDC Ethics Review Board (protocols #2003-08 and #2006-11) and the University of Washington's Institutional Review Board (IRB #23667; HSRC D committee). Because no identifying information was accessed by the investigators,

this study was considered exempt from Human Subjects by Vanderbilt University's Institutional Review Board (IRB #061062; HS2 committee).

Results

Population characteristics

Characteristics of the NHANES III study participants are shown in Table 2.2. Genetic NHANES III included 2,631 non-Hispanic whites, 2,108 non-Hispanic blacks, and 2,073 Mexican Americans. As expected (Marcovina et al., 1996), the mean Lp(a) level in non-Hispanic blacks was 43.4 mg/dL (SD, 32.8 mg/dL), a twofold increase compared to non-Hispanic whites and a three-fold increase compared to Mexican Americans. Mexican Americans had significantly lower mean Lp(a) levels compared to whites ($p<0.0001$). Body mass index (BMI) was similar across all three populations ($p=0.093$). Demographic variables age and sex, along with other blood lipid traits LDL-C, HDL-C, and triglycerides, differed significantly ($p<0.0001$) across populations.

TagSNP allele frequencies are presented in Table 2.1, by population. We calculated the Pearson correlation coefficient (r) between each of the three populations. Not surprisingly (Crawford et al., 2008; Bamshad, 2005), *LPA* allele frequencies observed in non-Hispanic whites were highly correlated with allele frequencies observed in Mexican Americans ($r=0.80$). Also as expected (Bamshad, 2005; Crawford et al., 2008; Carlson et al., 2003), we observed weaker correlation between allele frequencies in non-Hispanic blacks compared with non-Hispanic whites ($r=0.60$) and Mexican Americans ($r=0.48$). Furthermore, compared with non-Hispanic whites, the proportion of SNPs that

differed in allele frequency by more than ±0.10 was smaller in Mexican Americans (2/19 SNPs; 11%) than in blacks (11/19 SNPs; 58%).

Table 2.2. NHANES III study population characteristics. Study characteristics are shown for participants greater than 18 years of age who had non-missing Lp(a) levels**.** Samples sizes shown are the DNA samples available in Genetic NHANES III for each subpopulation. Values are shown as unweighted mean (SD). P-values are based on one-way unweighted ANOVA.

Variable	**Non-Hispanic Whites** (n=2,631)	**Non-Hispanic Blacks** (n=2,108)	**Mexican Americans** (n=2,073)	**P-value**
Males, %	40.0%	42.6%	49.4%	<0.0001
Age (yr)	50.2 (22.3)	36.0 (18.3)	37.1 (18.7)	<0.0001
BMI (kg/m^2)	26.3 (5.6)	27.3 (6.8)	27.0 (5.6)	0.093
Lp(a) (mg/dL)	20.3 (24.1)	43.4 (32.8)	14.9 (8.5)	<0.0001
HDL-C (mg/dL)	50.2 (15.6)	53.8 (16.4)	48.0 (13.1)	0.112
LDL-C (mg/dL)	127.0 (38.0)	118.8 (39.5)	116.3 (34.1)	<0.0001
TG (mg/dL)	147.63 (116.8)	108.8 (79.9)	154.1 (121.2)	<0.0001

We also compared the allele frequencies of these *LPA* SNPs in NHANES III to those in HapMap (Frazer et al., 2007) (Table 2.3). Among the 12 *LPA* SNPs that overlapped this dataset and HapMap, we observed extremely high correlations (r=0.99) in allele frequencies between non-Hispanic whites and HapMap CEU (US individuals of northern and western European ancestry) and between non-Hispanic blacks and both HapMap YRI (Yoruba from West Africa, r>0.99) and ASW (individuals with African ancestry from the Southwest USA, r=0.99). Mexican American allele frequencies were also very similar (r=0.93) to those of HapMap MEX (individuals with Mexican ancestry in Los Angeles, California). Because Mexican Americans are a historically admixed

population, a comparison with HapMap Asian populations was performed. The correlation between NHANES Mexican Americans and HapMap Han Chinese (HCB) and Japanese (JPT) was 0.77 and 0.78, respectively.

Haplotype frequencies were inferred for the 19 tagSNPs in *LPA* by NHANES III subpopulation. We observed eight common haplotypes (frequency >5%) in at least one subpopulation (Table 2.4). While two haplotypes (#1 and #2) were common across all three populations, the remaining haplotypes were either common only to non-Hispanic blacks (#7 and #8), only non-Hispanic whites (#6), or shared between whites and Mexican Americans (#3, #4, #5). As expected (Crawford et al., 2004), the majority of chromosomes from non-Hispanic whites (71.5%) and Mexican Americans (72.6%) were represented by common haplotypes inferred from *LPA* tagSNPs. Only approximately half of the chromosomes from non-Hispanics blacks (55.7%) were represented by common haplotypes, and the remaining half are scattered across rare haplotypes.

Table 2.3. Frequency of *LPA* variants in HapMap samples. Abbreviations: Utah residents with Northern and Western European ancestry (CEU); Yoruba in Ibadan, Nigeria (YRI); Han Chinese in Beijing, China (HCB); Japanese in Tokyo, Japan (JPT); African ancestry in Southwest USA (ASW); Mexican ancestry in Los Angeles, California (MEX); minor allele (MA); minor allele frequency (MAF); monomorphic (mono); SNPs designated not applicable (N/A) were not genotyped in HapMap in that particular population.

SNP	Alleles	CEU n=60		YRI n=60		HCB n=45		JPT n=45		ASW n=47		MEX n=47	
		MA	MAF	MA	MAF	MA	MAF	MA	MAF	MA	MAF	MA	MAF
rs1321196	C/T	C	0.38	C	0.43	T	0.49	C	0.49	C	0.47	C	0.26
rs1321195	A/G	A	0.17	mono	0.00	A	0.22	A	0.28	N/A	N/A	N/A	N/A
rs1367211	T/C	T	0.30	C	0.46	T	0.19	T	0.29	T	0.50	T	0.17
rs1652507	C/T	C	0.13	C	0.06	C	0.38	C	0.38	C	0.06	C	0.41
rs6907156	C/T	C	0.01	N/A	N/A	mono	0.00	mono	0.00	N/A	N/A	N/A	N/A
rs6919346	T/C	T	0.15	T	0.01	mono	0.00	mono	0.00	T	0.06	T	0.16
rs6926458	G/A	G	0.25	G	0.10	A	0.49	G	0.48	G	0.16	G	0.16
rs7755463	T/C	T	0.01	T	0.40	mono	0.00	mono	0.00	T	0.30	N/A	N/A
rs7767084	C/T	C	0.14	mono	0.00	C	0.26	C	0.31	C	0.03	C	0.09
rs10945682	A/G	A	0.38	G	0.38	G	0.49	A	0.50	G	0.49	A	0.26
rs7450261	C/T	mono	0.00	T	0.08	mono	0.00	mono	0.00	T	0.05	N/A	N/A
rs7765803	C/G	C	0.33	G	0.40	G	0.50	G	0.48	G	0.48	C	0.22

Table 2.4. ***LPA*** **common haplotypes and haplotype frequencies.** Only haplotypes with frequencies > 5% in at least one population are displayed. Alleles are ordered based on chromosomal location (5' to 3'). Frequencies > 5% are in bold.

Haplotype Number	Haplotype Alleles	Frequency in Non-Hispanic Whites	Frequency in Non-Hispanic Blacks	Frequency in Mexican Americans
1	A-C-T-C-G-G-C-C-A-T-A-C-T-G-G-A-G-G-C	**0.076**	**0.062**	**0.391**
2	A-C-T-C-G-G-C-C-A-T-A-C-T-G-G-A-A-G-C	**0.207**	**0.253**	**0.142**
3	A-T-T-C-G-G-C-C-A-T-A-C-T-G-G-A-A-G-C	**0.151**	0.030	**0.078**
4	A-C-C-C-A-C-A-C-A-T-G-C-C-A-G-G-A-A-T	**0.088**	0.016	**0.058**
5	A-C-T-C-G-C-C-C-A-T-A-C-C-A-G-G-A-A-C	**0.107**	0.018	**0.057**
6	T-C-T-C-G-G-C-C-A-T-A-C-T-G-G-A-A-G-C	**0.086**	0.012	0.040
7	A-C-T-C-G-C-C-T-T-C-A-C-T-A-G-G-A-A-C	0.002	**0.091**	0.013
8	A-C-T-C-G-C-C-T-A-T-A-C-T-A-G-G-A-A-C	0.001	**0.151**	0.007

LPA *SNP associations with Lp(a) levels*

Each SNP was tested for an association with transformed Lp(a) levels. Results from this analysis are presented in Figure 2.2 and Table 2.5. After adjusting for age and sex, 15 of the 19 SNPs tested were significantly associated with Lp(a) levels in at least one subpopulation at $p<0.0001$, meeting the standard Bonferroni p-value threshold for multiple testing. Among non-Hispanic whites, we confirmed previous evidence of a strong association with rs6919346 ($p=1.2\times10^{-30}$) (Clarke et al., 2009; Ober et al., 2009), which explained approximately 6% of the trait variance ($R^2=0.057$) in our dataset. We also identified two novel associations with rs6926458 and rs12194138 ($p=5.3\times10^{-6}$ and 2.1×10^{-13}, respectively). To evaluate the combined effects of significantly associated

variants, we calculated a continuous Genetic Risk Score (GRS) for each participant based on his or her total number of risk (i.e. Lp(a) increasing) alleles at each associated SNP. Based on the GRS, the additive effect of rs6919646, rs6926458, and rs12194138 explained 7% of the variation in transformed Lp(a) levels in non-Hispanic whites (Table 2.6).

Mexican Americans had twice the number of significant associations compared with non-Hispanic whites, with six SNPs associated with transformed Lp(a) levels at $p<0.0001$. One SNP in particular, rs1652507, was strongly associated at $p=5.44 \times 10^{-34}$ and had the largest effect size of all the associations ($R^2=0.086$). Two of the six associated SNPs (rs1321195 and rs7765803) have previously been associated with Lp(a) in a cohort of Europeans (Clarke et al., 2009). The joint effect of all six associated SNPs, as measured by the GRS, explained 11% of the variance in Lp(a) trait distribution observed in Mexican Americans.

Of the three subpopulations, non-Hispanic blacks had the greatest number of significant associations at $p<0.0001$ with 12 SNPs. Each associated SNP contributed 1% to 4.5% of the trait variance, with the additive effect of the SNPs contributing up to 9% of the total variance in Lp(a) levels. Five of the 12 associated SNPs (rs1321195, rs1652507, rs6919346, rs6926458, and rs7755463) were also associated in one of the two other racial/ethnic groups, non-Hispanic whites or Mexican Americans, and the directions of the effect (beta) were consistent across the associated subpopulations.

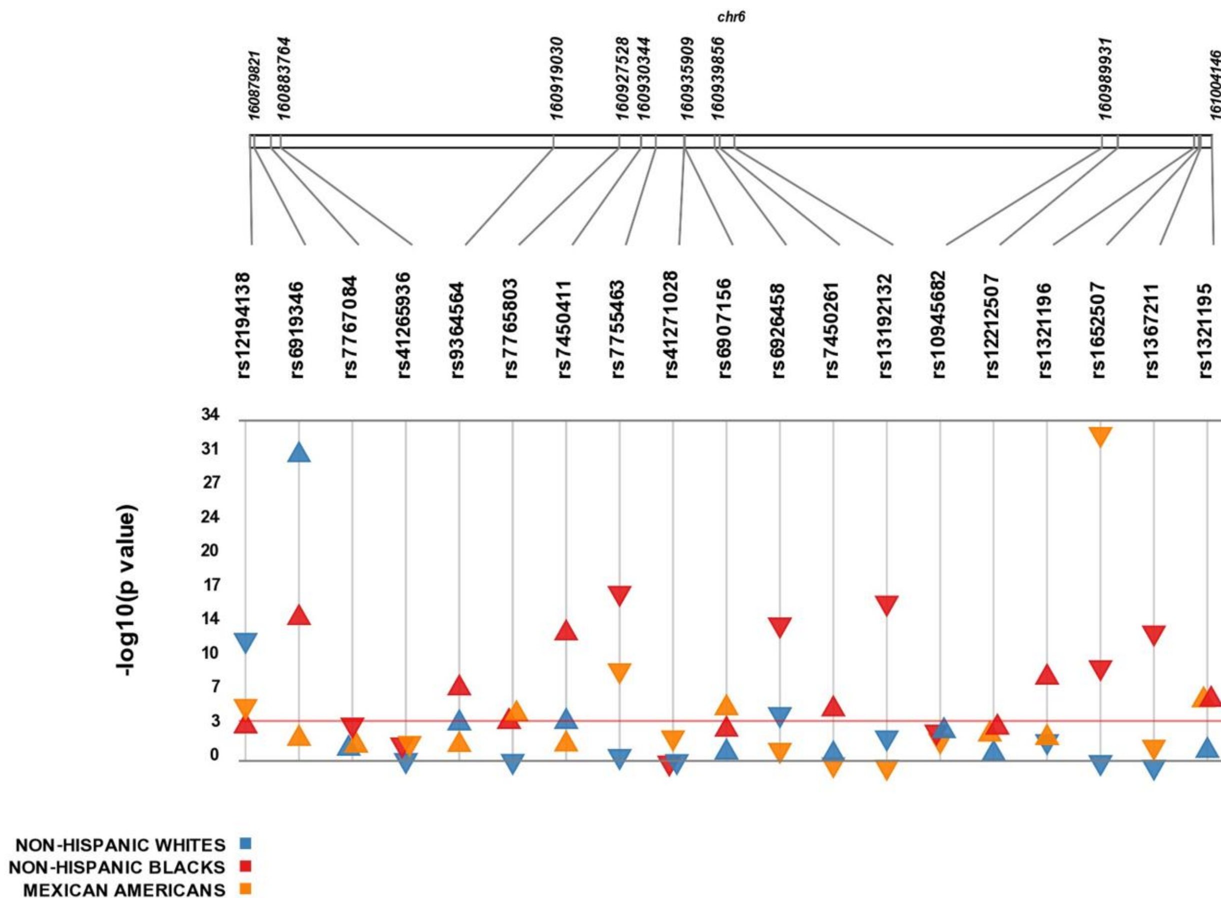


Figure 2.2. Overview of association results between *LPA* SNPs and Lp(a) levels. Plot showing the significance of all single-SNP associations with transformed Lp(a) levels. All results are unweighted adjusted for age and sex and are stratified by race/ethnicity. SNPs are plotted on top along the *x* axis in order from 5′ to 3′, and association with Lp(a) is indicated on the *y* axis (as $-\log_{10}$ p-value). Red line indicates p-value of 1×10^{-4}. Direction of the triangle indicates direction of effect (β-coefficient).

Table 2.5. Associations between *LPA* SNPs and Lp(a) levels. The association of *LPA* SNPs with log transformed Lp(a) levels is shown by a regression coefficient (beta, β) and 95% confidence interval (CI) for each SNP, adjusted for age and sex. Measures of variance explained (R^2) are provided for each SNP based on unadjusted regressions. Significant associations (P-value < 0.0001) are in bold.

SNPs	Non-Hispanic Whites n=2,397			Non-Hispanic Blacks n=1,711			Mexican Americans n=1,749		
	β (95% CI)	R^2	P-value	β (95% CI)	R^2	P-value	β (95% CI)	R^2	P-value
rs1321196	-0.13 (-0.22, -0.05)	0.0036	0.0026	0.21 (0.14, 0.28)	0.0193	**1.88×10^{-8}**	0.12 (0.02, 0.22)	0.0031	0.0204
rs1321195	0.05 (-0.07, 0.18)	0.0003	0.3863	0.53 (0.31, 0.75)	0.0123	**3.03×10^{-6}**	0.40 (0.23, 0.56)	0.0127	**3.49×10^{-6}**
rs1367211	-0.004 (-0.10, 0.09)	0.0000	0.9330	-0.27 (-0.34, -0.20)	0.0345	**3.67×10^{-14}**	-0.15 (-0.26, 0.04)	0.0039	0.0092
rs1652507	-0.05 (-0.16, 0.05)	0.0005	0.3295	-0.45 (-0.59, -0.32)	0.0247	**1.06×10^{-10}**	-0.54 (-0.63, -0.46)	0.0858	**5.44×10^{-34}**
rs6907156	0.21 (-0.51, 0.92)	0.0001	0.5722	0.15 (0.05, 0.25)	0.0053	0.0031	0.73 (0.39, 1.06)	0.0103	**2.14×10^{-5}**
rs6919346	0.61 (0.51, 0.71)	0.0565	**1.18×10^{-30}**	0.75 (0.56, 0.94)	0.0344	**2.16×10^{-14}**	0.18 (0.02, 0.33)	0.0029	0.0248
rs6926458	-0.23 (-0.32, -0.13)	0.0087	**5.29×10^{-6}**	-0.45 (-0.57, -0.34)	0.0364	**5.90×10^{-15}**	-0.15 (-0.28, 0.03)	0.0032	0.0189
rs7755463	-0.47 (-0.99, 0.06)	0.0014	0.0823	-0.33 (-0.40, -0.25)	0.0449	**4.01×10^{-18}**	-0.89 (-1.16. -0.62)	0.0230	**2.14×10^{-10}**
rs7767084	0.07 (-0.05, 0.18)	0.0006	0.2467	-0.41 (-0.61, -0.21)	0.0097	**5.88×10^{-5}**	0.10 (-0.03, 0.23)	0.0015	0.1187
rs9364564	0.19 (0.08, 0.29)	0.0049	0.0007	0.33 (0.21, 0.46)	0.0156	**2.29×10^{-7}**	0.11 (-0.02, 0.24)	0.0016	0.1053

rs12212507	0.03 (-0.15, 0.21)	0.0001	0.7633	0.63 (0.23, 1.03)	0.0063	0.0019	0.57 (0.14, 1.00)	0.0037	0.0093
rs13192132	-0.14 (-0.23, -0.06)	0.0043	0.0011	-0.43 (-0.53, -0.33)	0.0425	**3.85x10^{-17}**	-0.003 (-0.11, 0.10)	0.0000	0.9508
rs10945682	0.12 (0.04, 0.21)	0.0033	0.0041	-0.14 (-0.21, -0.06)	0.0079	0.0003	-0.16 (-0.26, -0.06)	0.0055	0.0022
rs12194138	-0.41 (-0.52, -0.30)	0.0229	**2.05x10^{-13}**	0.34 (0.12, 0.56)	0.0020	0.0014	-0.47 (-0.66, -0.28)	0.0149	**7.97x10^{-7}**
rs7450261	0.10 (-1.29, 1.48)	0.0000	0.8907	0.35 (0.18, 0.51)	0.0108	**2.84x10^{-5}**	-0.33 (-1.34, 0.68)	0.0003	0.5176
rs7450411	0.19 (0.08, 0.30)	0.0051	0.0005	0.37 (0.27, 0.47)	0.0308	**7.37x10^{-13}**	0.12 (-0.02, 0.25)	0.0017	0.0910
rs7765803	-0.05 (-0.14, 0.04)	0.0005	0.2471	0.13 (0.06, 0.20)	0.0071	0.0005	0.21 (0.10, 0.31)	0.0089	**8.54x10^{-5}**
rs41265936	-0.80 (-2.04, 0.44)	0.0007	0.2037	-0.22 (-0.37, -0.06)	0.0047	0.0054	-0.88 (-1.48, -0.28)	0.0046	0.0039
rs41271028	-0.45 (-1.22, 0.32)	0.0006	0.2556	-0.06 (-0.18, 0.07)	0.0007	0.3608	-0.66 (-1.06, -0.27)	0.0061	0.0011

Table 2.6. Additive effects of *LPA* risk alleles on Lp(a) levels. The amount of variance explained (R^2) in transformed Lp(a) levels by the Genetic Risk Score (GRS) is displayed, along with the regression coefficient (beta, β) and 95% confidence interval (CI) for each association.

	Non-Hispanic Whites	**Non-Hispanic Blacks**	**Mexican Americans**
Total n	2269	1605	1665
No. SNPs used in GRS	3	12	6
β (95% CI)	0.36 (0.31-0.42)	0.11 (0.09-0.13)	0.34 (0.29-0.39)
P-value	$<10^{-35}$	$<10^{-34}$	$<10^{-41}$
R^2	0.07	0.09	0.11

LPA *risk allele distribution*

The proportion of risk alleles (i.e. the total number of risk alleles divided by the number of risk alleles possible) was examined across all NHANES subpopulations (Figure 2.3). In general, the distributions differed greatly among non-Hispanic whites, non-Hispanic blacks, and Mexican Americans. In non-Hispanic whites, the proportion of risk alleles followed a normal distribution and the average (mean) number of risk alleles was 3.5 out of the possible six risk alleles (58.3%). In contrast, the distribution of risk alleles was skewed to the left in non-Hispanic blacks and to the right in Mexican Americans (Figure 2.3). The average number of risk alleles in non-Hispanic black participants was 17 out of 24 possible risk alleles (70.8%) while the average number in Mexican American participants was 5.5 out of the 12 possible risk alleles (45.8%). Overall, non-Hispanic blacks had the largest genetic burden of all three subpopulations defined by these alleles, with 99.0% percent of participants possessing greater than 50% of the possible risk alleles. This genetic burden is significantly greater than that carried by non-Hispanic whites (51.4%, $p<0.001$) or Mexican Americans (2.7%, $p<0.001$). Figure 2.3 also illustrates mean Lp(a) levels in participants with various proportions of risk alleles. As expected, mean Lp(a) is higher in participants with a greater proportion of risk alleles, again reflecting the role that these variants may play in contributing to both between- and within-population Lp(a) trait variation.

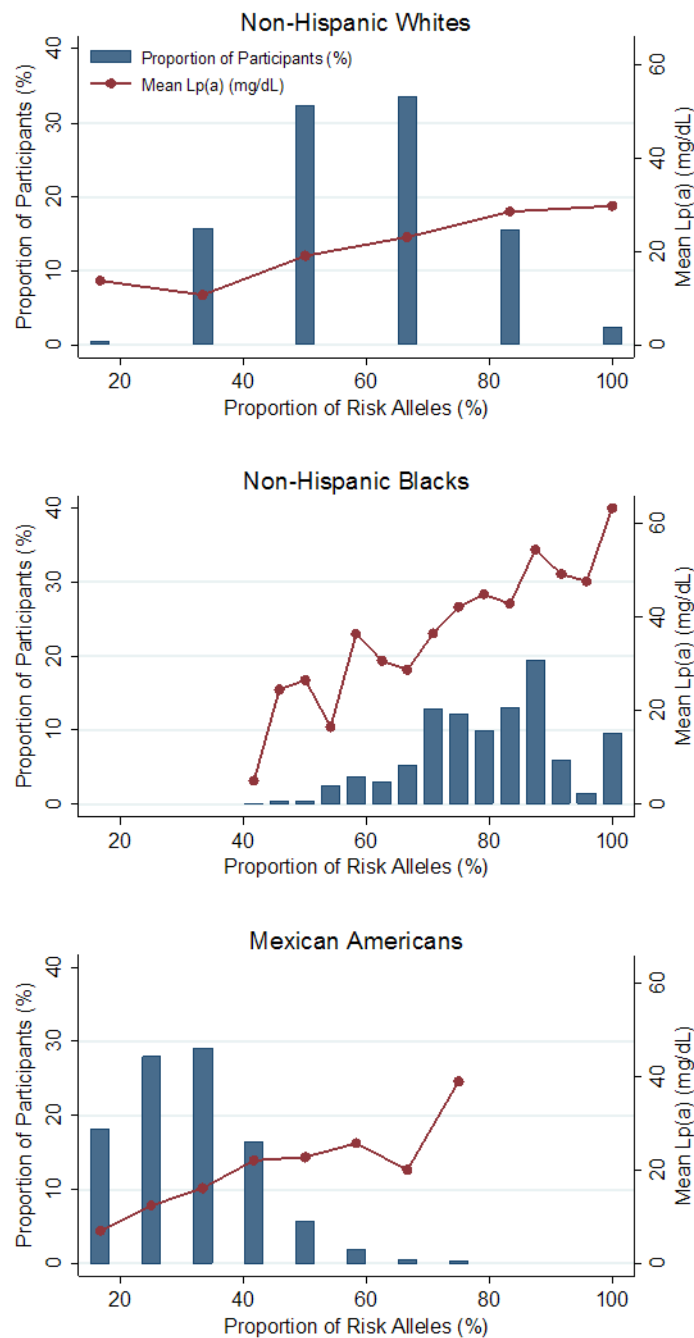


Figure 2.3. Distribution of *LPA* risk alleles in non-Hispanic whites, non-Hispanic blacks, and Mexican Americans. Plots showing the frequency distributions of the proportion of risk alleles in the three NHANES III subpopulations. Proportion of risk alleles was calculated by dividing the total number of *LPA* risk alleles (i.e. the GRS) by the total number of possible risk alleles in each population, multiplied by 100%. Mean Lp(a) values are also plotted for each corresponding proportion.

Discussion

In this study, we identified several variants in the *LPA* gene that are strongly associated with Lp(a) levels in a diverse epidemiologic study. More specifically, three SNPs in non-Hispanic whites, twelve SNPs in non-Hispanic blacks, and six SNPs in Mexican Americans were strongly associated at $p<0.0001$. While no single *LPA* variant was significantly associated in all three racial/ethnic groups, six SNPs were significantly associated in two subpopulations and the directions of effects were consistent.

Most previously published studies characterizing the relationship between Lp(a) and *LPA* have focused on the effects of the kringle IV-2 copy number polymorphism. More recently, a genome-wide association study in Hutterites identified one SNP in *LPA* (rs6919346) that associated with Lp(a) levels, independent of kringle IV-2 copy number (Ober et al., 2009). Subsequent studies have found this variant to be independently associated with increased Lp(a) levels in European Caucasians (Clarke et al., 2009; Lanktree et al., 2010) and South Asians and Chinese (Lanktree et al., 2010). In our study, the same allele (G) was also strongly associated with increased trait levels not only in non-Hispanic whites ($\beta=0.61$, $p=1.18\times10^{-30}$) but also in non-Hispanic blacks ($\beta=0.75$, $p=2.16\times10^{-14}$). In contrast, the association in Mexican Americans was much less robust ($p=0.02$), but the effect trended in the same direction ($\beta=0.18$). This intronic tagSNP is not in linkage disequilibrium (LD) with any genotyped SNP (Figure 2.4), nor with the kringle IV repeat (Crawford et al., 2008). As others have suggested (Ober et al., 2009), rs6919346 may be tagging the causal variant or, due to the fact that it resides in a CRE-binding site, may play a role in gene expression.

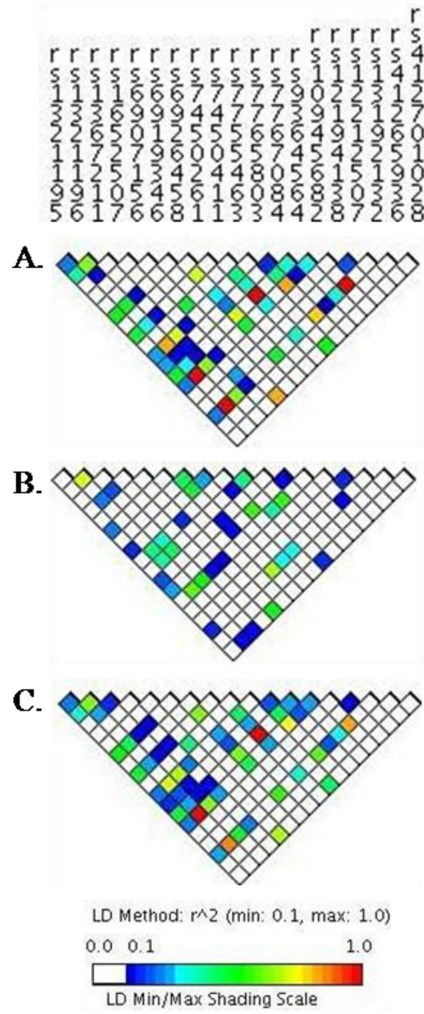


Figure 2.4. Pair-wise linkage disequilibrium (r^2) calculated for 19 *LPA* SNPs in non-Hispanic whites (A), non-Hispanic blacks (B), and Mexican Americans (C) in NHANES III.

It is interesting to note that while we did replicate the association between Lp(a) levels and rs6919346, we did not necessarily replicate the associations reported recently for rs10945682 and rs7765803 (Lanktree et al., 2010). *LPA* rs10945682 was not associated with Lp(a) levels in NHANES III at the significant threshold of $p<0.0001$ (Table 2.5). Furthermore, while the direction of effect in non-Hispanic whites was consistent with that observed for Europeans and Asians studied by Lanktree et al (taking into account the coded allele), the direction of effect was opposite in the non-Hispanic black and Mexican American subpopulations in NHANES III. *LPA* rs7765803 was not associated with Lp(a) levels in non-Hispanic whites ($p=0.2471$) while it was strongly associated in European and Asian populations in Lanktree et al. Finally, the data reported here are not consistent with the linkage disequilibrium data reported by Lanktree et al. *LPA* rs10945682 and rs6919346 are reported to be in the same linkage disequilibrium block (Lanktree et al., 2010) but the LD calculated in our non-Hispanic white samples and in HapMap CEU suggests there is little LD ($r^2=0.06$ and 0.03 in non-Hispanic whites and CEU, respectively) between the two SNPs. It is possible that this discrepancy can be explained by unidentified population substructure or by the use of different LD measures, but this is unclear from the literature and requires further investigation.

As alluded to above, the relationship between *LPA* tagSNPs and Lp(a) levels may represent a direct (i.e. causal) or indirect (i.e. proxy for true causal variant) relationship. The latter situation most likely applies to the majority of SNPs genotyped in this study. Of the 19 *LPA* SNPs, 17 are located in introns, and the two nonsynonymous SNPs (rs7765803 and rs41265936, Table 2.1) are not predicted to alter protein function using SIFT (Kumar et al., 2009). Additional studies are needed to determine if these variants regulate *LPA* expression *in vivo*. However, since apo(a) is present only in humans, Old

World primates, and the hedgehog, resources for these studies are limited to transgenic mice and rabbits as models (Boffa et al., 2004).

In an attempt to evaluate the joint effect of significantly associated variants, a genetic risk score (GRS) was calculated. Based on this GRS, these variants together explained 7%, 9%, and 11% of the variance in Lp(a) levels in non-Hispanic whites, non-Hispanic blacks, and Mexican Americans, respectively. In comparison to the effect attributed to the kringle repeat region based on previous studies (Ali et al., 1998; Boerwinkle et al., 1992; Boomsma et al., 2000; Chiu et al., 2000; Crawford et al., 2008; Kraft et al., 1996; Rainwater et al., 1997; Schmidt et al., 2006), the effect of these SNPs is considerably small.

This study has several strengths and limitations. The greatest strength is the use of a large and diverse population. While there have been several studies of *LPA* SNPs and its association with Lp(a) that have included both European and African descent populations, no single study, to our knowledge, has also included Mexican Americans genotyped for the same *LPA* SNPs. This latter point cannot be under emphasized as the Hispanic or Latino population is the fastest growing minority population in the United States yet remains relatively underrepresented in genetic association studies (Choudhry et al., 2007).

A limitation is that the method of measuring serum Lp(a) levels in NHANES III does not account for apo(a) isoform size. While accurate measurement of apo(a) isoform is ideal, the reliability of the Lp(a) measurement used here has been adequately demonstrated (Center for Disease Control and Prevention, 1996). Furthermore, there is no generally accepted laboratory procedure or national standardization program for

Lp(a) measurement, which may help to explain the lack of generalizabilty across studies (Marcovina et al., 2003).

A second major limitation is that NHANES III does not have data on kringle repeat size for each participant. Several methods are used to measure kringle repeat size such as Southern blot (Crawford et al., 2008) and quantitative PCR (Lanktree et al., 2009), neither of which can be used in NHANES III DNA samples given investigators are aliquoted limited amounts of DNA from crude cell lysates. Without these data, it is unclear if the associations between *LPA* SNPs and Lp(a) levels reported here are independent of the KIV-2 copy number variant, which has a well-established, large effect on Lp(a) levels.

The amount of linkage disequilibrium, or lack thereof, between the KIV-2 region and other *LPA* variants is a controversial issue. Previous studies have reported strong LD between the KIV-2 alleles and SNPs in or around *LPA* (Clarke et al., 2009; Kraft et al., 1995; Luke et al., 2007; Ogorelkova et al., 2001). In contrast, additional studies indicate the lack of strong LD (Crawford et al., 2008; Ober et al., 2009). More specifically, the tagSNPs genotyped in this study had been selected from a previous study (Crawford et al., 2008) that provided data on kringle IV-2 repeat size, and no strong LD (r^2>0.80) was found for any of the SNPs tested (Crawford et al., 2008). However, there was moderate LD (r^2=0.45 in European American and r^2=0.57 in African American samples) between kringle repeat sizes 10 and 14 and *LPA* SNPs 74970 and rs41271028, respectively (Crawford et al., 2008). *LPA* 74970 was not genotyped here. *LPA* rs41271028 was genotyped here but was not significantly associated with Lp(a) levels in any of the three subpopulations after correction for multiple testing (Table 2.5). Thus, the tagSNPs genotyped here and significantly associated with Lp(a) levels after correction for

multiple testing are not in high or moderate LD with specific kringle repeat sizes examined in the original dataset reported by Crawford et al. Further studies are needed in NHANES and other large datasets to characterize the full spectrum of *LPA* genetic variation and its impact on Lp(a) levels in diverse populations.

Another limitation of this study is that only approximately 30-35% of the LDSelect "bins" for European Americans and African Americans are represented by tagSNPs as many *LPA* SNPs failed assay design or genotyping in NHANES III. And, tagSNPs selection was limited to common variation, leaving rarer variation such as *LPA* rs10455872 (<5% MAF) untested. Thus, much of the genetic variation in *LPA* and its association with Lp(a) levels in these populations remains to be explored. Furthermore, tagSNPs were not selected specifically for the Mexican American subpopulation. At the time of tagSNPs selection, HapMap 3 Mexican American samples were not available, and it was unclear which populations should be used for tagSNPs selection to adequately represent this admixed population. It is important to note that, however, that while our tagSNPs selection process may have been biased for populations of European- and African-descent, the allele frequencies observed in NHANES III Mexican Americans were very similar to that of non-Hispanic whites. Furthermore, our lack of Mexican American specific tagSNPs does not undermine the observation that there is an excess of significant variants associated only in non-Hispanic blacks compared to non-Hispanic whites.

Because of these strengths, and despite these limitations, we have taken an important step in understanding how *LPA* genetic variants contribute to Lp(a) levels in a diverse population. One of the major findings of our study was that there were notably more significant associations between Lp(a) and *LPA* SNPs in non-Hispanic blacks

compared to non-Hispanic whites and Mexican Americans. Moreover, nearly half of these associations were exclusive to non-Hispanic blacks. Our results suggest that between-population differences in Lp(a) levels can be explained, in part, by multiple population-specific *cis*-acting variants in *LPA*. While the role of multiple *trans*-acting factors in Lp(a) trait distribution has been disputed (Barkley et al., 2003; Mooser et al., 1997; Scholz et al., 1999) and cannot be ruled out, our results reaffirm the need for more comprehensive studies of the effects of *LPA* variants in large, diverse populations.

Acknowledgements

We would like to thank Dr. Geraldine McQuillan and Jody McLean for their help in accessing the Genetic NHANES III data. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for this work.

CHAPTER III

USE OF A GWAS OF CHILDREN TO IDENTIFY NOVEL LIPID-ASSOCIATED VARIANTS[2]

Introduction

Genome-wide association studies (GWAS) have identified many common genetic variants that contribute to normal variation in lipid traits. The largest GWAS meta-analysis to date, containing greater than 100,000 individuals of European descent, identified 95 loci that were independently associated with total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C) and triglycerides (TG) (Teslovich et al., 2010). Combined, these loci explained only ~25-30% of the genetic variance, leaving a majority of the genetic variance unexplained.

Some have argued that the "missing heritability" observed for most complex human traits is due, in part, to unidentified genetic and non-genetic modifiers (McCarthy and Hirschhorn, 2008). Very few GWAS have tested for and identified genetic modifiers as the statistical methods and approaches have yet to be developed that will fully exploit these data (Cordell, 2009; Thomas, 2010). Also, few GWAS have collected the data necessary to test for modifiers, particularly environmental exposures that modify genetic associations.

---

[2] Adapted from: Dumitrescu L, Brown-Gentry K, Glenn K, Young W, Kornegay N, Cai J, Relling MV, Crawford DC. Evidence for age as a modifier of genetic associations for lipid levels. *Annals of Human Genetics*. In Review.

Age, a potential modifier of the lipid trait distribution due to genetics, has been relatively unexplored. It is known that children and adolescents have different lipid distributions compared with adults (Hickman et al., 1998; Jolliffe and Janssen, 2006). Indeed, the National Cholesterol and Education Program (NCEP) provides a set of lipid and lipoprotein guidelines specific for children and adolescents (Expert Panel Blood Cholesterol Levels Child Adolesc, 1992). Furthermore, the increase in the means and variances of most lipid parameters as humans age has been well established (Snieder et al., 1997; Boomsma et al., 1996; Ericsson et al., 1991; Heller et al., 1993; Reilly et al., 1990). How much of that age dependency is due to increases in environmental or genetic variance, or both, is still left to be determined. Preliminary studies of complex traits such as body mass index (BMI) and systolic blood pressure suggest that over time, the contribution of genes to a phenotype remains relatively constant, but the relative contribution of genes decreases (Brown et al., 2003). Thus, the accumulation effects of environmental exposures, such as diet, exercise, smoking, are thought to be the factors that increase the phenotypic variance of complex traits such as BMI and lipids over time.

All published GWAS studies of the lipid traits (HDL-C, LDL-C, and TG) have been performed in adults (>18 years of age) (Hindorff et al., 2010), a population exposed to environmental factors that influence lipid trait distributions for at least two decades. While these GWAS have been successful in identifying genetic variants associated with lipid traits, we propose that the study of younger participants such as children will identify additional variants masked by the age dependency for these traits. To identify these novel variants associated with HDL-C, LDL-C, and TG levels, we first performed a GWAS in 411 children of European, African, or Mexican-descent ascertained at St. Jude Children's Research Hospital followed by replication in an additional dataset of youths

from the Third National Health and Nutrition Examination Survey (n=1,040; NHANES III). Replicated genetic variants were formally tested for an interaction with age in the larger NHANES III dataset with adults (n=3,508). Gene discovery in children followed by testing for interactions in adults has identified one lipid trait-associated locus potentially modified by age and represents a modest step towards identifying the full genetic architecture of the lipid trait distributions in human populations.

Methods

Subjects in the discovery GWAS were drawn from the Total Therapy Study XV, a prospective study of children undergoing treatment for acute lymphoblastic leukemia (ALL), initiated in 2000 at St. Jude Children's Research Hospital (Pui et al., 2009). From June 2000 to October 2007, a total of 501 newly diagnosed patients aged 1 to 18 years were enrolled (after informed consent and assent, as appropriate) in the study, 411 of whom were evaluable for both serum lipids and genome-wide genotyping. Race/ethnicity was inferred using germline genotypes as previously described (Yang et al., 2009) based on genotype-based hierarchical clustering of patients and using data from the HapMap cell lines (n=90 CEU (Caucasian), 90 YRI (African), 30 CHB and 30 JPT (Asians) as references. Patients exhibiting >90% European, African, or Asian ancestries were classified as white, black, or Asian respectively; "Hispanic" status was inferred for those patients who were self-declared Hispanics and had less than 90% of European, African and Asian ancestries. The remaining patients were labeled as "other" and were not included in this analysis. Serum levels of high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides (TG) were measured directly using standard

enzymatic techniques (Roche Diagnostics) as described in Kawedia et al (Kawedia et al., 2010). All lipid measurements were taken on consolidation day 15 of treatment (at least 4 weeks from the last dose of glucocorticoid or asparaginase) and were non-fasting.

Participants in the St. Jude study were genotyped with either the Affymetrix GeneChip 500K and 100K or 6.0 array. Only common SNPs genotyped across both platforms were considered for analysis. SNPs were excluded from the analysis based on genotyping call rates per SNP (≤ 98%) and minor allele frequency (≤ 0.01). From a total of 532,546 SNPs genotyped, 420,005 (79%) passed quality control thresholds. Tests of association between each genetic variant assuming an additive genetic model and natural-log transformed lipid trait were implemented in PLINK (Purcell et al., 2007) using linear regression, stratified by race/ethnicity.

Ascertainment of the Third National Health and Nutrition Examination Survey (NHANES III) and method of DNA collection have been previously described (Centers for Disease Control and Prevention, 1994). Briefly, NHANES III is a cross-sectional survey that was conducted from 1988-1994 by the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). Like all the NHANES, NHANES III is a complex survey design that over-sampled minorities (non-Hispanic blacks and Mexican Americans), the young, and the elderly. Blood samples were obtained at a central location known as the Mobile Examination Center (MEC). Serum total cholesterol, triglycerides, and HDL cholesterol were measured using standard enzymatic methods and LDL cholesterol was calculated using the Friedewald equation, with missing values assigned for samples with triglyceride levels greater than 400 mg/dl (Center for Disease Control and Prevention, 1996). Beginning with phase 2 of NHANES III, DNA samples were collected from study participants aged 12 years and

older. Race/ethnicity was self-identified as non-Hispanic white, non-Hispanic black, Mexican American, or other. All NHANES III procedures were approved by the CDC Ethics Review Board and written informed consent was obtained from all participants. The present study was approved by the CDC Ethics Review Board. Because the study investigators did not have access to personal identifiers, this study was considered non-human subjects research by the Vanderbilt University Institutional Review Board.

65 SNPs were carried forward for replication in NHANES III and were targeted for genotyping using Sequenom's iPLEX Gold assay on the MassARRAY platform (San Diego, CA) according to manufacturer's instructions (www.sequenom.com). One SNP, rs4811011, failed assay design and was subsequently genotyped using Applied Biosystem's TaqMan (Foster City, CA). 57 SNPs were genotyped successfully (genotyping call rate >95%); however, five SNPs failed blinded duplicates quality control measures as required by CDC and were excluded from further analyses. The remaining SNPs were tested for deviations from Hardy Weinberg Equilibrium (HWE) within each racial/ethnic group and all had HWE p-values > 0.001 in at least two populations, as required by the CDC. All genotype data reported here were deposited into the NHANES III Genetic database and are available for secondary analysis through CDC.

Unweighted genotype-phenotype tests of associations for the 52 SNPs and the three natural-log transformed lipid traits were performed in SAS version 9.2 (SAS Institute, Cary, NC) using linear regression, assuming an additive genetic model. Single-SNP analyses were stratified by race/ethnicity, adjusted for age and sex and, for the adult cohort (>18 years), were limited to fasting individuals not on lipid-lowering medications. Tests for interactions between age and genotype were also considered in

these regression models, adjusted for age, sex, and SNP genotype. Replication and interaction associations were deemed statistically significant at $p<0.05$. Data were accessed remotely from the CDC's Research Data Center (RDC) in Hyattsville, Maryland using Analytic Data Research by Email (ANDRE).

Differences in lipid levels between the St. Jude and NHANES III cohorts and between genotype groups were tested in STATA 10 using a standard two-sample t-test with unequal variances. Manhattan plots were produced using code provided by http://gettinggeneticsdone.blogspot.com.

Results

Study population characteristics

Across all three racial/ethnic subpopulations, St. Jude children with ALL had consistently lower mean LDL-C, and higher mean HDL-C and mean triglycerides compared to NHANES III youths (Table 3.1; Appendix C). The means of some lipid levels (HDL-C in Whites and Hispanics, LDL-C in Whites and Hispanics) were significantly different between the two studies of children, with differences in means of up to 18.8 mg/dL (Table 3.1; Appendix C). These differences in lipid levels are most likely related to the older mean age and/or the larger percentage of females in NHANES III youths compared with St. Jude children rather than health status, as none of the medications used during the consolidation stage of therapy for ALL are known to affect lipid levels, *per se*.

Table 3.1. Participant Characteristics.

Characteristic	GWAS cohort St. Jude Children			Replication cohort 1 NHANES III Youths			Replication cohort 2 NHANES III Adults		
racial/ethnic group	**Whites**	**Blacks**	**Hispanics**	**Whites**	**Blacks**	**Hispanics**	**Whites**	**Blacks**	**Hispanics**
Total n	282	66	63	241	439	360	1,386	1,039	1,083
Age (years)	7.0 ± 5	7.2 ± 4	6.4 ± 4	14.6 ± 2	14.8 ± 2	14.8 ± 2	52.2 ± 21	39.8 ± 16	40.0 ± 17
% Female	41.1	43.9	47.6	54.4	54.0	52.2	60.6	58.7	48.9
HDL-C (mg/dL)	55.1 ± 30	59.0 ± 27	67.9 ± 50	47.2 ± 10	52.6 ± 12	49.1 ± 11	50.3 ± 15	53.6 ± 17	47.4 ± 13
LDL-C (mg/dL)	84.5 ± 27	81.3 ± 37	87.4 ± 35	92.9 ± 28	98.5 ± 27	94.4 ± 24	129.4 ± 38	123.0 ± 40	120.1 ± 34
Triglycerides (mg/dL)	108.2 ± 79	88.3 ± 47	121.7 ± 101	97.6 ± 56	80.0 ± 43	99.0 ± 57	145.3 ± 111	107.6 ± 82	151.8 ± 110

Values are listed as mean ± sd unless otherwise indicated

When comparing the two NHANES III replication datasets, not surprisingly (Expert Panel Blood Cholesterol Levels Child Adolesc, 1992; Hickman et al., 1998; Jolliffe and Janssen, 2006; Snieder et al., 1999), adults had different lipid profiles compared to youths with statistically higher mean LDL-C and TG levels across all subpopulations (Table 3.1 and Appendix D). Mean HDL-C levels were significantly higher in White adults and significantly lower in Hispanic adults compared to youths, although the differences were small (2.3 and 1.7 mg/dL, respectively).

Discovery GWAS results

To identify novel common variants associated with HDL-C, LDL-C, and TG levels in children, we performed a genome-wide association screen in all three subpopulations of the St. Jude cohort. No SNP surpassed genome-wide statistical significance ($p<5x10^{-8}$) for any of the analyses; however, four SNPs approached that threshold and were significantly associated at $p<1x10^{-6}$ (Figure 3.1 and Appendix E). In Whites, two SNPs (intergenic rs4742455 and *USH2A* rs17026635) were associated with transformed HDL-C ($p=5.18x10^{-7}$ and $p=6.33x10^{-7}$, respectively). In Hispanics, intergenic rs7790255 and *GBAS/MRPS17* rs15892 were associated with transformed TG ($p=2.38x10^{-7}$ and $p=9.42x10^{-7}$, respectively). The latter two SNPs are in very high LD ($r^2=0.91$) in the Hispanic subpopulation and therefore likely represent the same association. Across all three lipid traits and all three racial/ethnic groups, 65 associations were significantly associated at $p<1x10^{-5}$ and were carried forward for replication in NHANES III youths.

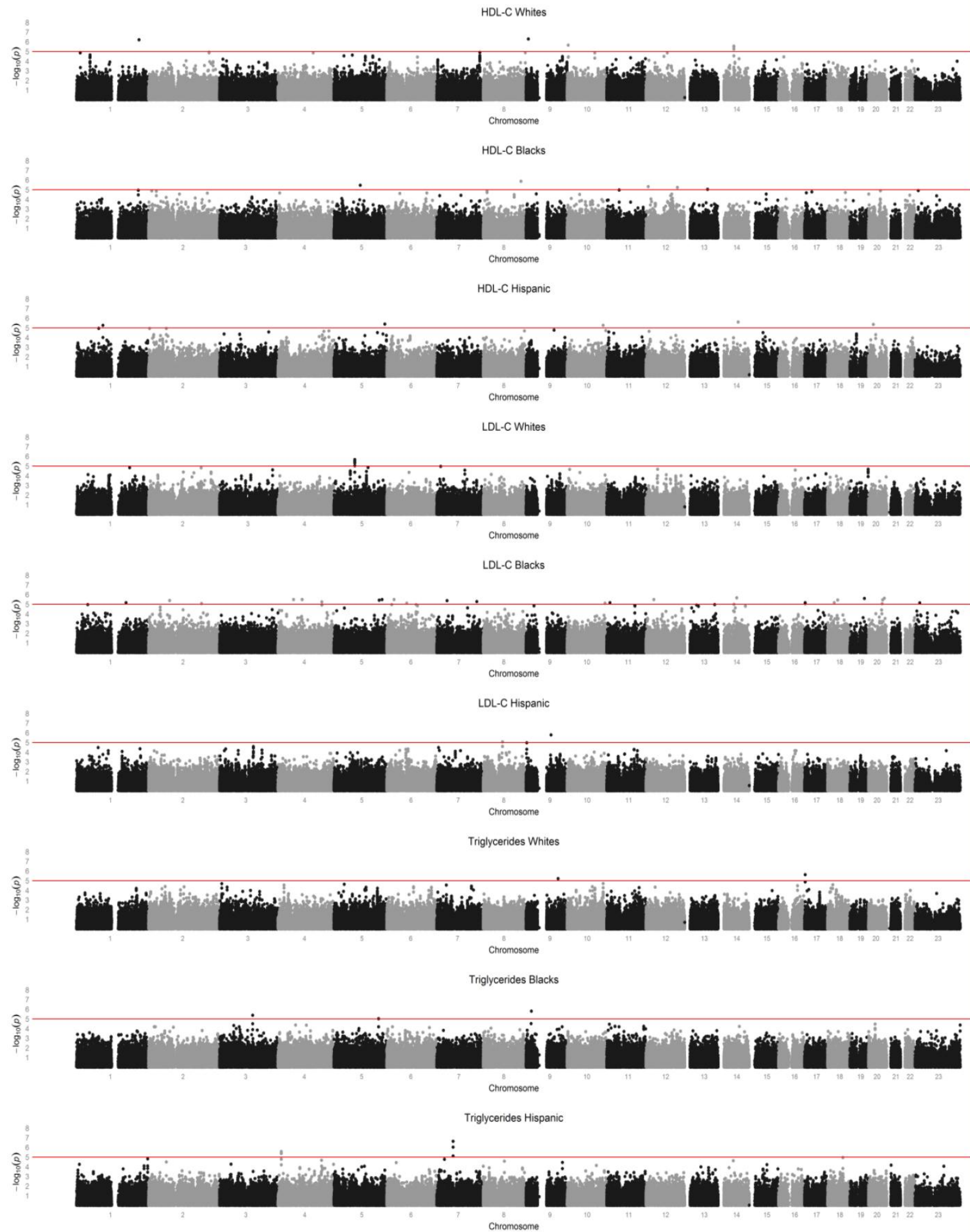


Figure 3.1. Discovery GWAS association results. The figure summarizes genome-wide association results for three lipid traits, HDL-C, LDL-C, and triglycerides in the three St. Jude cohort populations, Whites, Blacks, and Hispanics. Values are plotted as $-\log_{10}$P value on the y-axis. Chromosomal location is designated on the x-axis. The red line corresponds to a $p=1x10^{-5}$, the threshold used for determining which associations would be carried further for replication.

Replication in NHANES III youths and adults

Next we tested if the novel associations observed in the discovery GWAS would replicate in an independent study of children. For the replication study, a total of 1,040 youths (12-18 years of age) in NHANES III were genotyped for 65 SNPs, 52 of which passed assay design and quality control measures (see Methods). Three of the 52 associations replicated at $p<0.05$ in NHANES youths (Table 3.2). Two SNPs, *SGSM2* rs2429917 and intergenic rs12190789, were associated with transformed LDL-C at $p=0.009$ and $p=0.047$, respectively, and *CD96* rs16858329 was associated with transformed triglycerides at $p=0.048$. All three of the significant associations were in Blacks and the directions of effects were concordant with those of the discovery cohort, although they were of smaller magnitude.

We also conducted a second replication study in a large cohort of fasting adults (>18 years of age) in NHANES III to determine if associations initially discovered in children would generalize to adults. Despite the considerable increase in sample size, none of the three associations that replicated in NHANES youths were associated in NHANES adults (Table 3.2). There were, however, three significant associations in NHANES adults that were not replicated in NHANES youths (TG and intergenic rs6477578 in Whites, LDL and *FRMD3* rs10868008 in Hispanics, and LDL and *FRMD3* rs1140077 in Hispanics; Table 3.2). The former association was the most significant at $p=0.006$, yet the direction of effect was in the opposing direction compared with the GWAS discovery study of St. Jude children. The latter two associations most likely represent the same association as rs10868008 and rs114007 are in complete LD ($r^2=1.00$) in our discovery GWAS among Hispanic samples.

Table 3.2. Significant Replication Results in NHANES III

Trait	Population	SNP	Gene/ location	Coded Allele (Freq)	GWAS cohort St. Jude Children		Replication cohort 1 NHANES III Youths		Replication cohort 2 NHANES III Adults	
					Beta (SE)	p-value	Beta (SE)	p-value	Beta (SE)	p-value
LDL-C	Blacks	rs2429917	*SGSM2* intron	T (0.04)	-1.45 (0.30)	7.01E-06	-0.20 (0.08)	***0.009***	0.02 (0.05)	0.662
LDL-C	Blacks	rs12190789	intergenic	T (0.03)	-1.46 (0.29)	3.32E-06	-0.17 (0.08)	***0.047***	-0.04 (0.05)	0.413
TG	Blacks	rs16858329	*CD96* intron	T (0.41)	-0.36 (0.07)	4.29E-06	-0.06 (0.03)	***0.048***	0.03 (0.02)	0.127
TG	Whites	rs6477578	intergenic	G (0.27)	0.25 (0.05)	6.13E-06	0.03 (0.05)	0.425	-0.06 (0.02)	***0.006***
LDL	Hispanics	rs10868008	*FRMD3* intron	C (0.04)	-0.69 (0.13)	1.66E-06	-0.06 (0.06)	0.368	-0.08 (0.05)	***0.049***
LDL	Hispanics	rs11140077	*FRMD3* intron	A (0.06)	-0.69 (0.13)	1.66E-06	-0.06 (0.06)	0.368	-0.08 (0.05)	***0.049***

Significant ($p \leq 0.05$) associations in the Replication cohorts are bolded and italicized. Coded allele frequencies are based on the whole genetic NHANES III in the designated racial/ethnic population

Age as a potential modifier

As mentioned previously and shown in Table 3.1 and Appendix D, mean HDL-C, LDL-C, and TG levels differed significantly between adults and children/youths. Given that three associations were discovered and replicated in children but failed to generalize to adults, we explored the hypothesis that genetic associations observed for the lipid traits are modified by age. After testing for SNP-age interactions, we observed a significant interaction (p=0.024) between age and *SGSM2* rs2429917 with transformed LDL-C in Blacks.

Up until about the fourth decade of life, participants homozygous for the major allele of rs2429917 (C, frequency=0.96) had consistently lower LDL-C levels compared to CT heterozygotes (Table 3.3) with the largest significant difference in the 12 to 21 age group (mean difference = 17.25 mg/dl, p=0.002). Any trend in mean differences is harder to detect later in life due to smaller sample sizes for participants with the CT genotype (n ranges from 0 to 7; Table 3.3). However, participants with the CC genotype had significantly lower mean LDL-C (mean difference = 24.14 mg/dl, p=0.016) in the older 72 to 81age group, which is opposite that observed in the 12 to 21 year olds. These results suggest that early in life (and possibly much later in life) LDL-C concentration is dependent, in some small part, upon rs2429917 genotype.

Table 3.3. Mean LDL-C levels in Blacks, stratified by rs2429917 genotype and age group. Age was categorized into age groups spanning 10 years, beginning at age 12 (the age at which NHANES III began collecting genetic data on participants). For each age group, mean and standard deviations of LDL-C concentrations were calculated for participants with a CC genotype or a CT genotype at the rs2429917 locus, separately. Data for the TT genotype is not presented here due to small sample size (n=3). T-tests with unequal variances were calculated to test for differences in mean LDL-C between genotypes, within the same age group.

age group (years)	rs2429917 Genotype				p-value
	CC n=734		CT n=52		
	N	LDL-C mg/dl (mean ± sd)	N	LDL-C mg/dl (mean ± sd)	
12 to <21	157	100.25 ± 26.8	16	82.75 ± 19.0	***0.003***
≤21 to <30	126	113.79 ± 37.8	8	103.25 ± 32.1	0.398
≤30 to <39	169	115.80 ± 35.3	12	111.41 ± 24.7	0.574
≤39 to <48	125	124.80 ± 40.4	7	139.71 ± 52.5	0.486
≤48 to <57	42	132.12 ± 42.7	2	152.50 ± 37.5	0.580
≤57 to <63	40	151.00 ± 47.1	3	120.67 ± 63.5	0.497
≤63 to <72	43	138.55 ± 44.1	2	153.50 ± 55.9	0.770
≤72 to <81	22	145.86 ± 42.2	2	170.00 ±2.8	***0.016***
≥81	10	126.20 ± 41.6	0	--	--

Discussion

The aim of this study was to discover novel variants associated with lipid levels in children and to test if those associations were also significant in adults. We performed a GWAS of children undergoing treatment for ALL at St. Jude Children's Research Hospital, and attempted replication in an independent population of youths and adults from NHANES III. Three of the 52 lipid-genotype associations tested in NHANES III children replicated at $p<0.05$, including intronic *SGSM2* rs2429917 at $p=0.009$. However, these associations did not generalize to NHANES III adults. We also identified a genotype x age interaction with *SGSM2* rs2429917 for transformed LDL-C in Blacks, supporting a genetic basis for the differences observed in lipid levels in children compared to older individuals.

Age as a modifier of genetic association studies has only recently been highlighted in the literature. In one study of the 100K data of the longitudinal Framingham Heart Study, Lasky-Su et al described an age-dependent association between *ROBO1* and obesity where the association was stronger among the pediatric cohorts compared with adult cohorts (Lasky-Su et al., 2008). Although Lasky-Su et al did not speculate on the mechanism behind the age-dependent interaction, it is interesting to note that heritability estimates for obesity in children tend to be higher (Haworth et al., 2008) compared with estimates in adults (Brown et al., 2003). Somewhat consistent with these observations in obesity are the observations of heritability estimates for the lipid traits. That is, some studies have found that heritability of select lipid levels tended to decrease with age (Beekman et al., 2002; Heller et al., 1993).

However, the review by Snieder et al (Snieder et al., 1999) concluded that heritability estimates for HDL-C, LDL-C, and triglycerides remained relatively stable with age.

While the magnitude of the genetic influence on lipid metabolism may not change with age, the importance of different genes may. Different genes may be expressed in childhood and adolescence compared to adulthood. In regards to lipid metabolism, longitudinal twin studies support this possibility (Friedlander et al., 1997; Nance et al., 1998; Williams and Wijesiri, 1993), and an extended parent-twin study determined that different genes are expressed in adolescence compared to adulthood (Snieder et al., 1997). It is also possible that the same genes function throughout life, but are expressed at different levels depending on the decade of life. Supporting this latter hypothesis is the observation that younger patients heterozygous for *ABCA1* mutations that cause Tangier disease have significantly higher HDL-C levels than older patients heterozygous for *ABCA1* mutations (Clee et al., 2000). There is evidence that normal *ABCA1* function increases over time (Clee et al., 2000), which suggests pronounced HDL-C deficiency between age groups may be highlighting the heterozygous carriers' inability to do so.

Given the proposed and observed differences between children and adults for these traits, we purposefully performed a discovery study in children as this subset may allow for discovery of novel genes associated with lipid levels compared with previously published GWAS from adults. For our work presented here, the most promising novel candidate as a result of this study is rs2429917, located in the intron of *SGSM2*, or small G protein signaling modulator 2. *SGSM2* is ubiquitously expressed in various tissues, including the liver, and as the name implies, acts as a modulator of G-protein signaling through its interaction with a subfamily of RAS proteins (Yang et al.,

2007). Proteins involved in G protein-mediated signal transduction are associated with a number of cellular mechanisms, including differentiation and proliferation. It is also important to note that rs2429917 is located in a fairly gene-dense region of chromosome 17, including *SMG6*, *SRR*, *TSR1*, *MNT*, and *METT10D*, all within ~100 kb flanking *SGSM2*. Based on their biological functions, none of these neighboring genes are compelling candidates for association with lipid metabolism. However, *SMG6* is an intriguing candidate given its essential association with telomerase activity (Reichenbach et al., 2003; Snow et al., 2003) and, thereby, aging. Deletion of Est1p (the *S. cervisiae* homolog to human *SMG6*) in yeast leads to ever-shorter telomeres over time despite functional telomerase activity (Lundblad and Szostak, 1989). Telomere shortening occurs in all mitotic tissues (excluding germline tissue) as humans age and has been shown to contribute to mortality in many age-related diseases, including heart disease (Cawthon et al., 2003). Although these findings require further study, it is interesting to speculate that these data may point to the involvement of previously unsuspected pathways contributing to lipid metabolism.

This study had several limitations, including the fact the children genotyped for the discovery GWAS were currently undergoing treatment for ALL. Lipid levels were collected during a particular phase of therapy (consolidation) which lasts for approximately eight weeks and consists of doses of methotrexate, hydrocortisone, and cytarabine every other week, and daily doses of mercaptopurine. To our knowledge, changes in lipid or lipoprotein concentrations have not been reported during this treatment period. While changes in lipid profiles have been reported in children during combination therapy with L-asparaginase (Halton et al., 1998; Parsons et al., 1997) and

with glucorticoids (Boers et al., 2003), lipid measurements were taken at least four weeks from the last dose of these drugs and, therefore, should reflect baseline measurements.

A second limitiation is that the discovery GWAS was underpowered due to its small sample size. Even with our largest population (n=282 in Whites), and an allele frequency of 5%, we had 80% power to detect only large effect sizes (R^2>11%) at genome-wide significance. The majority of published lipid GWAS-identified variants have small effect sizes and explain only a small percent of the variance of lipid traits in the population (Teslovich et al., 2010; Manolio, 2009). However, to our knowledge, no GWAS has been performed on children with lipid levels; therefore, it is unknown whether the effect size and/or the significance of these well-known variants remain constant over a lifetime.

Despite the small discovery sample size, we were able to detect nominally significant associations, of which three replicated at $p<0.05$ in an independent dataset. Furthermore, examination of genetic variation known to influence lipids in European-descent populations demonstrates that true associations can be detected in spite of the low power of the study. That is, of the 26 established lipid-associated SNPs in 23 genes (including *CETP, LPL, GCKR, APOB,* etc; Appendix F), we detected seven associations with p-values ≤ 0.05. As an example, rs328 is a non-synonymous SNP in *LPL* and has been shown to have a reproducible effect (~19 mg/dL in one study) (Kathiresan et al., 2008) on lowering triglycerides. In our GWAS of children, rs1741102, a proxy for rs328 (r^2=1 in HapMap CEU), was associated at p=0.02 with β=-0.17, corresponding to -21.3 mg/dL, which is consistent in both the previously reported magnitude and direction of effect.

A benefit of using NHANES III data is that it allows for genetic studies in a large, diverse population with a wide age-range. However, it is a cross-sectional study. Since our data suggest that there may be age-specific genetic influences, longitudinal data (with lipid concentrations measured at numerous ages for the same participant) are necessary to derive further conclusions and to replicate this interaction.

Post-mortem studies on young adults and children have shown that atherosclerosis starts early in life (Expert Panel Blood Cholesterol Levels Child Adolesc, 1992), even though clinical symptoms usually do not manifest until decades later. The potential temporal nature of factors, both genetic and non-genetic, that contribute to cardiovascular disease is important for better understanding of the etiology of the disease. While it is often assumed in genotype-phenotype association studies that genetic effects are stable over a lifetime, the possibility of important age-effects should not be ignored when studying the genetics of lipid metabolism.

Acknowledgements

Genotyping for replication in NHANES III was supported in part by the Vanderbilt Institute for Clinical and Translational Research (VICTR), as well as a Clinical and Translational Science Award (CTSA) grant (1UL1 RR024975-01 from the National Center for Research Resources and the National Institutes of Health [NIRR/NIH]). We would like to thank Dr. Geraldine McQuillan and Jody McLean for their help in accessing the Genetic NHANES data and Dr. Jean Cai at St. Jude for serum lipids measurements. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for

this work. The NHANES DNA samples are stored and plated by the Vanderbilt DNA Resources Core. Genotyping was performed by Ping Mayo and Dr. Nathalie Schnetz-Boutaud in the laboratory of Dr. Jonathan Haines and Ping Mayo under the direction of Dr. Cara Sutcliffe in the Vanderbilt DNA Resources Core. Supported by NCI grant CA 21765, the NIH/NIGMS Pharmacogenomics Research Network (U01 GM92666), and by the American Lebanese Syrian Associated Charities (ALSAC).

CHAPTER IV

REPLICATION AND GENERALIZATION OF LIPID GWAS-IDENTIFIED VARIANTS FROM THE LITERATURE[3]

Introduction

Since its introduction in 2005, the genome-wide association study (GWAS) design has become a powerful tool in human genetics to identify single nucleotide polymorphisms (SNPs) associated with common diseases or traits using an experimental design that does not require *a priori* biological knowledge. As of September 2010, greater than 1,000 SNPs across the genome have been reported as genome-wide significant ($p \leq 5x10^{-8}$) for 165 traits (Hindorff et al., 2010). An early analysis of the GWAS-reported SNPs demonstrated that most GWAS-identified variants were intergenic or intronic (Hindorff et al., 2009), suggesting either novel biology or that the functional variant has yet to be found.

While GWA studies have been successful in identifying novel associations, there are several limitations. First, the majority of GWAS have been conducted in populations of European-descent. While there are several GWAS in populations of Asian-descent, GWAS are just emerging for other populations such as African Americans (Genovese et

[3]Adapted from: Dumitrescu L, Carty CL, Taylor K, Schumacher FR, Hindorff LA, Ambite JL, Anderson G, Best LG, Brown-Gentry K, Bůžková P, Carlson CS, Cochran B, Cole SA, Devereux RB, Duggan D, Eaton CB, Fornage M, Franceschini N, Haessler J, Howard BV, Johnson KC, Laston S, Kolonel LN, Lee ET, MacCluer JW, Manolio TA, Pendergrass SA, Quibrera M, Shohet RV, Wilkens LR, Haiman CA, Le Marchand L, Buyske S, Kooperberg C, North KE, and Crawford DC. Genetic determinants of lipid traits in diverse populations from the Population Architecture using Genomics and Epidemiology (PAGE) Study. *PLoS Genetics*. In Review.

al., 2010; Hallmayer et al., 2009; Himes et al., 2009; Smith et al., 2009; Shi et al., 2009; Adeyemo et al., 2009; Ge et al., 2009; Sebastiani et al., 2010; Mathias et al., 2010; Edenberg et al., 2010; Bierut et al., 2010; Adkins et al., 2010; Pelak et al., 2010; Kang et al., 2010; Sleiman et al., 2010; Nielsen et al., 2010; Bostrom et al., 2010; Kariuki et al., 2010), Mexican Americans/Hispanics (Kariuki et al., 2010; Hayes et al., 2007; Norris et al., 2009; Kanetsky et al., 2009; Ge et al., 2009; Hancock et al., 2009; Palmer et al., 2010; Bozaoglu et al., 2010), and American Indians (Hodgkinson et al., 2010). It is possible that novel associations await discovery in these populations given the differing linkage disequilibrium (LD) patterns when compared with populations of European-descent (Rosenberg et al., 2010). Second, much work is needed to test SNPs discovered in case-control studies in more population-based, representative cohorts to determine if the associations generalize. Data on generalization will inform future fine-mapping (Teo et al., 2010) and discovery studies as well as provide clues to whether GWAS-identified SNPs are simply tagSNPs or are more likely to be true functional SNP(s).

A major goal of the Population Architecture using Genomics and Epidemiology (PAGE) study is to determine whether GWAS-identified variants generalize to diverse groups drawn from population-based studies (Matise et al., 2010). Generalization is defined here as a significant association ($p<0.05$, uncorrected for multiple testing) in a non-European population and a direction of genetic effect in the same direction as that of European Americans. In PAGE, variants identified in GWAS and well replicated in multiple studies are chosen for targeted genotyping in hundreds to thousands of European Americans (~20,000), African Americans (~9.000), American Indians (~6,000), Mexican Americans/Hispanics (~2,500), Japanese/East Asians (~690), and Native Hawaiians/Pacific Islanders (~175). All samples are linked to extensive demographic,

health, and exposure data, making the PAGE study a rich resource for post-discovery generalization and characterization for common human diseases and traits.

We present here PAGE study data on the replication and generalization for 49 SNPs associated with three common lipid traits: low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides. Each of these three traits has numerous GWAS published in European ancestry individuals (Pollin et al., 2008; Teslovich et al., 2010; Aulchenko et al., 2009; Wallace et al., 2008; Sandhu et al., 2008; Heid et al., 2008; Sabatti et al., 2009; Ridker et al., 2009; Saxena et al., 2007; Willer et al., 2008; Kooner et al., 2008; Kathiresan et al., 2008; Kathiresan et al., 2009) but only a handful published in other populations, such as Asians (Hiura et al., 2009) and Micronesians (Burkhardt et al., 2008). Additional data are just now emerging from large samples sizes of diverse populations for generalization (Keebler et al., 2009; Gupta et al., 2010; Waterworth et al., 2010; Chang et al., 2010; Nakayama et al., 2009; Deo et al., 2009; Teslovich et al., 2010) and fine-mapping (Keebler et al., 2010) of these lipid GWAS-identified SNPs. We demonstrate that the majority of the targeted GWAS-identified SNPs replicate in European Americans in PAGE and that many generalize to diverse populations. Both power and LD are explored as explanations of non-generalization, highlighting the complexities involved in properly interpreting results of even robust genetic associations such as these.

Methods

Study populations and phenotypes

All studies were approved by Institutional Review Boards at their respective sites (details are given in Appendix G). PAGE study samples were drawn from four large population-based studies or consortia: EAGLE (Epidemiologic Architecture for Genes Linked to Environment), based on three National Health and Nutrition Examination Surveys (NHANES) (Centers for Disease Control and Prevention, 2010; Centers for Disease Control and Prevention, 2004; Centers for Disease Control and Prevention (CDC), 2002), the Multiethnic Cohort (MEC) (Kolonel et al., 2004), the Women's Health Initiative (WHI) (1998; Anderson et al., 2003), and Causal Variants Across the Life Course (CALiCo), a consortium of several cohort studies: Atherosclerosis Risk in Communities Study (ARIC) (The ARIC Investigators, 1989), Coronary Artery Risk in Young Adults (CARDIA) (Friedman et al., 1988), Cardiovascular Health Study (CHS) (Fried et al., 1991), Strong Heart Family Study (SHFS) (North et al., 2003), and Strong Heart Cohort Study (SHS) (Lee et al., 1990). The PAGE study design is detailed in Matise et al (Matise et al., 2010).

Serum HDL-C, triglycerides, and total cholesterol were measured using standard enzymatic methods. LDL-C was calculated using the Friedewald equation (Friedewald et al., 1972), with missing values assigned for samples with triglyceride levels greater than 400 mg/dl. For PAGE study sites with longitudinal data, the baseline measurement was used for analysis. A full description of each study, along with population-specific study characteristics, is presented in Appendix G and Appendix H.

SNP selection and genotyping

All SNPs considered for genotyping were previously associated with HDL-C, LDL-C, and/or triglycerides in published (as of 2008) candidate gene and genome-wide association studies. A total of 52 SNPs were targeted for genotyping by two or more PAGE study sites. The 52 targeted variants are located in or nearby 32 different genes/gene regions, with 12 of the gene/gene regions represented by two or more SNPs. Six SNPs are nonsynonymous, one SNP is a nonsense variant, two SNPs are synonymous, and three SNPs are in untranslated regions; the remainder are located in introns, flanking, or intergenic regions. The full list of targeted SNPs, their locations, and their previously associated lipid trait can be found in Appendix I.

Cohorts were genotyped using either commercially available genotyping arrays (Affymetrix 6.0, Illumina 370CNV BeadChip), custom mid- and low-throughput assays (TaqMan, Sequenom, Illumina GoldenGate or BeadXpress), or a combination thereof. Quality control was implemented at each study site independently. In addition to site-specific quality control, all PAGE study sites genotyped 360 DNA samples for all SNPs from the International HapMap Project and submitted these data to the PAGE Coordinating Center for concordance statistics (Matise et al., 2010). Study specific genotyping details are described in Appendix G. Of the 52 targeted SNPs, three (*CETP* rs1800775, *APOE* rs429358, and *APOE* rs7412) failed at all PAGE study sites that attempted genotyping; therefore, a total of 49 SNPs were tested in this analysis.

Statistical methods

All tests of association were performed by each PAGE study site using the same analysis protocol prior to meta-analysis. The study protocol excluded participants <18

years of age as well as non-fasting samples (defined here as <8 hours). When triglyceride level was the dependent variable, participants with >1,000 mg/dl were excluded from analyses. Triglyceride (TG) levels were natural-log transformed (ln) prior to analysis.

Linear regression was performed for fasting adults regardless of lipid lowering medication use with HDL-C, LDL-C, or ln(TG) as the dependent variable and a SNP as the independent variable, assuming an additive genetic model, stratified by race/ethnicity. The beta estimate is per additional copy of the coded allele. For each SNP, four models were considered: 1) unadjusted, 2) adjusted for age (continuous in years) and sex, 3) adjusted for age, body mass index (continuous in kg/m^2), current smoking (yes/no; binary), type 2 diabetes (yes/no; binary), post-menopausal status (yes/no for females only; binary), and current hormone use (yes/no for females only; binary), and 4) adjusted for age, body mass index, current smoking, type 2 diabetes, post-menopausal status, current hormone use, and previous myocardial infarction (yes; no; binary). All PAGE study sites (except for WHI, which is female only) stratified models 3 and 4 by sex given the sex-specific variables (post-menopausal status and hormone use) prior to meta-analysis. Select PAGE study sites also included study site or site of ascertainment as a covariate in all models. Results from Model 2 (adjusted for age and sex) are reported in the main text while results from Models 1, 3, and 4 are presented in Appendices J-L. Model 2 excluding participants on lipid-lowering medications are presented in Appendices M-O.

Meta-analyses, using a fixed-effects inverse-variance weighted approach and tests for effect size heterogeneity across studies, were performed using METAL (Willer et al., 2010). P-values were not adjusted for multiple testing, and association results

were plotted using Synthesis-View (Pendergrass et al., 2010), where indicated. Power calculations were performed using Quanto (Gauderman and Morrison, 2006; Gauderman, 2002) assuming unrelated participants, an additive genetic model, the published effect size from European-descent populations listed in Appendix H, and the population-specific allele frequencies. Linkage disequilibrium was calculated using HapMap European (CEU) and West African (YRI) data accessed through the Genome Variation Server.

Results

Study population characteristics

The PAGE study sites are diverse across multiple variables (Table 4.1 and Appendix H). Together, the PAGE study consists of several populations: European Americans, African Americans, Mexican Americans/Hispanics, American Indians, Japanese/East Asians, and Native Hawaiians/Pacific Islanders. All PAGE study sites except WHI ascertained both men and women. Participant age varies widely across PAGE. For example, CHS ascertained on average older adults (median age = 74 and 72 years for European and African Americans, respectively), CARDIA ascertained younger adults (median age = 26 and 24.5 years for European and African Americans, respectively), and NHANES ascertained all ages of adults (18 years to 90 years; median age = 51, 39, and 40 years for European, African, and Mexican Americans, respectively). In addition to demographic differences, lifestyles and health differed across the PAGE study sites by population, including lipid lowering medication use and current smoking

Table 4.1. Characteristics of PAGE study populations

	EAGLE	MEC	WHI	CALiCo			
				ARIC	CARDIA	CHS	SHS
Type of Study	Cross-sectional	Nested Case Control	Cohort and Clinical Trials	Longitudinal	Longitudinal	Longitudinal	Longitudinal
Focus of Cohort	N/A	Cancer	Women's Health	Cardiovascular Disease	Cardiovascular Disease	Cardiovascular Disease	Cardiovascular Disease
Years Collected	1991-1994, 1999-2002	1993-1996	1993-1998	1987-2007	1986-2006	1989-1999	1988-present
Median Age	43	67	63	54	25	73	47
Age Range	18-90	48-86	50-79	44-66	18-35	64-96	14-93
% Women	54	36	100	57	56	62	59.3
Race/Ethnicity(n_{max})							
European Americans	3,909	317	4,688	11,178	2,134	2,787	--
African Americans	1,896	552	1,840	3,770	2,035	550	--
American Indians	--	--	113	--	--	--	6,021
Mexican Americans	2,361	299	762	--	--	--	--
Japanese/ East Asian	--	576	251	--	--	--	--
Native Hawaiian/ Pacific Islander	--	87	113	--	--	--	--

status. More Japanese participants ascertained by MEC reported lipid lowering medication use compared with other populations ascertained by other PAGE study sites: 38.3% versus <5-10%. American Indians from the Dakotas reported more smoking (42.2-47.8%) than other American Indians (25-33%) or other PAGE study site populations (6.3% to 35.3%). The differences in demographics, lifestyle, and health characteristics observed across the PAGE study sites and populations are reflected in the three traits studied here (Appendix H). Given the diversity observed across the PAGE study sites, we performed all tests of association for HDL-C, LDL-C, and triglycerides unadjusted, minimally adjusted (for age and sex), and adjusted for various demographic, lifestyle, and health variables.

Allele frequencies

Coded allele frequencies are presented in Tables 4.2-4.4 and in Appendix P, by population. We calculated the Pearson correlation coefficient (r) between European American coded allele frequencies and all other groups. The highest correlation was observed in the comparison with Mexican Americans/Hispanics (0.97) followed by American Indians (0.92), Native Hawaiians/Pacific Islanders (0.90), Japanese/East Asians (0.87), and African Americans (0.84). Compared with European Americans, the proportion of SNPs that differed in allele frequency by more than ±0.10 was smallest in Mexican Americans/Hispanics (seven SNPs; 14%) and largest in African Americans (25 SNPs; 51%). For the remaining populations, approximately half of the SNPs genotyped differed in allele frequency compared with European Americans by more than ±0.10. A striking example of population differences in allele frequencies is *FADS1* rs174547. The T allele of *FADS1* rs174547 is the major allele in three populations (allele frequency =

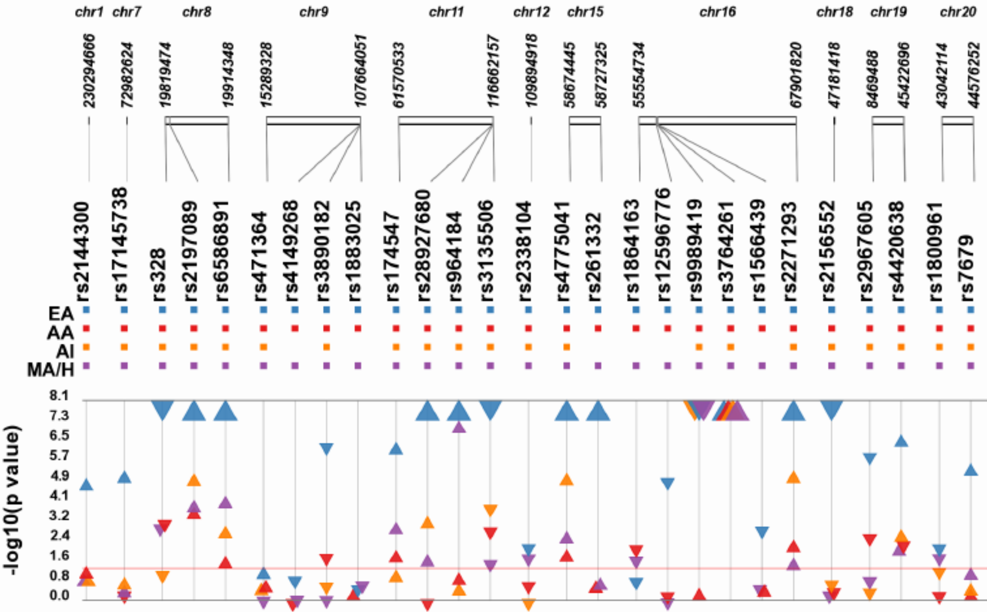
0.66, 0.91, and 0.59 in European Americans, African Americans, and Japanese/East Asians, respectively), but is the minor allele in the other three populations (allele frequency = 0.39, 0.21, and 0.42 in Mexican Americans/Hispanics, American Indians, and Native Hawaiians/Pacific Islanders, respectively).

Replication in European-descent populations

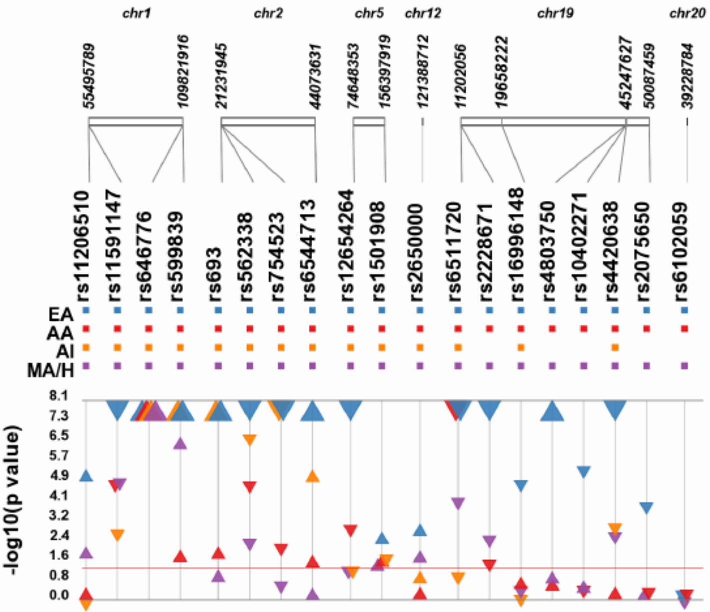
We meta-analyzed tests of association for 27, 19, and 14 SNPs previously associated with HDL-C, LDL-C, and/or triglycerides, respectively, across European American populations collected by individual PAGE study sites (Appendix I). For HDL-C, 23 of the 27 (85%) SNPs tested were associated at $p<0.05$ assuming an additive genetic model and adjusting for age and sex (Figure 4.1 and Table 4.1). The four SNPs that did not replicate at this liberal significance threshold were rs471364 (*TTC39B*), rs1883025 (*ABCA1*), rs4149268 (*ABCA1*), and rs1864163 (*CETP*), all of which are intronic (Appendix I). For LDL-C, only one (intergenic *MAFB* rs6102059) of the 19 SNPs tested was not significantly associated at $p<0.05$ (Figure 4.1 and Table 4.2). Finally, for ln(TG), all 14 SNPs tested were associated at $p<0.05$ (Figure 4.1 and Table 4.3).

Of the associations that did not replicate in the European-descent populations from PAGE, four out of five had sufficient power (>80%) to detect the previously reported effect size: *TTC39B* rs471364 (>99% power; HDL-C), *CETP* rs1864163 (80% power; HDL-C); *MAFB* rs6102059 (>90% power; LDL-C), and *ABCA1* rs4149268 (99% power; HDL-C). *ABCA1* rs1883025, which did not replicate the expected association with HDL-C, did not have sufficient power to detect the reported effect size (68% power; n=3,865).

HDL-C



LDL-C



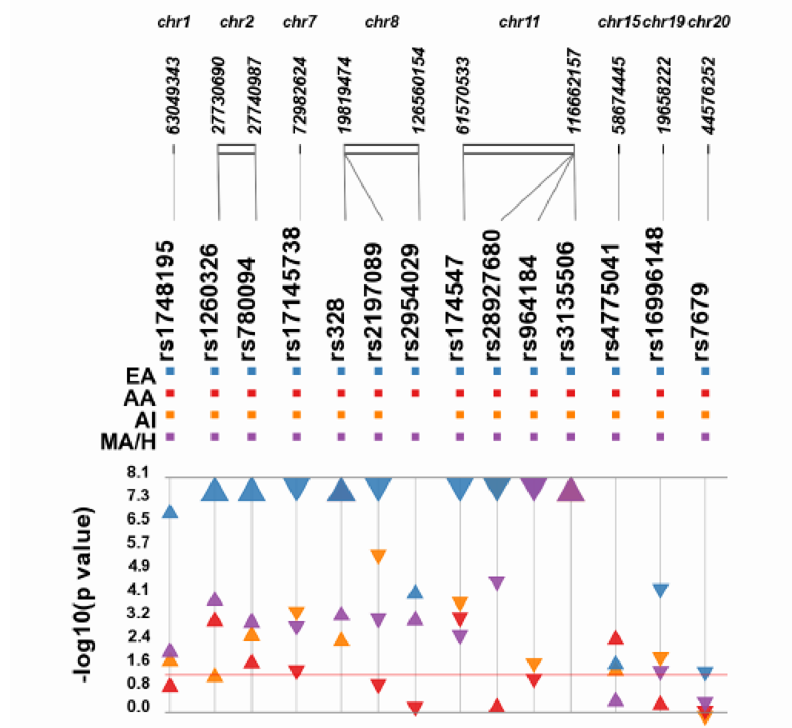


Figure 4.1. Meta-analysis results for GWAS-identified SNPs by population. Each SNP was tested for an association with the indicated trait assuming an additive genetic model adjusted for age and sex. Meta-analysis was performed, and p-values (–$\log_{10}$ transformed) of the meta-analysis are plotted along the y-axis. SNP location is given on the x-axis. Each triangle represents a meta-analysis p-value for each population. Populations are color-coded as follows: European Americans (blue; EA), African Americans (red; AA), Mexican Americans/Hispanics (orange; MA/H), and American Indians (purple; AI). Large triangles represent p-values at or smaller than genome-wide significance ($p<10^{-8}$). The direction of the arrows corresponds to the direction of the beta coefficient. The significance threshold is indicated by the red bar at p=0.05.

Table 4.2. Meta-analysis of GWAS-identified HDL-C SNPs. Abbreviations: coded allele (CA); coded allele frequency (CAF); beta coefficient (β); standard error (SE); data not available (--).

SNP	Nearest Gene	CA	European Americans (n_{max}=25,167)			African Americans (n_{max}=10,436)			American Indians (n_{max}=6,134)			Mexican Americans and Hispanics (n_{max}=3,371)		
			CAF	β (SE)	P-value	CAF	β (SE)	P-value	CAF	β (SE)	P-value	CAF	β (SE)	P-value
rs2144300	*GALNT2*	T	0.60	0.59 (0.14)	3.33E-05	0.15	0.48 (0.31)	0.12	0.55	0.29 (0.25)	0.25	0.56	0.39 (0.34)	0.25
rs17145738	*MLXIPL*	T	0.12	0.91 (0.21)	1.64E-05	0.09	-0.28 (0.40)	0.46	0.08	0.46 (0.48)	0.34	0.07	0.25 (0.67)	0.71
rs328	*LPL*	C	0.90	-2.29 (0.24)	5.60E-22	0.93	-1.79 (0.52)	5.84E-04	0.97	-1.55 (0.85)	0.07	0.94	-2.31 (0.69)	8.80E-04
rs2197089	*LPL*	A	0.55	0.90 (0.13)	9.49E-11	0.78	0.95 (0.27)	4.79E-04	0.40	1.10 (0.26)	2.19E-05	0.47	1.22 (0.33)	2.56E-04
rs6586891	*LPL*	A	0.66	0.96 (0.14)	5.88E-11	0.84	0.60 (0.30)	4.76E-02	0.44	0.76 (0.26)	2.86E-03	0.53	1.27 (0.34)	1.73E-04
rs471364	*TTC39B*	T	0.89	0.35 (0.23)	0.13	0.81	0.24 (0.31)	0.45	0.97	0.43 (0.77)	0.58	0.92	-0.23 (0.69)	0.74
rs4149268	*ABCA1*	T	0.37	-0.30 (0.18)	0.12	0.67	-0.03 (0.35)	0.92	--	--	--	0.32	-0.17 (0.40)	0.67
rs3890182	*ABCA1*	A	0.12	-1.06 (0.20)	4.53E-07	0.12	-0.83 (0.34)	1.39E-02	0.05	-0.92 (0.72)	0.20	0.09	-0.24 (0.63)	0.70
rs1883025	*ABCA1*	A	0.26	-0.44 (0.38)	0.25	0.34	0.02 (0.56)	0.97	--	--	--	0.27	-0.59 (0.44)	0.18
rs174547	*FADS1*	T	0.66	0.84 (0.17)	1.14E-06	0.91	0.94 (0.42)	2.73E-02	0.21	0.56 (0.41)	0.17	0.39	1.17 (0.38)	1.98E-03
rs28927680	*APOA1/C3/A4/A5*	C	0.93	1.51 (0.26)	8.61E-09	0.84	-0.03 (0.30)	0.93	0.83	1.19 (0.37)	1.13E-03	0.86	1.00 (0.48)	3.98E-02

rs964184	*APOA1/C3/A4/A5*	C	0.86	1.57 (0.25)	6.08E-10	0.80	0.48 (0.39)	0.22	0.78	1.32 (2.48)	0.60	0.71	1.98 (0.38)	1.55E-07
rs3135506	*APOA1/C3/A4/A5*	C	0.06	-1.86 (0.31)	1.42E-09	0.06	-1.94 (0.60)	1.17E-03	0.17	-1.41 (0.37)	1.40E-04	0.14	-1.22 (0.54)	2.45E-02
rs2338104	*MMAB-MVK*	C	0.46	-0.40 (0.14)	5.64E-03	0.27	-0.35 (0.27)	0.19	0.58	-0.03 (0.26)	0.91	0.52	-0.92 (0.38)	1.46E-02
rs4775041	*LIPC*	C	0.29	1.31 (0.16)	1.03E-16	0.14	0.79 (0.35)	2.55E-02	0.21	1.34 (0.47)	2.05E-05	0.18	1.34 (0.47)	4.66E-03
rs261332	*LIPC*	A	0.20	1.76 (0.24)	1.99E-13	0.24	0.31 (0.43)	0.46	--	--	--	0.15	0.66 (0.72)	0.35
rs1864163	*CETP*	A	0.23	-2.07 (1.36)	0.13	0.27	-2.79 (1.02)	6.19E-03	--	--	--	0.28	-2.98 (1.26)	1.78E-02
rs12596776	*CETP*	C	0.90	-1.36 (0.31)	1.18E-05	0.94	-0.48 (0.70)	0.50	--	--	--	0.94	-0.13 (0.75)	0.86
rs9989419	*CETP*	A	0.39	-2.17 (0.14)	1.71E-53	0.59	0.02 (0.24)	0.93	0.26	-1.62 (0.30)	4.42E-08	0.32	-2.29 (0.39)	5.29E-09
rs3764261	*CETP*	A	0.32	3.64 (0.15)	8.83E-129	0.32	2.79 (0.25)	5.98E-28	0.31	2.81 (0.27)	5.00E-25	0.33	2.68 (0.40)	2.53E-11
rs1566439	*CETP*	T	0.60	-0.54 (0.16)	1.07E-03	0.78	0.16 (0.37)	0.67	--	--	--	0.53	-0.42 (0.37)	0.25
rs2271293	*LCAT*	A	0.12	1.45 (0.22)	8.40E-11	0.68	1.11 (0.43)	1.05E-02	0.26	1.26 (0.29)	1.65E-05	0.14	0.99 (0.52)	5.65E-02
rs2156552	*LIPG*	A	0.17	-1.27 (0.19)	5.11E-11	0.04	-0.59 (0.62)	0.34	0.05	-0.96 (0.69)	0.17	0.08	-0.49 (0.68)	0.47
rs2967605	*ANGPTL4*	T	0.18	-0.90 (0.18)	1.12E-06	0.21	-0.89 (0.29)	2.24E-03	0.30	-0.26 (0.28)	0.35	0.23	-0.67 (0.43)	0.12
rs4420638	*APOE/C1/C4*	A	0.82	1.00 (0.20)	5.69E-07	0.80	-1.01 (0.35)	4.29E-03	0.90	1.38 (0.48)	3.95E-03	0.90	1.45 (0.59)	1.47E-02

rs1800961	*HNF4A*	T	0.03	-1.14 (0.41)	5.78E-03	0.01	-1.01 (1.46)	0.49	0.03	-1.43 (0.73)	0.05	0.04	-2.33 (0.95)	1.42E-02
rs7679	*PLTP*	T	0.82	0.95 (0.21)	8.42E-06	0.96	0.01 (0.58)	0.99	0.94	0.31 (0.58)	0.60	0.89	0.89 (0.60)	0.14

Table 4.3. Meta-analysis of GWAS-identified LDL-C SNPs. Abbreviations: coded allele (CA); coded allele frequency (CAF); beta coefficient (β); standard error (SE); data not available (--).

SNP	Nearest Gene	CA	European Americans (n_{max}=21,986)			African Americans (n_{max}=9,328)			American Indians (n_{max}=6,144)			Mexican Americans and Hispanics (n_{max}=2,532)		
			CAF	β (SE)	P-value	CAF	β (SE)	P-value	CAF	β (SE)	P-value	CAF	β (SE)	P-value
rs11206510	*PCSK9*	T	0.81	1.98 (0.45)	1.44E-05	0.86	0.09 (0.84)	0.91	0.93	-0.07 (1.30)	0.96	0.88	3.36 (1.44)	1.97E-02
rs11591147	*PCSK9*	T	0.02	-16.92 (1.42)	1.00E-32	4.10E-03	-22.64 (5.21)	1.41E-05	0.01	-15.66 (4.92)	1.44E-03	0.01	-23.39 (5.34)	1.19E-05
rs646776	*CELSR2/PSCR1/ SORT*	T	0.78	5.74 (0.44)	1.44E-37	0.65	4.46 (0.63)	1.48E-12	--	--	--	0.81	7.70 (1.41)	4.49E-08
rs599839	*CELSR2/PSRC1/ SORT*	A	0.77	5.67 (0.45)	3.61E-36	0.28	1.60 (0.72)	2.67E-02	0.78	6.17 (0.67)	3.94E-20	0.78	8.68 (1.75)	6.99E-07
rs693	*APOB*	A	0.50	3.45 (0.36)	3.38E-21	0.24	1.60 (0.69)	2.04E-02	0.34	4.02 (0.59)	7.08E-12	0.38	1.38 (1.02)	0.18
rs562338	*APOB*	A	0.19	-5.52 (0.45)	1.05E-33	0.59	-2.54 (0.59)	1.57E-5	0.09	-5.44 (1.05)	1.93E-07	0.16	-3.90 (1.33)	3.42E-03
rs754523	*APOB*	A	0.68	-3.64 (0.40)	3.44E-19	0.78	-2.12 (0.76)	5.52E-03	0.66	-4.26 (0.61)	2.17E-12	0.72	-1.63 (1.23)	0.19
rs6544713	*ABCG8*	T	0.31	2.98 (0.42)	1.17E-12	0.17	1.49 (0.74)	4.45E-02	0.11	4.76 (1.10)	1.51E-05	0.18	0.06 (1.22)	0.96
rs12654264	*HMGCR*	A	0.62	-2.66 (0.37)	6.56E-13	0.67	-2.02 (0.61)	9.39E-04	0.58	-1.17 (0.59)	4.55E-02	0.62	-2.06 (1.04)	4.68E-02
rs1501908	*TIMD4*	C	0.64	1.23 (0.44)	4.961E-03	0.37	1.31 (0.64)	4.18E-02	0.85	-2.18 (0.89)	1.46E-02	0.76	2.40 (1.28)	6.19E-02
rs2650000	*HNF1A*	A	0.35	1.20 (0.39)	2.338E-03	0.12	0.15 (0.97)	0.88	0.41	0.73 (0.57)	0.20	0.37	2.58 (1.17)	2.84E-02

rs6511720	*LDLR*	T	0.12	-7.32 (0.52)	2.99E-44	0.13	-8.10 (0.80)	7.05E-24	0.07	-2.48 (1.41)	0.08	0.09	-6.43 (1.62)	7.34E-05
rs2228671	*LDLR*	T	0.12	-5.83 (0.97)	1.96E-09	0.04	-6.62 (2.94)	2.43E-02	--	--	--	0.08	-6.14 (2.03)	2.53E-03
rs16996148	*CILP2/PBX4/ NCAN1*	T	0.08	-2.88 (0.66)	1.40E-05	0.15	0.77 (0.80)	0.34	0.04	-0.70 (1.51)	0.64	0.06	-2.12 (2.01)	0.29
rs4803750	*BCL3*	A	0.93	5.57 (0.92)	1.37E-09	0.92	1.52 (1.77)	0.39	--	--	--	0.86	5.95 (4.68)	0.20
rs10402271	*APOE/C1/C4*	T	0.67	-2.27 (0.49)	3.86E-06	0.84	-1.38 (1.28)	0.28	--	--	--	0.61	2.39 (3.38)	0.48
rs4420638	*APOE/C1/C4*	A	0.82	-5.34 (0.51)	2.16E-25	0.79	0.16 (0.92)	0.87	0.90	-3.57 (1.07)	8.00E-04	0.90	-5.35 (1.72)	1.82E-03
rs2075650	*TOMM40*	A	0.88	-4.77 (1.23)	1.14E-04	0.87	-2.26 (2.40)	0.35	--	--	--	0.90	0.02 (5.27)	1.00
rs6102059	*MAFB*	T	0.30	-0.41 (0.52)	0.42	0.43	-0.77 (0.88)	0.38	--	--	--	0.29	-0.30 (1.22)	0.80

Table 4.4. Meta-analysis of GWAS-identified Triglyceride SNPs. Abbreviations: coded allele (CA); coded allele frequency (CAF); beta coefficient (β); standard error (SE); data not available (--).

SNP	Nearest Gene	CA	European Americans (n_{max}=21,986)			African Americans (n_{max}=9,328)			American Indians (n_{max}=6,144)			Mexican Americans and Hispanics (n_{max}=2,532)		
			CAF	β (SE)	P-value	CAF	β (SE)	P-value	CAF	β (SE)	P-value	CAF	β (SE)	P-value
rs1748195	*ANGPTL3*	C	0.66	0.03 (0.01)	1.93E-07	0.35	0.01 (0.01)	0.19	0.61	0.16 (0.07)	2.44E-02	0.60	0.04 (0.01)	1.17E-02
rs1260326	*GCKR*	T	0.42	0.05 (0.01)	6.44E-13	0.16	0.05 (0.02)	9.98E-04	0.28	0.15 (0.09)	8.52E-02	0.33	0.06 (0.02)	1.97E-04
rs780094	*GCKR*	T	0.40	0.06 (0.01)	1.69E-32	0.18	0.02 (0.01)	2.91E-02	0.25	0.04 (0.01)	3.23E-03	0.33	0.06 (0.02)	1.13E-03
rs17145738	*MLXIPL*	T	0.12	-0.07 (0.01)	5.71E-24	0.09	-0.03 (0.01)	2.53E-02	0.08	-0.07 (0.02)	2.30E-04	0.07	-0.09 (0.03)	7.40E-04
rs328	*LPL*	C	0.90	0.09 (0.01)	4.16E-30	0.93	0.08 (0.02)	2.62E-08	0.97	0.09 (0.03)	4.83E-03	0.93	0.09 (0.03)	6.31E-04
rs2197089	*LPL*	A	0.55	-0.03 (0.01)	4.97E-15	0.78	-0.01 (0.01)	7.45E-02	0.41	-0.05 (0.01)	2.57E-06	0.48	-0.05 (0.01)	4.01E-04
rs2954029	*TRIB1*	A	0.54	0.05 (0.01)	1.13E-04	0.68	-0.01 (0.02)	0.46	--	--	--	0.62	0.06 (0.02)	9.28E-04
rs174547	*FADS1*	T	0.66	-0.03 (0.01)	3.82E-10	0.91	-0.05 (0.01)	3.73E-04	0.21	-0.06 (0.02)	1.10E-04	0.39	-0.05 (0.02)	1.51E-03
rs28927680	*APOA1/C3/A4/A5 gene cluster*	C	0.93	-0.12 (0.01)	2.88E-38	0.84	<0.001 (0.01)	0.95	0.83	-0.13 (0.01)	6.33E-19	0.86	-0.08 (0.02)	2.15E-05
rs964184	*APOA1/C3/A4/A5 gene cluster*	C	0.86	-0.14 (0.01)	1.91E-59	0.80	-0.02 (0.01)	4.87E-02	0.78	-0.17 (0.07)	1.43E-02	0.72	-0.14 (0.02)	1.04E-19
rs3135506	*APOA1/C3/A4/A5 gene cluster*	C	0.06	0.13 (0.01)	2.59E-33	0.06	0.11 (0.02)	2.06E-10	0.17	0.13 (0.01)	4.28E-20	0.14	0.13 (0.02)	3.08E-08

rs4775041	*LIPC*	C	0.29	0.01 (0.01)	3.15E-02	0.14	0.03 (0.01)	4.29E-03	0.21	0.02 (0.01)	5.15E-02	0.18	0.01 (0.02)	0.58
rs16996148	*CILP2/PBX4/ NCAN*	T	0.08	-0.04 (0.01)	3.91E-05	0.15	<0.001 (0.01)	0.77	0.04	-0.07 (0.03)	8.86E-03	0.06	-0.06 (0.03)	2.69E-02
rs7679	*PLTP*	T	0.82	-0.02 (0.01)	2.84E-02	0.96	-0.01 (0.02)	0.61	0.94	-2.0E-03 [illegible]	0.93	0.89	-0.03 (0.03)	0.31

We then compared the genetic effect sizes reported in the literature to the genetic effect sizes estimated from the meta-analysis of these population-based studies. We observed that the majority of the point estimates of effect size (β) were smaller than previously reported estimates. Using the HDL-C association results as an example, 15 out of the 23 (65%) significant associations had effect estimates smaller than published effect estimates. We caution, however, that we did not formally test for significant differences between estimates and that these smaller effect estimates may or may not be significantly different than the published reports. However, it is interesting to note that 11 of our effect estimates differed from previous reports by more than 25%, including two HDL-C associations whose effect sizes differed by 50% or more from those in the literature (*ANGPTL4* rs2967605 and *MLXIPL* rs17145738; Table 4.2 and Appendix I).

Associations in non-European-descent populations

We meta-analyzed tests of association performed in African Americans for the same 27, 19, and 14 SNPs previously associated with HDL-C, LDL-C, and/or triglycerides in populations of European-descent. For all three traits studied, assuming an additive genetic model and adjusting for age and sex, approximately half of the tested GWAS-identified SNPs were associated at $p<0.05$: 12/27 (44%) for HDL-C, 11/19 (58%) for LDL-C, and 8/14 (57%) for ln(TG) (Figure 4.1, Tables 4.2-4.4). The majority of SNPs that failed to replicate in the meta-analysis for European Americans also failed to associate in the meta-analysis for African Americans. Interestingly, one SNP (*CETP* rs1864163) was significantly associated with HDL-C in African Americans (n=451; CAF=0.27; $\beta = -2.79$; $p=6.19 \times 10^{-3}$) but not in European Americans (n=291; CAF=0.23; $\beta = -2.07$; $p=0.13$).

Other populations that were examined for select SNPs included American Indians, Mexican Americans/Hispanics, Japanese/East Asians, and Native Hawaiians/Pacific Islanders. Among American Indians, 9/21 (43%), 10/14 (71%), and 10/13 (77%) of the SNPs tested for association with HDL-C, LDL-C, and ln(TG), respectively, were associated at the liberal significance threshold of $p<0.05$. For Mexican Americans/Hispanics, 14/27 (52%), 10/19 (53%), and 12/14 (86%) SNPs were significantly associated at $p<0.05$ with HDL-C, LDL-C, and ln(TG), respectively. Despite a small sample size, intronic *CETP* rs1864163 was significantly associated with HDL-C in Mexican Americans/Hispanics (n=265; CAF=0.28; β = -2.98; $p=1.78\times10^{-2}$) but not in European Americans (n=291; CAF=0.27; β = -2.07; p=0.13), although the size and the direction of effect were similar.

The sample sizes for Japanese/East Asians and Native Hawaiians/Pacific Islanders are considerably smaller compared with the other populations examined. Despite the lower power to detect associations, significant associations were observed for both groups at a liberal significance threshold of $p<0.05$. Among the 26, 18, and 13 SNPs tested for associations with HDL-C, LDL-C, and ln(TG), respectively, there were nine (35%), three (17%), and three (23%) SNPs significantly associated in the combined Japanese/East Asian group.

For Native Hawaiians/Pacific Islanders, the group with the smallest sample size considered here, one SNP each was associated with HDL-C (*APOA1/C3/A4/A5* gene cluster rs28927680) and LDL-C (*APOB* rs754523) out of the 24 and 18 SNPs tested for association, respectively. Three out of 12 SNPs tested for an association with ln(TG) were associated at $p<0.05$ (*PLTP* rs7679, *MLXIPL* rs17145738, and *APOA1/C3/A4/A5* gene cluster rs28927680), with the latter at a significance of $p<10^{-19}$.

Generalization across non-European-descent populations

For the 55 SNP-trait associations that replicated in European Americans, we determined which associations generalized across all four of our largest populations (European Americans, African Americans, American Indians, and Mexican Americans/Hispanics). Generalization was based on two criteria: 1) level of significance (i.e. p-value) and 2) direction of effect (i.e. positive or negative beta) (Appendix Q). SNPs that were significantly associated at $p<0.05$ and had the same direction of effect as European Americans in all populations studied were considered to have generalized. For HDL-C, five SNPs (*CETP* rs3764261, *LPL* rs6586891, *LIPC* rs4775041, *LPL* rs2197089, and *APOA1/C3/A4/A5* gene cluster rs3135506) met these criteria (Appendix Q), and two SNPs (*LCAT* rs2271293 and *LPL* rs328) were associated in three groups and trended towards significance in a fourth group ($p=0.06$ and $p=0.07$ in Mexican Americans/Hispanics and American Indians, respectively; Table 4.2).

For LDL-C, six SNPs generalized across all four groups, if available: *CELSR2/PSRC1/SORT* rs599839 and rs646776, *APOB* rs562338, *PCSK9* rs11591147, *HMGCR* rs12654264, and *LDLR* rs2228671 (Appendix Q). Similarly for ln(TG), six SNPs were significantly associated across the four largest populations: *APOA1/C3/A4/A5* gene cluster rs964184 and rs3135506, *GCKR* rs780094, *LPL* rs328, *MLXIPL* rs1714573, and *FADS1* rs174547 (Appendix Q). In addition, for ln(TG), two SNPs (*LPL* rs2197089 and *GCKR* rs1260326) were associated in three groups and trended towards significance in a fourth group ($p=0.07$ in African Americans and $p=0.09$ in American Indians, respectively). Among the 16 SNPs that generalized across the largest groups among the three lipid traits, only four (25%) were either nonsense (rs328) or missense SNPs (rs3135506, rs11591147, and rs1260326; Appendix I).

Power

Based on our definition of generalization, several SNPs discovered and replicated in European-descent populations failed to generalize to other populations. There are several possible explanations for non-generalization including power. To further investigate potential lack of power, for SNPs that associated in European Americans but not in African Americans, Mexican Americans/Hispanics, and/or American Indians, we performed post hoc power calculations assuming the genetic effect size estimated in European Americans and the allele frequency and sample sizes observed in non-European Americans. In African Americans, four HDL-C (*APOA1/C3/A4/A5* gene cluster rs28927680 and rs964184, *LIPC* rs261332, and *CETP* rs9989419), three LDL-C (*CILP2/PBX4* rs16996148, *BCL3* rs4803750, and *APOE/C1/C4* rs4420638) and four ln(TG) (*ANGPTL3* rs1748195, *LPL* rs2197089, *TRIB1* rs2954029, and *APOA1/C3/A4/A5* rs28927680) tests of association were properly powered (>80%) but did not generalize (Appendix Q). Interestingly, *APOE/C1/C4* rs4420638 was also properly powered and significantly associated with HDL-C levels in African Americans (β = -1.01; p=4.29×10^{-3}), but effect size was in the opposite direction compared with European Americans (β = 1.00; p=5.69×10^{-7}).

For American Indians, two HDL-C (*LPL* rs328 and *ANGPTL4* rs2967605) and one LDL-C (*LDLR* rs6511720) tests of association did not generalize despite being properly powered. In this population, we observed one properly powered significant association that failed to generalize due to inconsistent direction of effect: *TIMD4* rs1501908 for LDL-C. In Mexican-Americans/Hispanics two LDL-C (*APOB* rs693 and rs754523) tests of association failed to generalize despite sufficient power. All other tests of association that did not generalize to African Americans, American Indians, or Mexican

Americans/Hispanics were underpowered assuming that the expected genetic effect sizes were similar to those in European Americans.

Linkage disequilibrium

To examine whether LD can account for the lack of generalization of the properly powered tests of association in African Americans, we examined LD patterns in HapMap Europeans (CEU) and West Africans (YRI) as well as those published in the literature for the genotyped SNPs and surrounding variation. For *APOA1/C3/A4/A*5 rs28927680, previous studies in European-descent populations have noted that this SNP is in strong LD (r^2=0.98) with missense *APOA5* rs3135506 (Kathiresan et al., 2008). *APOA1/C3/A4/A5* rs964184 is also in moderate LD with missense rs3135506 (r^2=0.510 in CEU). However, neither rs28927680 nor rs964184 are in LD with missense rs3135506 (r^2=0.039 and r^2=0.048) in YRI. Furthermore, *APOA5* rs3135506 is associated with HDL-C in European Americans, African Americans, Mexican Americans/Hispanics, and American Indians (Tables 4.1 and 4.2). Generalization of rs3135506 coupled with non-generalization and differences in YRI LD patterns for rs28927680 and rs964184 suggest that *APOA5* rs3135506 is either the putative functional SNP for the association with HDL-C or in LD with the functional SNP.

Other interpretations of LD patterns are more difficult. For example, *CETP* rs9989419, which failed to generalize in African Americans for HDL-C despite sufficient power, is not in strong LD with obvious functional SNPs in CEU within 50kb flanking the genotyped SNP. The strongest pair-wise LD (r^2=0.251) consists of intergenic and intronic SNPs, and these same SNPs have weak LD (r^2<0.03) or are not found in YRI. Similarly, *LIPC* rs261332 associated with HDL-C levels in European Americans but

failed to generalize in African Americans. *LIPC* rs261332 is in strong LD (r^2>0.80 in CEU) with SNPs in the 5′ flanking region of *LIPC*, but not in LD with these same SNPs in YRI (r^2<0.15).

Adjustments for exposures and co-morbidities

Genetic variations in isolation are not the sole determinants of lipid trait distributions. Many environmental exposures and demographic variables are associated with lipid traits. To account for these variables, we meta-analyzed all tests of association for HDL-C, LDL-C, and ln(TG) adjusted for age, sex, body mass index, current smoking, type 2 diabetes, post-menopausal status, current hormone use, and previous myocardial infarction (Models 3 and 4) did not appreciably alter the results compared with the models minimally adjusted for age and sex (Appendices J-L), except for some associations with LDL-C in Japanese/East Asians (Appendix K.e). More specifically, four SNPs (rs562338, rs6544713, rs1501908, and rs10402271) were not associated in the minimally adjusted models (Models 1 and 2) but they were associated at $p<0.05$ in the more fully adjusted models (Models 3 and 4). This discrepancy simply may be due to decreased sample sizes between the models (from n=690 in Models 1 and 2; n=653 in Models 3 and 4), or it may highlight the need to account for these coviariates in this particular racial/ethnic population.

Effect of including versus excluding by medication use

All analyses presented thus far include fasting adult participants regardless of lipid lowering medication use. Many GWA studies conducted for the lipid traits excluded participants on lipid lowering medication (Kathiresan et al., 2008; Kathiresan

et al., 2009; Willer et al., 2008) given that these medications substantially lower LDL-C levels. We have included these participants for analysis as participants on lipid lowering medication could represent the upper extreme of the normal LDL-C distribution associated with a genetic profile found in a general population. Exclusion of these participants would preclude these meta-analyses from fully describing the extent and strength of associations relevant to these traits in a population-based setting. However, if genetic variation is associated with lipid concentrations and medication use lowers lipid concentrations, inclusion of participants on lipid lowering medications could bias associations towards the null. As a sensitivity analysis, WHI used detailed medication data available on a subset of participants, and performed the tests of association for HDL-C, LDL-C, and ln(TG) excluding and including participants on lipid lowering medication with the latter adjusted for medication usage using average effects estimated in Wu et al (Wu et al., 2007) for specific drug classes. Appendix R suggests that both the point estimates and the confidence intervals of the genetic effects are similar for this female-only study whether participants are excluded or included and adjusted for medication use.

We also performed a second sensitivity analysis: tests of association excluding participants on lipid lowering medication for all models. As detailed in Appendices M-O, excluding participants on lipid lowering medication usage does not appreciably alter the results, with the possible exception of LDL-associations in Japanese/East Asians. More specifically, two SNPs (rs11206510 and rs1501908) became significantly associated with LDL-C after excluding individuals on medications while two other SNPs (rs562338 amd rs6544713) were no longer significantly associated. Much like in the different modeling schemes mentioned earlier, this discrepancy simply may be due to decreased

sample sizes after excluding individuals on lipid lowering medications (from n=690 to n=467). Of note, use of lipid-lowering medications was low (<10%) in the ARIC, CHS, NHANES, and WHI studies since the majority of study recruitment occurred before the introduction or widespread use of the recent generation of lipid-lowering medications. However, medication use was higher in the MEC study (20-38% depending on the population), which contributed the majority of Japanese/East Asian samples.

Discussion

We have performed an extensive replication and generalization effort for HDL-C, LDL-C, and TG GWAS-identified SNPs. The PAGE study consists of six racial/ethnic groups: European American, African American, Mexican American/Hispanic, American Indian, Japanese/East Asian, and Native Hawaiian/Pacific Islander, with population-specific sample sizes ranging from ~100 to >20,000 for any one test of association. Although power to detect associations varied across the lipid traits and populations, we observed general patterns worth noting for future genetic epidemiological studies.

Replication in European-descent populations

Perhaps not unexpectedly, we were able to replicate most reported associations in European Americans. Regardless of significance, all but one of the tested SNPs had effect estimates in the same direction as the previously reported association from the literature. *FADS1* rs174547, which was significantly associated with decreased ln(TG) in this meta-analysis for European Americans, was associated with increased TG in

European Americans from the Framingham Heart Study (n=7,423) (Kathiresan et al., 2009). HDL-C had proportionally (15%) the greatest number of SNPs that failed to replicate in European Americans compared with LDL-C (5%) and TG (0%) despite the fact that we had sufficient power to detect the reported genetic effect size for many of these tests. *TTC39B* rs471364 was not associated with HDL-C levels despite a sample size of 18,089 and >99% power to detect the reported effect size. Neither *ABCA1* rs4149268 nor rs1883025 was associated with HDL-C, although the latter test of association was underpowered (68%; n=3,865). Finally, as previously discussed, *CETP* rs1864163 was not associated with HDL-C in this European American dataset although we had 80% power to detect the reported genetic effect size. For LDL-C, only *MAFB* rs6102059 was not associated despite >90% power to detect the reported effect size.

The reasons for non-replication in this European American dataset for properly powered tests of association are unclear. It is possible that we have overestimated our power to detect reported associations. The "winner's curse" and inflated genetic effect estimates from initial discovery are well known (Goring et al., 2001; Zollner and Pritchard, 2007). Indeed, for the five SNPs that did not replicate in this meta-analysis for European Americans, the association was described in only one GWAS each despite the fact that numerous GWAS (Aulchenko et al., 2009; Heid et al., 2008; Kathiresan et al., 2008; Kathiresan et al., 2009; Kooner et al., 2008; Pollin et al., 2008; Ridker et al., 2009; Sabatti et al., 2009; Sandhu et al., 2008; Saxena et al., 2007; Wallace et al., 2008; Willer et al., 2008) and a large meta-analysis (Teslovich et al., 2010) for these three traits have been conducted in populations of European-descent. The meta-analysis recently reported by Teslovich et al (Teslovich et al., 2010) did report significant associations between *TTC39B* rs581080 for HDL-C and *MAFB* rs2902940 for LDL-C. *TTC39B* rs581080 is in moderate

linkage disequilibrium (LD) with rs471364 (r^2 = 0.49 in CEU HapMap), but *MAFB* rs2902940 is not in LD with rs6102059 (r^2 = 0.03 in HapMap CEU).

A second possibility for our observed non-replication is heterogeneity among the PAGE studies. A test of heterogeneity suggested that the association results for *TTC39B* rs471364 had significant evidence for heterogeneity across studies ($p_{heterogeneity}$=0.048; I^2=58.25%). Only two other association results had evidence for heterogeneity: *FADS1* rs174547 for HDL-C ($p_{heterogeneity}$=0.006; I^2=75.73%) and *PCSK9* rs11206510 for LDL-C ($p_{heterogeneity}$=0.048; I^2=55.34%). However, for both of these loci, the tests of association were significant in European Americans.

Generalization to non-European populations

When taking into account power, significance, and direction of effect, most SNPs discovered in European Americans generalized to African Americans, Mexican Americans, and American Indians. Of note are the eleven tests of association significant in European Americans that did not generalize to African Americans despite having adequate power. Given that GWAS products are a mixture of tagSNPs and functional SNPs, it is likely that discovery in European Americans represents tagSNPs rather than the true functional SNP. Because linkage disequilibrium patterns differ across populations, tagSNPs genotyped directly in populations of non-European descent may not recapitulate the association observed in European-descent populations depending on the pattern of LD. The association of HDL-C and nonsynonymous rs3135506 versus tagSNPs rs28927680 in the *APOA1/C3/A4/A5*gene cluster in this analysis is an example of the effects of LD and the ability to generalize across populations.

Evoking LD as an explanation for lack of generalization is appealing, but it does have limitations given that the functional SNP is not often obvious. All tests of association that did not generalize to African Americans had evidence of LD differences between CEU and YRI using the HapMap data. However, most of these SNPs are located in the intergenic and intronic regions. Further fine-mapping in both the discovery population as well as other diverse populations will be needed along with a better understanding of genetic variation and its relationship to biological function to identify the true functional SNPs for these traits.

Among the five putative functional SNPs genotyped (nonsynonymous rs11591147, rs1260326, rs3135506, and rs1800961 and nonsense rs328), all five replicated in populations of European-descent, and three of the five generalized to populations of non-European descent. One putative functional SNP that did not replicate across ethnicities was *HNF4A* rs1800961, likely due to low power because of the very low minor allele frequency in all subpopulations (0.0065 to 0.0398). Both the direction and magnitude of effect, however, were consistent across groups. *GCKR* rs1260326 did not generalize to all populations of non-European descent but did generalize in three of the four populations tested and trended towards significance in American Indians (p=0.085; Table 4.4).

<u>Limitations and strengths</u>

The major strengths and limitations of the PAGE study for lipids are sample size and diversity. The largest sample size is for samples of European-descent (~20,000), followed by African Americans and American Indians. The sample sizes for Mexican Americans, Japanese/East Asians, and Pacific Islanders/Native Hawaiians are smaller

and consequently underpowered for tests of association as estimated from genetic effect sizes in the published European-descent discovery studies. Also, not all SNPs were genotyped in all PAGE studies, further affecting the power of the meta-analyses.

An additional limitation is the lack of data related to lipid lowering medication. Ideally, all analyses would be adjusted for use of lipid lowering medication based on the type and dose of medication. In most PAGE studies, these data were not available and in many, use was low at baseline when blood samples were obtained. As we demonstrate in Supplementary material, inclusion of participants using lipid-lowering medication did not appreciably alter the results of the meta-analysis for the majority of racial/ethnic populations when compared with excluding these participants.

In general, the cohorts and surveys included in PAGE are diverse with regard to demographics, genetic ancestry, lifestyle, health, and environmental exposure. Despite this diversity, very few tests of association from the meta-analysis exhibited evidence of heterogeneity.

Conclusions

Overall, the majority of GWAS-identified SNPs for HDL-C, LDL-C, and TG replicated in European Americans and generalized to non-European-descent populations. These results suggest that the genotyped SNP either tags the functional SNP(s) common across these populations or that the genotyped SNP represents the risk SNP directly. SNPs that replicated in European Americans but did not generalize in the largest non-European-descent populations, despite adequate power, could represent priority associations that require fine-mapping and re-sequencing to identify the functional variant(s).

Acknowledgements

The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALiCo), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004790 (WHI), and U01HG004801 (Coordinating Center). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The complete list of PAGE members can be found at http://www.pagestudy.org.

The data and materials included in this report result from a collaboration between the following studies:

The "Epidemiologic Architecture for Genes Linked to Environment (EAGLE)" is funded through the NHGRI PAGE program (U01HG004798-01). Genotyping services for select NHANES III SNPs presented here were also provided by the Johns Hopkins University under federal contract number (N01-HV-48195) from NHLBI. The study participants derive from the National Health and Nutrition Examination Surveys (NHANES), and these studies are supported by the Centers for Disease Control and Prevention. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

The Multiethnic Cohort study (MEC) characterization of epidemiological architecture is funded through the NHGRI PAGE program (U01HG004802). The MEC study is funded through the National Cancer Institute (R37CA54281, R01 CA63, P01CA33619, U01CA136792, and U01CA98758).

Funding support for the "Epidemiology of putative genetic variants: The Women's Health Initiative" study is provided through the NHGRI PAGE program (U01HG004790). The WHI program is funded by the National Heart, Lung, and Blood Institute; NIH; and U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf.

Funding support for the Genetic Epidemiology of Causal Variants Across the Life Course (CALiCo) program was provided through the NHGRI PAGE program (U01HG004803). The following studies contributed to this manuscript and are funded by the following agencies: The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022. The Coronary Artery Risk Development in Young Adults (CARDIA) study is supported by the following National Institutes of Health, National Heart, Lung and Blood Institute contracts: N01-HC-95095; N01-HC-48047; N01-HC-48048; N01-HC-48049; N01-HC-48050; N01-HC-45134; N01-HC-05187; and N01-HC-45205. The Cardiovascular Health Study (CHS) is supported by contracts N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01-HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, grants U01HL080295 and R01 HL087652 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke. CHS GWAS DNA handling

and genotyping was supported in part by National Center for Research Resources grant M01-RR00425 to the Cedars-Sinai General Clinical Research Center Genotyping core and National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The Strong Heart Study (SHS) is supported by NHLBI grants U01 HL65520, U01 HL41642, U01 HL41652, U01 HL41654, and U01 HL65521. The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Indian Health Service.

Assistance with phenotype harmonization, SNP selection and annotation, data cleaning, data management, integration and dissemination, and general study coordination was provided by the PAGE Coordinating Center (U01HG004801-01). The National Institutes of Mental Health also contributes to the support for the Coordinating Center.

The PAGE consortium thanks the staff and participants of all PAGE studies for their important contributions.

EAGLE would like to thank Dr. Geraldine McQuillan and Jody McLean for their help in accessing the Genetic NHANES data. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for this work. The EAGLE/NHANES DNA samples are stored and plated by the Vanderbilt DNA Resources Core. Genotyping was performed by Ping Mayo, Melissa Allen, and Dr. Nathalie Schnetz-Boutaud in the laboratory of Dr. Jonathan Haines and Hailing Jin and Nila Gillani under the direction of Dr. Holli Dilks in the Vanderbilt DNA Resources Core.

CHAPTER V

ENVIRONMENTAL MODIFIERS OF GWAS-IDENTIFIED LIPID-ASSOCIATED VARIANTS

Introduction

The importance of both genetics and environment in shaping an individual's lipid profile is intuitively obvious. However, the search for gene-environment interactions that influence levels of HDL-C, LDL-C, and triglycerides has only been relatively recent. One driving force for expanding beyond the standard single-SNP models is the observation that single-SNP main effects do not account for the majority of the heritability attributed to additive genetics for most complex human traits (Manolio et al., 2009). For the lipid traits, heritability estimates are as high as 80% (O'Connell et al., 1988; Heller et al., 1993; Snieder et al., 1999), yet the largest and most comprehensive lipid meta-analysis to date was only able to explain about 25-30% of the genetic variance (Teslovich et al., 2010). The identification of gene-environment interactions may help find a proportion of this "missing heritability".

Within a statistical framework, a gene-environment interaction describes the effect of a genotype and an environmental factor that deviates from their additive effects. Within a biological framework, the environment (or its by-product) modifies the function or amount of a gene product (Hunter, 2005). The latter approach to identify gene-environment interactions is difficult in outbred populations such as humans given that both genetic background and environmental exposures vary within and across

populations. Model organisms are more suited to identify biological interactions, but it is difficult to automate these studies, and the findings of these experiments may not generalize to humans (Ober and Vercelli, 2011). In contrast, methods to identify statistical interactions can be automated, making them an attractive option for detecting gene-environment interactions important for complex human traits (Hunter, 2005).

A number of candidate environmental factors affect lipoprotein phenotypes, including smoking and diet. Cigarette smoking has been associated in many studies with decreased HDL-C and increased LDL-C and triglycerides (Craig et al., 1989; Chelland et al., 2008). One proposed mechanism to explain the link between smoking and changes in lipid and lipoprotein concentrations is that nicotine stimulates the release of adrenaline, leading to increased concentrations of free fatty acids which can then stimulate hepatic secretion of very low density lipoprotein (VLDL) and hence triglycerides (Brischetto et al., 1983). In addition to its unfavorable alterations of the lipid profile, smoking is a strong, independent predictor of coronary heart disease (CHD) (and is a major public health concern (Third Report of the National Cholesterol Education Program (NCEP) Expert Panel, 2002).

Vitamin E and vitamin A are fat-soluble micronutrients. Antioxidant vitamin E may play a role in the prevention of atherosclerosis, through inhibition of oxidation of LDL. While one randomized control trial demonstrated an inverse association between vitamin E intake and relative risk of coronary artery disease (Stephens et al., 1996), others were unable to replicate this protective effect, as reviewed in Nicolosi et al (Nicolosi et al., 2001). Discrepancies between studies may be due to the fact that vitamin E can also function as a prooxidant (Stocker, 1999). The antioxidant and anti-atherogenic properties of vitamin A are less studied, although, it is known that high

doses of vitamin A in the form of isotretinoin (better known as the acne drug Accutane) increase triglycerides and cholesterol levels and lower HDL-C levels (Bershad et al., 1985; Marsden, 1986; Murray et al., 1983; Zech et al., 1983).

Despite evidence that genetic variants and environmental factors are independently associated with lipid traits, relatively few studies have been published investigating the interaction between the two (Hagberg et al., 2000; Weinberg, 2002; Bernstein et al., 2002; Lai et al., 2006; Corella et al., 2001b; Corella et al., 2001a). And, to our knowledge, no studies have been published explicitly testing for interactions between lipid-associated SNPs and smoking, vitamin E, and vitamin A. We present here an investigation of the effects of 23 lipid-associated SNPs in the context of smoking and dietary intake of vitamins A and E using data from the National Health and Nutrition Examination Surveys (NHANES). Analysis of ~15,000 participants from this diverse population-based survey reveals nine significant interactions between lipid-associated SNPs and dietary intake of vitamins A and E, along with a SNP-smoking interaction that trended towards significance. These significant interactions explain 0.35-0.39%, 0.67-1.28%, and 0.36-0.80% of the variability in HDL-C, LDL-C, and triglyceride levels, respectively. Overall, these data provide the first steps in finding the "missing heritability" for lipid traits by accounting for the environment.

Methods

Study population

Study samples were drawn from three National Health and Nutrition Examination Surveys (NHANES III, NHANES 1999-200, and NHANES 2001-2002).

Participant ascertainment and data collection for NHANES has been outlined in previous chapters (Chapters III and IV) and, therefore, will not be discussed here. Only fasting adults (age ≥ 18 years) were included in this analysis. Race/ethnicity was self-described.

Laboratory measurements

Cotinine levels in the blood were measured using a highly sensitive method developed by the National Center for Environmental Health (Bernert, Jr. et al., 1997) using high-performance liquid chromatography and atmospheric-pressure chemical ionization tandem mass spectrometry. Serum HDL-C, triglycerides, and total cholesterol were measured using standard enzymatic methods. LDL-C was calculated using the Friedewald equation, with missing values assigned for samples with triglyceride levels greater than 400 mg/dl. Serum levels of vitamin E and vitamin A were measured with isocratic high-performance liquid chromatography (Center for Disease Control and Prevention, 1996; Centers for Disease Control and Prevention (CDC), 2002).

Determination of smoking status

Smoking status was determined using measured serum cotinine levels. Current smoking was defined has having cotinine levels ≥15ng/ml, a cutoff used by previous studies and shown to give approximately the same overall estimates of self-reported smoking behavior (Caraballo et al., 2001; Jarvis et al., 1987; Patrick et al., 1994; Pirkle et al., 1996). In a study NHANES III participants, 92.5% of self-reported smokers had cotinine levels greater than 15.0 ng/ml and 98.6% of self-reported non-smokers had

cotinine levels less than 15.0 ng/ml (Caraballo et al., 2001). The value of using cotinine over self-reported smoking status is it can be used as a marker of both active and passive smoking. While nicotine is the best marker of tobacco exposure, it has a short half-life (2-3 hours) in the body. Cotinine is the major metabolite of nicotine and has a much longer half-life (18-20 hours) (Caraballo et al., 2001) and, therefore, is widely used as a marker for tobacco smoke exposure.

SNP selection and genotyping

A total of 23 SNPs were considered in this analysis (Appendix S). All SNPs were previously associated with HDL-C, LDL-C, and/or triglycerides in published (as of 2008) candidate gene and genome-wide association studies and were subsequently analyzed for single-SNP associations with lipid levels in a large meta-analysis by the Population Architecture using Genomics and Epidemiology (PAGE) study (refer to Chapter IV for details). The 23 SNPs tested for gene-environment interactions were either accessed from existing data in the Genetic NHANES database (Keebler et al., 2009) or were directly genotyped by the Epidemiological Architecture of Genes Linked to Environment (EAGLE), one of the four large population-based studies of the PAGE network, using Sequenom or Illumina BeadXpress. Genotyping was performed in the Vanderbilt DNA Resources Core. In addition to genotyping experimental NHANES samples, we genotyped blind duplicates provided by CDC and HapMap controls (n=360). All EAGLE SNPs considered here were genotyped in all three NHANES (NHANES III, NHANES 1999-2000, and NHANES 2001-2002), had minor allele frequencies >5% in all three racial/ethnic populations, passed CDC quality control metrics, and are available for secondary analyses through NCHS/CDC.

Statistical analysis

Regression modeling was used to investigate the effect of any interaction between lipid-associated variants and smoking on levels of HDL-C, LDL-C, and transformed TG. Gene-environment interactions were modeled using a multiplicative interaction term between the environmental variable and an additive-encoded SNP. All models were adjusted for the main effect of the SNP and the environmental variable, along with the potential confounding effects of age and sex. Smoking status was dichotomized into non-smokers (cotinine levels<15 ng/ml) and smokers (cotinine levels ≥ 15 ng/ml). Triglycerides and vitamin E levels were natural-log transformed due to a skewed, non-normal distribution. HDL-C, LDL-C, and vitamin A levels were left as continuous and untransformed. All statistical analyses were conducted unweighted and remotely in SAS v9.2 (SAS Institute, Cary, NC) using the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center in Hyattsville, MD.

Results

Population characteristics

Table 5.1 displays descriptive statistics for the key variables in this study, stratified by smoking status. Of the maximum total sample size (n=8,088), there were 2,079 (26%) current smokers (cotinine levels greater or equal to 15 ng/ml). Non-Hispanic blacks had the highest proportion of smokers (35%) compared to the other two racial/ethnic populations (26% and 18% in non-Hispanic whites and Mexican Americans, respectively). Mean age was similar between non-Hispanic blacks and Mexican Americans, regardless of smoking status (~42 yrs), while mean age was overall

higher in non-Hispanic whites, particularly among non-smokers (~54 yrs). There was also a higher prevalence of smoking in men compared to women across all three racial/ethnic populations.

Mean triglyceride levels were consistently and significantly ($p<0.01$) higher in smokers versus non-smokers across all racial/ethnic groups, after adjusting for the effects of age and sex. Mean LDL-C levels were similar among both smoking categories and across all three populations. Interestingly, only in non-Hispanic whites were mean HDL-C concentrations significantly higher in non-smokers compared to smokers.

Table 5.2 also presents descriptive statistics of dietary measures (serum vitamin levels) of NHANES participants, regardless of smoking status. Both vitamin A and vitamin E levels were significantly different among the three racial/ethnic groups ($p<0.001$, one-way ANOVA). Non-Hispanic whites have both higher mean vitamin A and vitamin E levels (60.6 ug/dl and 1,322 ug/dl, respectively) compared to non-Hispanic blacks (53.1 ug/dl and 1,002 ug/dl) and Mexican Americans (52.8 ug/dl and 1,135 ug/dl). Non-Hispanic blacks and Mexican Americans have similar mean vitamin A levels, although vitamin E levels are higher in Mexican Americans.

Table 5.1. NHANES participant characteristics, stratified by smoking status. Values are represented as mean ± sd unless otherwise indicated. Non-smokers were defined as having measured serum cotinine levels <15 ng/ml; smokers were defined as having serum cotinine levels ≥15 ng/ml. *Represents a statistically significant difference ($p<0.01$) in the lipid trait between non-smokers and smokers in that specific racial/ethnic group, as determined using linear regression, adjusted for age and sex.

Trait	**Non-Hispanic Whites**		**Non-Hispanic Blacks**		**Mexican Americans**	
	Non-Smokers	Smokers	Non-Smokers	Smokers	Non-Smokers	Smokers
N	2,874	992	1,227	656	1,908	431
Age (years)	54.2 ± 20	45.2 ± 17	42.3 ± 18	42.4 ± 14	42.9 ± 18	42.2 ± 16
Female (%)	57	43	60	46	53	35
Cotinine (ng/ml)	0.36 ± 1	225.7 ± 147	0.73 ± 2	255.5 ± 171	0.51 ± 2	134.7 ± 105
HDL-C (mg/dl)	52.3 ± 16*	48.2 ± 15*	53.9 ± 16	54.5 ± 18	48.5 ± 13	47.4 ± 15
LDL-C (mg/dl)	126.6 ± 35	127.2 ± 37	122.7 ± 37	120.4 ± 40	120.9 ± 33	121.6 ± 33
Triglycerides (mg/dl)	145.2 ± 90*	150.1 ± 96*	103.8 ± 70*	113.0 ± 77*	153.8 ± 100*	167.8 ± 120*

Table 5.2. NHANES participant characteristics, including serum vitamin levels. Values are represented as mean ± sd unless otherwise indicated.

Trait	Non-Hispanic Whites	Non-Hispanic Blacks	Mexican Americans
N	2,435	1,407	1,734
Age (years)	51.9 ± 20	42.5 ± 17	42.8 ± 18
Female (%)	54	56	51
Vitamin A (ug/dl)	60.6 ± 16	53.1 ± 17	52.8 ± 15
Vitamin E(ug/dl)	1,322 ± 615	1,002 ± 379	1,135 ± 459

It is important to note that vitamins A and E are highly correlated with the majority of lipid levels in all three NHANES populations. More specifically, vitamin A is associated with all three lipid traits in the majority of participants. For triglycerides, the amount of variance explained (R^2) by vitamin A was as high as 14% in non-Hispanic whites. R^2 was smaller for the other two lipid traits (max R^2<5% between LDL-C and vitamin A in Mexican Americans; Table 5.3) although it was still larger than the average amount of variance explained by single common genetic variants (~3%). Vitamin E is also very strongly correlated with LDL-C and triglyceride levels ($p<4.05 \times 10^{-45}$) across all racial/ethnic groups. Furthermore, vitamin E levels explain 17-24% of the variance in LDL-C levels and 25-40% of the variance in triglyceride levels (Table 5.3).

Table 5.3. Associations between lipid traits and vitamins A and E. The association of lipid traits and vitamin levels were performed using linear regression, adjusted for age and sex. Both triglycerides and vitamin E levels were natural-log transformed. Measures of variance explained (R^2) are also provided for each association based on unadjusted regressions. Significant associations ($p<0.01$) are in bold.

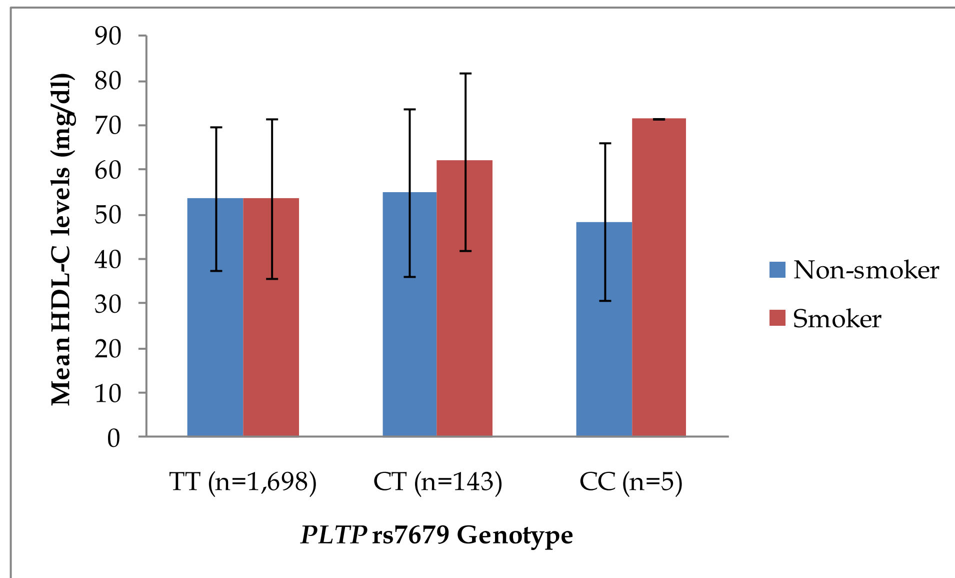
Vitamin A									
Lipid Trait	**Non-Hispanic Whites**			**Non-Hispanic Blacks**			**Mexican American**		
	Beta (SE)	p-value	R^2	Beta (SE)	p-value	R^2	Beta (SE)	p-value	R^2
HDL-C	0.05 (0.02)	**2.40E-03**	<0.01	0.08 (0.03)	**4.37E-03**	0.01	0.06 (0.02)	**9.28E -03**	<0.01
LDL-C	0.24 (0.24)	**2.88E-05**	0.02	0.17 (0.08)	0.03	0.02	0.38 (0.07)	**1.58E-08**	0.05
TG	0.01 (0.001)	**5.50E-56**	0.14	0.01 (0.001)	**2.14E-27**	0.11	0.01 (0.001)	**1.26E-30**	0.12
Vitamin E									
Lipid Trait	**Non-Hispanic Whites**			**Non-Hispanic Blacks**			**Mexican American**		
	Beta (SE)	p-value	R^2	Beta (SE)	p-value	R^2	Beta (SE)	p-value	R^2
HDL-C	1.25 (0.82)	0.13	<0.01	0.65 (1.53)	0.67	<0.01	-2.18 (0.97)	0.02	<0.01
LDL-C	36.21 (2.29)	**1.03E-52**	0.17	57.87 (3.88)	**4.05E-45**	0.24	46.91 (2.83)	**2.04E-55**	0.23
TG	0.68 (0.03)	**4.95E-116**	0.26	0.78 (0.04)	**1.91E-68**	0.25	1.01 (0.03)	**8.23E-151**	0.40

SNPxSmoking interactions

No gene-environment interaction was significantly associated with HDL-C, LDL-C, or triglyceride levels at the bonferroni corrected p-value of 0.002 ($0.05/23 = 2.1x10^{-3}$) (Appendices T, U, and V). However, one interaction term (*PLTP* rs7679xsmoking) was associated with increased HDL-C levels (β=8.42) at p=0.003 (Appendix T) in non-Hispanic blacks. Interestingly, this interaction term was not associated in the other two populations (p=0.529 and p=0.532 in non-Hispanic whites and Mexican Americans, respectively). Figure 5.1 displays the relationship between rs7679 genotype, smoking status, and mean HDL-C levels in non-Hispanic blacks. Smoking and non-smoking participants homozygous for the major allele (T) had very similar mean HDL-C levels. However, heterozygous smokers had significantly higher mean HDL-C levels (p=0.045, t-test) compared to non-smokers. While, the sample size for CC homozygous participants was very small (n=5), the trend continued with smokers having higher mean HDL-C levels.

Neither the main effect of *PLTP* rs7679 nor the main effect of smoking status was significant, although their directions of effect were positive (β=0.53 for both; Appendix T), consistent with that of the interaction term (β=8.42). The full model (including the interaction term rs7679xsmoking, along with age, sex, and the main effects of rs7679 and smoking) explained 3.84% of the variance in HDL-C levels in non-Hispanic blacks. The variance explained by the SNP alone, after adjusting for age and sex, was 3.29%, suggesting the interaction term does not appreciably add to the variance of HDL-C levels observed in this population.

Figure 5.1. *PLTP*xSmoking interaction effect on HDL-C levels in non-Hispanic blacks. Mean HDL-C levels are displayed for each rs7679 genotype (TT, CT, or CC), stratified by smoking status. Error bars represent standard deviations.



Eight SNPxsmoking interactions associated with one of three lipid traits at p<0.05 (Appendices T-V). The interactions between two SNPs, *FADS*1 rs174547 in non-Hispanic blacks and *MAFB* rs6102059 in Mexican Americans, were associated with HDL-C levels at p=0.041 and p=0.014, respectively. SNPxsmoking interactions including *MVK/MMAB* rs2338104 and *ABCA1* rs4149268 were associated with LDL-C levels at p=0.036 and p=0.028, respectively, in non-Hispanic whites. Lastly, two interactions in non-Hispanic blacks (*ANGPTL4* rs2967605xsmoking and *APOB* rs693xsmoking) and two in Mexican Americans (*MAFB* rs6102059xsmoking and *PLTP* rs7679xsmoking) were associated with transformed triglycerides at p<0.042.

SNPxVitamin A and E interactions

We tested for gene-environment interaction effects between our 23 lipid-associated variants and vitamins A and E on HDL-C, LDL-C, and triglyceride levels. A total of nine gene-environment interactions were statistically significant at $p<2.16 \times 10^{-3}$ and are summarized in Table 5.4. The association between LDL-C and *APOB* rs693xvitamin E in Mexican Americans was the most significant at $p=8.94 \times 10^{-7}$. This same interaction was significant in non-Hispanic whites ($p=2.67 \times 10^{-4}$) but not in non-Hispanic blacks (p=0.11, Appendix AA). Additionally, other interactions with this *APOB* variant (rs693xvitamin A and rs693xvitamin E) were significantly associated with triglyceride levels among non-Hispanic whites at $p=2.16 \times 10^{-3}$ and 4.65×10^{-5}, respectively.

Interactions between *ANGPTL3* rs1748195 and both vitamin A and E were associated with HDL-C levels in non-Hispanic whites ($p=1.16 \times 10^{-3}$ and $p=2.06 \times 10^{-3}$). The *ANGPTL3* rs1748195xvitamin A interaction trended towards significance in non-Hispanic blacks (p=0.01) but was not associated with HDL-C in Mexican Americans (p=0.64, Appendix W). Similarly, the rs1748195xvitamin E interaction was not associated with HDL-C in the other two populations.

Two interactions with a variant in *PCSK9* are also listed in Table 5.4. The *PCSK9* rs11206510xvitamin A interaction was associated with LDL-C in Mexican Americans at $p=7.65 \times 10^{-5}$. In addition, the *PCSK9* rs11206510xvitamin E interaction was associated with transformed triglycerides in non-Hispanic whites at $p=1.27 \times 10^{-3}$. Lastly, the only significant gene-environment interaction observed in non-Hispanic blacks was between the *APOA1/C3/A4/A5* cluster variant rs313550 and vitamin E, which was associated with triglyceride levels at $p=2.45 \times 10^{-4}$.

Table 5.4. Significant SNPxenvironment interactions in NHANES. Associations with significant interaction terms ($p<2.17E-03$, bonferroni corrected p-value for 23 SNPs) are listed. Betas, standard errors (SE), and p-values for main effects of the SNP and the environment are represented, along with the amount of trait variance explained (R^2) by interaction term.

Interaction	Associated Lipid Trait	Population	SNP Main Effect		Environment Main Effect		SNPxEnvironment Interaction Effect		
			Beta (SE)	P-value	Beta (SE)	P-value	Beta (SE)	P-value	R^2 (%)
rs693xVitA	TG	Non-Hispanic Whites	-0.16 (0.06)	6.11E-03	0.01 (0.001)	1.01E-22	0.003 (0.001)	**2.16E-03**	0.39
rs693xVitE	LDL-C	Non-Hispanic Whites	-74.86 (21.54)	5.22E-04	31.86 (2.76)	1.39E-29	11.11 (3.04)	**2.67E-04**	0.67
rs693xVitE	LDL-C	Mexican Americans	-155.52 (31.82)	1.17E-06	38.98 (3.25)	2.51E-31	22.71 (4.60)	**8.94E-07**	1.28
rs693xVitE	TG	Non-Hispanic Whites	-0.99 (0.25)	8.59E-09	0.60 (0.03)	3.48E-62	0.14 (0.04)	**4.65E-05**	0.60
rs1748195xVitA	HDL-C	Non-Hispanic Whites	-5.15 (1.67)	2.07E-03	-0.05 (0.04)	0.18	0.09 (0.03)	**1.16E-03**	0.39
rs1748195xVitE	HDL-C	Non-Hispanic Whites	-23.13 (7.55)	2.22E-03	-3.12 (0.65)	0.06	3.28 (1.06)	**2.06E-03**	0.35
rs11206510xVitA	LDL-C	Mexican Americans	-30.58 (8.08)	1.63E-04	0.25 (0.07)	9.22E-04	0.58 (0.15)	**7.65E-05**	1.26
rs11206510xVitE	TG	Non-Hispanic Whites	1.03 (0.33)	1.63E-03	0.74 (0.03)	1.89E-49	-1.15 (0.05)	**1.27E-03**	0.36
rs3135506xVitE	TG	Non-Hispanic Blacks	-3.02 (0.85)	4.16E-04	0.74 (0.04)	1.14E-56	0.46 (0.12)	**2.45E-04**	0.80

The nine significant interaction models individually explained 0.35-1.28% of the variation in one of the lipid traits. Interactions rs693xvitamin E and rs11206510xvitamin A had the greatest R^2 values and contributed to 1.28% and 1.26%, respectively, of the variation in LDL-C among Mexican Americans. The seven other interaction terms had R^2 values <1%.

Discussion

Considerable attention has been paid to the discovery of common variants important in lipid metabolism. Despite many successes, lipid-associated variants discovered through GWAS do not account for the majority of heritability estimated for these traits. Epidemiological studies have long indicated that certain environmental factors are capable of shaping lipid distributions in the population. However, environmental modifiers of known genotype-phenotype associations are just recently emerging in the literature. In this study we tested for interactions between 23 GWAS-identified lipid-associated variants and three environmental factors: smoking status, and vitamin E and A levels across three racial/ethnic populations in NHANES. We discovered nine significant gene-environment interactions, along with several interesting interactions that trended towards significance. Interpretation of our results is detailed below, divided by the two categories of environmental variables, smoking and dietary vitamins.

SNPxSmoking interactions

This study provided evidence for a potentially interesting lipid gene-smoking interaction on between *PLTP* rs7679xsmoking on HDL-C levels in non-Hispanic blacks. We observed that among nonsmokers, rs7679 had little influence on mean HDL-C, although participants homozygous for the minor allele (CC) had the lowest levels compared with participants homozygous for the major allele or heterozygous. Smokers with at least one copy of the minor allele (C) had, on average, higher mean HDL-C compared to smokers homozygous for the major allele. This finding suggests that the effects of rs7679 genotype on HDL-C metabolism are more important in smokers than non-smokers.

The variant rs7679 lies in the 3′ untranslated region of *PCIF1*, approximately 50 kb downstream of *PLTP*, and has been significantly associated with *PLTP* expression in human liver, along with increased HDL-C (Kathiresan et al., 2009). *PLTP* encodes plasma phospholipid transfer protein (PLTP), which plays a major role in the transfer of phospholipids between lipoproteins and in the modulation of HDL-C particle size and composition (Huuskonen et al., 2001). PLTP knock-out mice show a 60-70% reduction in HDL-C (Jiang et al., 1999), and reduced PLTP activity has been reported in patients with Tangier disease, a genetic disorder characterized by significantly reduced levels of HDL-C (von Eckardstein et al., 1998).

PLTP transfer activity may also be affected by smoking. In one study by Dullaart et al, normolipidemic cigarette-smoking men had increased PLTP activity compared to non-smoking controls (Dullaart et al., 1994). However, in another study, PLTP activity levels were similar among fasting smokers and controls (Mero et al., 1998).

In the lipids PAGE study (discussed in Chapter IV), the single-SNP association between rs7679 and HDL-C was not significant in African Americans (p=0.992). In contrast, in the GWAS by Kathiresan et al (Kathiresan et al., 2009) and the lipids PAGE study, the T allele of rs7679 was significantly associated with increased HDL-C in European-descent populations. The gene-environment interaction described here for non-Hispanic blacks may be spurious or it may help to explain the lack of generalizability of rs7679 across racial/ethnic groups.

It is interesting to note that the most significant SNPxsmoking interactions were found in non-Hispanic blacks, despite their smaller sample (n=1,883) size compared to non-Hispanic whites (n=3,866) and Mexican Americans (n=2,339). This racial/ethnic specificity could possibly be due to a difference in nicotine metabolism between blacks and whites. Black smokers have higher levels of serum cotinine compared to whites, even after adjusting for number and yield of cigarettes (Perez-Stable et al., 1998; Wagenknecht et al., 1990). This difference can be explained by both slower cotinine clearance and higher intake of nicotine per cigarette in blacks (Perez-Stable et al., 1998). Our results demonstrate the need for further gene-environment interaction studies in diverse populations to better understand how these differences may impact lipid levels.

SNPxVitamin A and E interactions

In this study we have identified three novel SNPxvitamin A and six novel SNPxvitamin E interactions. A majority of the significant interactions were associated with triglycerides (4/9) and were among non-Hispanic whites (6/9). Our most significant finding (*APOB* rs693xvitamin E), however, explained less than 1.3% of the variance in LDL-C among Mexican Americans, a trait that is up to 80% heritable. In

comparison, the effect of age and sex together accounted for 5.9% of the variance in LDL-C among Mexican Americans.

All of the genes implicated here play key roles in lipid metabolism. The gene products of *APOB*, apoB-48 and apo-100, are the main apoliprotein of chylomicrons and LDL particles, respectively. *ANGPTL3* encodes a protein which can suppress lipoprotein lipase (LPL) activity, leading to increases in plasma triglycerides and HDL-C. *PCSK9* encodes protein convertase subtilisin kexin 9, a protein that binds the LDL receptor and induces its degradation. Lastly, the *APOA1/C3/A4/A5* gene cluster lies within a 17kb region on chromosome 11. Proteins made by this gene cluster are major constituents of very low density lipoprotein (VLDL) and/or HDL, act to inhibit LPL activity, and influence dietary fat absorption and chylomicron synthesis (Delgado-Lista et al., 2010).

Both vitamin E and A are incorporated into lipoproteins and are delivered to peripheral tissues. Additionally, both are found exclusively in plasma lipoproteins (VLDL, LDL, and HDL) (Borel et al., 2007). The interdependence of these vitamins and lipids (as demonstrated in Table 5.3) suggests that the interactions described in this study may be either just reflective of the strong correlation between vitamins and lipids or point to biological relevance. In support of the latter interpretation, micronutrients such as vitamin A and E have previously been implicated in affecting the gene expression of import lipid-metabolizing genes (Mooradian et al., 2006b; Mooradian et al., 2006a; Hagberg et al., 2000; Oliveros et al., 2007; Gatica et al., 2006). For example, Mooradian et al demonstrated that high concentrations of vitamin E were associated with significant decreases in apoA-I expression (which is sensitive to the oxidative state

of the cell) in hepatic HepG2 cells by reducing apoA-I promoter activity (Mooradian et al., 2006a).

It has been argued that gene-environment heterogeneity may be, in part, to blame for the lack of replication among GWAS studies and among different ancestral populations (Lasky-Su et al., 2008; Ober and Vercelli, 2011). In the single-SNP PAGE meta-analysis detailed in Chapter IV, *APOB* rs693 was strongly associated in European Americans ($p=3.38 \times 10^{-21}$), marginally associated with LDL-C in African Americans ($p=0.02$), but not associated in Mexican Americans/Hispanics ($p=0.18$; Chapter IV, Table 4.3). However, in this analysis, which represents a subset of the PAGE study sample, the main effect of rs693 was significantly associated in Mexican Americans ($p=1.17 \times 10^{-6}$, Table 5.4) after adjusting for the interaction with vitamin E. Accounting for environmental modifiers in genetic studies of lipid levels may not only uncover new biology, it may also improve the generalizabilty of findings from genome-wide association studies.

Strengths and limitations

In interpreting our findings, we should consider several aspects. First, NHANES is a cross-sectional study and, therefore, we are unable to determine the temporal sequence of our results. Second, the issue of sample size and the 'curse of dimesionality' (Bellman, 1961) is relevant to this study. As the number of factors under study increases (as with the addition of interaction terms), so do the number of strata. With a set sample size, increasing the number of terms in the model quickly increases the degrees of freedom and reduces the per-stratum sample size, thus decreasing statistical power. For this reason, even with relatively large sample sizes in NHANES, we had to restrict our

analysis to SNPs with minor allele frequencies greater than 5%. To better study less-common variants, collaborative studies and/or other non-regression based approaches (such as multifactor dimensionality reduction) (Ritchie et al., 2001) may be appropriate, although they are not without their own limitations. Lastly, other potential confounding environmental factors, such as physical activity and alcohol consumption, were not included in the analysis because they are difficult to measure quantitatively and without recall error.

A major strength of the study is that NHANES systematically collects environmental exposures in a diverse population. It is important to keep in mind that, beyond sample size, the power to detect gene-environment interactions is influenced by the accuracy of the measurement of the outcome and the environmental exposure (Wong et al., 2003). In general, environmental variables are notoriously difficult to collect and quantify. Most environmental factors are assessed by questionnaire, which can lead to certain biases, including under-reporting of risky behaviors. Therefore, biomarkers as quantitative measures of the environmental exposures are preferred. For example, blood cotinine levels can be used as biomarkers of current smoking status. While NHANES does contain data on self-reported tobacco use, serum cotinine is considered a better marker of smoking status because, in certain populations, self-reports underestimate the actual smoking prevalence (Caraballo et al., 2001). Measures of dietary intake may be assessed by collection of daily food diaries or 24-hour dietary recalls. From these recall data, calculation of fat, vitamin, and mineral content is available in NHANES but these estimates are subject to poor recall. However, serum vitamin A and E levels are easily measured from a blood draw and may be used as a measure of dietary compliance.

Conclusions

The results presented here highlight the fact that effect sizes of gene-environment interactions tend to be small and large sample sizes are needed to detect them. Nevertheless, understanding the mechanism of the interaction between these lipid-associated variants and environmental factors, such as smoking and dietary vitamin E and A intake, is imperative to determining the etiology of a poor lipid profile and could, therefore, have implications in clinical care.

Acknowledgements

We would like to thank Dr. Geraldine McQuillan and Jody McLean for their help in accessing the Genetic NHANES data. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for this work. The EAGLE/NHANES DNA samples are stored and plated by the Vanderbilt DNA Resources Core. Genotyping was performed by Ping Mayo, Melissa Allen, and Dr. Nathalie Schnetz-Boutaud in the laboratory of Dr. Jonathan Haines and Hailing Jin and Nila Gillani under the direction of Dr. Holli Dilks in the Vanderbilt DNA Resources Core.

CHAPTER VI

CONCLUSION

Summary

Lipid and lipoproteins are products of a complex molecular network that is modulated by a number of genetic loci and environmental factors. While a substantial number of genes have been implicated in lipid metabolism, the proportion of trait variation they collectively explain is still lacking. This "missing heritability" problem is not unique to lipid traits. Indeed, nearly every common, complex disease or phenotype studied to date share this quandary. Both the source of this missing heritability and the appropriate strategy to uncover it has been hotly debated.

One step towards explaining a proportion of the missing heritability of lipid traits may be the identification of novel variants. In this work, I identify novel variants associated with lipid and lipoprotein levels using two different study design methods: candidate gene studies and GWAS. For example, in Chapter II, 19 variants in the candidate gene *LPA* were tested for associations with Lp(a) levels across the three different NHANES subpopulations (non-Hispanic whites, non-Hispanic blacks, and Mexican Americans). At a significance threshold of $p<0.0001$, 15 of the 19 SNPs tested were strongly associated with Lp(a) levels in at least one subpopulation. While these findings replicated previously known associations, such as rs6919246 and Lp(a) levels, the majority of significant associations were novel. Overall, the additive effects of these

associated alleles explained up to 11% of the variance observed for Lp(a) levels in the different racial/ethnic populations.

GWAS are the most commonly used agnostic approach to identify novel susceptibility genes for common, complex traits, including lipid levels. In Chapter III, I utilized this approach to discover new loci that affect HDL-C, LDL-C, and triglyceride concentrations in more than 400 children under 18 years of age. A total of 52 promising associations ($p<1x10^{-5}$) were subsequently examined in 1,040 additional youths and 3,508 adults from NHANES III. Three genotype-phenotype associations replicated in NHANES III youths and three associated in NHANES III adults at $p<0.05$; however, no single association was significant in both youths and adults. The most significant association ($p=0.009$) in NHANES III youths was between LDL-C and intronic *SGSM2* rs2429917 among participants of African-descent. Given the known age-dependency of lipid levels, I also tested for gene-age interactions in NHANES III participants and identified a significant ($p=0.024$) age-dependent association between *SGSM2* rs2429917 and LDL-C. This finding illustrates the utility of using children to discover novel variants associated with complex phenotypes and the importance of considering age-dependent genetic effects in association studies of lipid levels.

It has also been proposed that gene-environment interactions, which are usually not explicitly modeled in GWAS, could account for a significant proportion of the missing heritability (Ober and Vercelli, 2011). Therefore, in Chapter V, I tested for environmental modifiers of associations between 23 GWAS-identified variants from the literature (a subset of those explored in Chapter IV) and HDL-C, LDL-C, and triglycerides. The environmental factors I focused on were smoking status and serum levels of two nutritional antioxidants, vitamin E and vitamin A. While no SNPxsmoking

interaction was significantly associated with any lipid levels, nine SNPxvitamin interactions were associated at $p<2.17 \times 10^{-3}$. However, these interactions explained only 0.35 - 1.28% of the variance of their respective lipid trait.

Another important issue discussed throughout this work is the importance of sampling different population groups in genotype-phenotype association studies. Chapter I demonstrated how mean lipid levels may differ among different ancestral populations. These differences are likely due to both differences in allele frequencies and environmental exposures among populations. The studies presented here explore both of these factors. In the *LPA* study presented in Chapter II, only six of the 19 SNPs tested were significant in at least two of the three populations studied and none were significant in all three populations. The lack of generalization across all subpopulations suggests that specific *LPA* variants may be contributing to the observed Lp(a) between-population variance.

Lipid-associated genetic variants are being discovered in GWAS in samples of European descent, but insufficient data exist for other populations. Therefore, there is a strong need to characterize the effect of these GWA-identified variants in more diverse cohorts. In Chapter IV, I selected over forty genetic loci previously associated with lipid levels and tested for replication in a large European American cohort. I also investigated if the effect of these variants generalizes to non-European descent populations, including African Americans, American Indians, and Mexican Americans/Hispanics. A majority (92%) of these GWAS-identified associations replicated at $p<0.05$ in our European American cohort. Based on significance and consistent direction of effect, 16 associations generalized across all three non-European descent populations. The ability of only 16 of 42 associations to generalize across racial/ethnic populations indicates that

some of these GWAS-identified variants may not be functional and are more likely to be in linkage disequilibrium with the functional variant(s).

Inconsistencies in disease associations in different populations, like those mentioned in Chapter IV, may be due to, in part, different environmental exposures that modify the effect of a genetic variant or different frequencies of genetic variants that modify the effect of an environmental exposure. In other words, gene-environment interactions, when not accounted for, may mask detection of a genetic effect in some populations. I observed evidence of this when comparing the results of Chapter IV and Chapter V. For example, in the single-SNP PAGE meta-analysis detailed in Chapter IV, *APOB* rs693 was strongly associated in European Americans ($p=3.38x10^{-21}$) but not associated in Mexican Americans/Hispanics ($p=0.18$). However, in the gene-environment interaction study detailed in Chapter V, the main effect of rs693 was significantly associated in Mexican Americans ($p=1.17x10^{-6}$) after adjusting for the interaction with vitamin E levels.

This work explored associations with common genectic variants to uncover some of the missing heritability attributed to lipid and lipoprotein concentrations. Unfortunately, most common genetic variants (individually and in total) implicated here explain only a small fraction of the genetic variation. Futhermore, similiarly small effect sizes were observed in the investigation of gene-environment interactions, with the interaction explaining much less of the lipid trait distribution than the environment alone. However, having a small effect size does not necessarily mean that a genetic variant is of no interest. Genetic variants with small effects can still point invesitigators to important biological pathways and become targets for pharmacologic intervention, which may result in larger, clinically relevant effects.

Future Directions

Besides those utilized in this work, there are several additional approaches future researchers must employ if we are to fully illuminate the genetic architecture of lipid levels and other complex traits. While GWAS are the most commonly used agnostic approach to identify novel susceptibility genes for common disease, they have very limited potential to capture rare and low frequency variants (i.e. MAF<5%). Such variants, which may also not be detectable through traditional linkage studies due to low penetrance, could individually and collectively impact familial risk. While detection of rare and low frequency variants remains a challenge, recent technological advances in high-throughput sequencing should enable progress on this front. In fact, the goal of the 1000 Genomes Project is to uncover most of the genetic variants with frequencies of at least 1% using a combination of low-coverage whole-genome and high-coverage whole-exome sequencing (Durbin et al., 2010). Completion will, presumably, result in a detailed map of our genetic landscape and fill in the gaps in our knowledge about how genetic variation is related to disease.

Currently, our understanding of the mechanisms by which risk alleles contribute to disease lags behind the ferocious pace at which new loci are being discovered. It is also in our best interest to refine association signals identified by GWAS so that we may uncover the true "causal" variant. Due to strong linkage disequilibrium throughout most of the genome, there will often be several candidate variants that have equivalent evidence of association. One way to refine GWAS signals is to perform fine-mapping in populations of different ancestries. Using this approach, the PAGE consortium is presently genotyping thousands of African Americans on the MetaboChip, a high-

density custom array of about 200K SNPs that captures and fine maps a number of GWAS-identified variants (p values < 5.0 x 10-8 in European-descent samples) from phenotypes related to type 2 diabetes and cardiovascular diseases (such as BMI, lipids, blood pressure). Presumably, these data will allow PAGE investigators to determine the actual variant responsible for the GWAS signal, or at least narrow down the list of possible functional variants. Of course, to determine if a specific variant is truly causal, functional and animal studies are necessary and cannot be overlooked.

Progress in understanding the genes important in determining lipid and lipoprotein levels has accelerated rapidly in the last decade following the completion of the Human Genome Project. However, much of the potential impact of these advances has yet to be realized. The challenges of moving from associated variant to mechanism of action to changes in clinical practice are substantial. A coordinated effort among geneticists, statisticians, physicians, and basic biologists is necessary to improve our understanding of human disease and biology. Only then will the promises of GWAS, whole-genome sequencing, and the general field of human genetics be realized.

Appendix A. Associations between *LPA* SNPs and Lp(a) levels, weighted for selection and non-response biases. The association of *LPA* SNPs with log transformed Lp(a) levels is shown by a regression coefficient (beta, β) and 95% confidence interval (CI) for each SNP, adjusted for age and sex. Measures of variance explained (R^2) are provided for each SNP based on unadjusted regressions. Significant associations (P-value < 0.0001) are in bold.

SNPs	Non-Hispanic Whites n=2,397			Non-Hispanic Blacks n=1,711			Mexican Americans n=1,749		
	β (95% CI)	R^2	P-value	β (95% CI)	R^2	P-value	β (95% CI)	R^2	P-value
rs1321196	-0.17 (-0.28, -0.06)	0.0061	0.0043	0.19 (0.09, 0.29)	0.0171	0.0005	0.18 (0.07, 0.28)	0.0069	0.0018
rs1321195	0.10 (-0.01, 0.21)	0.0012	0.0651	0.49 (0.16, 0.81)	0.0108	0.0055	0.38 (0.16, 0.60)	0.0114	0.0015
rs1367211	0.02 (-0.09, 0.12)	0.0000	0.7427	-0.24 (-0.36, -0.12)	0.0271	0.0005	-0.19 (-0.30, -0.08)	0.0068	0.0016
rs1652507	-0.04 (-0.17, 0.10)	0.0002	0.6083	-0.44 (-0.62, -0.26)	0.0225	**3.44×10^{-5}**	-0.58 (-0.69, -0.47)	0.0977	**$< 2 \times 10^{-10}$**
rs6907156	0.03 (-0.82, 0.88)	0.0000	0.9404	0.14 (0.04, 0.24)	0.0047	0.0111	0.67 (0.34, 1.00)	0.0088	0.0003
rs6919346	0.51 (0.36, 0.66)	0.0418	**2.23×10^{-7}**	0.77 (0.62, 0.91)	0.0381	**$< 2 \times 10^{-10}$**	0.21 (-0.02, 0.44)	0.0040	0.0762
rs6926458	-0.29 (-0.42, -0.15)	0.0145	0.0002	-0.44 (-0.60, -0.28)	0.0325	**1.01×10^{-5}**	-0.11 (-0.24, 0.02)	0.0016	0.0860
rs7755463	-0.40 (-1.04, 0.24)	0.0007	0.2088	-0.33 (-0.41, -0.24)	0.0457	**4.59×10^{-8}**	-0.89 (-1.08, -0.71)	0.0251	**$< 7 \times 10^{-10}$**
rs7767084	0.01 (-0.08, 0.11)	0.0000	0.8006	-0.39 (-0.67, -0.11)	0.0086	0.0089	0.08 (-0.05, 0.21)	0.0010	0.2143
rs9364564	0.24 (0.08, 0.39)	0.0084	0.0044	0.33 (0.20, 0.46)	0.0161	**2.00×10^{-5}**	0.05 (-0.09, 0.20)	0.0003	0.4614

rs12212507	0.03 (-0.15, 0.22)	0.0001	0.6958	0.71 (0.13, 1.28)	0.0074	0.0179	0.55 (0.18, 0.92)	0.0030	0.0057
rs13192132	-0.18 (-0.28, -0.08)	0.0067	0.0013	-0.43 (-0.54, -0.32)	0.0408	**3.90×10^{-8}**	0.06 (-0.06, 0.18)	0.0007	0.3151
rs10945682	0.16 (0.05, 0.27)	0.0055	0.0048	-0.13 (-0.21, -0.05)	0.0079	0.0019	-0.22 (-0.33, -0.11)	0.0106	0.0005
rs12194138	-0.43 (-0.54, -0.32)	0.0255	**3.52×10^{-8}**	0.30 (-0.05, 0.64)	0.0045	0.0871	-0.49 (-0.69, -0.29)	0.0155	**4.53×10^{-5}**
rs7450261	-0.49 (-2.52, 1.54)	0.0002	0.6209	0.31 (0.15, 0.48)	0.0087	0.0007	-0.03 (-1.92, 1.87)	0.0000	0.97843
rs7450411	0.24 (0.09, 0.4)	0.0088	0.0033	0.37 (0.27, 0.47)	0.0308	**1.69×10^{-7}**	0.06 (-0.07, 0.20)	0.0005	0.3555
rs7765803	-0.08 (-0.2, 0.03)	0.0013	0.1436	0.12 (0.01, 0.23)	0.0064	0.0302	0.26 (0.15, 0.38)	0.0152	**9.09×10^{-5}**
rs41265936	-1.11 (-1.61, -0.62)	0.0020	0.0001	-0.18 (-0.33, -0.02)	0.0032	0.0280	-0.82 (-1.22, -0.41)	0.0043	0.0004
rs41271028	-0.39 (-0.92, 0.14)	0.0003	0.1421	-0.05 (-0.15, 0.06)	0.0005	0.3700	-0.58 (-0.95, -0.22)	0.0049	0.0031

Appendix B. Additive effects of *LPA* alleles associated with increased Lp(a) levels. The amount of variance explained (R^2) in transformed Lp(a) levels by the Weighted Genetic Risk Score (WGRS) is displayed, along with the median WGRS score, WGRS interquartile range (IQR), regression coefficient (beta, β) and 95% confidence interval (CI) for each association.

	Non-Hispanic Whites	**Non-Hispanic Blacks**	**Mexican Americans**
Total n	2269	1605	1665
No. SNPs used in WGRS	3	12	6
Median WGRS (IQR)	1.50 (0.45)	8.24 (1.01)	1.56 (0.53)
β (95% CI)	0.87 (0.75-0.99)	0.33 (0.28-0.38)	0.87 (0.75-0.99)
P-value	$<10^{-40}$	$<10^{-38}$	$<10^{-47}$
R^2	0.08	0.10	0.12

Appendix C. Differences in mean HDL-C, LDL-C, and triglycerides among St. Jude Children and NHANES III youths, stratified by race/ethnicity. P-values were calculated using a standard two-sample t-test with unequal variances.

Whites				
HDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
ST. JUDE Children	282	55.063	30.128	<0.001
NHANES III Youth	240	47.279	10.235	
LDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
ST. JUDE Children	282	84.512	26.926	0.024
NHANES III Youth	75	92.853	28.424	
Triglycerides (mg/dL)				
Study	N	Mean	Std. Dev	P-value
ST. JUDE Children	282	108.180	79.133	0.075
NHANES III Youth	241	97.614	55.708	

Blacks				
HDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
ST. JUDE Children	66	58.964	27.043	0.065
NHANES III Youth	436	52.619	12.496	
LDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
ST. JUDE Children	66	81.318	36.929	0.001
NHANES III Youth	153	98.464	26.965	
Triglycerides (mg/dL)				
Study	N	Mean	Std. Dev	P-value
ST. JUDE Children	66	88.288	47.399	0.181
NHANES III Youth	436	79.954	42.594	

Hispanics				
HDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
ST. JUDE Children	63	67.892	49.720	0.004
NHANES III Youth	360	49.144	10.809	
LDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
ST. JUDE Children	63	87.381	35.381	0.225
NHANES III Youth	134	93.373	23.723	
Triglycerides (mg/dL)				
Study	N	Mean	Std. Dev	P-value
ST. JUDE Children	63	121.687	101.468	0.089
NHANES III Youth	360	99.006	57.335	

Appendix D. Differences in mean HDL-C, LDL-C, and triglycerides among NHANES III adults and NHANES III youths, stratified by race/ethnicity. P-values were calculated using a standard two-sample t-test with unequal variances.

Whites				
HDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
NHANES III Adults	1372	50.281	15.154	<0.001
NHANES III Youth	240	47.279	10.235	
LDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
NHANES III Adults	996	129.374	37.508	<0.001
NHANES III Youth	75	92.853	28.424	
Triglycerides (mg/dL)				
Study	N	Mean	Std. Dev	P-value
NHANES III Adults	1378	145.324	111.390	<0.001
NHANES III Youth	241	97.614	55.708	

Blacks				
HDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
NHANES III Adults	1025	53.578	16.691	0.227
NHANES III Youth	436	52.619	12.496	
LDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
NHANES III Adults	669	122.976	40.285	<0.001
NHANES III Youth	153	98.464	26.965	
Triglycerides (mg/dL)				
Study	N	Mean	Std. Dev	P-value
NHANES III Adults	1029	107.622	82.014	<0.001
NHANES III Youth	436	79.954	42.594	

Hispanics				
HDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
NHANES III Adults	1078	47.393	13.119	0.012
NHANES III Youth	360	49.144	10.809	
LDL-C (mg/dL)				
Study	N	Mean	Std. Dev	P-value
NHANES III Adults	710	120.061	34.160	<0.001
NHANES III Youth	134	93.373	23.723	
Triglycerides (mg/dL)				
Study	N	Mean	Std. Dev	P-value
NHANES III Adults	1080	151.816	109.668	<0.001
NHANES III Youth	360	99.006	57.335	

Appendix E. St. Jude discovery GWAS associations selected for replication

SNP	Chr.	Nearby Gene(s)	Location	Associated Trait	Associated Population	Coded Allele	Beta (SE)	GWAS P-value
rs4742455	9	--	intergenic	HDL	Whites	A	0.18 (0.03)	5.18E-07
rs17026635	1	*USH2A*	intron	HDL	Whites	A	0.72 (0.14)	6.33E-07
rs11252020	10	--	intergenic	HDL	Whites	A	0.32 (0.07)	2.24E-06
rs8009160	14	--	intergenic	HDL	Whites	C	-0.17 (0.04)	2.82E-06
rs8012357	14	--	intergenic	HDL	Whites	A	-0.17 (0.04)	2.82E-06
rs1541164	14	--	intergenic	HDL	Whites	G	-0.17 (0.04)	3.31E-06
rs1953743	14	--	intergenic	HDL	Whites	C	-0.16 (0.04)	5.77E-06
rs16902507	8	*PVT1*	intron	HDL	Blacks	A	-0.36 (0.07)	1.4E-06
rs2162987	5	--	intergenic	HDL	Blacks	G	0.59 (0.12)	3.56E-06
rs12815715	12	*ACSM4*	downstream	HDL	Blacks	G	1.02 (0.20)	5.03E-06
rs7958130	12	--	intergenic	HDL	Blacks	A	0.31 (0.06)	5.94E-06
rs1328810	13	--	intergenic	HDL	Blacks	A	0.46 (0.09)	9.39E-06
rs1561193	14	--	intergenic	HDL	Hispanics	A	1.26 (0.24)	2.49E-06
rs10516115	5	--	intergenic	HDL	Hispanics	C	0.59 (0.12)	4.1E-06
rs16998203	20	*OTOR*	downstream	HDL	Hispanics	G	0.63 (0.13)	4.42E-06
rs1693669	10	*ATE1*	downstream	HDL	Hispanics	G	1.11 (0.22)	5.52E-06
rs10754306	1	--	intergenic	HDL	Hispanics	C	1.03 (0.21)	5.62E-06
rs7709495	5	--	intergenic	LDL	Whites	C	-0.14 (0.03)	2.2E-06
rs1911937	5	--	intergenic	LDL	Whites	G	-0.14 (0.03)	3.25E-06
rs7446139	5	--	intergenic	LDL	Whites	T	-0.13 (0.03)	4.79E-06
rs7710638	5	--	intergenic	LDL	Whites	G	-0.13 (0.03)	9.43E-06
rs8007972	14	*SYNE2*	intron	LDL	Blacks	C	-0.55 (0.11)	2.14E-06
rs6069670	20	*MC3R*	promoter	LDL	Blacks	T	-0.98 (0.19)	2.43E-06
rs7258420	19	*FLJ41856*	intron	LDL	Blacks	A	-0.57 (0.11)	2.55E-06
rs11542053	4	*RASGEF1B*	downstream	LDL	Blacks	A	-1.49 (0.29)	3.29E-06
rs17698672	4	--	intergenic	LDL	Blacks	A	-1.49 (0.29)	3.29E-06
rs2043283	5	--	intergenic	LDL	Blacks	C	-1.49 (0.29)	3.29E-06
rs11048606	12	*ITPR2*	intron	LDL	Blacks	T	-1.49 (0.29)	3.29E-06
rs17558056	4	*RASGEF1B*	intron	LDL	Blacks	G	-1.49 (0.29)	3.29E-06
rs12190789	6	--	intergenic	LDL	Blacks	T	-1.46 (0.29)	3.32E-06
rs4811011	20	--	intergenic	LDL	Blacks	A	-0.39 (0.08)	3.51E-06
rs16971165	18	--	intergenic	LDL	Blacks	T	-1.22 (0.24)	3.69E-06

rs6580132	5	--	intergenic	LDL	Blacks	T	-0.62 (0.12)	3.88E-06
rs17486127	2	--	intergenic	LDL	Blacks	G	-0.93 (0.19)	4.15E-06
rs340406	7	--	intergenic	LDL	Blacks	A	-1.22 (0.24)	4.35E-06
rs7786318	7	*PTN*	intron	LDL	Blacks	C	-0.55 (0.11)	5.45E-06
rs4586907	4	--	intergenic	LDL	Blacks	A	-0.87 (0.18)	5.68E-06
rs2429917	17	*SGSM2;MNT*	intron; downstream	LDL	Blacks	A	-1.45 (0.30)	7.01E-06
rs16942835	18	--	intergenic	LDL	Blacks	A	-1.45 (0.3)	7.01E-06
rs17429652	11	--	intergenic	LDL	Blacks	A	-1.45 (0.30)	7.01E-06
rs11871926	17	*SMG6*	intron	LDL	Blacks	T	-1.45 (0.30)	7.01E-06
rs6673413	1	*DNM3*	intron	LDL	Blacks	T	-1.45 (0.3)	7.01E-06
rs1615079	23	--	intergenic	LDL	Blacks	T	-0.46 (0.09)	7.32E-06
rs1634652	23	--	intergenic	LDL	Blacks	A	-0.46 (0.09)	7.32E-06
rs2702237	23	--	intergenic	LDL	Blacks	A	-0.46 (0.09)	7.32E-06
rs2702239	23	--	intergenic	LDL	Blacks	G	-0.46 (0.09)	7.32E-06
rs11016132	10	--	intergenic	LDL	Blacks	G	-1.45 (0.30)	7.68E-06
rs9346050	6	--	intergenic	LDL	Blacks	G	-1.45 (0.30)	7.68E-06
rs2887172	2	--	intergenic	LDL	Blacks	A	-1.45 (0.30)	8.3E-06
rs34166400	17	*SMG6*	intron	LDL	Blacks	G	-1.45 (0.3)	8.41E-06
rs6125463	20	*PREX1*	intron	LDL	Blacks	A	-0.69 (0.14)	8.77E-06
rs10868008	9	*FRMD3*	intron	LDL	Hispanics	A	-0.69 (0.13)	1.66E-06
rs11140077	9	*FRMD3*	intron	LDL	Hispanics	C	-0.69 (0.13)	1.66E-06
rs11140092	9	*FRMD3*	intron	LDL	Hispanics	A	-0.69 (0.13)	1.66E-06
rs16931326	8	*LOC401463;BHLHE22*	promoter; 3'UTR	LDL	Hispanics	T	-0.76 (0.16)	8.64E-06
rs11078231	17	--	intergenic	TG	Whites	A	0.24 (0.05)	2.39E-06
rs6477578	9	--	intergenic	TG	Whites	G	0.25 (0.05)	6.13E-06
rs1360414	9	--	intergenic	TG	Blacks	C	0.53 (0.10)	1.6E-06
rs16858329	3	*CD96*	intron	TG	Blacks	T	-0.36 (0.07)	4.29E-06
rs177001	5	--	intergenic	TG	Blacks	A	-0.52 (0.11)	9.89E-06
rs7790255	7	--	intergenic	TG	Hispanics	C	0.81 (0.14)	2.38E-07
rs15892	7	*GBAS; MRPS17*	promoter; 3'UTR	TG	Hispanics	A	0.73 (0.13)	9.42E-07
rs11722485	4	*LOC644753*	intron	TG	Hispanics	T	0.49 (0.09)	2.67E-06
rs2866056	4	*LOC644753*	intron	TG	Hispanics	G	0.48 (0.10)	4.28E-06
rs6593296	7	*CCT6A; PSPH; SNORA15*	promoter; intron; promoter	TG	Hispanics	A	0.66 (0.14)	7.76E-06

Appendix F. Association results of lipid candidate genes and HDL-C, LDL-C, and TG in the St. Jude discovery GWAS, Whites

Trait	Previously reported GWAS SNP	Nearby Lipid Candidate Gene(s)	Coded Allele	Beta (SE) in Discovery St. Jude GWAS	P-value in Discovery St. Jude GWAS
HDL	rs3890182	*ABCA1*	A	-0.12 (0.06)	***0.05***
HDL	rs4149268	*ABCA2*	T	-0.04 (0.04)	0.28
HDL	rs28927680	*APOA1/C3/A4/A5*	G	-0.05 (0.07)	0.50
HDL	rs4420638	*APOE/C1/C4*	C	-0.06 (0.05)	0.25
HDL	rs1566439	*CETP*	A	0.08 (0.04)	***0.04***
HDL	rs9989419	*CETP*	A	-0.05 (0.04)	0.18
HDL	rs174547	*FADS1*	C	-0.01 (0.04)	0.81
HDL	rs2271293	*LCAT*	A	0.01 (0.06)	0.89
HDL	rs4775041	*LIPC*	C	0.06 (0.04)	0.13
HDL	rs2156552	*LIPG*	A	-0.05 (0.04)	0.26
HDL	rs17411024 (r^2=1 with rs328)	*LPL*	A	0.06 (0.05)	0.24
HDL	rs2338104	*MMAB-MVK*	C	0.06 (0.04)	0.11
HDL	rs7679	*PLTP*	C	0.003 (0.04)	0.95
HDL	rs471364	*TTC39B*	G	-0.12 (0.05)	***0.02***
LDL	rs562338	*APOB*	T	-0.06 (0.04)	0.12
LDL	rs693	*APOB*	C	-0.05 (0.03)	0.07
LDL	rs754523	*APOB*	G	0.06 (0.03)	***0.05***
LDL	rs10402271	*APOE/C1/C4*	C	0.02 (0.03)	0.46
LDL	rs4420638	*APOE/C1/C4*	C	0.07 (0.04)	0.07
LDL	rs599839	*CELSR2/PSRC1/SORT*	G	-0.08 (0.04)	***0.03***
LDL	rs16996148	*CILP2/PBX4; NCAN*	T	-0.10 (0.06)	0.07
LDL	rs12654264	*HMGCR*	A	0.01 (0.03)	0.68
LDL	rs11206510	*PCSK9*	G	-0.08 (0.04)	***0.04***
LDL	rs1501908	*TIMD4*	G	0.003 (0.03)	0.92
TG	rs1748195	*ANGPTL3*	C	-0.06 (0.05)	0.26
TG	rs28927680	*APOA1/C3/A4/A5*	G	0.07 (0.11)	0.52
TG	rs16996148	*CILP2/PBX4; NCAN*	T	0.04 (0.09)	0.66
TG	rs174547	*FADS1*	C	-0.02 (0.05)	0.73
TG	rs780094	*GCKR*	T	0.01 (0.05)	0.89
TG	rs4775041	*LIPC*	C	-0.04 (0.05)	0.45
TG	rs17411024 (r^2=1 with rs328)	*LPL*	A	-0.17 (0.07)	***0.02***
TG	rs17145738	*MLXIPL*	T	-0.04 (0.09)	0.69
TG	rs7679	*PLTP*	C	0.08 (0.06)	0.19

Significant ($p \leq 0.05$) p-values are bolded and italicized.

Appendix G. Descriptions of Participants, Data Collection, and Genotyping by Study:

I. **Causal Variants Across the Life Course (CALiCo).** CALiCo is a consortium of six demographically diverse population based studies and a central laboratory. This network contributes a maximum of approximately 58,000 men and women ranging in age from childhood to older adulthood. Four CALiCo studies are involved in this analysis:

A. **Atherosclerosis Risk in Communities (ARIC) Study.** The ARIC study is a multi-center prospective investigation of atherosclerotic disease in a predominantly bi-racial population. White and African American men and women aged 45-64 years at baseline were recruited from four communities: Forsyth County, North Carolina; Jackson, Mississippi; suburban areas of Minneapolis, Minnesota; and Washington County, Maryland (The ARIC Investigators, 1989). A total of 15,792 individuals participated in the baseline examination in 1987-1989, with follow-up examinations in approximate 3-year intervals, during 1990-1992, 1993-1995, and 1996-1998. After the institutional review board at every participating university approved the ARIC Study protocol, written informed consent was obtained from each participant.

Data Collection: Body mass index was calculated as the ratio of weight in kilograms to height in meters squared. Current smoking was defined by "Do you now smoke cigarettes?" Diabetes status at baseline (yes/no) was based on either fasting plasma glucose levels ≥126 mg/dL, non-fasting plasma glucose ≥200 mg/dL, anti-diabetic medication use within two weeks of the baseline interview, or self-report of a physician diagnosis of diabetes. Women reported their age at menopause, if applicable, and hormone therapy replacement use at baseline. Women were either classified as pre-menopausal, peri-menopausal, natural post-menopausal, and surgical post-menopausal (Luoto et al., 2000). Race and lipid-lowering medication use were self-reported. Previous myocardial infarction (MI) was indicated by self-

report of physician-diagnosed MI or silent MI identified by electrocardiography. Fasting blood was drawn while the participant was seated from an antecubital vein into tubes containing EDTA, and plasma was obtained by centrifugation at 4°C and stored at −70°C until analysis. All samples were sent to the ARIC Central Lipid Laboratory for processing. Triglycerides were determined by enzymatic methods (Nagele et al., 1984) using the Cobas Bioanalyzer (Roche). Plasma HDL-C levels were measured using an enzymatic cholesterol assay using dextran-magnesium precipitation (Warnick et al., 1982). The Friedewald equation was used to calculate LDL-C in those with triglyceride levels under 400 mg/dl (Friedewald et al., 1972). Blood chemistries were performed at the Central Chemistry Laboratory of the University of Minnesota, and blood lipid analyses were performed at the University of Texas, Houston.

Genotyping: ARIC Study samples were genotyped using two approaches: *de novo* genotyping with TaqMan 6.0 (Applied Biosystems) and accessing previous genome-wide association study data from the Affymetrix Genome-Wide Human SNP Array 6.0 (Santa Clara, California). For the *de novo* genotyping data, the genotyping calls were made using the Applied Biosystem Autocaller 3.1 software. Internal QC's were included on every plate and across the full genotyping sample sets. QC genotypes were examined for consistency within the SNP genotyped. Other criteria include: 1) Internal genomic DNA, which is examined for replication across the study set; 2) Fingerprint blanks, which are used to identify the plate and verify cross-contamination and or sample error; 3) Genomic DNA pools to examine consistency of genotypes and plate validation; 4) No template controls (NTC), which serve as a background detector and review of any reagent problems; and 5) Autocaller confidence score, which determines percent genotyping call reliability. All SNPs are tested for departure from Hardy-Weinberg Equilibrium (HWE), and SNPs with

HWE $\chi2>3.84$ in unrelated cohorts (excluding duplicates) were excluded. SNP genotype data was also obtained from previous GWAS data. Genotyping was conducted using the Affymetrix Genome-Wide Human SNP Array 6.0 (Santa Clara, California). Sample exclusion criteria included discordant with previous genotype data, genotypic and phenotypic sex mismatch, suspected first-degree relative of an included individual based on genotype data (n=297), genetic outlier as assessed by Identity by State (IBS) using PLINK (Purcell et al., 2007; Friedewald et al., 1972), and >8 SD along any of the first 10 principal components in EIGENSTRAT (Price et al., 2006) with 5 iterations. Autosomal SNPs were used for imputation after exclusion of SNPs with HWE deviation $p<5 \times 10^{-5}$, call rate <95%, or MAF<1%.

B. The Coronary Artery Risk in Young Adults (CARDIA) Study. CARDIA is a multicenter longitudinal study of the development and determinants of cardiovascular disease in 5,115 young adults initially aged 18 to 30 years from 1985 to 1986. Black and white adults were recruited from four U.S. cities (Birmingham, Alabama; Chicago, Illinois; Minneapolis, Minnesota; and Oakland, California) with population-based samples approximately balanced within center by sex, age (18 to 24 or 25 to 30 years), race (white or black), and education (high school graduate or less or greater than high school graduate). Participants have been reexamined 2, 5, 7, 10, 15, and 20 years after baseline; and retention rates across examinations were 91%, 86%, 81%, 79%, 74%, and 72%, respectively. Further details of study recruitment and design are available (Friedman et al., 1988). All participants provided written informed consent at each examination, and institutional review boards from each field center and the coordinating center approved the study annually.

Data Collection: Each participant's age, race, and sex were self-reported during the recruitment phase and verified during the baseline clinic visit. Structured interviews or self-administered questionnaires were used to collect information on demographic

characteristics, lifestyle habits, physical activity, and medical history. The detailed methods, instruments and quality control procedures have been previously described (Friedman et al., 1988). Body weight was measured to the nearest 0.1 kg, using a calibrated scale, with the participant in light clothing without shoes. Height was measured to the nearest 0.5 cm with a vertical ruler. Body mass index (BMI) was computed as body weight / height^2 (kg/m^2). Blood samples were drawn after an overnight fast. Total plasma cholesterol, triglycerides, HDL-, and LDL-cholesterol were measured according to standardized methods. Triglycerides were measured enzymatically within 6 weeks of collection (Warnick, 1986). HDL-C was determined after precipitation with dextran sulfate/magnesium chloride of lipoproteins containing low-density lipoprotein cholesterol (Warnick et al., 1982).

Genotyping: Genotypes were obtained from the Central Texas laboratory using TaqMan (as described above for ARIC).

C. **The Cardiovascular Heart Study (CHS).** The CHS is a population-based longitudinal study of risk factors for cardiovascular disease in adults 65 years of age or older, recruited at four field centers (Forsyth County, North Carolina; Sacramento County, California; Washington County, Maryland; Pittsburgh, Pennsylvania) (Fried et al., 1991). Overall, 5,201 predominantly white individuals were recruited in 1989-1990 from random samples of Medicare eligibility lists, followed by an additional 687 African Americans recruited in 1992-1993 (total n=5,888).

Data Collection: CHS participants completed standardized clinical examinations and questionnaires at study enrollment and at nine annual follow-up visits. Height and weight were measured at the baseline examination. Current smoking status was self-reported at baseline. LDL cholesterol, HDL cholesterol, and triglycerides were measured under fasting conditions by enzymatic methods at a central laboratory (Cushman et al., 1995). Diabetes was defined as history of diabetes, use of

hypoglycemic agent or insulin, or fasting glucose 126 mg/dL. All women in CHS are postmenopausal. Race/ethnicity was based on self-report. MI was defined as evolving Q-wave MI or cardiac pain plus abnormal enzymes and either an evolving ST-T pattern or new left bundle branch block.

Genotyping: DNA was extracted from blood samples drawn on all participants at their baseline examination. Like ARIC, the CHS SNP genotypes were also obtained from two sources. First, SNP genotyping data was conducted in the Houston central lab using TaqMan (please see details above provided for the ARIC study). The second source of genotyping data from CHS was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV BeadChip system. Genotypes were called using the Illumina BeadStudio software as above. The following exclusions were applied to identify a final set of 306,655 autosomal SNPs: call rate < 97%, HWE $p < 1x10^{-5}$, > 1 duplicate error or Mendelian inconsistency (for reference CEPH trios), heterozygote frequency = 0, SNP not found in dbSNP. A total of 1,908 persons were excluded from the GWAS study sample due to the presence at study baseline of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke, or transient ischemic attack (Psaty et al., 2009).

D. **The Strong Heart Study (SHS).** SHS is composed of both a community-based and family-based study (Lee et al., 1990). The Strong Heart Community Study (SHCS) is a community-based study of CVD and its risk factors. The family component of the study, the SHFS, began in 1998 (phase III) with a pilot study that recruited and examined at least 300 members (North et al., 2003). In phase IV of the SHFS, an additional ~900 family members were recruited from each center yielding sample sizes of more than 1,200 participants at each site. In all centers, some individuals are

descended from more than one tribe and/or from non-Indian ancestors. Informed consent was obtained from all participants.

Data Collection: During the clinic visit for SHFS participants, a personal interview and physical exam of family members were performed. Tobacco exposure was quantified using standardized questionnaires. Anthropometric measures of height and weight were recorded and used to estimate BMI.

Genotyping: Genotypes were obtained from the Central Texas laboratory using TaqMan (as described above for ARIC).

II. **Epidemiologic Architecture of Genes Linked to Environment (EAGLE).** The EAGLE study accesses DNA samples and data collected for the National Health and Nutrition Examination Surveys (NHANES) by the National Center on Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). NHANES is a collection of diverse, population-based cross-sectional surveys of non-institutionalized Americans regardless of health status at the time of ascertainment. NHANES is considered a complex survey given that specific age groups (such as the elderly) and racial/ethnic groups (non-Hispanic blacks and Mexican-Americans) are oversampled. The NHANES data accessed for this work includes phase 2 of NHANES III (collected between 1991 and 1994), NHANES 1999-2000, and NHANES 2000-2001. Collectively, these surveys contain 14,998 DNA samples linked to demographic, health, and lifestyle data. Participants were consented by the CDC at the time of the survey and sample collection, and consent included the storage of data and biological specimens such as blood for future research (Centers for Disease Control and Prevention, 2002). The present study was approved by the CDC Ethics Review Board. Because the study investigators did not have access to personal identifiers, this study was considered non-human subjects research by the Vanderbilt University Internal Review Board.

Data Collection: In EAGLE/NHANES, race/ethnicity is self-described and was categorized as non-Hispanic white, non-Hispanic black, Mexican-American, and others. Serum HDL-C, triglycerides, and total cholesterol were measured using standard enzymatic methods. LDL-C was calculated using the Friedewald equation, with missing values assigned for samples with triglyceride levels greater than 400 mg/dl. Body mass index was calculated from height and weight measured in the Mobile Examination Center by CDC medical personnel. Current smoking was defined by "do you smoke cigarettes now?" or cotinine levels > 15ng/ml. Post-menopausal status was defined as a woman >60 years of age answering "no" to "have you had a period or regular periods in the past 12 months" or as a woman with bilateral oophorectomy. Current hormone use in NHANES III was defined as "yes" to "have you ever taken or used estrogen or female hormones in any form? Include pills, vaginal cream, suppositories, injections, or skin patches" and "still taking" to "how many months ago did you stop taking or using the estrogen or female hormones?". In NHANES 1999-2002, hormone use is defined as "yes" to "ever used female hormones?" and "yes" to "ever take estrogen/progestin?" Participants were considered to have type 2 diabetes if they answered "yes" to "Ever been told you have sugar/diabetes?" and "Are you now taking insulin?" or if they had fasting blood glucose levels >126 mg/dL. Previous myocardial infarction was defined by "Doctor ever told you had a heart attack?" (NHANES III) or "Ever told you had heart attack" (NHANES 1999-2002). Lipid lowering medication use was defined by "take prescribed med to lower cholesterol?" (NHANES III) and "now taking prescribed medicine to lower blood pressure" (NHANES 1999-2002).

Genotyping: In EAGLE, rs693 (*APOB*), rs673548 (*APOB*), rs2228671 (*LDLR*), and rs6511720 (*LDLR*) genotyping in NHANES III was performed using the Illumina GoldenGate assay (as part of a custom 384 OPA) by the Center for Inherited Disease

Research (CIDR) through the National Heart Lung and Blood Institute's Resequencing and Genotyping Service. For rs3890182 (*ABCA1*), rs3135506 (*APOA5*), rs1800775 (*CETP*), rs1260326 (*GCKR*), rs1323432 (*GRIN3A*), rs12654264 (*HMGCR*), rs1800588 (*LIPC*), rs1529729 (*LDLR*), rs328 (*LPL*), and rs11591147 (*PCSK9*) in NHANES III, we accessed existing data (Keebler et al., 2009) in the Genetic NHANES database. Genotyping was performed for the remaining EAGLE SNPs in NHANES III (n=7,159) and NHANES 1999-2002 (n=7,839) using Sequenom or Illumina BeadXpress. Genotyping was performed in the Vanderbilt DNA Resources Core. In addition to genotyping experimental NHANES samples, we genotyped blind duplicates provided by CDC and HapMap controls (n=360). All EAGLE SNPs reported here passed CDC quality control metrics and are available for secondary analyses through NCHS/CDC. All statistical analyses were conducted remotely in SAS v9.2 (SAS Institute, Cary, NC) using the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center in Hyattsville, MD.

III. **Multiethnic Cohort (MEC).** The MEC is a population-based prospective cohort study consisting of 215,251 men and women, and comprises mainly five self-reported racial/ethnic populations: African Americans, Japanese Americans, Latinos, Native Hawaiians and European Americans (Kolonel et al., 2004). The MEC was designed to provide prospective data on exposures and biomarkers potentially involved in cancer initiation and progression across groups with distinct cultural and dietary patterns. Between 1993 and 1996, adults between 45 and 75 years old were enrolled by completing a 26-page, self-administered questionnaire asking detailed information about dietary habits, demographic factors, level of education, personal behaviors, and history of prior medical conditions (e.g. diabetes). Between 1995 and 2004, blood specimens were collected from ~67,000 MEC participants at which time a short questionnaire was administered to update certain exposures, and collect

current information about medication use. Study protocols and consent forms were approved by the institutional review boards at all participating institutions.

Data Collection: MEC baseline questionnaires queried ever smokers (defined as >20 packs in lifetime) about the average number of cigarettes smoked per day, duration in years, and years since quitting if not currently smoking. **S**elf-reported height and weight were used to calculate baseline BMI. Medication use and fasting information was collected at the time of blood collections. LDL, HDL, and triglycerides concentrations were measured using standard clinical assays.

Genotyping: Genotyping of 43 SNPs was conducted by the OpenArray platform (Life Technologies, Carlsbad, CA) at the Cancer Research Center in Hawaii following the recommended protocol. For all SNPs genotype call rates were >90% and individual call rates were >90%. Concordance rates between duplicate samples were 100%. HWE was examined and p-values were >0.01 in at least 4 of the 5 ethnic groups.

IV. Women's Health Initiative (WHI). WHI is a long-term national health study that focuses on strategies for preventing heart disease, breast and colorectal cancer and fracture in postmenopausal women. A total of 161, 838 women aged 50–79 yrs old were recruited from 40 clinical centers in the US between 1993 and 1998(Anderson et al., 2003). WHI consists of an observational study, two clinical trials of postmenopausal hormone therapy (estrogen alone or estrogen plus progestin), a calcium and vitamin D supplement trial, and a dietary modification trial. Trial exclusion criteria have been described previously(1998). Study protocols and consent forms were approved by the institutional review boards at all participating institutions. A subset of 21,000 WHI women were selected for genotyping and inclusion in these PAGE analyses, of those, approximately 8,000 have at least one baseline lipid measurement available. Women were selected based on self-reported

history of disease, incident event outcomes, DNA availability and consent, and racial/ethnic diversity.

Data Collection: Self-reported demographic, lifestyle and general health characteristics (current smoking, history of myocardial infarction, type 2 diabetes) were collected at baseline as described previously (1998). BMI was calculated from measured weight and height at time of enrollment. Baseline medication use (lipid-lowering and hormone replacement medications) was ascertained using a computer-driven medication inventory system at the first screening visit. Race/ethnicity was self reported as one of white, black, Hispanic, Asian/Pacific Islander, Native American, or other (this last category was not genotyped for PAGE). Additional self-reported race/ethnicity data were available for a subset of participants and were used to subset Asian/Pacific Islanders into East Asians and Pacific Islanders. Fasting LDL, HDL, and triglycerides concentrations were measured using standard clinical assays in a variety of core and ancillary WHI studies; measurements were normalized to correct for laboratory and study effect.

Genotyping: In WHI, 20 lipids SNPs were genotyped at the Translational Genomics Research Institute (TGen) (Phoenix, AZ) on Illumina's BeadXpress Reader using Illumina's Veracode GoldenGate genotyping assay, following the manufacturer's recommended protocol (www.illumina.com). Study protocol included calculation of concordance rates among duplicates, genotyping of HapMap samples, re-genotyping of failing samples, and other extensive QA procedures. One HDL SNP, rs1883025, did not pass quality control and was excluded from analyses; all other genotyping data reported here passed QA with individual and SNP call rates exceeding 95% and 97%, respectively.

Appendix H. Study characteristics by PAGE study and population. Descriptive statistics for fasting (≥8 hours) adults (≥18 years of age) are expressed as percentage, median, and standard deviation (SD) for each variable.

a) CALiCo (ARIC)

	European Americans	**African Americans**
N	11,178	3,770
% Female	52.93	62.02
Median Age (SD)	54 years (5.71)	53 years (5.84)
Median BMI (SD)	26 kg/m^2 (4.86)	29 kg/m^2 (6.13)
Median HDL-C (SD)	47 mg/dl (16.74)	52 mg/dl (17.56)
Median LDL-C (SD)	135 mg/dl (37.74)	135 mg/dl (43.18)
Median TG (SD)	115 mg/dl (91.16)	95 mg/dl (76.60)
% Lipid lowering medication use	3.41	1.41
% Current smokers	24.40	29.54
% Post menopausal	59.52	58.35
% Hormone use	20.49	13.73
% Previous MI	4.01	3.23

b) CALiCo (CARDIA)

	European Americans	**African Americans**
N	2,134	2,035
% Female	53.30	58.40
Median Age (SD)	26 years (3.4)	24.5 years (3.8)
Median BMI (SD)	23 kg/m^2 (4.1)	24 kg/m^2 (5.75)
Median HDL-C (SD)	51 mg/dl (12.97)	53 mg/dl (13.01)
Median LDL-C (SD)	105 mg/dl (29.8)	109 mg/dl (31.85)
Median TG (SD)	66 mg/dl (55.7)	57 mg/dl (36)
% Lipid lowering medication use	0	0
% Current smokers	25.4	32.3
% Post menopausal	0	0
% Hormone use	0	0
% Previous MI	0	0

c) CALiCo (CHS)

	European Americans	African Americans
N	2,787	550
% Female	61	63
Median Age (SD)	74 years (5.15)	72 years (5.63)
Median BMI (SD)	26 kg/m^2 (4.48)	28 kg/m^2 (5.58)
Median HDL-C (SD)	51 mg/dl (14.18)	55 mg/dl (15.11)
Median LDL-C (SD)	126 mg/dl (33.44)	128 mg/dl (36.13)
Median TG (SD)	128 mg/dl (88.18)	102 mg/dl (58.35)
% Lipid lowering medication use	8	7
% Current smokers	9	14
% Post menopausal	100	100
% Hormone use	9	5
% Previous MI	11	8

d) EAGLE

	European Americans	African Americans	Mexican Americans
N	3,909	1,896	2,361
% Female	53.83	55.89	50.76
Median Age (SD)	51 years (19.69)	39 years (16.65)	40 years (17.52)
Median BMI (SD)	27 kg/m^2 (5.84)	28 kg/m^2 (6.96)	28 kg/m^2 (5.51)
Median HDL-C (SD)	48 mg/dl (15.76)	51 mg/dl (16.82)	46 mg/dl (13.35)
Median LDL-C (SD)	125 mg/dl (36.08)	118 mg/dl (38.75)	119 mg/dl (33.54)
Median TG (SD)	123 mg/dl (124.53)	85 mg/dl (75.85)	128 mg/dl (123.47)
% Lipid lowering medication use	4.61	2.15	2.37
% Current smokers	261.0	35.30	20.90
% Post menopausal	18.77	9.36	10.01
% Hormone use	4.99	3.56	2.07
% Previous MI	5.17	3.29	2.28

e) MEC

	European Americans	African Americans	Mexican Americans	Native Hawaiians	Japanese
N	317	552	299	87	576
% Female	38.85	18.08	32.78	48.86	38.99
Median Age (SD)	67 years (8.01)	69 years (7.13)	68 years (6.92)	62 years (7.02)	70 years (8.28)
Median BMI (SD)	26 kg/m^2 (5.11)	27 kg/m^2 (4.71)	27 kg/m^2 (4.54)	28 kg/m^2 (5.32)	24 kg/m^2 (3.61)
Median HDL-C (SD)	53 mg/dl (15.57)	50 mg/dl (14.69)	48 mg/dl (12.97)	48 mg/dl (13.66)	54 mg/dl (14.92)
Median LDL-C (SD)	116 mg/dl (32.71)	123 mg/dl (41.93)	118 mg/dl (36.65)	108 mg/dl (27.16)	111 mg/dl (34.71)
Median TG (SD)	109 mg/dl (61.27)	103 mg/dl (54.47)	136 mg/dl (70.37)	130 mg/dl (69.85)	126 mg/dl (69.31)
% Lipid lowering medication use	28.39	20.8	24.08	27.27	38.3
% Current smokers	6.31	21.05	15.36	9.09	8.7
% Post menopausal	76.32	62.22	71.43	67.5	69.16
% Hormone use	67.8	50	51.04	57.14	61.26
% Previous MI	5.67	9.95	8.36	6.82	5.72

f) CALiCo (SHFS)

	American Indians (Arizona)	American Indians (Oklahoma)	American Indians (the Dakotas)
N	1,191	1,196	1,178
% Female	62	59	59
Median Age (SD)	36 years (15.96)	42 years (17.29)	38 years (17.08)
Median BMI (SD)	35 kg/m^2 (8.80)	30 kg/m^2 (6.89)	29 kg/m^2 (6.83)
Median HDL-C (SD)	45 mg/dl (14.11)	50 mg/dl (15.38)	49 mg/dl (13.74)
Median LDL-C (SD)	93 mg/dl (25.91)	97 mg/dl (30.41)	98 mg/dl (30.92)
Median TG (SD)	139 mg/dl (134.80)	144 mg/dl (170.62)	126 mg/dl (201.74)
% Lipid lowering medication use	3.91	5.14	6.07
% Current smokers	25.06	33.22	42.22
% Post menopausal	15.82	21.96	15.94
% Hormone use	2.41	9.44	4.97
% Previous MI	--	--	--

g) CALiCo (SHS-C)

	American Indians (Arizona)	American Indians (Oklahoma)	American Indians (the Dakotas)
N	950	943	939
% Female	68	60	59
Median Age (SD)	54 years (7.88)	56 years (8.19)	55 years (7.91)
Median BMI (SD)	32 kg/m^2 (7.07)	30 kg/m^2 (6.04)	29 kg/m^2 (5.46)
Median HDL-C (SD)	43 mg/dl (12.31)	44 mg/dl (13.61)	44 mg/dl (14.41)
Median LDL-C (SD)	97 mg/dl (29.43)	109 mg/dl (31.20)	113 mg/dl (32.29)
Median TG (SD)	121 mg/dl (138.42)	122 mg/dl (120.59)	113 mg/dl (198.42)
% Lipid lowering medication use	--	--	--
% Current smokers	18.34	32.63	47.79
% Post menopausal	--	--	--
% Hormone use	2.41	9.44	4.97
% Previous MI	0.83	2.94	3.37

h) WHI

	European Americans	African Americans	Hispanics	Asian/Pacific Islander	American Indians
N	4,688	1,840	762	359	113
% Female	100	100	100	100	100
Median Age (SD)	67 years (6.86)	60 years (7.13)	60 years (6.67)	66 years (7.19)	59 years (6.72)
Median BMI (SD)	27 kg/m^2 (6.53)	30 kg/m^2 (8.12)	28 kg/m^2 (5.79)	24 mg/k^2 (4.10)	29 kg/m^2 (5.89)
Median HDL-C (SD)	55 mg/dl (15.26)	55 mg/dl (14.36)	51 mg/dl (14.34)	59 mg/dl (16.28)	52 mg/dl (13.53)
Median LDL-C (SD)	140 mg/dl (34.31)	137 mg/dl (38.46)	124 mg/dl (38.22)	124 mg/dl (29.57)	124 mg/dl (31.70)
Median TG (mg/dl)	140 mg/dl (114.12)	110 mg/dl (55.73)	151 mg/dl (90.38)	140 mg/dl (87.98)	140 mg/dl (64.34)
% Lipid lowering medication use	9.3	7.3	7.4	16.2	4.4
% Current smokers	8.1	10.3	6.4	2.3	8.9
% Post menopausal	100	100	100	100	100
% Hormone use	28.2	22.3	31.8	40.7	41.6
% Previous MI	1.6	1.5	0	1.1	0.9

Appendix I. List of candidate gene and GWAS-identified SNPs targeted for genotyping in PAGE. For each SNP (denoted by rs number), we list the chromosomal and genomic location, the putative function of the SNP (based on SNP location) and the nearest gene, the number of PAGE studies that genotyped the SNP, the trait associated with the SNP based on the literature, the effect allele and effect size based on the literature, and the reference for these data. Abbreviations: Base-pair (bp), Untranslated region (UTR). *Number of PAGE studies that genotyped the SNP. PAGE studies include CALiCo, EAGLE, MEC, and WHI. §SNP failed genotyping on the BeadXpress at one PAGE site (WHI); therefore, only data from one PAGE site was available for analysis (EAGLE using Sequenom). ‡Effect size (beta) is reported in s.d. (standard deviation) units. †Effect size originally expressed in mmol/L. #Ariza et al 2010 collapsed heterozygotes and homozygotes for the minor allele into the same bin and compared them to the referent (homozygotes for the major allele) to estimate the genetic effect size.

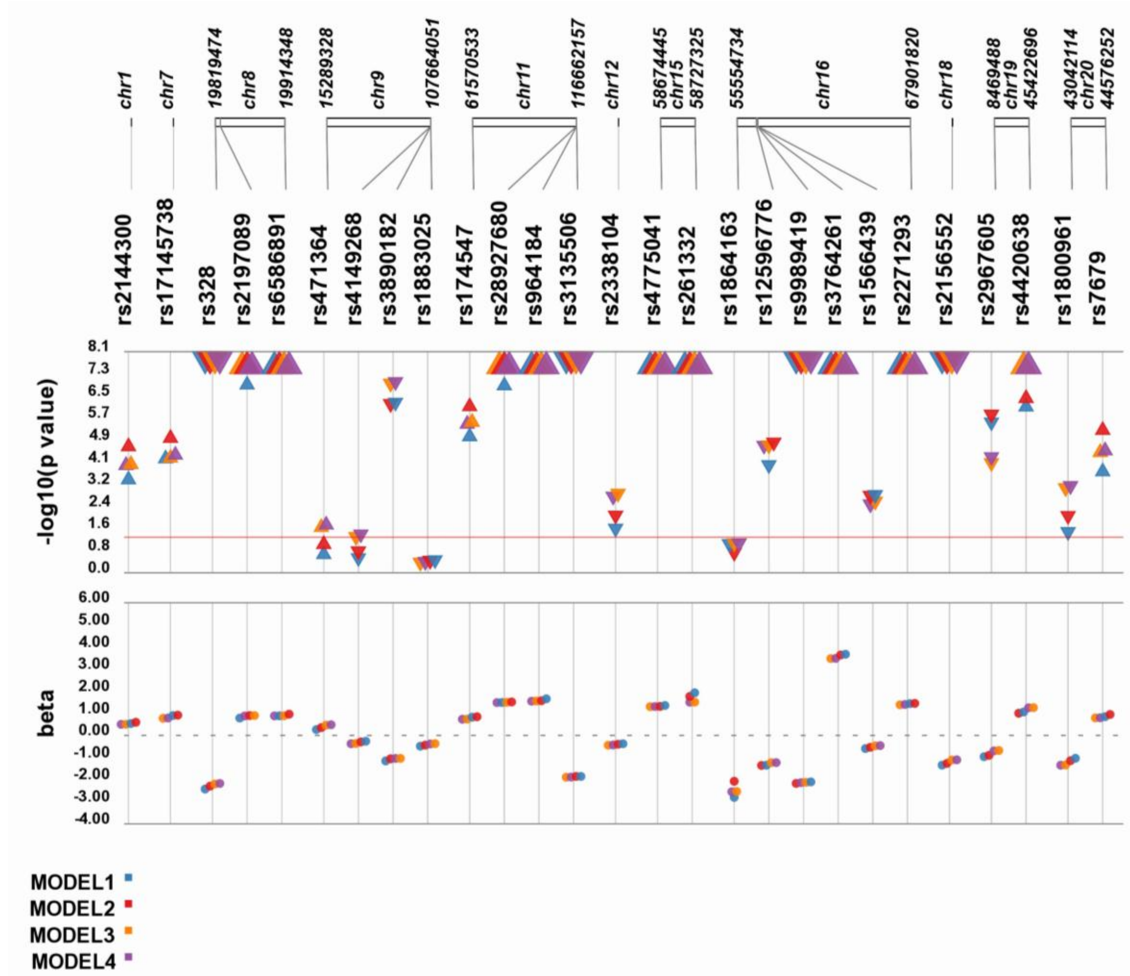
SNP	Chr.	Build 37 location (bp)	Function	Nearest Gene	# PAGE Studies*	Previously Associated Trait	Effect Allele	Effect Size‡ (mg/dl)	Reference
rs11206510	1	55495789	Intergenic	*PCSK9*	4	LDL-C	T	3.04	Willer et al 2008
rs11591147	1	55505397	Non-synonymous	*PCSK9*	3	LDL-C	T	-17.1	Kathiresan et al 2009
rs1748195	1	63049343	Intronic	*ANGPTL3*	2	TG	C	7.12	Willer et al 2008
rs646776	1	109818280	Downstream	*CELSR2/PSRC/SORT*	4	LDL-C	T	6.18	Kathiresan et al 2008
rs599839	1	109821916	Downstream	*CELSR2/PSRC/SORT*	3	LDL-C	A	5.48	Willer et al 2008
rs2144300	1	230294666	Intronic	*GALNT2*	4	HDL-C	T	1.11	Willer et al 2008
rs693	2	21231945	Synonymous	*APOB*	4	LDL-C	A	2.44	Willer et al 2008
rs562338	2	21288071	Intergenic	*APOB*	4	LDL-C	A	-4.89	Willer et al 2008
rs754523	2	21311441	Intergenic	*APOB*	4	LDL-C	A	-2.78	Willer et al 2008
rs1260326	2	27730690	Non-synonymous splice site	*GCKR*	2	TG	T	8.76	Teslovich et al 2010
rs780094	2	27740987	Intronic	*GCKR*	3	TG	T	8.59	Willer et al 2008
rs6544713	2	44073631	Intronic	*ABCG8*	4	LDL-C	T	5.1	Kathiresan et al 2009
rs12654264	5	74648353	Intronic	*HMGCR*	4	LDL-C	A	-3.86	Kathiresan et al 2008
rs1501908	5	156397919	Intergenic	*TIMD4*	3	LDL-C	C	2.38	Kathiresan et al 2009
rs17145738	7	72982624	Downstream	*MLXIPL*	4	HDL-C	T	0.57	Teslovich et al 2010
						TG	T	-9.32	Teslovich et al 2010

rs328	8	19819474	Nonsense	*LPL*	4	HDL-C	C	-2.62	Kathiresan et al 2008
						TG	C	19.47	Kathiresan et al 2008
rs2197089	8	19826123	Downstream	*LPL*	4	HDL-C	A	1.38	Willer et al 2008
						TG	A	-3.38	Willer et al 2008
rs6586891	8	19914348	Intergenic	*LPL*	4	HDL-C	A	1	Willer et al 2008
rs2954029	8	126560154	Intergenic	*TRIB1*	2	TG	A	5.64	Teslovich et al 2010
rs471364	9	15289328	Intronic	*TTC39B*	3	HDL-C	T	1.2	Kathiresan et al 2009
rs4149268	9	107647220	Intronic	*ABCA1*	4	HDL-C	T	-0.82	Willer et al 2008
rs3890182	9	107647405	Intronic	*ABCA1*	4	HDL-C	A	-1.54	Kathiresan et al 2008
rs1883025	9	107664051	Intronic	*ABCA1*	2[§]	HDL-C	A	-0.94	Teslovich et al 2010
rs174547	11	61570533	Intronic	*FADS1*	4	HDL-C	T	1.35	Kathiresan et al 2009
						TG	T	5.46	Kathiresan et al 2009
rs28927680	11	116618823	3' UTR	*APOA1/C3/A4/A5* gene cluster	4	HDL-C	C	2.01	Kathiresan et al 2008
						TG	C	-16.95	Teslovich et al 2010
rs964184	11	116648667	3' UTR	*APOA1/C3/A4/A5* gene cluster	2	HDL-C	C	1.5	Teslovich et al 2010
						TG	C	-27.3	Kathiresan et al 2009
rs3135506	11	116662157	Non-synonymous	*APOA1/C3/A4/A5* gene cluster	3	HDL-C	C	-2.65	Lu et al 2008
						TG	C	1.13#	Ariza et al 2010
rs2338104	12	109894918	Intronic	*MMAB*-MVK	3	HDL-C	C	-0.48	Willer et al 2008
rs2650000	12	121388712	Intergenic	*HNF1A*	3	LDL-C	A	2.38	Kathiresan et al 2009
rs4775041	15	58674445	Intergenic	*LIPC*	4	HDL-C	C	1.38	Willer et al 2008
						TG	C	3.62	Willer et al 2008
rs261332	15	58727325	Intronic	*LIPC*	3	HDL-C	A	1.41	Willer et al 2008
rs1800775	16	555552737	Upstream	*CETP*	3	HDL-C	SNP failed genotyping at all PAGE sites that attempted genotyping		
rs1864163	16	55554734	Intronic	*CETP*	2	HDL-C	A	-4.12	Willer et al 2008
rs12596776	16	56919098	Intronic	*CETP*	2	HDL-C	C	-1.26	Willer et al 2008
rs9989419	16	56984889	Upstream	*CETP*	4	HDL-C	A	-1.72	Willer et al 2008
rs3764261	16	56993074	Upstream	*CETP*	4	HDL-C	A	3.47	Willer et al 2008
rs1566439	16	57024412	Intronic	*CETP*	2	HDL-C	T	-0.93	Willer et al 2008
rs2271293	16	67901820	Intronic	*LCAT*	4	HDL-C	A	1.05	Kathiresan et al 2009
rs2156552	18	47181418	Intergenic	*LIPG*	4	HDL-C	A	-1.2	Willer et al 2008
rs2967605	19	8469488	Downstream	*ANGPTL4*	3	HDL-C	T	-1.8	Kathiresan et al 2009

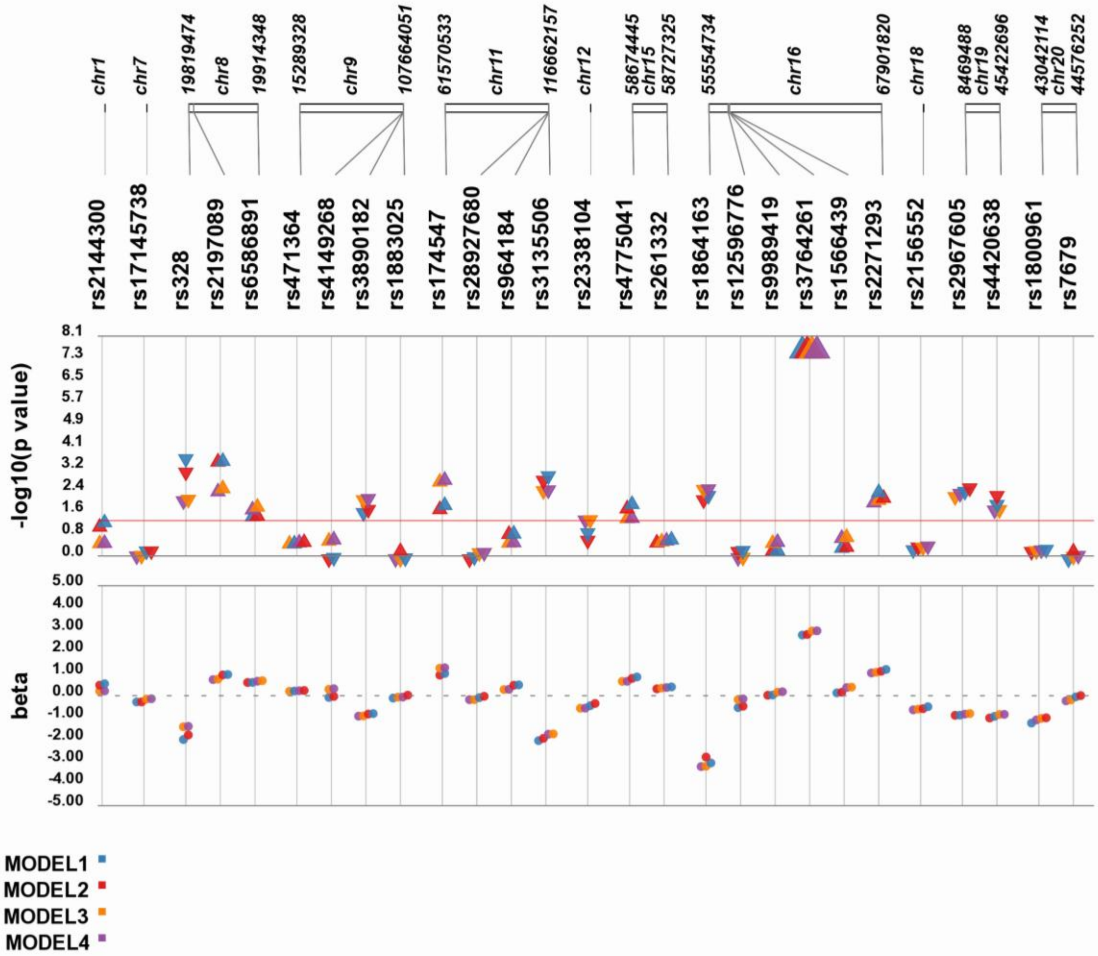
rs6511720	19	11202056	Intronic	*LDLR*	4	LDL-C	T	-6.99	Teslovich et al 2010
rs2228671	19	11210662	Synonymous	*LDLR*	2	LDL-C	T	-0.14‡	Aulchenko et al 2009
rs16996148	19	19658222	Downstream	*CILP2/PBX4/NCAN*	4	LDL-C	T	-3.32	Willer et al 2008
						TG	T	-6.1	Willer et al 2008
rs4803750	19	45247627	Upstream	*BCL3*	2	LDL-C	A	10.9†	Sandhu et al 2008
rs10402271	19	45329214	Downstream	*APOE/C1/C4* gene cluster	2	LDL-C	T	-2.62	Willer et al 2008
rs4420638	19	45422696	Downstream	*APOE/C1/C4* gene cluster	4	HDL-C	A	1.06	Teslovich et al 2010
						LDL-C	A	-7.14	Teslovich et al 2010
rs2075650	19	50087459	Intronic	*TOMM40*	2	LDL-C	A	-0.16‡	Aulchenko et al 2009
rs429358	19	50103781	Non-synonymous	*APOE*	2	LDL-C	SNP failed genotyping at all PAGE sites that attempted genotyping		
rs7412	19	50103919	Non-synonymous	*APOE*	2	LDL-C	SNP failed genotyping at all PAGE sites that attempted genotyping		
rs6102059	20	39228784	Intergenic	*MAFB*	3	LDL-C	T	-2.04	Kathiresan et al 2009
rs1800961	20	43042114	Non-synonymous	*HNF4A*	3	HDL-C	T	-1.88	Teslovich et al 2010
rs7679	20	44576252	Downstream	*PLTP*	3	HDL-C	T	1.05	Kathiresan et al 2009
						TG	T	-6.37	Kathiresan et al 2009

Appendix J. Comparison of unadjusted, minimally adjusted, adjusted models for HDL-C, by population. Results of tests of association for four regression models are plotted: model 1 (unadjusted), model 2(adjusted for age and sex; and site of ascertainment for select PAGE studies), model 3 (adjusted for age, sex, body mass index, current smoking, type 2 diabetes, post-menopausal status, and current hormone use), and model 4 (model 3 with the addition of previous myocardial infarction). Each SNP was tested for an association with HDL-C. Meta-analysis was performed, and p-values ($-\log_{10}$ transformed) of the meta-analysis are plotted along the y-axis. SNP location is given on the x-axis. Each triangle represents a meta-analysis p-value for each population. Models are color coded. Large triangles represent p-values at or smaller than genome-wide significance ($p<10^{-8}$). The direction of the arrows corresponds to the direction of the beta coefficient. The exact beta coefficients are reported on the bottom panel. The significance threshold is indicated by the red bar at p=0.05.

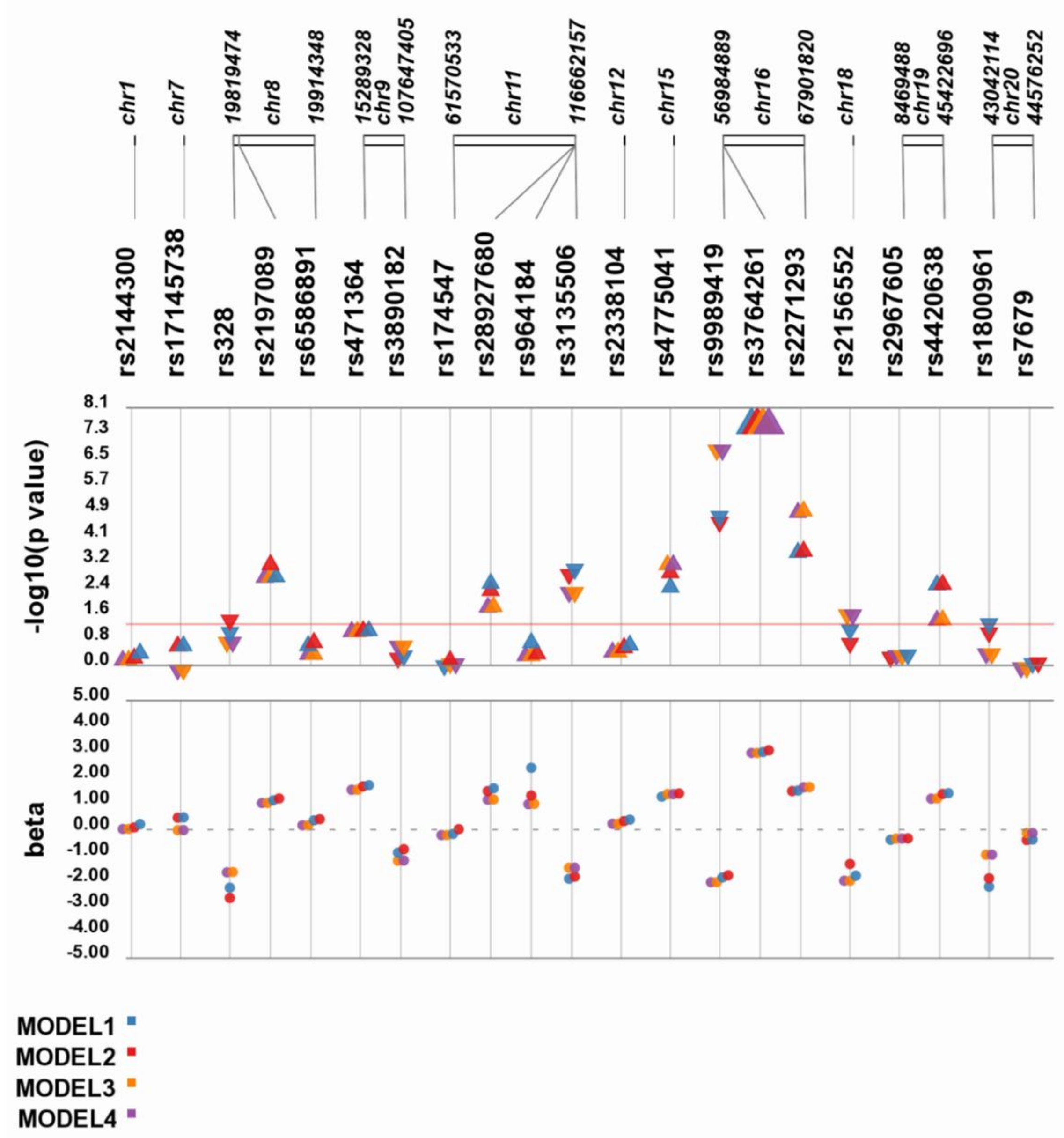
a) **European Americans**



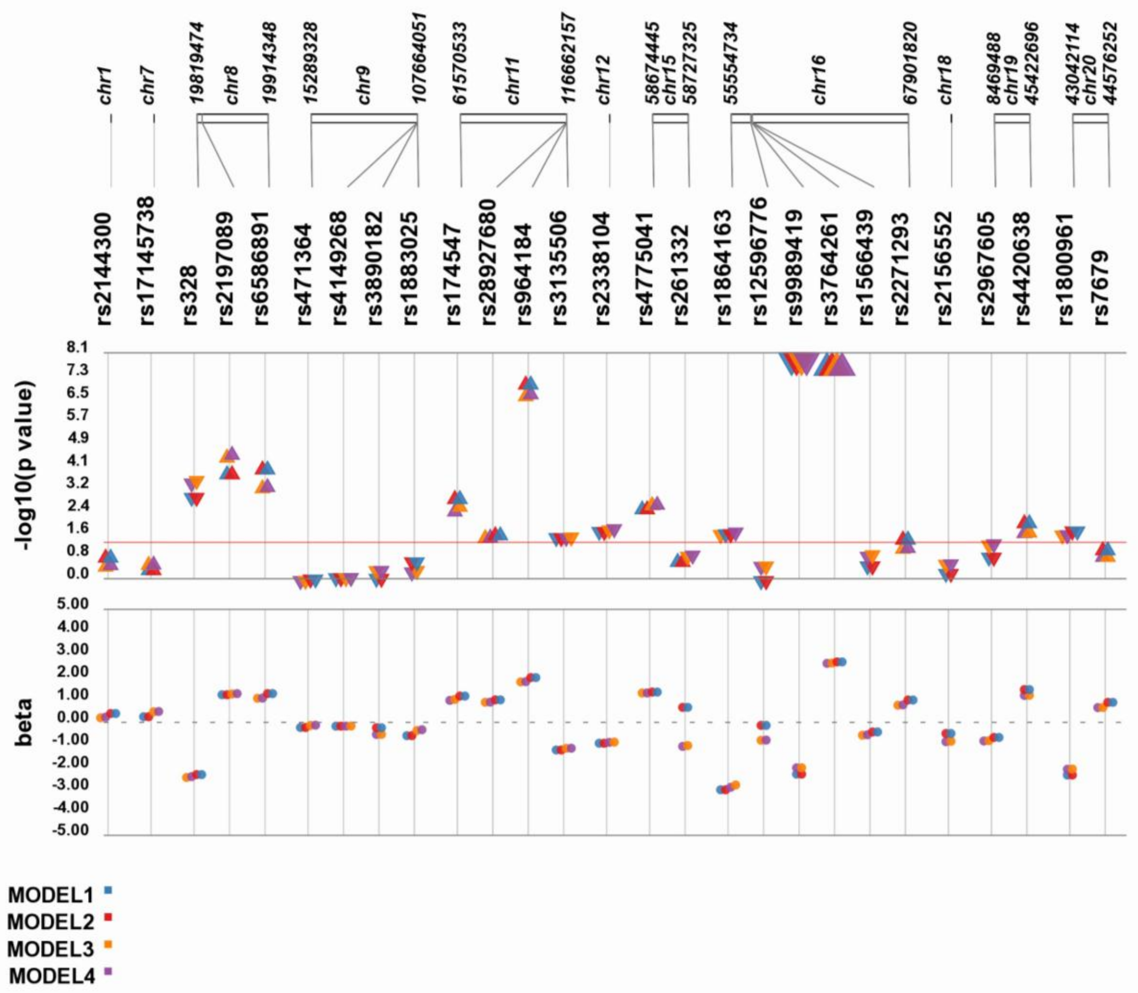
b) African Americans



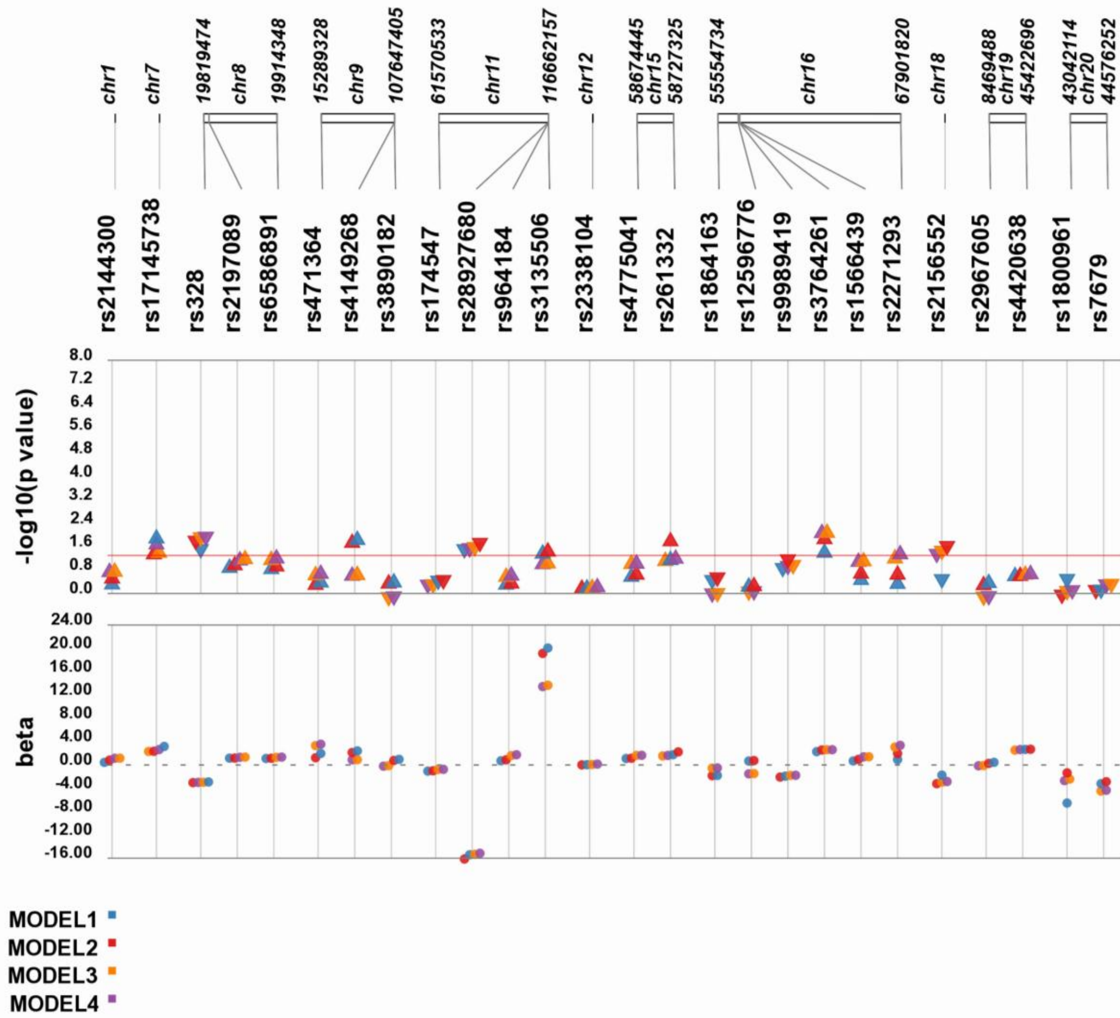
c) **American Indians**



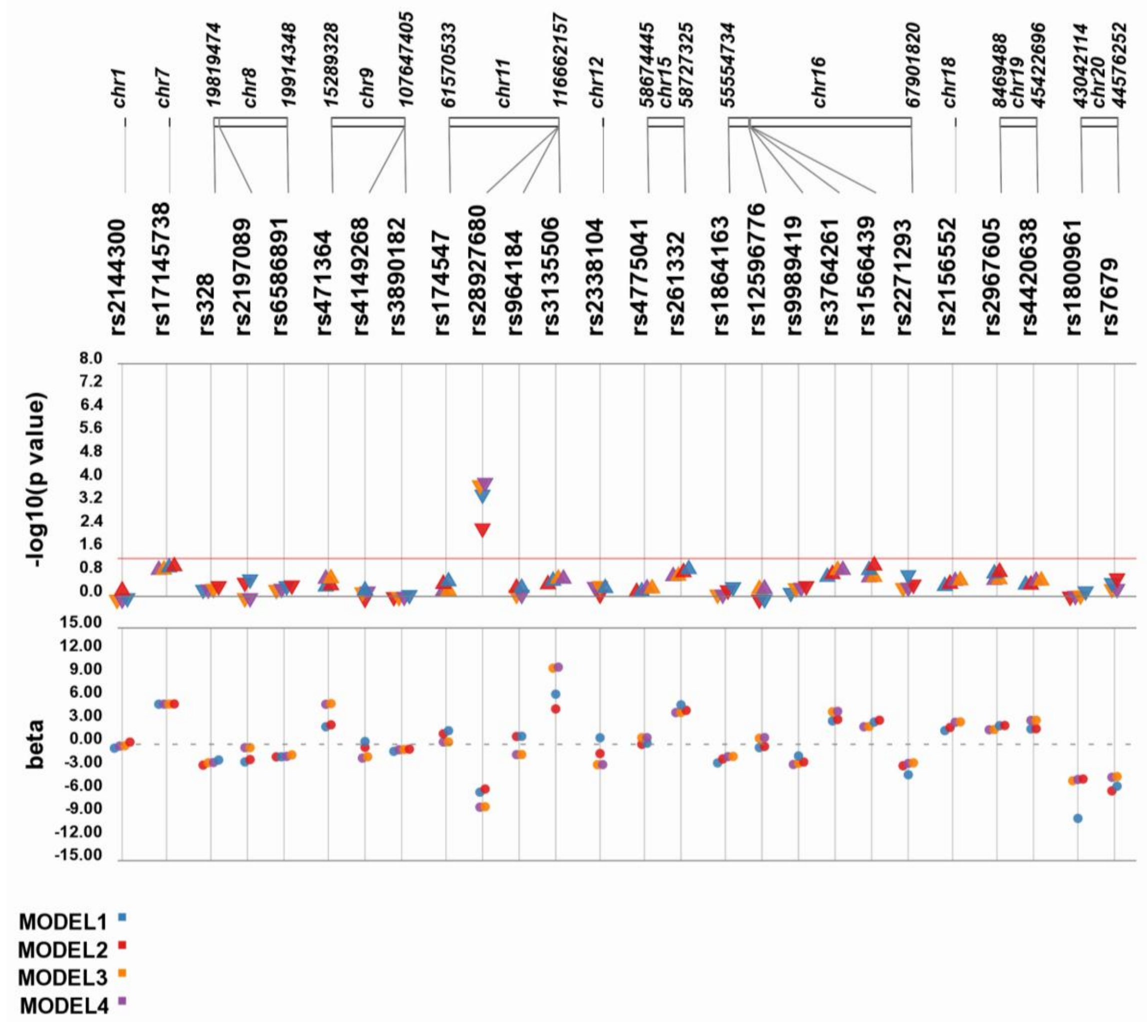
d) Mexican Americans/Hispanics



e) Japanese/East Asians

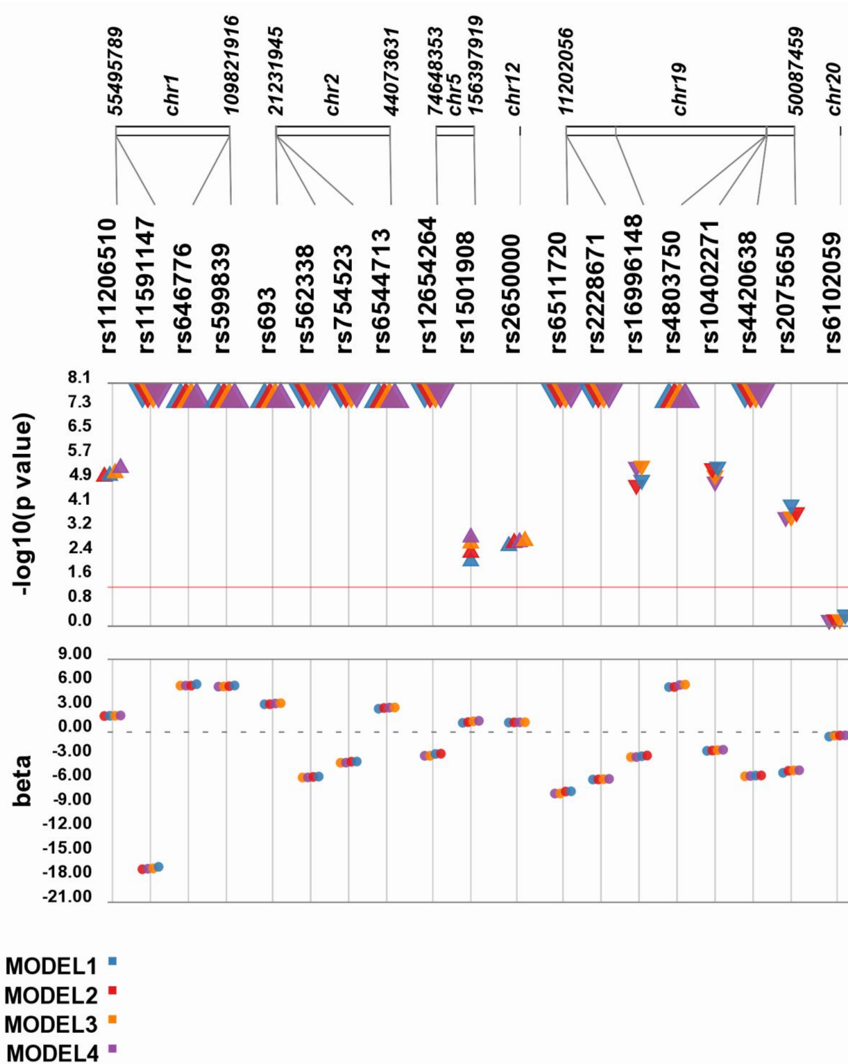


f) Native Hawaiians/Pacific Islanders

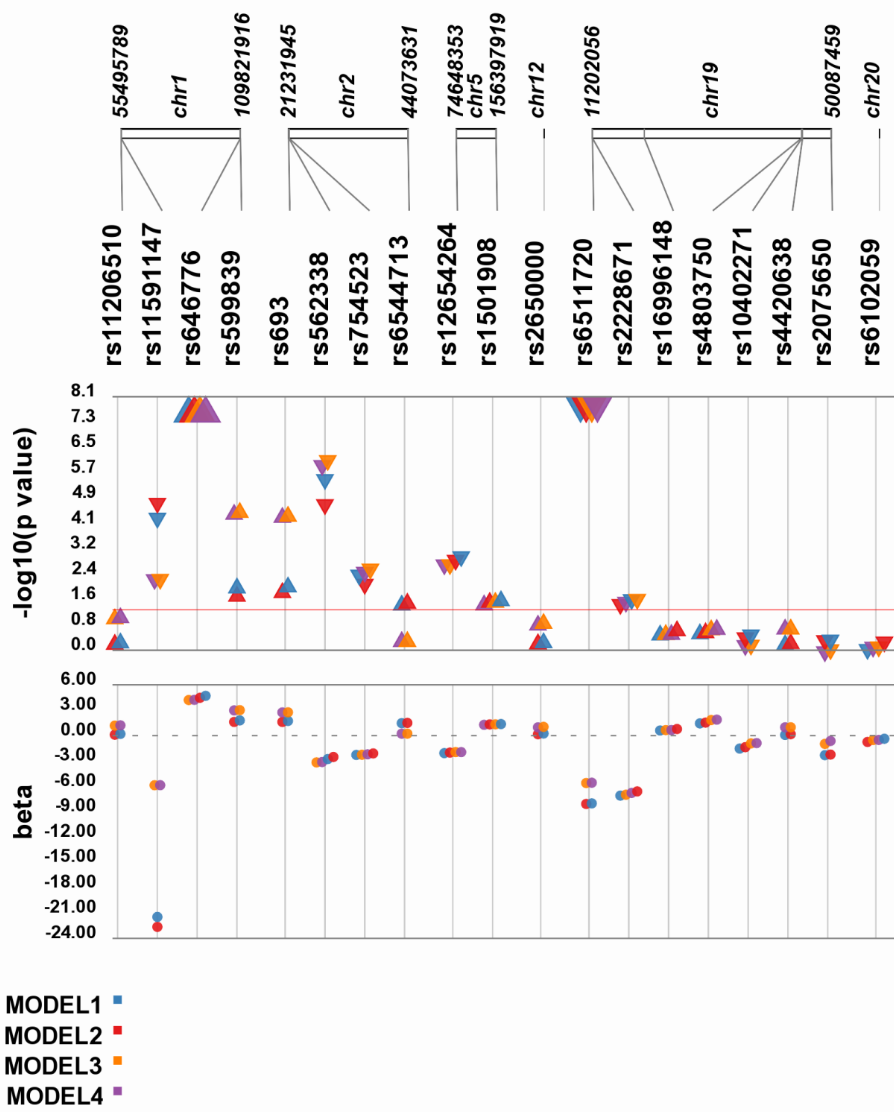


Appendix K. Comparison of unadjusted, minimally adjusted, adjusted models for LDL-C, by population. Results of tests of association for four regression models are plotted: model 1 (unadjusted), model 2(adjusted for age and sex; and site of ascertainment for select PAGE studies), model 3 (adjusted for age, sex, body mass index, current smoking, type 2 diabetes, post-menopausal status, and current hormone use), and model 4 (model 3 with the addition of previous myocardial infarction). Each SNP was tested for an association with LDL-C. Meta-analysis was performed, and p-values (–$\log_{10}$ transformed) of the meta-analysis are plotted along the y-axis. SNP location is given on the x-axis. Each triangle represents a meta-analysis p-value for each population. Models are color coded. Large triangles represent p-values at or smaller than genome-wide significance ($p<10^{-8}$). The direction of the arrows corresponds to the direction of the beta coefficient. The exact beta coefficients are reported on the bottom panel. The significance threshold is indicated by the red bar at p=0.05.

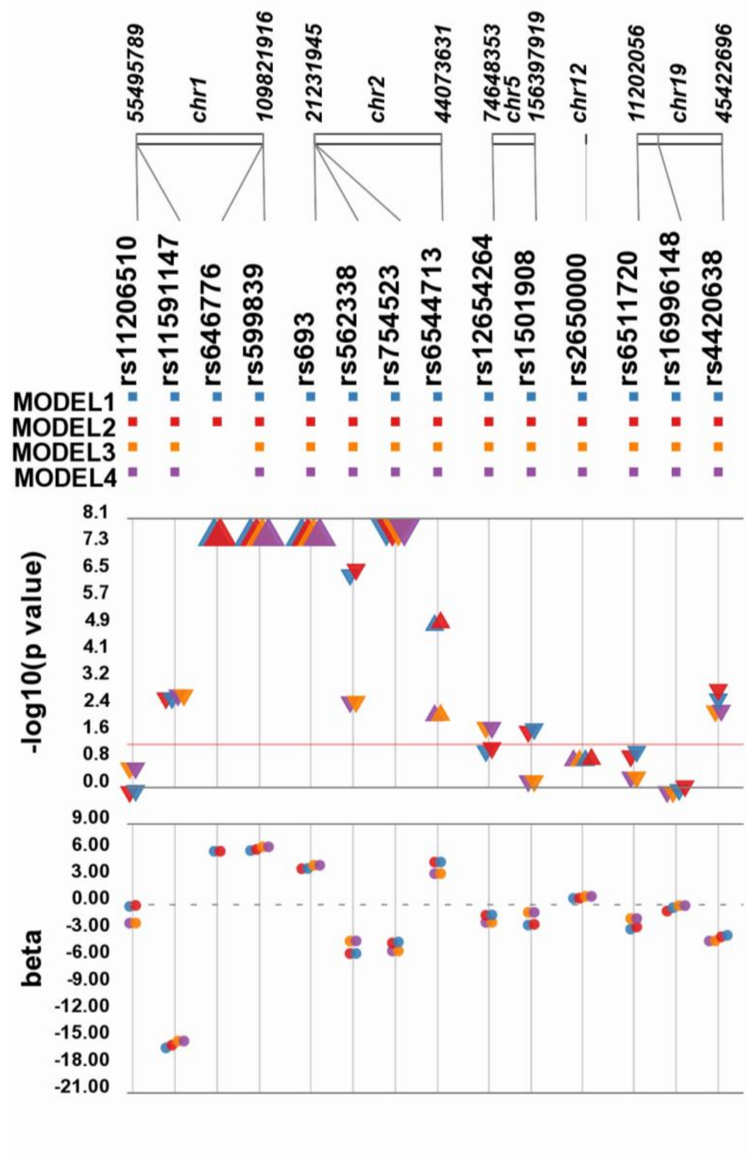
a) European Americans



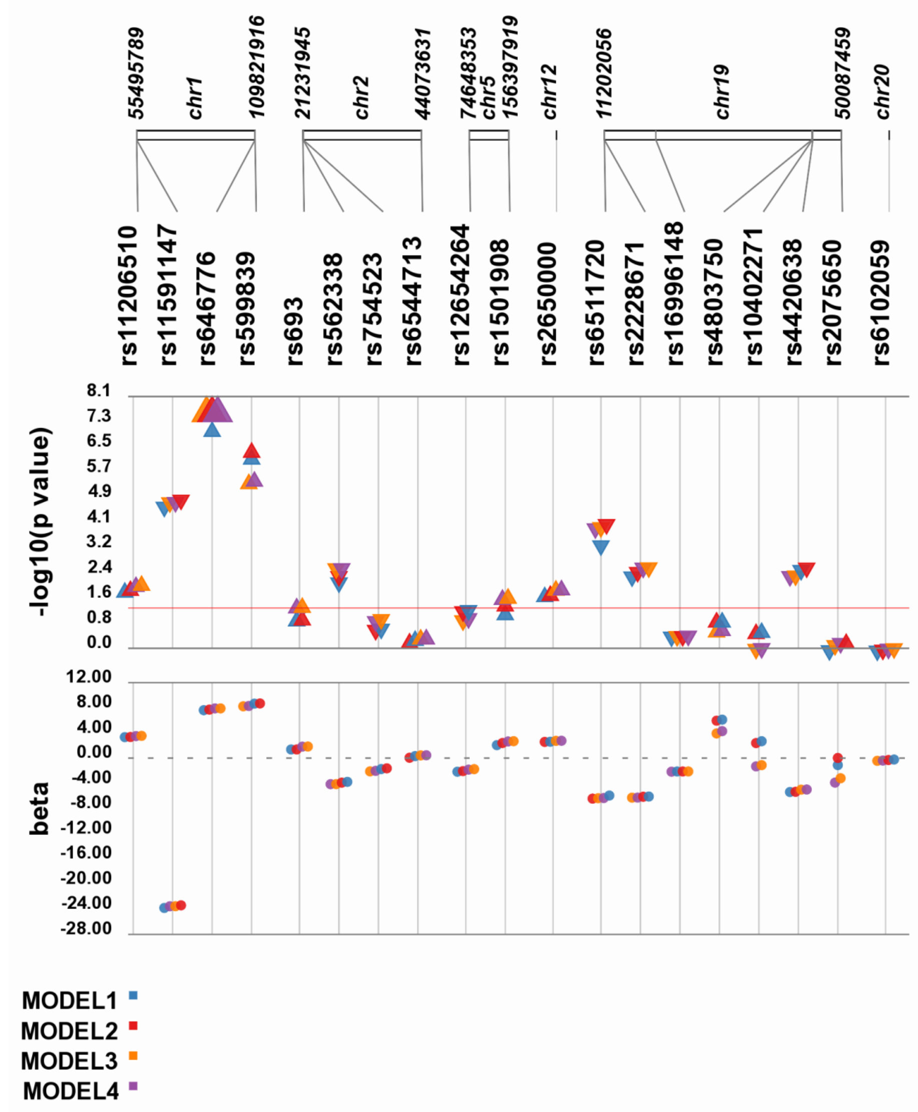
b) African Americans



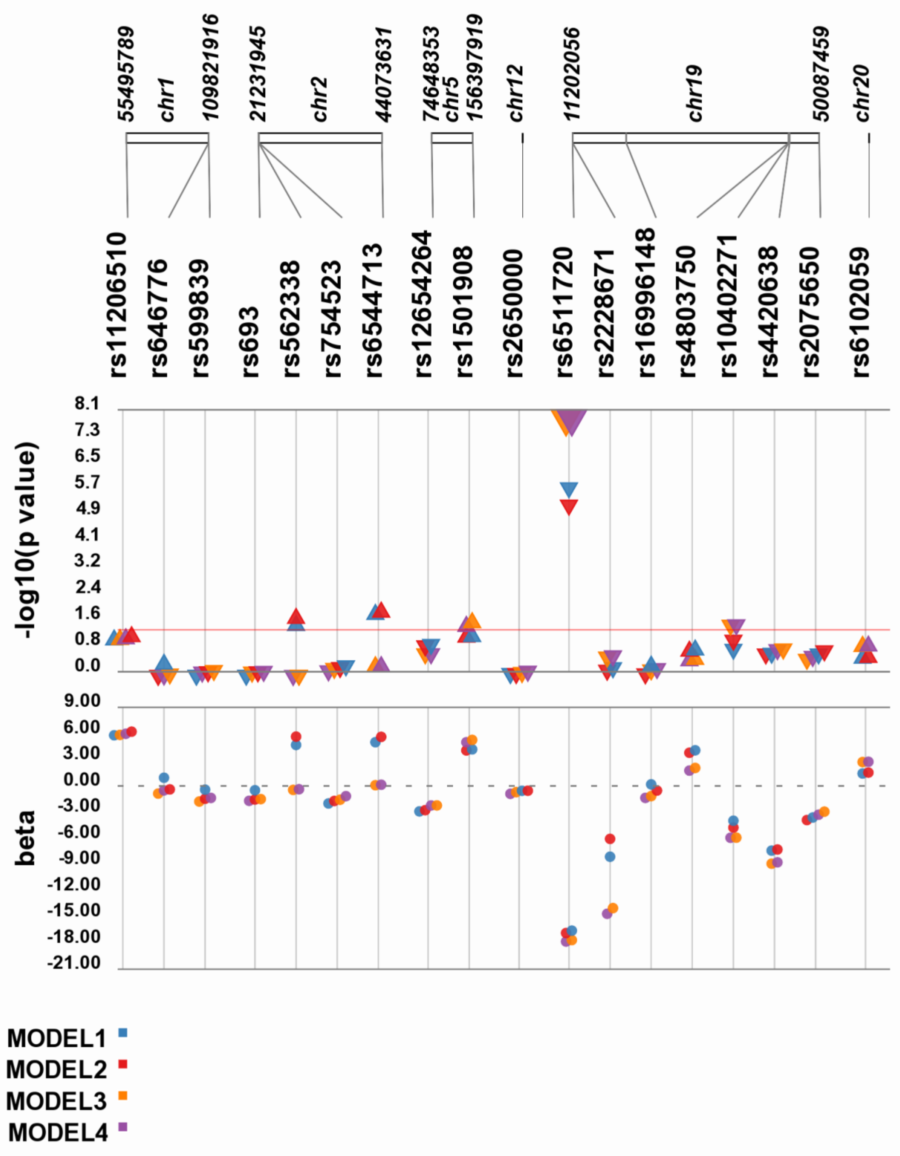
c) American Indians



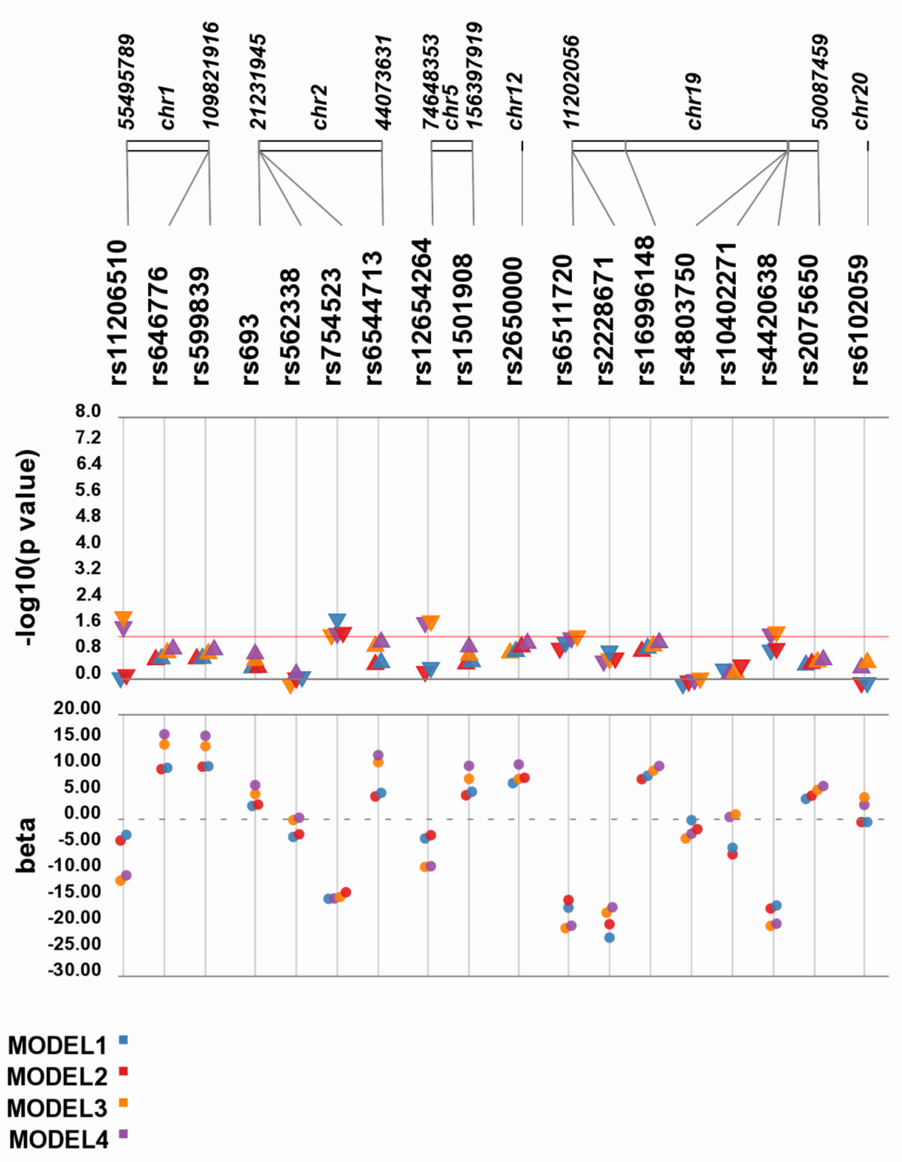
d) Mexican Americans/Hispanics



e) Japanese/East Asians

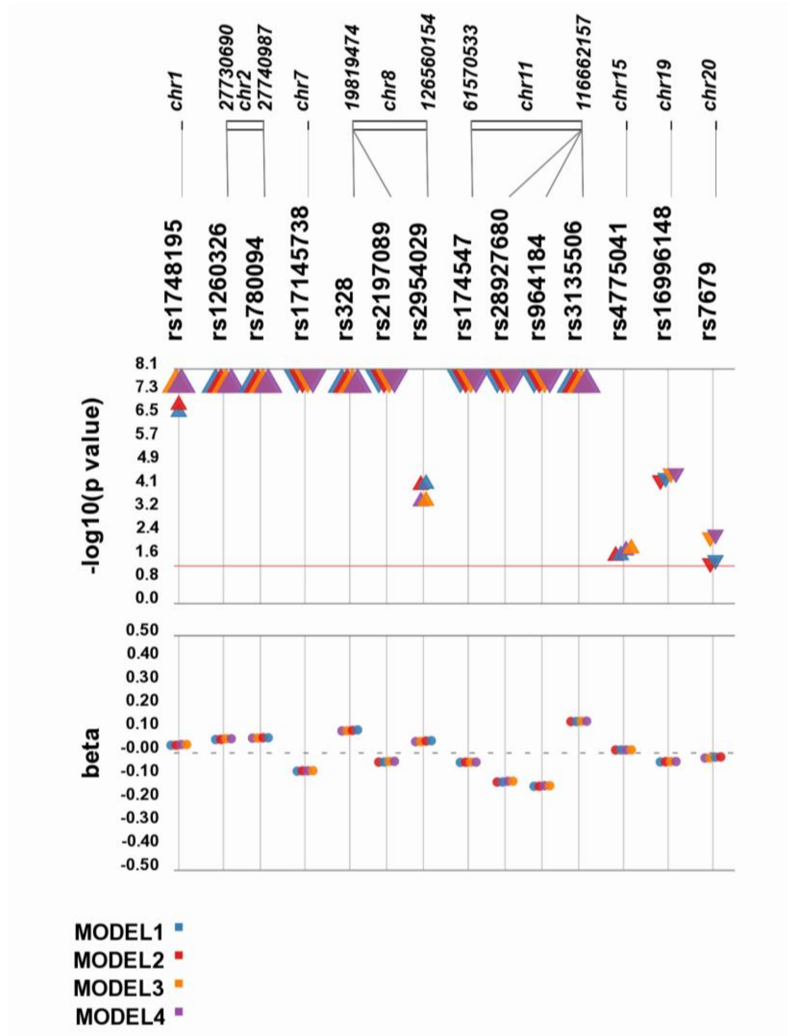


f) Native Hawaiians/Pacific Islanders

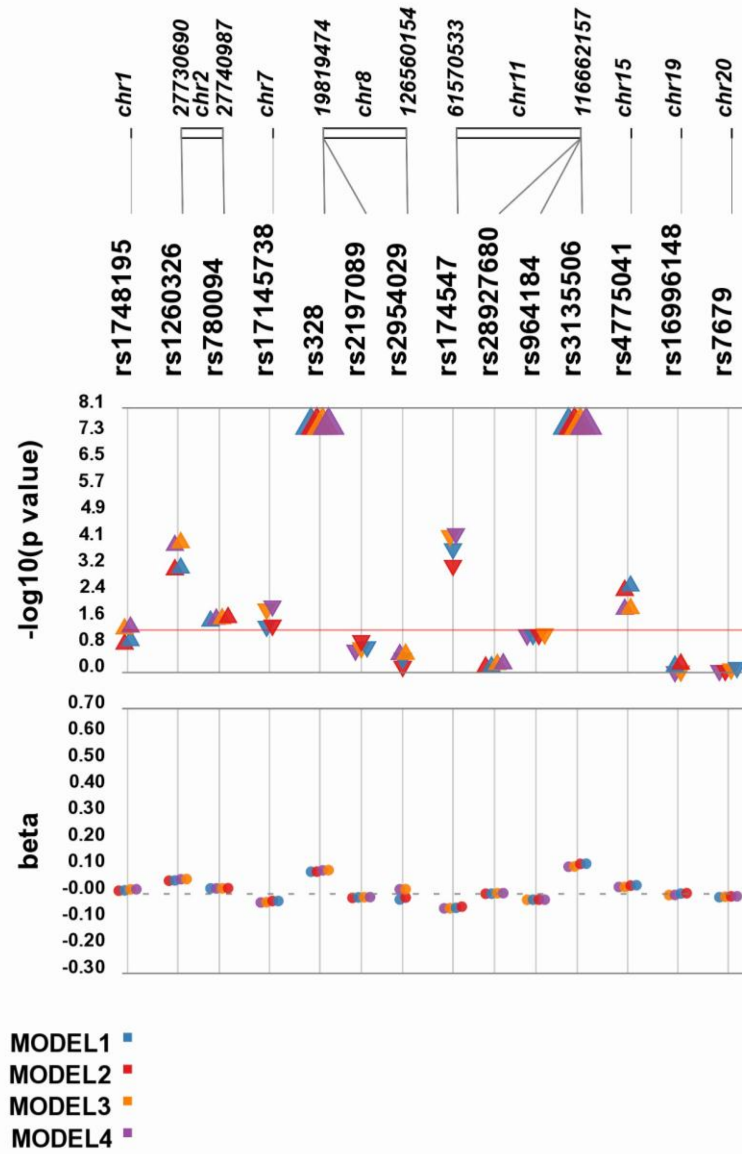


Appendix L. Comparison of unadjusted, minimally adjusted, adjusted models for triglyceride concentrations, by population. Results of tests of association for four regression models are plotted: model 1 (unadjusted), model 2(adjusted for age and sex; and site of ascertainment for select PAGE studies), model 3 (adjusted for age, sex, body mass index, current smoking, type 2 diabetes, post-menopausal status, and current hormone use), and model 4 (model 3 with the addition of previous myocardial infarction). Each SNP was tested for an association with triglycerides. Meta-analysis was performed, and p-values (–$\log_{10}$ transformed) of the meta-analysis are plotted along the y-axis. SNP location is given on the x-axis. Each triangle represents a meta-analysis p-value for each population. Models are color coded. Large triangles represent p-values at or smaller than genome-wide significance ($p<10^{-8}$). The direction of the arrows corresponds to the direction of the beta coefficient. The exact beta coefficients are reported on the bottom panel. The significance threshold is indicated by the red bar at p=0.05.

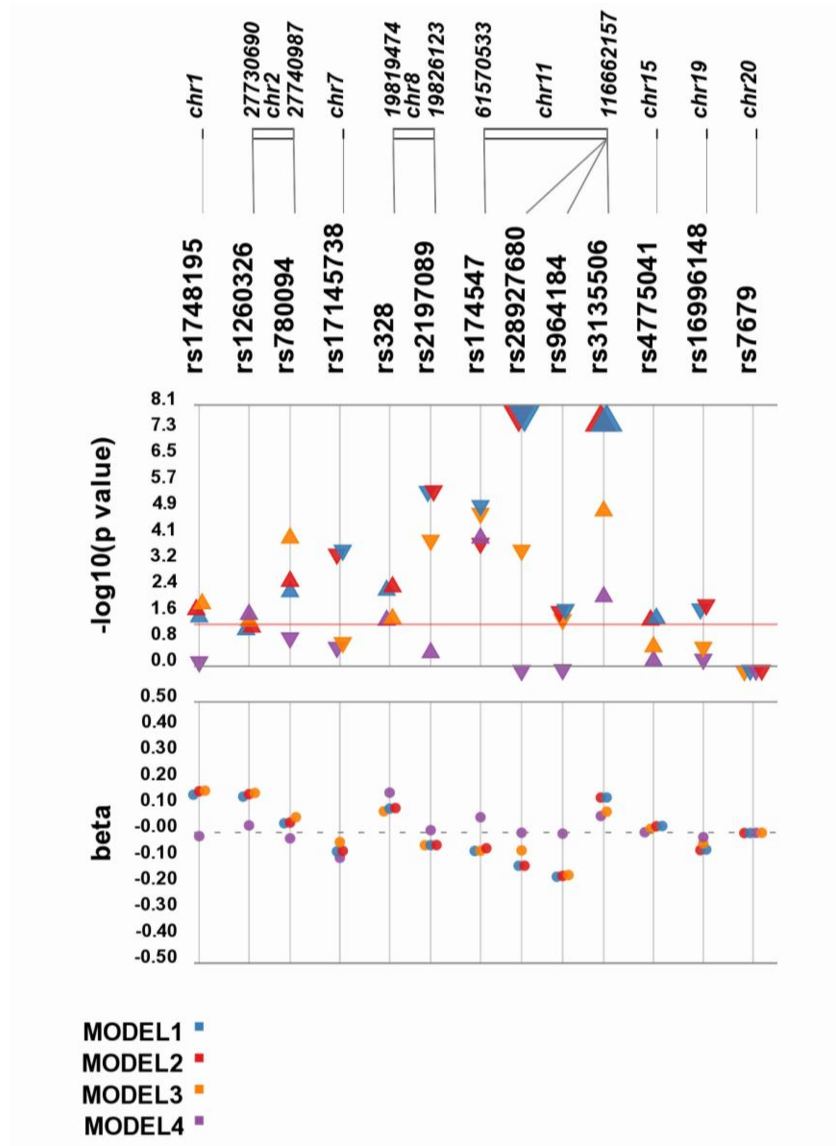
a) **European Americans**



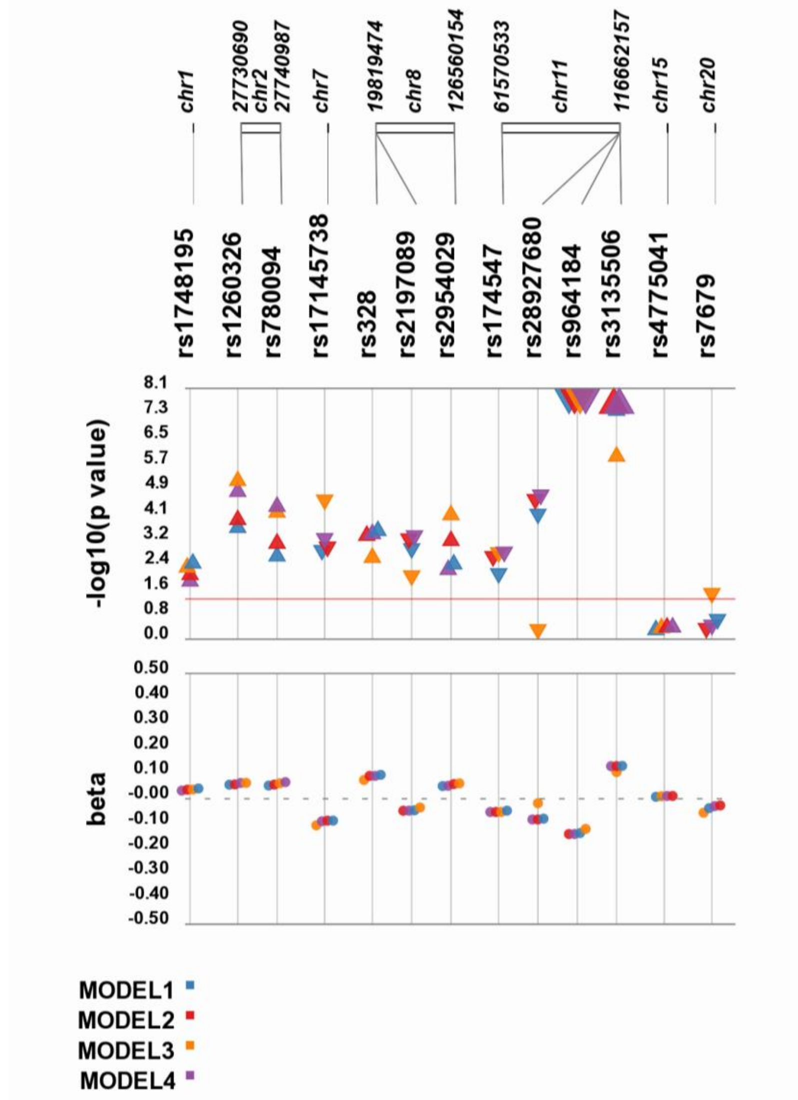
b) African Americans



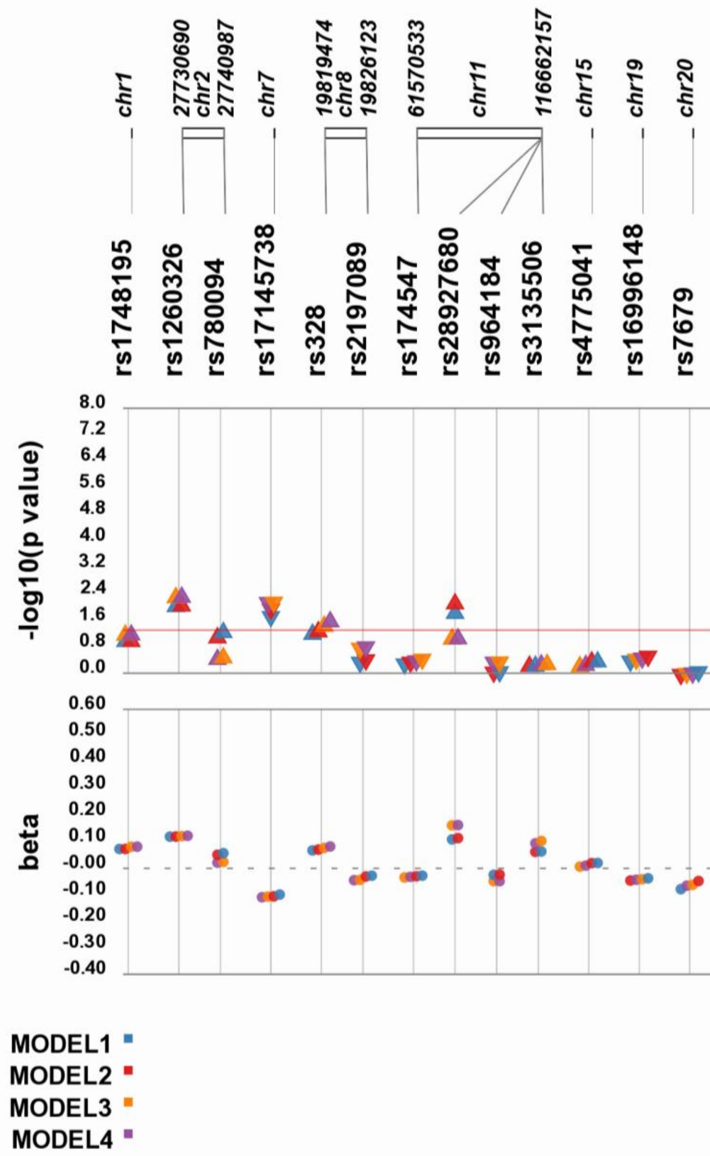
c) American Indians



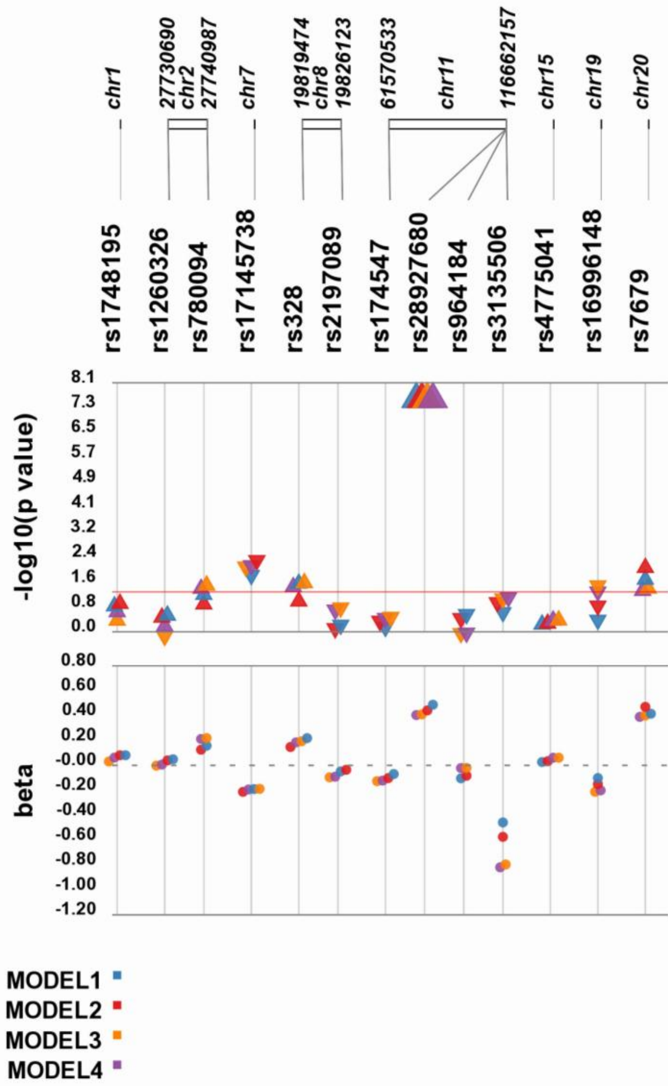
d) Mexican Americans/Hispanics



e) **Japanese/East Asians**

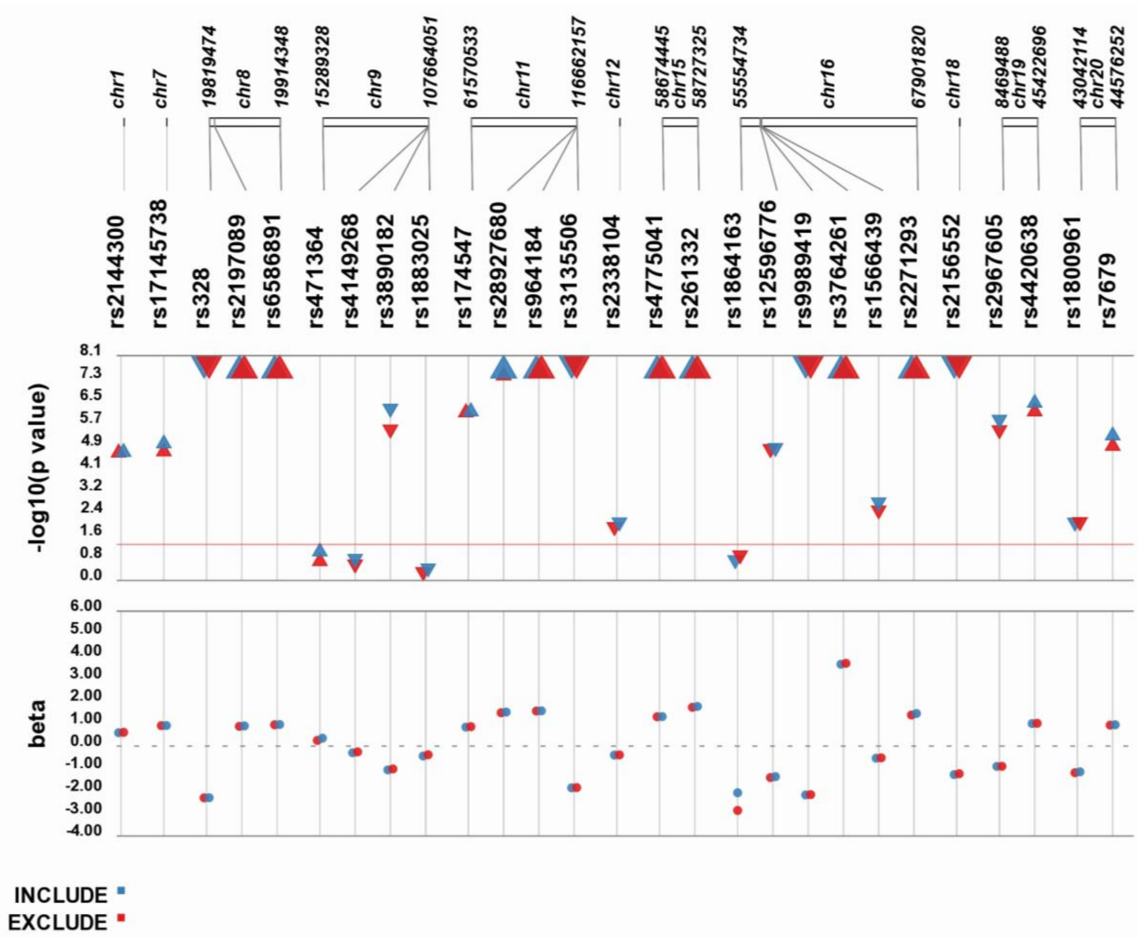


f) Native Hawaiians/Pacific Islanders

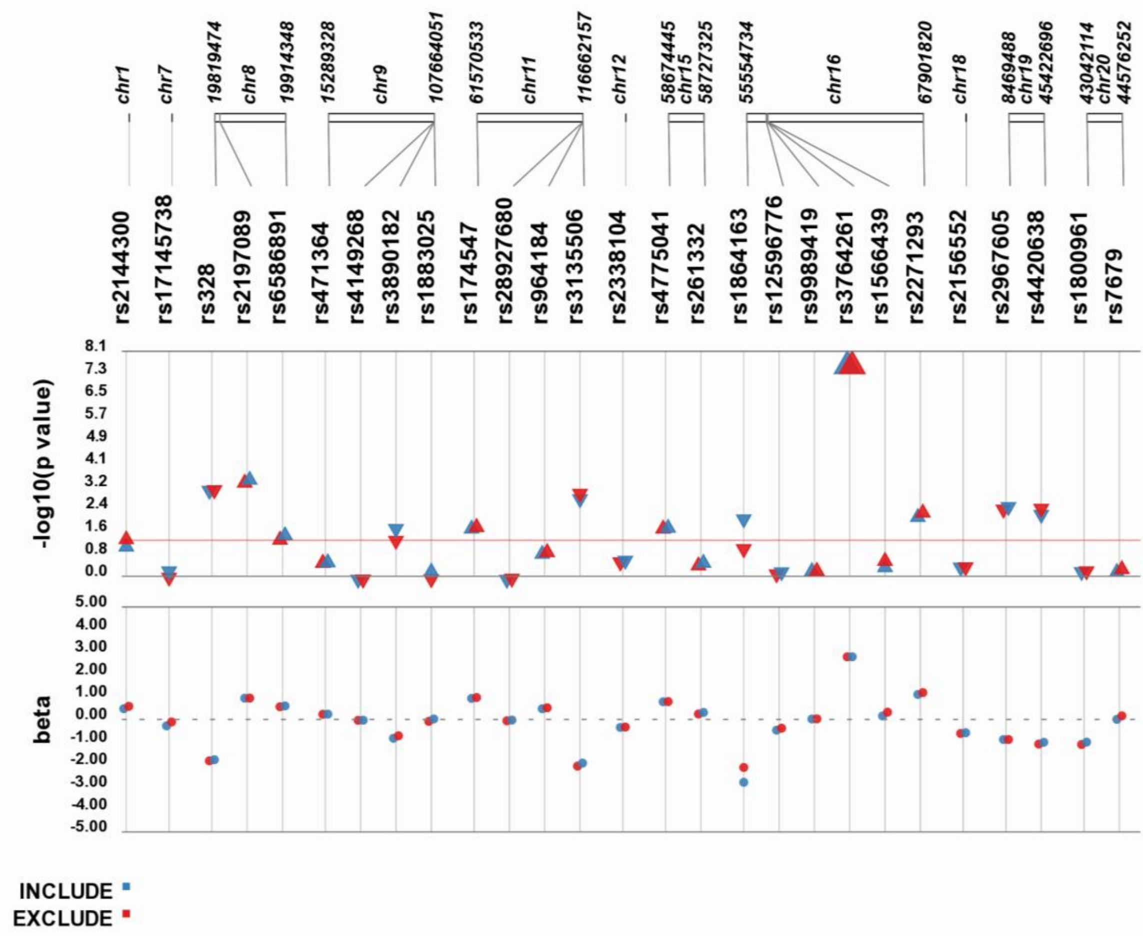


Appendix M. HDL-C and the effects of lipid lowering medication use on genetic associations, by population. Comparison of genetic effects and significance when tests of association are performed within fasting adults regardless of lipid lowering medication (Include) versus fasting adults not on lipid lowering medication (Exclude). All tests of association results shown here are minimally adjusted for age and sex.

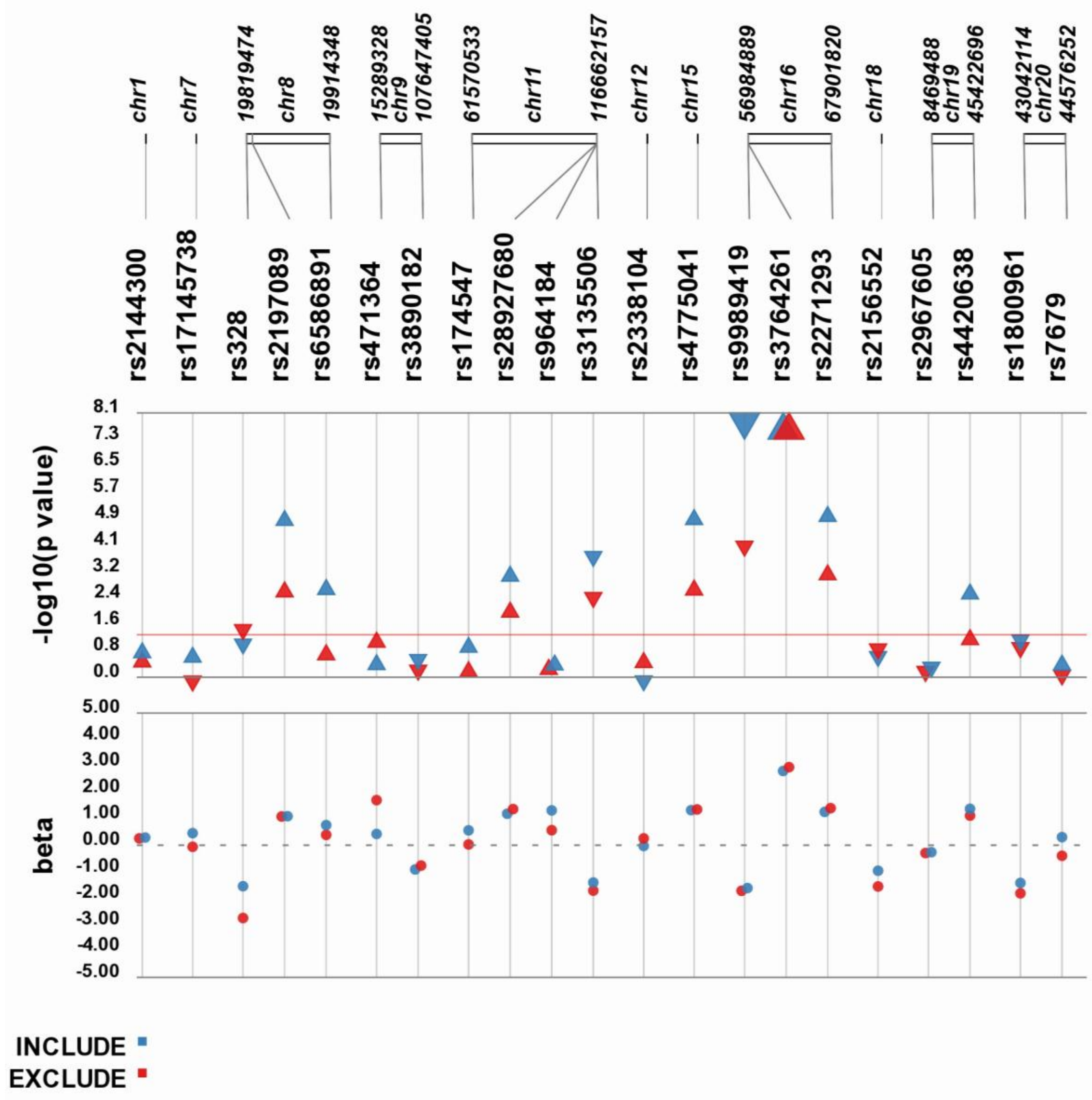
a) **European Americans**



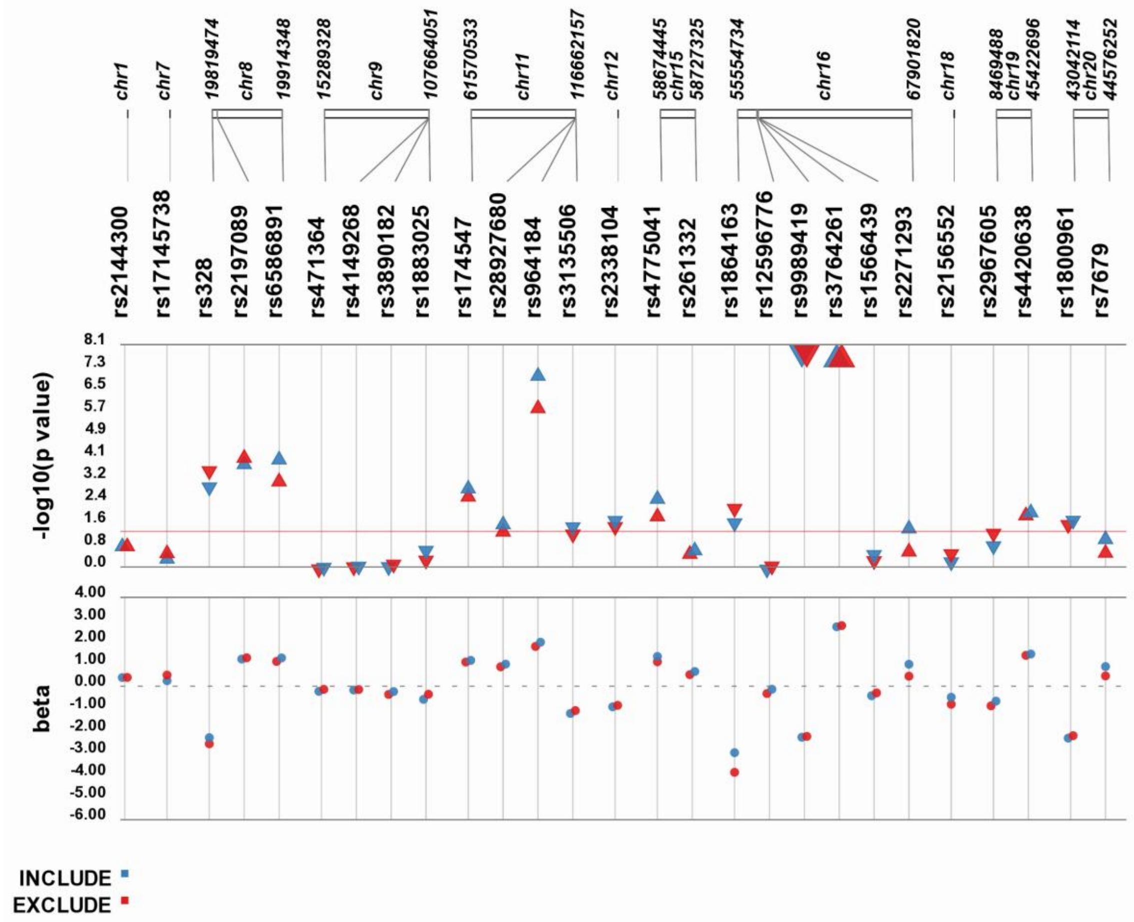
b) African Americans



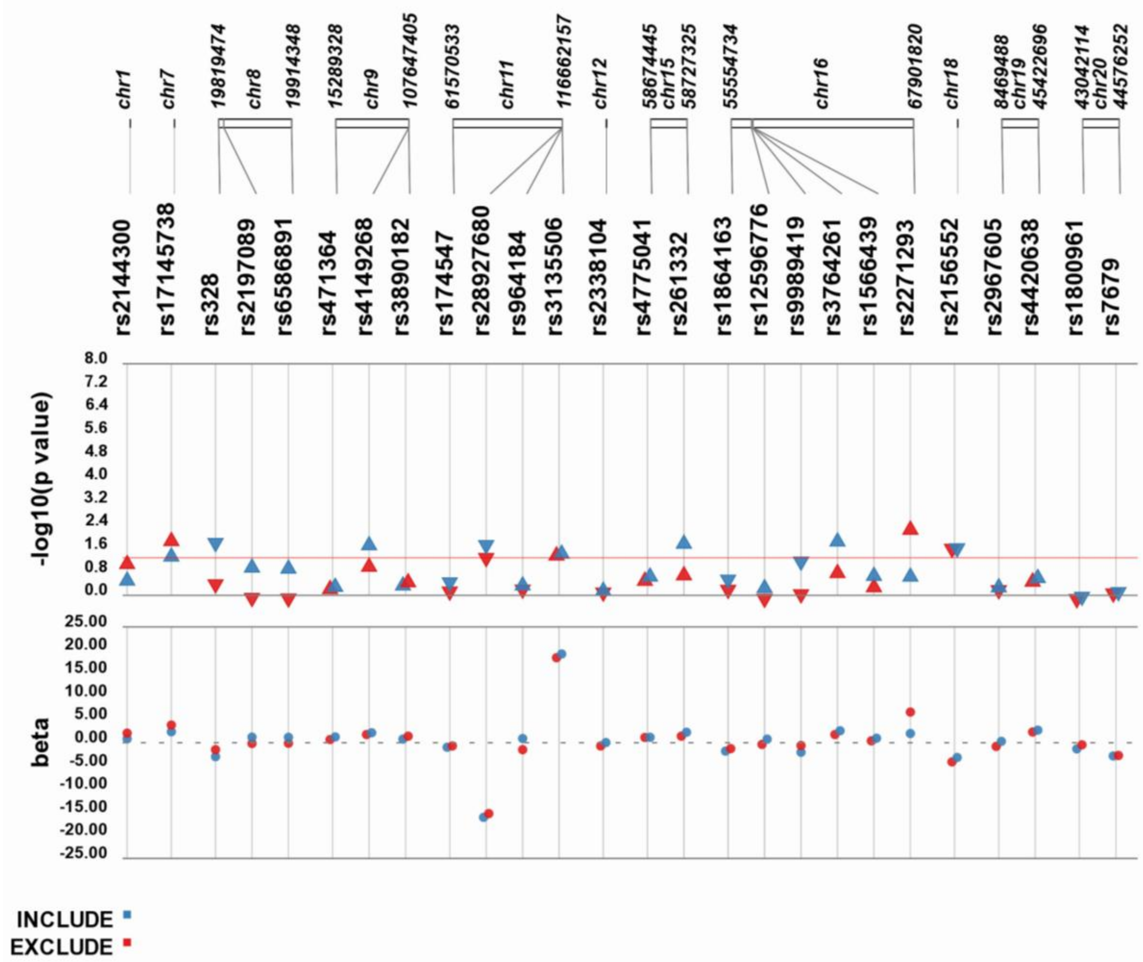
c) **American Indians**



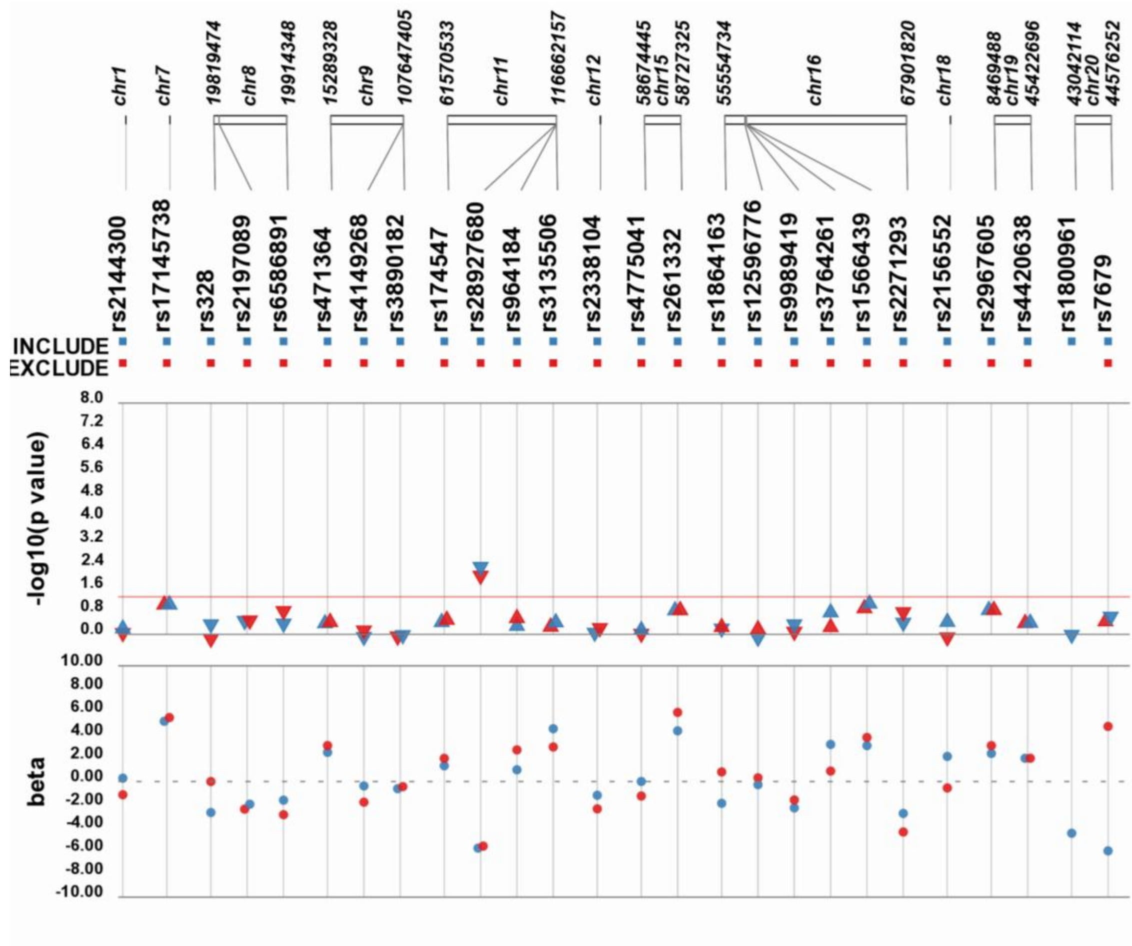
d) **Mexican Americans/Hispanics**



e) Japanese/East Asians

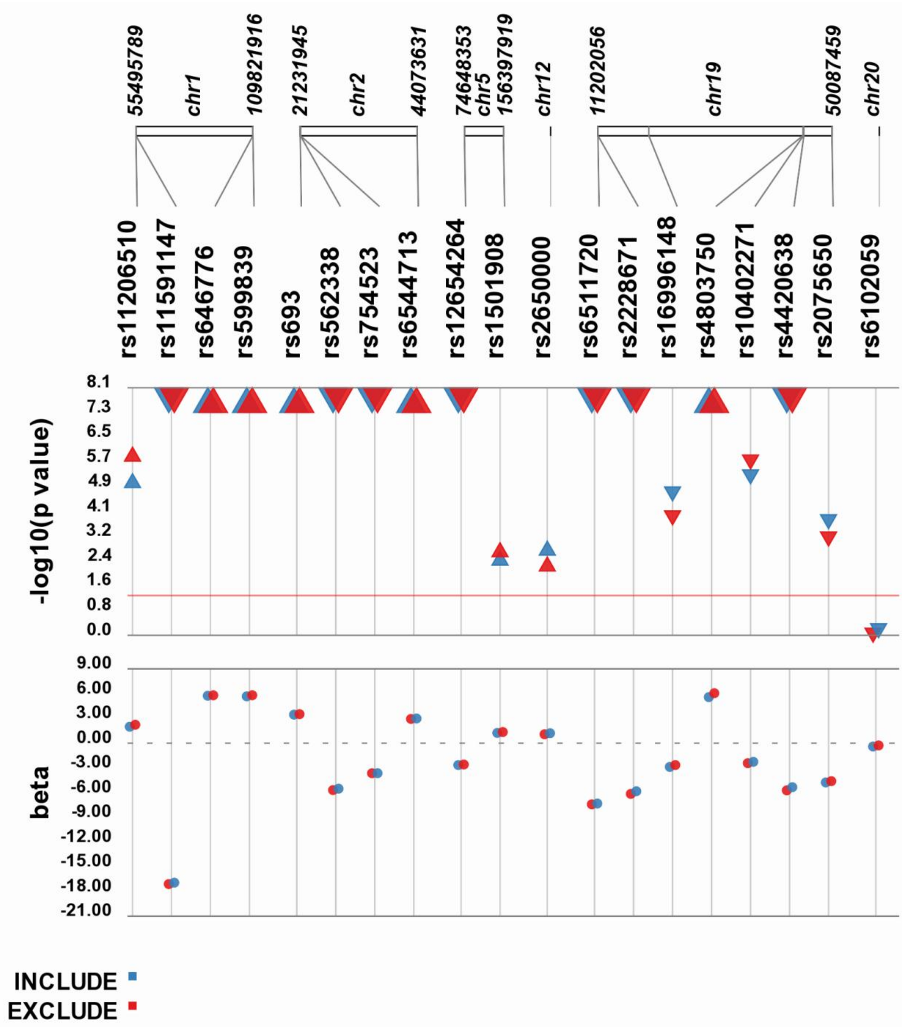


f) **Native Hawaiians/Pacific Islanders**

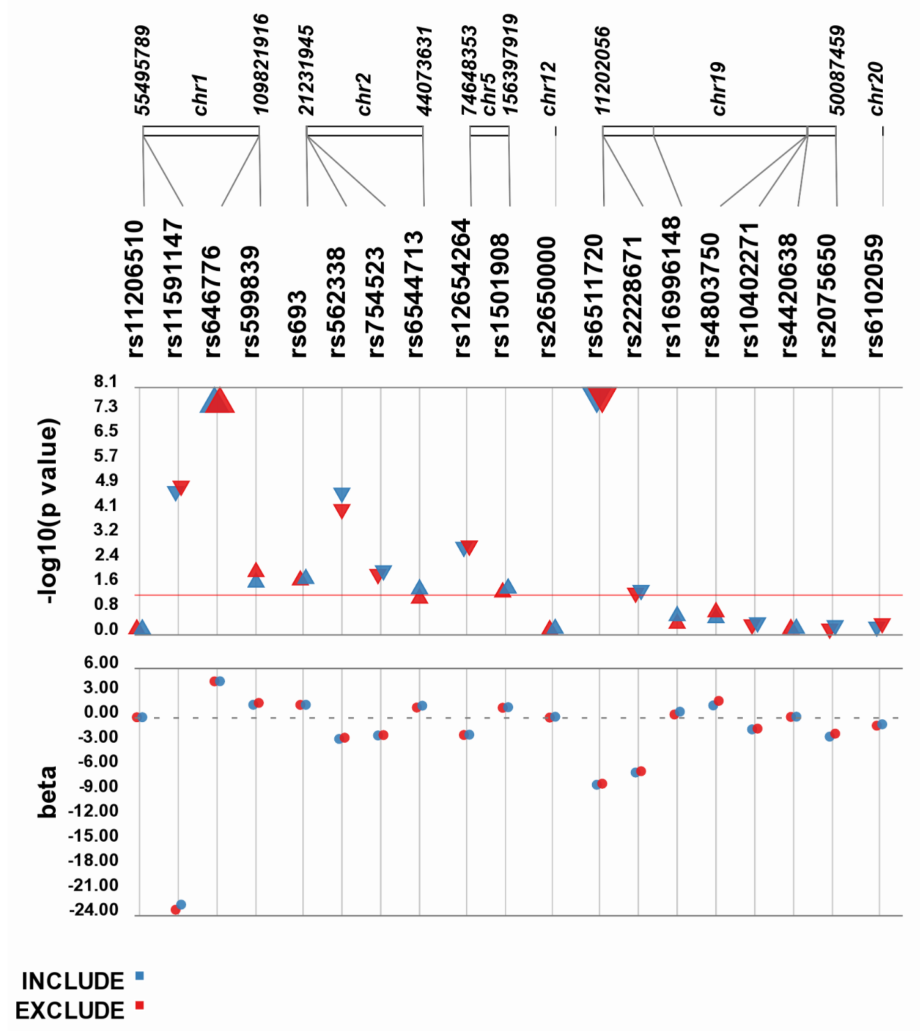


Appendix N. LDL-C and the effects of lipid lowering medication use on genetic associations, by population. Comparison of genetic effects and significance when tests of association are performed within fasting adults regardless of lipid lowering medication versus fasting adults not on lipid lowering medication. All tests of association results shown here are minimally adjusted for age and sex.

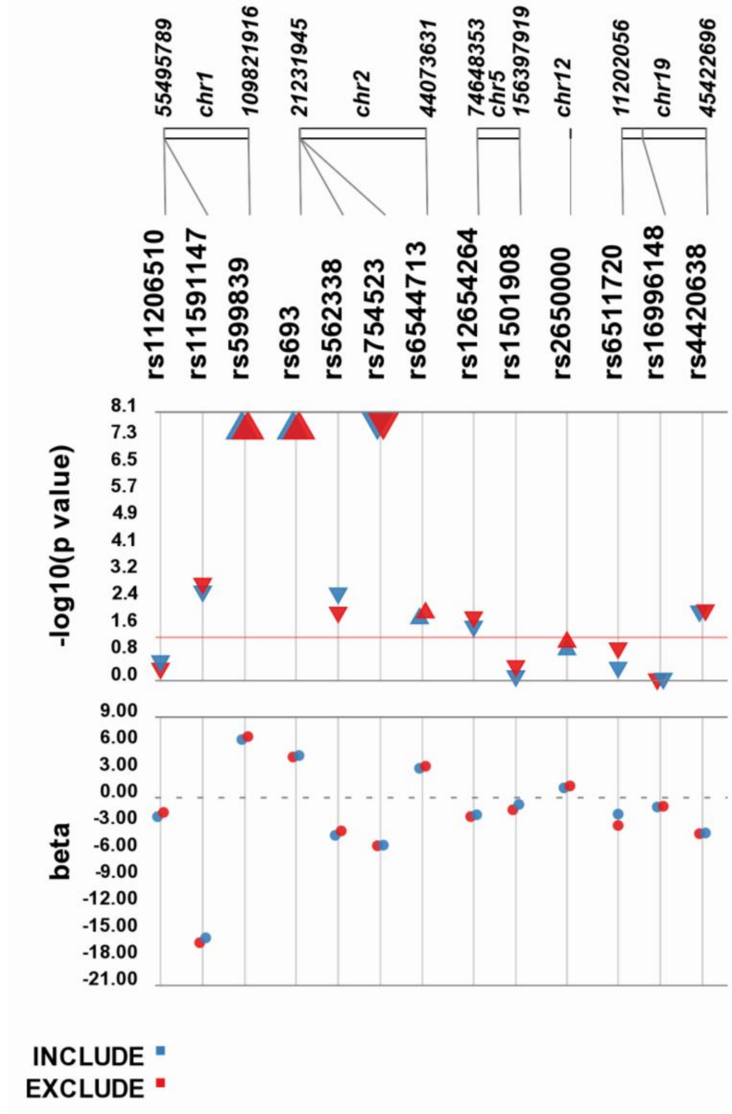
a) European Americans



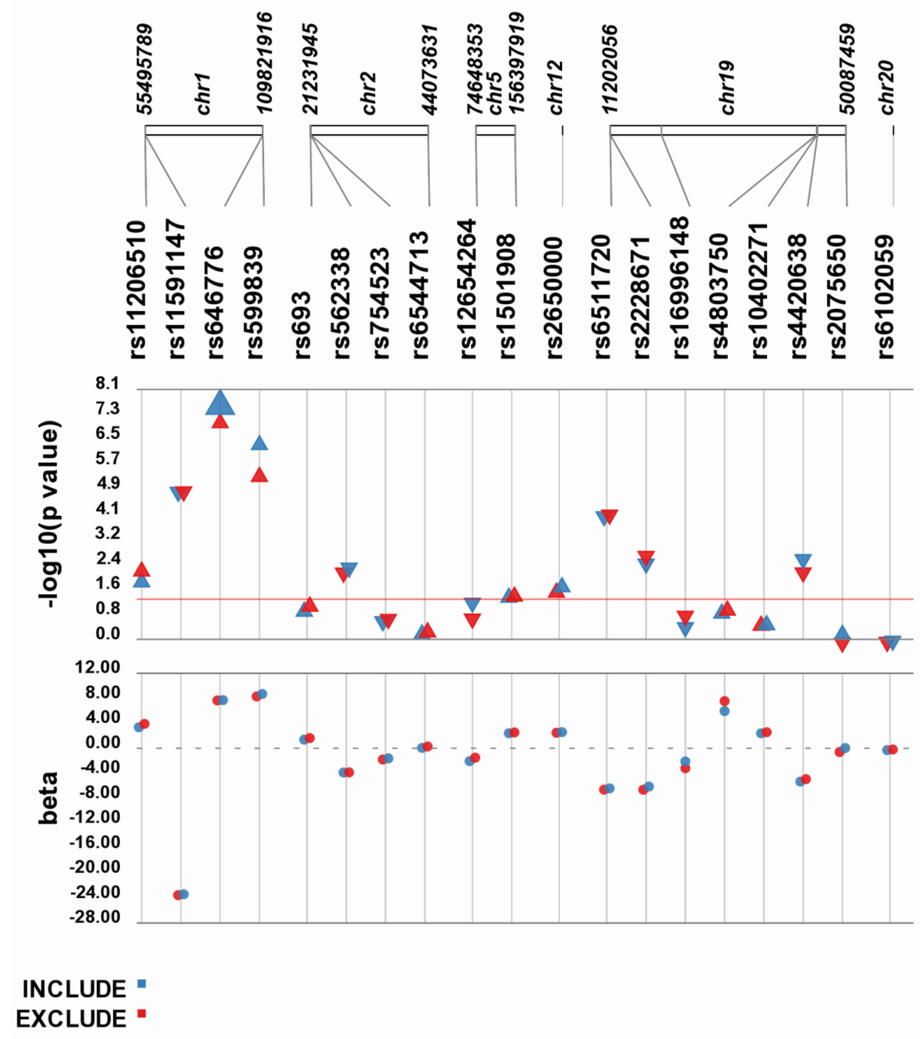
b) African Americans



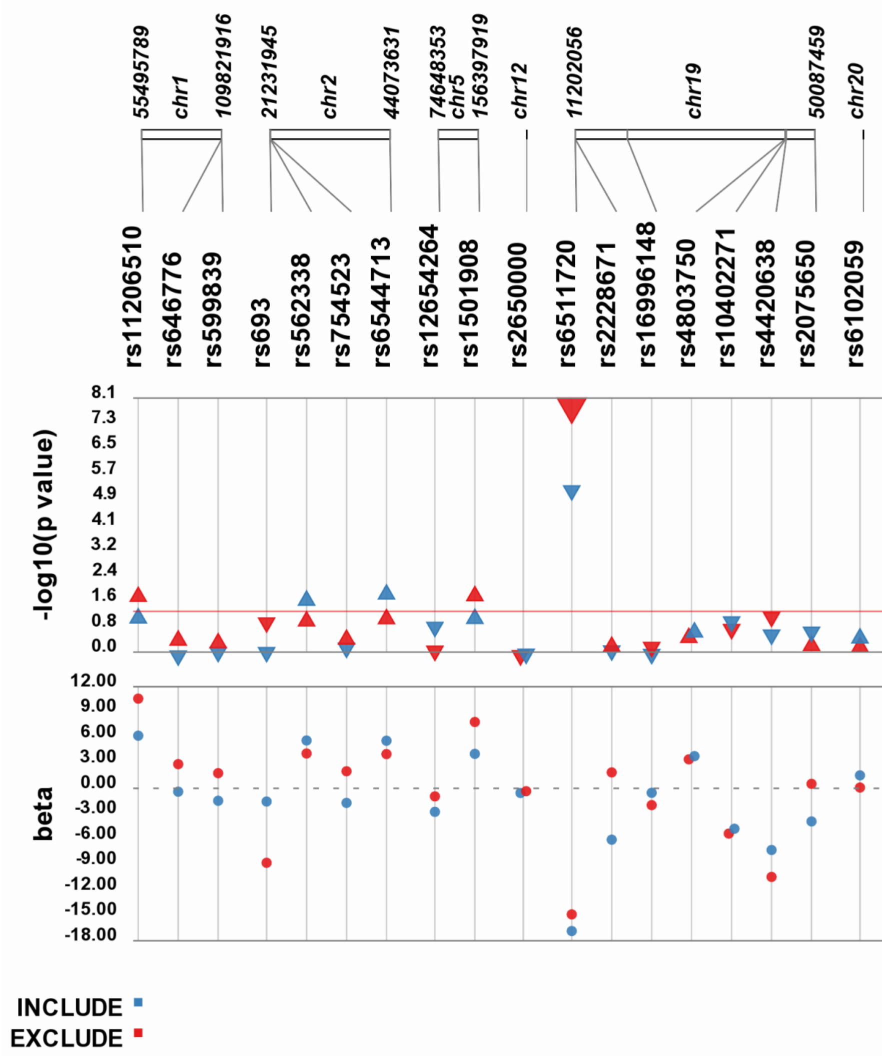
c) **American Indians**



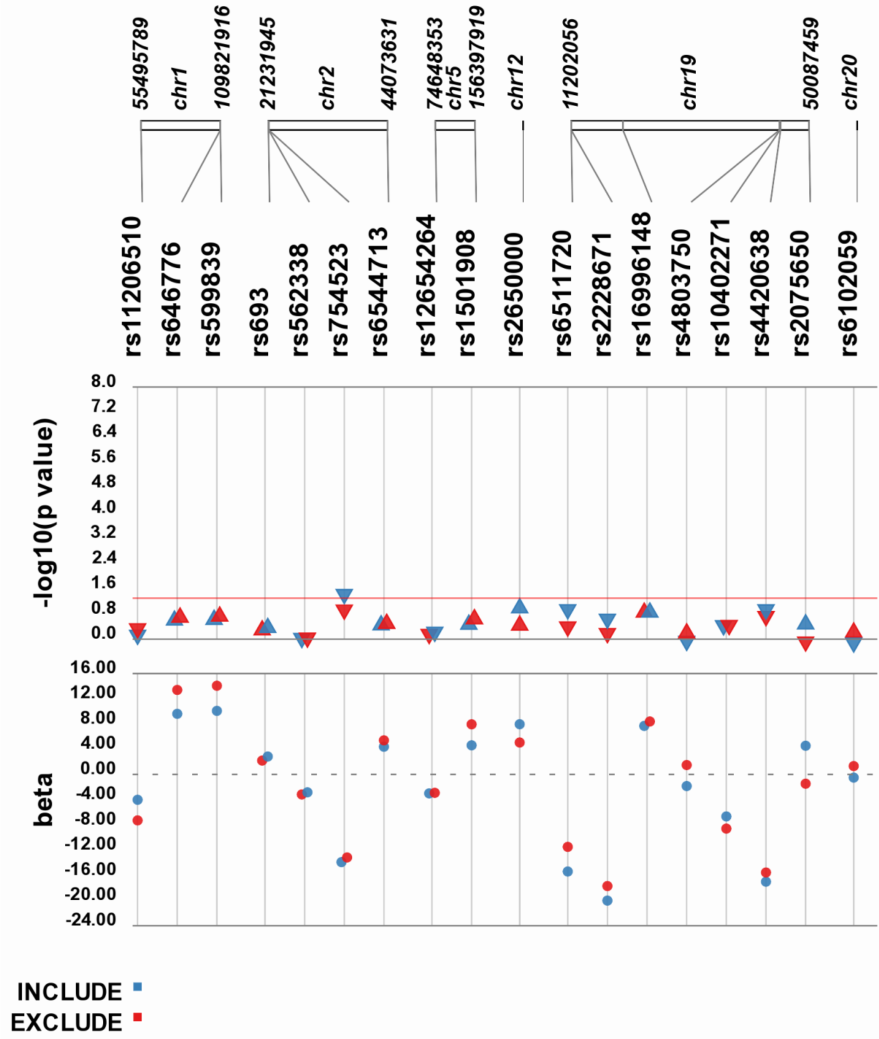
d) Mexican Americans/Hispanics



e) **Japanese/East Asians**

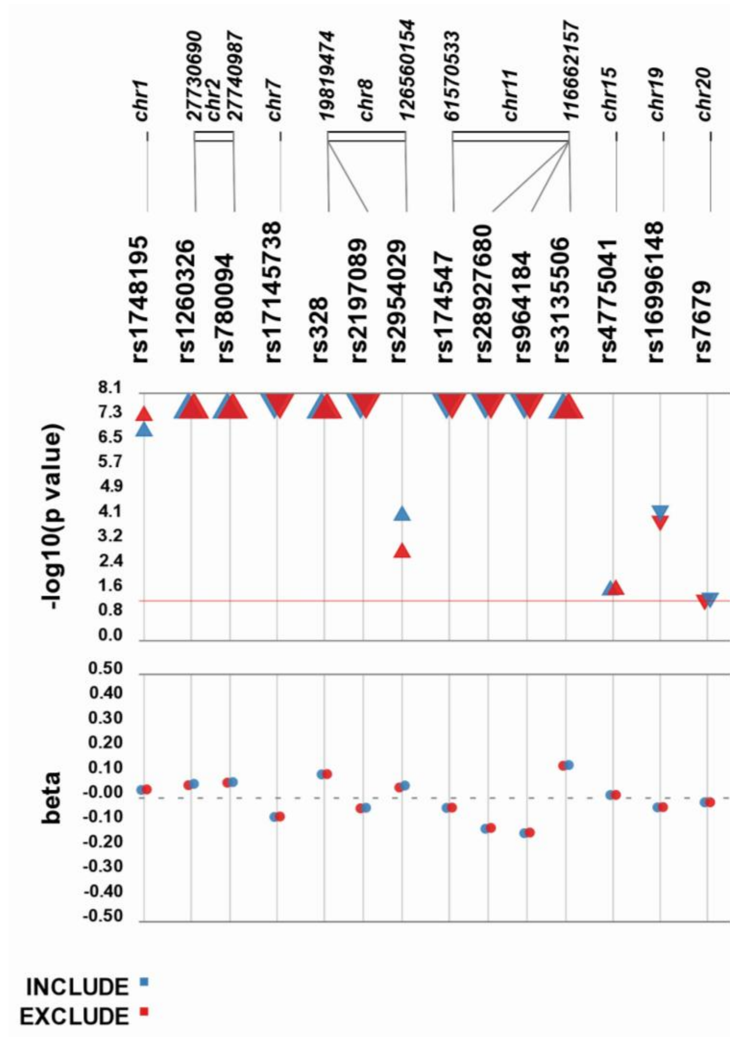


f) Native Hawaiians/Pacific Islanders

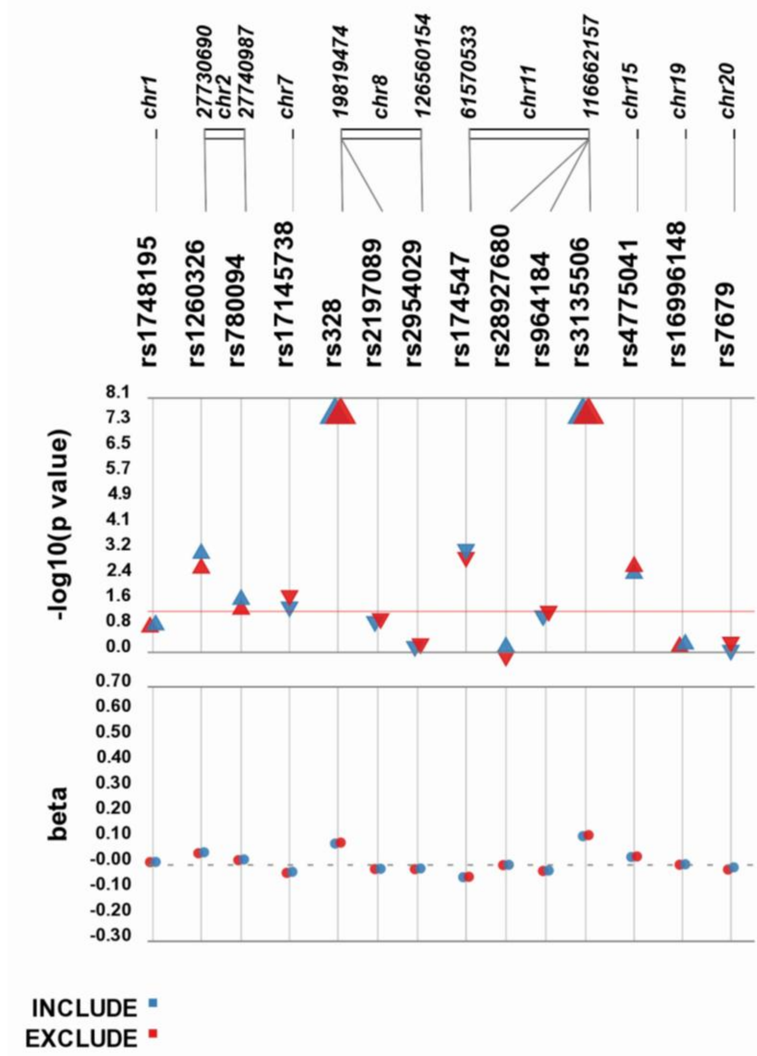


Appendix O. Transformed triglycerides and the effects of lipid lowering medication use on genetic associations, by population. Comparison of genetic effects and significance when tests of association are performed within fasting adults regardless of lipid lowering medication versus fasting adults not on lipid lowering medication. All tests of association results shown here are minimally adjusted for age and sex.

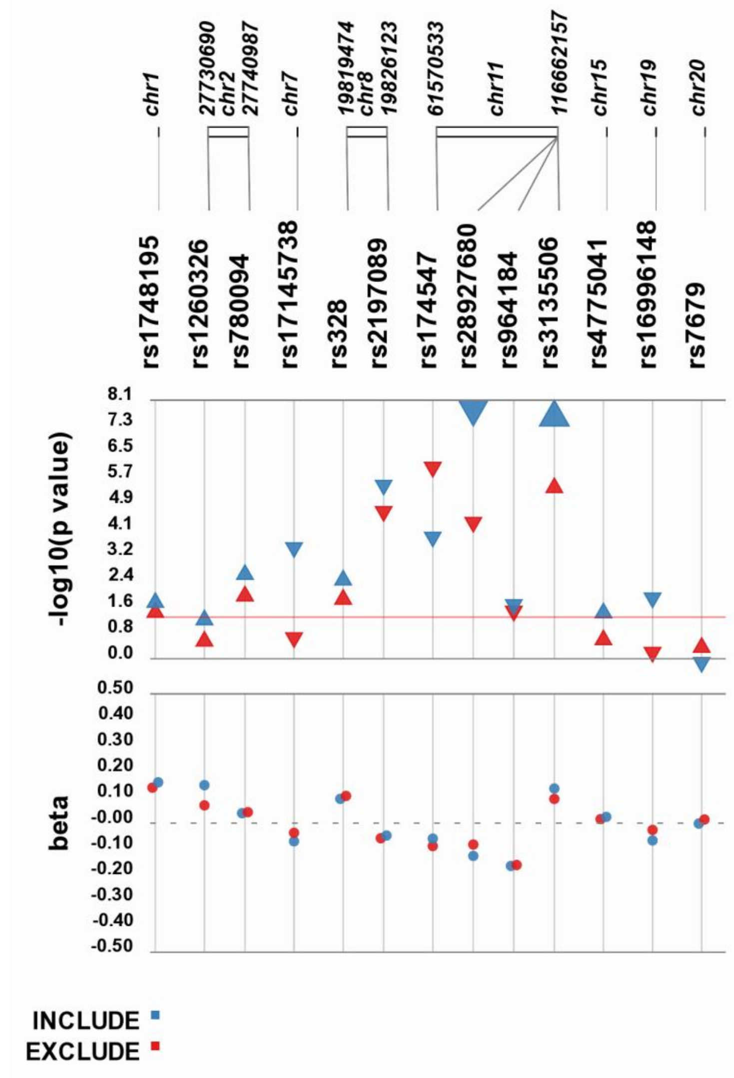
a) **European Americans**



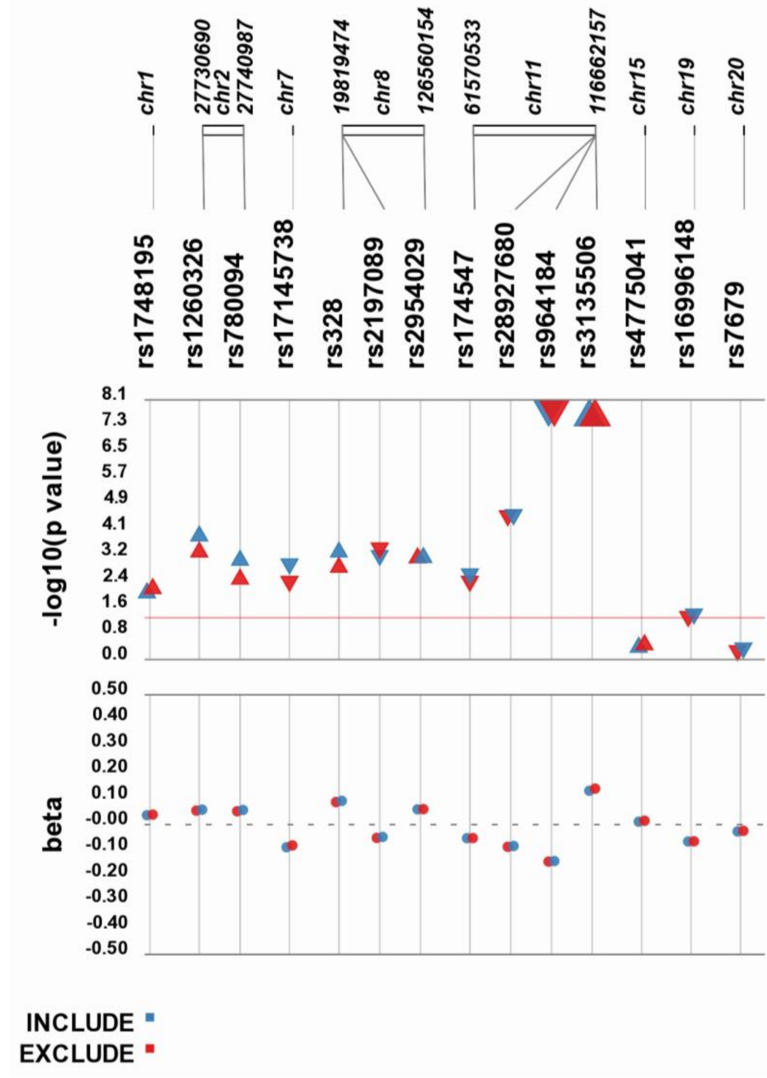
b) African Americans



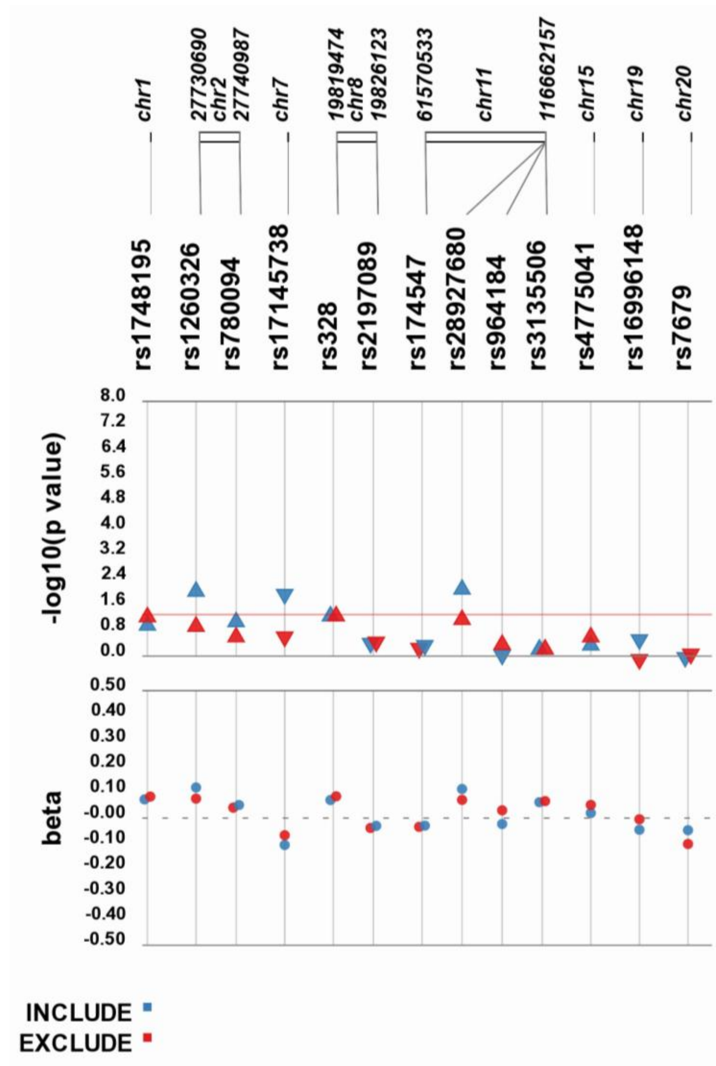
c) **American Indians**



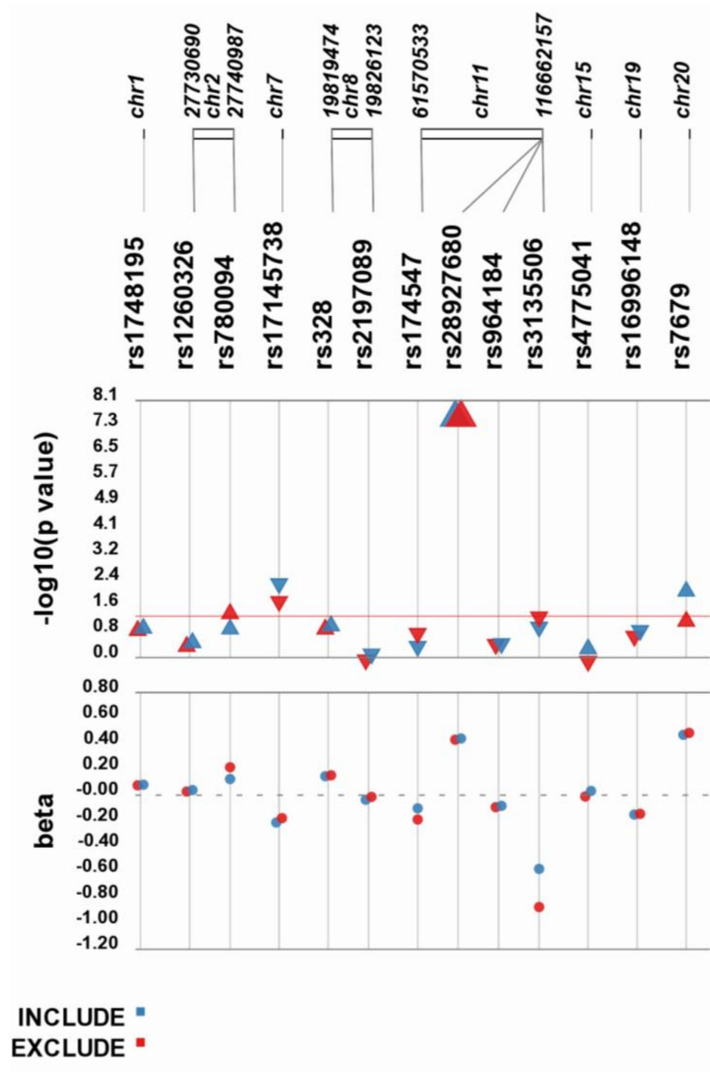
d) Mexican Americans/Hispanics



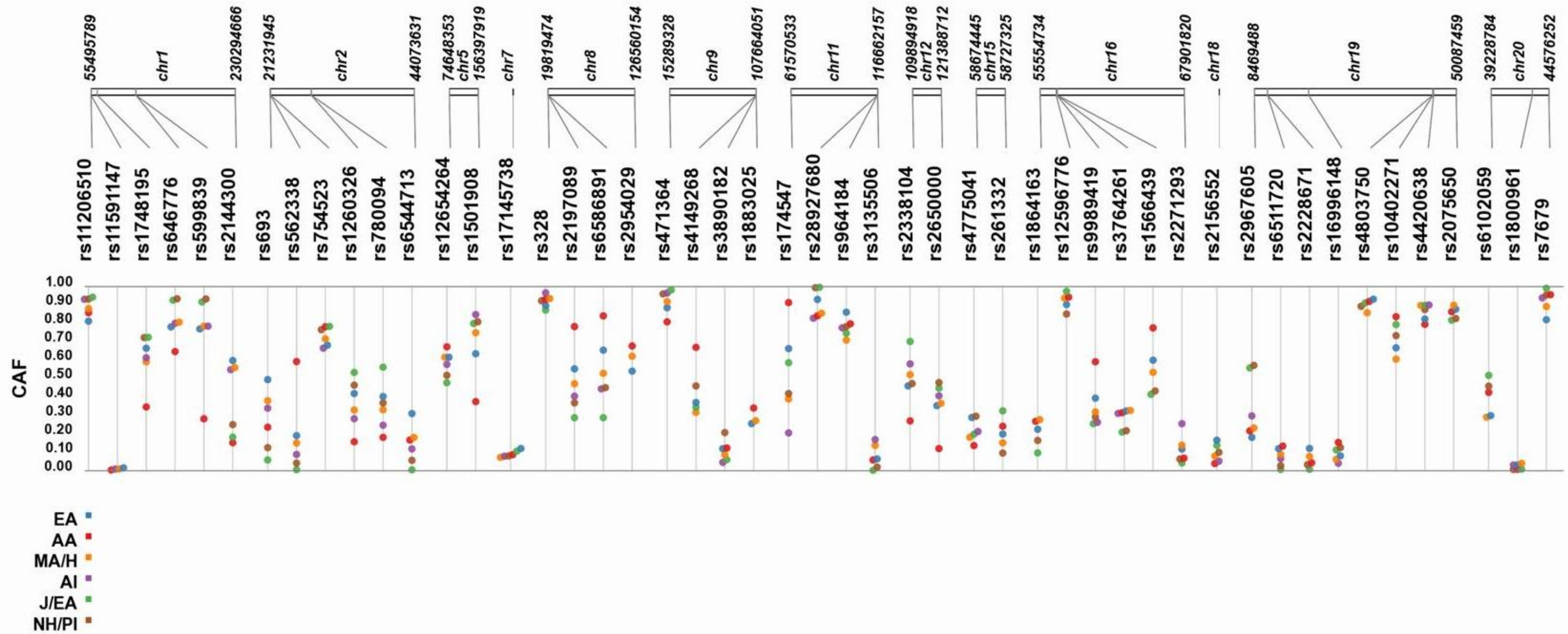
e) Japanese/East Asians



f) Native Hawaiians/Pacific Islanders



Appendix P. Coded allele frequency, by population. The coded allele frequency (CAF) is plotted for each of the 49 SNPs by population using Synthesis-View (Pendergrass et al., 2010). The populations include European Americans (EA), African Americans (AA), Mexican Americans/Hispanics (MA/H), American Indians (AI), Japanese/East Asians (J/EA), and Native Hawaiians/Pacific Islanders (NH/PI).



Appendix Q. Generalization of associations across non-European populations for HDL-C, LDL-C and ln(TG). In the significance column (S), "+" denotes that the SNP was associated with the lipid trait at p<0.05 and "-" denotes a p>0.05. In the effect column (E), "+" denotes that the direction of effect (beta) was in the same direction compared with European Americans and "-" denotes that the directions of effect were discordant. In the power column (P), "+" denotes that the association had at least 80% power to detect the effect size observed in European Americans, based on a p=0.05 and the allele frequency of the studied population. A "-" denotes power<80% and "‡" denotes that the association was significant in the studied population. In the generalize column (G), "+" denotes that the SNP generalized in the studied population based on our criteria (significance AND direction of effect if powered), "-" denotes a lack of generalization, and "?" denotes that we are unable to determine generalization based on power. Generalization here is defined as a significant association (p<0.05) and a similar direction of effect (β) compared with European Americans for the same test of association. Tests of association that generalize across all non-European populations tested are bolded. Abbreviations: Significant (S); Effect (E); Power (P); Generalize (G); Generalize across all populations (G_{all}); and Not applicable (NA).

a) HDL-C

SNP	Gene	African Americans				American Indians				Mexican Americans/Hispanics				G_{all}
		S	E	P	G	S	E	P	G	S	E	P	G	
rs2144300	*GALNT2*	-	+	-	?	-	+	-	?	-	+	-	?	?
rs17145738	*MLXIPL*	-	-	-	?	-	+	-	?	-	+	-	?	?
rs328	*LPL*	+	+	‡	+	-	+	+	-	+	+	‡	+	-
rs2197089	***LPL***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs6586891	***LPL***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs3890182	*ABCA1*	+	+	‡	+	-	+	-	?	-	+	-	?	?
rs174547	*FADS1*	+	+	‡	+	-	+	-	?	+	+	‡	+	?
rs28927680	*APOA1/C3/A4/A5*	-	-	+	-	+	+	‡	+	+	+	‡	+	-
rs964184	*APOA1/C3/A4/A5*	-	+	+	-	-	+	-	?	+	+	‡	+	-
rs3135506	***APOA1/C3/A4/A5***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs2338104	*MMAB-MVK;*	-	+	-	?	-	+	-	?	+	+	‡	+	?
rs4775041	***LIPC***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs261332	*LIPC*	-	+	+	-	NA	NA	NA	NA	-	+	-	?	-

rs12596776	*CETP*	-	+	-	?	NA	NA	NA	NA	-	+	-	?	?
rs9989419	*CETP*	-	-	+	-	+	+	‡	+	+	+	‡	+	-
rs3764261	***CETP***	**+**	**+**	‡	**+**	**+**	**+**	‡	**+**	**+**	**+**	‡	**+**	**+**
rs1566439	*CETP*	-	-	-	?	NA	NA	NA	NA	-	+	-	?	?
rs2271293	*LCAT*	+	+	‡	+	+	+	‡	+	-	+	-	?	?
rs2156552	*LIPG*	-	+	-	?	-	+	-	?	-	+	-	?	?
rs2967605	*ANGPTL4*	+	+	‡	+	-	+	+	-	-	+	-	?	-
rs4420638	*APOE/C1/C4*	+	-	‡	-	+	+	‡	+	+	+	‡	+	-
rs1800961	*HNF4A*	-	+	-	?	-	+	-	?	+	+	‡	+	?
rs7679	*PLTP*	-	+	-	?	-	+	-	?	-	+	‡	?	?

b) LDL-C

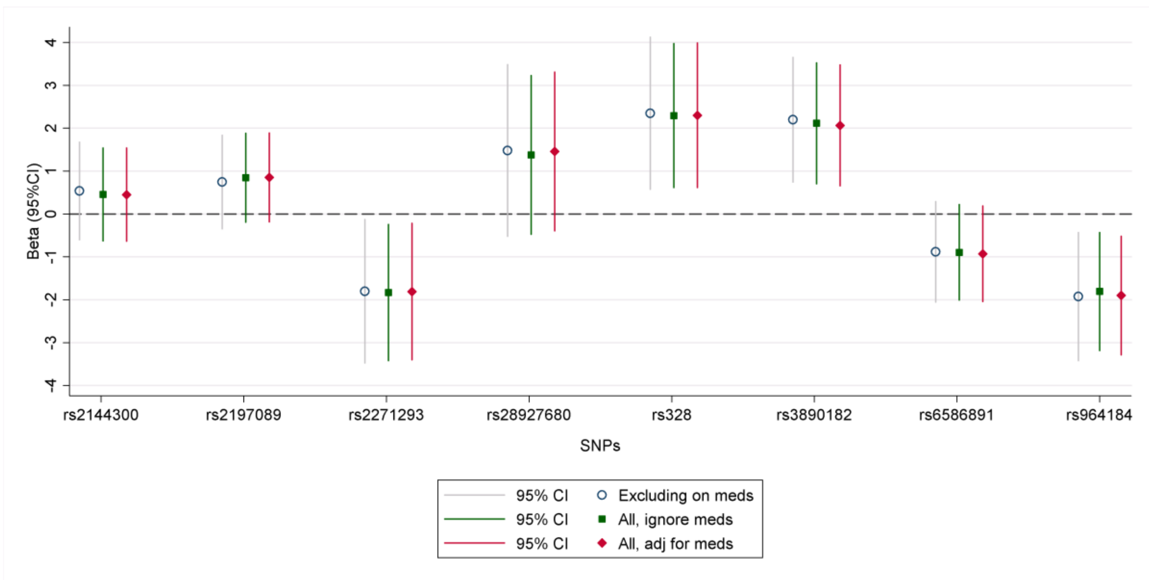
SNP	Gene	African Americans				American Indians				Mexican Americans/Hispanics				G_{all}
		S	E	P	G	S	E	P	G	S	E	P	G	
rs11206510	*PCSK9*	-	+	-	?	-	-	-	?	+	+	‡	+	?
rs11591147	***PCSK9***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs646776	***CELSR2/PSRC1/SORT***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs599839	***CELSR2/PSRC1/SORT***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs693	*APOB*	+	+	‡	+	+	+	‡	+	-	+	+	-	-
rs562338	***APOB***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs754523	*APOB*	+	+	‡	+	+	+	‡	+	-	+	+	-	-
rs6544713	*ABCG8*	+	+	‡	+	+	+	‡	+	-	+	-	?	?
rs12654264	***HMGCR***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs1501908	*TIMD4*	+	+	‡	+	+	-	‡	-	-	+	-	?	-
rs2650000	*HNF1A*	-	+	-	?	-	+	-	?	+	+	‡	+	?
rs6511720	*LDLR*	+	+	‡	+	-	+	+	-	+	+	‡	+	-
rs2228671	***LDLR***	**+**	**+**	**‡**	**+**	**NA**	**NA**	**NA**	**NA**	**+**	**+**	**‡**	**+**	**+**
rs16996148	*CILP2/PBX4*	-	-	+	-	-	+	-	?	-	+	-	?	-
rs4803750	*BCL3*	-	+	+	-	NA	NA	NA	NA	-	+	-	?	-
rs10402271	*APOE/C1/C4*	-	+	-	?	NA	NA	NA	NA	-	-	-	?	?
rs4420638	*APOE/C1/C4*	-	-	+	-	+	+	‡	+	+	+	‡	+	-
rs2075650	*TOMM40*	-	+	-	?	NA	NA	NA	NA	-	-	-	?	?

c) ln(TG)

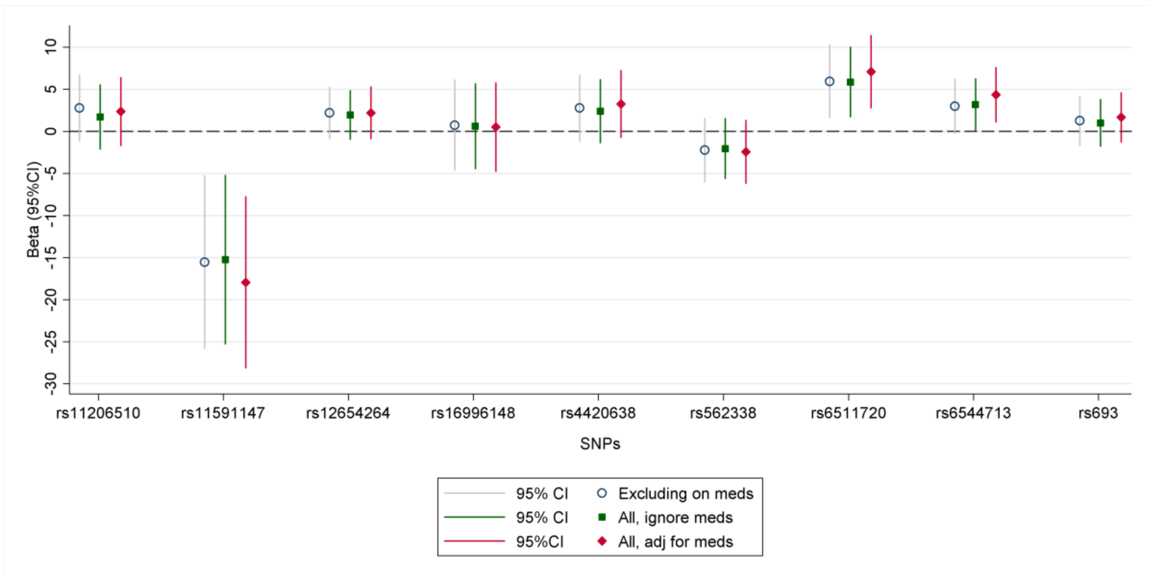
SNP	Gene	African Americans				American Indians				Mexican Americans/Hispanics				G_{all}
		S	E	P	G	S	E	P	G	S	E	P	G	
rs1748195	*ANGPTL3*	-	+	+	-	+	+	‡	+	+	+	‡	+	-
rs1260326	*GCKR*	+	+	‡	+	-	+	-	?	+	+	‡	+	?
rs780094	***GCKR***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs17145738	***MLXIPL***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs328	***LPL***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs2197089	*LPL*	-	+	+	-	+	+	‡	+	+	+	‡	+	-
rs2954029	*TRIB1*	-	-	+	-	NA	NA	NA	NA	+	+	‡	+	-
rs174547	***FADS1***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs28927680	*APOA1/C3/A4/A5*	-	-	+	-	+	+	‡	+	+	+	‡	+	-
rs964184	***APOA1/C3/A4/A5***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs3135506	***APOA1/C3/A4/A5***	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**	**+**	**‡**	**+**	**+**
rs4775041	*LIPC*	+	+	‡	+	-	+	-	?	-	+	-	?	?
rs16996148	*CILP2/PBX4*	-	-	-	?	+	+	‡	+	+	+	‡	+	?
rs7679	*PLTP*	-	+	-	?	-	+	-	?	-	+	-	?	?

Appendix R. Comparison of genetic effect estimates when participants are excluded or included based on medication use with adjustments in WHI. Genetic effect estimates (β) and 95% confidence interval are plotted for each SNP tested for an association. The tests of association were performed on fasting European Americans adjusted for age and sex and excluding participants on lipid lowering medication (blue), including all participants regardless of medication use (green), and all participants on lipid lowering medication, adjusted for the average HDL-C, LDL-C, and ln(TG) effects estimated by Wu et al (Wu et al., 2007).

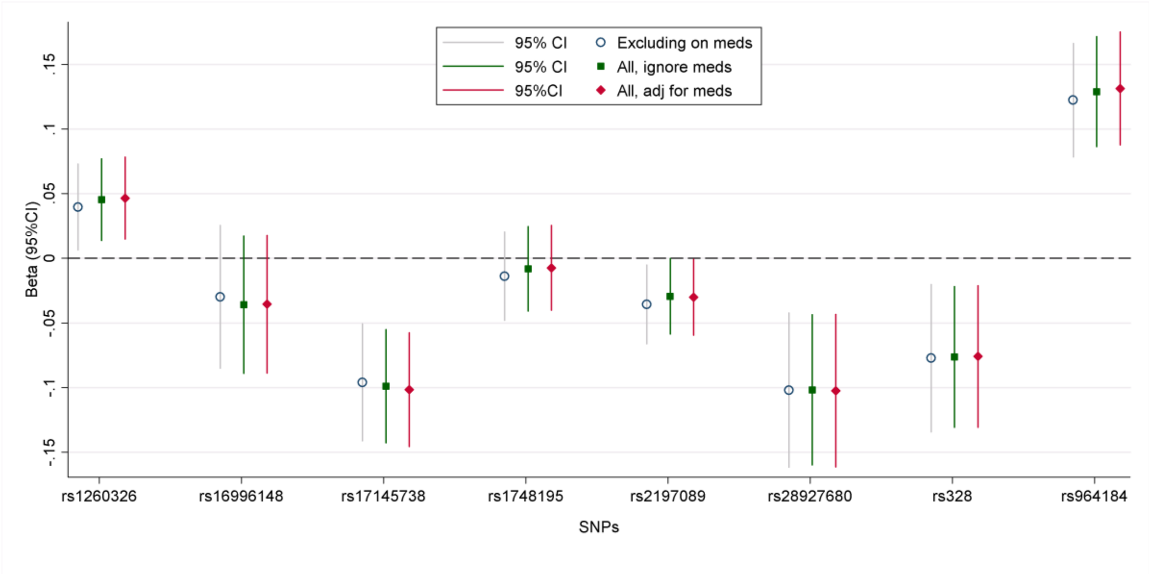
a) HDL-C



b) LDL-C



c) ln(TG)



Appendix Table S. List of 23 candidate gene and GWAS-identified SNPs genotyped in EAGLE. For each SNP (denoted by rs number), we list the chromosomal and genomic location, the putative function of the SNP (based on SNP location) and the nearest gene, the number of PAGE studies that genotyped the SNP, the trait associated with the SNP based on the literature, the effect allele and effect size based on the literature, and the reference for these data. Abbreviations: Base-pair (bp), Untranslated region (UTR).

SNP	Chr.	Build 37 location (bp)	Function	Nearest Gene	Previously Associated Trait	Reference
rs11206510	1	55495789	Intergenic	*PCSK9*	LDL-C	Willer et al 2008
rs1748195	1	63049343	Intronic	*ANGPTL3*	TG	Willer et al 2008
rs693	2	21231945	Synonymous	*APOB*	LDL-C	Willer et al 2008
rs754523	2	21311441	Intergenic	*APOB*	LDL-C	Willer et al 2008
rs780094	2	27740987	Intronic	*GCKR*	TG	Willer et al 2008
rs12654264	5	74648353	Intronic	*HMGCR*	LDL-C	Kathiresan et al 2008
rs1501908	5	156397919	Intergenic	*TIMD4*	LDL-C	Kathiresan et al 2009
rs2197089	8	19826123	Downstream	*LPL*	HDL-C	Willer et al 2008
					TG	Willer et al 2008
rs2954029	8	126560154	Intergenic	*TRIB1*	TG	Teslovich et al 2010
rs4149268	9	107647220	Intronic	*ABCA1*	HDL-C	Willer et al 2008
rs3890182	9	107647405	Intronic	*ABCA1*	HDL-C	Kathiresan et al 2008
rs1883025	9	107664051	Intronic	*ABCA1*	HDL-C	Teslovich et al 2010
rs174547	11	61570533	Intronic	*FADS1*	HDL-C	Kathiresan et al 2009
					TG	Kathiresan et al 2009
rs3135506	11	116662157	Non-synonymous	*APOA1/C3/A4/A5* gene cluster	HDL-C	Lu et al 2008
					TG	Ariza et al 2010
rs2338104	12	109894918	Intronic	*MMAB*-MVK	HDL-C	Willer et al 2008
rs4775041	15	58674445	Intergenic	*LIPC*	HDL-C	Willer et al 2008
					TG	Willer et al 2008
rs9989419	16	56984889	Upstream	*CETP*	HDL-C	Willer et al 2008
rs3764261	16	56993074	Upstream	*CETP*	HDL-C	Willer et al 2008
rs2271293	16	67901820	Intronic	*LCAT*	HDL-C	Kathiresan et al 2009

rs2156552	18	47181418	Intergenic	*LIPG*	HDL-C	Willer et al 2008
rs2967605	19	8469488	Downstream	*ANGPTL4*	HDL-C	Kathiresan et al 2009
rs6102059	20	39228784	Intergenic	*MAFB*	LDL-C	Kathiresan et al 2009
rs7679	20	44576252	Downstream	*PLTP*	HDL-C	Kathiresan et al 2009
					TG	Kathiresan et al 2009

Appendix Table T. Gene-smoking interaction results for HDL-C levels, adjusted for age and sex.

SNP	Non-Hispanic Whites SNPxSmoking Interaction			Non-Hispanic Blacks SNPxSmoking Interaction			Mexican Americans SNPxSmoking Interaction		
	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
rs11206510	-0.74	1.00	0.462	0.11	1.69	0.948	-1.40	1.55	0.364
rs12654264	0.34	0.80	0.674	-0.17	1.21	0.889	0.47	1.05	0.653
rs1501908	1.17	0.80	0.146	0.94	1.18	0.423	-0.19	1.18	0.875
rs174547	-1.09	0.83	0.188	4.16	2.04	0.041	0.37	0.98	0.702
rs1748195	-1.11	0.80	0.164	1.04	1.21	0.391	-0.97	1.00	0.334
rs1883025	-0.09	0.89	0.921	-0.37	1.18	0.754	-0.55	1.08	0.613
rs2156552	-0.91	1.04	0.383	3.02	2.80	0.280	0.80	1.79	0.656
rs2197089	0.08	0.76	0.919	-0.62	1.42	0.663	0.95	0.98	0.329
rs2271293	-1.66	1.14	0.145	-1.10	2.20	0.615	-1.12	1.49	0.454
rs2338104	-0.40	0.77	0.609	0.21	1.30	0.872	0.71	0.99	0.473
rs2954029	-0.07	0.77	0.932	0.82	1.19	0.491	-0.45	1.01	0.657
rs2967605	0.28	0.99	0.778	2.32	1.41	0.098	-1.22	1.18	0.303
rs3135506	2.04	1.63	0.211	1.29	2.28	0.573	1.49	1.43	0.298
rs3764261	-0.12	0.83	0.886	-0.70	1.24	0.575	-0.27	1.07	0.798
rs3890182	1.92	1.19	0.108	1.19	1.76	0.500	-1.11	1.99	0.578
rs4149268	-0.63	0.79	0.426	1.38	1.21	0.255	-0.31	1.07	0.772
rs4775041	0.34	0.84	0.683	1.96	1.66	0.240	0.62	1.23	0.617
rs6102059	-1.13	0.85	0.184	-0.19	1.11	0.867	2.61	1.06	0.014
rs693	-1.08	0.89	0.225	-0.76	1.57	0.627	-0.32	1.18	0.786
rs754523	-0.30	0.82	0.719	0.44	1.38	0.749	0.88	1.13	0.433
rs7679	-0.63	1.00	0.529	8.42	2.88	0.003	0.97	1.55	0.532
rs780094	-1.03	0.77	0.182	-1.51	1.51	0.317	-0.62	1.07	0.562
rs9989419	-1.47	0.78	0.061	-1.84	1.16	0.114	-0.98	1.03	0.339

Appendix Table U. Gene-smoking interaction results for LDL-C levels, adjusted for age and sex.

SNP	Non-Hispanic Whites SNPxSmoking Interaction			Non-Hispanic Blacks SNPxSmoking Interaction			Mexican Americans SNPxSmoking Interaction		
	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
rs11206510	-4.62	2.90	0.111	-2.28	4.69	0.627	4.45	4.70	0.344
rs12654264	3.53	2.30	0.125	-0.32	3.49	0.927	1.26	3.45	0.714
rs1501908	2.28	2.30	0.322	-3.71	3.31	0.263	1.17	3.65	0.748
rs174547	-2.13	2.43	0.381	7.38	5.67	0.194	-0.53	3.05	0.862
rs1748195	-0.78	2.37	0.742	-2.69	3.44	0.433	-2.84	3.08	0.356
rs1883025	-2.87	2.57	0.265	2.90	3.34	0.386	0.63	3.36	0.852
rs2156552	4.22	2.94	0.151	-4.67	7.78	0.548	-1.07	5.46	0.844
rs2197089	-0.23	2.17	0.914	0.36	4.02	0.929	3.53	3.08	0.252
rs2271293	3.39	3.34	0.310	-10.53	5.87	0.073	0.76	4.50	0.867
rs2338104	4.73	2.25	0.035	-5.60	3.58	0.118	-2.84	3.12	0.363
rs2954029	-1.13	2.20	0.609	1.33	3.46	0.702	2.52	3.11	0.418
rs2967605	4.87	2.86	0.089	-4.07	3.97	0.306	-2.33	3.74	0.533
rs3135506	-0.64	4.72	0.892	-3.73	6.60	0.572	3.33	4.57	0.466
rs3764261	2.54	2.41	0.291	2.67	3.49	0.444	5.95	3.37	0.078
rs3890182	-6.33	3.53	0.073	-7.92	4.99	0.113	-0.03	6.22	0.997
rs4149268	-5.09	2.32	0.028	-2.76	3.40	0.417	-2.29	3.29	0.488
rs4775041	4.30	2.44	0.078	3.32	4.61	0.472	-0.55	3.80	0.884
rs6102059	0.70	2.44	0.775	1.68	3.14	0.592	-0.08	3.37	0.981
rs693	2.22	2.63	0.398	1.04	4.66	0.823	-0.76	3.70	0.838
rs754523	0.35	2.41	0.885	-6.54	4.00	0.102	-1.62	3.53	0.646
rs7679	1.49	2.89	0.606	9.89	8.45	0.242	3.46	4.96	0.486
rs780094	1.19	2.23	0.594	-0.02	4.14	0.996	-1.55	3.37	0.646
rs9989419	1.99	2.29	0.384	-3.06	3.29	0.354	4.19	3.20	0.191

Appendix Table V. Gene-smoking interaction results for transformed triglyceride levels, adjusted for age and sex.

SNP	Non-Hispanic Whites SNPxSmoking Interaction			Non-Hispanic Blacks SNPxSmoking Interaction			Mexican Americans SNPxSmoking Interaction		
	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
rs11206510	-0.037	0.020	0.387	-0.032	0.033	0.457	-0.001	0.030	0.360
rs12654264	0.017	0.016	0.926	-0.001	0.023	0.267	-0.002	0.020	0.346
rs1501908	-0.018	0.016	0.728	-0.039	0.022	0.219	-0.014	0.023	0.953
rs174547	0.060	0.016	0.318	0.067	0.037	0.700	0.068	0.020	0.058
rs1748195	0.073	0.016	0.077	0.020	0.023	0.700	0.053	0.019	0.517
rs1883025	-0.032	0.018	0.766	-0.001	0.023	0.527	-0.002	0.022	0.237
rs2156552	0.016	0.020	0.789	-0.005	0.053	0.310	-0.035	0.036	0.189
rs2197089	0.024	0.015	0.675	0.016	0.027	0.849	0.037	0.019	0.353
rs2271293	0.000	0.024	0.868	0.077	0.042	0.294	0.002	0.030	0.558
rs2338104	0.000	0.015	0.369	0.023	0.024	0.542	0.023	0.020	0.263
rs2954029	0.048	0.015	0.584	-0.003	0.023	0.300	0.045	0.020	0.093
rs2967605	0.049	0.020	0.273	0.011	0.026	0.042	0.020	0.022	0.409
rs3135506	0.116	0.032	0.374	0.169	0.043	0.169	0.111	0.028	0.317
rs3764261	0.002	0.016	0.483	-0.024	0.024	0.403	0.024	0.021	0.487
rs3890182	-0.023	0.024	0.796	-0.021	0.033	0.441	-0.067	0.036	1.000
rs4149268	-0.026	0.016	0.683	0.015	0.023	0.852	-0.017	0.021	0.467
rs4775041	-0.009	0.017	0.317	0.062	0.031	0.774	0.009	0.025	0.861
rs6102059	-0.006	0.017	0.469	-0.011	0.021	0.785	0.029	0.022	0.022
rs693	-0.004	0.016	0.849	-0.073	0.029	0.036	-0.019	0.022	0.558
rs754523	0.022	0.016	0.725	-0.048	0.027	0.104	0.030	0.021	0.372
rs7679	0.031	0.020	0.845	0.004	0.052	0.700	0.004	0.031	0.026
rs780094	0.049	0.016	0.975	-0.005	0.028	0.552	0.044	0.020	0.236
rs9989419	-0.005	0.016	0.099	0.003	0.022	0.547	0.026	0.021	0.165

Appendix Table W. Gene-vitamin A interaction results for HDL-C levels, adjusted for age and sex.

SNP	Non-Hispanic Whites SNPxVitA Interaction			Non-Hispanic Blacks SNPxVitA Interaction			Mexican Americans SNPxVitA Interaction		
	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
rs11206510	-0.025	0.035	0.471	-0.023	0.056	0.684	0.007	0.046	0.873
rs12654264	-0.036	0.026	0.174	0.014	0.039	0.717	-0.018	0.031	0.567
rs1501908	0.046	0.026	0.075	0.019	0.037	0.610	-0.007	0.035	0.832
rs174547	0.015	0.028	0.598	-0.029	0.050	0.563	0.003	0.031	0.924
rs1748195	0.087	0.027	**1.16E-03**	0.096	0.039	0.014	0.014	0.030	0.642
rs1883025	-0.052	0.029	0.074	-0.080	0.037	0.032	0.000	0.033	0.996
rs2156552	-0.022	0.037	0.558	-0.081	0.099	0.414	-0.027	0.061	0.656
rs2197089	0.009	0.026	0.739	-0.006	0.044	0.888	0.003	0.031	0.911
rs2271293	-0.066	0.040	0.104	-0.146	0.070	0.037	-0.071	0.046	0.121
rs2338104	0.009	0.026	0.719	-0.041	0.040	0.307	-0.015	0.031	0.620
rs2954029	-0.003	0.025	0.893	0.008	0.036	0.826	0.017	0.030	0.585
rs2967605	-0.043	0.032	0.183	-0.056	0.047	0.233	0.037	0.035	0.280
rs3135506	-0.029	0.052	0.582	0.070	0.072	0.331	-0.093	0.040	0.019
rs3764261	-0.043	0.027	0.114	-0.045	0.039	0.245	0.004	0.032	0.907
rs3890182	-0.020	0.039	0.607	-0.019	0.060	0.749	0.047	0.057	0.408
rs4149268	-0.038	0.027	0.163	0.006	0.037	0.871	-0.066	0.032	0.040
rs4775041	0.008	0.028	0.776	0.047	0.052	0.361	0.013	0.041	0.745
rs6102059	-0.028	0.028	0.312	-0.023	0.036	0.526	0.035	0.035	0.314
rs693	-0.032	0.026	0.212	0.031	0.046	0.496	-0.026	0.032	0.418
rs754523	-0.028	0.028	0.320	0.096	0.047	0.039	-0.016	0.033	0.618
rs7679	0.027	0.034	0.423	-0.003	0.085	0.970	-0.057	0.041	0.164
rs780094	0.026	0.026	0.307	0.022	0.047	0.646	0.005	0.031	0.877
rs9989419	-0.036	0.027	0.177	-0.005	0.036	0.890	-0.021	0.032	0.503

Appendix Table X. Gene-vitamin A interaction results for LDL-C levels, adjusted for age and sex.

SNP	Non-Hispanic Whites SNPxVitA Interaction			Non-Hispanic Blacks SNPxVitA Interaction			Mexican Americans SNPxVitA Interaction		
	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
rs11206510	0.002	0.102	0.982	0.342	0.155	0.027	0.583	0.147	**7.65E-05**
rs12654264	0.056	0.078	0.471	0.048	0.108	0.659	-0.130	0.098	0.185
rs1501908	0.071	0.078	0.365	-0.081	0.107	0.449	-0.211	0.106	0.047
rs174547	-0.068	0.084	0.421	0.307	0.176	0.081	-0.102	0.093	0.269
rs1748195	0.043	0.082	0.606	-0.267	0.112	0.018	0.000	0.091	1.000
rs1883025	0.181	0.087	0.038	0.013	0.112	0.908	0.001	0.104	0.994
rs2156552	-0.227	0.110	0.040	-0.052	0.278	0.853	-0.070	0.181	0.699
rs2197089	-0.004	0.076	0.961	0.132	0.123	0.283	-0.136	0.095	0.155
rs2271293	-0.075	0.125	0.546	0.024	0.184	0.894	-0.103	0.135	0.443
rs2338104	-0.037	0.078	0.636	-0.005	0.109	0.962	-0.201	0.097	0.038
rs2954029	-0.010	0.072	0.885	-0.031	0.107	0.770	0.099	0.090	0.275
rs2967605	0.063	0.092	0.494	-0.005	0.135	0.973	-0.123	0.101	0.223
rs3135506	0.255	0.151	0.093	-0.502	0.225	0.026	-0.132	0.119	0.268
rs3764261	-0.033	0.081	0.679	0.141	0.110	0.203	0.031	0.099	0.755
rs3890182	-0.008	0.122	0.947	-0.268	0.169	0.113	0.199	0.182	0.274
rs4149268	0.146	0.080	0.068	-0.167	0.099	0.093	0.025	0.099	0.803
rs4775041	-0.012	0.081	0.886	0.065	0.141	0.644	0.107	0.128	0.403
rs6102059	-0.005	0.081	0.952	0.110	0.102	0.282	-0.029	0.107	0.785
rs693	0.146	0.079	0.066	-0.133	0.135	0.327	0.255	0.101	0.012
rs754523	0.011	0.083	0.892	-0.095	0.137	0.487	-0.150	0.104	0.150
rs7679	0.055	0.103	0.594	-0.562	0.305	0.066	0.020	0.129	0.875
rs780094	-0.044	0.079	0.577	-0.088	0.132	0.505	0.162	0.100	0.106
rs9989419	0.060	0.079	0.449	-0.028	0.099	0.776	0.110	0.096	0.250

Appendix Table Y. Gene-vitamin A interaction results for transformed triglyceride levels, adjusted for age and sex.

SNP	Non-Hispanic Whites SNPxVitA Interaction			Non-Hispanic Blacks SNPxVitA Interaction			Mexican Americans SNPxVitA Interaction		
	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
rs11206510	-0.0003	0.0013	0.789	0.0001	0.0017	0.971	0.0037	0.0020	0.071
rs12654264	0.0009	0.0009	0.330	0.0001	0.0012	0.943	-0.0011	0.0014	0.430
rs1501908	0.0005	0.0009	0.634	-0.0017	0.0011	0.128	-0.0012	0.0015	0.401
rs174547	0.0001	0.0010	0.905	-0.0005	0.0019	0.781	0.0013	0.0013	0.312
rs1748195	0.0003	0.0010	0.797	-0.0006	0.0012	0.584	-0.0001	0.0013	0.919
rs1883025	0.0015	0.0011	0.163	0.0010	0.0012	0.412	0.0003	0.0014	0.828
rs2156552	-0.0022	0.0013	0.097	0.0033	0.0030	0.278	0.0001	0.0026	0.976
rs2197089	0.0001	0.0009	0.950	-0.0008	0.0014	0.554	0.0006	0.0013	0.648
rs2271293	0.0033	0.0015	0.027	-0.0024	0.0021	0.252	0.0000	0.0020	0.986
rs2338104	0.0018	0.0009	0.056	0.0008	0.0012	0.527	0.0007	0.0013	0.618
rs2954029	0.0002	0.0009	0.800	-0.0003	0.0012	0.796	-0.0003	0.0013	0.847
rs2967605	0.0008	0.0012	0.487	0.0037	0.0014	0.009	-0.0029	0.0015	0.046
rs3135506	-0.0048	0.0019	0.011	0.0007	0.0024	0.766	0.0027	0.0016	0.102
rs3764261	0.0009	0.0010	0.363	-0.0009	0.0012	0.419	-0.0030	0.0014	0.036
rs3890182	-0.0011	0.0014	0.450	0.0020	0.0018	0.275	0.0013	0.0026	0.616
rs4149268	0.0000	0.0010	0.990	0.0008	0.0011	0.444	0.0019	0.0014	0.194
rs4775041	-0.0001	0.0010	0.959	0.0009	0.0015	0.536	-0.0006	0.0018	0.716
rs6102059	-0.0005	0.0010	0.606	0.0014	0.0011	0.194	0.0002	0.0016	0.877
rs693	0.0028	0.0009	**2.16E-03**	0.0000	0.0014	0.975	0.0009	0.0013	0.506
rs754523	-0.0005	0.0010	0.591	-0.0022	0.0014	0.114	-0.0022	0.0014	0.110
rs7679	0.0000	0.0012	0.976	0.0015	0.0026	0.548	-0.0004	0.0018	0.841
rs780094	0.0000	0.0009	0.990	0.0005	0.0015	0.751	0.0024	0.0014	0.078
rs9989419	0.0022	0.0010	0.027	0.0001	0.0011	0.921	-0.0001	0.0014	0.935

Appendix Table Z. Gene-vitamin E interaction results for HDL-C levels, adjusted for age and sex.

SNP	Non-Hispanic Whites SNPxVitE Interaction			Non-Hispanic Blacks SNPxVitE Interaction			Mexican Americans SNPxVitE Interaction		
	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
rs11206510	1.834	1.376	0.183	-5.272	3.026	0.082	-0.921	2.070	0.657
rs12654264	-1.334	1.108	0.229	-0.414	2.166	0.848	1.020	1.383	0.461
rs1501908	0.247	1.102	0.823	-1.320	2.061	0.522	-0.045	1.455	0.975
rs174547	0.575	1.138	0.613	-0.037	3.297	0.991	-2.094	1.333	0.116
rs1748195	3.279	1.063	**2.06E-03**	0.676	2.130	0.751	-1.498	1.329	0.260
rs1883025	-0.531	1.243	0.669	2.644	2.134	0.216	1.762	1.383	0.203
rs2156552	-1.519	1.430	0.288	2.041	4.586	0.656	-1.742	2.631	0.508
rs2197089	-2.016	1.077	0.061	-1.881	2.494	0.451	2.712	1.243	0.029
rs2271293	0.252	1.702	0.882	4.049	4.292	0.346	-2.066	2.038	0.311
rs2338104	0.953	1.077	0.376	-2.549	2.112	0.228	0.105	1.313	0.936
rs2954029	-0.663	1.057	0.530	2.050	2.107	0.331	-2.002	1.352	0.139
rs2967605	-1.089	1.355	0.422	2.425	2.506	0.333	0.848	1.534	0.580
rs3135506	-5.433	2.208	0.014	-2.779	4.517	0.539	-5.204	1.844	0.005
rs3764261	-0.834	1.133	0.462	0.303	2.351	0.897	-0.345	1.441	0.811
rs3890182	0.429	1.644	0.794	-5.264	3.215	0.102	6.950	2.409	0.004
rs4149268	0.823	1.135	0.468	0.177	2.001	0.930	0.668	1.428	0.640
rs4775041	-0.088	1.155	0.939	-1.690	2.976	0.570	-0.153	1.674	0.927
rs6102059	1.169	1.161	0.314	1.435	1.872	0.444	0.924	1.443	0.522
rs693	-1.092	1.093	0.318	1.502	2.640	0.569	0.396	1.441	0.784
rs754523	-0.944	1.112	0.396	1.108	2.536	0.662	-0.230	1.422	0.872
rs7679	1.210	1.364	0.375	13.123	5.165	0.011	-4.159	1.940	0.032
rs780094	2.547	1.046	0.015	1.438	2.791	0.606	-0.054	1.376	0.969
rs9989419	0.277	1.086	0.799	-2.350	1.998	0.240	0.052	1.420	0.971

Appendix Table AA. Gene-vitamin E interaction results for LDL-C levels, adjusted for age and sex.

SNP	Non-Hispanic Whites SNPxVitE Interaction			Non-Hispanic Blacks SNPxVitE Interaction			Mexican Americans SNPxVitE Interaction		
	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
rs11206510	-3.284	3.797	0.387	17.961	7.515	0.017	17.617	6.225	0.005
rs12654264	-3.327	3.009	0.269	1.072	5.676	0.850	-5.377	4.087	0.189
rs1501908	1.125	3.090	0.716	-10.704	5.068	0.035	-0.512	4.211	0.903
rs174547	-1.971	3.193	0.537	-2.235	8.043	0.781	3.754	3.793	0.323
rs1748195	-4.179	3.030	0.168	-1.313	5.458	0.810	2.922	3.824	0.445
rs1883025	-1.795	3.538	0.612	-6.115	5.332	0.252	-11.222	3.842	0.004
rs2156552	-6.600	3.897	0.091	-3.780	11.963	0.752	-1.331	7.376	0.857
rs2197089	1.116	3.049	0.714	7.244	6.460	0.262	-3.602	3.743	0.336
rs2271293	4.031	4.879	0.409	-2.141	10.145	0.833	-9.130	5.616	0.104
rs2338104	-0.165	3.005	0.956	-5.746	5.468	0.294	-3.743	3.835	0.329
rs2954029	3.404	2.853	0.233	-4.995	5.378	0.353	4.473	3.866	0.248
rs2967605	4.613	3.778	0.222	1.328	6.429	0.836	1.950	4.215	0.644
rs3135506	5.830	6.240	0.350	20.031	11.789	0.090	6.106	5.532	0.270
rs3764261	0.507	3.192	0.874	2.744	5.723	0.632	-11.030	4.174	0.008
rs3890182	2.795	4.706	0.553	5.048	7.974	0.527	-9.546	7.013	0.174
rs4149268	-0.035	3.138	0.991	-7.470	5.167	0.149	-3.401	4.027	0.399
rs4775041	-4.510	3.128	0.150	-2.991	7.275	0.681	-2.509	4.684	0.592
rs6102059	-1.702	3.163	0.591	5.333	4.675	0.254	-2.413	4.312	0.576
rs693	11.105	3.039	**2.67E-04**	10.875	6.878	0.114	22.708	4.595	**8.94E-07**
rs754523	-1.758	3.123	0.574	7.625	6.429	0.236	1.332	4.306	0.757
rs7679	-4.225	3.833	0.270	-6.495	12.818	0.612	-2.828	5.615	0.615
rs780094	-2.484	2.923	0.396	-19.277	6.766	0.004	-1.304	4.068	0.749
rs9989419	1.584	2.962	0.593	-2.435	5.069	0.631	-9.543	4.104	0.020

Appendix Table BB. Gene-vitamin E interaction results for transformed triglyceride levels, adjusted for age and sex.

SNP	Non-Hispanic Whites SNPxVitE Interaction			Non-Hispanic Blacks SNPxVitE Interaction			Mexican Americans SNPxVitE Interaction		
	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value
rs11206510	-0.1483	0.0459	**1.27E-03**	0.0877	0.0829	0.290	0.1326	0.0735	0.071
rs12654264	0.0096	0.0374	0.797	0.1272	0.0585	0.030	0.1081	0.0491	0.028
rs1501908	0.0550	0.0376	0.144	0.0486	0.0565	0.390	-0.0160	0.0510	0.754
rs174547	-0.0500	0.0387	0.196	0.1363	0.0885	0.124	0.0869	0.0472	0.066
rs1748195	-0.0117	0.0367	0.751	0.0387	0.0577	0.502	0.0326	0.0476	0.494
rs1883025	-0.0067	0.0425	0.875	-0.0396	0.0586	0.499	-0.0577	0.0487	0.236
rs2156552	0.0140	0.0483	0.773	-0.0424	0.1228	0.730	0.1495	0.0935	0.110
rs2197089	0.0222	0.0366	0.544	-0.0015	0.0702	0.984	-0.0738	0.0455	0.105
rs2271293	0.0494	0.0573	0.388	0.0423	0.1154	0.714	0.0888	0.0727	0.222
rs2338104	-0.0421	0.0368	0.253	0.0580	0.0581	0.319	0.0514	0.0464	0.267
rs2954029	0.0327	0.0357	0.360	-0.0629	0.0593	0.289	0.0829	0.0477	0.082
rs2967605	0.0218	0.0460	0.636	-0.0293	0.0690	0.671	0.0321	0.0536	0.550
rs3135506	0.0456	0.0778	0.558	0.4555	0.1238	**2.45E-04**	0.1918	0.0641	0.003
rs3764261	0.0099	0.0394	0.801	-0.0261	0.0640	0.684	-0.1325	0.0509	0.009
rs3890182	-0.0353	0.0581	0.543	-0.0487	0.0899	0.588	-0.0856	0.0881	0.331
rs4149268	-0.0470	0.0393	0.232	-0.0101	0.0535	0.850	0.0102	0.0514	0.842
rs4775041	-0.0223	0.0388	0.566	-0.0552	0.0727	0.448	-0.0911	0.0587	0.121
rs6102059	-0.0668	0.0394	0.090	-0.0360	0.0519	0.488	0.0611	0.0532	0.251
rs693	0.1449	0.0355	**4.65E-05**	-0.0850	0.0703	0.227	0.1238	0.0492	0.012
rs754523	0.0290	0.0377	0.442	-0.0921	0.0687	0.180	-0.0055	0.0505	0.913
rs7679	-0.0234	0.0466	0.615	-0.0531	0.1417	0.708	-0.0169	0.0691	0.807
rs780094	0.0084	0.0356	0.814	-0.0106	0.0753	0.888	0.0324	0.0491	0.510
rs9989419	0.0202	0.0372	0.586	0.0861	0.0545	0.114	0.0183	0.0500	0.715

REFERENCES

(1998). Design of the Women's Health Initiative Clinical Trial and Observational Study. Controlled Clinical Trials *19*, 61-109.

Adeyemo,A., Gerry,N., Chen,G., Herbert,A., Doumatey,A., Huang,H., Zhou,J., Lashley,K., Chen,Y., Christman,M., and Rotimi,C. (2009). A Genome-Wide Association Study of Hypertension and Blood Pressure in African Americans. PLoS Genet *5*, e1000564.

Adkins,D.E., Aberg,K., McClay,J.L., Bukszar,J., Zhao,Z., Jia,P., Stroup,T.S., Perkins,D., McEvoy,J.P., Lieberman,J.A., Sullivan,P.F., and van den Oord,E.J.C.G. (2010). Genomewide pharmacogenomic study of metabolic side effects to antipsychotic drugs. Mol Psychiatry.

Ali,S., Bunker,C.H., Aston,C.E., Ukoli,F.A., and Kamboh,M.I. (1998). Apolipoprotein A kringle 4 polymorphism and serum lipoprotein (a) concentrations in African blacks. Hum. Biol. *70*, 477-490.

Anderson,G.L., Manson,J., Wallace,R., Lund,B., Hall,D., Davis,S., Shumaker,S., Wang,C.Y., Stein,E., and Prentice,R.L. (2003). Implementation of the women's health initiative study design. Annals of Epidemiology *13*, S5-S17.

Aulchenko,Y.S., Ripatti,S., Lindqvist,I., Boomsma,D., Heid,I.M., Pramstaller,P.P., Penninx,B.W., Janssens,A.C., Wilson,J.F., Spector,T., Martin,N.G., Pedersen,N.L., Kyvik,K.O., Kaprio,J., Hofman,A., Freimer,N.B., Jarvelin,M.R., Gyllensten,U., Campbell,H., Rudan,I., Johansson,A., Marroni,F., Hayward,C., Vitart,V., Jonasson,I., Pattaro,C., Wright,A., Hastie,N., Pichler,I., Hicks,A.A., Falchi,M., Willemsen,G., Hottenga,J.J., de Geus,E.J., Montgomery,G.W., Whitfield,J., Magnusson,P., Saharinen,J., Perola,M., Silander,K., Isaacs,A., Sijbrands,E.J., Uitterlinden,A.G., Witteman,J.C., Oostra,B.A., Elliott,P., Ruokonen,A., Sabatti,C., Gieger,C., Meitinger,T., Kronenberg,F., Doring,A., Wichmann,H.E., Smit,J.H., McCarthy,M.I., van Duijn,C.M., and Peltonen,L. (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat. Genet. *41*, 47-55.

Austin,M.A., Hutter,C.M., Zimmern,R.L., and Humphries,S.E. (2004). Genetic causes of monogenic heterozygous familial hypercholesterolemia: a HuGE prevalence review. Am. J. Epidemiol. *160*, 407-420.

Bamshad,M. (2005). Genetic influences on health: does race matter? JAMA *294*, 937-946.

Barkley,R.A., Brown,A.C., Hanis,C.L., Kardia,S.L., Turner,S.T., and Boerwinkle,E. (2003). Lack of genetic linkage evidence for a trans-acting factor having a large effect on plasma lipoprotein[a] levels in African Americans. J. Lipid Res. *44*, 1301-1305.

Beekman,M., Heijmans,B.T., Martin,N.G., Pedersen,N.L., Whitfield,J.B., DeFaire,U., van Baal,G.C., Snieder,H., Vogler,G.P., Slagboom,P.E., and Boomsma,D.I. (2002). Heritabilities of apolipoprotein and lipid levels in three countries. Twin. Res. *5*, 87-97.

Bellman,R. (1961). Adaptive Control Processes. Princeton University Press).

Bennet,A., Di,A.E., Erqou,S., Eiriksdottir,G., Sigurdsson,G., Woodward,M., Rumley,A., Lowe,G.D., Danesh,J., and Gudnason,V. (2008). Lipoprotein(a) levels and risk of future coronary heart disease: large-scale prospective data. Arch. Intern. Med. *168*, 598-608.

Berglund,L. and Ramakrishnan,R. (2004). Lipoprotein(a): an elusive cardiovascular risk factor. Arterioscler. Thromb. Vasc. Biol. *24*, 2219-2226.

Bernert,J.T., Jr., Turner,W.E., Pirkle,J.L., Sosnoff,C.S., Akins,J.R., Waldrep,M.K., Ann,Q., Covey,T.R., Whitfield,W.E., Gunter,E.W., Miller,B.B., Patterson,D.G., Jr., Needham,L.L., Hannon,W.H., and Sampson,E.J. (1997). Development and validation of sensitive method for determination of serum cotinine in smokers and nonsmokers by liquid chromatography/atmospheric pressure ionization tandem mass spectrometry. Clin. Chem. *43*, 2281-2291.

Bernstein,M.S., Costanza,M.C., James,R.W., Morris,M.A., Cambien,F., Raoux,S., and Morabia,A. (2002). Physical activity may modulate effects of ApoE genotype on lipid profile. Arterioscler. Thromb. Vasc. Biol. *22*, 133-140.

Bershad,S., Rubinstein,A., Paterniti,J.R., Le,N.A., Poliak,S.C., Heller,B., Ginsberg,H.N., Fleischmajer,R., and Brown,W.V. (1985). Changes in plasma lipids and lipoproteins during isotretinoin therapy for acne. N. Engl. J. Med. *313*, 981-985.

Bierut,L.J., Agrawal,A., Bucholz,K.K., Doheny,K.F., Laurie,C., Pugh,E., Fisher,S., Fox,L., Howells,W., Bertelsen,S., Hinrichs,A.L., Almasy,L., Breslau,N., Culverhouse,R.C., Dick,D.M., Edenberg,H.J., Foroud,T., Grucza,R.A., Hatsukami,D., Hesselbrock,V., Johnson,E.O., Kramer,J., Krueger,R.F., Kuperman,S., Lynskey,M., Mann,K., Neuman,R.J., N+¦then,M.M., Nurnberger,J.I., Porjesz,B., Ridinger,M., Saccone,N.L., Saccone,S.F., Schuckit,M.A., Tischfield,J.A., Wang,J.C., Rietschel,M., Goate,A.M., Rice,J.P., and as part of the Gene,E.A.S.G.C. (2010). A genome-wide association study of alcohol dependence. PNAS *107*, 5082-5087.

Boers,M., Nurmohamed,M.T., Doelman,C.J., Lard,L.R., Verhoeven,A.C., Voskuyl,A.E., Huizinga,T.W., van de Stadt,R.J., Dijkmans,B.A., and van der,L.S. (2003). Influence of glucocorticoids and disease activity on total and high density lipoprotein cholesterol in patients with rheumatoid arthritis. Ann. Rheum. Dis. *62*, 842-845.

Boerwinkle,E., Leffert,C.C., Lin,J., Lackner,C., Chiesa,G., and Hobbs,H.H. (1992). Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. J. Clin. Invest *90*, 52-60.

Boffa,M.B., Marcovina,S.M., and Koschinsky,M.L. (2004). Lipoprotein(a) as a risk factor for atherosclerosis and thrombosis: mechanistic insights from animal models. Clin. Biochem. *37*, 333-343.

Boomsma,D.I., Kempen,H.J., Gevers Leuven,J.A., Havekes,L., de,K.P., and Frants,R.R. (1996). Genetic analysis of sex and generation differences in plasma lipid, lipoprotein, and apolipoprotein levels in adolescent twins and their parents. Genet. Epidemiol. *13*, 49-60.

Boomsma,D.I., Knijff,P., Kaptein,A., Labeur,C., Martin,N.G., Havekes,L.M., and Princen,H.M. (2000). The effect of apolipoprotein(a)-, apolipoprotein E-, and apolipoprotein A4- polymorphisms on quantitative lipoprotein(a) concentrations. Twin. Res. *3*, 152-158.

Borel,P., Moussa,M., Reboul,E., Lyan,B., Defoort,C., Vincent-Baudry,S., Maillot,M., Gastaldi,M., Darmon,M., Portugal,H., Planells,R., and Lairon,D. (2007). Human plasma levels of vitamin E and carotenoids are associated with genetic polymorphisms in genes involved in lipid metabolism. J. Nutr. *137*, 2653-2659.

Bostrom,M., Lu,L., Chou,J., Hicks,P., Xu,J., Langefeld,C., Bowden,D., and Freedman,B. (2010). Candidate genes for non-diabetic ESRD in African Americans: a genome-wide association study using pooled DNA. Human Genetics *128*, 195-204.

Bozaoglu,K., Curran,J.E., Stocker,C.J., Zaibi,M.S., Segal,D., Konstantopoulos,N., Morrison,S., Carless,M., Dyer,T.D., Cole,S.A., Goring,H.H.H., Moses,E.K., Walder,K., Cawthorne,M.A., Blangero,J., and Jowett,J.B.M. (2010). Chemerin, a Novel Adipokine in the Regulation of Angiogenesis. J Clin Endocrinol Metab *95*, 2476-2485.

Brischetto,C.S., Connor,W.E., Connor,S.L., and Matarazzo,J.D. (1983). Plasma lipid and lipoprotein profiles of cigarette smokers from randomly selected families: enhancement of hyperlipidemia and depression of high-density lipoprotein. Am. J. Cardiol. *52*, 675-680.

Brown,M.S. and Goldstein,J.L. (1986). A receptor-mediated pathway for cholesterol homeostasis. Science *232*, 34-47.

Brown,W.M., Beck,S.R., Lange,E.M., Davis,C.C., Kay,C.M., Langefeld,C.D., and Rich,S.S. (2003). Age-stratified heritability estimation in the Framingham Heart Study families. BMC Genet. *4 Suppl 1*, S32.

Burkhardt,R., Kenny,E.E., Lowe,J.K., Birkeland,A., Josowitz,R., Noel,M., Salit,J., Maller,J.B., Pe'er,I., Daly,M.J., Altshuler,D., Stoffel,M., Friedman,J.M., and Breslow,J.L. (2008). Common SNPs in HMGCR in Micronesians and Whites Associated With LDL-Cholesterol Levels Affect Alternative Splicing of Exon13. Arterioscler Thromb Vasc Biol *28*, 2078-2084.

Caraballo,R.S., Giovino,G.A., Pechacek,T.F., and Mowery,P.D. (2001). Factors associated with discrepancies between self-reports on cigarette smoking and measured serum

cotinine levels among persons aged 17 years or older: Third National Health and Nutrition Examination Survey, 1988-1994. Am. J. Epidemiol. *153*, 807-814.

Carlson,C.S., Eberle,M.A., Kruglyak,L., and Nickerson,D.A. (2004a). Mapping complex disease loci in whole-genome association studies. Nature *429*, 446-452.

Carlson,C.S., Eberle,M.A., Rieder,M.J., Smith,J.D., Kruglyak,L., and Nickerson,D.A. (2003). Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat. Genet. *33*, 518-521.

Carlson,C.S., Eberle,M.A., Rieder,M.J., Yi,Q., and Kruglyak,L. (2004b). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet *74*, 106.

Cawthon,R.M., Smith,K.R., O'Brien,E., Sivatchenko,A., and Kerber,R.A. (2003). Association between telomere length in blood and mortality in people aged 60 years or older. Lancet *361*, 393-395.

Center for Disease Control and Prevention (1996). The Third National Health and Nutrition Examination Survey, 1988-94. Reference manuals and reports (CD ROM). US Department of Health and Human Services National Center for Health Statistics.

Centers for Disease Control and Prevention. Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94. 1994. Bethesda, MD, Centers for Disease Control and Prevention.
Ref Type: Report

Centers for Disease Control and Prevention. Third National Health and Nutrition Examination Survey, 1988-94, Plan and Operations Procedures Manuals. 1996. Hyattsville, Md, US Department of Health and Human Services (DHHS), National Center for Health Statistics, Centers for Disease Control and Prevention.
Ref Type: Audiovisual Material

Centers for Disease Control and Prevention (2002). National Health and Nutrition Examination Survey III (NHANES) DNA Specimens: Guidelines for Proposals to Use Samples and Proposed Cost Schedule. Federal Register *67*, 51585-51589.

Centers for Disease Control and Prevention (2010). National Health and Nutrition Examination Survey (NHANES) DNA Samples: Guidelines for Proposals to Use Samples and Cost Schedule. Federal Register *75*, 32191-32195.

Centers for Disease Control and Prevention (2004). Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94. (Bethesda, MD).

Centers for Disease Control and Prevention (CDC), National Center for Health Statistics NCHS. National Health and Nutrition Examination Survey Data. Centers for Disease Control and Prevention . 2002. Hyattsville, MD, U.S. Department of Health and Human

Services.
Ref Type: Electronic Citation

Chang,M.h., Lindegren,M.L., Butler,M.A., Chanock,S.J., Dowling,N.F., Gallagher,M., Moonesinghe,R., Moore,C.A., Ned,R.M., Reichler,M.R., Sanders,C.L., Welch,R., Yesupriya,A., Khoury,M.J., and for the CDC/NCI NHANES III Genomics Working Group (2009). Prevalence in the United States of Selected Candidate Gene Variants: Third National Health and Nutrition Examination Survey, 1991-1994. Am. J. Epidemiol. *169*, 54-66.

Chang,M.h., Yesupriya,A., Ned,R., Mueller,P., and Dowling,N. (2010). Genetic variants associated with fasting blood lipids in the U.S. population: Third National Health and Nutrition Examination Survey. BMC Medical Genetics *11*, 62.

Chelland,C.S., Moffatt,R.J., and Stamford,B.A. (2008). Smoking and smoking cessation -- the relationship between cardiovascular disease and lipoprotein metabolism: a review. Atherosclerosis *201*, 225-235.

Chiu,L., Hamman,R.F., and Kamboh,M.I. (2000). Apolipoprotein A polymorphisms and plasma lipoprotein(a) concentrations in non-Hispanic Whites and Hispanics. Hum. Biol. *72*, 821-835.

Choudhry,S., Seibold,M.A., Borrell,L.N., Tang,H., Serebrisky,D., Chapela,R., Rodriguez-Santana,J.R., Avila,P.C., Ziv,E., Rodriguez-Cintron,W., Risch,N.J., and Burchard,E.G. (2007). Dissecting complex diseases in complex populations: asthma in latino americans. Proc Am. Thorac. Soc. *4*, 226-233.

Chretien,J.P., Coresh,J., Berthier-Schaad,Y., Kao,W.H., Fink,N.E., Klag,M.J., Marcovina,S.M., Giaculli,F., and Smith,M.W. (2006). Three single-nucleotide polymorphisms in LPA account for most of the increase in lipoprotein(a) level elevation in African Americans compared with European Americans. J. Med. Genet. *43*, 917-923.

Clark,L.T., Ferdinand,K.C., Flack,J.M., Gavin,J.R., III, Hall,W.D., Kumanyika,S.K., Reed,J.W., Saunders,E., Valantine,H.A., Watson,K., Wenger,N.K., and Wright,J.T. (2001). Coronary heart disease in African Americans. Heart Dis. *3*, 97-108.

Clarke,R., Peden,J.F., Hopewell,J.C., Kyriakou,T., Goel,A., Heath,S.C., Parish,S., Barlera,S., Franzosi,M.G., Rust,S., Bennett,D., Silveira,A., Malarstig,A., Green,F.R., Lathrop,M., Gigante,B., Leander,K., de,F.U., Seedorf,U., Hamsten,A., Collins,R., Watkins,H., and Farrall,M. (2009). Genetic variants associated with Lp(a) lipoprotein level and coronary disease. N. Engl. J. Med. *361*, 2518-2528.

Clee,S.M., Kastelein,J.J., van,D.M., Marcil,M., Roomp,K., Zwarts,K.Y., Collins,J.A., Roelants,R., Tamasawa,N., Stulc,T., Suda,T., Ceska,R., Boucher,B., Rondeau,C., DeSouich,C., Brooks-Wilson,A., Molhuizen,H.O., Frohlich,J., Genest,J., Jr., and Hayden,M.R. (2000). Age and residual cholesterol efflux affect HDL cholesterol levels and coronary artery disease in ABCA1 heterozygotes. J. Clin. Invest *106*, 1263-1270.

Cohen,J., Pertsemlidis,A., Kotowski,I.K., Graham,R., Garcia,C.K., and Hobbs,H.H. (2005). Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat. Genet *37*, 161-165.

Cohen,J.C., Kiss,R.S., Pertsemlidis,A., Marcel,Y.L., and McPherson,R. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science *305*, 869.

Collins,F.S., Guyer,M.S., and Charkravarti,A. (1997). Variations on a theme: cataloging human DNA sequence variation. Science *278*, 1580.

Cordell,H.J. (2009). Detecting gene-gene interactions that underlie human diseases. Nat. Rev. Genet. *10*, 392-404.

Corella,D., Guillen,M., Saiz,C., Portoles,O., Sabater,A., Cortina,S., Folch,J., Gonzalez,J.I., and Ordovas,J.M. (2001a). Environmental factors modulate the effect of the APOE genetic polymorphism on plasma lipid concentrations: ecogenetic studies in a Mediterranean Spanish population. Metabolism *50*, 936-944.

Corella,D., Tucker,K., Lahoz,C., Coltell,O., Cupples,L.A., Wilson,P.W., Schaefer,E.J., and Ordovas,J.M. (2001b). Alcohol drinking determines the effect of the APOE locus on LDL-cholesterol concentrations in men: the Framingham Offspring Study. Am. J. Clin. Nutr. *73*, 736-745.

Craig,W.Y., Palomaki,G.E., and Haddow,J.E. (1989). Cigarette smoking and serum lipid and lipoprotein concentrations: an analysis of published data. BMJ *298*, 784-788.

Crawford,D.C., Carlson,C.S., Rieder,M.J., Carrington,D.P., Yi,Q., Smith,J.D., Eberle,M.A., Kruglyak,L., and Nickerson,D.A. (2004). Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. Am. J. Hum. Genet. *74*, 610-622.

Crawford,D.C., Peng,Z., Cheng,J.F., Boffelli,D., Ahearn,M., Nguyen,D., Shaffer,T., Yi,Q., Livingston,R.J., Rieder,M.J., and Nickerson,D.A. (2008). LPA and PLG sequence variation and kringle IV-2 copy number in two populations. Hum. Hered. *66*, 199-209.

Crawford,D.C., Sanders,C.L., Qin,X., Smith,J.D., Shephard,C., Wong,M., Witrak,L., Rieder,M.J., and Nickerson,D.A. (2006). Genetic Variation Is Associated With C-Reactive Protein Levels in the Third National Health and Nutrition Examination Survey. Circulation *114*, 2458-2465.

Cushman,M., Cornell,E.S., Howard,P.R., Bovill,E.G., and Tracy,R.P. (1995). Laboratory methods and quality assurance in the Cardiovascular Health Study. Clin Chem *41*, 264-270.

Danesh,J., Collins,R., and Peto,R. (2000). Lipoprotein(a) and coronary heart disease. Meta-analysis of prospective studies. Circulation *102*, 1082-1085.

Delgado-Lista,J., Perez-Jimenez,F., Ruano,J., Perez-Martinez,P., Fuentes,F., Criado-Garcia,J., Parnell,L.D., Garcia-Rios,A., Ordovas,J.M., and Lopez-Miranda,J. (2010). Effects of variations in the APOA1/C3/A4/A5 gene cluster on different parameters of postprandial lipid metabolism in healthy young men. J. Lipid Res. *51*, 63-73.

Deo,R.C., Reich,D., Tandon,A., Akylbekova,E., Patterson,N., Waliszewska,A., Kathiresan,S., Sarpong,D., Taylor,H.A., and Wilson,J.G. (2009). Genetic Differences between the Determinants of Lipid Profile Phenotypes in African and European Americans: The Jackson Heart Study. PLoS Genet *5*, e1000342.

Dullaart,R.P., Hoogenberg,K., Dikkeschei,B.D., and van,T.A. (1994). Higher plasma lipid transfer protein activities and unfavorable lipoprotein changes in cigarette-smoking men. Arterioscler. Thromb. *14*, 1581-1585.

Durbin,R.M., Abecasis,G.R., Altshuler,D.L., Auton,A., Brooks,L.D., Durbin,R.M., Gibbs,R.A., Hurles,M.E., and McVean,G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.

Edenberg,H.J., Koller,D.L., Xuei,X., Wetherill,L., McClintick,J.N., Almasy,L., Bierut,L.J., Bucholz,K.K., Goate,A., Aliev,F., Dick,D., Hesselbrock,V., Hinrichs,A., Kramer,J., Kuperman,S., Nurnberger,J.I., Rice,J.P., Schuckit,M.A., Taylor,R., Todd Webb,B., Tischfield,J.A., Porjesz,B., and Foroud,T. (2010). Genome-Wide Association Study of Alcohol Dependence Implicates a Region on Chromosome 11. Alcoholism: Clinical and Experimental Research *34*, 840-852.

Ericsson,S., Eriksson,M., Vitols,S., Einarsson,K., Berglund,L., and Angelin,B. (1991). Influence of age on the metabolism of plasma low density lipoproteins in healthy males. J. Clin. Invest *87*, 591-596.

Expert Panel Blood Cholesterol Levels Child Adolesc (1992). National Cholesterol Education Program Report of the Expert Panel on Blood Cholesterol Levels in Children and Adolescents. Pediatrics *89*, I-584.

Fless,G.M., Snyder,M.L., Furbee,J.W., Jr., Garcia-Hedo,M.T., and Mora,R. (1994). Subunit composition of lipoprotein(a) protein. Biochemistry *33*, 13492-13501.

Frazer,K.A., Ballinger,D.G., Cox,D.R., Hinds,D.A., Stuve,L.L., Gibbs,R.A., Belmont,J.W., Boudreau,A., Hardenbol,P., Leal,S.M., Pasternak,S., Wheeler,D.A., Willis,T.D., Yu,F., Yang,H., Zeng,C., Gao,Y., Hu,H., Hu,W., Li,C., Lin,W., Liu,S., Pan,H., Tang,X., Wang,J., Wang,W., Yu,J., ZHANG,B., Zhang,Q., Zhao,H., Zhao,H., Zhou,J., Gabriel,S.B., Barry,R., Blumenstiel,B., Camargo,A., Defelice,M., Faggart,M., Goyette,M., Gupta,S., Moore,J., Nguyen,H., Onofrio,R.C., Parkin,M., Roy,J., Stahl,E., Winchester,E., Ziaugra,L., Altshuler,D., Shen,Y., Yao,Z., Huang,W., Chu,X., He,Y., Jin,L., Liu,Y., Shen,Y., Sun,W., Wang,H., Wang,Y., Wang,Y., Xiong,X., Xu,L., Waye,M.M., Tsui,S.K., Xue,H., Wong,J.T., Galver,L.M., Fan,J.B., Gunderson,K., Murray,S.S., Oliphant,A.R., Chee,M.S., Montpetit,A., Chagnon,F., Ferretti,V., Leboeuf,M., Olivier,J.F., Phillips,M.S., Roumy,S., Sallee,C., Verner,A., Hudson,T.J., Kwok,P.Y., Cai,D., Koboldt,D.C., Miller,R.D., Pawlikowska,L., Taillon-Miller,P., Xiao,M., Tsui,L.C., Mak,W., Song,Y.Q., Tam,P.K.,

Nakamura,Y., Kawaguchi,T., Kitamoto,T., Morizono,T., Nagashima,A., Ohnishi,Y., Sekine,A., Tanaka,T., Tsunoda,T., Deloukas,P., Bird,C.P., Delgado,M., Dermitzakis,E.T., Gwilliam,R., Hunt,S., Morrison,J., Powell,D., Stranger,B.E., Whittaker,P., Bentley,D.R., Daly,M.J., De Bakker,P.I., Barrett,J., Chretien,Y.R., Maller,J., McCarroll,S., Patterson,N., Pe'er,I., Price,A., Purcell,S., Richter,D.J., Sabeti,P., Saxena,R., Schaffner,S.F., Sham,P.C., Varilly,P., Altshuler,D., Stein,L.D., Krishnan,L., Smith,A.V., Tello-Ruiz,M.K., Thorisson,G.A., Chakravarti,A., Chen,P.E., Cutler,D.J., Kashuk,C.S., Lin,S., Abecasis,G.R., Guan,W., Li,Y., Munro,H.M., Qin,Z.S., Thomas,D.J., McVean,G., Auton,A., Bottolo,L., Cardin,N., Eyheramendy,S., Freeman,C., Marchini,J., Myers,S., Spencer,C., Stephens,M., Donnelly,P., Cardon,L.R., Clarke,G., Evans,D.M., Morris,A.P., Weir,B.S., Tsunoda,T., Mullikin,J.C., Sherry,S.T., Feolo,M., Skol,A., Zhang,H., Zeng,C., Zhao,H., Matsuda,I., Fukushima,Y., Macer,D.R., Suda,E., Rotimi,C.N., Adebamowo,C.A., Ajayi,I., Aniagwu,T., Marshall,P.A., Nkwodimmah,C., Royal,C.D., Leppert,M.F., Dixon,M., Peiffer,A., Qiu,R., Kent,A., Kato,K., Niikawa,N., Adewole,I.F., Knoppers,B.M., Foster,M.W., Clayton,E.W., Watkin,J., Gibbs,R.A., Belmont,J.W., Muzny,D., Nazareth,L., Sodergren,E., Weinstock,G.M., Wheeler,D.A., Yakub,I., Gabriel,S.B., Onofrio,R.C., Richter,D.J., Ziaugra,L., Birren,B.W., Daly,M.J., Altshuler,D., Wilson,R.K., Fulton,L.L., Rogers,J., Burton,J., Carter,N.P., Clee,C.M., Griffiths,M., Jones,M.C., McLay,K., Plumb,R.W., Ross,M.T., Sims,S.K., Willey,D.L., Chen,Z., Han,H., Kang,L., Godbout,M., Wallenburg,J.C., L'Archeveque,P., Bellemare,G., Saeki,K., Wang,H., An,D., Fu,H., Li,Q., Wang,Z., Wang,R., Holden,A.L., Brooks,L.D., McEwen,J.E., Guyer,M.S., Wang,V.O., Peterson,J.L., Shi,M., Spiegel,J., Sung,L.M., Zacharia,L.F., Collins,F.S., Kennedy,K., Jamieson,R., and Stewart,J. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851-861.

Fried,L.P., Borhani,N.O., Enright,P., Furberg,C.D., Gardin,J.M., Kronmal,R.A., Kuller,L.H., Manolio,T.A., Mittelmark,M.B., Newman,A., O'Leary,D., Psaty,B., Rautaharju,P., and Tracy,R. (1991). The Cardiovascular Health Study: design and rationale. Ann Epidemiol *3*, 263-276.

Friedewald,W.T., Levy,R.I., and Fredrickson,D.S. (1972). Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem *18*, 499-501.

Friedlander,Y., Austin,M.A., Newman,B., Edwards,K., Mayer-Davis,E.I., and King,M.C. (1997). Heritability of longitudinal changes in coronary-heart-disease risk factors in women twins. Am. J. Hum. Genet. *60*, 1502-1512.

Friedman,G.D., Cutter,G.R., Donahue,R.P., Hughes,G.H., Hulley,S.B., Jacobs,D.R., Liu,K., and Savage,P.J. (1988). CARDIA: Study design, recruitment and some characteristics of the examined subjects. J Clin Epidemiol *41*, 1105-1116.

Gatica,L.V., Vega,V.A., Zirulnik,F., Oliveros,L.B., and Gimenez,M.S. (2006). Alterations in the lipid metabolism of rat aorta: effects of vitamin a deficiency. J. Vasc. Res. *43*, 602-610.

Gauderman, W. and Morrison, J. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. 2006.
Ref Type: Computer Program

Gauderman,W.J. (2002). Sample Size Requirements for Association Studies of Gene-Gene Interaction. Am. J. Epidemiol. *155*, 478-484.

Ge,D., Fellay,J., Thompson,A.J., Simon,J.S., Shianna,K.V., Urban,T.J., Heinzen,E.L., Qiu,P., Bertelsen,A.H., Muir,A.J., Sulkowski,M., McHutchison,J.G., and Goldstein,D.B. (2009). Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature *461*, 399-401.

Genovese,G., Tonna,S.J., Knob,A.U., Appel,G.B., Katz,A., Bernhardy,A.J., Needham,A.W., Lazarus,R., and Pollak,M.R. (2010). A risk allele for focal segmental glomerulosclerosis in African Americans is located within a region containing APOL1 and MYH9. Kidney Int *78*, 698-704.

Gordon,T., Castelli,W.P., Hjortland,M.C., Kannel,W.B., and Dawber,T.R. (1977). High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. Am. J. Med. *62*, 707-714.

Goring,H.H., Terwilliger,J.D., and Blangero,J. (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet *69*, 1357-1369.

Gupta,R., Ejebe,K., Butler,J., Lettre,G., Lyon,H., Guiducci,C., Wilks,R., Bennett,F., Forrester,T., Tayo,B., Musunuru,K., Hirschhorn,J., Kathiresan,S., Cooper,R., and McKenzie,C. (2010). Association of common DNA sequence variants at 33 genetic loci with blood lipids in individuals of African ancestry from Jamaica. Human Genetics 1-5.

Guyton,J.R., Dahlen,G.H., Patsch,W., Kautz,J.A., and Gotto,A.M., Jr. (1985). Relationship of plasma lipoprotein Lp(a) levels to race and to apolipoprotein B. Arterioscler Thromb Vasc Biol *5*, 265-272.

Haffner,S.M., Gruber,K.K., Morales,P.A., Hazuda,H.P., Valdez,R.A., Mitchell,B.D., and Stern,M.P. (1992). Lipoprotein(a) concentrations in Mexican Americans and non-Hispanic whites: the San Antonio Heart Study. Am. J. Epidemiol. *136*, 1060-1068.

Hagberg,J.M., Wilund,K.R., and Ferrell,R.E. (2000). APO E gene and gene-environment effects on plasma lipoprotein-lipid levels. Physiol Genomics *4*, 101-108.

Hallmayer,J., Faraco,J., Lin,L., Hesselson,S., Winkelmann,J., Kawashima,M., Mayer,G., Plazzi,G., Nevsimalova,S., Bourgin,P., Hong,S.S.-C., Honda,Y., Honda,M., Hogl,B., Longstreth,W.T., Montplaisir,J., Kemlink,D., Einen,M., Chen,J., Musone,S.L., Akana,M., Miyagawa,T., Duan,J., Desautels,A., Erhardt,C., Hesla,P.E., Poli,F., Frauscher,B., Jeong,J.H., Lee,S.P., Ton,T.G.N., Kvale,M., Kolesar,L., Dobrovolna,M., Nepom,G.T., Salomon,D., Wichmann,H.E., Rouleau,G.A., Gieger,C., Levinson,D.F., Gejman,P.V., Meitinger,T., Young,T., Peppard,P., Tokunaga,K., Kwok,P.Y., Risch,N., and Mignot,E.

(2009). Narcolepsy is strongly associated with the T-cell receptor alpha locus. Nat Genet *41*, 708-711.

Halton,J.M., Nazir,D.J., McQueen,M.J., and Barr,R.D. (1998). Blood lipid profiles in children with acute lymphoblastic leukemia. Cancer *83*, 379-384.

Hancock,D.B., Romieu,I., Shi,M., Sienra-Monge,J.J., Wu,H., Chiu,G.Y., Li,H., Rio-Navarro,B.E.d., Willis-Owens,S.A.G., Weiss,S.T., Raby,B.A., Gao,H., Eng,C., Chapela,R., Burchard,E.G., Tang,H., Sullivan,P.F., and London,S.J. (2009). Genome-Wide Association Study Implicates Chromosome 9q21.31 as a Susceptibility Locus for Asthma in Mexican Children. PLoS Genet *5*, e1000623.

Haworth,C.M., Butcher,L.M., Docherty,S.J., Wardle,J., and Plomin,R. (2008). No evidence for association between BMI and 10 candidate genes at ages 4, 7 and 10 in a large UK sample of twins. BMC Med. Genet. *9*, 12.

Hayes,M.G., Pluzhnikov,A., Miyake,K., Sun,Y., Ng,M.C.Y., Roe,C.A., Below,J.E., Nicolae,R.I., Konkashbaev,A., Bell,G.I., Cox,N.J., and Hanis,C.L. (2007). Identification of Type 2 Diabetes Genes in Mexican Americans Through Genome-wide Association Studies. Diabetes db07-0482.

Hegele,R.A. (2009). Plasma lipoproteins: genetic influences and clinical implications. Nat. Rev. Genet. *10*, 109-121.

Hegele,R.A. (2010). Genome-wide association studies of plasma lipids: have we reached the limit? Arterioscler. Thromb. Vasc. Biol. *30*, 2084-2086.

Heid,I.M., Boes,E., Muller,M., Kollerits,B., Lamina,C., Coassin,S., Gieger,C., Doring,A., Klopp,N., Frikke-Schmidt,R., Tybjaerg-Hansen,A., Brandstatter,A., Luchner,A., Meitinger,T., Wichmann,H.E., and Kronenberg,F. (2008). Genome-Wide Association Analysis of High-Density Lipoprotein Cholesterol in the Population-Based KORA Study Sheds New Light on Intergenic Regions. Circ Cardiovasc Genet *1*, 10-20.

Heiss,G., Schonfeld,G., Johnson,J.L., Heyden,S., Hames,C.G., and Tyroler,H.A. (1984). Black-white differences in plasma levels of apolipoproteins: the Evans County Heart Study. Am Heart J *108*, 807-814.

Heller,D.A., de,F.U., Pedersen,N.L., Dahlen,G., and McClearn,G.E. (1993). Genetic and environmental influences on serum lipid levels in twins. N. Engl. J. Med. *328*, 1150-1156.

Hickman,T.B., Briefel,R.R., Carroll,M.D., Rifkind,B.M., Cleeman,J.I., Maurer,K.R., and Johnson,C.L. (1998). Distributions and trends of serum lipid levels among United States children and adolescents ages 4-19 years: data from the Third National Health and Nutrition Examination Survey. Prev. Med. *27*, 879-890.

Himes,B.E., Hunninghake,G.M., Baurley,J.W., Rafaels,N.M., Sleiman,P., Strachan,D.P., Wilk,J.B., Willis-Owen,S.A.G., Klanderman,B., Lasky-Su,J., Lazarus,R., Murphy,A.J., Soto-Quiros,M.E., Avila,L., Beaty,T., Mathias,R.A., Ruczinski,I., Barnes,K.C., Celedn,J.C.,

Cookson,W.O.C., Gauderman,W.J., Gilliland,F.D., Hakonarson,H., Lange,C., Moffatt,M.F., O'Connor,G.T., Raby,B.A., Silverman,E.K., and Weiss,S.T. (2009). Genome-wide Association Analysis Identifies PDE4D as an Asthma-Susceptibility Gene. Am J Hum Genet *84*, 581-593.

Hindorff, L. A., Junkins, H. A., Mehta, J. P., and Manolio, T. A. A Catalog of Published Genome-Wide Association Studies. National Human Genome Research Institute . 2010. 6-30-2010.
Ref Type: Electronic Citation

Hindorff,L.A., Sethupathy,P., Junkins,H.A., Ramos,E.M., Mehta,J.P., Collins,F.S., and Manolio,T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. PNAS *106*, 9362-9367.

Hiura,Y., Shen,C.S., Kokubo,Y., Okamura,T., Morisaki,T., Tomoike,H., Yoshida,T., Sakamoto,H., Goto,Y., Nonogi,H., and Iwai,N. (2009). Identification of genetic markers associated with high-density lipoprotein-cholesterol by genome-wide screening in a Japanese population: the Suita study. Circ J *73*, 1119-1126.

Hodgkinson,C.A., Enoch,M.A., Srivastava,V., Cummins-Oman,J.S., Ferrier,C., Iarikova,P., Sankararaman,S., Yamini,G., Yuan,Q., Zhou,Z., Albaugh,B., White,K.V., Shen,P.H., and Goldman,D. (2010). Genome-wide association identifies candidate genes that influence the human electroencephalogram. PNAS *107*, 8695-8700.

Hoggart,C.J., Clark,T.G., De,I.M., Whittaker,J.C., and Balding,D.J. (2008). Genome-wide significance for dense SNP and resequencing data. Genet. Epidemiol. *32*, 179-185.

Hokanson,J.E. and Austin,M.A. (1996). Plasma triglyceride level is a risk factor for cardiovascular disease independent of high-density lipoprotein cholesterol level: a meta-analysis of population-based prospective studies. J. Cardiovasc. Risk *3*, 213-219.

Howard,B.V., Le,N.A., Belcher,J.D., Flack,J.M., Jacobs,D.R., Jr., Lewis,C.E., Marcovina,S.M., and Perkins,L.L. (1994). Concentrations of Lp(a) in black and white young adults: relations to risk factors for cardiovascular disease. Ann. Epidemiol. *4*, 341-350.

Howard,B.V., Lee,E.T., Cowan,L.D., Devereux,R.B., Galloway,J.M., Go,O.T., Howard,W.J., Rhoades,E.R., Robbins,D.C., Sievers,M.L., and Welty,T.K. (1999). Rising tide of cardiovascular disease in American Indians. The Strong Heart Study. Circulation *99*, 2389-2395.

Huang,L.S., Voyiaziakis,E., Markenson,D.F., Sokol,K.A., Hayek,T., and Breslow,J.L. (1995). apo B gene knockout in mice results in embryonic lethality in homozygotes and neural tube defects, male infertility, and reduced HDL cholesterol ester and apo A-I transport rates in heterozygotes. J. Clin. Invest *96*, 2152-2161.

Hunter,D.J. (2005). Gene-environment interactions in human diseases. Nat. Rev. Genet. *6*, 287-298.

Huuskonen,J., Olkkonen,V.M., Jauhiainen,M., and Ehnholm,C. (2001). The impact of phospholipid transfer protein (PLTP) on HDL metabolism. Atherosclerosis *155*, 269-281.

Ioannidis,J.P.A., Ntzani,E.E., Trikalinos,T.A., and Contopoulos-Ioannidis,D.G. (2001). Replication validity of genetic association studies. Nat Genet *29*, 306-309.

Jarvis,M.J., Tunstall-Pedoe,H., Feyerabend,C., Vesey,C., and Saloojee,Y. (1987). Comparison of tests used to distinguish smokers from nonsmokers. Am. J. Public Health *77*, 1435-1438.

Jiang,X.C., Bruce,C., Mar,J., Lin,M., Ji,Y., Francone,O.L., and Tall,A.R. (1999). Targeted mutation of plasma phospholipid transfer protein gene markedly reduces high-density lipoprotein levels. J. Clin. Invest *103*, 907-914.

Johnson,C.L., Rifkind,B.M., Sempos,C.T., Carroll,M.D., Bachorik,P.S., Briefel,R.R., Gordon,D.J., Burt,V.L., Brown,C.D., Lippel,K., and . (1993). Declining serum total cholesterol levels among US adults. The National Health and Nutrition Examination Surveys. JAMA *269*, 3002-3008.

Jolliffe,C.J. and Janssen,I. (2006). Distribution of lipoproteins by age and gender in adolescents. Circulation *114*, 1056-1062.

Kamboh,M.I., Rewers,M., Aston,C.E., and Hamman,R.F. (1997). Plasma apolipoprotein A-I, apolipoprotein B, and lipoprotein(a) concentrations in normoglycemic Hispanics and non-Hispanic whites from the San Luis Valley, Colorado. Am. J. Epidemiol. *146*, 1011-1018.

Kamstrup,P.R., Tybjaerg-Hansen,A., Steffensen,R., and Nordestgaard,B.G. (2009). Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. JAMA *301*, 2331-2339.

Kanetsky,P.A., Mitra,N., Vardhanabhuti,S., Li,M., Vaughn,D.J., Letrero,R., Ciosek,S.L., Doody,D.R., Smith,L.M., Weaver,J., Albano,A., Chen,C., Starr,J.R., Rader,D.J., Godwin,A.K., Reilly,M.P., Hakonarson,H., Schwartz,S.M., and Nathanson,K.L. (2009). Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. Nat Genet *41*, 811-815.

Kang,S.J., Chiang,C.W.K., Palmer,C.D., Tayo,B.O., Lettre,G., Butler,J.L., Hackett,R., Adeyemo,A.A., Guiducci,C., Berzins,I., Nguyen,T.T., Feng,T., Luke,A., Shriner,D., Ardlie,K., Rotimi,C., Wilks,R., Forrester,T., McKenzie,C.A., Lyon,H.N., Cooper,R.S., Zhu,X., and Hirschhorn,J.N. (2010). Genome-wide association of anthropometric traits in African- and African-derived populations. Human Molecular Genetics *19*, 2725-2738.

Kariuki,S., Franek,B., Kumar,A., Arrington,J., Mikolaitis,R., Utset,T., Jolly,M., Crow,M., Skol,A., and Niewold,T. (2010). Trait-stratified genome-wide association study identifies novel and diverse genetic associations with serologic and cytokine phenotypes in systemic lupus erythematosus. Arthritis Research & Therapy *12*, R151.

Kathiresan,S., Melander,O., Guiducci,C., Surti,A., Burtt,N.P., Rieder,M.J., Cooper,G.M., Roos,C., Voight,B.F., Havulinna,A.S., Wahlstrand,B., Hedner,T., Corella,D., Tai,E.S., Ordovas,J.M., Berglund,G., Vartiainen,E., Jousilahti,P., Hedblad,B., Taskinen,M.R., Newton-Cheh,C., Salomaa,V., Peltonen,L., Groop,L., Altshuler,D.M., and Orho-Melander,M. (2008). Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat. Genet. *40*, 189-197.

Kathiresan,S., Willer,C.J., Peloso,G.M., Demissie,S., Musunuru,K., Schadt,E.E., Kaplan,L., Bennett,D., Li,Y., Tanaka,T., Voight,B.F., Bonnycastle,L.L., Jackson,A.U., Crawford,G., Surti,A., Guiducci,C., Burtt,N.P., Parish,S., Clarke,R., Zelenika,D., Kubalanza,K.A., Morken,M.A., Scott,L.J., Stringham,H.M., Galan,P., Swift,A.J., Kuusisto,J., Bergman,R.N., Sundvall,J., Laakso,M., Ferrucci,L., Scheet,P., Sanna,S., Uda,M., Yang,Q., Lunetta,K.L., Dupuis,J., De Bakker,P.I., O'Donnell,C.J., Chambers,J.C., Kooner,J.S., Hercberg,S., Meneton,P., Lakatta,E.G., Scuteri,A., Schlessinger,D., Tuomilehto,J., Collins,F.S., Groop,L., Altshuler,D., Collins,R., Lathrop,G.M., Melander,O., Salomaa,V., Peltonen,L., Orho-Melander,M., Ordovas,J.M., Boehnke,M., Abecasis,G.R., Mohlke,K.L., and Cupples,L.A. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. Nat. Genet. *41*, 56-65.

Kawedia,J.D., Kaste,S.C., Pei,D., Panetta,J.C., Cai,X., Cheng,C., Neale,G., Howard,S.C., Evans,W.E., Pui,C.H., and Relling,M.V. (2010). Pharmacokinetic, pharmacodynamic and pharmacogenetic determinants of osteonecrosis in children with acute lymphoblastic leukemia. Blood.

Keebler,M.E., Sanders,C.L., Surti,A., Guiducci,C., Burtt,N.P., and Kathiresan,S. (2009). Association of blood lipids with common DNA sequence variants at 19 genetic loci in the multiethnic United States National Health and Nutrition Examination Survey III. Circ. Cardiovasc. Genet. *2*, 238-243.

Keebler,M.E., Deo,R.C., Surti,A., Konieczkowski,D., Guiducci,C., Burtt,N., Buxbaum,S.G., Sarpong,D.F., Steffes,M.W., Wilson,J.G., Taylor,H.A., and Kathiresan,S. (2010). Fine-Mapping in African Americans of 8 Recently Discovered Genetic Loci for Plasma Lipids: The Jackson Heart Study. Circ Cardiovasc Genet *3*, 358-364.

Kolonel,L.N., Altshuler,D., and Henderson,B.E. (2004). The multiethnic cohort study: exploring genes, lifestyle and cancer risk. Nat. Rev. Cancer *4*, 519-527.

Kooner,J.S., Chambers,J.C., guilar-Salinas,C.A., Hinds,D.A., Hyde,C.L., Warnes,G.R., Gomez Perez,F.J., Frazer,K.A., Elliott,P., Scott,J., Milos,P.M., Cox,D.R., and Thompson,J.F. (2008). Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. Nat. Genet. *40*, 149-151.

Korn,E.L. and Graubard,B.I. (1991). Epidemiologic studies utilizing surveys: accounting for the sampling design. Am. J. Public Health *81*, 1166-1173.

Kraft,H.G., Haibach,C., Lingenhel,A., Brunner,C., Trommsdorff,M., Kronenberg,F., Muller,H.J., and Utermann,G. (1995). Sequence polymorphism in kringle IV 37 in

linkage disequilibrium with the apolipoprotein (a) size polymorphism. Hum. Genet. *95*, 275-282.

Kraft,H.G., Lingenhel,A., Pang,R.W., Delport,R., Trommsdorff,M., Vermaak,H., Janus,E.D., and Utermann,G. (1996). Frequency distributions of apolipoprotein(a) kringle IV repeat alleles and their effects on lipoprotein(a) levels in Caucasian, Asian, and African populations: the distribution of null alleles is non-random. Eur. J. Hum. Genet. *4*, 74-87.

Kumar,P., Henikoff,S., and Ng,P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protocols *4*, 1073-1081.

Lai,C.Q., Corella,D., Demissie,S., Cupples,L.A., Adiconis,X., Zhu,Y., Parnell,L.D., Tucker,K.L., and Ordovas,J.M. (2006). Dietary intake of n-6 fatty acids modulates effect of apolipoprotein A5 gene on plasma fasting triglycerides, remnant lipoprotein concentrations, and lipoprotein particle size: the Framingham Heart Study. Circulation *113*, 2062-2070.

Lanktree,M., Oh,J., and Hegele,R.A. (2008). Genetic testing for atherosclerosis risk: inevitability or pipe dream? Can. J. Cardiol. *24*, 851-854.

Lanktree,M.B., Anand,S.S., Yusuf,S., and Hegele,R.A. (2010). Comprehensive analysis of genomic variation in the LPA locus and its relationship to plasma lipoprotein(a) in South Asians, Chinese, and European Caucasians. Circ. Cardiovasc. Genet. *3*, 39-46.

Lanktree,M.B., Rajakumar,C., Brunt,J.H., Koschinsky,M.L., Connelly,P.W., and Hegele,R.A. (2009). Determination of lipoprotein(a) kringle repeat number from genomic DNA: copy number variation genotyping using qPCR. J. Lipid Res. *50*, 768-772.

Lasky-Su,J., Lyon,H.N., Emilsson,V., Heid,I.M., Molony,C., Raby,B.A., Lazarus,R., Klanderman,B., Soto-Quiros,M.E., Avila,L., Silverman,E.K., Thorleifsson,G., Thorsteinsdottir,U., Kronenberg,F., Vollmert,C., Illig,T., Fox,C.S., Levy,D., Laird,N., Ding,X., McQueen,M.B., Butler,J., Ardlie,K., Papoutsakis,C., Dedoussis,G., O'Donnell,C.J., Wichmann,H.E., Celedon,J.C., Schadt,E., Hirschhorn,J., Weiss,S.T., Stefansson,K., and Lange,C. (2008). On the replication of genetic associations: timing can be everything! Am. J. Hum. Genet. *82*, 849-858.

Lee,E.T., Welty,T.K., Fabsitz,R., Cowan,L.D., Le,N.A., Oopik,A.J., Cucchiara,A.J., Savage,P.J., and Howard,B.V. (1990). The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. Am J Epidemiol *132*, 1141-1155.

Liao,Y., Cooper,R.S., Cao,G., Kaufman,J.S., Long,A.E., and McGee,D.L. (1997). Mortality from coronary heart disease and cardiovascular disease among adult U.S. Hispanics: findings from the National Health Interview Survey (1986 to 1994). J. Am. Coll. Cardiol. *30*, 1200-1205.

Lloyd-Jones,D., Adams,R.J., Brown,T.M., Carnethon,M., Dai,S., de,S.G., Ferguson,T.B., Ford,E., Furie,K., Gillespie,C., Go,A., Greenlund,K., Haase,N., Hailpern,S., Ho,P.M., Howard,V., Kissela,B., Kittner,S., Lackland,D., Lisabeth,L., Marelli,A., McDermott,M.M., Meigs,J., Mozaffarian,D., Mussolino,M., Nichol,G., Roger,V.L., Rosamond,W., Sacco,R., Sorlie,P., Roger,V.L., Thom,T., Wasserthiel-Smoller,S., Wong,N.D., and Wylie-Rosett,J. (2010). Heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation *121*, e46-e215.

Lohr,S.L. (1999). Sampling: Design and Analysis. (Pacific Grove, Calif.: Duxbury Press).

Luke,M.M., Kane,J.P., Liu,D.M., Rowland,C.M., Shiffman,D., Cassano,J., Catanese,J.J., Pullinger,C.R., Leong,D.U., Arellano,A.R., Tong,C.H., Movsesyan,I., Naya-Vigne,J., Noordhof,C., Feric,N.T., Malloy,M.J., Topol,E.J., Koschinsky,M.L., Devlin,J.J., and Ellis,S.G. (2007). A polymorphism in the protease-like domain of apolipoprotein(a) is associated with severe coronary artery disease. Arterioscler. Thromb. Vasc. Biol. *27*, 2030-2036.

Lundblad,V. and Szostak,J.W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. Cell *57*, 633-643.

Luoto,R., Sharrett,A.R., Schreiner,P., Sorlie,P.D., Arnett,D., and Ephross,S. (2000). Blood pressure and menopausal transition: the Atherosclerosis Risk In Communities study (1987-95). Journal of Hypertension *18*.

Lusis,A.J., Fogelman,A.M., and Fonarow,G.C. (2004). Genetic basis of atherosclerosis: part I: new genes and pathways. Circulation *110*, 1868-1873.

Lusis,A.J. and Pajukanta,P. (2008). A treasure trove for lipoprotein biology. Nat. Genet. *40*, 129-130.

Manninen,V., Tenkanen,L., Koskinen,P., Huttunen,J.K., Manttari,M., Heinonen,O.P., and Frick,M.H. (1992). Joint effects of serum triglyceride and LDL cholesterol and HDL cholesterol concentrations on coronary heart disease risk in the Helsinki Heart Study. Implications for treatment. Circulation *85*, 37-45.

Manolio,T.A. (2009). Cohort studies and the genetics of complex disease. Nat. Genet. *41*, 5-6.

Manolio,T.A., Collins,F.S., Cox,N.J., Goldstein,D.B., Hindorff,L.A., Hunter,D.J., McCarthy,M.I., Ramos,E.M., Cardon,L.R., Chakravarti,A., Cho,J.H., Guttmacher,A.E., Kong,A., Kruglyak,L., Mardis,E., Rotimi,C.N., Slatkin,M., Valle,D., Whittemore,A.S., Boehnke,M., Clark,A.G., Eichler,E.E., Gibson,G., Haines,J.L., Mackay,T.F., McCarroll,S.A., and Visscher,P.M. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753.

Manolio,T.A., Bailey-Wilson,J.E., and Collins,F.S. (2006). Genes, environment and the value of prospective cohort studies. Nat Rev Genet *7*, 812-820.

Marcovina,S.M., Albers,J.J., Wijsman,E., Zhang,Z.H., Chapman,N.H., and Kennedy,H. (1996). Differences in Lp(a) concentrations and apo(a) polymorphs between black and white Americans. J Lipid Res *37*, 2569-2585.

Marcovina,S.M., Koschinsky,M.L., Albers,J.J., and Skarlatos,S. (2003). Report of the National Heart, Lung, and Blood Institute Workshop on Lipoprotein(a) and Cardiovascular Disease: recent advances and future directions. Clin. Chem. *49*, 1785-1796.

Marsden,J. (1986). Hyperlipidaemia due to isotretinoin and etretinate: possible mechanisms and consequences. Br. J. Dermatol. *114*, 401-407.

Mathias,R.A., Grant,A.V., Rafaels,N., Hand,T., Gao,L., Vergara,C., Tsai,Y.J., Yang,M., Campbell,M., Foster,C., Gao,P., Togias,A., Hansel,N.N., Diette,G., Adkinson,N.F., Liu,M.C., Faruque,M., Dunston,G.M., Watson,H.R., Bracken,M.B., Hoh,J., Maul,P., Maul,T., Jedlicka,A.E., Murray,T., Hetmanski,J.B., Ashworth,R., Ongaco,C.M., Hetrick,K.N., Doheny,K.F., Pugh,E.W., Rotimi,C.N., Ford,J., Eng,C., Burchard,E.G., Sleiman,P.M.A., Hakonarson,H., Forno,E., Raby,B.A., Weiss,S.T., Scott,A.F., Kabesch,M., Liang,L., Abecasis,G., Moffatt,M.F., Cookson,W.O.C., Ruczinski,I., Beaty,T.H., and Barnes,K.C. (2010). A genome-wide association study on African-ancestry populations for asthma. Journal of Allergy and Clinical Immunology *125*, 336-346.

Matise, T., Ambite, J. L., Buyske, S., Cole, S. A., Crawford, D. C., Haiman, C., Heiss, H., Kooperberg, C., Le Marchand, L., Manolio, T. A., North, K., Peters, R., Ritchie, M. D., Hindorff, L., Haines, J. L., and for PAGE. The next PAGE in understanding complex traits: study design for analysis of Population Architecture using Genomics and Epidemiology. 2010.
Ref Type: Unpublished Work

McCarthy,M.I., Abecasis,G.R., Cardon,L.R., Goldstein,D.B., Little,J., Ioannidis,J.P., and Hirschhorn,J.N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. *9*, 356-369.

McCarthy,M.I. and Hirschhorn,J.N. (2008). Genome-wide association studies: potential next steps on a genetic journey. Hum. Mol. Genet. *17*, R156-R165.

Mero,N., van,T.A., Scheek,L.M., Van,G.T., Labeur,C., Rosseneu,M., and Taskinen,M.R. (1998). Decreased postprandial high density lipoprotein cholesterol and apolipoproteins A-I and E in normolipidemic smoking men: relations with lipid transfer proteins and LCAT activities. J. Lipid Res. *39*, 1493-1502.

Metcalf,P.A., Sharrett,A.R., Folsom,A.R., Duncan,B.B., Patsch,W., Hutchinson,R.G., Szklo,M., Davis,C.E., and Tyroler,H.A. (1998). African American-white differences in lipids, lipoproteins, and apolipoproteins, by educational attainment, among middle-aged adults: the Atherosclerosis Risk in Communities Study. Am. J. Epidemiol. *148*, 750-760.

Moffatt,R.J. and Stamford,B. (2006). Lipid Metabolism and Health. (Boca Raton: CRC Press).

Moliterno,D.J., Jokinen,E.V., Miserez,A.R., Lange,R.A., Willard,J.E., Boerwinkle,E., Hillis,L.D., and Hobbs,H.H. (1995). No association between plasma lipoprotein(a) concentrations and the presence or absence of coronary atherosclerosis in African-Americans. Arterioscler. Thromb. Vasc. Biol. *15*, 850-855.

Mooradian,A.D., Haas,M.J., and Wadud,K. (2006a). Ascorbic acid and alpha-tocopherol down-regulate apolipoprotein A-I gene expression in HepG2 and Caco-2 cell lines. Metabolism *55*, 159-167.

Mooradian,A.D., Haas,M.J., and Wong,N.C. (2006b). The effect of select nutrients on serum high-density lipoprotein cholesterol and apolipoprotein A-I levels. Endocr. Rev. *27*, 2-16.

Moore,J.H. and Williams,S.M. (2002). New strategies for identifying gene-gene interactions in hypertension. Ann Med *34*, 88-95.

Mooser,V., Scheer,D., Marcovina,S.M., Wang,J., Guerra,R., Cohen,J., and Hobbs,H.H. (1997). The Apo(a) gene is the major determinant of variation in plasma Lp(a) levels in African Americans. Am. J. Hum. Genet. *61*, 402-417.

Murray,J.C., Gilgor,R.S., and Lazarus,G.S. (1983). Serum triglyceride elevation following high-dose vitamin A treatment for pityriasis rubra pilaris. Arch. Dermatol. *119*, 675-676.

Nagele,U., Hagele,E.O., Sauer,G., Wiedemann,E., Lehmann,P., Wahlefeld,A.W., and Gruber,W. (1984). Reagent for the enzymatic determination of serum total triglycerides with improved lipolytic efficiency. J Clin Chem Clin Biochem *22*, 165-174.

Nakayama,K., Bayasgalan,T., Yamanaka,K., Kumada,M., Gotoh,T., Utsumi,N., Yanagisawa,Y., Okayama,M., Kajii,E., Ishibashi,S., Iwamoto,S., and The Jichi Community Genetics Team ((2009). Large scale replication analysis of loci associated with lipid concentrations in a Japanese population. J Med Genet *46*, 370-374.

Nance,W.E., Bodurtha,J., Eaves,L.J., Hewitt,J., Maes,H., Segrest,J., Meyer,J., Neale,M., and Schieken,R. (1998). Models for the longitudinal genetic analysis of same-age twins: application to HDL cholesterol. Twin. Res. *1*, 3-8.

Nelson,R.G., Sievers,M.L., Knowler,W.C., Swinburn,B.A., Pettitt,D.J., Saad,M.F., Liebow,I.M., Howard,B.V., and Bennett,P.H. (1990). Low incidence of fatal coronary heart disease in Pima Indians despite high prevalence of non-insulin-dependent diabetes. Circulation *81*, 987-995.

Nicolosi,R.J., Wilson,T.A., Lawton,C., and Handelman,G.J. (2001). Dietary effects on cardiovascular disease risk factors: beyond saturated fatty acids and cholesterol. J. Am. Coll. Nutr. *20*, 421S-427S.

Nielsen,D.A., Ji,F., Yuferov,V., Ho,A., He,C., Ott,J., and Kreek,M.J. (2010). Genome-wide association study identifies genes that may contribute to risk for developing heroin addiction. Psychiatr Genet *20*, 207-214.

Norris,J.M., Langefeld,C.D., Talbert,M.E., Wing,M.R., Haritunians,T., Fingerlin,T.E., Hanley,A.J.G., Ziegler,J.T., Taylor,K.D., Haffner,S.M., Chen,Y.d., Bowden,D.W., and Wagenknecht,L.E. (2009). Genome-wide Association Study and Follow-up Analysis of Adiposity Traits in Hispanic Americans: The IRAS Family Study. Obesity *17*, 1932-1941.

North,K.E., Howard,B.V., Welty,T.K., Best,L.G., Lee,E.T., Yeh,J.L., Fabsitz,R.R., Roman,M.J., and MacCluer,J.W. (2003). Genetic and Environmental Contributions to Cardiovascular Disease Risk in American Indians. Am. J. Epidemiol. *157*, 303-314.

O'Connell,D.L., Heller,R.F., Roberts,D.C., Allen,J.R., Knapp,J.C., Steele,P.L., and Silove,D. (1988). Twin study of genetic and environmental effects on lipid levels. Genet. Epidemiol. *5*, 323-341.

Ober,C., Nord,A.S., Thompson,E.E., Pan,L., Tan,Z., Cusanovich,D., Sun,Y., Nicolae,R., Edelstein,C., Schneider,D.H., Billstrand,C., Pfaffinger,D., Phillips,N., Anderson,R.L., Philips,B., Rajagopalan,R., Hatsukami,T.S., Rieder,M.J., Heagerty,P.J., Nickerson,D.A., Abney,M., Marcovina,S., Jarvik,G.P., Scanu,A.M., and Nicolae,D.L. (2009). Genome-wide association study of plasma lipoprotein(a) levels identifies multiple genes on chromosome 6q. J. Lipid Res. *50*, 798-806.

Ober,C. and Vercelli,D. (2011). Gene-environment interactions in human disease: nuisance or opportunity? Trends Genet.

Ogorelkova,M., Kraft,H.G., Ehnholm,C., and Utermann,G. (2001). Single nucleotide polymorphisms in exons of the apo(a) kringles IV types 6 to 10 domain affect Lp(a) plasma concentrations and have different patterns in Africans and Caucasians. Hum. Mol. Genet. *10*, 815-824.

Oliveros,L.B., Domeniconi,M.A., Vega,V.A., Gatica,L.V., Brigada,A.M., and Gimenez,M.S. (2007). Vitamin A deficiency modifies lipid metabolism in rat liver. Br. J. Nutr. *97*, 263-272.

Ordovas,J.M. (2002). HDL genetics: candidate genes, genome wide scans and gene-environment interactions. Cardiovasc. Drugs Ther. *16*, 273-281.

Palmer,N., Langefeld,C., Ziegler,J., Hsu,F., Haffner,S., Fingerlin,T., Norris,J., Chen,Y., Rich,S., Haritunians,T., Taylor,K., Bergman,R., Rotter,J., and Bowden,D. (2010). Candidate loci for insulin sensitivity and disposition index from a genome-wide association analysis of Hispanic participants in the Insulin Resistance Atherosclerosis (IRAS) Family Study. Diabetologia *53*, 281-289.

Parsons,S.K., Skapek,S.X., Neufeld,E.J., Kuhlman,C., Young,M.L., Donnelly,M., Brunzell,J.D., Otvos,J.D., Sallan,S.E., and Rifai,N. (1997). Asparaginase-associated lipid abnormalities in children with acute lymphoblastic leukemia. Blood *89*, 1886-1895.

Patrick,D.L., Cheadle,A., Thompson,D.C., Diehr,P., Koepsell,T., and Kinne,S. (1994). The validity of self-reported smoking: a review and meta-analysis. Am. J. Public Health *84*, 1086-1093.

Pelak,K., Goldstein,D., Walley,N., Fellay,J., Ge,D., Shianna,K., Gumbs,C., Gao,X., Maia,J., Cronin,K., Hussain,S., Carrington,M., Michael,N., and Weintrob,A. (2010). Host Determinants of HIV–1 Control in African Americans. The Journal of Infectious Diseases *201*, 1141-1149.

Pendergrass,S.A., Dudek,S.M., Crawford,D.C., and Ritchie,M.D. (2010). Synthesis-View: visualization and interpretation of SNP association results for multi-cohort, multi-phenotype data and meta-analysis. BioData. Min *3*, 10.

Pennacchio,L.A., Olivier,M., Hubacek,J.A., Krauss,R.M., Rubin,E.M., and Cohen,J.C. (2002). Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. Hum Mol Genet *11*, 3031-3038.

Perez-Stable,E.J., Herrera,B., Jacob,P., III, and Benowitz,N.L. (1998). Nicotine metabolism and intake in black and white smokers. JAMA *280*, 152-156.

Pirkle,J.L., Flegal,K.M., Bernert,J.T., Brody,D.J., Etzel,R.A., and Maurer,K.R. (1996). Exposure of the US population to environmental tobacco smoke: the Third National Health and Nutrition Examination Survey, 1988 to 1991. JAMA *275*, 1233-1240.

Pollin,T.I., Damcott,C.M., Shen,H., Ott,S.H., Shelton,J., Horenstein,R.B., Post,W., McLenithan,J.C., Bielak,L.F., Peyser,P.A., Mitchell,B.D., Miller,M., O'Connell,J.R., and Shuldiner,A.R. (2008). A Null Mutation in Human APOC3 Confers a Favorable Plasma Lipid Profile and Apparent Cardioprotection. Science *322*, 1702-1705.

Price,A.L., Patterson,N.J., Plenge,R.M., Weinblatt,M.E., Shadick,N.A., and Reich,D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet *38*, 904-909.

Psaty,B.M., O'Donnell,C.J., Gudnason,V., Lunetta,K.L., Folsom,A.R., Rotter,J.I., Uitterlinden,A.G., Harris,T.B., Witteman,J.C.M., Boerwinkle,E., and on Behalf of the CHARGE Consortium (2009). Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts. Circ Cardiovasc Genet *2*, 73-80.

Pui,C.H., Campana,D., Pei,D., Bowman,W.P., Sandlund,J.T., Kaste,S.C., Ribeiro,R.C., Rubnitz,J.E., Raimondi,S.C., Onciu,M., Coustan-Smith,E., Kun,L.E., Jeha,S., Cheng,C., Howard,S.C., Simmons,V., Bayles,A., Metzger,M.L., Boyett,J.M., Leung,W., Handgretinger,R., Downing,J.R., Evans,W.E., and Relling,M.V. (2009). Treating childhood acute lymphoblastic leukemia without cranial irradiation. N. Engl. J. Med. *360*, 2730-2741.

Purcell,S., Neale,B., Todd-Brown,K., Thomas,L., Ferreira,M.A.R., Bender,D., Maller,J., Sklar,P., de Bakker,P.I.W., Daly,M.J., and Sham,P.C. (2007). PLINK: A Tool Set for

Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet *81*, 559-575.

Rader,D.J., Cohen,J., and Hobbs,H.H. (2003). Monogenic hypercholesterolemia: new insights in pathogenesis and treatment. J Clin. Invest *111*, 1795-1803.

Rainwater,D.L., Kammerer,C.M., VandeBerg,J.L., and Hixson,J.E. (1997). Characterization of the genetic elements controlling lipoprotein(a) concentrations in Mexican Americans. Evidence for at least three controlling elements linked to LPA, the locus encoding apolipoprotein(a). Atherosclerosis *128*, 223-233.

Reich,D.E. and Lander,E.S. (2001). On the allelic spectrum of human disease. Trends in Genetics *17*, 502-510.

Reichenbach,P., Hoss,M., Azzalin,C.M., Nabholz,M., Bucher,P., and Lingner,J. (2003). A human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed. Curr. Biol. *13*, 568-574.

Reilly,S.L., Kottke,B.A., and Sing,C.F. (1990). The effects of generation and gender on the joint distributions of lipid and apolipoprotein phenotypes in the population at large. J. Clin. Epidemiol. *43*, 921-940.

Rhoads,G.G., Dahlen,G., Berg,K., Morton,N.E., and Dannenberg,A.L. (1986). Lp(a) lipoprotein as a risk factor for myocardial infarction. JAMA *256*, 2540-2544.

Ridker,P.M., Pare,G., Parker,A.N., Zee,R.Y.L., Miletich,J.P., and Chasman,D.I. (2009). Polymorphism in the CETP Gene Region, HDL Cholesterol, and Risk of Future Myocardial Infarction: Genomewide Analysis Among 18 245 Initially Healthy Women From the Women's Genome Health Study. Circ Cardiovasc Genet *2*, 26-33.

Risch,N. and Merikangas,K. (1996). The future of genetic studies of complex human diseases. Science *273*, 1516.

Ritchie,M.D., Hahn,L.W., Roodi,N., Bailey,L.R., Dupont,W.D., Parl,F.F., and Moore,J.H. (2001). Multi-factor dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet *69*, 138-147.

Robbins,D.C., Welty,T.K., Wang,W.Y., Lee,E.T., and Howard,B.V. (1996). Plasma lipids and lipoprotein concentrations among American Indians: comparison with the US population. Curr. Opin. Lipidol. *7*, 188-195.

Roger,V.L., Go,A.S., Lloyd-Jones,D.M., Adams,R.J., Berry,J.D., Brown,T.M., Carnethon,M.R., Dai,S., de,S.G., Ford,E.S., Fox,C.S., Fullerton,H.J., Gillespie,C., Greenlund,K.J., Hailpern,S.M., Heit,J.A., Ho,P.M., Howard,V.J., Kissela,B.M., Kittner,S.J., Lackland,D.T., Lichtman,J.H., Lisabeth,L.D., Makuc,D.M., Marcus,G.M., Marelli,A., Matchar,D.B., McDermott,M.M., Meigs,J.B., Moy,C.S., Mozaffarian,D., Mussolino,M.E., Nichol,G., Paynter,N.P., Rosamond,W.D., Sorlie,P.D., Stafford,R.S., Turan,T.N., Turner,M.B., Wong,N.D., Wylie-Rosett,J., Roger,V.L., and Turner,M.B. (2011). Heart

Disease and Stroke Statistics--2011 Update: A Report From the American Heart Association. Circulation *123*, e18-e209.

Rosenberg,N.A., Huang,L., Jewett,E.M., Szpiech,Z.A., Jankovic,I., and Boehnke,M. (2010). Genome-wide association studies in diverse populations. Nat Rev Genet *11*, 356-366.

Sabatti,C., Service,S.K., Hartikainen,A.L., Pouta,A., Ripatti,S., Brodsky,J., Jones,C.G., Zaitlen,N.A., Varilo,T., Kaakinen,M., Sovio,U., Ruokonen,A., Laitinen,J., Jakkula,E., Coin,L., Hoggart,C., Collins,A., Turunen,H., Gabriel,S., Elliot,P., McCarthy,M.I., Daly,M.J., Jarvelin,M.R., Freimer,N.B., and Peltonen,L. (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet *41*, 35-46.

Sandhu,M.S., Waterworth,D.M., Debenham,S.L., Wheeler,E., Papadakis,K., Zhao,J.H., Song,K., Yuan,X., Johnson,T., Ashford,S., Inouye,M., Luben,R., Sims,M., Hadley,D., McArdle,W., Barter,P., Kesaniemi,Y.A., Mahley,R.W., McPherson,R., Grundy,S.M., Bingham,S.A., Khaw,K.T., Loos,R.J., Waeber,G., Barroso,I., Strachan,D.P., Deloukas,P., Vollenweider,P., Wareham,N.J., and Mooser,V. (2008). LDL-cholesterol concentrations: a genome-wide association study. Lancet *371*, 483-491.

Sarwar,N., Danesh,J., Eiriksdottir,G., Sigurdsson,G., Wareham,N., Bingham,S., Boekholdt,S.M., Khaw,K.T., and Gudnason,V. (2007). Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. Circulation *115*, 450-458.

Saxena,R., Voight,B.F., Lyssenko,V., Burtt,N.P., De Bakker,P.I., Chen,H., Roix,J.J., Kathiresan,S., Hirschhorn,J.N., Daly,M.J., Hughes,T.E., Groop,L., Altshuler,D., Almgren,P., Florez,J.C., Meyer,J., Ardlie,K., Bengtsson,B.K., Isomaa,B., Lettre,G., Lindblad,U., Lyon,H.N., Melander,O., Newton-Cheh,C., Nilsson,P., Orho-Melander,M., Rastam,L., Speliotes,E.K., Taskinen,M.R., Tuomi,T., Guiducci,C., Berglund,A., Carlson,J., Gianniny,L., Hackett,R., Hall,L., Holmkvist,J., Laurila,E., Sjogren,M., Sterner,M., Surti,A., Svensson,M., Svensson,M., Tewhey,R., Blumenstiel,B., Parkin,M., Defelice,M., Barry,R., Brodeur,W., Camarata,J., Chia,N., Fava,M., Gibbons,J., Handsaker,B., Healy,C., Nguyen,K., Gates,C., Sougnez,C., Gage,D., Nizzari,M., Gabriel,S.B., Chirn,G.W., Ma,Q., Parikh,H., Richardson,D., Ricke,D., and Purcell,S. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science *316*, 1331-1336.

Schmidt,K., Kraft,H.G., Parson,W., and Utermann,G. (2006). Genetics of the Lp(a)/apo(a) system in an autochthonous Black African population from the Gabon. Eur. J. Hum. Genet. *14*, 190-201.

Scholz,M., Kraft,H.G., Lingenhel,A., Delport,R., Vorster,E.H., Bickeboller,H., and Utermann,G. (1999). Genetic control of lipoprotein(a) concentrations is different in Africans and Caucasians. Eur. J. Hum. Genet. *7*, 169-178.

Sebastiani,P., Solovieff,N., Hartley,S.W., Milton,J.N., Riva,A., Dworkis,D.A., Melista,E., Klings,E.S., Garrett,M.E., Telen,M.J., shley-Koch,A., Baldwin,C.T., and Steinberg,M.H.

(2010). Genetic modifiers of the severity of sickle cell anemia identified through a genome-wide association study. Am. J. Hematol. *85*, 29-35.

Sharrett,A.R., Ballantyne,C.M., Coady,S.A., Heiss,G., Sorlie,P.D., Catellier,D., and Patsch,W. (2001). Coronary heart disease prediction from lipoprotein cholesterol levels, triglycerides, lipoprotein(a), apolipoproteins A-I and B, and HDL density subfractions: The Atherosclerosis Risk in Communities (ARIC) Study. Circulation *104*, 1108-1113.

Shi,J., Levinson,D.F., Duan,J., Sanders,A.R., Zheng,Y., PeΓÇÖer,I., Dudbridge,F., Holmans,P.A., Whittemore,A.S., Mowry,B.J., Olincy,A., Amin,F., Cloninger,C.R., Silverman,J.M., Buccola,N.G., Byerley,W.F., Black,D.W., Crowe,R.R., Oksenberg,J.R., Mirel,D.B., Kendler,K.S., Freedman,R., and Gejman,P.V. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature *460*, 753-757.

Shriner,D., Vaughan,L.K., Padilla,M.A., and Tiwari,H.K. (2007). Problems with genome-wide association studies. Science *316*, 1840-1842.

Sleiman,P.M.A., Flory,J., Imielinski,M., Bradfield,J.P., Annaiah,K., Willis-Owen,S.A.G., Wang,K., Rafaels,N.M., Michel,S., Bonnelykke,K., Zhang,H., Kim,C.E., Frackelton,E.C., Glessner,J.T., Hou,C., Otieno,F.G., Santa,E., Thomas,K., Smith,R.M., Glaberson,W.R., Garris,M., Chiavacci,R.M., Beaty,T.H., Ruczinski,I., Orange,J.M., Allen,J., Spergel,J.M., Grundmeier,R., Mathias,R.A., Christie,J.D., von Mutius,E., Cookson,W.O.C., Kabesch,M., Moffatt,M.F., Grunstein,M.M., Barnes,K.C., Devoto,M., Magnusson,M., Li,H., Grant,S.F.A., Bisgaard,H., and Hakonarson,H. (2010). Variants of DENND1B Associated with Asthma in Children. N Engl J Med *362*, 36-44.

Smith,E.N., Bloss,C.S., Badner,J.A., Barrett,T., Belmonte,P.L., Berrettini,W., Byerley,W., Coryell,W., Craig,D., Edenberg,H.J., Eskin,E., Foroud,T., Gershon,E., Greenwood,T.A., Hipolito,M., Koller,D.L., Lawson,W.B., Liu,C., Lohoff,F., McInnis,M.G., McMahon,F.J., Mirel,D.B., Murray,S.S., Nievergelt,C., Nurnberger,J., Nwulia,E.A., Paschall,J., Potash,J.B., Rice,J., Schulze,T.G., Scheftner,W., Panganiban,C., Zaitlen,N., Zandi,P.P., Zollner,S., Schork,N.J., and Kelsoe,J.R. (2009). Genome-wide association study of bipolar disorder in European American and African American individuals. Mol Psychiatry *14*, 755-763.

Snieder,H., van Doornen,L.J., and Boomsma,D.I. (1997). The age dependency of gene expression for plasma lipids, lipoproteins, and apolipoproteins. Am. J. Hum. Genet. *60*, 638-650.

Snieder,H., van Doornen,L.J., and Boomsma,D.I. (1999). Dissecting the genetic architecture of lipids, lipoproteins, and apolipoproteins: lessons from twin studies. Arterioscler. Thromb. Vasc. Biol. *19*, 2826-2834.

Snow,B.E., Erdmann,N., Cruickshank,J., Goldman,H., Gill,R.M., Robinson,M.O., and Harrington,L. (2003). Functional conservation of the telomerase protein Est1p in humans. Curr. Biol. *13*, 698-704.

Sorlie,P.D., Backlund,E., Johnson,N.J., and Rogot,E. (1993). Mortality by Hispanic status in the United States. JAMA *270*, 2464-2468.

Sorrentino,M.J., Vielhauer,C., Eisenbart,J.D., Fless,G.M., Scanu,A.M., and Feldman,T. (1992). Plasma lipoprotein (a) protein concentration and coronary artery disease in black patients compared with white patients. Am. J. Med. *93*, 658-662.

Srinivasan,S.R., Dahlen,G.H., Jarpa,R.A., Webber,L.S., and Berenson,G.S. (1991). Racial (black-white) differences in serum lipoprotein (a) distribution and its relation to parental myocardial infarction in children. Bogalusa Heart Study. Circulation *84*, 160-167.

Stamler,J., Daviglus,M.L., Garside,D.B., Dyer,A.R., Greenland,P., and Neaton,J.D. (2000). Relationship of baseline serum cholesterol levels in 3 large cohorts of younger men to long-term coronary, cardiovascular, and all-cause mortality and to longevity. JAMA *284*, 311-318.

Steinberg,K.K., Sanderlin,K.C., Ou,C.-Y., Hannon,W.H., McQuillan,G.M., and Sampson,E.J. (1997). DNA banking in epidemiologic studies. Epidemiol Rev *19*, 156-162.

Stephens,N.G., Parsons,A., Schofield,P.M., Kelly,F., Cheeseman,K., and Mitchinson,M.J. (1996). Randomised controlled trial of vitamin E in patients with coronary disease: Cambridge Heart Antioxidant Study (CHAOS). Lancet *347*, 781-786.

Stocker,R. (1999). The ambivalence of vitamin E in atherogenesis. Trends Biochem. Sci. *24*, 219-223.

Talmud,P.J., Hawe,E., Martin,S., Olivier,M., Miller,G.J., Rubin,E.M., Pennacchio,L.A., and Humphries,S.E. (2002). Relative contribution of variation within the APOC3/A4/A5 gene cluster in determining plasma triglycerides. Hum. Mol. Genet. *11*, 3039-3046.

Talmud,P.J. and Humphries,S.E. (2002). Gene:environment interaction in lipid metabolism and effect on coronary heart disease risk. Curr. Opin. Lipidol. *13*, 149-154.

Teo,Y.Y., Small,K.S., and Kwiatkowski,D.P. (2010). Methodological challenges of genome-wide association analysis in Africa. Nat Rev Genet *11*, 149-160.

Teslovich,T.M., Musunuru,K., Smith,A.V., Edmondson,A.C., Stylianou,I.M., Koseki,M., Pirruccello,J.P., Ripatti,S., Chasman,D.I., Willer,C.J., Johansen,C.T., Fouchier,S.W., Isaacs,A., Peloso,G.M., Barbalic,M., Ricketts,S.L., Bis,J.C., Aulchenko,Y.S., Thorleifsson,G., Feitosa,M.F., Chambers,J., Orho-Melander,M., Melander,O., Johnson,T., Li,X., Guo,X., Li,M., Shin,C.Y., Jin,G.M., Jin,K.Y., Lee,J.Y., Park,T., Kim,K., Sim,X., Twee-Hee,O.R., Croteau-Chonka,D.C., Lange,L.A., Smith,J.D., Song,K., Hua,Z.J., Yuan,X., Luan,J., Lamina,C., Ziegler,A., Zhang,W., Zee,R.Y., Wright,A.F., Witteman,J.C., Wilson,J.F., Willemsen,G., Wichmann,H.E., Whitfield,J.B., Waterworth,D.M., Wareham,N.J., Waeber,G., Vollenweider,P., Voight,B.F., Vitart,V., Uitterlinden,A.G., Uda,M., Tuomilehto,J., Thompson,J.R., Tanaka,T., Surakka,I., Stringham,H.M., Spector,T.D., Soranzo,N., Smit,J.H., Sinisalo,J., Silander,K., Sijbrands,E.J., Scuteri,A., Scott,J., Schlessinger,D., Sanna,S., Salomaa,V., Saharinen,J., Sabatti,C., Ruokonen,A.,

Rudan,I., Rose,L.M., Roberts,R., Rieder,M., Psaty,B.M., Pramstaller,P.P., Pichler,I., Perola,M., Penninx,B.W., Pedersen,N.L., Pattaro,C., Parker,A.N., Pare,G., Oostra,B.A., O'Donnell,C.J., Nieminen,M.S., Nickerson,D.A., Montgomery,G.W., Meitinger,T., McPherson,R., McCarthy,M.I., McArdle,W., Masson,D., Martin,N.G., Marroni,F., Mangino,M., Magnusson,P.K., Lucas,G., Luben,R., Loos,R.J., Lokki,M.L., Lettre,G., Langenberg,C., Launer,L.J., Lakatta,E.G., Laaksonen,R., Kyvik,K.O., Kronenberg,F., Konig,I.R., Khaw,K.T., Kaprio,J., Kaplan,L.M., Johansson,A., Jarvelin,M.R., Janssens,A.C., Ingelsson,E., Igl,W., Kees,H.G., Hottenga,J.J., Hofman,A., Hicks,A.A., Hengstenberg,C., Heid,I.M., Hayward,C., Havulinna,A.S., Hastie,N.D., Harris,T.B., Haritunians,T., Hall,A.S., Gyllensten,U., Guiducci,C., Groop,L.C., Gonzalez,E., Gieger,C., Freimer,N.B., Ferrucci,L., Erdmann,J., Elliott,P., Ejebe,K.G., Doring,A., Dominiczak,A.F., Demissie,S., Deloukas,P., de Geus,E.J., de,F.U., Crawford,G., Collins,F.S., Chen,Y.D., Caulfield,M.J., Campbell,H., Burtt,N.P., Bonnycastle,L.L., Boomsma,D.I., Boekholdt,S.M., Bergman,R.N., Barroso,I., Bandinelli,S., Ballantyne,C.M., Assimes,T.L., Quertermous,T., Altshuler,D., Seielstad,M., Wong,T.Y., Tai,E.S., Feranil,A.B., Kuzawa,C.W., Adair,L.S., Taylor,H.A., Jr., Borecki,I.B., Gabriel,S.B., Wilson,J.G., Holm,H., Thorsteinsdottir,U., Gudnason,V., Krauss,R.M., Mohlke,K.L., Ordovas,J.M., Munroe,P.B., Kooner,J.S., Tall,A.R., Hegele,R.A., Kastelein,J.J., Schadt,E.E., Rotter,J.I., Boerwinkle,E., Strachan,D.P., Mooser,V., Stefansson,K., Reilly,M.P., Samani,N.J., Schunkert,H., Cupples,L.A., Sandhu,M.S., Ridker,P.M., Rader,D.J., van Duijn,C.M., Peltonen,L., Abecasis,G.R., Boehnke,M., and Kathiresan,S. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707-713.

The ARIC Investigators (1989). The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol *129*, 687-702.

The Emerging Risk Factors Collaboration (2009). Lipoprotein(a) Concentration and the Risk of Coronary Heart Disease, Stroke, and Nonvascular Mortality. JAMA: The Journal of the American Medical Association *302*, 412-423.

The International HapMap Consortium (2005). A haplotype map of the human genome. Nature *437*, 1299-1320.

Third Report of the National Cholesterol Education Program (NCEP) Expert Panel (2002). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation *106*, 3143-3421.

Thomas,D. (2010). Gene-environment-wide association studies: emerging approaches. Nat. Rev. Genet. *11*, 259-272.

von Eckardstein,A., Chirazi,A., Schuler-Luttmann,S., Walter,M., Kastelein,J.J., Geisel,J., Real,J.T., Miccoli,R., Noseda,G., Hobbel,G., and Assmann,G. (1998). Plasma and fibroblasts of Tangier disease patients are disturbed in transferring phospholipids onto apolipoprotein A-I. J. Lipid Res. *39*, 987-998.

Wagenknecht,L.E., Cutter,G.R., Haley,N.J., Sidney,S., Manolio,T.A., Hughes,G.H., and Jacobs,D.R. (1990). Racial differences in serum cotinine levels among smokers in the

Coronary Artery Risk Development in (Young) Adults study. Am. J. Public Health *80*, 1053-1056.

Wallace,C., Newhouse,S.J., Braund,P., Zhang,F., Tobin,M., Falchi,M., Ahmadi,K., Dobson,R.J., Marcano,A.C., Hajat,C., Burton,P., Deloukas,P., Brown,M., Connell,J.M., Dominiczak,A., Lathrop,G.M., Webster,J., Farrall,M., Spector,T., Samani,N.J., Caulfield,M.J., and Munroe,P.B. (2008). Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. Am. J. Hum. Genet. *82*, 139-149.

Wang,W., Hu,D., Lee,E.T., Fabsitz,R.R., Welty,T.K., Robbins,D.C., Yeh,J.L., and Howard,B.V. (2002). Lipoprotein(a) in American Indians is low and not independently associated with cardiovascular disease. The Strong Heart Study. Ann. Epidemiol. *12*, 107-114.

Warnick,G. (1986). Enzymatic methods for quantification of lipoprotein lipids. Methods Enzymol *129*, 101-123.

Warnick,G.R., Benderson,J., and Albers,J.J. (1982). Dextran sulfate-Mg2+ precipitation procedure for quantitation of high- density-lipoprotein cholesterol. Clin Chem *28*, 1379-1388.

Waterworth,D.M., Ricketts,S.L., Song,K., Chen,L., Zhao,J.H., Ripatti,S., Aulchenko,Y.S., Zhang,W., Yuan,X., Lim,N., Luan,J., Ashford,S., Wheeler,E., Young,E.H., Hadley,D., Thompson,J.R., Braund,P.S., Johnson,T., Struchalin,M., Surakka,I., Luben,R., Khaw,K.T., Rodwell,S.A., Loos,R.J.F., Boekholdt,S.M., Inouye,M., Deloukas,P., Elliott,P., Schlessinger,D., Sanna,S., Scuteri,A., Jackson,A., Mohlke,K.L., Tuomilehto,J., Roberts,R., Stewart,A., Kesaniemi,Y.A., Mahley,R.W., Grundy,S.M., Wellcome Trust Case Control Consortium, McArdle,W., Cardon,L., Waeber,G., Vollenweider,P., Chambers,J.C., Boehnke,M., Abecasis,G.R., Salomaa,V., Jarvelin,M.R., Ruokonen,A., Barroso,I., Epstein,S.E., Hakonarson,H.H., Rader,D.J., Reilly,M.P., Witteman,J.C.M., Hall,A.S., Samani,N.J., Strachan,D.P., Barter,P., van Duijn,C.M., Kooner,J.S., Peltonen,L., Wareham,N.J., McPherson,R., Mooser,V., and Sandhu,M.S. (2010). Genetic Variants Influencing Circulating Lipid Levels and Risk of Coronary Artery Disease. Arterioscler Thromb Vasc Biol *30*, 2264-2276.

Weinberg,R.B. (2002). Apolipoprotein A-IV polymorphisms and diet-gene interactions. Curr. Opin. Lipidol. *13*, 125-134.

Welty,T.K., Rhoades,D.A., Yeh,F., Lee,E.T., Cowan,L.D., Fabsitz,R.R., Robbins,D.C., Devereux,R.B., Henderson,J.A., and Howard,B.V. (2002). Changes in cardiovascular disease risk factors among American Indians. The Strong Heart Study. Ann. Epidemiol. *12*, 97-106.

Willer,C.J., Sanna,S., Jackson,A.U., Scuteri,A., Bonnycastle,L.L., Clarke,R., Heath,S.C., Timpson,N.J., Najjar,S.S., Stringham,H.M., Strait,J., Duren,W.L., Maschio,A., Busonero,F., Mulas,A., Albai,G., Swift,A.J., Morken,M.A., Narisu,N., Bennett,D., Parish,S., Shen,H., Galan,P., Meneton,P., Hercberg,S., Zelenika,D., Chen,W.M., Li,Y.,

Scott,L.J., Scheet,P.A., Sundvall,J., Watanabe,R.M., Nagaraja,R., Ebrahim,S., Lawlor,D.A., Ben-Shlomo,Y., vey-Smith,G., Shuldiner,A.R., Collins,R., Bergman,R.N., Uda,M., Tuomilehto,J., Cao,A., Collins,F.S., Lakatta,E., Lathrop,G.M., Boehnke,M., Schlessinger,D., Mohlke,K.L., and Abecasis,G.R. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat. Genet. *40*, 161-169.

Willer,C.J., Li,Y., and Abecasis,G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190-2191.

Williams,C.J. and Wijesiri,U.W. (1993). Lipid data from NHLBI veteran twins: interpreting genetic analyses when model assumptions fail. Genet. Epidemiol. *10*, 551-556.

Williams,S.M., Canter,J.A., Crawford,D.C., Moore,J.H., Ritchie,M.D., and Haines,J.L. (2007). Problems with genome-wide association studies. Science *316*, 1840-1842.

Wong,M.Y., Day,N.E., Luan,J.A., Chan,K.P., and Wareham,N.J. (2003). The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? Int. J. Epidemiol. *32*, 51-57.

Wu,J., Province,M., Coon,H., Hunt,S., Eckfeldt,J., Arnett,D., Heiss,G., Lewis,C., Ellison,R.C., Rao,D., Rice,T., and Kraja,A. (2007). An investigation of the effects of lipid-lowering medications: genome-wide linkage analysis of lipids in the HyperGEN study. BMC Genetics *8*, 60.

Yang,H., Sasaki,T., Minoshima,S., and Shimizu,N. (2007). Identification of three novel proteins (SGSM1, 2, 3) which modulate small G protein (RAP and RAB)-mediated signaling pathway. Genomics *90*, 249-260.

Yang,J.J., Cheng,C., Yang,W., Pei,D., Cao,X., Fan,Y., Pounds,S.B., Neale,G., Trevino,L.R., French,D., Campana,D., Downing,J.R., Evans,W.E., Pui,C.H., Devidas,M., Bowman,W.P., Camitta,B.M., Willman,C.L., Davies,S.M., Borowitz,M.J., Carroll,W.L., Hunger,S.P., and Relling,M.V. (2009). Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. JAMA *301*, 393-403.

Zech,L.A., Gross,E.G., Peck,G.L., and Brewer,H.B. (1983). Changes in plasma cholesterol and triglyceride levels after treatment with oral isotretinoin. A prospective study. Arch. Dermatol. *119*, 987-993.

Zollner,S. and Pritchard,J.K. (2007). Overcoming the winner's curse: estimating penetrance parameters from case-control data. Am J Hum Genet *80*, 605-615.