8-11-2011 12:00 AM

Elucidating the genetic determinants of the archetypal complex disease hypertriglyceridemia

Christopher T. Johansen
University of Western Ontario

Supervisor
Dr. Robert A. Hegele
The University of Western Ontario

Graduate Program in Biochemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

ELUCIDATING THE GENETIC DETERMINANTS OF THE ARCHETYPAL COMPLEX DISEASE HYPERTRIGLYCERIDEMIA

Spine title: Elucidating the genetic determinants of hypertriglyceridemia

(Format: Monograph)

by

Christopher T. Johansen

Graduate program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor	External Examiner
Dr. Robert A. Hegele	Dr. Ruth McPherson
Supervisory Committee	**University Examiners**
Dr. Peter K. Rogan	Dr. Richard B. Kim
Dr. Richard B. Kim	Dr. David W. Litchfield
Dr. Stephen M. Sims	Dr. C. Anthony Rupar

The thesis submitted by

Christopher T. Johansen

entitled:

Elucidating the Genetic Determinants of the Archetypal Complex Disease Hypertriglyceridemia

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date:_____

_____
Chair of Thesis Examination Board

ABSTRACT

Cardiovascular disease (CVD) is the leading cause of morbidity and mortality in Canada. Among non-traditional risk factors, plasma triglyceride (TG) concentration is re-emerging as a significant risk factor. Patients with hypertriglyceridemia (HTG) – an archetypal complex phenotype defined by fasting plasma TG concentration >95th percentile – thus have significantly increased CVD risk, compounded by associated co-morbidities such as obesity, metabolic syndrome and type 2 diabetes. However, the molecular pathways contributing to HTG susceptibility are incompletely defined. A better understanding of the genetic determinants that underlie the phenotypic spectrum of plasma TG and HTG susceptibility is necessary to identify novel genes and pathways that could be targeted to effectively lower plasma TG and improve cardiovascular risk. Accordingly, we sought to characterize the genetic architecture of HTG susceptibility and phenotypic heterogeneity using several modern genomic technologies, including high-density microarray genotyping and high-throughput resequencing of candidate genes in HTG patients and healthy controls. We demonstrate that a broad allelic spectrum of common small effect variants and rare large effect variants is associated with HTG. Furthermore, we demonstrate that significant overlap exists between genes and variants that modulate plasma TG and increase HTG susceptibility. Taken together, we suggest that HTG susceptibility is the result of a genetic burden of TG-raising alleles in genes that modulate plasma TG concentration. These findings provide a breadth of novel targets for pharmaceutical development in hopes of reducing plasma TG concentration and improving cardiovascular risk in HTG patients.

KEYWORDS: cardiovascular disease, atherosclerosis, plasma triglycerides, hypertriglyceridemia, hyperlipoproteinemia, complex diseases, genetic variation, single nucleotide polymorphisms, mutations, genome-wide association studies, rare variant accumulation, high-throughput resequencing, genetic risk scores

CO-AUTHORSHIP

The references for material published in this dissertation are listed at the beginning of each respective chapter. This section will describe the contributions of co-authors.

Dr. Robert A. Hegele (supervisor) provided funding, supervision, and patient- and population-based samples for all studies. He also contributed to study design, and manuscript preparation and critical revision for all chapters.

Dr. Jian Wang managed clinical databases, provided excellent technical assistance, and contributed to critical revision of manuscripts in chapters 2-4. Dr. Matthew B. Lanktree contributed to statistical analyses, and provided critical review comments for manuscripts in chapters 2-4.

Dr. Henian Cao, Adam D. McIntyre, Matthew R. Ban, Rebecca A. Martins, Brooke A. Kennedy, Reina G. Hassell, and Maartje E. Visser provided technical assistance for chapters 2-4.

Dr. Murray W. Huff conducted lipoprotein concentration measurements and contributed to the critical review of manuscripts in chapters 2-4.

Dr. Sekar Kathiresan contributed population-based samples through the Myocardial Infarction Genetics (MIGen) Consortium, contributed to study design in chapter 2, and provided advice and critical revision of manuscripts in chapters 2-4.

Dr. Miklós Péterfy, Dr. Margarete Mehrabian, Dr. Aldons J. Lusis, Dr. Ann-Hwee Lee, and Dr. Laurie H. Glimcher shared unpublished data regarding mouse models of hypertriglyceridemia, upon which sequencing was based, in chapter 3.

Dr. Geesje M. Dallinga-Thie contributed patient samples to chapters 2 and 4. Dr. Stephen M. Schwartz, Dr. Roberto Elosua, Dr. Veikko Salomaa, Dr. Christopher J. O'Donnell, Dr. Benjamin F. Voight contributed population-based control samples to the MIGen Consortium used in chapters 2 and 4.

Dr. Sonia S. Anand and Dr. Salim Yusuf provided population-based control samples from the Study of Health Assessment and Risk in Ethnicities (SHARE) used in chapters 2-4.

DEDICATION

For my family Debbie, Shan, Donna and Herb,
and my mentor Rob.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ADH	autosomal dominant familial hypercholesterolemia
ANGPTL3	angiopoietin-like 3
APO	apolipoprotein
APOA5	apolipoprotein A-V
APOB	apolipoprotein B
APOC2	apolipoprotein C-II
APOE	apolipoprotein E
ARIC	Atherosclerosis Risk in Communities
ATP-III	Adult Treatment Panel III
BMI	body mass index
CAD	coronary artery disease
CDCV	common-disease common-variant
CDRV	common-disease rare-variant
CE	cholesteryl ester
CETP	cholesteryl ester transfer protein
CILP2	cartilage intermediate layer protein 2
CNP	copy number polymorphism
CNV	copy number variant
CREBH	cyclic AMP responsive element binding protein H
CREB3L3	cyclic AMP responsive element binding protein 3-like 3
CVD	cardiovascular disease
DGAT	diacylglycerol acyltransferase

DNAH11	dynein, axonemal, heavy chain 11
eQTL	expression quantitative trait locus
ER	endoplasmic reticulum
FA	fatty acid
FADS1	fatty acid desaturase 1
FADS2	fatty acid desaturase 2
FADS3	fatty acid desaturase 3
FH	familial hypercholesterolemia
G1K	1000 Genomes Project
GCKR	glucokinase regulatory protein
GK	glucokinase
GKRP	glucokinase regulatory protein
GLGC	Global Lipids Genetics Consortium
GOF	gain-of-function
GPIHBP1	glycosylphosphatidylinositol-anchored HDL binding protein 1
GWAS	genome-wide association study
HDL	high-density lipoprotein
HDL-C	HDL cholesterol
HL	hepatic lipase
HLP	hyperlipoproteinemia
HTG	hypertriglyceridemia
IDL	Intermediate-density lipoprotein
IRS1	insulin receptor substrate 1

KLHL8	kelch-like 8
LD	linkage disequilibrium
LDL	low-density lipoprotein
LDL-C	LDL cholesterol
LDLR	LDL receptor
LMF1	lipase maturation factor 1
LOF	loss-of-function
LPL	lipoprotein lipase
LRP	LDLR-related protein
MAF	minor allele frequency
MGAT	monoacylglycerol acyltransferase
MI	myocardial infarction
MLXIPL	MLX interacting protein-like
MR	Mendelian randomization
MTP	microsomal triglyceride transfer protein
NAT2	N-acetyltransferase 2
NCAN	neurocan
NCEP	National Cholesterol Education Program
Polyphen-2	Polymorphism Phenotyping v2
PPARα	peroxisome proliferator-activated receptor alpha
PROCAM	Prospective Cardiovascular Munster
QTL	quantitative trait locus
RFLP	restriction fragment length polymorphisms

sdLDL	small dense LDL
SIFT	sorting intolerant from tolerant
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SORT1	sortilin
TC	total cholesterol
TG	triglyceride
TRIB1	tribbles homolog 1
T2D	type 2 diabetes
VLDL	very low-density lipoprotein
WHO	World Health Organization
ZHX2	zinc fingers and homeoboxes 2
ZHX3	zinc fingers and homeoboxes 3

CHAPTER 1

INTRODUCTION

This chapter is based on material from the following publications: (1) **Johansen CT** and Hegele RA. (2009). Predictive genetic testing for coronary artery disease. *Crit Rev Clin Lab Sci* 46(5-6): 343-60; (2) **Johansen CT**, Lanktree MB, and Hegele RA. (2010). Translating genomic analyses into improved management of coronary artery disease. *Future Cardiol* 6(4): 507-21; (3) **Johansen CT**, Kathiresan S, and Hegele RA. (2011). Genetic determinants of plasma triglycerides. *J Lipid Res* 52(2): 189-206; and (4) **Johansen CT** and Hegele RA. (2011). Genetic bases of hypertriglyceridemic phenotypes. *Curr Opin Lidiol* 22(4): 247-53; and (5) **Johansen CT** and Hegele RA. (2011). Allelic and phenotypic spectrum of plasma triglycerides. *Biochim Biophys Acta*. Submitted.

1.1 Cardiovascular disease

Cardiovascular disease (CVD) is a broad term referring to diseases involving blood vessels and the heart. Ischemic CVD is defined by a pathological process called atherosclerosis, a chronic inflammatory-metabolic disease in which accumulation of plaque causes thickening of the arterial wall (Roy et al., 2009). Acute cardiovascular events occur when atherosclerosis is sufficiently advanced to result in complete arterial occlusion from either narrowing of the artery or plaque rupture and subsequent thrombosis, processes that both prevent delivery of oxygenated blood to cardiac and

cerebral tissues, leading to ischemia and necrosis unless blood flow is restored. Coronary artery disease (CAD) refers to plaque formation specifically in the coronary arteries, which impairs cardiac blood flow causing chronic conditions such as angina pectoris or acute conditions such as myocardial infarction (MI). Cerebrovascular disease refers to occlusion of the carotid arteries, which impairs cerebral blood flow resulting in chronic conditions such as transient ischemic attacks and acute conditions such as stoke.

According to the World Health Organization (WHO), CVD is the primary global cause of death (http://www.who.int/mediacentre/factsheets/fs317/en/index.html). In 2004, 17.1 million people world-wide died from CVD (29% of deaths), including 7.2 million deaths resulting from CAD and 5.7 million deaths resulting from stroke. Projection of these statistics suggests that in the year 2030, ~23.6 million deaths will be caused by CVD. In Canada, the incidence of disease is equivalent to WHO rates, causing 1/3 of Canadian deaths and resulting in $22 billion in direct and indirect health care costs (Genest et al., 2009). Therefore, understanding of CVD susceptibility and identification of effective prevention strategies are essential to reduce the burden of this prevalent, asymptomatic and irreversible disease.

1.1.1. Pathophysiology

The formation of an atherosclerotic lesion is a complex pathophysiological process involving multiple cellular and molecular interactions between endothelial cells, monocytes, and smooth muscle cells (Wagenseil and Mecham, 2009). These cell types are found in distinct layers of the arterial wall, including the intima (inner layer), media

(middle layer) and adventia (outer layer) **(Figure 1.1)**. The intima is a layer of extracellular matrix composed of elastic fibers, proteoglycans and collagen. It is bounded by a monolayer of endothelial cells that form a selective barrier between arterial blood flow and subsequent layers of the artery wall. The media is composed of smooth muscle cells that regulate vascular tone through contraction and dilation in response to external signaling cues. The adventia is the strongest layer, composed predominantly of connective tissues such as collagen, which provide arterial structure and shape.

Lesion formation begins with endothelial cell dysfunction **(Figure 1.2)**. Physical forces such as fluid shear stresses resulting from blood flow lead to disruption of intercellular tight junctions between endothelial cells (Gimbrone, 1999). Dysregulation of the selective barrier allows passage of cholesteryl ester (CE)-rich low-density lipoprotein (LDL) into the intima-media space, which become retained, oxidized and capable of eliciting an immune response from endothelial cells (Lusis, 2000). Simultaneous exposure of endothelial cells to environmental stresses that accentuate endothelial dysfunction, including traditional risk factors such as dyslipidemia, hypertension, and dysglycemia, also stimulate endothelial cell expression of cell adhesion molecules that facilitate transmigration of monocytes from blood into the intima-media space (Libby and Theroux, 2005). Together, these processes recruit monocytes in the intima-media, which differentiate into macrophages in response to pro-inflammatory cytokines. Macrophages internalize lipoproteins via scavenger receptors on the cell surface and transform into metabolically-inactive lipid-loaded foam cells. It is the accumulation of foam cells in the intima-media which form a fatty streak, the initiation of an atherosclerotic plaque.

Figure 1.1. Simplified cross section of a normal artery. Endothelial cells provide a selective barrier between the capillary lumen and the arterial intima. The intima is a layer of extracellular matrix composed of elastic fibers, proteoglycans and collagen. The media is a layer of smooth muscle cells that regulate contractility and dilation. The adventitia is a strong layer of connective tissue that provides artery structure and shape. Atherosclerotic plaques develop between the intimal and medial layers of the artery wall. Figure adapted from (Lusis, 2000).

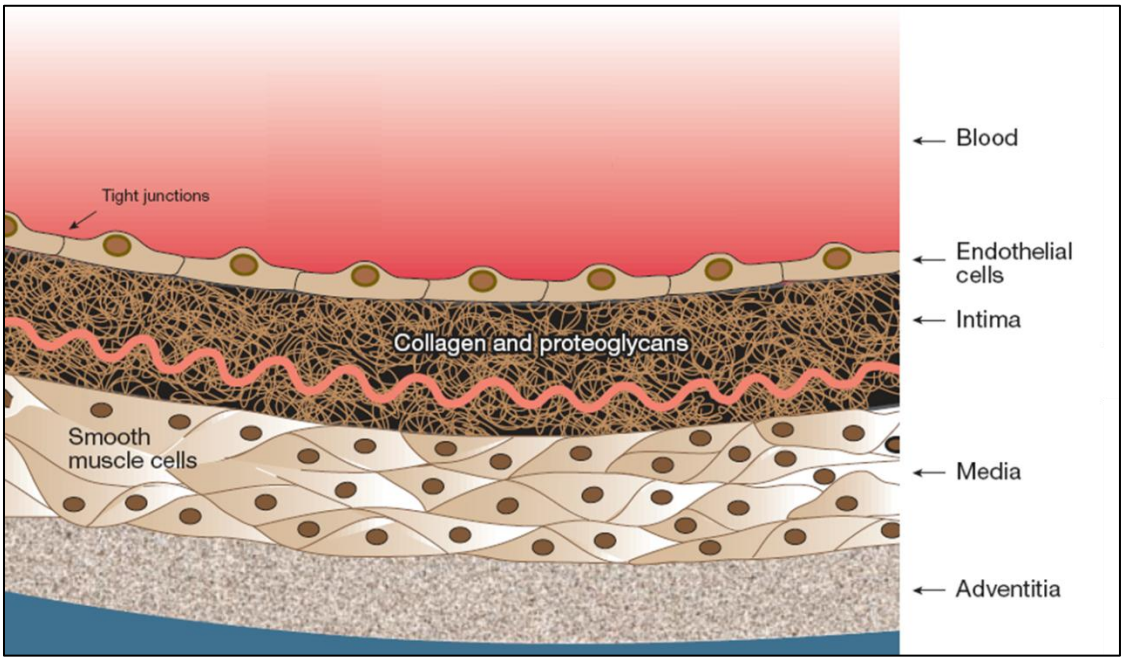
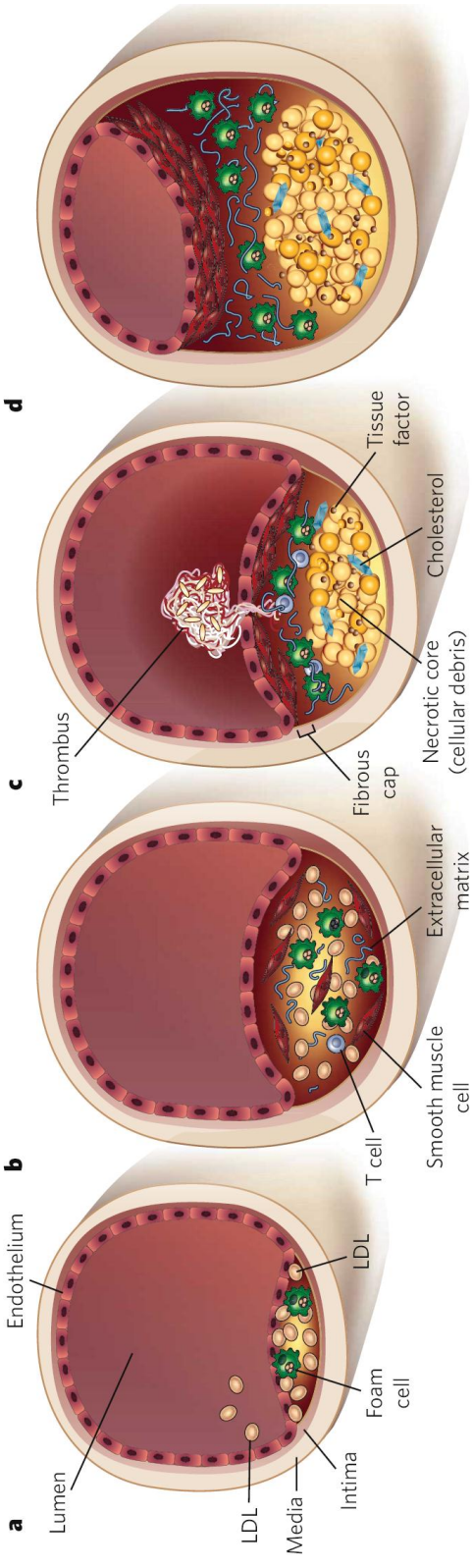

Blood
Tight junctions
Endothelial cells
Intima
Collagen and proteoglycans
Smooth muscle cells
Media
Adventitia

Figure 1.2. Simplified model of atherosclerosis progression (Rader and Daugherty, 2008). (A) Endothelial dysfunction permits movement of CE-rich LDL into the intima-media space. Oxidized LDL causes cytokine release from endothelial cells, stimulating recruitment of monocytes that differentiate into macrophages, phagocytose LDL, and become lipid-laden foam cells. (B) Smooth muscle cells are recruited to the growing lesion, where they proliferate and deposit extracellular matrix. Retention of LDL is enhanced, generating a positive feedback loop of cytokine secretion, inflammatory cell recruitment, and foam cell deposition. T-cells eventually become recruited to the growing lesion, resulting in a chronic state of inflammation. Acute cardiovascular endpoints result from two potential mechanisms: (C) plaque rupture releases tissue factor into the arterial lumen which stimulates a blood clot (thrombus) that blocks oxygenated blood flow, or (D) plaque progression narrows the arterial lumen such that complete occlusion prevents delivery of sufficient oxygenated blood. Chronic and acute cardiovascular events affecting the coronary arteries include angina pectoris and MI, whereas chronic and acute cardiovascular events affecting the carotid arteries include transient ischemic attacks and stroke.


a
b
c
d
Lumen
Endothelium
LDL
Foam cell
LDL
Media
Intima
T cell
Smooth muscle cell
Extracellular matrix
Thrombus
Fibrous cap
Necrotic core (cellular debris)
Cholesterol
Tissue factor

Accumulation of foam cells initiates a feed-forward loop of plaque progression involving cytokine secretion, monocyte recruitment and foam cell formation that progresses atherosclerotic lesion development. Eventually, the proinflammatory state of the developing atherosclerotic plaque stimulates migration of smooth muscle cells into the intima-media space, which proliferate and attempt to stabilize the growing lesion by secreting extra-cellular matrix proteins, including proteoglycans, collagen and fibronectin (Doran et al., 2008). These extra-cellular matrix components enhance retention of plasma lipoproteins and secretion of pro-inflammatory cytokines by smooth muscle cells. Recruitment of leukocytes such as T-cells leads to a chronic inflammatory state. Lipid-laden foam cells at the center of an ischemic plaque eventually die and release cellular debris and crystalline lipid into the necrotic core of the atherosclerotic plaque (Libby and Theroux, 2005). Eventually, acute complications such as MI or stroke will occur, resulting from one of two mechanisms (Rader and Daugherty, 2008): 1) expansion of the atherosclerotic plaque until arterial blood flow is completely obstructed; or 2) plaque rupture causes release of cellular debris including tissue factor, which stimulates formation of an intra-arterial blood clot called a thrombus that obstructs blood flow.

1.1.2. Traditional risk factors

Many epidemiological studies have attempted to identify clinical and biochemical variables or “biomarkers” that accurately predict atherosclerosis susceptibility in large populations. These studies employ different statistical analyses to estimate the effect of each variable on CAD risk, such that risk prediction algorithms can calculate the likelihood of experiencing an acute coronary event within a 10-year period, based on

each patient's individual data. Such risk prediction algorithms are used to stratify patients according to CAD risk to guide intensity of evidence-based treatments. For example, a low risk patient with a 10-year risk of CAD <5% would have a liberal target LDL cholesterol value and would likely not require any intervention for primary prevention of CAD in the short term. Conversely, a high-risk patient with a 10-year risk >20% would have strict target values for LDL cholesterol, blood pressure and glucose, requiring aggressive treatment to reduce risk through both therapeutic lifestyle changes and pharmacological intervention. Another emerging perspective includes implementation of 'lifetime' risk prediction algorithms to stratify patients early in life according to their potential CAD risk. Such a strategy could help guide lifestyle and potentially pharmacological interventions to reduce a patient's cumulative exposure to CAD risk factors over their lifetime, and presumably improve cardiovascular outcomes (Lloyd-Jones, 2006).

Risk prediction algorithms include the Framingham risk score (Wilson et al., 1998), the Reynolds risk score (Ridker et al., 2007, Ridker et al., 2008), the Prospective Cardiovascular Munster (PROCAM) (Assmann et al., 2002), and the Atherosclerosis Risk in Communities (ARIC) Study Cardiovascular Risk Score (Morrison et al., 2007). These algorithms are very similar, differing mainly in details due to the populations in which they were derived. In both Canada and United States, the Framingham risk score is currently used to classify and treat patients according to their potential CAD risk as indicated by the 2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of dyslipidemia and cardiovascular disease in the adult (Genest et

al., 2009) and National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP-III) guidelines (Expert Panel on Detection; Evaluation; and Treatment of High Blood Cholesterol in Adults, 2001), respectively. It is calculated using a patient's age, gender, blood pressure, total cholesterol, high-density lipoprotein (HDL)-cholesterol, diabetes status, and smoking status. These variables are robustly associated with CAD across multiple epidemiological studies; however additional variables are occasionally implemented in other algorithms, including family history, additional lipid markers, apolipoproteins, and novel biomarkers such as high sensitivity C-reactive protein.

Select CAD risk factors worth discussion are age, gender, dyslipidemia, hypertension, type 2 diabetes (T2D), and family history of CAD. Advanced age is a non-modifiable risk factor for CAD, but the strongest predictor of cardiovascular events (Sniderman and Furberg, 2008). Direct effects of advanced age include increased stiffening of the artery walls, vascular calcification, endothelial cell dysfunction and cell adhesion molecule expression (Lee and Oh, 2010). Other traditional biochemical risk factors such as total cholesterol, blood pressure, body mass index and sex hormones are also correlated with increasing age, suggesting that indirect effects of advanced age may also increase cardiovascular risk (Jousilahti et al., 1999).

Gender is also a non-modifiable risk factor for CAD. Female sex appears to have a protective effect on atherosclerosis progression in pre-menopausal women compared to age-matched men; CAD onset appears to be delayed by approximately 10-years. This gender disparity has been attributed to the cardioprotective effects of estrogen. These

include antioxidant effects, improved lipid profiles including decreased LDL cholesterol and increased HDL cholesterol, and reduced vascular inflammation (Villablanca et al., 2010). The mechanisms underlying the protective effect of estrogen are poorly understood, as cardioprotection is only observed in pre-menopausal women, while paradoxically hormone replacement therapy does not appear to reduce cardiovascular risk if initiated more than 10 years post-menopause (Perez-Lopez et al., 2010).

Dyslipidemia is the most well-established modifiable atherosclerosis risk factor (Kannel et al., 1979). Increased total cholesterol, increased LDL cholesterol and decreased HDL cholesterol all contribute significantly to elevated cardiovascular risk. Increased concentrations of LDL cholesterol have a direct role in atherosclerosis progression, as discussed above. Randomized controlled trials have demonstrated that pharmacological lowering of LDL cholesterol is an effective measure to prevent atherosclerosis, decreasing risk of CAD proportionally to the decrease in LDL cholesterol (Baigent et al., 2010, Baigent et al., 2005). By comparison, decreased concentrations of HDL cholesterol may impair effective reverse cholesterol transport, a process that effluxes unesterified cholesterol from peripheral tissues for delivery to the liver and subsequent excretion (Rader, 2007). HDL particles mediate the efflux of cholesterol from macrophages in atherosclerotic plaques; thus, HDL cholesterol concentration has a strong inverse relationship with intima-media thickness (Khera et al., 2011). Interestingly, pharmacological strategies used to increase HDL cholesterol would be expected to attenuate foam cell formation and confer atheroprotection in humans, however only modest evidence supporting a positive therapeutic effect for increasing HDL cholesterol

concentration currently exists (Singh et al., 2007). Future investigation of different pharmacological agents with various biological effects on components of the reverse-cholesterol transport pathway may eventually reveal therapies with potential atheroprotective effects.

Other modifiable CAD risk factors including increased blood pressure and T2D also accentuate the metabolic dysfunction that perpetuates CAD. For instance, normal fluid dynamics cause shear stresses that disrupt intercellular tight junctions between endothelial cells. Thus, it is not surprising that high blood pressure (or hypertension) is also strongly associated with CAD (Vasan et al., 2001). Increased systolic blood pressure accentuates sheer stresses on the vascular endothelium, which stimulates a pro-adhesive endothelial cell phenotype causing greater recruitment of monocytes into growing atherosclerotic plaques (Malyszko, 2010). Similarly, impaired insulin signaling that accompanies increased plasma concentrations of glucose and insulin in T2D disrupts metabolic processes in multiple tissues that enhance conditions such as dyslipidemia, hypertension, arterial hardening, and endothelial dysfunction (Ginsberg, 2000). T2D is considered a CAD equivalent by the Canadian and ATP-III Guidelines, meaning that a patient with established T2D has a similar chance of experiencing an acute cardiovascular event as a patient without T2D who has already experienced an event (Juutilainen et al., 2005).

Interestingly, a very strong independent CAD risk factor not included in most risk prediction algorithms is a family history of CAD (Lusis et al., 2004). A family history of

premature CAD is thought to encompass several genetic variables that contributes to intermediate trait modulation and CAD susceptibility, including genetic variants, epigenetic factors, and higher-order gene-gene and gene-environment interactions that are essential for disease pathogenesis but would otherwise be difficult to directly implicate in disease susceptibility (Lanktree and Hegele, 2009). Many studies have shown that family history is an independent risk factor for CAD (Arnett et al., 2007, Lloyd-Jones et al., 2004, Michos et al., 2005, Murabito et al., 2005, Taraboanta et al., 2008). Even in the Framingham Heart Study, family history was determined to be an independent risk factor for CAD, although due to a lack of thorough documentation for all study subjects, it was not included in the multivariate regression equation that gave rise to the Framingham risk score (Myers et al., 1990). Regardless, family history is still considered to be a major risk factor for CAD by the Canadian and ATP-III guidelines (Expert Panel on Detection; Evaluation; and Treatment of High Blood Cholesterol in Adults, 2001, Genest et al., 2009); it is also included as a variable in the Reynolds risk score algorithm (Ridker et al., 2007, Ridker et al., 2008). Similar associations continue to show that family history of CAD is a risk factor for disease even in multi-ethnic samples, and for various endophenotypes of cardiovascular disease, including coronary artery calcification, ischemic cerebrovascular disease, and carotid intima-media thickness (Choi et al., 2009, Nasir et al., 2007, Wang et al., 2003).

1.1.3. Plasma triglyceride concentration as an emerging risk factor

Traditional risk factors are well established in CAD susceptibility, however many novel or emerging risk factors have been proposed to contribute to atherosclerosis progression.

For instance, multiple lines of evidence suggest that plasma triglyceride (TG) concentration is another significant risk factor for CAD, independent of other plasma lipoproteins. Most historical studies have studied fasting plasma TG concentration as a risk factor, demonstrating that it is associated with an increased burden of both CAD (Cullen, 2000, Morrison and Hokanson, 2009) and ischemic stroke (Labreuche et al., 2009). A longstanding estimate originating from meta-analysis of 17 studies including 46,413 men and 10,864 women indicated that each 1 mM increase in fasting plasma TG concentration increased risk of CAD by 32% and 76%, respectively (Hokanson and Austin, 1996). More recently, investigations have demonstrated that nonfasting plasma TG concentration is additionally associated with cardiovascular events in women, increasing cardiovascular risk by 67% per 1 mM increase in plasma TG concentration, (Bansal et al., 2007). Similar associations have emerged from additional epidemiological studies in both men and women, reporting significant associations with increased risk of both CAD (Nordestgaard et al., 2007) and ischemic stroke (Freiberg et al., 2008). Together with historical data, these recent observations strongly underscore the re-emergence of plasma TG concentrations as a risk factor for CAD.

1.2. Plasma triglycerides and triglyceride-rich lipoprotein metabolism

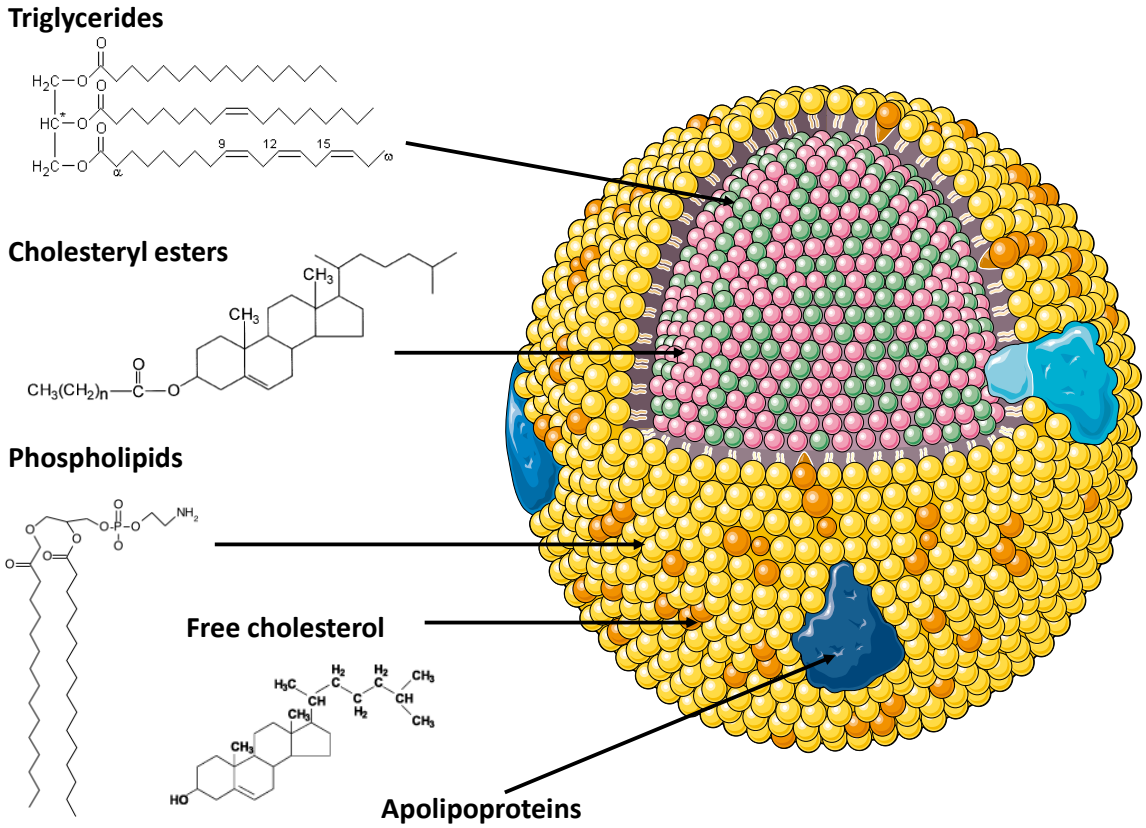
1.2.1. Plasma lipids

Three essential types of lipids are transported in plasma by lipoproteins: cholesterol, phospholipids and TG (**Figure 1.3**). Cholesterol has several important biological functions, including maintenance of fluidity in cell membranes, steroid synthesis in steroidogenic tissues, and synthesis of bile acids. It is a hydrophobic molecule primarily synthesized by the liver but also obtained exogenously from diet and transported through the plasma in lipoproteins to target organs. Phospholipids are amphipathic molecules composed of a 3-carbon glycerol molecule esterified to 2 hydrophobic fatty acids (FAs) and 1 hydrophilic phosphate group. Their amphipathic properties make them an essential component of cell membranes, lipoproteins and micelles. TG is an insoluble molecule comprised of three FAs esterified to a glycerol backbone. TG molecules serve multiple biological functions, including transport of high energy FAs to muscle cells for β-oxidation and to adipocytes for storage in lipid droplets. TG-mediated sequestration of FAs is also necessary to protect cells from lipotoxicity due to their inherent detergent properties, and to regulate their action as energy sources, ligands and signaling molecules. TG is obtained primarily from our diet, although tissues such as adipose and muscle can synthesize TG *de novo*.

1.2.2. Plasma lipoproteins

Plasma lipids are primarily transported through plasma in specialized structures called lipoproteins **(Figure 1.3)**. Conceptually, lipoproteins are spherical biochemical particles

Figure 1.3. Plasma lipids are transported through plasma in structures called lipoproteins. Plasma lipids including CE and TG are transported in the hydrophobic core of plasma lipoproteins, solubilized by amphipathic phospholipids and free cholesterol. Apolipoproteins embedded in the surface of lipoproteins mediate interactions of different lipoprotein species with specific cellular receptors and enzymes. This figure was created using Servier Medical Art.


Triglycerides
Cholesteryl esters
Phospholipids
Free cholesterol
Apolipoproteins

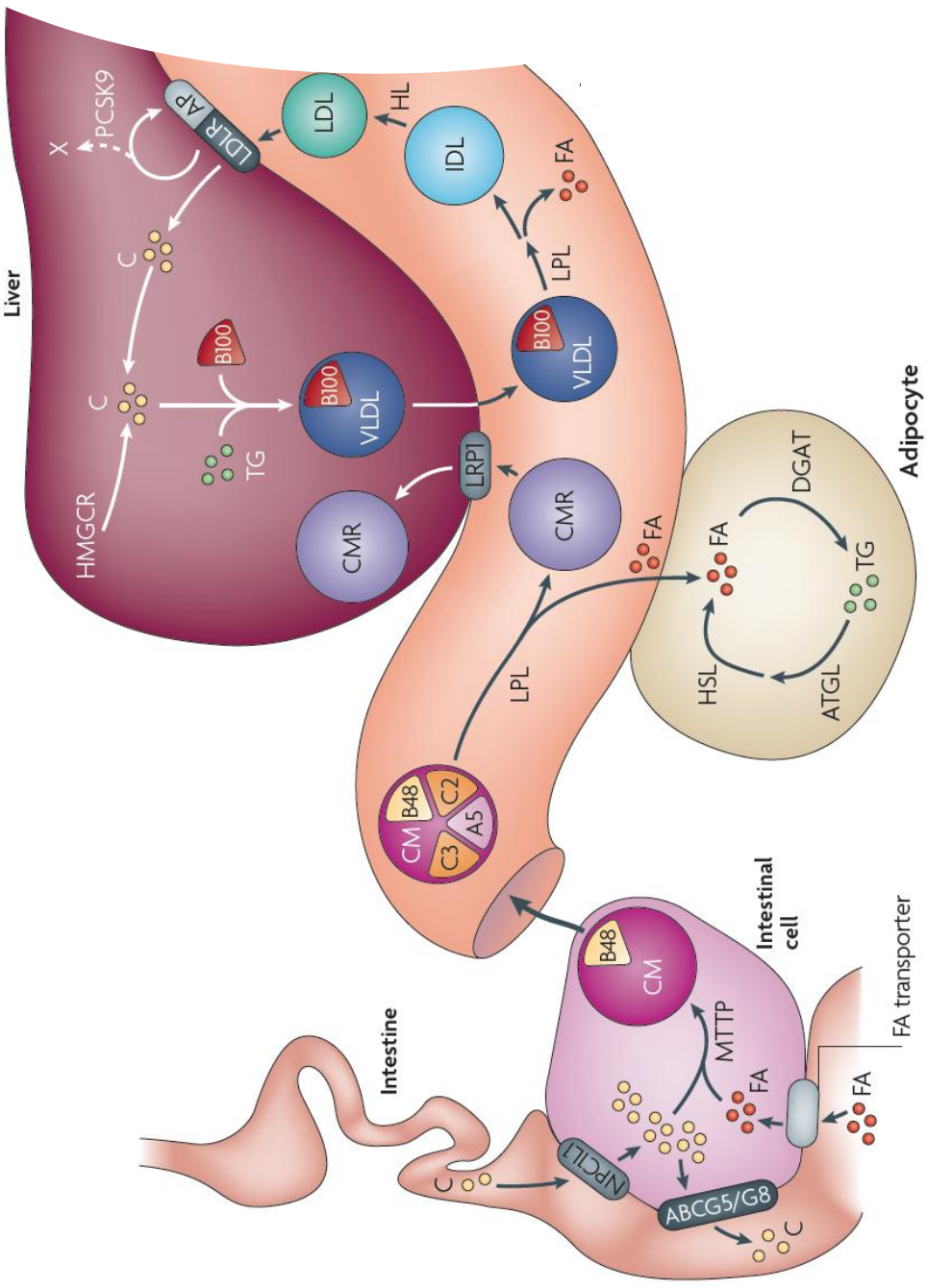
of variable size that contain a hydrophobic core of non-polar lipids including TG or CE solubilized by an outer membrane of phospholipid and free cholesterol. Lipoproteins also contain accessory proteins called apolipoproteins embedded in their surface that modulate lipoprotein interactions with their environment. Depending upon their location of synthesis, target destination, apolipoprotein content and specific lipid composition, lipoproteins have a range of physiological roles mediated through interactions with different cell surface receptors and metabolizing enzymes.

Importantly, biochemical measurement of plasma TG concentration is effectively an integrated measurement of TG-rich lipoprotein concentration, including intestinally-synthesized chylomicrons in the post-prandial state and hepatically-synthesized very low-density lipoproteins (VLDL) in the fasted state. Chylomicrons are composed of 80-95% TG obtained exogenously from diet, whereas VLDL is composed of 45-65% TG synthesized endogenously or recycled from chylomicron-remnants.

1.2.3. Plasma lipoprotein metabolism

TG metabolism can be considered as being compartmentalized into three separate processes: exogenous TG absorption and endogenous TG synthesis, intracellular TG-rich lipoprotein processing and circulating TG-rich lipoprotein metabolism **(Figure 1.4)**. Exogenous dietary TG is catabolized by gastric lipases into glycerol and FA that are subsequently emulsified into mixed micelles by phospholipid, free cholesterol and bile salts (Iqbal and Hussain, 2009). From these structures, FAs are transported across the

Figure 1.4. Plasma lipoprotein metabolism. Dietary TG is hydrolyzed by gastric lipases into free FAs. FA is subsequently transported into intestinal enterocytes by both passive and active mechanisms. Within enterocytes, TG is re-assembled from its components and packaged into chylomicrons by the lipid-loading action of microsomal TG transfer protein (MTP) and apolipoprotein (apo) B-48. Chylomicrons reach plasma via the superior vena cava where they acquire apolipoproteins from HDL that facilitate their metabolism. Chylomicrons are hydrolyzed by lipoprotein lipase (LPL), to release free FAs to peripheral cells such as adipocytes for TG storage or myocytes for oxidation. Chylomicron-remnants are removed from circulation by interaction of apo E with hepatic low-density lipoprotein receptor (LDLR)-related protein (LRP). In hepatocytes, endogenously synthesized TG and residual TG from chylomicron-remnants are packaged into VLDL by the lipid-loading action of MTP and apo B-100. VLDL is secreted into circulation for delivery of TG to peripheral tissues similarly to chylomicrons. VLDL is metabolized into intermediate-density lipoprotein (IDL) by LPL, and further into low-density lipoprotein (LDL) by hepatic lipase (HL). LDL is removed from circulation by the interaction of apo B-100 with LDLR. Figure adapted from (Hegele, 2009).


Liver
HMGCR
C
TG
B100
VLDL
CMR
LRP1
LDLR
AP
PCSK9
X
LDL
HL
IDL
LPL
FA
CM
B48
C2
A5
C3
Intestine
Intestinal cell
NPC1L1
ABCG5/G8
MTTP
FA transporter
Adipocyte
HSL
ATGL
DGAT

intestinal brush border of the duodenum by passive diffusion and FA binding protein-mediated mechanisms (Mansbach and Gorelick, 2007). Within the intestinal enterocytes, TG is re-synthesized from its components via progressive enzymatic reactions involving monoacylglycerol acyltransferase (MGAT) and diacylglycerol acyltransferase (DGAT) in the endoplasmic reticulum (ER) (Shi and Cheng, 2009).

In parallel, dietary carbohydrate is converted into FA for esterification in TG and storage in adipose tissue in response to glucose and insulin signaling. Key hepatic genes involved in this process include the sterol regulatory element binding protein 1c (*SREBP1*), carbohydrate response element binding protein (*CHREBP*; also known as *MLXIPL*), and liver X receptor (*LXR*). *SREBP1* is a transcription factor that mediates the effect of insulin by increasing expression of genes required for FA and TG synthesis including fatty acid synthase (*FAS*) and acetyl CoA carboxylase (*ACC*) (Dentin et al., 2005). Similarly, *CHREBP* is a transcription factor that mediates the effects of glucose by increasing expression of lipogenic genes *FAS* and *ACC*, but also genes involved in glycolysis such as liver pyruvate kinase (Poupeau and Postic, 2011). *LXR* is a master regulator of many genes required for lipid metabolism in response to cholesterol metabolites (Ducheix et al., 2011), including both *SREBP1* (Repa et al., 2000) and *CHREBP* (Cha and Repa, 2007). Together, these three regulators of hepatic TG synthesis stimulate coordinated expression of genes required to convert glucose into acetyl-CoA for FA synthesis and further endogenous TG synthesis in the fed state, whereas FA flux

from TG stored in adipose tissue is used to synthesize endogenous hepatic TG in the fasted state (Barrows and Parks, 2006).

Exogenous TG is subsequently packaged into chylomicrons in enterocytes, whereas endogenously synthesized TG is packaged into VLDL in hepatocytes for delivery to peripheral tissues (Iqbal and Hussain, 2009). This is accomplished by cotranslation of apolipoprotein (apo) B with the lipid loading action of the ER chaperone microsomal TG transfer protein (MTP) (Jiang et al., 2008). In enterocytes, the apo B-100 backbone is truncated by an RNA editing enzyme called APOBEC1, which deaminates the full length *APOB* transcript encoding apo B-100 and introduces a stop codon at 48% of the full protein length to produce apo B-48 (Conticello, 2008). Apo B-48 is thus a smaller protein product lacking a crucial receptor binding domain. Hepatocytes do not express APOBEC1, and thus only synthesize apo B-100. Chylomicrons are secreted into lymph and reach systemic circulation via the thoracic duct and the superior vena cava, whereas VLDL is secreted directly into circulation.

Within plasma, chylomicrons and VLDL acquire additional apolipoproteins from HDL that mediate their interaction with membrane receptors and other enzymes. Chylomicrons and VLDL are hydrolyzed by lipoprotein lipase (LPL) in capillaries of muscle and adipose tissue, where free FAs are delivered for oxidation or re-synthesized and storage, respectively. The process of hydrolysis is mediated by several proteins. First, glycosylphosphatidylinositol-anchored HDL binding protein 1 (GPIHBP1) facilitates interaction of the chylomicron with LPL (Beigneux et al., 2007). GPIHBP1 tethers

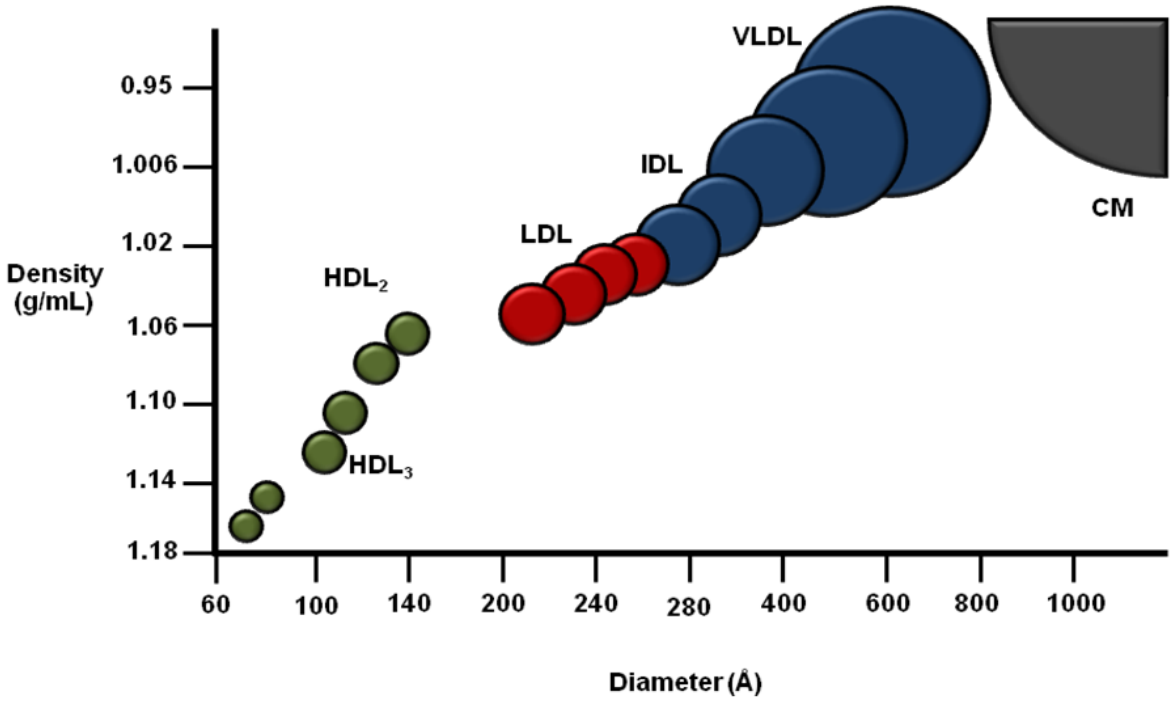
chylomicrons to the endothelial capillary lumen, binding both LPL and apo A-V, and allowing efficient TG hydrolysis by LPL. Apo C-II is a cofactor required for LPL activity, whereas apo C-III inhibits LPL activity. TG-rich lipoproteins eventually lose their apolipoproteins to HDL as hydrolysis alters lipoprotein surface area.

TG-rich lipoproteins are metabolized into a heterogeneous spectrum of other plasma lipoproteins **(Figure 1.5)**. Chylomicron-remnants largely depleted of TG are removed from circulation through the interaction of apo E with the hepatic LDL receptor (LDLR) and LDLR-related protein (LRP). Their remnant TG is reassembled into VLDL for secretion and redistribution to peripheral tissues in the fasted state. Conversely, VLDL is metabolized into intermediate density lipoprotein (IDL), which is either removed from circulation via interaction of apo B-100 with the LDLR or remodeled by the enzyme hepatic lipase (HL) into LDL. LDL is necessary to transport cholesterol to peripheral cells, but it is eventually cleared from plasma by interaction of apo B-100 with hepatic LDLR.

1.2.4. TG-rich lipoproteins and atherosclerosis progression

Multiple direct and indirect mechanisms have been proposed that could explain why plasma TG concentration is associated with increased cardiovascular risk (Botham et al., 2007, Kannel and Vasan, 2009, Lopez-Miranda et al., 2007). It has been demonstrated in human subjects that lipoprotein remnants (including chylomicron-remnants and VLDL-remnants

Figure 1.5. Spectrum of plasma lipoproteins. Intestinally-synthesized chylomicrons are the largest and least dense lipoproteins. Chylomicrons transport predominantly dietary TG; they are metabolized by LPL into chylomicron-remnants. Hepatically-synthesized VLDL transport endogenously synthesized TG; they are metabolized by LPL into IDL and further into LDL by HL. The size differences between lipoprotein subspecies are a result of relative CE and TG composition, in addition to time spent in circulation. HDL is atheroprotective and involved in efflux of cholesterol from peripheral tissues including macrophages. Chylomicrons are likely the only lipoprotein species that are unable to cross the selective endothelial cell barrier into the arterial wall, whereas their remnants and smaller lipoproteins are able to cross the barrier and contribute directly to atherogenesis (or cholesterol efflux for HDL). Figure courtesy of M. Beyea, adapted from (Vance and Vance, 2004).


VLDL
IDL
LDL
CM
HDL2
HDL3
Density (g/mL)
0.95
1.006
1.02
1.06
1.10
1.14
1.18
60
100
140
200
240
280
400
600
800
1000
Diameter (Å)

are small enough to cross the arterial wall similar to LDL (Shaikh et al., 1991), where they can induce macrophage foam-cell formation without prior oxidation (as is required for LDL uptake by macrophages). Intestinally-derived chylomicrons are not likely capable of crossing the arterial wall without prior metabolism due to their relatively large size (**Figure 1.5**), however their remnants are small enough to pass through a damaged endothelium (Mamo et al., 1998). The CE-rich core of remnants, including IDL which has an equimolar ratio of TG to CE, makes them potently atherogenic. Furthermore, increased concentrations of plasma TG-rich lipoproteins may saturate LPL-mediated lipolysis resulting in prolonged residence of remnants in circulation, extending their opportunity to interact with the arterial wall and contribute to atherosclerosis.

Another hypothesis is that elevated plasma concentrations of TG-rich lipoproteins could give rise to increased numbers of potently atherogenic small dense LDL (sdLDL) (Kannel and Vasan, 2009). Since VLDL is metabolized into LDL, elevated plasma concentration of TG-rich lipoproteins would give rise to increased concentrations of LDL. The increase in both TG-rich lipoproteins and LDL also result in competitive inhibition for LDLR binding sites, and thus prolonged residence in plasma. This increases their opportunity for interaction with the CE transfer protein (CETP), which exchanges CE in LDL and HDL for TG in TG-rich lipoproteins. Elevated concentrations of TG-rich lipoproteins cause greater exchange of CE/TG, and greater deposition of TG in LDL and HDL, resulting in increased remodeling by HL to deplete TG from the lipoprotein core, which produces sdLDL. The decreasing size of these LDL derivatives is thought to confer increased atherogenicity because they are less amenable to uptake by the LDLR,

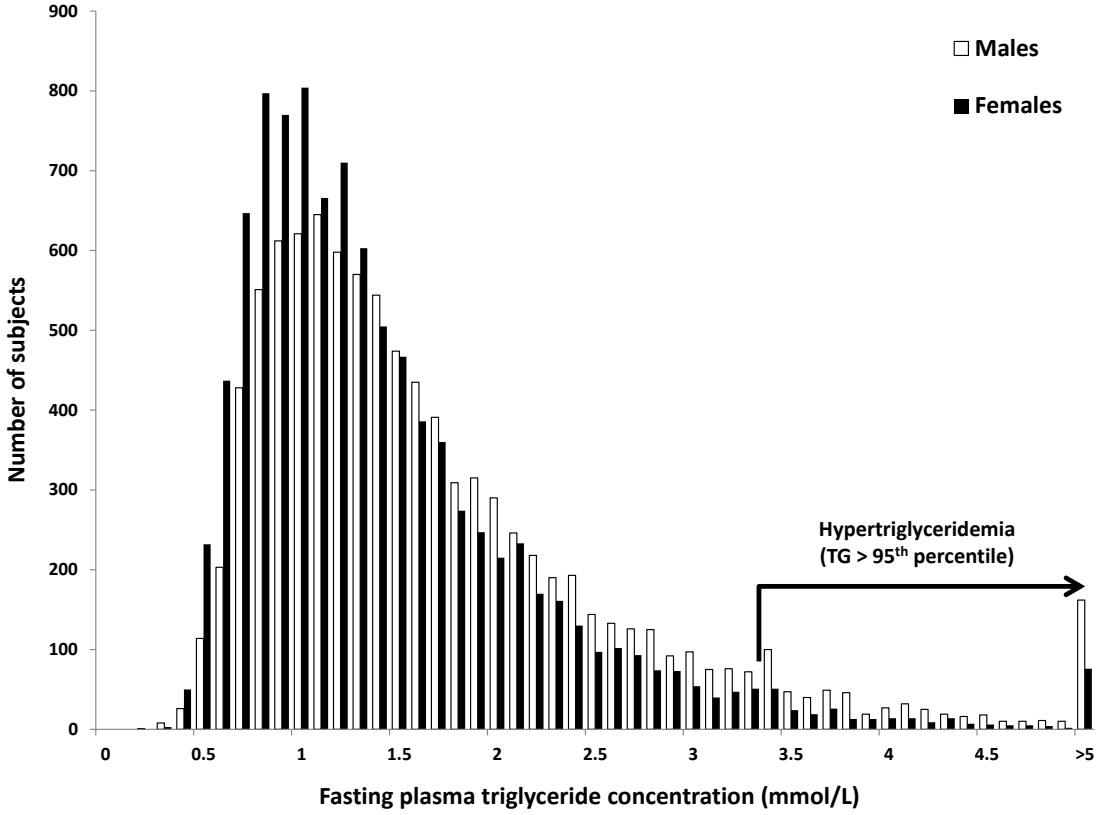
more easily taken into the arterial wall and are more prone to oxidation. The same mechanism by which LDL is remodeled into sdLDL via CETP also causes depletion of CE from HDL, which enhances its clearance from plasma and subsequently attenuates the capacity for reverse cholesterol transport and resultant atheroprotection. Together, these mechanisms provide multiple hypotheses by which increased concentrations of plasma TG-rich lipoproteins could directly and indirectly increase cardiovascular risk. However, the true mechanism is likely a complex interaction of several variables that will be difficult to prove definitively.

1.2.5 Clinical hypertriglyceridemia

A severe clinical phenotype of elevated plasma TG concentration is hypertriglyceridemia (HTG). It is defined as an age and sex adjusted fasting plasma TG concentration >95th percentile (Yuan et al., 2007), corresponding generally to a plasma concentration >3.37 mmol/L in the adult Canadian population (MacLean et al., 1992) **(Figure 1.6)**. HTG can occur as a primary genetic disease or as a symptom secondary to associated conditions such as obesity, metabolic syndrome, T2D, excessive alcohol consumption, renal disease treatment with pharmacological agents, and exposure to estrogen either exogenously or endogenously in late pregnancy, which either result in over-production or improper catabolism of TG-rich lipoproteins.

The greatest clinical consequence of HTG is increased cardiovascular risk resulting from significantly increased concentrations and prolonged residence of TG-rich

Figure 1.6. Frequency distribution of fasting plasma TG concentrations. White bars represent male subjects, black bars represent female subjects. Subjects with plasma TG concentration >3.37 mmol·L^{-1} are in the 95^{th} percentile, considered the threshold for HTG. The maximum plasma TG concentration in this sample was 45 mmol·L^{-1}. Data were obtained from the Canadian Heart Health Survey, a cross-sectional population-based study including >26,000 participants of multiple ancestries and ages 18-74, from metropolitan, urban and rural areas of Canada (MacLean et al., 1992).


900
800
700
600
500
400
300
200
100
0
Number of subjects
Males
Females
Hypertriglyceridemia
(TG > 95th percentile)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
>5
Fasting plasma triglyceride concentration (mmol/L)

lipoproteins in the plasma (Hopkins et al., 2003). Other visible clinical consequences include eruptive, tuberous, and palmar crease xanthomas, lipemic plasma and lipemia retinalis, neurologic symptoms such as irritability, hepatosplenomegaly, and recurrent epigastric pain associated with pancreatitis (Yuan et al., 2007). Xanthomas are CE engorged foam cells that deposit under the skin, observed in patients with severely elevated plasma TG >10 mmol/L. Eruptive xanthomas are small red papules that present on surfaces such as the back and neck. Tuberous xanthomas are larger yellowish nodules that tend to deposit on extensor surfaces. Palmar crease xanthomas are yellowish deposits within or adjoining the palmar creases. Lipemic plasma and lipemia retinalis are caused by the accumulation of TG-rich chylomicrons that become visible in plasma following blood sampling or when performing fundoscopy. Hepatosplenomegaly is the enlargement of both liver and spleen resulting from lipid accumulation in these tissues.

Severe HTG is also sufficient to cause hypertriglyceridemic pancreatitis in ~10% of patients, although the precise mechanisms by which this occurs is poorly understood (Ewald et al., 2009). Pancreatitis refers to acute or chronic inflammation of the pancreas caused by abnormal secretion of digestive enzymes that damage pancreatic tissues causing nausea and epigastric pain. One proposed mechanism is that the abundance of free FA released from TG-rich lipoproteins is able to fully saturate plasma albumin, such that the detergent properties of unbound FA cause lipotoxicity that damages the pancreatic acinar cells causing ischemia and pancreatic injury. Alternatively, the increased viscosity of plasma rich in TG may cause ischemia and acidosis that damages the pancreas. Prevention of pancreatitis in HTG patients is accomplished by treatment

with medications including fibrates, niacin, or fish oil, in addition to absolute restriction of dietary fat, refined carbohydrates and alcohol.

Few effective pharmacologic interventions exist to reduce elevated plasma TG concentration. Available options include fibrates, niacin, and certain polyunsaturated fatty acids (Yuan et al., 2007). Fibrates are peroxisome proliferator-activated receptor alpha (PPARα) agonists. They activate a genetic program responsible for coordinated FA oxidation, which decreases VLDL secretion and increases TG-rich lipoprotein hydrolysis, effectively reducing plasma TG concentration by 50%. Fibrates are the most commonly used treatment option for patient with HTG. By comparison, niacin (nicotinic acid) can lower plasma TG to a comparable extent while simultaneously increasing HDL cholesterol, but with clinically significant side-effects including light-headedness, sweating, cutaneous flushing and pruritis. The mechanism through which niacin lowers plasma TG concentration is not well understood. Polyunsaturated FAs, particularly omega-3 FAs, are natural compounds that can lower TG concentrations by ~20%; however they are rarely effective as monotherapy. The mechanism whereby omega-3 FAs lower plasma TG levels is not fully understood. Other pharmacotherapies including statins are effective at lowering plasma concentrations of LDL cholesterol in HTG patients with elevated cardiovascular risk, however they are not primarily indicated for prevention of plasma TG concentration in HTG patients (Drexel, 2009). Combination therapy with fibrates/niacin and statins are indicated for patients with plasma TG <5 mmol/L for CAD prevention, whereas fibrate therapy becomes increasingly important for pancreatitis prevention when plasma TG exceeds 10 mmol/L.

Also, HTG often exists in the presence of other medical conditions such as poorly controlled diabetes, obesity, non-alcoholic steatohepatitis, certain autoimmune diseases, and chronic renal disease. Management of these conditions frequently helps to improve HTG. Further, certain medications such as corticosteroids, antihypertensive agents, and anti-rejection medications can raise TG in susceptible patients. Discontinuation or adjusting these treatments can also have beneficial effects on TG levels.

1.3. Genetic variation and human diseases

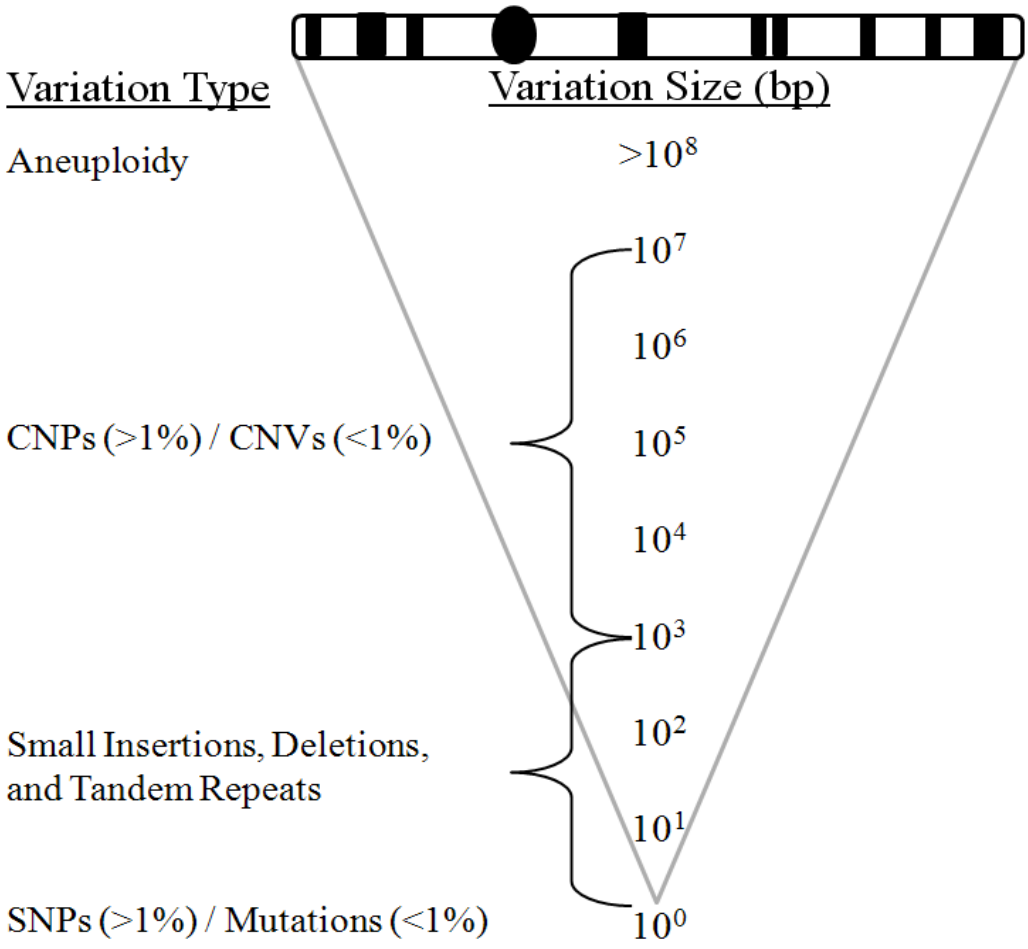
1.3.1 Genetic variation

The average human genome contains 3.1 billion base pairs of DNA, containing over 12 million common sequence variants and 11,000 structural variants documented across world populations (Conrad et al., 2009, Frazer et al., 2007, Lander et al., 2001). Differences in DNA sequence between individuals are broadly encompassed by the term "genetic variation". Human genetic variation can range from single base-pair differences to entire chromosome duplications or deletions **(Figure 1.7)**.

1.3.2. Single nucleotide variants

Single nucleotide variants (SNVs) are DNA base pair substitutions that occur throughout the genome (Conrad et al., 2009, Frazer et al., 2009). They are the most abundant source of inter-individual variation, having the potential to alter metabolism either by changing patterns of gene expression or protein function. Nomenclature for SNVs is based on their allele frequencies. Rare variants have frequencies <1%, uncommon variants have frequencies 1%-5%, and common variants have frequencies >5%. Moreover, SNVs with frequencies >1% are often referred to as single nucleotide polymorphism (SNPs). SNPs are generally not damaging to protein function, meaning they have small effect sizes, and thus typically persist in the population with varying allele frequencies. SNVs with frequencies <1% are often referred to as mutations, especially when proven to attenuate protein function. Mutations often have significant functional impact leading to very large

Figure 1.7. Selected types of genetic variation found throughout the human genome. Variant size is indicated in base-pairs (bp). Single-nucleotide polymorphisms (SNPs) and mutations refer to changes in a single DNA base at a particular position in the genome. Structural variation includes inversions, translocations, and CNPs and CNVs, ranging from <1 kilobase to >1 megabase in size. Typically, any DNA segment is found once per chromosome, giving a copy number of 2. Deletions result in a copy number <2, and insertions result in a copy number >2. Inversions and translocation do not change DNA dosage, although they change the structural arrangement of a genome.

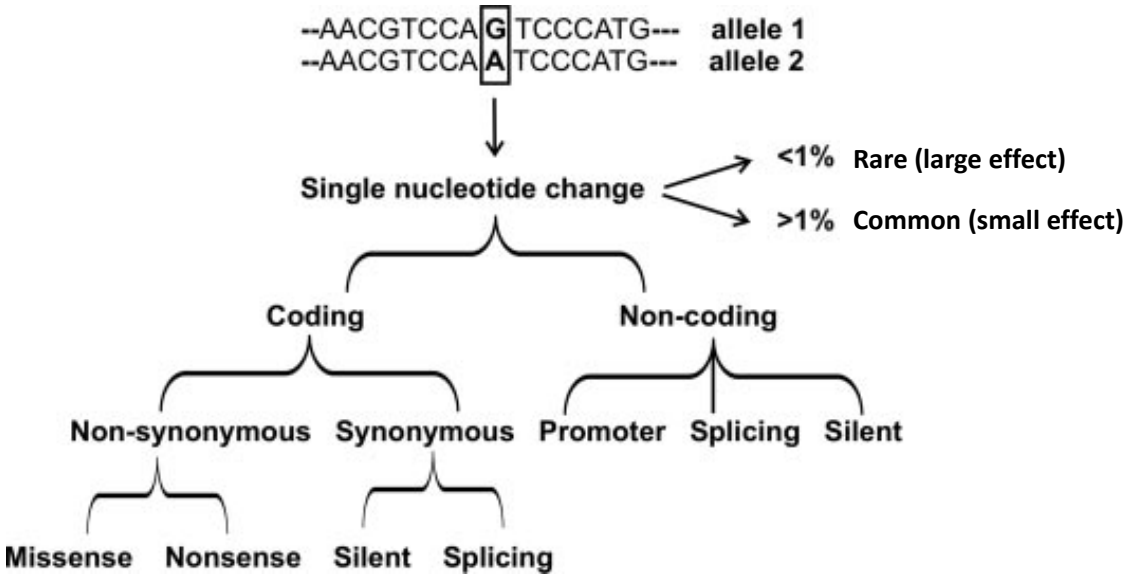

Variation Type
Variation Size (bp)
Aneuploidy
>10^8
10^7
10^6
CNPs (>1%) / CNVs (<1%)
10^5
10^4
10^3
10^2
Small Insertions, Deletions,
and Tandem Repeats
10^1
SNPs (>1%) / Mutations (<1%)
10^0

effect sizes that cause monogenic diseases. Due to evolutionary selection pressure, mutations are less likely to persist in the population at high frequency. Rare variants should only be referred to as mutations in situations where the SNV has a confirmed deleterious effect.

SNVs that are associated with an intermediate trait or disease phenotype are not necessarily sufficient to cause disease in their carriers. This property is referred to as penetrance; the likelihood that carriers of a disease-causing allele will express the disease phenotype. This concept differs from expressivity which refers to variations in phenotypic expression or severity of a disease phenotype in variant carriers. For example, the penetrance of mutations causing monogenic diseases is virtually 100%. Carriers of a completely penetrant disease-causing mutation will manifest the disease, regardless of environmental or metabolic variables. The penetrance of common variants in complex diseases is comparably much lower and carriers do not necessarily express the disease phenotype. Alleles with low penetrance are much more likely to be affected by additional gene-gene or gene-environment interactions to result in phenotype expression. The penetrance of disease-causing alleles is thus related to SNV frequency and effect size, which is dependent upon the variant properties and gene product in question.

Different types of SNVs are able to modulate intermediate phenotypes and disease susceptibility **(Figure 1.8)**. For instance, SNVs in the protein coding sequence can introduce nonsynonymous amino acid substitutions (missense variants) into a

Figure 1.8. Types of SNVs. SNVs refer to individual sequence variations that occur between subjects. SNVs with frequency >1% generally have modest effects, whereas SNVs with frequency <1% generally have large effects on intermediate traits or disease susceptibility. Different types of SNVs have the capacity to influence gene and protein function. For instance, SNVs that occur in protein coding sequence are called non-synonymous variants, which may introduce incorrect amino acids (missense variants) or premature stop codons (nonsense variants) into the protein sequence. Alternatively, they may not affect the amino acid sequence (synonymous variants) resulting in benign effects (silent variants) or affecting splicing or regulatory elements that occur in exons. SNVs that occur out of protein coding regions have the capacity to influence promoter activity, regulatory elements or intron-exon splicing patterns.


--AACGTCCAGTCCCATG--- allele 1
--AACGTCCAATCCCATG--- allele 2
Single nucleotide change
<1% Rare (large effect)
>1% Common (small effect)
Coding
Non-coding
Non-synonymous
Synonymous
Promoter
Splicing
Silent
Missense
Nonsense
Silent
Splicing

growing polypeptide that could potentially alter protein structure and function. These may have mild or severe deleterious effects, depending on the position of the mutation in the protein and the type of amino acid change. SNV can also introduce nonsense mutations into the coding sequence that can prematurely truncate the protein and likely alter its function. Similarly, small insertions or deletions that shift the reading frame of a protein can abnormally shorten or lengthen the product and likely alter its function. SNVs that border on intron-exon boundaries can interrupt the proper splicing pattern of a gene's transcript, resulting downstream in an incorrectly synthesized protein, via exon skipping or intron retention. Finally, SNVs in regulatory elements that control gene expression, including promoters, enhancers, or repressors, can disrupt the proper dosage and expression patterns of genes whose products act in pathways that affect a phenotype. Ultimately, many genetic variants are also silent within the genome, as they occur in regions that presumably have no functional consequence.

1.3.3. Functional evaluation of single nucleotide variants

Rapid thorough functional evaluation of genetic variation is not always possible following discovery of disease-associated variants. Thus, *in silico* prediction algorithms are useful tools to evaluate the functional consequence of amino acid substitutions on protein products. Two commonly employed tools are SIFT (Sorting intolerant from tolerant) and Polyphen-2 (Polymorphism Phenotyping v2). SIFT (Kumar et al., 2009) and Polyphen-2 (Adzhubei et al., 2010) both predict the impact of amino acid substitutions on protein function by evaluating protein sequence homology, assessing amino acid conservation at the mutated residue throughout evolution. Polyphen-2 additionally

considers the impact of altered amino acid properties and position effects of substitutions based on known functional domains and protein modeling. These algorithms produce very straightforward interpretable results, using a composite score to indicate whether the variants are predicted to be benign or deleterious. These algorithms are quite effective at identifying functional from benign variants; however variant predictions ultimately require follow-up using functional assays in vitro to determine the true effects of variants on protein function.

1.3.4. Structural variation

Structural variants are also prevalent in the genome and can also influence disease susceptibility (Conrad et al., 2009). Structural variants are much larger in scale than SNVs, typically ranging from 1-5000 kilobases, encompassing copy number variants (CNVs), inversions and translocations (Hurles et al., 2008, Pollex and Hegele, 2007, Redon et al., 2006). They are found with varying frequencies, and thus have nomenclature similar to SNVs. CNVs have frequencies <1%, and copy number polymorphisms (CNPs) have frequencies >1%. CNVs change the quantity or dose of a section of a chromosome. A human diploid genome has 2 copies of each chromosome, but CNVs deviate from a copy number of 2: deletions reduce the DNA copy number to 1 or 0, and duplications increase the copy number to 3 or even $\geq$4.

Since structural variants are large, they can contain entire genes or *cis*-acting elements that have the capacity to alter gene dosage in carriers, which can result in clinical phenotypes (Hurles et al., 2008). Inversions and translocations are changes in the

organization of DNA, without changing copy number. Inversions change the orientation of a DNA segment without changing its position on the chromosome. Translocations change the physical position of a DNA segment; it is relocated onto another chromosome. All structural variants are able to disrupt the integrity of genes, and sometimes alter their function or expression. CNVs can further affect the proper dosage of genes, as they can contain entire genes or regulatory elements. However, while structural variation undoubtedly contributes to disease phenotypes, it is more difficult to detect with high resolution and prove statistical association with disease phenotypes (Kathiresan et al., 2009a). As a result, few studies to date have investigated the impact of structural variation as a contributor to disease pathophysiology. Therefore, SNVs are the primary focus of this dissertation.

1.3.5. Monogenic diseases

Monogenic (or Mendelian) diseases are caused by highly penetrant mutations that are usually sufficient to cause disease. Such mutations attenuate gene function, disrupting normal metabolism so drastically that environmental influences have relatively negligible effects. Monogenic diseases are often inherited in an autosomal recessive manner, in which two loss-of-function mutations must be inherited to disrupt both alleles. However, monogenic diseases can also be inherited in an autosomal dominant or co-dominant manner, in which a single mutation causing haplo-insufficiency causes a disease phenotype, whereas two mutations cause a more severe phenotype. For instance, consider the genetic disease familial hypercholesterolemia in which plasma LDL concentration is $>95^{th}$ percentile (Yuan et al., 2006). Autosomal recessive hypercholesterolemia is caused

by two mutations inherited in *ARH* (encoding the LDLR adapter protein). Abrogation of both *ARH* alleles is required to produce the disease phenotype, whereas a single mutation is insufficient to cause disease. Conversely, the related phenotype autosomal dominant familial hypercholesterolemia (ADH) is caused primarily by heterozygous mutations in *LDLR*, whereas disruption of both *LDLR* alleles produces a more severe phenotype (homozygous ADH) in which LDL cholesterol concentrations are doubled compare to heterozygous ADH. Mutations in multiple different genes that participate in the same metabolic pathway often cause the same monogenic phenotype, as in familial hypercholesterolemia where mutations in *LDLR*, *APOB*, *PCSK9*, and *ARH* may all contribute to disease pathophysiology. Studies of patients and families with familial hypercholesterolemia have been instrumental in the identification of genes and gene products with important physiological roles in LDL metabolism.

1.3.6. Complex diseases

Complex (or polygenic) disease phenotypes are the product of both genetic and environmental components, in which multiple low penetrance genetic variants in hundreds of genes contribute to the expressed phenotype. In contrast to monogenic disorders, no single genetic variant is necessary or sufficient to result in manifestation of a complex phenotype. Three genetic hypotheses have been proposed to explain familial inheritance of complex diseases, including the common-disease common-variant (CDCV) model, the common-disease rare-variant (CDRV) model, and the heterogeneity model. The CDCV model predicts that a spectrum of common variants in multiple genes, each with a subtle effects but high population allele frequency, additively contribute to

the manifestation of complex diseases (Reich and Lander, 2001). Alternatively, the CDRV model predicts that fewer rare variants with large effects and higher penetrance are the underlying cause of disease (Pritchard, 2001). Importantly, neither model suggests that the inheritance of any specific risk variant is deterministic of disease; rather they incrementally contribute to increased disease susceptibility. The heterogeneity model predicts that a combination of both common and rare variants contribute to complex phenotypes (Schork et al., 2009). The heterogeneity model seems to be emerging as a more accurate model that will explain larger proportions of genetic variation of discrete and quantitative traits (Altshuler et al., 2008).

1.4 Genetic investigation of monogenic and polygenic diseases

1.4.1. Linkage analysis and autozygosity mapping

Linkage analysis traces the cosegregation of genetic markers with a disease phenotype through large multi-generational families to identify regions of the genome that are statistically associated with disease susceptibility (Morton, 1955). For instance, restriction fragment length polymorphisms (RFLPs) were among genetic markers once used to map diseases; however technological advances and better understanding of the human genome have evolved such that predominantly SNP genotyping is now used to increase both statistical power and resolution of linkage analyses. Linkage analyses are used primarily to identify the basis of Mendelian disorders in large multi-generational pedigrees, where a single highly penetrant large effect mutation is the cause of disease (Belmont and Leal, 2005). Genetic and allelic heterogeneity are a source of confounding in linkage analyses of Mendelian traits. Linkage analyses are less powerful when examining multifactorial diseases because they require intact family units and are confounded by small effect sizes, genetic and allelic heterogeneity, and multiple different gene-gene and gene-environment interactions (Pollex and Hegele, 2005).

Autozygosity mapping (also known as homozygosity mapping) is a type of linkage analysis used in autosomal recessive diseases in which consanguinity has been implicated in disease causation. It traces regions of the genome that are identical-by-descent, having arisen from a single ancestor, becoming homozygous as the result of consanguinity. Homozygosity is identified across regions of the genome that are shared

by affected family members (Lander and Botstein, 1987). This is a powerful approach that functions well using small pedigrees with few affected family members, identifying strong associations with potentially causal loci. However, the limitations of general linkage analyses also pertain to autozygosity mapping.

1.4.2. Candidate gene association studies

Candidate gene association studies are based on *a priori* knowledge of genes involved in pathophysiology of complex diseases. Genetic variation is identified in a candidate gene, typically by resequencing in a subset of subjects, followed by genotyped in a large sample. Various different statistical analyses are employed to determine whether a genetic variant is associated with an intermediate trait or disease phenotype. Candidate gene studies were once used frequently to implicate candidate genes in disease processes, however they are notorious for failure to replicate in subsequent studies, likely due to confounding effects of population stratification, ethnicity, inadequate phenotyping, inadequate sample size, or inappropriate case-control matching (Hegele, 2002). Replication of study findings in independent cohorts is the gold standard for true positive association. Therefore, these challenges have limited the effectiveness of candidate gene studies over the years.

1.4.3. Genome-wide association studies

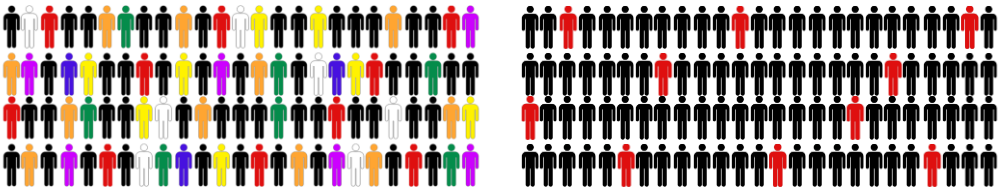
Unlike candidate gene studies, genome-wide association studies (GWAS) are a recently established unbiased technique capable of powerfully detecting reproducible associations between common genetic variants and complex quantitative or discrete phenotypes (Attia

et al., 2009a, b, c). Briefly, GWAS genotype millions of independent SNPs covering the entire genome in large populations and tests them each independently for association with a target phenotype **(Figure 1.9)**. Either quantitative intermediate biochemical phenotypes such as plasma TG concentration or discrete clinical phenotypes such as CAD can be used in GWAS. The result is an estimate of effect and strength of association measured by a P-value for each SNP (adjusted for covariates and multiple testing). Statistically robust associations represent regions of the genome containing genes or regulatory elements in which common variation that modulate the target phenotype. The GWAS framework is an intricate part of analyses conducted for this dissertation. Accordingly, the following sections will review concepts and analyses integral to GWAS.

1.4.3.1. Obtaining genotypes

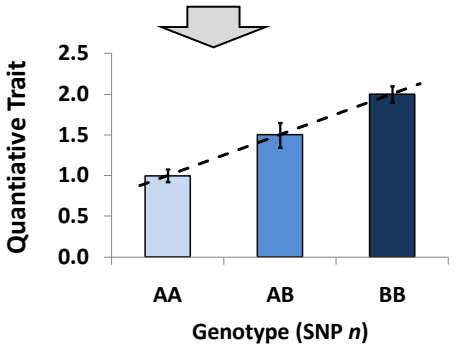
GWAS are enabled by inexpensive high-density microarray platforms that simultaneously genotype millions of known variants across the genome. For instance, the Affymetrix Genome-Wide Human SNP Array 6.0 is capable of simultaneously genotyping ~1.8 million markers representing ~900,000 SNPs and ~900,000 copy number probes, at positions strategically chosen across the genome. A variety of molecular techniques are used to segment, amplify and label genomic DNA for hybridization to the arrays. Affymetrix-based microarrays fluorescently-label sample DNA such that hybridization to a complementary probe, whose sequence and position on the array is known, allows genotype calling. For SNPs, probes are designed to target each allele of a known variant, such that fluorescence from either probe alone represents a

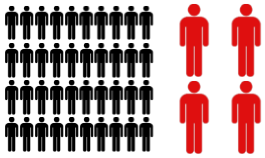
Figure 1.9. A path for discovery and translation of GWAS findings.



1. Collect quantitative trait data (left) or dichotomize cases and controls (right)

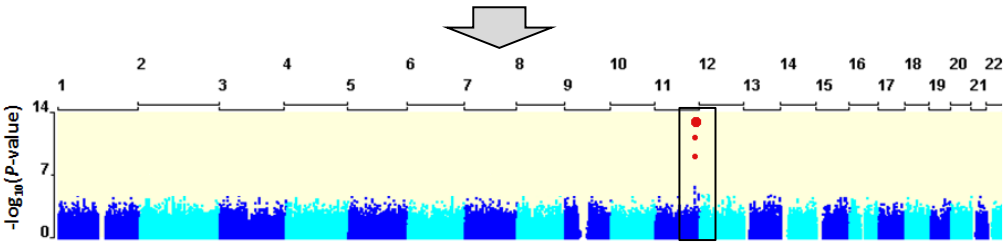
2. Genotype discovery cohort using high-density microarrays





Allele	Controls	Cases
A	80%	65%
B	20%	35%

3. Test for association between each SNP and phenotype of interest



4. Rank P-values to identify most strongly associated loci ($P<5X10^{-8}$)

5. Replicate associated loci in independent cohort (from step 1 using P<0.05)

---

6. Replicate associations in additional cohorts/ethnicities/phenotypes

7. Fine-mapping to identify specific associated gene/genomic element

8. Re-sequence to identify common and rare functional variants

9. Functional validation, gene-gene and gene-environment interaction studies

10.Diagnostic, pharmacogenetic and therapeutic applications

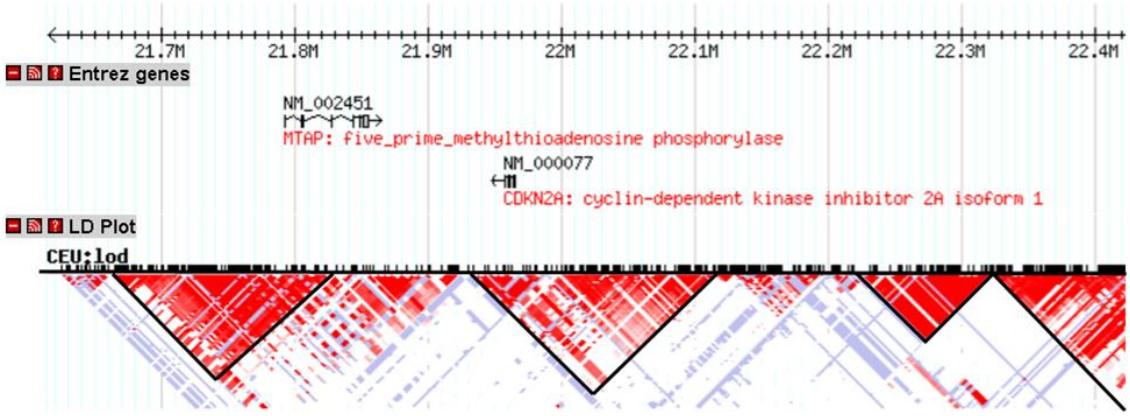
homozygote for either allele, whereas fluorescence from both probes represents a heterozygote for the target allele. For CNVs, signal intensities of each probe are compared to the average intensity across the genome to identify regions that have relatively increased or decreased signal, which correspond to changes in copy number. Advances in genome-wide genotyping now allow increased density, increased samples per chip, and custom design of targeted arrays to facilitate different experimental designs. However, these technologies are likely to become obsolete once next-generation sequencing becomes affordable and commonplace.

1.4.3.2. Linkage disequilibrium and the HapMap project

SNPs are chosen for inclusion on microarrays based on their position in haplotype blocks across each chromosome. Haplotype blocks are groups of SNPs that are inherited together on a chromosome **(Figure 1.10)**. They have arisen historically in regions of the genome where homologous recombination has not occurred. As a result, SNPs within haplotype blocks have a non-random association among them known as linkage disequilibrium (LD), a feature that can be quantified and evaluated statistically using the metric r^2, which ranges from 0 (complete linkage equilibrium) to 1 (complete LD). These properties of common variants allow sophisticated statistical analysis to estimate un-genotyped SNPs with a high degree of accuracy based on panels of genotyped and patterns of LD in reference cohorts, a process known as imputation (Li et al., 2009).

Patterns of human genetic variation have been thoroughly catalogued by the International HapMap Consortium, a large scale organization whose goal is to produce a

Figure 1.10. Illustration of LD and haplotype block structure. The axis below the "LD Plot" heading indicates where SNPs were genotyped by the International HapMap Consortium; each vertical black line indicates a SNP. The colour at the intersection points of two SNPs indicates the LD, or non-random association, between them. Red squares indicate high linkage, pink squares indicate moderate linkage, blue squares are uninformative markers, and white squares indicate no linkage. Large red blocks indicate extended regions of LD between multiple SNPs that form haplotype blocks which are likely inherited together on a chromosome. Single SNPs may be genotyped in haplotype blocks to effectively infer the genotype of surrounding SNPs. A limitation of this approach is that an association with a SNP in a haplotype block does not identify the precise functional variant mediating the association; it simply indicates the presence of a functional variant somewhere within that block. This figure is a screenshot of the 9p21 locus modified from the International HapMap Consortium website.


21.7M
21.8M
21.9M
22M
22.1M
22.2M
22.3M
22.4M
Entrez genes
NM_002451
MTAP: five_prime_methylthioadenosine phosphorylase
NM_000077
CDKN2A: cyclin-dependent kinase inhibitor 2A isoform 1
LD Plot
CEU: lod

high resolution human haplotype map outlining the location of haplotype blocks and the LD within (International HapMap Consortium, 2003). The HapMap project has been invaluable to GWAS, allowing few "tagging" SNPs to be selected from within the haplotype block to serve as a surrogate for many more, such that modern microarrays can achieve maximally 90% coverage of the genome for testing association with disease traits (Altshuler et al., 2008). Imputation further increases power in large scale GWAS to detect associations between genotype and phenotypes. While effective at maximizing the likelihood of identifying significant associations, a limitation of this strategy is that statistically associated SNPs are rarely the specific variants that mediate a phenotypic effect; they simply indicate whether a haplotype block presumably containing a functional variant is associated with the phenotype. More recently, the 1000 Genomes Project was launched to more comprehensively characterize the majority of genetic variation in the studied population (Durbin et al., 2010). Next generation sequencing technologies were used to identify all variants with frequency >1% in European, East Asian and African populations. The project consisted of 3 phases, including low coverage (2-4X) sequencing of 180 subjects, moderate depth (30X) sequencing in 2 trios, and high depth (>50X) sequencing of 1000 genes in 900 samples. Data obtained by the 1000 Genomes Project has been made available in the form of a genome browser and public databases, which have facilitated genetic analyses of both monogenic and complex diseases.

1.4.3.3. Statistics in GWAS

Once typed and imputed genotypes are obtained, regression is commonly used to test for association between SNPs and quantitative or discrete phenotypes. Linear and logistic regression are often used to measure the effect of multiple independent variables, including SNP genotypes and other covariates, on the intermediate traits or discrete disease phenotypes, respectively. Generally, these analyses assume an additive model of risk alleles, meaning that each risk allele contributes additively to the phenotype. These analyses produce an effect size for each variant, measured by frequency ratios for discrete phenotypes, and a P-value representing the strength of association with the phenotype.

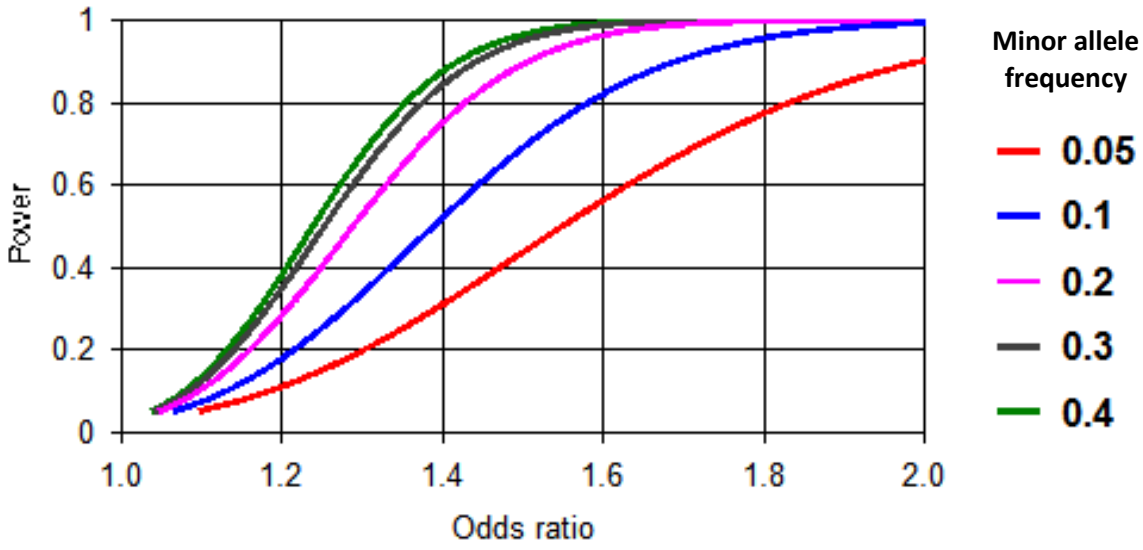
The statistical analysis used to detect associations between genotype and phenotype in GWAS is relatively straightforward. However, the simultaneous analysis of millions of variants introduces a problem of multiple testing. Typically, statistical tests accept a false-positive rate of 5% (α=0.05), however this would correspond to an overwhelming 50,000 false positives occurring by chance in any particular GWAS that studied 1 million SNPs. A conservative approach to correct for multiple testing is the Bonferroni correction, which divides the accepted false positive rate by the number of independent tests. SNPs that surpass the accepted Bonferroni-corrected genome-wide significance threshold of $P=5X10^{-8}$ are considered robustly associated with the target phenotype.

GWAS also require implementation of other rigorous statistical analyses to prevent potentially confounding variables from identifying false positive associations.

One such confounding variable is population stratification. Population stratification refers to differences in allele frequencies that occur between different groups of subjects (Kraft, 2011). These systematic differences in allele frequencies could be misconstrued as false positive associations if the sample distribution (e.g. cases and controls) is composed of unequal proportions of subjects from different geographical regions or ancestries. Two commonly employed statistical analyses are used to control population stratification, which include principal component analysis (Patterson et al., 2006, Price et al., 2006) and genomic control (Devlin and Roeder, 1999).

An important concept in GWAS is statistical power. Power refers to the probability that a statistical test will reject the null hypothesis - that no association exists between genotype and phenotype – when no true association exists. In GWAS, power calculations are used to determine the sample size required to detect a given effect size; alternatively, they are used to determine the effect size that can be detected in a fixed sample size. Several variables affect statistical power in association studies, including sample size, effect size, allele frequency and threshold for false positive association. For example, a sample of 500 case patients and 1200 controls would have 100% power to detect associations with an effect size (measured by odds ratio [OR]) of 1.6 given an allele frequency of 30% **(Figure 1.11)**. Decreasing allele frequencies reduce power to detect the same odds ratio, such that 100% power is achieved for an OR=1.8 at 20% allele frequency, OR=2.0 at 10% allele frequency, and OR=2.5 at 5% allele frequency. The experiments conducted for this dissertation have ~70% power to detect an OR=1.5 at an allele frequency of 10%, given a false positive rate of 5%.

Figure 1.11. Statistical power in genetic association studies. The plot above demonstrates the statistical power of a Fisher's exact test to detect an association between genotype and phenotype at a false-positive rate of 5% (α=0.05), given various allele frequencies in a population of 500 case patients and 1200 healthy controls.


Power
1
0.8
0.6
0.4
0.2
0
1.0
1.2
1.4
1.6
1.8
2.0
Odds ratio
Minor allele frequency
0.05
0.1
0.2
0.3
0.4

1.4.4. Resequencing studies

Identification of rare variation is becoming increasingly important to explain the unattributed variation in complex phenotypes. Increasingly affordable high-throughput massively paralleled sequencing technologies permit identification of all genetic variants in a subject's exome or whole genome (Metzker, 2010), however the recent introduction of these platforms has not allowed sufficient time for the development of robust analytical strategies (Bansal et al., 2010). Low frequency variants cannot be tested individually for association with a phenotype, given extremely limited power to detect a true association, especially given the need for multiple testing corrections. Hypothesis-driven multi-locus grouping of rare variants is emerging as a standard approach to detect associations between groups of genes and a phenotype of interest, whereas gene-centric analyses facilitated by novel statistical analyses are quickly emerging to enhance the resolution of rare variant association studies (Neale et al., 2011, Price et al., 2010). However, rare variant analyses are a relatively novel avenue of investigation in genetic association studies resulting in few established analysis strategies or experimental designs.

1.5. Genetic architecture of plasma triglycerides and hypertriglyceridemia

1.5.1 Common genetic determinants of plasma triglycerides

Our understanding of the genetic architecture - that is the integrated contribution of all genetic variants that contribute to inter-individual variation – of plasma TG concentrations has increased substantially following the reports of GWAS in the literature. Fasting plasma TG is a stable biochemical analyte allowing GWAS to combine findings across multiple epidemiologic studies to identify associated genetic loci. Prior to the outset of the projects described in this dissertation, GWAS had identified association signals at 11 TG-associated loci **(Table 1.1)**, containing both classically established genes and previously unknown genomic regions, as determinants of plasma TG concentration (Aulchenko et al., 2009, Kathiresan et al., 2008, Kathiresan et al., 2009b, Kooner et al., 2008, Sabatti et al., 2009, Wallace et al., 2008, Willer et al., 2008). However, the genes and functional variants at remaining newly-identified loci are mostly unknown. GWAS is not able to identify the specific genes or variants that mediate each association. Thus, the majority of genes associated with plasma TG by GWAS represent the most likely biological candidates for the association based on *a priori* functional evidence. These associations merely represent a starting point for subsequent functional characterization necessary to identify novel genes and pathways involved in modulation of TG-rich lipoprotein metabolism. Such discoveries could take many years to be validated, even with a novel straightforward mechanism without inherent biological complexities.

Table 1.1. Genetic variation associated with plasma TG concentration or HTG by GWAS or resequencing studies in January 2009.

	Common variation - GWAS		Rare variation - Sequencing	
	Normal TG	Polygenic HTG	Polygenic HTG	Monogenic HTG
APOA5	X	X	X	X
LPL	X	X	X	X
APOC2			X	X
GPIHBP1				X
LMF1				X
GCKR	X	X		
TRIB1	X	X		
MLXIPL	X	X		
APOB	X			
NCAN	X			
ANGPTL3	X	X		
FADS1	X			
PLTP	X			
XKR6	X			

1.5.2. Rare genetic determinants of monogenic hypertriglyceridemia

The genetic determinants of Mendelian HTG syndromes are fairly well defined. Characterization of these gene products has contributed significantly to our current understanding of TG metabolism. Monogenic HTG has a population frequency of $<1:10^6$ and follows autosomal recessive inheritance resulting from homozygous (or compound heterozygous) large-effect disease-causing mutations in genes that modulate plasma TG concentration. Most monogenic HTG patients have mutations in LPL leading to complete LPL deficiency (OMIM 238600). LPL hydrolyses TG molecules transported in TG-rich lipoproteins to liberate free FAs for TG re-synthesis and storage in adipose tissue or beta-oxidation in skeletal muscle and heart (Kirchgessner et al., 1989). LPL is matured in the ER by LMF1 and transported to the capillary lumen by GPIHBP1. Thus, mutations in the N-terminal domain that attenuate LPL catalytic activity (Rahalkar et al., 2009), in addition to mutations in the C-terminal domain that prevent binding to GPIHBP1 (Voss et al., 2011), are sufficient to cause monogenic HTG. In total, >114 homozygous or compound heterozygous mutations have been shown to cause LPL deficiency in patients, including 72 missense mutations, 14 nonsense mutations, 13 small scale insertions and deletions, 9 splicing variants, 3 regulatory variants, and 3 large scale structural rearrangements (**Figure 1.12a**).

Interestingly, remaining genes that cause monogenic HTG are all essential for efficient activity, assembly or transport of LPL, which effectively block the lipolytic cascade (**Figure 1.12b**). For instance, *APOC2* (encoding apo C-II) and *APOA5* (encoding

Figure 1.12a. Homozygous and compound heterozygous mutations in *LPL* that cause Mendelian HTG. Mutations shown in red are nonsense mutations or frameshift mutations causing premature truncation of the protein product.

Lipoprotein lipase (*LPL*)

Nonsynonymous:

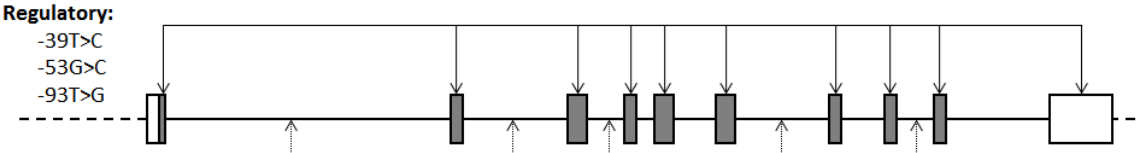
M1L	R102S	G181S	R197L	G222E	E269K	S286R	L313P	W409X
W14X	W113R	G181V	S199C	D231E	R270C	S286G	Y315X	W421X
W41X	W113G	D183N	D201V	I232S	R270H	A288T	N318S	E437V
C54X	A125T	D183G	A203T	P234L	R270L	Y289H	M328R	E437K
N70S	T128A	D183H	D207E	P241S	S271T	C291X	M328T	C445Y
S72G	G132R	P184R	H210Q	C243S	I276T	C291Y	Y329X	E448K
Y88X	Q133X	A185T	G215R	I252T	D277N	S293P	L330P	
W91X	H163R	G186E	G215E	Q262X	S278C	F297L	A361T	
V96L	G166S	G188E	S220R	C266X	S279R	L303F	T379I	
Y100X	G169E	E190G	I221T	C266W	L279V	C305R	L392V	

Frameshifts:

T45Hfs2X	V96Gfs50X	N147Tfs23X	L380Afs1X	S423Gfs51X
G62Rfs27X	A97Afs49X	G236Gfs15X	G314Gfs1X	
T83Xfs0X	T128Tfs45X	A248Lfs3X	K339Nfs14X	

Regulatory:

-39T>C
-53G>C
-93T>G



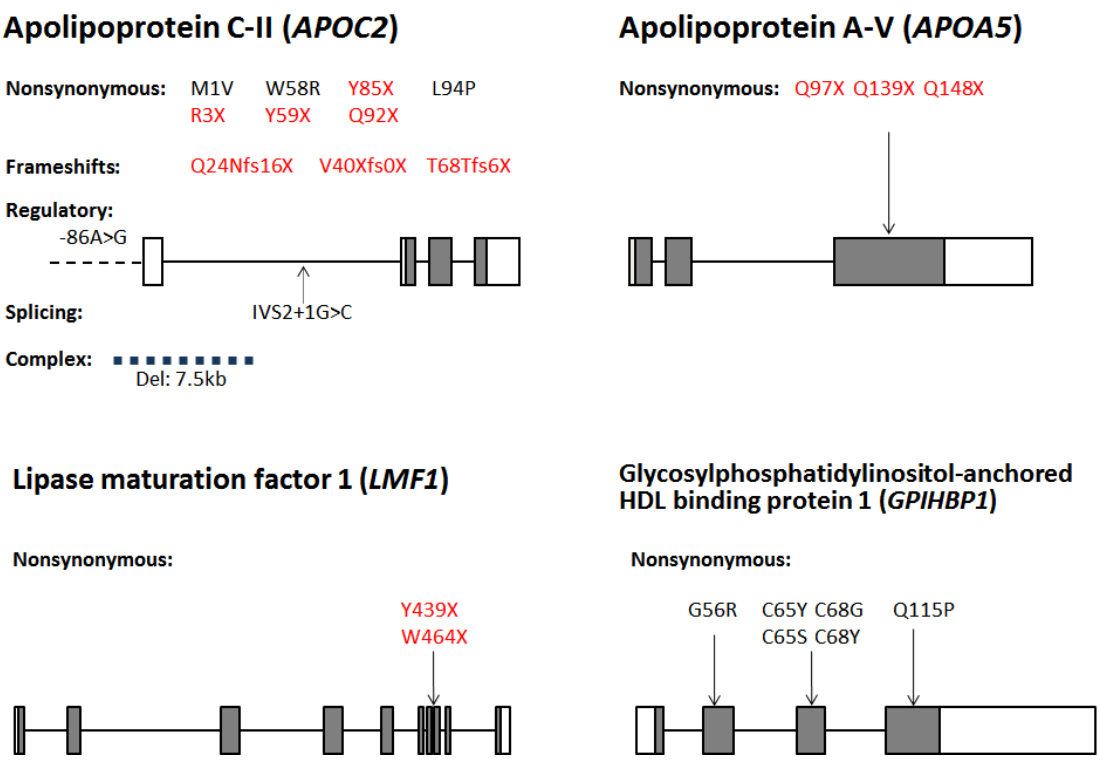
Splicing:

IVS1+1G>C
IVS1+2insT
IVS2-1G>A
IVS2+1G>A
IVS3-6C>T
IVS6-3C>A
IVS6-3C>T
IVS8-90T>G
IVS8+2T>C

Complex:

Del: 2.3kb
Del: 2.1kb
Del: exon 8-10

Figure 1.12b. Homozygous and compound heterozygous mutations in *APOC2, APOA5, LMF1,* and *GPIHBP1* that cause Mendelian HTG. Mutations shown in red are nonsense mutations or frameshift mutations causing premature truncation of the protein product.


Apolipoprotein C-II (APOC2)
Nonsynonymous: M1V W58R Y85X L94P
R3X Y59X Q92X
Frameshifts: Q24Nfs16X V40Xfs0X T68Tfs6X
Regulatory:
-86A>G
Splicing:
IVS2+1G>C
Complex:
Del: 7.5kb
Apolipoprotein A-V (APOA5)
Nonsynonymous: Q97X Q139X Q148X
Lipase maturation factor 1 (LMF1)
Nonsynonymous:
Y439X
W464X
Glycosylphosphatidylinositol-anchored HDL binding protein 1 (GPIHBP1)
Nonsynonymous:
G56R
C65Y C68G
C65S C68Y
Q115P

apo A-V) also cause monogenic HTG. Apo C-II is an essential LPL coactivator absolutely required for hydrolysis TG-rich lipoprotein hydrolysis (LaRosa et al., 1970), thus homozygous LOF mutations in *APOC2* cause apo C-II deficiency (OMIM 207750) and monogenic HTG (Connelly et al., 1987). *APOC2* mutations were first definitively shown to cause HTG through characterization at the amino acid level (Breckenridge et al., 1978). Since then, approximately 13 mutations have been identified that cause apo C-II deficiency, including 4 nonsense mutations, 3 missense mutations, 1 splicing mutation, 1 regulatory regions mutation, 1 gross deletion, 3 indels each causing a translational frameshift and predicted premature truncation of the protein product (**Figure 1.12b**). These variants result in complete absence of apo C-II from plasma, whereas missense mutations usually produced protein defective in LPL binding. Apo C-II variants cause significant increases in plasma TG concentration >10 mmol/L, accompanied by epigastric pain and pancreatitis, resulting from defective LPL activity.

Apo A-V is also required for efficient lipolysis of TG-rich particles by LPL (Charlton-Menys and Durrington, 2005), however the precise mechanism by which it confers this activity is unknown. *Apoa5* overexpression in mice has been demonstrated to decrease plasma TG concentrations, whereas *Apoa5* knockout results in significantly increased plasma TG concentration in mice. Thus, it is logical to assume that *APOA5* deficiency (OMIM 144650) in humans would similarly cause HTG. However, surprisingly only 3 mutations have been associated with apo A-V deficiency, each a distinct glutamine to nonsense mutation (Marcais et al., 2005, Priore Oliva et al., 2008, Priore Oliva et al., 2005) (**Figure 1.12b**), although many heterozygous mutations have

been found in polygenic HTG patients. One monogenic patient with abdominal pain, hepatosplenomegaly and plasma TG concentration >40 mmol/L was a carrier of a glutamine to stop mutation at position 148 (Q148X) (Priore Oliva et al., 2005), which prematurely truncated the protein such that it lacked critical hydrophobic lipid binding domains and putative heparin-binding domains, although apo A-V was completely devoid from plasma. This truncation mutation cosegregated with increased plasma TG concentration in the proband's family but without causing HTG. Another glutamine to stop mutation at position 139 (Q139X) was also identified in a HTG subject with recurrent pancreatitis, which was also identified as a heterozygous mutation in additional families (Marcais et al., 2005). Finally, a glutamine to stop mutation at position 97 (Q97X) was identified in a male proband (Priore Oliva et al., 2008). This mutation resulted in complete absence of apo A-V from plasma, but interestingly his plasma TG was controlled greatly by a low-fat low-oligosaccharide diet (Priore Oliva et al., 2008). These variants were associated with increased plasma TG concentration but were not sufficient to cause HTG without the contribution of additional variants that attenuated TG metabolism in *APOA5*, *APOC3*, or *APOE*, or perhaps due to the contribution of environmental exposures not necessarily described by these studies (Calandra et al., 2006). A more thorough understanding of the clinical consequences of apo A-V deficiency may eventually explain the incomplete penetrance of mutations in *APOA5*.

Genes that are required for efficient assembly and transport of LPL, including *LMF1* and *GPIHBP1*, were also recently shown to cause monogenic HTG. Only 2 nonsense mutations have been identified in *LMF1* (**Figure 1.12b**). One tyrosine to

nonsense (Y439X) mutation was identified in a HTG patient with recurrent pancreatitis, which resulted in a 93% reduction in post heparin plasma activity of LPL and plasma TG concentration of ~30 mmol/L (Peterfy et al., 2007). A tryptophan to nonsense (W464X) mutation identified in another patient also with recurrent pancreatitis was attributed to LPL activities of 76% causing a plasma TG concentration ~27 mmol/L (Cefalu et al., 2009). Similarly, only 6 mutations have been identified in *GPIHBP1* that cause monogenic HTG (**Figure 1.12b**). Most of these LOF mutations have been identified in the cysteine-rich Ly6 domain, which is required for proper GPIHBP1 folding and interaction with LPL. These variants included compound heterozygotes for a cysteine to serine mutation at amino acid residue 65 (C65S) and a cysteine to glycine mutation at amino acid residue 68 (C68G) (Coca-Prieto et al., 2011, Franssen et al., 2010b, Olivecrona et al., 2010). Another patient was homozygous for a cysteine to tyrosine mutation also at residue 65 (C65Y) (Franssen et al., 2010a). These variants all result in plasma TG concentrations between 18 and 27 mmol/L. A homozygous glutamine to proline substitution at amino acid residue 115 (Q115P) was also identified in the Ly6 domain, causing a plasma TG concentration ~38 mmol/L (Beigneux et al., 2009). Finally, a homozygous glycine to arginine mutation at residue 56 (G56R) was identified in two siblings with pancreatitis and plasma TG concentration >40 mmol/L (Wang and Hegele, 2007). In vitro analyses did not show a specific defect affecting GPIHBP1 function (Gin et al., 2007), however the allele carrying this mutation cosegregated well with the HTG phenotype in affected family members.

1.5.3. Common genetic determinants of polygenic hypertriglyceridemia

In contrast to monogenic HTG, the etiology of polygenic HTG was not well-defined. It was presumed to result from the deleterious effects of multiple low penetrance genetic variants combined with environmental and metabolic stressors (Ballantyne et al., 2000, Yuan et al., 2007). Prior to this dissertation, few replication studies had attempted replication of loci associated with population-based plasma TG concentrations **(Table 1.1)** (Wang et al., 2008a, Wang et al., 2008b). However, these studies were conducted in relatively small sample sizes with poor power to detect biologically relevant associations. It was well established that common variants such as *LPL* D9N were significantly over-represented in patients with HTG, whereas *LPL* S447X was less prevalent in HTG patients (Wang et al., 2008b). Similarly, both *APOA5* S19W and -1131T>C variants were strong predictors of polygenic HTG, such that carriers of *APOA5* S19W were >6 times more likely to be diagnosed with severe HTG than controls (Wang et al., 2008a).

Rare variants were also weakly associated with polygenic HTG. Preliminary analyses that sequenced *LPL*, *APOA5*, and *APOC2* revealed 9 heterozygous rare variants in 110 HTG patients, suggesting that large effect mutations may also contribute to increased disease susceptibility. However, a comprehensive and systematic analysis of the genetic determinants and genetic architecture of HTG had never been conducted for patients with polygenic HTG, mainly because severe disorders of TG metabolism are relatively rare in the general population, making it difficult to accrue samples with sufficient statistical power to elucidate the mechanisms of disease.

1.5.4. Phenotypic heterogeneity within polygenic hypertriglyceridemia

Phenotypic heterogeneity among polygenic HTG patients is the result of qualitative and quantitative biochemical differences in plasma lipoproteins that form the basis of the WHO or Fredrickson hyperlipoproteinemia (HLP) phenotypes **(Table 1.2)**. Five of six Frederickson phenotypes are defined by HTG; the exception is familial hypercholesterolemia (HLP type 2A) which most often results from mutations in *LDLR* encoding the LDL receptor (Rader et al., 2003). Among HLP phenotypes defined by HTG, there is one monogenic phenotype and four polygenic phenotypes. Monogenic HTG (HLP type 1) is defined by fasting chylomicronemia as described previously. The remaining polygenic familial phenotypes are adult phenotypes called combined hyperlipidemia (HLP type 2B), dysbetalipoproteinemia (HLP type 3), primary hypertriglyceridemia (HLP type 4) and mixed hyperlipidemia (HLP type 5), each differ in their plasma lipoprotein profile. Only one replication study assessing the association of GWAS-identified TG-associated loci in the Fredrickson phenotypes identified *APOA5* as strongly associated with all phenotypes, whereas varying patterns of association were observed between remaining TG-associated loci and each phenotype (Hegele et al., 2009). The genetic basis for the biochemically-defined Fredrickson phenotypes has never been characterized.

Table 1.2. Summary of the classical Fredrickson HLP phenotypes.

WHO ICD number	Frederickson phenotype	MIM number	Lipids	Lipoproteins	Genetics
E78.3	HLP type 1 Familial chylomicronemia	238600	↑TG	↑CM	Primarily pediatric; young adults; Monogenic; AR due to mutant *LPL* or *APOC2*; Phenocopies seen with homozygous mutations in *APOA5*, *LMF1* or *GPIHBP1*
E78.0	HLP type 2A Familial hypercholesterolemia	143890	↑TC	↑LDL	Monogenic, Heterozygous form due to mutant *LDLR*, *APOB* or *PCSK9;* Homozygous form due to mutant *LDLR* or *ARH*
E78.4	HLP type 2B Combined hyperlipoproteinemia	144250	↑TC, ↑TG	↑VLDL, ↑LDL	Polygenic - combined SNPs and excess of rare variants in HTG associated genes from GWAS together with LDL-associated SNPs
E78.2	HLP type 3 Dysbetalipoproteinemia	107741	↑TC, ↑TG	↑IDL	Polygenic - combined SNPs and excess of rare variants in HTG associated genes from GWAS together *APOE* E2/E2 homozygosity or mutation.
E78.1	HLP type 4 Primary hypertriglyceridemia	144600; 145750	↑TG	↑VLDL	Polygenic - combined SNPs and excess of rare variants in HTG associated genes from GWAS
E78.3	HLP type 5 Mixed hyperlipidemia	144650	↑TC, ↑TG	↑VLDL, ↑CM	Polygenic - combined SNPs and excess of rare variants in HTG associated genes from GWAS

AR, autosomal recessive; CM, chylomicrons; GWAS, genome-wide association study; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; ICD, International Classification of Diseases; ; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; MIM, Mendelian Inheritance in Man; TC, total cholesterol; TG, triglyceride; VLDL, very-low density lipoprotein; WHO, World Health Organization.

1.6. Summary

Plasma TG concentration is re-emerging as a risk factor for CVD. HTG patients with plasma TG concentration >95th percentile have significantly elevated cardiovascular risk, compounded by accompanying co-morbidities such as obesity, metabolic syndrome and T2D. However, the genetic architecture underlying HTG susceptibility is incompletely understood. Recent technological advances in genetic analyses now permit high-throughput evaluation of common and rare genetic variation contributing to complex phenotypes. The overall hypothesis of experiments conducted in this dissertation is that modern genomic analyses strategies will be capable of elucidating the genetic determinants of HTG and the phenotypic heterogeneity therein. We sought to conduct a systematic and comprehensive analysis of polygenic HTG patients according to the following objectives: (1) conduct genome-wide genotyping and association to identify novel genes containing small effect common variants and genes associated with HTG; (2) conduct high-throughput resequencing of HTG-associated genes to identify containing large effect rare variants associated with HTG; and (3) conduct subgroup analyses within Fredrickson HLP phenotypes to evaluate the contribution of common and rare variants to the phenotypic heterogeneity among HTG patients. Together, these studies were able to identify specific variants associated with HTG and provide a model for the genetic architecture of HTG susceptibility and phenotypic heterogeneity. A long-term index of the success of this thesis will be its contribution to the discovery of biological pathways containing potential pharmacological targets capable of lowering plasma TG concentration and improving cardiovascular risk in HTG patients.

1.7. References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods *7*, 248-249.

Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic mapping in human disease. Science *322*, 881-888.

Arnett, D.K., Baird, A.E., Barkley, R.A., Basson, C.T., Boerwinkle, E., Ganesh, S.K., Herrington, D.M., Hong, Y., Jaquish, C., McDermott, D.A., *et al.* (2007). Relevance of genetics and genomics for prevention and treatment of cardiovascular disease: a scientific statement from the American Heart Association Council on Epidemiology and Prevention, the Stroke Council, and the Functional Genomics and Translational Biology Interdisciplinary Working Group. Circulation *115*, 2878-2901.

Assmann, G., Cullen, P., and Schulte, H. (2002). Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. Circulation *105*, 310-315.

Attia, J., Ioannidis, J.P., Thakkinstian, A., McEvoy, M., Scott, R.J., Minelli, C., Thompson, J., Infante-Rivard, C., and Guyatt, G. (2009a). How to use an article about genetic association: A: Background concepts. JAMA *301*, 74-81.

Attia, J., Ioannidis, J.P., Thakkinstian, A., McEvoy, M., Scott, R.J., Minelli, C., Thompson, J., Infante-Rivard, C., and Guyatt, G. (2009b). How to use an article about genetic association: B: Are the results of the study valid? JAMA *301*, 191-197.

Attia, J., Ioannidis, J.P., Thakkinstian, A., McEvoy, M., Scott, R.J., Minelli, C., Thompson, J., Infante-Rivard, C., and Guyatt, G. (2009c). How to use an article about genetic association: C: What are the results and will they help me in caring for my patients? JAMA *301*, 304-308.

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., *et al.* (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet *41*, 47-55.

Baigent, C., Blackwell, L., Emberson, J., Holland, L.E., Reith, C., Bhala, N., Peto, R., Barnes, E.H., Keech, A., Simes, J., *et al.* (2010). Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. Lancet *376*, 1670-1681.

Baigent, C., Keech, A., Kearney, P.M., Blackwell, L., Buck, G., Pollicino, C., Kirby, A., Sourjina, T., Peto, R., Collins, R., *et al.* (2005). Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet *366*, 1267-1278.

Ballantyne, C.M., Grundy, S.M., Oberman, A., Kreisberg, R.A., Havel, R.J., Frost, P.H., and Haffner, S.M. (2000). Hyperlipidemia: diagnostic and therapeutic perspectives. J Clin Endocrinol Metab *85*, 2089-2112.

Bansal, S., Buring, J.E., Rifai, N., Mora, S., Sacks, F.M., and Ridker, P.M. (2007). Fasting compared with nonfasting triglycerides and risk of cardiovascular events in women. JAMA *298*, 309-316.

Bansal, V., Libiger, O., Torkamani, A., and Schork, N.J. (2010). Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet *11*, 773-785.

Barrows, B.R., and Parks, E.J. (2006). Contributions of different fatty acid sources to very low-density lipoprotein-triacylglycerol in the fasted and fed states. J Clin Endocrinol Metab *91*, 1446-1452.

Beigneux, A.P., Davies, B.S., Gin, P., Weinstein, M.M., Farber, E., Qiao, X., Peale, F., Bunting, S., Walzem, R.L., Wong, J.S., *et al.* (2007). Glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1 plays a critical role in the lipolytic processing of chylomicrons. Cell Metab *5*, 279-291.

Beigneux, A.P., Franssen, R., Bensadoun, A., Gin, P., Melford, K., Peter, J., Walzem, R.L., Weinstein, M.M., Davies, B.S., Kuivenhoven, J.A., *et al.* (2009). Chylomicronemia with a mutant GPIHBP1 (Q115P) that cannot bind lipoprotein lipase. Arterioscler Thromb Vasc Biol *29*, 956-962.

Belmont, J.W., and Leal, S.M. (2005). Complex phenotypes and complex genetics: an introduction to genetic studies of complex traits. Curr Atheroscler Rep *7*, 180-187.

Botham, K.M., Moore, E.H., De Pascale, C., and Bejta, F. (2007). The induction of macrophage foam cell formation by chylomicron remnants. Biochem Soc Trans *35*, 454-458.

Breckenridge, W.C., Little, J.A., Steiner, G., Chow, A., and Poapst, M. (1978). Hypertriglyceridemia associated with deficiency of apolipoprotein C-II. N Engl J Med *298*, 1265-1273.

Calandra, S., Priore Oliva, C., Tarugi, P., and Bertolini, S. (2006). APOA5 and triglyceride metabolism, lesson from human APOA5 deficiency. Curr Opin Lipidol *17*, 122-127.

Cefalu, A.B., Noto, D., Arpi, M.L., Yin, F., Spina, R., Hilden, H., Barbagallo, C.M., Carroccio, A., Tarugi, P., Squatrito, S., *et al.* (2009). Novel LMF1 nonsense mutation in a patient with severe hypertriglyceridemia. J Clin Endocrinol Metab *94*, 4584-4590.

Cha, J.Y., and Repa, J.J. (2007). The liver X receptor (LXR) and hepatic lipogenesis. The carbohydrate-response element-binding protein is a target gene of LXR. J Biol Chem *282*, 743-751.

Charlton-Menys, V., and Durrington, P.N. (2005). Apolipoprotein A5 and hypertriglyceridemia. Clin Chem *51*, 295-297.

Choi, J.C., Lee, J.S., Kang, S.Y., Kang, J.H., and Bae, J.M. (2009). Family history and risk for ischemic stroke: Sibling history is more strongly correlated with the disease than parental history. J Neurol Sci.

Coca-Prieto, I., Kroupa, O., Gonzalez-Santos, P., Magne, J., Olivecrona, G., Ehrenborg, E., and Valdivielso, P. (2011). Childhood-onset chylomicronaemia with reduced plasma lipoprotein lipase activity and mass: identification of a novel GPIHBP1 mutation. J Intern Med.

Connelly, P.W., Maguire, G.F., and Little, J.A. (1987). Apolipoprotein CIISt. Michael. Familial apolipoprotein CII deficiency associated with premature vascular disease. J Clin Invest *80*, 1597-1606.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., *et al.* (2009). Origins and functional impact of copy number variation in the human genome. Nature *464*, 704-712.

Conticello, S.G. (2008). The AID/APOBEC family of nucleic acid mutators. Genome Biol *9*, 229.

Cullen, P. (2000). Evidence that triglycerides are an independent coronary heart disease risk factor. Am J Cardiol *86*, 943-949.

Dentin, R., Girard, J., and Postic, C. (2005). Carbohydrate responsive element binding protein (ChREBP) and sterol regulatory element binding protein-1c (SREBP-1c): two key regulators of glucose metabolism and lipid synthesis in liver. Biochimie *87*, 81-86.

Devlin, B., and Roeder, K. (1999). Genomic control for association studies. Biometrics *55*, 997-1004.

Doran, A.C., Meller, N., and McNamara, C.A. (2008). Role of smooth muscle cells in the initiation and early progression of atherosclerosis. Arterioscler Thromb Vasc Biol *28*, 812-819.

Drexel, H. (2009). Statins, fibrates, nicotinic acid, cholesterol absorption inhibitors, anion-exchange resins, omega-3 fatty acids: which drugs for which patients? Fundam Clin Pharmacol *23*, 687-692.

Ducheix, S., Lobaccaro, J.M., Martin, P.G., and Guillou, H. (2011). Liver X Receptor: an oxysterol sensor and a major player in the control of lipogenesis. Chem Phys Lipids *164*, 500-514.

Durbin, R.M., Abecasis, G.R., Altshuler, D.L., Auton, A., Brooks, L.D., 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.

Ewald, N., Hardt, P.D., and Kloer, H.U. (2009). Severe hypertriglyceridemia and pancreatitis: presentation and management. Curr Opin Lipidol *20*, 497-504.

Expert Panel on Detection; Evaluation; and Treatment of High Blood Cholesterol in Adults, T. (2001). Executive Summary of The 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). JAMA *285*, 2486-2497.

Franssen, R., Young, S.G., Peelman, F., Hertecant, J., Sierts, J.A., Schimmel, A.W., Bensadoun, A., Kastelein, J.J., Fong, L., Dallinga-Thie, G.M., *et al.* (2010a). Chylomicronemia with Low Postheparin Lipoprotein Lipase Levels in the Setting of GPIHBP1 Defects. Circ Cardiovasc Genet *3*, 169-178.

Franssen, R., Young, S.G., Peelman, F., Hertecant, J., Sierts, J.A., Schimmel, A.W., Bensadoun, A., Kastelein, J.J., Fong, L.G., Dallinga-Thie, G.M., *et al.* (2010b). Chylomicronemia with low postheparin lipoprotein lipase levels in the setting of GPIHBP1 defects. Circ Cardiovasc Genet *3*, 169-178.

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., *et al.* (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851-861.

Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. Nat Rev Genet *10*, 241-251.

Freiberg, J.J., Tybjaerg-Hansen, A., Jensen, J.S., and Nordestgaard, B.G. (2008). Nonfasting triglycerides and risk of ischemic stroke in the general population. JAMA *300*, 2142-2152.

Genest, J., McPherson, R., Frohlich, J., Anderson, T., Campbell, N., Carpentier, A., Couture, P., Dufour, R., Fodor, G., Francis, G.A., *et al.* (2009). 2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of dyslipidemia and prevention of cardiovascular disease in the adult - 2009 recommendations. Can J Cardiol *25*, 567-579.

Gimbrone, M.A., Jr. (1999). Endothelial dysfunction, hemodynamic forces, and atherosclerosis. Thromb Haemost *82*, 722-726.

Gin, P., Beigneux, A.P., Davies, B., Young, M.F., Ryan, R.O., Bensadoun, A., Fong, L.G., and Young, S.G. (2007). Normal binding of lipoprotein lipase, chylomicrons, and apo-AV to GPIHBP1 containing a G56R amino acid substitution. Biochim Biophys Acta *1771*, 1464-1468.

Ginsberg, H.N. (2000). Insulin resistance and cardiovascular disease. J Clin Invest *106*, 453-458.

Hegele, R.A. (2002). SNP judgments and freedom of association. Arterioscler Thromb Vasc Biol *22*, 1058-1061.

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

Hegele, R.A., Ban, M.R., Hsueh, N., Kennedy, B.A., Cao, H., Zou, G.Y., Anand, S., Yusuf, S., Huff, M.W., and Wang, J. (2009). A polygenic basis for four classical Fredrickson hyperlipoproteinemia phenotypes that are characterized by hypertriglyceridemia. Hum Mol Genet *18*, 4189-4194.

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.

Hopkins, P.N., Heiss, G., Ellison, R.C., Province, M.A., Pankow, J.S., Eckfeldt, J.H., and Hunt, S.C. (2003). Coronary artery disease risk in familial combined hyperlipidemia and familial hypertriglyceridemia: a case-control comparison from the National Heart, Lung, and Blood Institute Family Heart Study. Circulation *108*, 519-523.

Hurles, M.E., Dermitzakis, E.T., and Tyler-Smith, C. (2008). The functional impact of structural variation in humans. Trends Genet *24*, 238-245.

International HapMap Consortium, T. (2003). The International HapMap Project. Nature *426*, 789-796.

Iqbal, J., and Hussain, M.M. (2009). Intestinal lipid absorption. Am J Physiol Endocrinol Metab *296*, E1183-1194.

Jiang, Z.G., Liu, Y., Hussain, M.M., Atkinson, D., and McKnight, C.J. (2008). Reconstituting initial events during the assembly of apolipoprotein B-containing lipoproteins in a cell-free system. J Mol Biol *383*, 1181-1194.

Jousilahti, P., Vartiainen, E., Tuomilehto, J., and Puska, P. (1999). Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14 786 middle-aged men and women in Finland. Circulation *99*, 1165-1172.

Juutilainen, A., Lehto, S., Ronnemaa, T., Pyorala, K., and Laakso, M. (2005). Type 2 diabetes as a "coronary heart disease equivalent": an 18-year prospective population-based study in Finnish subjects. Diabetes Care *28*, 2901-2907.

Kannel, W.B., Castelli, W.P., and Gordon, T. (1979). Cholesterol in the prediction of atherosclerotic disease. New perspectives based on the Framingham study. Ann Intern Med *90*, 85-91.

Kannel, W.B., and Vasan, R.S. (2009). Triglycerides as vascular risk factors: new epidemiologic insights. Curr Opin Cardiol *24*, 345-350.

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., *et al.* (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., Voight, B.F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P.M., Anand, S., Engert, J.C., Samani, N.J., Schunkert, H., *et al.* (2009a). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet *41*, 334-341.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., *et al.* (2009b). Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet *41*, 56-65.

Khera, A.V., Cuchel, M., de la Llera-Moya, M., Rodrigues, A., Burke, M.F., Jafri, K., French, B.C., Phillips, J.A., Mucksavage, M.L., Wilensky, R.L., *et al.* (2011). Cholesterol efflux capacity, high-density lipoprotein function, and atherosclerosis. N Engl J Med *364*, 127-135.

Kirchgessner, T.G., LeBoeuf, R.C., Langner, C.A., Zollman, S., Chang, C.H., Taylor, B.A., Schotz, M.C., Gordon, J.I., and Lusis, A.J. (1989). Genetic and developmental regulation of the lipoprotein lipase gene: loci both distal and

proximal to the lipoprotein lipase structural gene control enzyme expression. J Biol Chem *264*, 1473-1482.

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

Kraft, P. (2011). Population stratification bias: more widespread than previously thought. Epidemiology *22*, 408-409.

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 Protoc *4*, 1073-1081.

Labreuche, J., Touboul, P.J., and Amarenco, P. (2009). Plasma triglyceride levels and risk of stroke and carotid atherosclerosis: a systematic review of the epidemiological studies. Atherosclerosis *203*, 331-345.

Lander, E.S., and Botstein, D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science *236*, 1567-1570.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lanktree, M.B., and Hegele, R.A. (2009). Gene-gene and gene-environment interactions: new insights into the prevention, detection and management of coronary artery disease. Genome Med *1*, 28.

LaRosa, J.C., Levy, R.I., Herbert, P., Lux, S.E., and Fredrickson, D.S. (1970). A specific apoprotein activator for lipoprotein lipase. Biochem Biophys Res Commun *41*, 57-62.

Lee, H.Y., and Oh, B.H. (2010). Aging and arterial stiffness. Circ J *74*, 2257-2262.

Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype imputation. Annu Rev Genomics Hum Genet *10*, 387-406.

Libby, P., and Theroux, P. (2005). Pathophysiology of coronary artery disease. Circulation *111*, 3481-3488.

Lloyd-Jones, D.M. (2006). Short-term versus long-term risk for coronary artery disease: implications for lipid guidelines. Curr Opin Lipidol *17*, 619-625.

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

Lopez-Miranda, J., Williams, C., and Lairon, D. (2007). Dietary, physiological, genetic and pathological influences on postprandial lipid metabolism. Br J Nutr *98*, 458-473.

Lusis, A.J. (2000). Atherosclerosis. Nature *407*, 233-241.

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.

MacLean, D.R., Petrasovits, A., Nargundkar, M., Connelly, P.W., MacLeod, E., Edwards, A., and Hessel, P. (1992). Canadian heart health surveys: a profile of cardiovascular risk. Survey methods and data analysis. Canadian Heart Health Surveys Research Group. CMAJ *146*, 1969-1974.

Malyszko, J. (2010). Mechanism of endothelial dysfunction in chronic kidney disease. Clin Chim Acta *411*, 1412-1420.

Mamo, J.C., Proctor, S.D., and Smith, D. (1998). Retention of chylomicron remnants by arterial tissue; importance of an efficient clearance mechanism from plasma. Atherosclerosis *141 Suppl 1*, S63-69.

Mansbach, C.M., 2nd, and Gorelick, F. (2007). Development and physiological regulation of intestinal lipid absorption. II. Dietary lipid absorption, complex lipid synthesis, and the intracellular packaging and secretion of chylomicrons. Am J Physiol Gastrointest Liver Physiol *293*, G645-650.

Marcais, C., Verges, B., Charriere, S., Pruneta, V., Merlin, M., Billon, S., Perrot, L., Drai, J., Sassolas, A., Pennacchio, L.A., *et al.* (2005). Apoa5 Q139X truncation predisposes to late-onset hyperchylomicronemia due to lipoprotein lipase impairment. J Clin Invest *115*, 2862-2869.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat Rev Genet *11*, 31-46.

Michos, E.D., Vasamreddy, C.R., Becker, D.M., Yanek, L.R., Moy, T.F., Fishman, E.K., Becker, L.C., and Blumenthal, R.S. (2005). Women with a low Framingham risk score and a family history of premature coronary heart disease have a high prevalence of subclinical coronary atherosclerosis. Am Heart J *150*, 1276-1281.

Morrison, A., and Hokanson, J.E. (2009). The independent relationship between triglycerides and coronary heart disease. Vasc Health Risk Manag *5*, 89-95.

Morrison, A.C., Bare, L.A., Chambless, L.E., Ellis, S.G., Malloy, M., Kane, J.P., Pankow, J.S., Devlin, J.J., Willerson, J.T., and Boerwinkle, E. (2007). Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. Am J Epidemiol *166*, 28-35.

Morton, N.E. (1955). Sequential tests for the detection of linkage. Am J Hum Genet *7*, 277-318.

Murabito, J.M., Pencina, M.J., Nam, B.H., D'Agostino, R.B., Sr., Wang, T.J., Lloyd-Jones, D., Wilson, P.W., and O'Donnell, C.J. (2005). Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. JAMA *294*, 3117-3123.

Myers, R.H., Kiely, D.K., Cupples, L.A., and Kannel, W.B. (1990). Parental history is an independent risk factor for coronary artery disease: the Framingham Study. Am Heart J *120*, 963-969.

Nasir, K., Budoff, M.J., Wong, N.D., Scheuner, M., Herrington, D., Arnett, D.K., Szklo, M., Greenland, P., and Blumenthal, R.S. (2007). Family history of premature coronary heart disease and coronary artery calcification: Multi-Ethnic Study of Atherosclerosis (MESA). Circulation *116*, 619-626.

Neale, B.M., Rivas, M.A., Voight, B.F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S.M., Roeder, K., and Daly, M.J. (2011). Testing for an unusual distribution of rare variants. PLoS Genet *7*, e1001322.

Nordestgaard, B.G., Benn, M., Schnohr, P., and Tybjaerg-Hansen, A. (2007). Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. JAMA *298*, 299-308.

Olivecrona, G., Ehrenborg, E., Semb, H., Makoveichuk, E., Lindberg, A., Hayden, M.R., Gin, P., Davies, B.S., Weinstein, M.M., Fong, L.G., *et al.* (2010). Mutation of conserved cysteines in the Ly6 domain of GPIHBP1 in familial chylomicronemia. J Lipid Res *51*, 1535-1545.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet *2*, e190.

Perez-Lopez, F.R., Larrad-Mur, L., Kallen, A., Chedraui, P., and Taylor, H.S. (2010). Gender differences in cardiovascular disease: hormonal and biochemical influences. Reprod Sci *17*, 511-531.

Peterfy, M., Ben-Zeev, O., Mao, H.Z., Weissglas-Volkov, D., Aouizerat, B.E., Pullinger, C.R., Frost, P.H., Kane, J.P., Malloy, M.J., Reue, K., *et al.* (2007). Mutations in

LMF1 cause combined lipase deficiency and severe hypertriglyceridemia. Nat Genet *39*, 1483-1487.

Pollex, R.L., and Hegele, R.A. (2005). Complex trait locus linkage mapping in atherosclerosis: time to take a step back before moving forward? Arterioscler Thromb Vasc Biol *25*, 1541-1544.

Pollex, R.L., and Hegele, R.A. (2007). Copy number variation in the human genome and its implications for cardiovascular disease. Circulation *115*, 3130-3138.

Poupeau, A., and Postic, C. (2011). Cross-regulation of hepatic glucose metabolism via ChREBP and nuclear receptors. Biochim Biophys Acta *1812*, 995-1006.

Price, A.L., Kryukov, G.V., de Bakker, P.I., Purcell, S.M., Staples, J., Wei, L.J., and Sunyaev, S.R. (2010). Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet *86*, 832-838.

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.

Priore Oliva, C., Carubbi, F., Schaap, F.G., Bertolini, S., and Calandra, S. (2008). Hypertriglyceridaemia and low plasma HDL in a patient with apolipoprotein A-V deficiency due to a novel mutation in the APOA5 gene. J Intern Med *263*, 450-458.

Priore Oliva, C., Pisciotta, L., Li Volti, G., Sambataro, M.P., Cantafora, A., Bellocchio, A., Catapano, A., Tarugi, P., Bertolini, S., and Calandra, S. (2005). Inherited apolipoprotein A-V deficiency in severe hypertriglyceridemia. Arterioscler Thromb Vasc Biol *25*, 411-417.

Pritchard, J.K. (2001). Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet *69*, 124-137.

Rader, D.J. (2007). Mechanisms of disease: HDL metabolism as a target for novel therapies. Nat Clin Pract Cardiovasc Med *4*, 102-109.

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

Rader, D.J., and Daugherty, A. (2008). Translating molecular discoveries into new therapies for atherosclerosis. Nature *451*, 904-913.

Rahalkar, A.R., Giffen, F., Har, B., Ho, J., Morrison, K.M., Hill, J., Wang, J., Hegele, R.A., and Joy, T. (2009). Novel LPL mutations associated with lipoprotein lipase

deficiency: two case reports and a literature review. Can J Physiol Pharmacol *87*, 151-160.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., *et al.* (2006). Global variation in copy number in the human genome. Nature *444*, 444-454.

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

Repa, J.J., Liang, G., Ou, J., Bashmakov, Y., Lobaccaro, J.M., Shimomura, I., Shan, B., Brown, M.S., Goldstein, J.L., and Mangelsdorf, D.J. (2000). Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXRalpha and LXRbeta. Genes Dev *14*, 2819-2830.

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

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

Roy, H., Bhardwaj, S., and Yla-Herttuala, S. (2009). Molecular genetics of atherosclerosis. Hum Genet *125*, 467-491.

Sabatti, C., Service, S.K., Hartikainen, A.L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M., *et al.* (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet *41*, 35-46.

Schork, N.J., Murray, S.S., Frazer, K.A., and Topol, E.J. (2009). Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev *19*, 212-219.

Shaikh, M., Wootton, R., Nordestgaard, B.G., Baskerville, P., Lumley, J.S., La Ville, A.E., Quiney, J., and Lewis, B. (1991). Quantitative studies of transfer in vivo of low density, Sf 12-60, and Sf 60-400 lipoproteins between plasma and arterial intima in humans. Arterioscler Thromb *11*, 569-577.

Shi, Y., and Cheng, D. (2009). Beyond triglyceride synthesis: the dynamic functional roles of MGAT and DGAT enzymes in energy metabolism. Am J Physiol Endocrinol Metab *297*, E10-18.

Singh, I.M., Shishehbor, M.H., and Ansell, B.J. (2007). High-density lipoprotein as a therapeutic target: a systematic review. JAMA *298*, 786-798.

Sniderman, A.D., and Furberg, C.D. (2008). Age as a modifiable risk factor for cardiovascular disease. Lancet *371*, 1547-1549.

Taraboanta, C., Wu, E., Lear, S., DiPalma, S., Hill, J., Mancini, G.B., and Frohlich, J. (2008). Subclinical atherosclerosis in subjects with family history of premature coronary artery disease. Am Heart J *155*, 1020-1026 e1021.

Vance, D.E., and Vance, J.E., eds. (2004). Biochemistry of Lipids, Lipoproteins and Membranes, 4th edition, 4th edn (Paris, Elsevier Ltd).

Vasan, R.S., Larson, M.G., Leip, E.P., Evans, J.C., O'Donnell, C.J., Kannel, W.B., and Levy, D. (2001). Impact of high-normal blood pressure on the risk of cardiovascular disease. N Engl J Med *345*, 1291-1297.

Villablanca, A.C., Jayachandran, M., and Banka, C. (2010). Atherosclerosis and sex hormones: current concepts. Clin Sci (Lond) *119*, 493-513.

Voss, C.V., Davies, B.S., Tat, S., Gin, P., Fong, L.G., Pelletier, C., Mottler, C.D., Bensadoun, A., Beigneux, A.P., and Young, S.G. (2011). Mutations in lipoprotein lipase that block binding to the endothelial cell transporter GPIHBP1. Proc Natl Acad Sci U S A.

Wagenseil, J.E., and Mecham, R.P. (2009). Vascular extracellular matrix and arterial mechanics. Physiol Rev *89*, 957-989.

Wallace, C., Newhouse, S.J., Braund, P., Zhang, F., Tobin, M., Falchi, M., Ahmadi, K., Dobson, R.J., Marcano, A.C., Hajat, C., *et al.* (2008). Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. Am J Hum Genet *82*, 139-149.

Wang, J., Ban, M.R., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., and Hegele, R.A. (2008a). APOA5 genetic variants are markers for classic hyperlipoproteinemia phenotypes and hypertriglyceridemia. Nat Clin Pract Cardiovasc Med *5*, 730-737.

Wang, J., Ban, M.R., Zou, G.Y., Cao, H., Lin, T., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., *et al.* (2008b). Polygenic determinants of severe hypertriglyceridemia. Hum Mol Genet *17*, 2894-2899.

Wang, J., and Hegele, R.A. (2007). Homozygous missense mutation (G56R) in glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1 (GPI-HBP1) in two siblings with fasting chylomicronemia (MIM 144650). Lipids Health Dis *6*, 23.

Wang, T.J., Nam, B.H., D'Agostino, R.B., Wolf, P.A., Lloyd-Jones, D.M., MacRae, C.A., Wilson, P.W., Polak, J.F., and O'Donnell, C.J. (2003). Carotid intima-media

thickness is associated with premature parental coronary heart disease: the Framingham Heart Study. Circulation *108*, 572-576.

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., *et al.* (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet *40*, 161-169.

Wilson, P.W., D'Agostino, R.B., Levy, D., Belanger, A.M., Silbershatz, H., and Kannel, W.B. (1998). Prediction of coronary heart disease using risk factor categories. Circulation *97*, 1837-1847.

Yuan, G., Al-Shali, K.Z., and Hegele, R.A. (2007). Hypertriglyceridemia: its etiology, effects and treatment. CMAJ *176*, 1113-1120.

Yuan, G., Wang, J., and Hegele, R.A. (2006). Heterozygous familial hypercholesterolemia: an underrecognized cause of early cardiovascular disease. CMAJ *174*, 1124-1129.

CHAPTER 2

COMMON GENETIC VARIATION IN TRIGLYCERIDE-ASSOCIATED GENES IS ASSOCIATED WITH HYPERTRIGLYCERIDEMIA AND ACCUMULATES IN PATIENTS VERSUS HEALTHY CONTROLS

The work in this chapter originates from material in the following publications: (1) **Johansen, CT,** Wang, J, Lanktree, MB, *et al*. (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 42(8): 684-7; (2) Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, **Johansen CT**, *et al*. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466(7307): 707-13.

2.1. Introduction

Plasma TG concentration is re-emerging as a risk factor for cardiovascular disease (CVD). Epidemiological evidence has suggested that prolonged exposure to elevated concentrations of triglyceride (TG)-rich lipoproteins, especially in the post prandial state, increases cardiovascular risk independent of other plasma lipoproteins (Bansal et al., 2007, Cullen, 2000, Freiberg et al., 2008, Hokanson and Austin, 1996, Labreuche et al., 2009, Morrison and Hokanson, 2009, Nordestgaard et al., 2007).Thus, patients diagnosed with clinical hypertriglyceridemia (HTG) likely also have significantly elevated CVD risk (Yuan et al., 2007). However, the pathways that modulate plasma TG concentrations

are incompletely understood, as indicated by the lack of very few effective therapies to lower plasma TG. Elucidation of the genes and genetic variants associated with plasma TG concentration can enrich our understanding of biochemical pathways involved in TG-rich lipoprotein metabolism. Furthermore, identification of defects in patients with disordered TG metabolism could guide the development of therapeutic interventions to treat HTG and ameliorate CVD risk.

HTG is a complex polygenic disease defined by plasma TG concentrations >95th percentile. Pediatric HTG is caused exclusively by homozygous loss-of-function (LOF) mutations in genes such as *LPL* (Havel and Gordon, 1960), *APOC2* (Connelly et al., 1987), *APOA5* (Priore Oliva et al., 2005), *LMF1* (Peterfy et al., 2007), and *GPIHBP1* (Beigneux et al., 2009, Wang and Hegele, 2007). Polygenic HTG has a much more complex etiology. As with other complex diseases, it is thought to result from the accumulation of genetic determinants associated with defective TG-rich lipoprotein hydrolysis (Plomin et al., 2009), combined with environmental and metabolic stressors such as alcohol, medications, renal disease, nonalcoholic fatty-liver disease, pregnancy, obesity, metabolic syndrome, and type 2 diabetes (T2D) (Yuan et al., 2007). Both common and rare genetic variation in genes known to modulated plasma TG concentration have been associated with HTG (Hegele et al., 2009, Wang et al., 2008, Wang et al., 2007), however the majority of phenotypic variation underlying HTG susceptibility remains unattributed (Hegele, 2009).

Genome-wide association studies (GWAS) are a class of recently developed genomic tools capable of identifying regions of the genome containing genetic variation associated with a complex trait (Altshuler et al., 2008). Briefly, a GWAS tests for association between common genetic variants with frequencies >1% in the population (called single nucleotide polymorphisms [SNPs]) and quantitative or discrete traits, enabled by modern microarray technologies and statistical genetic techniques. SNPs included on microarrays are strategically chosen to "tag" additional variants by virtue of blocks of linkage disequilibrium (LD), which cover large regions of the genome and often contain multiple genomic elements including variants, genes, and regulatory elements. GWAS cannot directly ascribe causation to a particular gene or variant underlying an association, although it can implicate new genomic regions for further study, irrespective of prior direct biological plausibility or knowledge of variant function. Thus, GWAS is an unbiased method that surveys the entire genome to identify novel genomic elements, such as genes or regulatory elements, involved in biological processes and disease susceptibility.

GWAS have proven effective at identifying association signals containing both classically established genes and previously unknown genomic regions as determinants of plasma TG concentration. For example, before the start of our study, population-based studies had identified 30 loci associated with quantitative lipid traits, including 11 loci associated with plasma TG concentration (Kathiresan et al., 2009b). Here, we sought to comprehensively characterize the contribution of common genetic variation to HTG susceptibility. Our main objective was to perform a GWAS of patients with HTG in order

to identify novel genes and common variants associated with HTG. Our secondary objective was to extend associations between common variants in recently identified TG-associated loci to our HTG cohort in a collaborative effort with the Global Lipids Genetics Consortium (GLGC). Our final objective was to assess the combined impact of common variation across multiple TG-associated loci on HTG susceptibility. We demonstrate that loci containing common variation associated with HTG are the same loci associated with plasma TG concentration in population-based studies, and that these common variants also accumulate in HTG patients compared to healthy controls.

2.2. Methods

2.2.1. Study subjects

This study was approved by the ethics boards at all collaborating institutions. All subjects provided informed consent for blood sampling, DNA analysis, and collection of clinical, biochemical and other demographic data. The GWAS of HTG included 463 HTG patients and 1197 controls after removal of subjects due to non-European ancestry or genotyping failure. HTG patients were defined as having untreated 12 h fasting plasma TG concentrations >3.4 mmol/L on at least two occasions. They were unrelated European subjects diagnosed with Fredrickson hyperlipoproteinemia (HLP) phenotypes 2B (MIM 144250), 3 (MIM 107741), 4 (MIM 144600) or 5 (MIM 144650), ascertained primarily from a single tertiary referral lipid clinic (92% of patients) in London, Ontario, Canada, or from a tertiary referral lipid clinic in Amsterdam, Netherlands. Controls were subjects with familial hypercholesterolemia (4% of controls) obtained from a single tertiary referral lipid clinic in London, Ontario, Canada, or normal healthy controls obtained from population-based studies including the Study of Health Assessment and Risk in Ethnic Groups (Anand et al., 2000) (18%) or the Myocardial Infarction Genetics Consortium (Kathiresan et al., 2009a) (78%). Controls had plasma TG concentration <2.3 mmol/L to exclude potentially undiagnosed HTG, but were otherwise not phenotypically-selected. Subjects with familial hypercholesterolemia were included as negative controls to counterbalance the increased cholesterol phenotype that is observed in patients with HTG. Biochemical analyses were conducted separately in each cohort, as previously described (Anand et al., 2000, Hegele et al., 2009, Kathiresan et al., 2009a). The

replication cohort used for analyses conducted in conjunction with the GLGC included a subset of GWAS patients that were available when collaboration began, including 344 HTG patients and 144 healthy population-based.

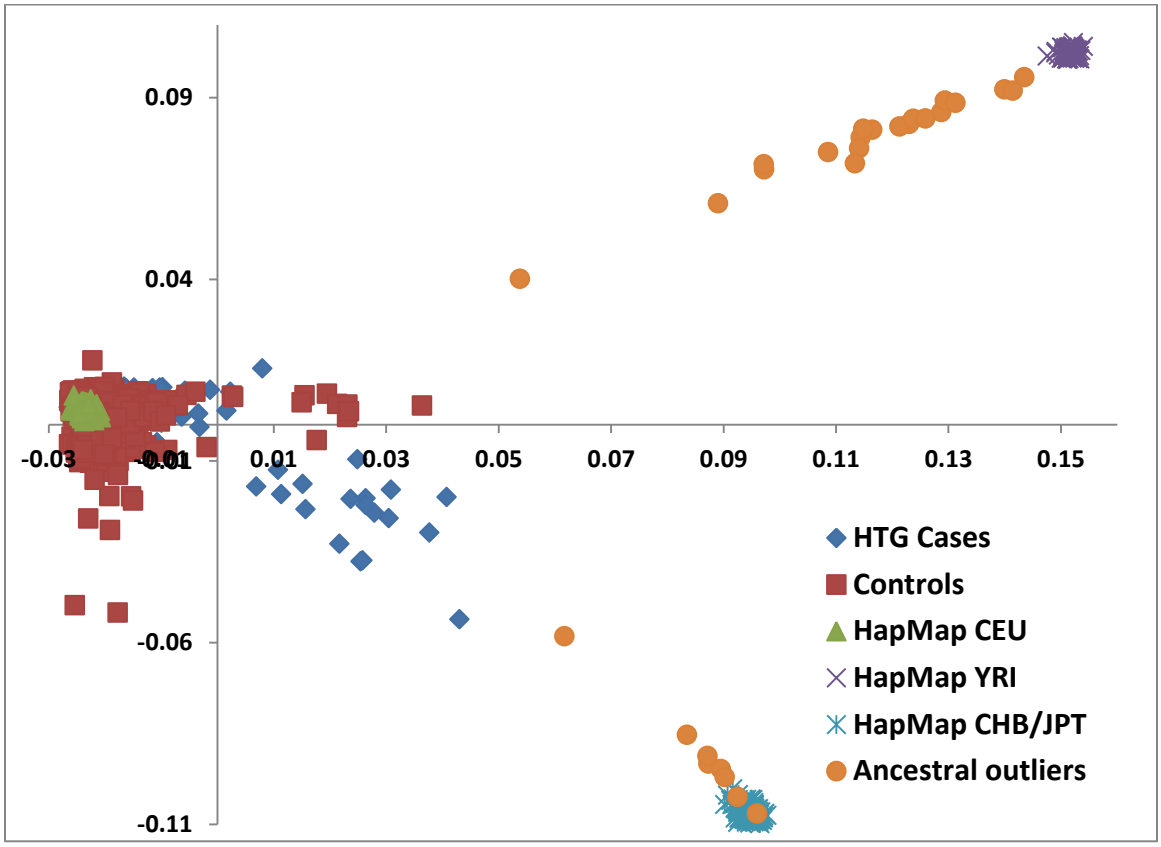
2.2.2. Genotyping and imputation

All subjects were genotyped using Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA) according to protocols of the London Regional Genomics Centre or the Broad Institute. Genotypes were called using Affymetrix Genotyping Console, setting quality control thresholds for subject call rate (>90%), SNP call rate (95%), Hardy-Weinberg equilibrium (P>0.0001) and minor allele frequency (>1%). Imputation was conducted using HapMap CEU phased haplotypes in MACH (Li and Abecasis, 2006). Imputed genotypes were also filtered for minor allele frequency >1%, Hardy Weinberg P>0.0001, and 95% call rate, in addition to imputation quality r^2>0.4. Microarray probe intensity cluster plots were manually verified to ensure robust genotype calling within the cohort for all significant associations.

2.2.3. Statistical analysis

Identity-by-state calculations, multi-dimensional scaling, and association testing were conducted as implemented in PLINK (Purcell et al., 2007). Identity-by-state calculations and multidimensional scaling were performed using freely available HapMap CEU (European ancestry), CHB/JPT (Asian ancestry) and YRI (African ancestry) subjects in order to demonstrate the relative ancestry of study participants (**Figure 2.1**). Subjects

Figure 2.1. Axes of genetic variation in the study sample. Identity-by-state and multidimensional scaling analyses were used to illustrate the relative genetic similarity between subjects, represented as closer relative distance on the plot above. HTG cases (blue diamonds) and controls (red squares) are predominantly of European ancestry, assessed by clustering with HapMap CEU subjects of European ancestry (green triangles), and distance from HapMap YRI subjects of African ancestry (purple X's) and CHB/JPT subject of East Asian ancestry (blue X's). Potential confounding effects of population stratification were minimized by removal of extreme ancestral outliers (orange circles); outliers were arbitrarily selected by x-axis values exceeding 0.05. Remaining population substructure was corrected using 10 principal components of ancestry as generated by Eigenstrat to adjust for potential differences in allele frequency arising from differences in ancestry among study subjects


0.09
0.04
-0.06
-0.11
-0.03
-0.01
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
HTG Cases
Controls
HapMap CEU
HapMap YRI
HapMap CHB/JPT
Ancestral outliers

exceeding an arbitrary x-axis value of 0.05 (14 HTG patients and 14 controls) were excluded from subsequent analyses, in order to remove subjects deviating from the cluster of European subjects, while maximizing inclusion of study participants. Principal component analysis was used to correct for remaining population substructure, as implemented by Eigenstrat (Patterson et al., 2006, Price et al., 2006). Association was tested using an additive multiple logistic regression model, with case or control status as the dependent variable. Genome-wide significance was pre-specified as a Bonferroni-corrected threshold $P<5X10^{-7}$. This threshold was selected to maximize our ability to identify novel loci. Statistical significance for replication of TG-associated SNPs was $P<0.05$, as all loci were previously implicated in TG metabolism. Covariates entered into all analyses included sex, BMI, T2D and 10 principal components of ancestry as generated by Eigenstrat (Patterson et al., 2006, Price et al., 2006). Explained variation calculations were performed using multiple logistic regression without a predefined modeling strategy. Case or control status was modeled as the dependent variable, entering clinical variables (age, sex, BMI, and T2D), and TG-raising risk alleles at 7 HTG-associated loci (*APOA5, GCKR, LPL, APOB, MLXIPL, TRIB1, ANGPTL3*) as independent variables.

Genetic risk scores were constructed as the sum of TG-raising alleles at all 32 TG-associated loci identified by the GLGC (Teslovich et al., 2010). Two risk scores were constructed: (1) unweighted risk scores composed of the raw number of TG-raising alleles; or (2) weighted risk scores composed of the weighted sum of TG-raising alleles each multiplied by their corresponding population-based effect estimate, as generated by

the GLGC (Teslovich et al., 2010). Units were initially reported as mg/dl but converted here to mmol/L (conversion factor 88.6 mg/dl = 1 mmol/L). All loci were included in risk score analysis regardless of statistical association in the replication phase. Frequency distributions (for both unweighted and weighted risk scores) were constructed by: (1) dividing the range of risk scores into equal bins; and (2) plotting the number of HTG patients or controls in each bin relative to total number of HTG patients or controls in the sample. Forest plots were generated by: (1) ranking all weighted risk scores from lowest to highest; (2) dividing risk scores into 4 bins of equal numbers of subjects; and (3) comparing the number of HTG patients to controls in each bin to the lowest risk score bin, which served as the referent group. The significance of increasing HTG susceptibility with increasing risk score bin was assessed using the Cochrane-Armitage test for trend.

2.3 Results

2.3.1 Study sample

Baseline clinical attributes of the 463 HTG patients and 1197 population-based controls are shown in **Table 2.1**. Generally, clinical attributes of HTG patients were less favourable than those of controls. HTG patients had increased BMI, increased plasma total cholesterol, decreased plasma HDL cholesterol and increased prevalence of T2D. Plasma TG concentrations were most greatly increased, with a mean of 14.3 mmol/L in HTG patients versus 1.1 mmol/L in healthy controls. Clinical attributes of the subset of HTG patients and controls used in the GLGC replication phase were essentially identical to the full cohort (**Table 2.2**).

2.3.2 Genome-wide association study of hypertriglyceridemia

The HTG phenotype was tested for association with >2.1 million genotyped and imputed SNPs using an additive multiple logistic regression model. Four loci were significantly associated with HTG at genome-wide significance levels ($P<5X10^{-7}$): *APOA5*, *GCKR*, *LPL* and *APOB* (**Figure 2.2**). Multiple SNPs at each locus were associated with HTG; however repeat analyses conditioned on the lead SNP at each locus attenuated each association signal, suggesting that only a single association signal was present at each locus. Finally, we determined that covariates entered into this model, which included sex, body mass index, diabetes status and 10 principal components of ancestry, appropriately adjusted for potential confounding effects of stratification as the genomic inflation factor

Table 2.1. Baseline clinical attributes of full GWAS cohort.

Characteristic	HTG	Controls
Number	463	1197
Female	30.7%	40.4%
Diabetes	25.7%	0.4%
Age (years)	50.9 ± 13.0	47.8 ± 11.1
Body mass index (kg/m^2)	29.9 ± 4.9	26.4 ± 4.6
Plasma total cholesterol (mmol/L)	8.2 ± 3.9	5.3 ± 1.3
Plasma HDL cholesterol (mmol/L)	0.9 ± 0.3	1.4 ± 0.4
Plasma LDL cholesterol (mmol/L)	-	3.4 ± 1.2
Plasma TG (mmol/L)	14.3 ± 19.8	1.1 ± 0.7

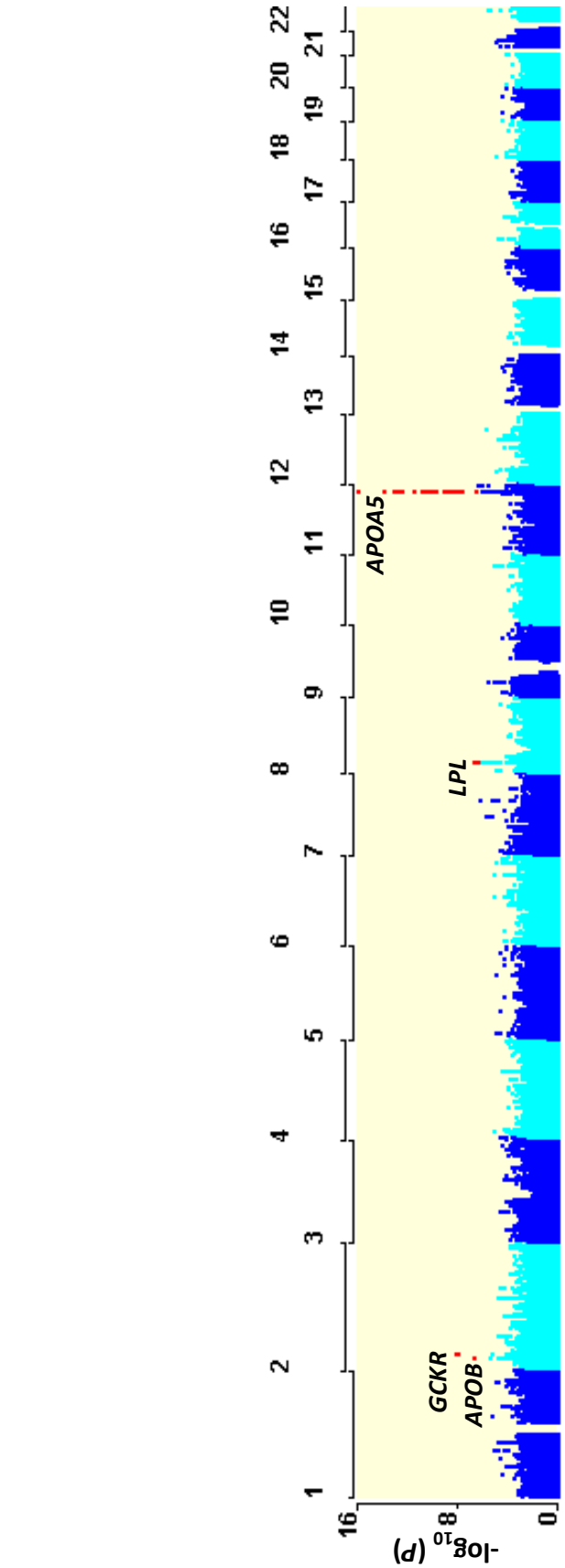
GWAS, genome-wide association study; HDL, high-density lipoprotein; HTG, hypertriglyceridemia; LDL, low-density lipoprotein; TG, triglycerides. Values are mean ± standard deviation. LDL cholesterol is not accurately calculated using the Friedewald equation for HTG patients when plasma TG concentration exceeds 4.5 mmol/L.

Table 2.2. Baseline clinical attributes of the replication cohort.

Characteristic	HTG	Controls
Number	344	144
Female	33.7%	52.1%
Diabetes	27.0%	3.5%
Age (years)	50.2 ± 13.1	45.3 ± 20.0
Body mass index (kg/m^2)	29.9 ± 4.4	26.9 ± 5.6
Plasma total cholesterol (mmol/L)	8.4 ± 4.0	6.2 ± 2.3
Plasma HDL cholesterol (mmol/L)	0.9 ± 0.3	1.3 ± 0.4
Plasma LDL cholesterol (mmol/L)	-	4.3 ± 1.9
Plasma TG (mmol/L)	12.1 ± 16.5	1.2 ± 0.5

GWAS, genome-wide association study; HDL, high-density lipoprotein; HTG, hypertriglyceridemia; LDL, low-density lipoprotein; TG, triglycerides. Values are mean ± standard deviation. LDL cholesterol is not accurately calculated using the Friedewald equation for HTG patients when plasma TG concentration exceeds 4.5 mmol/L.

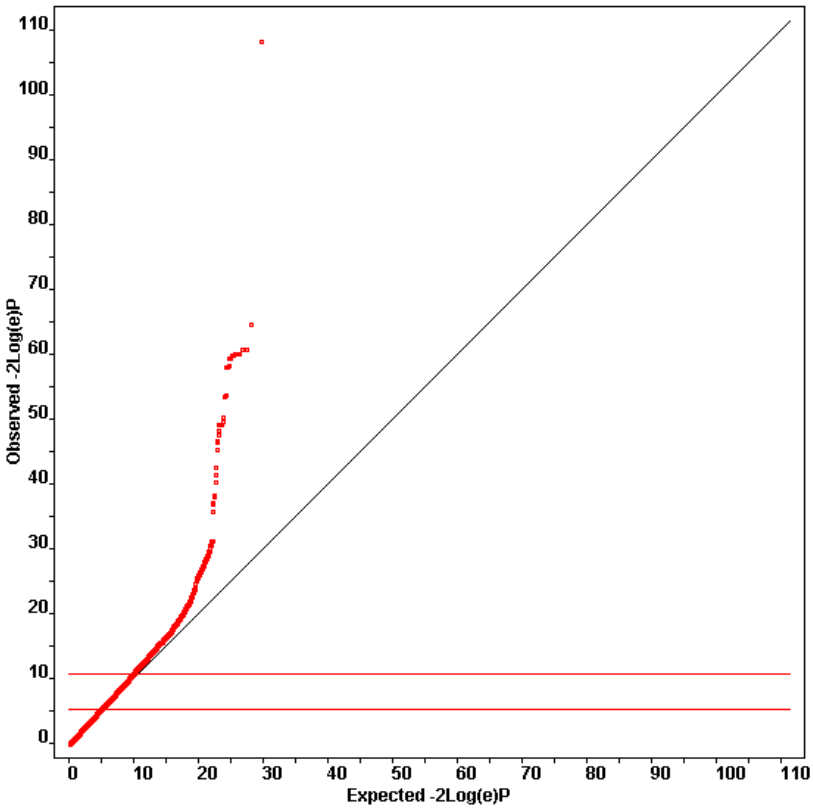
Figure 2.2. Manhattan plot of regression *P*-values. SNPs were independently tested for association with HTG using multiple logistic regression entering sex, body mass index, diabetes status and 10 principal components of ancestry as covariates. In this figure, SNPs are listed in increasing order of physical position along the x-axis (chromosomes are labeled at top), and their strength of association is plotted on the y-axis ($-\log_{10}$[P-value]). A threshold of $P < 5 \times 10^{-7}$ was considered genome-wide significant. Red data points represent SNPs surpassing this genome-wide significance threshold. Genome-wide associated loci are labeled. *APOA5* reached a maximum association statistic of 5.4×10^{-24}; however the y-axis scale is truncated for better visualization of other results. Data was visualized using WGAViewer (Ge et al., 2008).


-log10 (P)
16
8
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
GCKR
APOB
LPL
APOA5

was reduced from $\lambda = 1.25$ to $\lambda = 1.07$. Residual inflation was further reduced upon removal of those SNPs most significantly associated with HTG (**Figure 2.3**).

The HTG-associated loci ranged in strength of effect and association (**Table 2.3**). *APOA5* was the strongest associated locus ($P=5.4X10^{-24}$); each minor risk allele corresponded to a 3.3-fold increase in HTG susceptibility. The lead SNP at this locus is located in a block of LD containing several genes that are central in lipoprotein metabolism, including *APOA5*, *APOA4*, *APOC3* and *APOA1*, making it difficult to determine the precise genomic element containing variation associated with HTG. However, *APOA5* is known to be involved in both TG metabolism and HTG susceptibility, strongly suggesting that it is underlying the association with HTG. *GCKR* was second most strongly associated with HTG ($P=6.5X10^{-9}$); each minor allele increased HTG susceptibility 1.8-fold. The lead SNP at this locus causes a proline to leucine amino acid substitution in GCKR at residue 446 (P446L), suggesting that this functional variant may mediate increased HTG susceptibility. The *LPL* locus was less strongly associated with HTG compared to the preceding loci ($P=2.0X10^{-7}$), however each risk allele increases HTG susceptibility ~3.1-fold. Although the lead SNP is downstream of *LPL*, it is in a block of LD containing only *LPL*. Furthermore, *LPL* is well established in TG metabolism and HTG susceptibility, suggesting that common variants in this gene are underlying the GWAS association. *APOB* was associated with HTG with similar strength to *LPL* ($P=2.0X10^{-7}$), although each risk allele only increased susceptibility by 1.7-fold. This variant was located ~122-kb upsteam of *APOB* in a block of LD spanning the *APOB* promoter. While *APOB* is the only gene at this locus, the position of the associated

Figure 2.3. Quantile-quantile plot of regression *P*-values. Deviation of *P*-values from the null is caused predominantly by significant associations with HTG (top), which is eliminated when significantly associated loci are removed from analysis (bottom). Both plots show minimal residual inflation of association test statistics. Association P-values displayed here were visualized using WGAViewer (Ge et al., 2008).


Observed -2Log(e)P
Expected -2Log(e)P
0
10
20
30
40
50
60
70
80
90
100
110

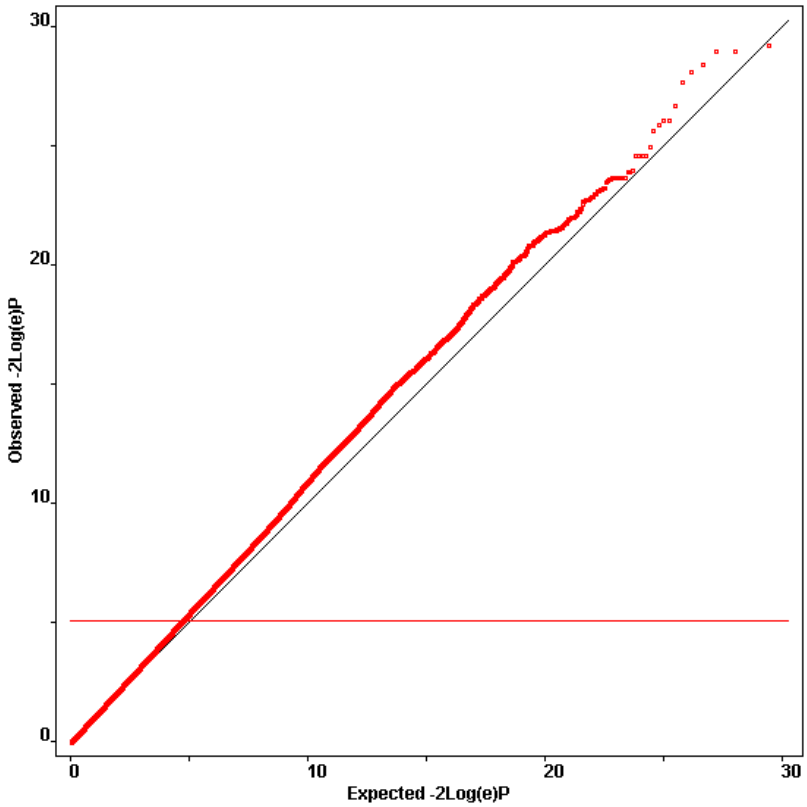

Observed -2Log(e)P
Expected -2Log(e)P
0
10
20
30

Table 2.3. Genetic loci associated with HTG.

Locus	SNP	CHR	Position	Minor Allele	HTG MAF	Control MAF	OR (95% CI)	P-value
APOA5	rs964184	11	116.2	G	0.33	0.14	3.28 (2.61-4.14)	5.4×10^{-24}
GCKR	rs1260326	2	2.8	T	0.52	0.41	1.75 (1.45-2.12)	6.5×10^{-9}
LPL	rs7016880	8	19.9	C	0.03	0.10	0.32 (0.21-0.49)	2.0×10^{-7}
APOB	rs4635554	2	21.2	G	0.39	0.31	1.67 (1.38-2.02)	2.0×10^{-7}
MLXIPL	rs714052	7	72.5	G	0.07	0.13	0.44 (0.31-0.62)	0.000003
TRIB1	rs2954029	8	126.6	T	0.37	0.46	0.71 (0.59-0.86)	0.0004
ANGPTL3	rs10889353	1	62.9	C	0.27	0.32	0.73 (0.59-0.89)	0.002
NCAN	rs17216525	19	19.5	T	0.07	0.09	0.71 (0.50-1.00)	0.05
FADS	rs174547	11	61.3	C	0.40	0.33	1.20 (0.99-1.44)	0.07
XKR6	rs7819412	8	11.1	G	0.46	0.50	0.87 (0.72-1.05)	0.14
PLTP	rs7679	20	44.0	C	0.20	0.19	1.17 (0.94-1.47)	0.16

CHR, chromosome; CI, confidence interval; HTG, hypertriglyceridemia; MAF, minor allele frequency; OR, odds ratio; SNP, single nucleotide polymorphism. Association was tested using an additive multiple logistic regression model, entering sex, body mass index, diabetes status, and 10 principal components of ancestry as covariates. The top four loci surpassed a pre-specified threshold for genome-wide significance of $P<5X10^{-7}$. Remaining loci were replicated from GWAS meta-analysis of population-based TG concentrations (Kathiresan et al., 2009b), using a significance threshold of $P<0.05$.

variant does not allow for common variation in genomic elements that regulate *APOB* expression to be ruled out.

The loci identified by our study had recently been identified by a GWAS of normal plasma TG concentrations in healthy controls (Kathiresan et al., 2009b). This GWAS of normolipidemic plasma TG concentrations conducted in ~40,000 healthy controls had identified 11 TG-associated loci. In addition to the 4 loci identified by our study, they identified associations with common variation in *MLXIPL*, *TRIB1*, *ANGPTL3*, *NCAN*, *FADS1*, and *PLTP*. We hypothesized that the same loci associated with population-based plasma TG concentration could also mediate HTG susceptibility. We reasoned that the strength of association at these loci likely precluded detection at our genome-wide significance threshold, given smaller effect sizes and allele frequencies. Thus, we tested for association between lead SNPs at these loci and HTG in the full GWAS cohort of HTG patients. Associations were replicated at only three loci using a significance threshold of $P<0.05$, including *MLXIPL*, *TRIB1*, and *ANGPTL3*, increasing HTG susceptibility by 1.4-fold to 2.3-fold (**Table 2.3**). Positive replication of these TG-associated loci, combined with trends towards significance at *FADS1-FADS2-FADS3* ($P=0.05$) and *NCAN-CILP2-PBX4* ($P=0.07$), suggested that additional TG-modulating loci may also be involved in HTG pathophysiology.

Finally, we sought to determine what proportion of variation was explained by clinical and genetic variants in HTG patients versus healthy controls. Together, clinical and genetic variables explained 40.5% of variation in HTG diagnoses: clinical variables

(age, sex, BMI and T2D) explained 19.7%, whereas genetic variables at 7 HTG-associated loci (*APOA5, GCKR, LPL, APOB, MLXIPL, TRIB1, and ANGPTL3*) explained an additional 20.8%. It is likely that additional genetic variants in TG-associated loci would explain more variation in case status.

2.3.3. Replication of TG-associated loci from the GLGC

While conducting the GWAS above, we simultaneously engaged in collaboration with the GLGC. Their meta-analysis of >100,000 healthy controls identified 95 loci associated with plasma lipid and lipoprotein traits, including 32 loci associated with plasma TG concentrations. Our contribution was to validate the clinical significance of TG-associated loci in patients with extreme levels of plasma TG. However, when we entered that collaboration only a subset of our sample was available for replication, which included 344 HTG patients and 144 healthy controls (**Table 2.2**). The baseline clinical characteristics of this subset of patients were virtually indistinguishable from the full GWAS cohort.

We first attempted to replicate associations at 32 TG-associated loci identified by the GLGC (**Table 2.4**). In total, 10 loci were associated with HTG at a significance threshold of $P<0.05$. The strongest associated loci were those identified by our initial GWAS of HTG (*APOA5*, *GCKR* and *LPL*) with similar allele frequencies and effect sizes. However, an association was not identified with *APOB* rs1042034. Investigation of this discrepancy revealed that the two variants were separated by ~180-kb and were not in LD ($r^2=0$). Re-analysis of each association using multiple logistic regression

Table 2.4. Replication of 32 TG-associated loci from the GLGC in HTG patients.

Locus	CHR	SNP	Minor allele (risk allele)	HTG MAF	Control MAF	OR	P
APOA5	11	rs964184	G (G)	0.32	0.14	2.75	4.7E-07
GCKR	2	rs1260326	T (T)	0.53	0.36	2.05	5.7E-06
LPL	8	rs12678919	G (A)	0.04	0.10	0.37	0.0013
ANGPTL3	1	rs2131925	G (T)	0.28	0.35	0.59	0.0017
TRIB1	8	rs2954029	T (A)	0.35	0.43	0.62	0.0031
MLXIPL	7	rs17145738	T (C)	0.07	0.13	0.49	0.0040
TYW1B	7	rs13238203	T (C)	0.02	0.04	0.28	0.0062
GALNT2	1	rs4846914	G (G)	0.42	0.35	1.55	0.0088
ZNF664	12	rs4765127	T (G)	0.34	0.27	1.48	0.020
KLHL8	4	rs442177	G (T)	0.35	0.41	0.71	0.041
APOE	19	rs439401	T (C)	0.31	0.34	0.72	0.10
PINX1	8	rs11776767	C (C)	0.39	0.35	1.24	0.17
CAPN3	15	rs2412710	A (A)	0.02	0.01	2.12	0.17
LIPC	15	rs1532085	A (A)	0.39	0.35	1.22	0.20
JMJD1C	10	rs10761731	T (A)	0.41	0.38	1.21	0.22
FRMD5	15	rs2929282	T (T)	0.06	0.05	1.49	0.24
TIMD4	5	rs6882076	T (C)	0.34	0.40	0.84	0.26
IRS1	2	rs2972146	G (T)	0.31	0.33	0.84	0.28
COBLL1	2	rs10195252	C (T)	0.38	0.40	0.85	0.30
NAT2	8	rs1495741	G (G)	0.24	0.22	1.20	0.33
FADS1	11	rs174546	T (T)	0.39	0.36	1.16	0.34
MAP3K1	5	rs9686661	T (T)	0.24	0.19	1.20	0.35
HLA	6	rs2247056	T (C)	0.23	0.25	0.87	0.41
MSL2L1	3	rs645040	G (T)	0.20	0.23	0.86	0.42
CILP2	19	rs10401969	C (T)	0.05	0.07	0.79	0.42
CYP26A1	10	rs2068888	A (G)	0.48	0.52	0.89	0.43
PLTP	20	rs6065906	C (C)	0.18	0.18	1.15	0.47
LRP1	12	rs11613352	T (C)	0.21	0.25	0.89	0.52
APOB	2	rs1042034	C (T)	0.19	0.23	0.91	0.63
KLF12	13	rs9592961	G (A)	0.01	0.01	0.69	0.69
CETP	16	rs3764261	A (C)	0.33	0.37	0.96	0.82
ACSS2	20	rs11546155	A (G)	0.15	0.16	1.03	0.88

CHR, chromosome; HTG, hypertriglyceridemia; GLGC, Global Lipids Genetics Consortium; GWAS, genome-wide association study; MAF, minor allele frequency; OR, odds ratio; SNP, single nucleotide polymorphism. Loci were replicated from the GLGC GWAS meta-analysis of >100,000 population-based TG concentrations (Teslovich et al., 2010), using a significance threshold of $P<0.05$. Association was tested using an additive multiple logistic regression model, entering sex, body mass index, diabetes status, and 10 principal components of ancestry as covariates. This analysis included 344 HTG patients and 144 healthy controls. Risk allele is the TG-raising allele identified by the GLGC.

conditioned upon the other SNP revealed negligible changes in strength of association. We confirmed that the three additional sub-threshold *APOB* variants identified in our initial GWAS were all in moderate LD with our lead SNP rs4635554 (r^2 from 0.35 to 0.51); furthermore, they were not affected by conditioning on the population-based TG-associated SNP *APOB* rs1042034, whereas conditioning upon our lead SNP rs4635554 completely ablated their associations with HTG. This totality of evidence suggested that the HTG-derived *APOB* signal and the population-based *APOB* signal likely represent two independent associations.

Replication was also successful but less robust at loci including *MLXIPL*, *TRIB1*, and *ANGPTL3* (**Table 2.4**), with allele frequencies and effect sizes similar to those observed in the full GWAS cohort. Additional replications were observed at 4 newly identified loci including *TYWIB*, *GALNT2*, *ZNF664*, and *KLHL8*. These variants ranged in effect, increasing HTG susceptibility by 1.4-fold to 3.6-fold. However, only ~1/3 of all variants were statistically associated with HTG ($P<0.05$). Interestingly, there was an over-representation of TG-raising risk alleles at 29/32 loci with directions of effect concordant with population-based studies ($P=1.2X10^{-6}$), suggesting that additional loci likely contribute to increased HTG susceptibility, but again small effect sizes likely limit their detection. Given the small sample size of this subset of our GWAS cohort, we did not assess the additional variation explained by newly discovered TG-associated loci.

We next compared the genetic burden of common TG-associated variants in HTG patients versus healthy controls. We hypothesized that common variants may accumulate

in HTG patients given that the same loci associated with population-based TG concentrations were also associated with increased HTG susceptibility. Unweighted TG risk scores were increased in HTG patients versus controls (**Figure 2.4**). Comparison of the number of HTG patients and healthy controls in each risk score bin revealed a striking trend towards increased proportions of HTG patients compared to healthy controls with increasing risk scores. The mean unweighted risk score was clearly increased in HTG versus controls, with 36.2 ± 0.2 risk alleles versus 33.6 ± 0.3 risk alleles ($P=6.2X10^{-15}$), respectively.

Weighted TG risk scores were also significantly increased in HTG patients versus controls (**Figure 2.5**). The relative frequency distribution was clearly shifted towards increased scores in HTG patients versus controls: 0.99±0.01 mmol/L versus 0.88±0.01 mmol/L ($P=4.7X10^{-32}$), respectively. Furthermore, weighted TG risk scores were an important predictor of HTG susceptibility (**Figure 2.6**). Increasing bins of HTG risk score were distinctly associated with increasing HTG susceptibility ($P_{trend}<0.001$), such that subjects in the highest risk score bin were 7.16 (95% CI: 3.86-13.2) times more likely to be an HTG patient than a healthy control compared to subjects in the lowest risk score bin ($P=8.5X10^{-14}$). Together, these data strongly suggested that an increase burden of common population-based TG associated loci is characteristics of patients with HTG.

Figure 2.4. Distribution of unweighted TG risk scores in HTG patients and healthy controls. Unweighted TG risk scores were constructed as the sum of raw TG-raising alleles at 32 TG-associated loci identified by the GLGC (Teslovich et al., 2010). All loci were included in risk score analysis regardless of statistical association in the replication phase. The figure was constructed by counting the proportion of HTG patients (n=344) to healthy controls (n=144) in each risk score bin.

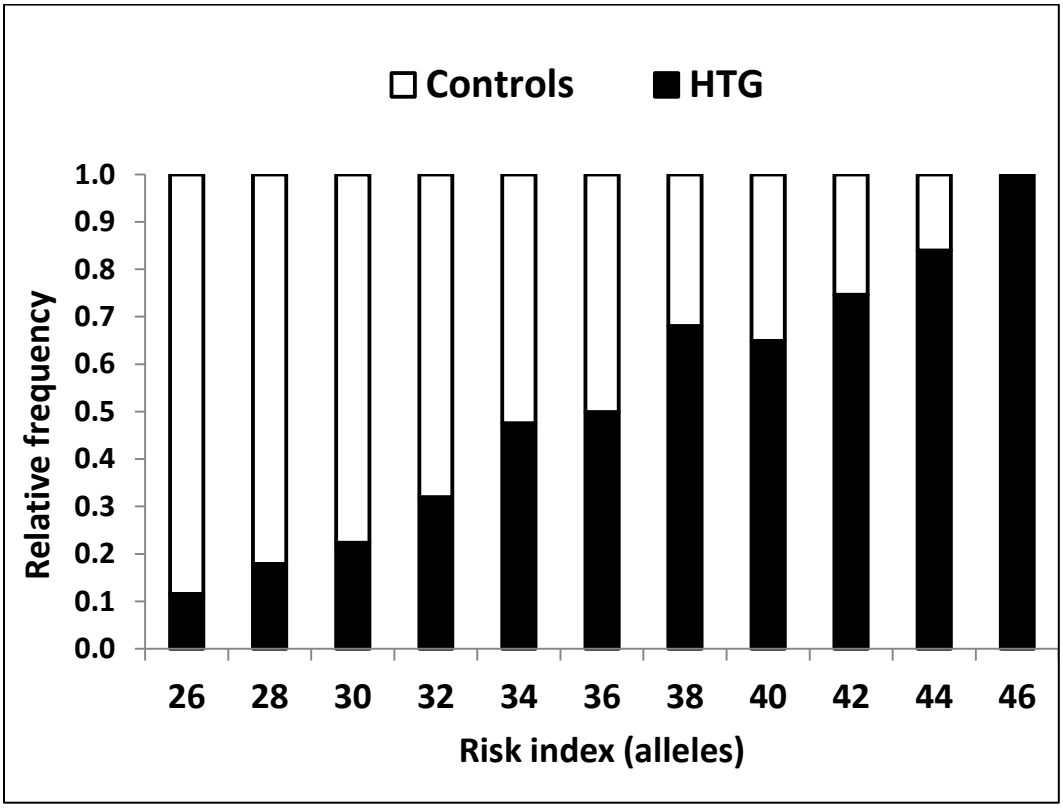

Controls
HTG
Relative frequency
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
26
28
30
32
34
36
38
40
42
44
46
Risk index (alleles)

Figure 2.5. Relative frequency distribution of weighted TG risk scores in HTG patients and healthy controls. Weighted TG risk scores constructed as the sum of TG-raising alleles at all 32 TG-associated loci, each multiplied by the corresponding population-based effect estimate, as generated by the GLGC (Teslovich et al., 2010). Units were initially reported as mg/dl but converted here to mmol/L (conversion factor 88.6 mg/dl = 1 mmol/L). All loci were included in risk score analysis regardless of statistical association in the replication phase. The frequency distributions was constructed by dividing the range of risk scores into equal bins, and plotting the number of HTG patients (n=344) or controls (n=144) in each bin relative to total number of HTG patients or controls in the sample.

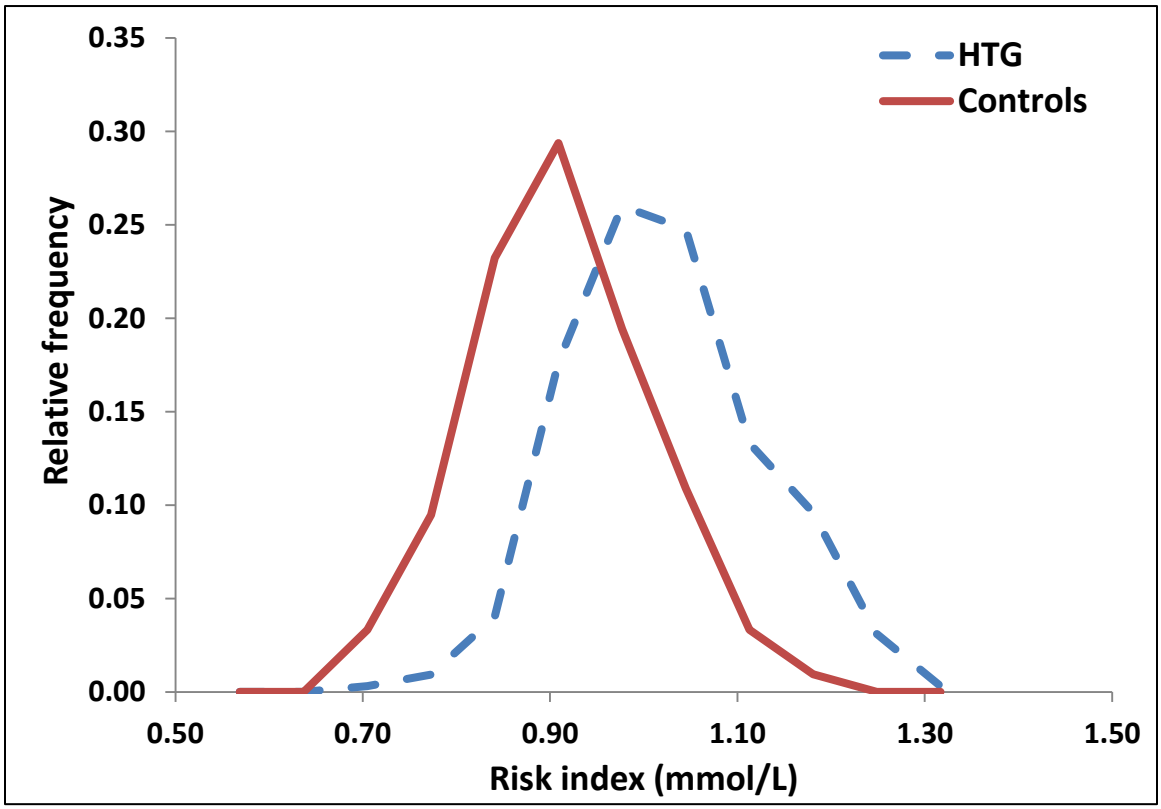
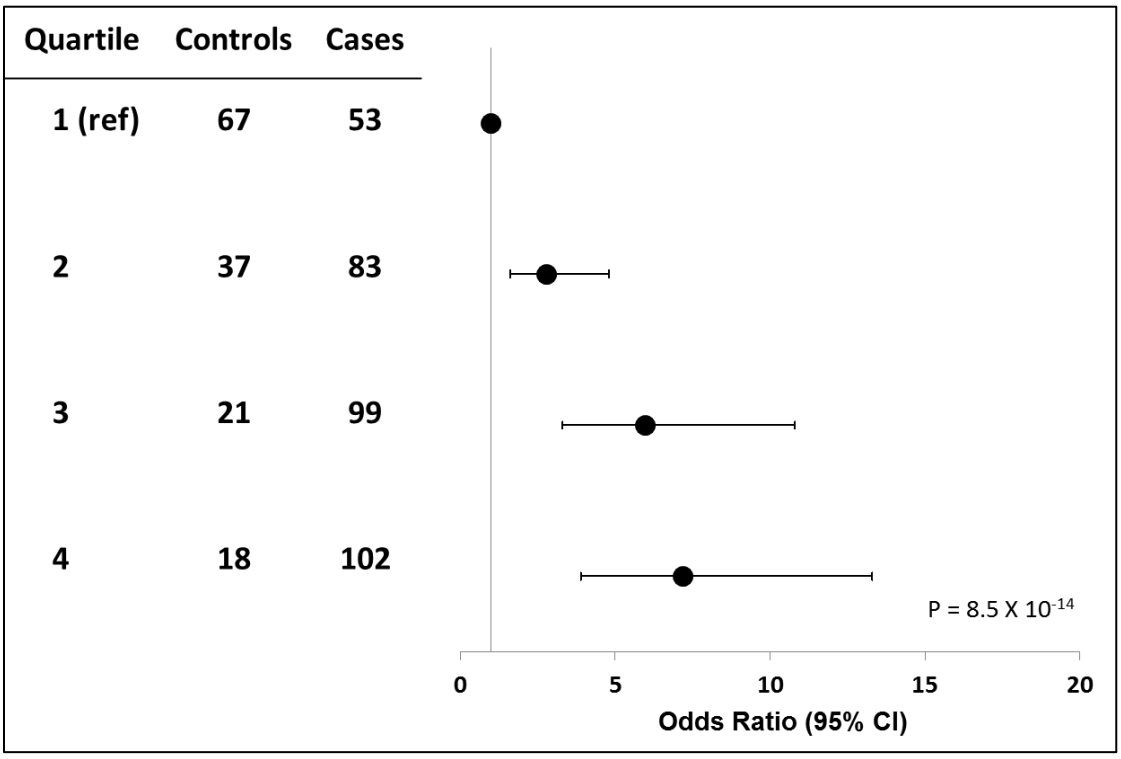

HTG
Controls
Relative frequency
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.50
0.70
0.90
1.10
1.30
1.50
Risk index (mmol/L)

Figure 2.6. Increasing TG risk score is associated with increased HTG susceptibility. Forest plots were generated by ranking all weighted risk scores in the combined sample of HTG patients (n=344) and healthy controls (n=144) from lowest to highest, dividing risk scores into 4 bins of equal numbers of subjects; comparing the number of HTG patients to controls in each bin to the reference bin using Fisher's exact test. The significance of increasing HTG susceptibility with increasing risk score bin was assessed using the Cochrane-Armitage test for trend.


Quartile
Controls
Cases
1 (ref)
67
53
2
37
83
3
21
99
4
18
102
P = 8.5 X 10-14
0
5
10
15
20
Odds Ratio (95% CI)

2.4 Discussion

We have conducted the first GWAS evaluating subjects at the extremes of a lipid phenotype distribution conducted to date. Using a case-control design comparing 463 HTG patients and 1197 healthy population-based controls, we showed genome-wide significant associations with *APOA5*, *GCKR*, *LPL* and *APOB*, and replicated associations at *MLXIPL*, *TRIB1*, *ANGPTL3*. Trends towards significance were also observed at *NCAN* and *FADS1*. Furthermore, we demonstrated that TG-raising alleles at 29/32 TG-associated loci identified by the GLGC were over-represented in HTG patients compared to healthy controls, suggesting that loci that modulate plasma TG concentration in the general population also underlie HTG susceptibility. Furthermore, TG-raising risk alleles appear to accumulate in HTG patients rather than healthy controls, suggesting that an increased genetic burden of common TG-associated variants could be a characteristic of HTG susceptibility. Together, common variants at 7 HTG-associated loci explain ~21% of variation in HTG diagnoses.

A common feature that has emerged from GWAS studies, including our GWAS of HTG patients, is the identification of both known and novel genes involved in the target phenotype. For instance, *LPL* (lipoprotein lipase) and *APOA5* (apolipoprotein A-V) were already appreciated as involved in monogenic and polygenic HTG phenotypes. *LPL* has long been recognized as the fulcrum for hydrolysis of plasma TG-rich lipoproteins (Havel, 2010). It hydrolyses TG-rich lipoproteins in tissues requiring fatty acids for TG re-synthesis in adipose tissue or beta-oxidation in muscle (Kirchgessner et al., 1989). *APOA5* is also a crucial determinant of TG-rich lipoprotein metabolism. It enhances LPL

activity, likely by affecting the distribution of apo C-III molecules on VLDL, thus promoting TG-rich lipoprotein lipolysis (Qu et al., 2007). Rare LOF variants in both *LPL* and *APOA5* are also associated with both monogenic and polygenic HTG (Hegele, 2009).

However, *GCKR* and *APOB* are newly implicated players in HTG susceptibility. *GCKR* encodes an allosteric regulator of glucokinase (GK) called the glucokinase regulatory protein (GKRP), which allows rapid mobilization of GK in response to increased cellular glucose concentrations (Agius, 2008). The P446L functional variant in GKRP is known to attenuate ligand-dependent GK inhibition (Beer et al., 2009, Orho-Melander et al., 2008), in turn promoting glucose uptake and mobilization, thus attenuating fatty acid oxidation and promoting *de novo* TG synthesis (Beer et al., 2009). *APOB* encodes the defining structural component of TG-rich lipoproteins called a B-100, traditionally implicated in extreme phenotypes involving plasma LDL-C, such as familial hypercholesterolemia and familial hypobetalipoproteinemia. Preliminary analyses of *APOB* association signals from population-based studies suggest that the functional variant affects an expression quantitative trait locus (QTL) that modulates transcript levels of both *APOB* and a novel noncoding RNA BU630349, in addition to being a QTL for plasma apo B concentrations (Haas et al., 2011). Although the precise contribution of common variation in *GCKR* and *APOB* to increased HTG susceptibility remains to be elucidated, these associations provide new leads and hypotheses that can be evaluated to better understand the biology and biochemistry of TG metabolism and HTG susceptibility.

Further examination of SNPs below our threshold for significance (i.e. between $P=5X10^{-7}$ and $P=5X10^{-6}$) revealed 82 additional associations localizing predominantly to *APOA5*, *LPL*, *APOB*, and *MLXIPL*. Otherwise, only two potentially novel associated SNPs were found in the *LHFPL3* ($P=5.7X10^{-7}$) and *TSC1* ($P=3.0X10^{-6}$) genes. *LHFPL3* is a poorly characterized gene, annotated as a homolog of *LHFP;* mutations in the latter were identified by mapping of non-syndromic deafness (Longo-Guess et al., 2005). While somewhat better characterized, mutations in *TSC1* have been associated with tuberous sclerosis (Sampson, 2003), and some cancers. Neither locus is thus a credible *a priori* candidate for involvement in HTG susceptibility or TG-metabolism.

A particularly important novel insight provided by our study is that the genetic determinants of plasma TG concentration also contribute to increased HTG susceptibility. For example, the lead variants at *APOA5* and *GCKR* are identical between both GWAS of HTG and population-based studies, and the lead variant in LPL is in the same block of LD between both studies. Furthermore, we successfully replicated TG-associated loci from population-based studies at *MLXIPL*, *TRIB1*, and *ANGPTL3*, and we demonstrated that 29/32 loci have similar directions of effect as those reported in the GLGC. Our interpretation of this trend is that, given sufficient power, the same loci that modulate plasma TG in population-based subjects will be shown also to modulate HTG susceptibility.

We also demonstrated that genetic risk scores composed of TG-associated risk alleles were increased in HTG patients versus controls. Whether unweighted or weighted

by population-based effect estimates, TG-raising alleles accumulated more often in HTG patients. This suggests that a genetic burden of TG-raising alleles may be important in HTG pathophysiology. However, the significant overlap between risk score distributions in HTG patients and healthy controls suggests that common variation contributes only in part to HTG susceptibility; our current estimate being ~21%.

Future studies will require larger sample sizes and different study designs to complement our analyses and elucidate remaining genetic determinants of HTG. Replication and risk score analyses of the 32 TG-associated loci identified by the GLGC are required in a larger HTG cohort to determine whether the hypotheses generated by our study are correct. Furthermore, a more exhaustive survey of genetic variation at HTG-associated loci, obtained by complete resequencing in many subjects, will be required to uncover additional genetic variation underlying HTG susceptibility.

In summary, we used a GWAS of HTG and replication of TG-associated loci to demonstrate that the same genetic determinants of plasma TG concentration are associated with increased HTG susceptibility. Furthermore, TG-raising alleles tend to accumulate in HTG patients, suggesting that a genetic burden of TG-associated risk alleles is characteristic of HTG. Common variation in TG-associated genes only explains ~21% of variation in HTG diagnoses, meaning that additional genes and genetic variants need to be identified to fully understand the complex nature of HTG pathophysiology.

2.5 References

Agius, L. (2008). Glucokinase and molecular aspects of liver glycogen metabolism. Biochem J *414*, 1-18.

Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic mapping in human disease. Science *322*, 881-888.

Anand, S.S., Yusuf, S., Vuksan, V., Devanesen, S., Teo, K.K., Montague, P.A., Kelemen, L., Yi, C., Lonn, E., Gerstein, H., *et al.* (2000). Differences in risk factors, atherosclerosis, and cardiovascular disease between ethnic groups in Canada: the Study of Health Assessment and Risk in Ethnic groups (SHARE). Lancet *356*, 279-284.

Bansal, S., Buring, J.E., Rifai, N., Mora, S., Sacks, F.M., and Ridker, P.M. (2007). Fasting compared with nonfasting triglycerides and risk of cardiovascular events in women. JAMA *298*, 309-316.

Beer, N.L., Tribble, N.D., McCulloch, L.J., Roos, C., Johnson, P.R., Orho-Melander, M., and Gloyn, A.L. (2009). The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. Hum Mol Genet *18*, 4081-4088.

Beigneux, A.P., Franssen, R., Bensadoun, A., Gin, P., Melford, K., Peter, J., Walzem, R.L., Weinstein, M.M., Davies, B.S., Kuivenhoven, J.A., *et al.* (2009). Chylomicronemia with a mutant GPIHBP1 (Q115P) that cannot bind lipoprotein lipase. Arterioscler Thromb Vasc Biol *29*, 956-962.

Connelly, P.W., Maguire, G.F., and Little, J.A. (1987). Apolipoprotein CIISt. Michael. Familial apolipoprotein CII deficiency associated with premature vascular disease. J Clin Invest *80*, 1597-1606.

Cullen, P. (2000). Evidence that triglycerides are an independent coronary heart disease risk factor. Am J Cardiol *86*, 943-949.

Freiberg, J.J., Tybjaerg-Hansen, A., Jensen, J.S., and Nordestgaard, B.G. (2008). Nonfasting triglycerides and risk of ischemic stroke in the general population. JAMA *300*, 2142-2152.

Ge, D., Zhang, K., Need, A.C., Martin, O., Fellay, J., Urban, T.J., Telenti, A., and Goldstein, D.B. (2008). WGAViewer: software for genomic annotation of whole genome association studies. Genome Res *18*, 640-643.

Haas, B.E., Weissglas-Volkov, D., Aguilar-Salinas, C.A., Nikkola, E., Vergnes, L., Cruz-Bautista, I., Riba, L., Stancakova, A., Kuusisto, J., Soininen, P., *et al.* (2011). Evidence of how rs7575840 influences apolipoprotein B-containing lipid particles. Arterioscler Thromb Vasc Biol *31*, 1201-1207.

Havel, R.J. (2010). Triglyceride-rich lipoproteins and plasma lipid transport. Arterioscler Thromb Vasc Biol *30*, 9-19.

Havel, R.J., and Gordon, R.S., Jr. (1960). Idiopathic hyperlipemia: metabolic studies in an affected family. J Clin Invest *39*, 1777-1790.

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

Hegele, R.A., Ban, M.R., Hsueh, N., Kennedy, B.A., Cao, H., Zou, G.Y., Anand, S., Yusuf, S., Huff, M.W., and Wang, J. (2009). A polygenic basis for four classical Fredrickson hyperlipoproteinemia phenotypes that are characterized by hypertriglyceridemia. Hum Mol Genet *18*, 4189-94.

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.

Kathiresan, S., Voight, B.F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P.M., Anand, S., Engert, J.C., Samani, N.J., Schunkert, H., *et al.* (2009a). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet *41*, 334-341.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., *et al.* (2009b). Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet *41*, 56-65.

Kirchgessner, T.G., LeBoeuf, R.C., Langner, C.A., Zollman, S., Chang, C.H., Taylor, B.A., Schotz, M.C., Gordon, J.I., and Lusis, A.J. (1989). Genetic and developmental regulation of the lipoprotein lipase gene: loci both distal and proximal to the lipoprotein lipase structural gene control enzyme expression. J Biol Chem *264*, 1473-1482.

Labreuche, J., Touboul, P.J., and Amarenco, P. (2009). Plasma triglyceride levels and risk of stroke and carotid atherosclerosis: a systematic review of the epidemiological studies. Atherosclerosis *203*, 331-345.

Li, Y., and Abecasis, G.R. (2006). Mach 1.0: rapid haplotype reconstruction and missing genotype inference. Am J Hum Genet *S79*, 2290.

Longo-Guess, C.M., Gagnon, L.H., Cook, S.A., Wu, J., Zheng, Q.Y., and Johnson, K.R. (2005). A missense mutation in the previously undescribed gene Tmhs underlies deafness in hurry-scurry (hscy) mice. Proc Natl Acad Sci U S A *102*, 7894-7899.

Morrison, A., and Hokanson, J.E. (2009). The independent relationship between triglycerides and coronary heart disease. Vasc Health Risk Manag *5*, 89-95.

Nordestgaard, B.G., Benn, M., Schnohr, P., and Tybjaerg-Hansen, A. (2007). Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. JAMA *298*, 299-308.

Orho-Melander, M., Melander, O., Guiducci, C., Perez-Martinez, P., Corella, D., Roos, C., Tewhey, R., Rieder, M.J., Hall, J., Abecasis, G., *et al.* (2008). Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. Diabetes *57*, 3112-3121.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet *2*, e190.

Peterfy, M., Ben-Zeev, O., Mao, H.Z., Weissglas-Volkov, D., Aouizerat, B.E., Pullinger, C.R., Frost, P.H., Kane, J.P., Malloy, M.J., Reue, K., *et al.* (2007). Mutations in LMF1 cause combined lipase deficiency and severe hypertriglyceridemia. Nat Genet *39*, 1483-1487.

Plomin, R., Haworth, C.M., and Davis, O.S. (2009). Common disorders are quantitative traits. Nat Rev Genet *10*, 872-878.

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.

Priore Oliva, C., Pisciotta, L., Li Volti, G., Sambataro, M.P., Cantafora, A., Bellocchio, A., Catapano, A., Tarugi, P., Bertolini, S., and Calandra, S. (2005). Inherited apolipoprotein A-V deficiency in severe hypertriglyceridemia. Arterioscler Thromb Vasc Biol *25*, 411-417.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.

Qu, S., Perdomo, G., Su, D., D'Souza, F.M., Shachter, N.S., and Dong, H.H. (2007). Effects of apoA-V on HDL and VLDL metabolism in APOC3 transgenic mice. J Lipid Res *48*, 1476-1487.

Sampson, J.R. (2003). TSC1 and TSC2: genes that are mutated in the human genetic disorder tuberous sclerosis. Biochem Soc Trans *31*, 592-596.

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., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707-713.

Wang, J., Ban, M.R., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., and Hegele, R.A. (2008). APOA5 genetic variants are markers for classic hyperlipoproteinemia phenotypes and hypertriglyceridemia. Nat Clin Pract Cardiovasc Med *5*, 730-737.

Wang, J., Cao, H., Ban, M.R., Kennedy, B.A., Zhu, S., Anand, S., Yusuf, S., Pollex, R.L., and Hegele, R.A. (2007). Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). Arterioscler Thromb Vasc Biol *27*, 2450-2455.

Wang, J., and Hegele, R.A. (2007). Homozygous missense mutation (G56R) in glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1 (GPI-HBP1) in two siblings with fasting chylomicronemia (MIM 144650). Lipids Health Dis *6*, 23.

Yuan, G., Al-Shali, K.Z., and Hegele, R.A. (2007). Hypertriglyceridemia: its etiology, effects and treatment. CMAJ *176*, 1113-1120.

CHAPTER 3

RARE GENETIC VARIATION IN TRIGLYCERIDE-ASSOCIATED GENES ACCUMULATES IN PATIENTS WITH HYPERTRIGLYCERIDEMIA

The work in this chapter originates from material in the following publications: (1) **Johansen CT,** Wang J, Lanktree MB, *et al*. (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 42 (8): 684-7; and (2) **Johansen CT,** Wang J, McIntyre AD, et al. (2011). Excess of rare variants in non-GWAS candidate genes in patients with hypertriglyceridemia. *Circ Cardiovasc Genet*. Submitted.

3.1 Introduction

Genome-wide association studies (GWAS) have reliably identified novel and known loci associated with population-based plasma lipid concentrations (Aulchenko et al., 2009, Kathiresan et al., 2008, Kathiresan et al., 2009, Sabatti et al., 2009, Teslovich et al., 2010, Willer et al., 2008). Despite the robustness of these associations, the proportion of variability explained by GWAS-identified loci is relatively modest, <10% in most studies (Aulchenko et al., 2009, Kathiresan et al., 2009, Teslovich et al., 2010), and <20% in our study of hypertriglyceridemia (HTG) (Johansen et al., 2010). While vastly expanded GWAS sample sizes continue to reveal new associations, each newly associated variant has an incrementally smaller effect size and contributes only marginally to the cumulative variation of each lipid phenotype (Manolio et al., 2009). This suggests that GWAS of

population-based subjects may be reaching their limits to explain heritability of complex traits.

The source of 'missing heritability' in complex traits is a topic of intense speculation (Manolio et al., 2009). One hypothesis is that additional forms of genetic variation may explain some unattributed variation. Rare variants of individually large effect are suspected to account for a large proportion of such variation. Furthermore, gene-gene and gene-environment interactions involving both common and rare variants could also contribute to the unexplained variation. Another explanation is that GWAS-identified genes represent only a fraction of all genes that contribute to complex traits. For instance, important disease-associated genes may not contain common variation, or common variants may have minute effect sizes, that would cause such genes to elude detection even in epidemiologically-sized GWAS (Johansen et al., 2011a). Regardless, such non-GWAS loci may harbour rare genetic variants that contribute to disease susceptibility. For instance, several non-GWAS genes have been implicated in triglyceride (TG) metabolism by pre-genome-era biochemical studies of patients with Mendelian HTG syndromes and naturally-occurring or genetically-engineered mutants causing HTG in mice, including *LPL*, *APOC2*, *APOA5*, *GPIHBP1*, and *LMF1* (Hegele, 2009). Extensive re-sequencing of such genes and evaluation of rare variant accumulation could help identify the contribution of additional genes and variation to polygenic HTG.

High-throughput resequencing studies are effective at detecting rare variant accumulation in case-control studies of complex disease (Bansal et al., 2010). Using

missense-accumulation analysis in genes defined *a priori* as likely to contain rare variants, a burden of mutations can be quantified statistically in subjects with severe phenotypes (Bansal et al., 2010). However, clear hypotheses are necessary to focus analyses towards a subset of genes and variants with increased *a priori* probability of determining the disease phenotype. One hypothesis is that genomic loci containing common small effect variants identified by GWAS also harbour rare variants that are associated with disease (Manolio et al., 2009). While the mechanistic basis for the association between lipid traits and most common variants discovered in GWAS is still largely unknown, it remains possible that rare variants in GWAS-identified genes may contribute significantly to lipid phenotypes and may help to identify specific disease-associated genes.

Studying subjects at the extremes of a quantitative phenotype distribution has proven useful to identify functional rare variants (Cohen et al., 2004, Nejentsev et al., 2009, Plomin et al., 2009, Romeo et al., 2007, Wang et al., 2007). In primary HTG, resequencing of TG-modulating candidate genes has implicated both common and rare variants in HTG disease pathophysiology (Hegele et al., 2009, Wang et al., 2008, Wang et al., 2007). Thus, our objectives were to resequence coding regions of candidate genes in 1) loci reaching genome-wide significance by GWAS of HTG; and in 2) non-GWAS candidate genes established in TG metabolism by mouse models and Mendelian syndromes of HTG, to evaluate the burden of rare variants in HTG patients compared with controls. We also sought to determine whether rare variants contribute additionally to the explained variation in HTG case status.

3.2 Methods

3.2.1 Study subjects

This study was approved by the University of Western Ontario Institutional Review Board (protocol #07920E), and ethics boards at collaborating institutions. All subjects provided informed consent for blood sampling, DNA analysis, and collection of clinical, biochemical and other demographic data. All subjects in this study were unrelated and of self-declared European ancestry. In total, 437 HTG patients and 326 population-based controls were included in these studies. HTG patients were clinically diagnosed with Fredrickson hyperlipoproteinemia (HLP) phenotypes 2B (MIM 144250), 3 (MIM 107741), 4 (MIM 144600) or 5 (MIM 144650), ascertained primarily from a single tertiary referral lipid clinic in London, Canada. HTG patients had plasma TG concentrations >95th percentile (3.37 mmol/L). Population-based controls were ascertained through the Study of Health Assessment and Risk in Ethnic Groups (Anand et al., 2000), as previously described. Control subjects with fasting plasma TG concentration >2.3 mmol/L were excluded from analyses due to potentially undiagnosed HTG, but were otherwise not phenotypically-selected, preventing the selection of potentially protective variants that might be expected in super-normal controls. The majority of subjects are common between both resequencing studies, including 421 HTG patients and 324 controls. Biochemical analyses of lipoprotein traits were conducted separately in each cohort, as previously described (Anand et al., 2000, Hegele et al., 2009).

3.2.2 Gene selection

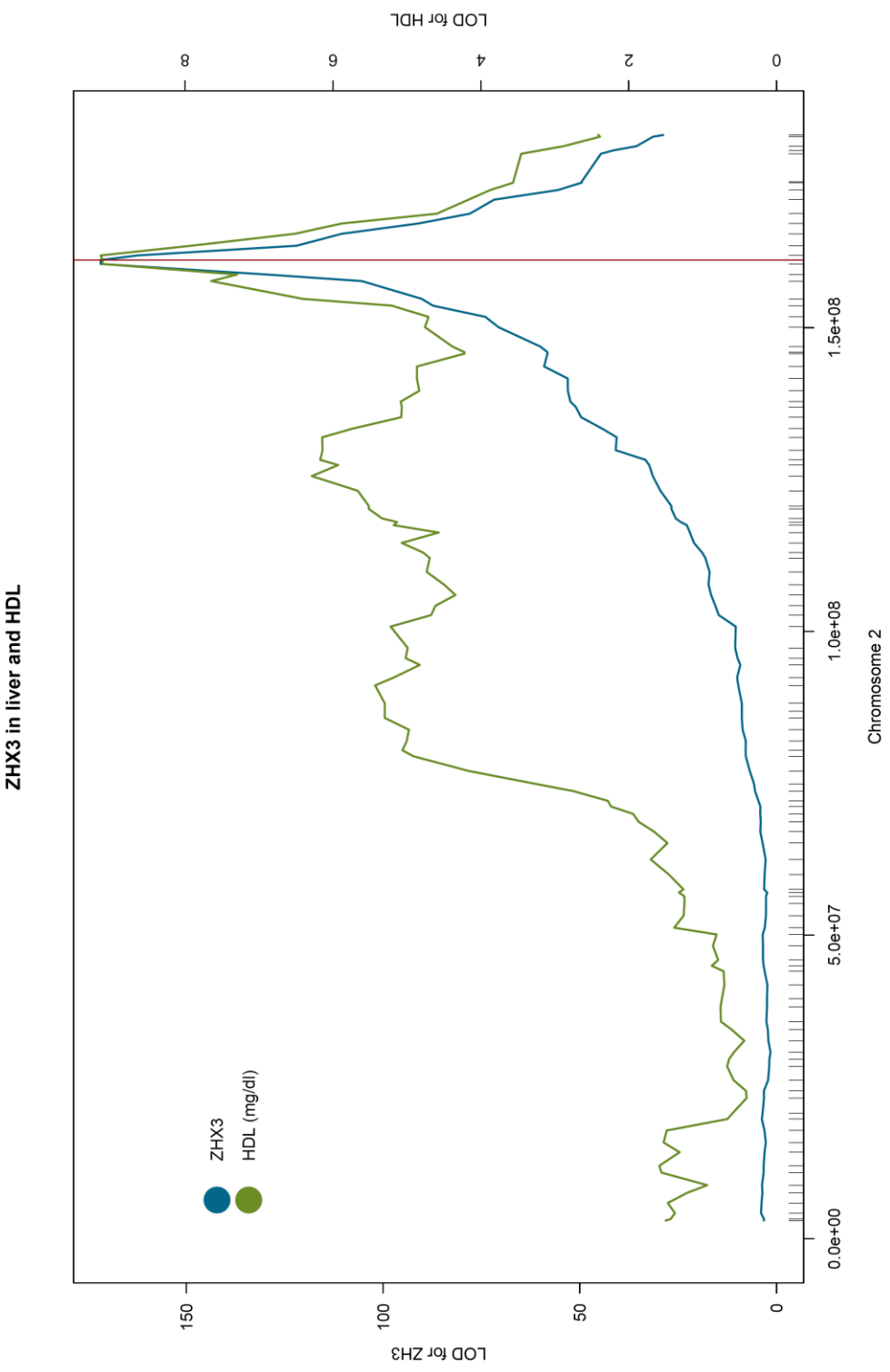
Two distinct hypotheses were tested in this chapter: (1) GWAS-identified genes would harbour an excess of rare variants in HTG patients versus controls; and (2) candidate genes identified by mouse and human models of HTG would harbour an excess of rare variants in HTG patients versus controls. GWAS-identified genes were selected from HTG-associated loci as the most probable gene to be involved in HTG pathophysiology (Hegele et al., 2009, Johansen et al., 2010), including *APOA5*, *GCKR*, *LPL*, and *APOB*. There were multiple genes at the *APOA5* locus with possible involvement in HTG pathophysiology, however only the *APOA5* gene was present in the HTG-associated block of linkage disequilibrium. Furthermore, homozygous rare variants in the *APOA5* gene have been established as causative for pediatric Mendelian HTG (Priore Oliva et al., 2005). The remaining GWAS-identified genes were the closest genes to each GWAS association signal.

Non-GWAS candidate genes were selected for sequencing based on mouse models and Mendelian syndromes of HTG. *APOC2* was selected based on knowledge that homozygous mutations in humans cause pediatric Mendelian HTG (Connelly et al., 1987). Heterozygous mutations in *APOC2* have also been identified in subjects with polygenic forms of HTG (Wang et al., 2007). *GPIHBP1* and *LMF1* were each selected as human genes orthologous to knockout mouse models characterized by HTG (Beigneux et al., 2007, Peterfy et al., 2007). Functionally characterized homozygous mutations in *GPIHBP1* (Beigneux et al., 2009, Franssen et al., 2010, Olivecrona et al., 2010) have also been reported in patients with pediatric Mendelian HTG.

Two novel HTG candidate genes were also sequenced in this study, including *CREB3L3* and *ZHX3*, both genes identified by novel mouse models of HTG. *Zhx3* was initially identified as a candidate gene for high-density lipoprotein (HDL) cholesterol concentration by a systems genetics analysis of HDL metabolism. It is a member of a family of three transcription factors, of which *Zhx2* has previously been shown to regulate both TG and cholesterol levels (Gargalovic et al., 2010). A linkage analysis of HDL cholesterol concentration and gene expression patterns in liver and adipose was used to identify candidate HDL-associated loci in the progeny of a cross between two inbred mouse strains differing vastly in their metabolism of HDL (CAST/EiJ and C57BL/6J) (**Figure 3.1**). *Zhx3* was directly under the linkage peak and exhibited a strong expression quantitative trait locus (eQTL) associated with HDL levels. Association of *Zhx3* with TG levels has also been observed in human genome-wide expression scans (Gargalovic et al., 2010).

Creb3l3, encoding the transcription factor Creb-h, was hypothesized to be an important regulator of nutrient and energy metabolism in mice (Lee et al., 2011). Recent studies demonstrated that Creb-h was involved in regulation of phosphoenolpyruvate carboxykinase 2 and glucose-6-phosphatase *in vivo*, and that *Creb3l3* expression is modulated by fasting and insulin resistant states (Lee et al., 2010). Studies aimed at elucidating the *in vivo* function of *Creb3l3* pursued the creation of a knockout mouse strain, which revealed a fasting HTG phenotype. The human orthologous gene *CREB3L3* was therefore suspected as a candidate gene for HTG susceptibility. A manuscript based on this work was recently published (Lee et al., 2011).

Figure 3.1. Linkage analysis of HDL cholesterol concentrations and gene expression changes in a cross between CAST/EiJ and C57BL/6J mice identifies *Zhx3* as a candidate gene for HDL metabolism. Subsequent knockout of *Zhx3* in mice revealed a fasting HTG phenotype, suggesting that the orthologous human gene *ZHX3* is a candidate for HTG. The x-axis shows increasing physical position near a linkage peak on chromosome 2; the y-axes represent increasing strengths of linkage between *Zhx3* transcript concentrations or HDL cholesterol concentrations with genetic markers using log of odds (LOD) scores (Figure courtesy of Jake Lusis).


ZHX3 in liver and HDL
LOD for ZH3
LOD for HDL
Chromosome 2
ZHX3
HDL (mg/dl)
0
50
100
150
2
4
6
8
0.0e+00
5.0e+07
1.0e+08
1.5e+08

3.2.3 Sequencing and rare variant definition

All genes were bidirectionally sequenced in individual samples using an ABI 3730 Automated DNA Sequencer and called using automated software (Applied Biosystems, Foster City, CA). Our main goal was the identification of protein coding variants, thus all exonic DNA was sequenced including ~50-bp of flanking intronic DNA; however only exons 26 and 29 of *APOB* (67.8% of protein sequence; corresponding to the receptor binding domain of apo B) were sequenced. Variants were manually curated, and confirmed by repeat analysis. Subjects missing sequencing data in any gene were removed from the study prior to analysis.

Variant frequency thresholds were calculated based on frequencies in healthy controls for GWAS genes, whereas variant frequency thresholds were calculated based on frequencies in the combined dataset for candidate genes. These methodological differences in rare variant definition reflect our evolving understanding of analyses in rare variants studies (as hypotheses were not tested simultaneously) that will be addressed in the overall discussion of this dissertation (Chapter 5). These differences in rare variant definition have virtually no difference on the results presented here (Johansen et al., 2011b). Rare variants were defined by frequency <1%, uncommon variants were defined by frequency ≥1% to <5%, and common variants were defined by frequency ≥5%. Nonsynonymous variants refer to sequence variations causing amino acid substitutions (missense variants) or introducing premature protein truncations (nonsense variants). Synonymous variants refer to DNA sequence variants that do not change protein coding

sequence (silent variants) or do not affect protein coding sequence (variants in noncoding variants).

3.2.4 Statistical analysis

Rare variant accumulation across all genes was compared between HTG patients and controls using Fisher's exact test, defining nominal significance as a 2-sided $P<0.05$, using SAS v.9.2 (Cary, NC, USA). Comparison of plasma TG concentrations between carriers and non-carriers of rare variants in nuclear families was conducted using the Wilcoxon signed-rank test for non-parametric data, defining nominal statistical significance as a 2-sided $P<0.05$. Mutation accumulation analyses compared either the number of observed rare alleles, or the number of rare variant carriers, between HTG patients and healthy controls, as indicated. Carriers were defined as having ≥ 1 rare variant. Enrichment for rare variants with high probability of function effects was attempted using two different strategies, where indicated: (1) filtration enriching for 'exclusive' variants not identified in controls; and (2) filtration enriching for variants predicted *in silico* to have deleterious functional effect. Exclusive variants were defined as rare variants found exclusively in HTG patients or controls (not both), deliberately excluding variants previously reported without demonstrated functional compromise. Filtration of rare variants using *in silico* prediction of deleterious effect was conducted using established software, including Polyphen-2 (Polymorphism Phenotyping v2) (Adzhubei et al., 2010) and SIFT (Sorting Intolerant from Tolerant) (Kumar et al., 2009) for variants resulting in amino acid substitutions at the protein level. Nonsense mutations were ranked as analogous to having 2 predictions of deleterious effect.

Rare variant accumulation was assessed across individual genes using Fisher's exact test and three additional permutation-based analyses of genetic burden (general case-control comparison, C-alpha, and the variable threshold approach) as implemented in PlinkSeq (http://atgu.mgh.harvard.edu/plinkseq). No single proposed test statistic has been proposed for rare variant accumulation studies; therefore we chose 3 distinct independent analyses for further *post hoc* evaluation of our dataset. The general burden test was chosen as a method that was parallel to Fisher's exact test, and could compare the number of alternate alleles between HTG patients and healthy controls using a permutation-based approach. The C-alpha test statistic was chosen because it was a recently proposed analysis to measure genetic burden across a locus considering variants with a mixture of protective and deleterious directions of effect (Neale et al., 2011). The variable threshold analysis was chosen as an analysis strategy that increases the weight of very rare variants, such as those likely to be functionally significant (Price et al., 2010). Empirical P-values were calculated in parallel for all statistical tests using 10,000 or 1,000,000 permutations, as indicated.

Summary statistics are displayed as mean ± standard deviation. Identity-by-state calculations, multi-dimensional scaling and association testing were conducted as implemented in PLINK (Purcell et al., 2007) and PlinkSeq. Identity-by-state calculations and multidimensional scaling were performed using freely available HapMap CEU (European ancestry), CHB/JPT (Asian ancestry) and YRI (African ancestry) subjects in order to demonstrate the relative ancestry of study participants. Association was tested between common variants and HTG using an additive multiple logistic regression model,

adjusted for clinical covariates age, sex, body mass index (BMI) and type 2 diabetes (T2D) where indicated. Statistical significance for association was defined as a nominal $P<0.05$. Statistical significance for association was defined as a nominal $P<0.05$. All variants were in Hardy-Weinberg equilibrium ($P>0.05$). Explained variation calculations were performed using multiple logistic regression without a predefined modeling strategy. Case or control status was modeled as the dependent variable, entering clinical variables (age, sex, BMI, and T2D), and common variants at 7 HTG-associated loci (0, 1 or 2 TG-raising alleles) as independent variables. Subjects included in the explained variation analysis were common to both GWAS (Chapter 2) and resequencing cohorts, including 346 HTG patients and 205 healthy controls. Independent variables included clinical covariates age, sex, BMI and T2D status as either continuous or discrete variables, common variants as continuous variables of HTG risk associated alleles at each of the 7 HTG-associated loci, and rare variants as a continuous variable including the number of rare variants carried by each subject.

3.3 Results

3.3.1. Study subjects

Baseline clinical attributes of the 437 HTG patients and 326 population-based controls are shown in **Table 3.1**. Generally, clinical attributes of HTG patients were less favourable than controls. HTG patients had increased BMI, increased plasma total cholesterol, decreased plasma HDL cholesterol and increased prevalence of T2D. Plasma TG concentration was most greatly increased, with a mean of 12.8 mmol/L in HTG patients versus 1.2 mmol/L in healthy controls. In subjects with whole genome genotyping data available, including 346 HTG patients and 205 healthy controls, we confirmed their European ancestry using identity-by-state calculations and multidimensional scaling (**Figure 3.2**).

3.3.2. Rare variant accumulation in GWAS-identified HTG-associated genes

Our first hypothesis was that rare variants related to HTG disease causation would accumulate in genes containing common variation associated with HTG by GWAS. The protein-coding sequences of *APOA5*, *GCKR*, *LPL* and exons 26 and 29 (67.8%) of *APOB* were resequenced in individual subjects as regions most likely to harbour protein-compromising mutations. Across the 4 genes, 80 distinct heterozygous rare variants were identified (**Table 3.2; Figure 3.3**). Only 29/80 (36%) rare variants were annotated in public databases such as dbSNP or 1000 Genomes Project. The strength of association between HTG and GWAS-associated loci did not predict the mutation accumulation in

Table 3.1. Baseline clinical attributes of study sample.

Characteristic	HTG	Controls
Number	437	326
Female	33.0%	54.6%
Diabetes	27.9%	1.9%
Age (years)	51.4 ± 13.1	49.7 ± 15.1
Body mass index (kg/m^2)	30.0 ± 4.9	26.7 ± 4.5
Plasma total cholesterol (mmol/L)	8.4 ± 3.9	4.9 ± 0.8
Plasma HDL cholesterol (mmol/L)	0.9 ± 0.3	1.3 ± 0.4
Plasma LDL cholesterol (mmol/L)	-	3.2 ± 0.9
Plasma TG (mmol/L)	12.8 ± 16.4	1.2 ± 0.4

The majority of study subjects were common among both sequencing projects, including 421 HTG patients and 324 healthy controls. Values are shown for the GWAS cohort, but are virtually identical between cohorts. Values are mean ± standard deviation. LDL cholesterol is not accurately calculated using the Friedewald equation for HTG patients when plasma TG concentration exceeds 4.5 mmol/L. GWAS, genome-wide association study; HDL, high-density lipoprotein; HTG, hypertriglyceridemia; LDL, low-density lipoprotein; TG, triglyceride.

Figure 3.2. Confirmation of European ancestry within the study sample. Identity-by-state and multidimensional scaling analyses were used to illustrate the relative genetic similarity between study subjects, represented by closer relative distance on the plot above. HTG cases (diamonds) and healthy controls (squares) are all of European ancestry, assessed by clustering with HapMap CEU subjects of European ancestry (green triangles), and distance from HapMap YRI subjects of African ancestry (purple X's) and CHB/JPT subject of East Asian ancestry (blue X's). Rare variant carriers are shown as orange symbols to illustrate that they do not deviate from European ancestry. This analysis only included subjects for whom both sequencing and whole genome genotyping data were available, which included 346 HTG patients and 205 healthy controls.

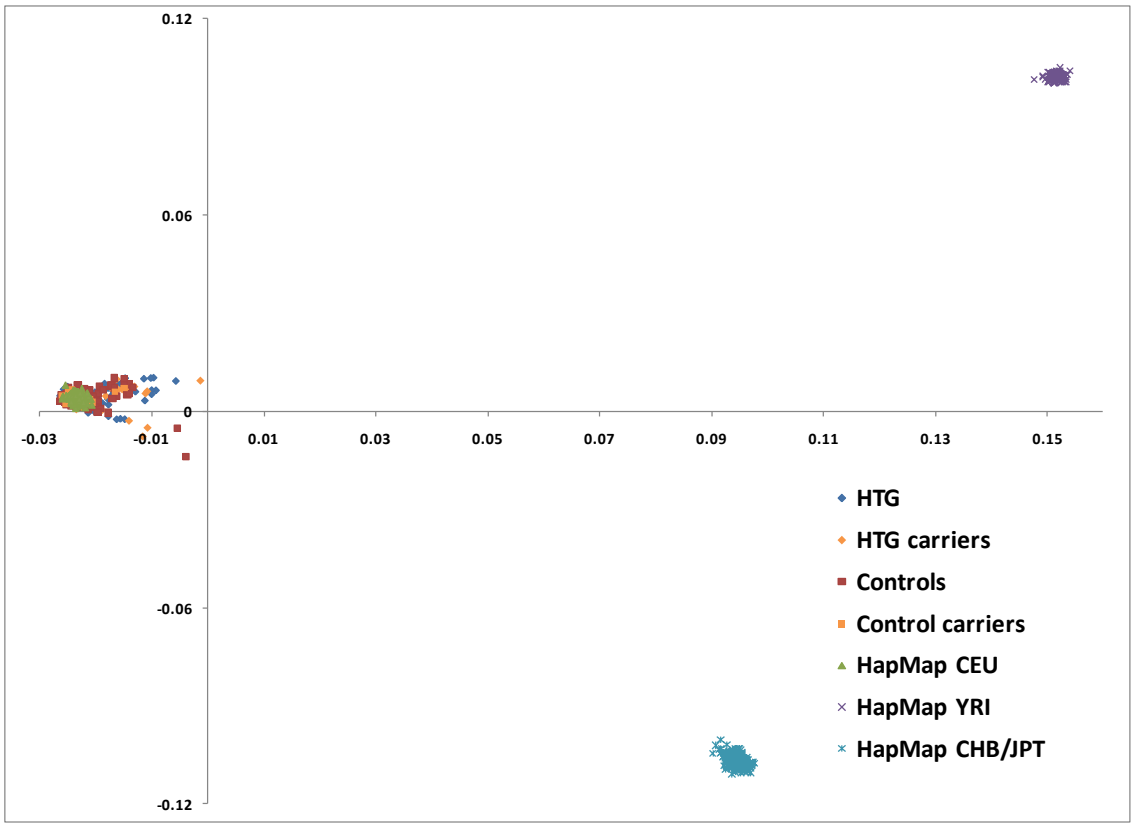

0.12
0.06
0
-0.06
-0.12
-0.03
-0.01
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
HTG
HTG carriers
Controls
Control carriers
HapMap CEU
HapMap YRI
HapMap CHB/JPT

Table 3.2. Annotation of rare variants found in GWAS-identified genes in HTG patients and healthy controls.

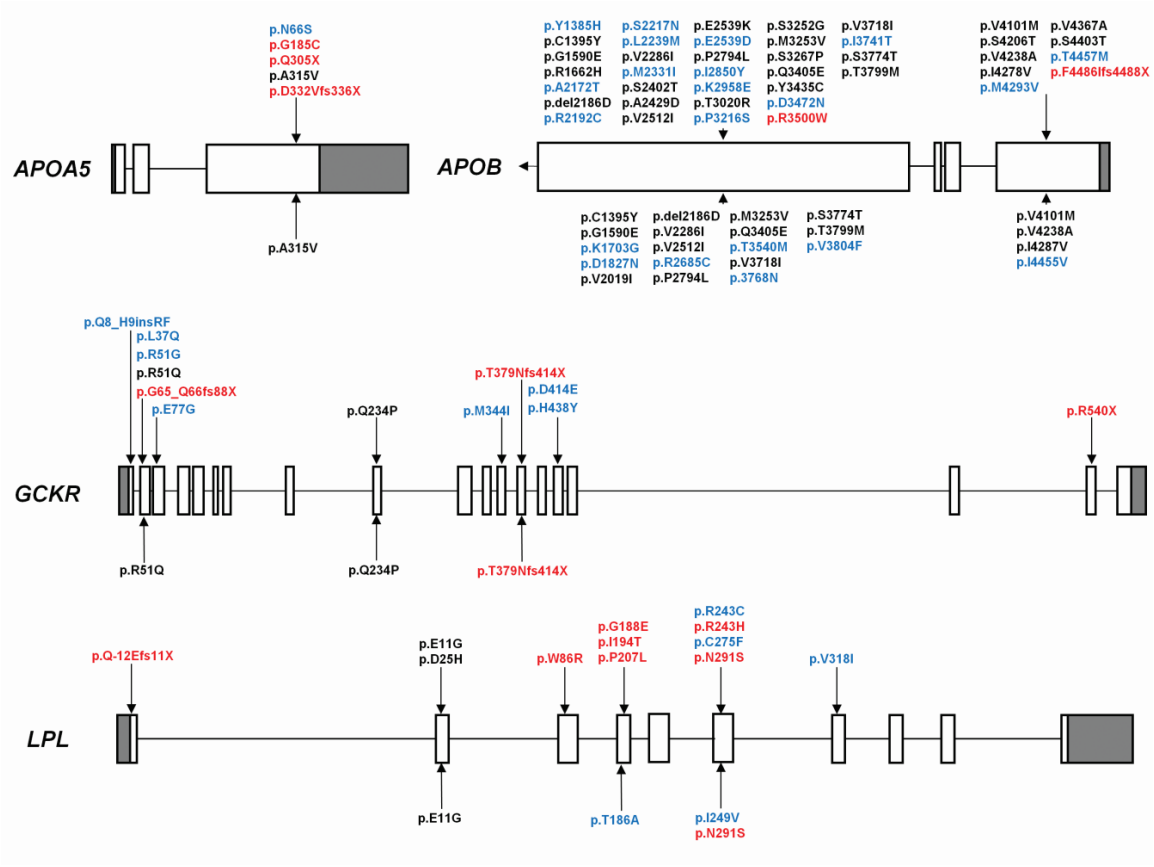
Gene	cDNA	Substitution	HTG (n=437)	Control (n=326)	Frequency	SIFT*	Polyphen2*	GIK/dbSNP	Published dysfunction
APOA5	c.197A>G	N66S	1	0	0.00065	Tolerated	Damaging	-/-	
	c.553G>T	G185C	1	0	0.00065	Damaging	Damaging	+/+	LPL activation reduced by 23% (Dorfmeister et al., 2008)
	c.913C>T	Q305X	1	0	0.00065	Damaging	Damaging	-/-	
	c.944C>T	A315V	1	1	0.0013	Damaging	Benign	-/-	
	c.995insT	D332Vfs336X	1	0	0.00065	Damaging	Damaging	-/-	
GCKR	c.25insRF	Q8_H9insRF	1	0	0.00065	Damaging	Damaging	-/-	
	c.110T>A	L37Q	1	0	0.00065	Damaging	Benign	-/-	
	c.151C>G	R51G	1	0	0.00065	Tolerated	Damaging	-/-	
	c.152G>A	R51Q	1	1	0.0013	Tolerated	Benign	-/-	
	c.196ins23nt	G65_Q66fs88X	1	0	0.00065	Damaging	Damaging	-/-	
	c.230A>G	E77G	1	0	0.00065	Damaging	Damaging	+/+	
	c.701A>C	Q234P	6	3	0.0059	Tolerated	Damaging	-/-	
	c.1032G>T	M344I	1	0	0.00065	Tolerated	Benign	-/-	
	c.1136insA	T379Nfs414X	4	1	0.0033	Damaging	Damaging	-/-	
	c.1242C>A	D414E	1	0	0.00065	Tolerated	Damaging	-/-	
	c.1312C>T	H438Y	1	0	0.00065	Tolerated	Damaging	-/-	
	c.1618C>T	R540X	1	0	0.00065	Damaging	Damaging	-/-	
LPL	c.46delCA	Q-12Efs11X	1	0	0.00065	Damaging	Damaging	-/-	
	c.113A>G	p.E11G	1	1	0.0013	Damaging	Damaging	-/-	
	c.154G>C	D25H	1	0	0.00065	Damaging	Damaging	-/-	
	c.337T>C	W86R	1	0	0.00065	Damaging	Damaging	-/+	LPL activity <3% (Ishimura-Oka et al., 1992)
	c.637A>G	p.T186A	0	1	0.00065	Damaging	Damaging	-/-	
	c.644G>A	p.G188E	10	0	0.0065	Tolerated	Damaging	-/+	LPL activity < 1% (Emi et al., 1990)
	c.662T>C	I194T	1	0	0.00065	Damaging	Damaging	-/+	LPL activity < 1% (Henderson et al., 1991)

	c.701C>T	p.P207L	1	0	0.00065	Damaging	Damaging	-/+	LPL activity < 1% (Ma et al., 1991)
	c.808C>T	p.R243C	1	0	0.00065	Damaging	Damaging	-/+	
	c.809G>A	R243H	1	0	0.00065	Damaging	Damaging	-/+	LPL activity < 1% (Gotoda et al., 1991)
	c.826A>G	p.I249V	0	1	0.00065	Tolerated	Benign	-/-	
	c.905G>T	p.C275F	1	0	0.00065	Damaging	Damaging	-/-	
	c.953A>G	p.N291S	24	5	0.019	Tolerated	Benign	+/+	LPL activity ~60% (Busca et al., 1995)
	c.1033G>A	p.V318I	1	0	0.00065	Tolerated	Benign	-/-	
APOB	c.4234T>C	Y1385H	1	0	0.00065	Tolerated	Damaging	-/-	
	c.4265A>G	Y1395C	1	1	0.0013	Tolerated	Benign	+/+	
	c.4850G>A	G1590E	1	1	0.0013	Damaging	Damaging	-/-	
	c.5066G>A	R1662H	2	0	0.0013	Damaging	Benign	+/-	
	c.5189A>C	K1703T	0	1	0.00065	Tolerated	Damaging	-/-	
	c.5560G>A	D1827N	0	1	0.00065	Damaging	Damaging	-/-	
	c.6136G>A	V2019I	0	1	0.00065	Tolerated	Damaging	-/+	
	c.6595G>A	A2172T	1	0	0.00065	Tolerated	Benign	-/-	
	c.6639delTG	D2186Efs	3	4	0.0046	Damaging	Damaging	-/-	
	c.6655C>T	R2192C	1	0	0.00065	Damaging	Damaging	-/-	
	c.6731G>A	S2217N	1	0	0.00065	Tolerated	Benign	-/-	
	c.6796C>A	L2239M	1	0	0.00065	Damaging	Benign	-/-	
	c.6937A>G	I2286V	2	1	0.0020	Tolerated	Benign	+/+	
	c.7074G>A	M2331I	1	0	0.00065	Damaging	Damaging	-/-	
	c.7285T>A	S2402T	1	0	0.00065	Tolerated	Benign	+/+	
	c.7367C>A	A2429D	1	0	0.00065	Damaging	Benign	+/+	
	c.7615G>A	V2512I	2	3	0.0033	Tolerated	Benign	-/-	
	c.7696G>A	E2539K	5	0	0.0033	Tolerated	Benign	-/+	
	c.7698G>C	E2539D	1	0	0.00065	Tolerated	Benign	-/-	
	c.8134C>T	R2685C	0	1	0.00065	Tolerated	Damaging	-/-	
	c.8462C>T	P2794L	6	2	0.0052	Tolerated	Damaging	+/+	
	c.8550T>G	I2823M	1	0	0.00065	Damaging	Damaging	-/-	
	c.8953A>G	K2958E	1	0	0.00065	Tolerated	Benign	-/-	

	c.9140C>G	T3020R	1	0	0.00065	Tolerated	Benign	+/+	
	c.9727C>T	P3216S	1	0	0.00065	Tolerated	Benign	-/-	
	c.9835A>G	S3252G	4	0	0.0026	Damaging	Damaging	+/+	
	c.9838A>G	M3253V	1	1	0.0013	Damaging	Damaging	-/-	
	c.9880T>C	S3267P	3	0	0.0020	Damaging	Damaging	+/+	
	c.10294G>C	Q3405E	5	3	0.0052	Not scored	Benign	-/-	
	c.10385A>G	Y3435C	1	0	0.00065	Damaging	Damaging	-/-	
	c.10495G>A	D3472N	1	0	0.00065	Tolerated	Benign	-/-	
	c.10579C>T	R3500W	1	0	0.00065	Damaging	Damaging	-/-	Familial defective apolipoprotein B-100 (Gaffney et al., 1995)
	c.10700C>T	T3540M	0	1	0.00065	Damaging	Benign	-/-	
	c.11233G>A	V3718I	1	1	0.0013	Tolerated	Benign	-/-	
	c.11303T>C	I3741T	1	0	0.00065	Damaging	Damaging	-/-	
	c.11383G>A	D3768N	0	1	0.00065	Tolerated	Damaging	-/-	
	c.11401T>A	S3774T	3	3	0.0039	Tolerated	Benign	+/+	
	c.11477C>T	T3799M	1	1	0.0013	Damaging	Damaging	+/+	
	c.11491G>T	V3804F	0	1	0.00065	Tolerated	Benign	-/-	
	c.12382G>A	V4101M	3	3	0.0039	Tolerated	Benign	+/+	
	c.12697T>A	S4206T	1	0	0.00065	Tolerated	Benign	+/+	
	c.12794T>C	V4238A	9	3	0.0079	Damaging	Benign	+/+	
	c.12940A>G	I4287V	7	4	0.0072	Tolerated	Benign	+/+	
	c.12958A>G	M4293V	1	0	0.00065	Tolerated	Benign	-/-	
	c.13181T>C	V4367A	1	0	0.00065	Damaging	Benign	-/+	
	c.13288T>A	S4403T	1	0	0.00065	Tolerated	Benign	-/+	
	c.13444A>G	I4455V	0	1	0.00065	Tolerated	Benign	-/-	
	c.13451C>T	T4457M	6	0	0.0039	Tolerated	Damaging	+/+	
	c.13537insA	F4486Ifs4488X	1	0	0.00065	Damaging	Damaging	-/-	

*Polyphen-2 was used to predict the deleterious nature of non-synonymous variants (Adzhubei et al., 2010); other mutation types are indicated. APOA5, apolipoprotein A-V; APOB, apolipoprotein B; dbSNP, Database of Single Nucleotide Polymorphism; G1K, 1000 Genomes Project; GCKR, glucokinase regulatory protein; GWAS, genome-wide association study; HTG, hypertriglyceridemia; LPL, lipoprotein lipase; Polyphen, Polymorphism Phenotyping; SIFT, Sorting Intolerant from Tolerant.

Figure 3.3. Rare variants identified by resequencing GWAS-identified genes in HTG patients and healthy controls. Variants above gene maps were identified in HTG patients and variants below gene maps were identified in controls. Rare variants are coloured according to their identification in control subjects or previous identification in subjects of unknown clinical status (black), exclusivity to HTG patients or controls (blue), or proven biological dysfunction or truncation (red). Nomenclature refers to functional protein sequences. Only exons 26 and 29 were resequenced in *APOB*. Gene maps are roughly to scale, although differ in scale between genes. GWAS, genome-wide association study; HTG, hypertriglyceridemia; TG, triglyceride.


APOA5
p.N66S
p.G185C
p.Q305X
p.A315V
p.D332Vfs336X
p.A315V
APOB
p.Y1385H p.S2217N p.E2539K p.S3252G p.V3718I
p.C1395Y p.L2239M p.E2539D p.M3253V p.I3741T
p.G1590E p.V2286I p.P2794L p.S3267P p.S3774T
p.R1662H p.M2331I p.I2850Y p.Q3405E p.T3799M
p.A2172T p.S2402T p.K2958E p.Y3435C
p.del2186D p.A2429D p.T3020R p.D3472N
p.R2192C p.V2512I p.P3216S p.R3500W
p.V4101M p.V4367A
p.S4206T p.S4403T
p.V4238A p.T4457M
p.I4278V p.F4486Ifs4488X
p.M4293V
p.C1395Y p.del2186D p.M3253V p.S3774T
p.G1590E p.V2286I p.Q3405E p.T3799M
p.K1703G p.V2512I p.T3540M p.V3804F
p.D1827N p.R2685C p.V3718I
p.V2019I p.P2794L p.3768N
p.V4101M
p.V4238A
p.I4287V
p.I4455V
GCKR
p.Q8_H9insRF
p.L37Q
p.R51G
p.R51Q
p.G65_Q66fs88X
p.E77G
p.Q234P
p.T379Nfs414X
p.D414E
p.M344I
p.H438Y
p.R540X
p.R51Q
p.Q234P
p.T379Nfs414X
LPL
p.Q-12Efs11X
p.E11G
p.D25H
p.W86R
p.G188E
p.I194T
p.P207L
p.R243C
p.R243H
p.C275F
p.N291S
p.V318I
p.E11G
p.T186A
p.I249V
p.N291S

resequenced genes. *LPL* harboured the largest relative proportion of rare variants, followed by *GCKR*, *APOB* and *APOA5*, with 30.9, 10.7, 9.3, and 4.5 rare variants per kilobase of coding sequence in HTG patients, whereas the same genes harboured 5.6, 2.7, 4.3 and 0.9 rare variants per kilobase of coding sequence in controls. Among these variants, 20 were identified in both HTG patients and controls, 50 were identified exclusively in HTG patients, and 10 exclusively in controls. The vast majority of rare variants were singletons (54/80), compared to doubletons (9/80), tripletons (2/80) or variants with other frequencies (15/80).

Overall, a significant accumulation of rare variants was identified in HTG patients (**Table 3.3**): 154 rare variant occurrences were observed among 437 HTG patients compared to 53 rare variant occurrences observed among 326 healthy controls, corresponding to a 2.4-fold (95% CI: 1.72-3.34) increase in rare variants in HTG patients ($P=6.2X10^{-8}$). There was also a significant increase in rare variant carriers in HTG patients versus controls: 28.1% of HTG patients and 15.3% in controls carried ≥1 rare variant, corresponding to a 2.2-fold (95% CI: 1.50-3.12) increased in carrier frequency ($P=2.6X10^{-5}$). Although the majority of subjects carried only 1 rare variant, subjects with multiple rare variants were also significantly over-represented in HTG patients (6.6% HTG carriers versus 0.9% control carriers; $P=3.7X10^{-5}$).

A more restricted analysis enriching for rare variants with potentially functional effects based on exclusivity in either HTG patients or controls, similarly revealed a

Table 3.3. Rare variant accumulation in GWAS-identified genes in HTG patients and healthy controls.

		All mutations		Missense/Indels		Nonsense	
		HTG	Controls	HTG	Controls	HTG	Controls
	Total alleles	876	654	876	654	876	654
All Variants	*APOA5*	5	1	3	1	2	0
	GCKR	20	5	14	4	6	1
	LPL	44	8	43	8	1	0
	APOB	85	39	84	39	1	0
	Total	154	53	144	52	10	1
		P = 6.2 X 10^{-8}		**P = 6.3 X 10^{-7}**		**P = 0.030**	
Exclusive Variants	*APOA5*	4	0	3	0	2	0
	GCKR	9	0	7	0	2	0
	LPL	19	2	18	2	1	0
	APOB	15	7	14	7	1	0
	Total	47	9	41	9	6	0
		P = 2.4 X 10^{-5}		**P = 2.3 X 10^{-4}**		**P = 0.041**	

GWAS, genome-wide association study; HTG, hypertriglyceridemia. Exclusive variants refer to rare variants found exclusively in HTG cases or low triglyceride controls; previously reported variants without characterized functional compromise are deliberately excluded. Fisher's exact test was used to calculate the significance of rare variant accumulation in HTG patients, defining nominal statistical significance as a two-sided $P < 0.05$. Mutation counts and annotations are found in Table 3.2.

significant burden of 47 rare variant occurrences in HTG patients compared to 9 rare variant occurrences in controls **(Table 3.3)**. This corresponded to a 4.1-fold (95% CI: 1.98-8.35) increase in exclusive variants in HTG patients ($P=2.4X10^{-5}$). Similarly, there was a significantly increased carrier frequency of exclusive variants, as 10.3% of HTG patients and 2.8% of controls carried ≥1 exclusive variant, corresponding to a 4.0-fold (95% CI: 1.95-8.40) increase in HTG patients versus healthy controls ($P=4.4X10^{-5}$). HTG carriers' fasting plasma TG concentrations ranged from 3.10–88.5 mmol/L, whereas control carriers' fasting plasma TG concentrations ranged from 0.45–1.93 mmol/L. No discernable patterns were observed between such attributes as the gene, mutation type or mutation position with plasma TG concentration or HTG phenotype.

Finally, we conducted a *post hoc* analysis to determine whether rare variant accumulation could be detected in any specific genes. Using Fisher's exact test, a significant excess of rare variants was detected in HTG patient versus healthy controls in *LPL*, *GCKR*, and *APOB*. *LPL* had a 4.3-fold enrichment (95% CI: 2.00-9.13; $P=2.47X10^{-5}$), *GCKR* had a 3.0-fold enrichment (95% CI: 1.13-8.12; P=0.024), *APOB* had a 1.7-fold enrichment (95% CI: 1.14-2.51; $P=7.99X10^{-3}$). We could not detect a statistically significant increase in rare variants in *APOA5* (OR=3.8; P=0.25), although the relative magnitude of this accumulation was second only to *LPL* in comparison to other genes. We repeated the analysis above considering carriers versus non-carriers, confirming statistical association between rare variant accumulation and HTG in only *LPL* and *GCKR*. *LPL* had a 4.3-fold enrichment (95% CI: 2.00-9.37; $P=2.9X10^{-5}$), and *GCKR* had a 2.9-fold enrichment (95% CI: 1.08-7.91; P=0.035). However, we were no longer able to

detect an accumulation of rare variants in *APOB* (OR=1.56; P=0.054), likely because the magnitude of the accumulation was attenuated by carriers of multiple *APOB* rare variants.

We repeated the gene-centric association analysis using three recently proposed permutation-based statistical analyses designed for rare variant accumulation studies. We employed 3 different analyses to differently assess rare variant accumulation in individual genes *post hoc*, including a general test of burden, a recently proposed variance-based analysis using the C-alpha test statistic (Neale et al., 2011), and a recently proposed variable-threshold analysis (Price et al., 2010), all of which employed permutation-based analyses to obtain empirical P-values for variant accumulation. The three statistical analyses used to assess genetic burden varied in their ability to detect rare variant accumulation in HTG patients versus healthy controls in various genes **(Table 3.4)**. A significant accumulation of rare variants was confirmed empirically across all loci by the general burden test ($P=7.0X10^{-7}$), the C-alpha test statistic (P=0.002), and the variable threshold approach ($P=4.1X10^{-6}$). Among individual genes, rare variant accumulation was detected in *LPL*, *APOB,* and *GCKR* using the general burden test ($P=1.1X10^{-5}$, P=0.0092, and P=0.013, respectively) and the variable threshold approach ($P=1.5X10^{-5}$, P=0.023, and P=0.014, respectively). These analyses did not detect rare variant accumulation in *APOA5* using the burden test (P=0.11) or the C-alpha test (P=0.97), although a trend toward significance was observed using the variable threshold approach (P=0.071). C-alpha could only detect rare variant accumulation in *LPL* (P=0.0014).

Table 3.4. Rare variant accumulation among individual genes measured using multiple *post hoc* analyses of genetic burden.

		Multi-locus test	*APOA5*	*GCKR*	*LPL*	*APOB*
Rare (<1%)	Burden	$7.0X10^{-7}$	0.11	0.013	$1.1X10^{-5}$	0.0092
	C-alpha	0.0022	0.97	0.75	0.0014	0.43
	Variable threshold	$4.1X10^{-6}$	0.071	0.014	$1.5X10^{-5}$	0.023

Multi-locus test includes all rare variants identified across 4 genes in a single burden analysis. Empirical P-values were calculated for each test statistic using 10,000,000 permutations. All statistical analyses were executed as implemented in PlinkSeq.

3.3.3. Rare variant accumulation in non-GWAS candidate genes

Our second hypothesis was that rare variants related to HTG disease causation would accumulate in candidate genes identified by mouse models and Mendelian syndromes of severe HTG. Re-sequencing of HTG-associated candidate genes in 437 HTG patients and 326 healthy controls identified 49 distinct missense or nonsense variants in HTG patients and healthy controls **(Table 3.5)**, including 43 heterozygous rare variants (minor allele frequency [MAF] <1%), 3 heterozygous uncommon variants (MAF 1-5%) and 3 common variants (MAF >5%). Only 6/43 (14%) rare variants were annotated in public databases such as dbSNP or 1000 Genomes Project, whereas all common and uncommon variants had been previously identified. *APOC2* had the largest number of rare variants relative to size, followed by *LMF1*, *CREB3L3*, *GPIHBP1*, and *ZHX3*, harbouring respectively 19.8, 9.4, 8.0, 7.2 and 5.6 variants per kilobase sequenced in HTG patients, and 0, 2.9, 0.7, 3.6 and 2.8 variants per kilobase sequenced in controls. Among these variants, 6 were identified in both HTG patients and controls, 30 were identified exclusively in HTG patients, and 7 exclusively in controls **(Figure 3.4)**. The vast majority of rare variants were again singletons (30/43), as opposed to doubletons (6/43), tripletons (3/43) or variants with other frequencies (4/43).

Rare nonsynonymous variants were found to be in large excess among HTG patients **(Table 3.6)**. In total, we found 53 rare variant occurrences among 437 HTG patients and 16 rare variant occurrences among 326 controls, corresponding to a 2.3-fold increase in variants found in HTG patients (P=0.0050). Only 6 HTG patients and 1 healthy control were carriers of 2 rare variants, which had virtually no effect on the

Table 3.5. Annotation of variants identified in candidate HTG genes in HTG patients and healthy controls.

Gene	cDNA	Substitution	HTG (n=437)	Control (n=326)	Frequency	SIFT*	Polyphen2*	G1K/dbSNP
GPIHBP1	c.106 C>T	L11F	0	1	0.00066	Damaging	Benign	-/-
	c.116 G>T	C14F	71	54	0.08191	Tolerated	Probably	+/+
	c.117 C>A	C14X	1	0	0.00066	Damaging	Probably	-/-
	c.241 G>C	G56R	1	0	0.00066	Damaging	Probably	-/-
	c.352 A>G	I93V	0	1	0.00066	Tolerated	Benign	-/-
	c.506 C>T	S144F	2	0	0.00131	Tolerated	Probably	+/+
LMF1	c.111 G>A	G36D	68	73	0.09240	Tolerated	Possibly	+/+
	c.306 G>C	R101T	1	0	0.00066	Tolerated	Benign	+/-
	c.368 A>G	M122V	1	0	0.00066	Tolerated	Possibly	-/-
	c.687 G>A	G228E	1	2	0.00197	Damaging	Benign	-/-
	c.980 G>A	A326T	1	1	0.00131	Tolerated	Benign	-/-
	c.1056 G>A	R351Q	2	2	0.00262	Tolerated	Benign	-/-
	c.1064 C>T	R354W	13	14	0.01769	Tolerated	Possibly	+/-
	c.1095 G>A	R364Q	15	8	0.01507	Tolerated	Probably	+/+
	c.1268 C>T	Q422X	1	0	0.00066	NA	NA	-/-
	c.1296 C>A	A431D	1	0	0.00066	Tolerated	Benign	+/+
	c.1355 C>T	R451W	4	0	0.00262	Tolerated	Probably	+/-
	c.1409 G>A	A469T	1	0	0.00066	Tolerated	Probably	-/-
	c.1569 G>T	S522I	1	0	0.00066	Damaging	Possibly	-/-
	c.1597 G>C	E531D	1	0	0.00066	Tolerated	Benign	-/-
	c.1598 G>A	G532S	1	0	0.00066	Damaging	Probably	-/-
	c.1689 C>G	P562R	15	3	0.01180	Tolerated	Probably	+/+
CREB3L3	c.254 G>A	W46X	1	0	0.00066	NA	NA	-/-
	c.429 G>A	G105R	2	1	0.00197	Tolerated	Possibly	+/+
	c.613C>T	P166L	2	0	0.00131	Damaging	Benign	-/-
	c.654 G>A	V180M	1	0	0.00066	Damaging	Benign	-/-

	c.660 G>A	D182N	1	0	0.00066	Tolerated	Possibly	-/-
	c.834 G>A	E240K	2	0	0.00131	Damaging	Probably	-/-
	c.249-254 AAAAAA> GAAAAAT	K245Efs374X	2	0	0.00131	NA	NA	-/-
APOC2	c.225 A>C	K19T	4	0	0.00262	Damaging	Benign	+/+
	c.377 C>T	Q70X	1	0	0.00066	NA	NA	-/-
	c.378 insC	Q70Pfs97X	1	0	0.00066	NA	NA	-/-
ZHX3	c.535 G>A	A52T	2	0	0.00131	Tolerated	Benign	-/-
	c.583 C>T	R69W	1	0	0.00066	Damaging	Probably	-/-
	c.611 C>T	S78F	1	2	0.00197	Tolerated	Benign	-/-
	c.1125 C>A	N249K	1	0	0.00066	Tolerated	Benign	-/-
	c.1141 G>A	G255R	0	1	0.00066	Damaging	Probably	-/-
	C1307 A>G	N310S	153	123	0.18087	Tolerated	Possibly	+/+
	c.1366 A>G	T330A	1	0	0.00066	Damaging	Probably	-/-
	c.1500 G>C	K374N	0	1	0.00066	Damaging	Probably	-/-
	c.1516 A>G	I380V	0	1	0.00066	Tolerated	Probably	-/-
	c.1922 G>A	R515Q	0	1	0.00066	Tolerated	Probably	-/-
	c.1945 G>A	E523K	1	0	0.00066	Damaging	Probably	-/-
	c.2138 G>T	C587F	1	0	0.00066	Tolerated	Probably	-/-
	c.2240 delC	P621Lfs645X	1	0	0.00066	NA	NA	-/-
	c.2389 G>A	E671K	5	1	0.00393	Tolerated	Benign	-/+
	c.2528 T>C	L717S	1	0	0.00066	Tolerated	Benign	-/-
	c.2918 G>A	R847Q	1	0	0.00066	Tolerated	Probably	-/-
	c.2948 A>C	H857P	0	1	0.00066	Damaging	Probably	-/-

NA, not applicable. *Prediction of deleterious effects was conducted using SIFT (http://sift.jcvi.org/) or Polyphen2 (http://genetics.bwh.harvard.edu/pph2/). *Polyphen-2 was used to predict the deleterious nature of non-synonymous variants (Adzhubei et al., 2010); other mutation types are indicated. APOA5, apolipoprotein A-V; APOB, apolipoprotein B; dbSNP, Database of Single Nucleotide Polymorphism; G1K, 1000 Genomes Project; GCKR, glucokinase regulatory protein; GWAS, genome-wide association study; HTG, hypertriglyceridemia; LPL, lipoprotein lipase; Polyphen, Polymorphism Phenotyping; SIFT, Sorting Intolerant from Tolerant.

Figure 3.4. Rare variants identified in HTG patients and healthy controls by resequencing candidate genes from mouse and human models of HTG. Variants above gene maps were identified in HTG patients and variants below gene maps were identified in controls. Rare variants are coloured according to their identification in both HTG patients and controls (black), exclusivity to HTG patients or controls (blue), or proven biological dysfunction or truncation (red). Nomenclature refers to functional protein sequences. Gene maps are roughly to scale, although differ in scale between genes. HTG, hypertriglyceridemia.

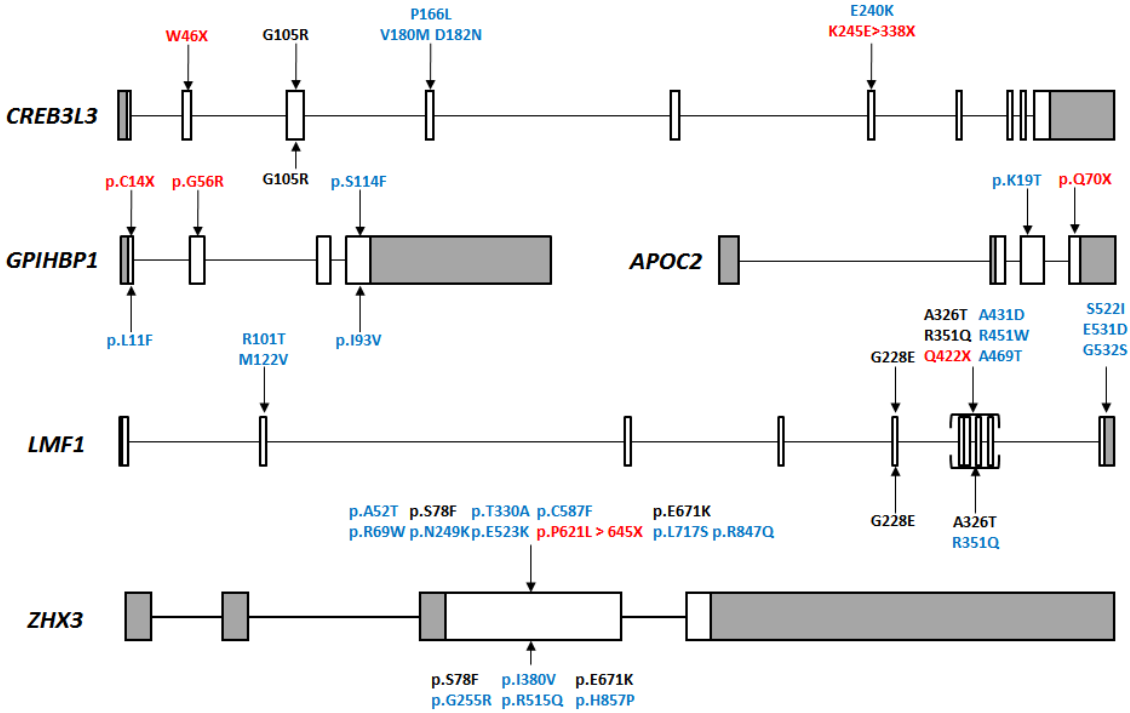

CREB3L3
W46X
G105R
P166L
V180M D182N
E240K
K245E>338X
G105R
GPIHBP1
p.C14X
p.G56R
p.S114F
p.L11F
p.I93V
APOC2
p.K19T
p.Q70X
LMF1
R101T
M122V
A326T A431D
R351Q R451W
G228E Q422X A469T
S522I
E531D
G532S
G228E
A326T
R351Q
ZHX3
p.A52T p.S78F p.T330A p.C587F p.E671K
p.R69W p.N249K p.E523K p.P621L > 645X p.L717S p.R847Q
p.S78F p.I380V p.E671K
p.G255R p.R515Q p.H857P

Table 3.6. Rare variant accumulation in candidate HTG genes in HTG patients and healthy controls.

Variant selection	*GPIHBP1*		*LMF1*		*CREB3L3*		*APOC2*		*ZHX3*		Total			
	HTG	Co	HTG	Co	HTG	Co	HTG	Co	HTG	Co	HTG	Co	OR (95% CI)	P-value
Total <1%	6	5	28	16	26	16	6	0	33	22	99	59	1.25 (0.88-1.78)	0.22
Nonsynonymous	4	2	16	5	11	1	6	0	16	8	53	16	2.47 (1.39-4.40)	0.0014
Synonymous/noncoding	2	3	12	11	15	15	0	0	17	14	46	43	0.80 (0.51-1.24)	0.37
Total <5%	6	5	97	68	41	25	6	0	33	22	183	120	1.14 (0.87-1.49)	0.37
Nonsynonymous	4	2	59	30	11	1	6	0	16	8	96	41	1.75 (1.18-2.59)	0.0060
Synonymous/noncoding	2	3	38	38	30	24	0	0	17	14	87	79	0.82 (0.59-1.15)	0.26

CI, confidence interval; Co, controls; HTG, hypertriglyceridemia; OR, odds ratio. Nonsynonymous variants cause missense or nonsense amino acid substitutions. Odds ratios and P-values are calculated based on carriers versus non-carriers in HTG patients (n=437) and population-based controls (n=326).

observed accumulation (OR=2.3, P=0.0046). These associations were maintained when uncommon variants with MAF <5% were included in the analysis **(Table 3.6)**. At a 5% allele frequency threshold, 96 variant occurrences were observed among HTG patients and 41 variant occurrences were observed among healthy controls, corresponding to a 1.8-fold increase in variants among HTG patients (P=0.0060). We identified 17 HTG patients and 6 healthy controls carrying 2 rare variants, which again had a minimal effect on the observed accumulation (OR=1.7; P=0.019). Rare variant accumulation was robust to potential sampling or technical biases in our cohort as no differences were observed between HTG patients and healthy controls in terms of benign synonymous and noncoding variants at either frequency threshold (1%: P=0.37; 5%: P=0.26).

We next hypothesized that enrichment for rare nonsynonymous variants with increased likelihood of full penetrance would accentuate the effect size of variants that accumulate in HTG patients. Accordingly, we assessed rare variant accumulation in HTG patients and healthy controls using predictions of deleterious effect using SIFT and Polyphen-2. The excess of rare variants in HTG patients was increased when nonsynonymous variants were filtered by prediction of deleterious effects *in silico* **(Table 3.7)**. There were 32/43 distinct rare variants predicted by ≥1 algorithm to have a deleterious effect *in silico*, including 37 variant occurrences in 437 HTG patients compared to 9 variant occurrences in 326 controls, which corresponded to a 3.1-fold increase in variants found in HTG patients (P=0.0018). Furthermore, 17/43 distinct rare variants were predicted by both algorithms to have deleterious effects *in silico*, including

Table 3.7. Enrichment for rare variants with *in silico* prediction of deleterious effect in HTG patients and healthy controls.

Variant selection	*In silico* filter	HTG	Co	OR (95% CI)	P-value
Nonsynonymous <1%	None	53	16	2.47 (1.39-4.40)	0.0014
	1	37	9	3.07 (1.46-6.44)	0.0018
	2	16	3	4.00 (1.15-13.8)	0.019
Nonsynonymous <5%	None	96	41	1.75 (1.18-2.59)	0.0060
	1	80	34	1.76 (1.18-3.01)	0.010
	2	16	3	4.00 (1.15-13.8)	0.019

CI, confidence interval; Co, controls; HTG, hypertriglyceridemia; OR, odds ratio. The *in silico* filter refers to the number of algorithms (0, 1 or 2) predicting deleterious functional effects of variants included in rare variant accumulation analysis. Polyphen-2 (possibly and probably damaging) or SIFT (damaging) were used to make predictions. Truncation mutations were given a score of 2. Odds ratios and P-values are calculated based on carriers versus non-carriers in HTG patients (n=437) and population-based controls (n=326).

16 variant observations in HTG patients and 3 variant observations in healthy controls, corresponding to a 4.0-fold increase in variants found in HTG patients (P=0.019). The accentuated accumulation of rare variants was essentially lost when considering variants with frequency <5% **(Table 3.7)**. There were 35/43 distinct uncommon variants predicted by ≥1 algorithm to have deleterious effect *in silico*, including 80 variants observed among HTG patients and 34 variants observed among healthy controls, corresponding to a 1.8-fold increase in variants observed in HTG patients (P=0.01). The subset of variants with 2 predictions of deleterious effect *in silico* was identical under both 1% and 5% frequency thresholds.

We conducted a *post hoc* analysis to determine whether rare variant accumulation could be detected in any one specific gene. Using Fisher's exact test, a significant excess of rare variants was detected in HTG patients versus healthy controls in *CREB3L3* and *APOC2*. *CREB3L3* had an 8.4-fold enrichment (95% CI: 1.08-64.4; P=0.016), and *APOC2* had a 9.9-fold enrichment (95% CI: 0.56-176.3; P=0.040). No evidence for rare variant accumulation was observed for *LMF1* (OR=2.3; p=0.11), *ZHX3* (OR=1.42; P=0.52) or *GPIHBP1* (OR=1.13; P=1.0). However, such experiments have relatively low power to detect associations. Accordingly, we sought to assess the individual burden of rare variants accumulation in specific genes among HTG patients and controls using the recently proposed permutation-based analyses designed for rare variant accumulation studies. Again, these three statistical analyses varied in their ability to detect rare variant accumulation in HTG patients versus healthy controls in several genes **(Table 3.8)**. A significant accumulation of rare variants was confirmed empirically across all loci by the

Table 3.8. Rare variant accumulation among individual genes measured using multiple *post hoc* analyses of genetic burden.

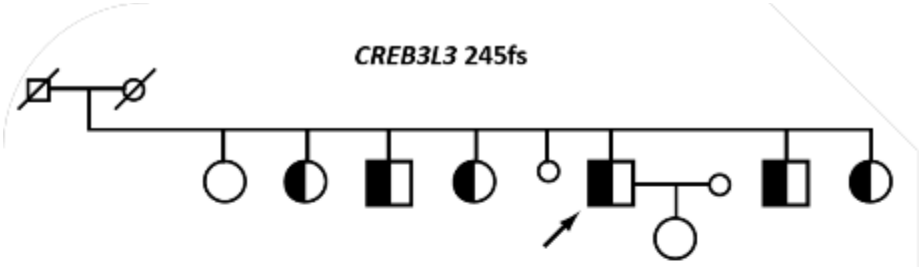
		Multi-locus test	*GPIHBP1*	*LMF1*	*CREB3L3*	*APOC2*	*ZHX3*
Rare (<1%)	Burden	0.0010	0.36	0.047	0.0056	0.018	0.21
	C-alpha	0.19	0.31	0.46	0.51	0.12	0.24
	Variable threshold	0.0029	0.50	0.021	0.0041	0.019	0.30

Multi-locus test includes all rare variants identified across 5 genes in a single burden analysis. Empirical P-values were calculated for each test statistic using 10,000 permutations. All statistical analyses were executed as implemented in PlinkSeq.

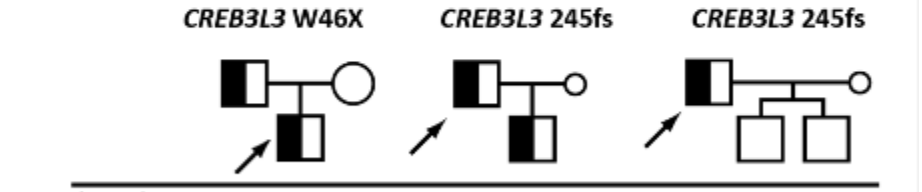
general burden test (p=0.0010) and the variable threshold approach (P=0.0029), but no differences were observed using the C-alpha test statistic (P=0.19). Virtually identical results were observed for a frequency threshold of <5%, as a significant accumulation was detected using the general burden test (P=0.0027) and the variable threshold approach (P=0.0066), but not with the C-alpha test statistic (P=0.11). Among individual genes, rare variant accumulation was detected in *LMF1*, *CREB3L3*, and *APOC2* using the general burden test (P=0.047, P=0.0056, and P=0.018, respectively) and the variable threshold approach (P=0.021, P=0.0041, and P=0.019, respectively). These analyses did not detect significant rare variant accumulation in either *GPIHBP1* (P=0.36-0.50) or *ZHX3* (P=0.21-0.30). The C-alpha test statistic could not detect significant rare variant accumulation in any particular gene.

Next, we sought to determine whether rare variants in non-GWAS genes showed evidence for cosegregation with HTG in nuclear families. Lipoprotein profiles of nuclear families related to 4 probands carrying either the *CREB3L3* E245Kfs397X (245fs) or W46X nonsense mutations were ascertained **(Figure 3.5)**. Within all pedigrees, probands had the highest plasma TG concentration of all subjects, whereas non-proband carriers all had somewhat elevated plasma TG concentrations. Overall, mean plasma TG concentrations were significantly elevated across 11 nonsense variant carriers versus 5 non-carrier first degree relatives (9.67±4.70 vs. 1.66±0.55 mmol/L, P=0.021). Interestingly, the father related to the W46X proband was also a carrier, but had normal plasma TG concentrations, whereas the mother related to the W46X proband who did not

Figure 3.5. Cosegregation of the *CREB3L3* K245Efs374X (245fs) and W46X variants with HTG through four pedigrees. Symbols for males and females are squares and circles, respectively. Small symbols are subjects not genotyped in our study. Symbols with a strike-through represent deceased subjects. White filled symbols are non-carriers; half-filled symbols are heterozygous carriers, black-filled symbols are homozygous carriers (none in this pedigree). Arrows indicate probands referred for tertiary referral.



Age (years)	68	67	65	64	62	33	58	56
BMI (kg/m^2)	21.5	27.5	27.5	27.0	30.3	21.9	25.1	23.6
TC (mmol/L)	3.49	4.49	2.92	3.32	12.3	3.59	7.00	4.11
TG (mmol/L)	0.77	2.14	2.66	2.70	51.0	1.61	3.17	1.16
LDL (mmol/L)	1.53	2.56	0.80	1.19	ND	1.95	4.23	2.24
HDL (mmol/L)	1.63	0.95	0.91	0.89	0.31	0.91	1.32	1.34



Age (years)	91	43	81	42	15	38	18	17
BMI (kg/m^2)	21.8	29.7	23.4	36.0	28.9	30.7	35.5	29.6
TC (mmol/L)	3.76	4.97	8.02	17.6	4.43	6.30	4.06	3.98
TG (mmol/L)	1.49	6.59	3.97	27.5	3.10	4.92	1.05	1.10
LDL (mmol/L)	2.40	ND	4.59	ND	2.10	ND	2.48	2. 22
HDL (mmol/L)	0.68	0.49	1.69	0.23	0.92	0.80	1.10	1.26

carry the variant, had elevated plasma TG concentration. These data suggest that the penetrance of heterozygous *CREB3L3* variants is incomplete, and that rare variants are individually insufficient to cause HTG, likely requiring additional environmental and genetic factors to significantly alter TG metabolism.

Finally, we hypothesized that variants with frequency >1% in HTG-associated candidate genes could also mediate increased susceptibility to HTG. Using logistic regression, we tested uncommon and common variants for association with HTG **(Table 3.9)**. Only the uncommon variant P562R in *LMF1* was modestly associated with HTG, corresponding to a 3.8-fold increase in disease susceptibility (P=0.04), which was robust to covariate adjusted analysis (OR=4.2; P=0.03). However, given the 1.2% MAF of P562R and modest association P-value, we believe that this variant is likely a false positive association that would not sustain a Bonferroni-corrected significance threshold. The common variant G36D in *LMF1* was also modestly associated with HTG (P=0.02), although this association was not robust to a covariate adjusted analysis (P=0.13). Generally, the relatively poor representation of common variation in these genes may provide a putative explanation for their lack of detection by large-scale GWAS of HTG or plasma TG concentration.

3.3.4. Combined analysis of all sequenced genes

As another *post hoc* experiment, we conducted a combined analysis of all rare variants identified in our resequencing studies. Not surprisingly, a significant accumulation of rare variants was again identified in HTG patients (**Table 3.10**). In total, 120 distinct

Table 3.9. Association between uncommon or common variants and HTG.

		Alleles		Frequency			Unadjusted			Adjusted		
Gene	Variant	Ref	Alt	Combined	HTG	Control	OR	SE	P	OR	SE	P
GPIHBP1	C14F	G	T	0.082	0.081	0.083	0.98	0.18	0.92	1.08	0.20	0.70
LMF1	G36D	C	T	0.092	0.078	0.11	0.66	0.18	0.02	0.73	0.21	0.13
LMF1	R354W	G	A	0.018	0.015	0.021	0.68	0.39	0.33	0.73	0.45	0.49
LMF1	R364Q	C	T	0.015	0.017	0.012	1.41	0.44	0.44	1.03	0.51	0.96
LMF1	P562R	G	C	0.012	0.017	0.0046	3.83	0.64	0.04	4.18	0.68	0.03
ZHX3	N310S	T	C	0.18	0.18	0.19	0.91	0.14	0.49	0.88	0.16	0.43

Alt, alternate; HTG, hypertriglyceridemia; OR, odds ratio; P, P-value; Ref, reference; SE, standard error. Association was tested using multiple logistic regression. Adjusted analyses entered age, sex, body mass index and diabetes status as covariates.

Table 3.10. Combined *post hoc* analysis of rare variants observed in HTG patients and healthy controls.

Hypothesis	Gene	Distinct variants	HTG	Control
GWAS	*APOA5*	5	5	1
	GCKR	12	20	5
	LPL	13	18	3
	APOB	46	84	38
Non-GWAS	*APOC2*	3	5	0
	CREB3L3	7	11	1
	GPIHBP1	6	5	2
	LMF1	12	15	5
	ZHX3	16	15	8
Post hoc	Total	120	178	63

Rare variants were defined by a combined frequency <1%. Both the number of distinct variants and total occurrences in 421 HTG patients and 324 healthy controls are shown. GWAS, genome-wide association study; HTG, hypertriglyceridemia.

heterozygous variants were observed, including 178 variants in 421 HTG patients and 63 variants in 324 healthy controls, corresponding to an overall 2.1-fold (95% CI: 1.53-2.91) increase in rare variant carriers in HTG patients versus healthy controls ($P=3.40X10^{-6}$). The majority of subjects carried only 1 rare variant, however subjects carrying multiple rare variants were also significantly over-represented in HTG patients, including 8.6% of HTG patients versus 2.4% of healthy controls, corresponding to a 3.5-fold (95% CI: 1.60-7.62) increase in carriers of multiple rare variants in HTG patients versus healthy controls ($P=8.0X10^{-4}$). Gene-centric association tests were not different from above, although the multi-locus association test detected a significant accumulation of variants in HTG patients versus controls using all three statistical analyses including the general burden test ($P=6.7X10^{-5}$), C-alpha ($P=0.049$), and the variable threshold approach ($P=1.0X10^{-6}$).

Finally, we assessed the contribution of different genetic and clinical variables to the total variation in HTG diagnosis, in subjects common between GWAS and resequencing cohorts. A comprehensive logistic regression model including clinical variables and both common and rare genetic variants explained maximally 41.6% of total variation in HTG diagnosis: clinical variables explained 19.7%, common genetic variants in 7 HTG-associated loci (from Chapter 2) explained 20.6% and rare genetic variants across 4-GWAS identified genes explained 1.3%. Interestingly, there was virtually no difference when rare variants from HTG-associated candidate genes were added to these calculations, contributing only 0.07% to the unattributed variation. Inclusion of a combined variable including any rare variant identified in either a GWAS or non-GWAS gene similarly explained only 1.3% of variation in HTG diagnoses.

3.4. Discussion

The principal finding of this chapter is that a significant excess of rare variants was identified in patients with polygenic adult-onset HTG. Extensive re-sequencing of GWAS-identified candidate genes and non-GWAS HTG-associated candidate genes selected based on mouse models and Mendelian syndromes of HTG identified a 2-fold increase in rare variants in HTG patients versus controls, including 178 variants in HTG patients and 63 variants in controls. Importantly, our study also provides valuable insight into the genetic architecture of complex diseases including HTG, demonstrating that: (1) GWAS-identified genes harbouring small effect common variants associated with complex disease also harbour rare variants that contribute additionally to disease susceptibility; and (2) non-GWAS candidate genes lacking common variation associated with complex diseases may harbour rare variants that significantly increase disease susceptibility.

In GWAS-identified genes, 80 distinct heterozygous missense and nonsense rare variants were identified across *APOA5*, *GCKR*, *LPL*, and *APOB* in 437 HTG patients and 326 healthy controls, corresponding to a 2.4-fold increase in rare variants in HTG patients ($P=6.2X10^{-8}$). Enrichment for variants with increased likelihood of mechanistic effect using exclusivity in HTG patients or controls enhanced this association: a 4.1-fold increase in exclusive variants was observed in HTG patients ($P=2.4X10^{-5}$). *Post hoc* analyses of genetic burden further implicated specific genes in polygenic HTG pathophysiology, including *LPL*, *APOB*, and *GCKR*. In HTG-associated candidate genes, 43 distinct heterozygous missense and nonsense rare variants were identified across

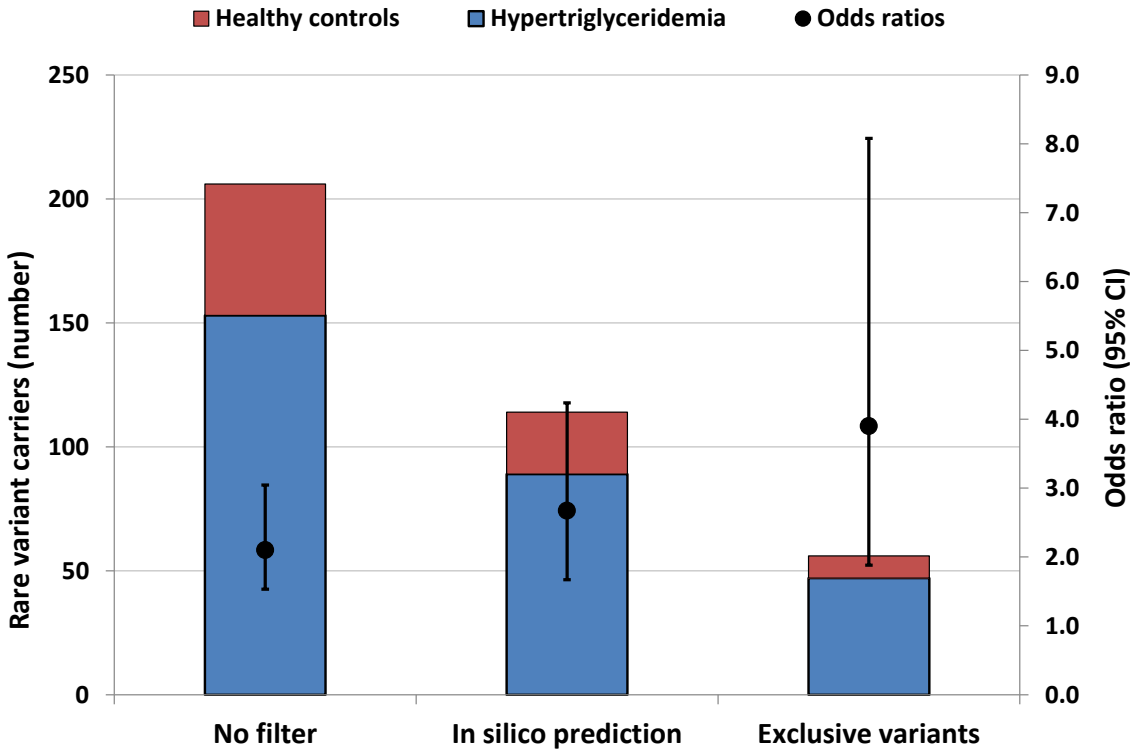
GPIHBP1, *LMF1*, *CREB3L3*, *APOC2*, and *ZHX3* in 437 HTG patients and 326 healthy controls, corresponding to a 2.5-fold increase of variants in HTG patients (P=0.001). Enrichment for variants with increased likelihood of mechanistic effect based on *in silico* predictions of deleterious effect accentuated this accumulation: variant enrichment in HTG patients was increased by 3.0- and 4.0-fold with 1 or 2 predictions of deleterious effect, respectively. *Post hoc* analyses of genetic burden assessed among individual genes suggested that rare variant accumulation was greater in *LMF1, APOC2* and *CREB3L3*, rather than in *GPIHBP1* or *ZHX3*. Together, these analyses suggest that rare variation in HTG-associated candidate genes is strongly associated with polygenic HTG susceptibility and the rare variants incrementally contribute to the unexplained variation in HTG susceptibility.

The burden of rare variants identified in HTG patients is very likely to increase HTG susceptibility. This is supported by several lines of evidence: (1) Genetic variants causing premature truncations that likely attenuate protein function were observed in 14 (3.2%) HTG patients versus 1 (0.30%) healthy control; (2) large effect singleton variants (that are only observed in 1 subject) were observed in 67 (15.3%) HTG patients versus 17 (5.2%) in healthy controls; (3) rare frameshift variants cosegregate with increased plasma TG concentration through families; and several *bonafide* deleterious mutations identified in HTG patients have been characterized in the literature and in preliminary studies in our laboratory (unpublished data). Furthermore, enrichment for rare variants with increased likelihood of fully penetrant deleterious effects shows an increase in effect size and significance of association. This filtering step seemed to increase specificity for

deleterious, potentially functional rare variants in HTG patients and reduce the noise contributed by benign variants in controls. In fact, the effect of likely deleterious rare variants more closely approximated the impact of carrying one APOA-V S19W variant (OR=5.35) (Hegele et al., 2009), which is the best single genetic marker so far reported for polygenic HTG. Filtration based on predicted dysfunction also reduced the contribution of any single gene to the overall rare variant tally and ensured a more uniform signal arising from all genes selected under our initial hypothesis. Interestingly, enrichment for exclusivity most greatly accentuated the accumulation observed in HTG patients (**Figure 3.6)**, supporting the hypothesis that variants absent from healthy controls are large effect variants that contribute to disease susceptibility. Conversely, including uncommon variants with MAF between 1% and 5% attenuated the estimated effect of rare variants on HTG susceptibility. This suggests that alleles with higher frequency are less likely to contribute to increased HTG susceptibility.

Rare variants in HTG patients may also contribute to the phenotypic heterogeneity characterized by the Fredrickson HLP phenotypes. For instance, the *APOB* p.R3500W variant causes hypercholesterolemia (Gaffney et al., 1995), but was found in a HTG patient with HLP phenotype 2B, defined by both plasma TG and total cholesterol in excess of the 95th percentile. For this individual case, *APOB* p.R3500W is more likely contributing to the elevated total cholesterol phenotype, but the mutation is part of the genetic background of this patient that led to his ascertainment through the lipid clinic. This patient exemplifies our working model that both common and rare genetic

Figure 3.6. Filtration of rare variants based on predictions of deleterious effect accentuates rare variant accumulation in HTG patients versus controls. Data presented are reflective of variants in genome-wide association study (GWAS)-identified genes *APOA5*, *GCKR*, *LPL*, and *APOB*, including 154 variants in 437 HTG patients and 53 variants in healthy controls. Deleterious effects were predicted using ≥1 *in silico* prediction algorithm (SIFT or Polyphen-2) or exclusivity in HTG patients or controls. Bars represent the number of rare variants among HTG patients and healthy controls (left y-axis); the corresponding odds ratio for rare variant accumulation is shown (right y-axis). Bars around odds ratios represent 95% confidence intervals.


Healthy controls
Hypertriglyceridemia
Odds ratios
Rare variant carriers (number)
Odds ratio (95% CI)
250
200
150
100
50
0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
No filter
In silico prediction
Exclusive variants

determinants in TG-associated genes together contribute to the phenotypic heterogeneity underlying HTG.

Our study provides a proof-of-principle that rare variant accumulation analyses can help implicate individual genes in complex phenotypes without prior identification by GWAS. Multi-locus association tests that group rare variants from many genes together provide increased power to detect an association with disease phenotypes, however they are unable to identify the contribution of specific genes to the overall association. We first conducted multi-locus association tests to evaluate our main hypotheses that GWAS-identified and non-GWAS candidate genes would harbour an excess of rare variants in HTG patients. We next conducted *post hoc* gene-centric analyses using three different methods to empirically implicate specific genes in HTG susceptibility. These latter analyses were of particular interest for non-GWAS genes, and specifically the 2 recently identified genes *CREB3L3* and *ZHX3*.

Both the general test of burden and the variable threshold approach were most successful at identifying rare variant accumulation in *LPL*, *APOB*, *GCKR*, *LMF1*, *CREB3L3* and *APOC2* as being associated with HTG susceptibility, whereas C-alpha only identified a specific burden in *LPL*. These differences in performance between statistical analyses are probably related to their respective designs and also to the distribution of variants in our sample. For instance, the variable threshold approach provides increased weight to very rare variants, which in retrospect was a favorable weighting scheme for our samples, since the majority of rare variants identified in these

studies were singletons. In contrast, C-alpha seems designed to detect genetic burden when genes have numerous rare variants, and when opposing directions of effect on phenotype are suspected for particular variants. C-alpha appears inadequately adapted to analyze numerous singleton variants in cases versus controls; recurrent rare variants seem to be optimal for this analysis. Thus, the general test of burden and the variable threshold approach seem to have been best suited to identify rare variant accumulation within our study.

Interestingly, rare variants in GWAS-identified and non-GWAS candidate genes explained only an additional ~1.1% of HTG diagnoses. However, we believe that this is likely an underestimate of the true biological effect of rare variants. In regression equations, the effect attributable to carrying a rare variant is estimated from the relative occurrence of variants in HTG patients versus controls, which essentially treats all variants as having an equal effect. More precise attribution of variation - weighting - to rare variants requires knowledge of whether they are truly benign or deleterious. Generation of an appropriate weighting scheme will require functional genomic analyses both *in vitro* and *in vivo*. Alternatively, surrogate weighting schemes, based on inverse allele frequencies for instance, may also prove useful in the interim. Such a weighting strategy will effectively minimize the contribution of benign variants identified in controls, and maximize the effect of pathogenic variants identified in HTG patients, allowing regression equations to accurately estimate the variation attributable to disease causing rare variants in HTG diagnoses. However, rare variants will probably not explain very large proportions of HTG diagnoses, even when their effects are large enough to

completely account for the disease pathophysiology. Such variants are rare and cumulatively affect very few HTG patients, which will limit the total proportion of variation attributable to rare variants.

Ultimately, our study provides evidence that rare variants contribute incrementally to the unexplained variation underlying HTG and that GWAS-identified and non-GWAS candidate genes in HTG pathophysiology, specifically *LPL*, *GCKR*, and *APOB*, *CREB3L3*, *APOC2*, and *LMF1*. *LPL* encoding lipoprotein lipase is well-established in TG metabolism, and both polygenic and monogenic forms of HTG. However, *GCKR* and *APOB* are more recently implicated players in HTG susceptibility. *GCKR* encodes an allosteric regulator of glucokinase (GK) called the glucokinase regulatory protein (GKRP), which allows rapid mobilization of GK in response to increased cellular glucose concentrations (Agius, 2008). The lead single nucleotide polymorphism (SNP) in *GCKR*, a common proline to leucine (P446L) functional variant in *GCKR* attenuates ligand-dependent GK inhibition (Beer et al., 2009, Orho-Melander et al., 2008), in turn promoting glucose uptake and mobilization, thus attenuating fatty acid oxidation and promoting de novo TG synthesis (Beer et al., 2009). *APOB* encodes apolipoprotein B-100, the backbone of TG-rich and other atherogenic lipoproteins, which was earlier implicated in phenotypes involving plasma low density lipoprotein (LDL) cholesterol concentrations, such as familial hypercholesterolemia (FH) and familial hypobetalipoproteinemia (Hegele, 2009). Furthermore, our study demonstrates that heterozygous rare variants in HTG candidate genes also contribute to HTG susceptibility. Most notably, we found that the recently identified *CREB3L3,* associated with TG

metabolism, is also significantly associated with increased HTG susceptibility. *CREB3L3* was recently implicated in TG metabolism as a transcription factor that regulates expression of cofactors required for TG-rich lipoprotein metabolism including *APOC2*, *APOA5* and *APOA4* (Lee et al., 2011), but has never identified in a pure Mendelian form of HTG. In contrast, *APOC2* and *LMF1* are established genes in TG metabolism: *APOC2* encodes an essential lipoprotein lipase (LPL) cofactor required for hydrolysis of TG-rich lipoproteins (Hegele et al., 1991); whereas *LMF1* encodes a maturation factor required for LPL activity (Peterfy et al., 2007).

Importantly, our findings cannot exclude the potential contribution of genetic variation in *APOA5*, *GPIHBP1* or *ZHX3* to HTG susceptibility, since these genes have been implicated in lipoprotein pathophysiology through other means. For instance, *APOA5* appears to play a focal role in hydrolysis of TG-rich lipoproteins by enhancing LPL activity (Charlton-Menys and Durrington, 2005), and has been implicated in HTG by *Apoa5* knock-out mice (Pennacchio et al., 2001) and human HTG patients homozygous for rare LOF mutations in *APOA5* (Marcais et al., 2005). *GPIHBP1* encodes a cell-surface scaffold required for transendothelial LPL transport (Beigneux et al., 2007), which is absolutely necessary for TG-rich lipoprotein metabolism. Functionally characterized heterozygous mutations in *GPIHBP1* (Beigneux et al., 2009, Franssen et al., 2010, Olivecrona et al., 2010) have been reported in patients with HTG. It is probable that we were unable to show a significant accumulation of rare variants in *APOA5* and *GPIHBP1*, due to small gene size, and relatively low power to detect increased genetic burden on an individual gene basis. Conversely, given the abundance of rare variants in

ZHX3 found in both HTG patients and healthy controls, statistical power was likely not an issue in this gene and any association with the HTG phenotype would be tenuous at best. *ZHX3* encodes a protein homologous to *ZHX2*, a developmental regulator of lipid metabolism (Gargalovic et al., 2010), although perhaps rare variants in this gene do not largely contribute to polygenic HTG susceptibility.

In summary, our results indicate the importance of framing hypotheses and selecting analytic strategies that apply not only for Sanger re-sequencing projects, but also for future exome and targeted resequencing studies, particularly in the absence of prior GWAS association signals for complex phenotypes. Specifically, re-sequencing four genes associated with HTG by GWAS and resequencing five HTG candidate genes identified by mouse models and Mendelian syndromes of HTG has helped to identify an accumulation of rare nonsynonymous variants in polygenic adult-onset HTG patients compared to normolipidemic population-based controls. Functional analyses may more accurately define the extent of dysfunction of rare variants identified in HTG patients and their role in disease causation, while higher level analyses including gene-gene and gene-environment interactions will determine the combined impact of multiple genetic variants on plasma TG concentration in patients with HTG. These findings support a cumulative burden of rare variants in both known and novel genes in human polygenic HTG that incrementally contribute to the unexplained variation in HTG. These studies foreshadow the importance of testing predefined hypotheses and provide a framework for whole-exome and genome resequencing studies of complex traits.

3.5. References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods *7*, 248-249.

Agius, L. (2008). Glucokinase and molecular aspects of liver glycogen metabolism. Biochem J *414*, 1-18.

Anand, S.S., Yusuf, S., Vuksan, V., Devanesen, S., Teo, K.K., Montague, P.A., Kelemen, L., Yi, C., Lonn, E., Gerstein, H., *et al.* (2000). Differences in risk factors, atherosclerosis, and cardiovascular disease between ethnic groups in Canada: the Study of Health Assessment and Risk in Ethnic groups (SHARE). Lancet *356*, 279-284.

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., *et al.* (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet *41*, 47-55.

Bansal, V., Libiger, O., Torkamani, A., and Schork, N.J. (2010). Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet *11*, 773-785.

Beer, N.L., Tribble, N.D., McCulloch, L.J., Roos, C., Johnson, P.R., Orho-Melander, M., and Gloyn, A.L. (2009). The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. Hum Mol Genet *18*, 4081-4088.

Beigneux, A.P., Davies, B.S., Gin, P., Weinstein, M.M., Farber, E., Qiao, X., Peale, F., Bunting, S., Walzem, R.L., Wong, J.S., *et al.* (2007). Glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1 plays a critical role in the lipolytic processing of chylomicrons. Cell Metab *5*, 279-291.

Beigneux, A.P., Franssen, R., Bensadoun, A., Gin, P., Melford, K., Peter, J., Walzem, R.L., Weinstein, M.M., Davies, B.S., Kuivenhoven, J.A., *et al.* (2009). Chylomicronemia with a mutant GPIHBP1 (Q115P) that cannot bind lipoprotein lipase. Arterioscler Thromb Vasc Biol *29*, 956-962.

Busca, R., Peinado, J., Vilella, E., Auwerx, J., Deeb, S.S., Vilaro, S., and Reina, M. (1995). The mutant Asn291-->Ser human lipoprotein lipase is associated with reduced catalytic activity and does not influence binding to heparin. FEBS Lett *367*, 257-262.

Charlton-Menys, V., and Durrington, P.N. (2005). Apolipoprotein A5 and hypertriglyceridemia. Clin Chem *51*, 295-297.

Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., and Hobbs, H.H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science *305*, 869-872.

Connelly, P.W., Maguire, G.F., and Little, J.A. (1987). Apolipoprotein CIISt. Michael. Familial apolipoprotein CII deficiency associated with premature vascular disease. J Clin Invest *80*, 1597-1606.

Dorfmeister, B., Zeng, W.W., Dichlberger, A., Nilsson, S.K., Schaap, F.G., Hubacek, J.A., Merkel, M., Cooper, J.A., Lookene, A., Putt, W., *et al.* (2008). Effects of six APOA5 variants, identified in patients with severe hypertriglyceridemia, on in vitro lipoprotein lipase activity and receptor binding. Arterioscler Thromb Vasc Biol *28*, 1866-1871.

Emi, M., Wilson, D.E., Iverius, P.H., Wu, L., Hata, A., Hegele, R., Williams, R.R., and Lalouel, J.M. (1990). Missense mutation (Gly----Glu188) of human lipoprotein lipase imparting functional deficiency. J Biol Chem *265*, 5910-5916.

Franssen, R., Young, S.G., Peelman, F., Hertecant, J., Sierts, J.A., Schimmel, A.W., Bensadoun, A., Kastelein, J.J., Fong, L., Dallinga-Thie, G.M., *et al.* (2010). Chylomicronemia with Low Postheparin Lipoprotein Lipase Levels in the Setting of GPIHBP1 Defects. Circ Cardiovasc Genet *3*, 169-178.

Gaffney, D., Reid, J.M., Cameron, I.M., Vass, K., Caslake, M.J., Shepherd, J., and Packard, C.J. (1995). Independent mutations at codon 3500 of the apolipoprotein B gene are associated with hyperlipidemia. Arterioscler Thromb Vasc Biol *15*, 1025-1029.

Gargalovic, P.S., Erbilgin, A., Kohannim, O., Pagnon, J., Wang, X., Castellani, L., LeBoeuf, R., Peterson, M.L., Spear, B.T., and Lusis, A.J. (2010). Quantitative trait locus mapping and identification of Zhx2 as a novel regulator of plasma lipid metabolism. Circ Cardiovasc Genet *3*, 60-67.

Gotoda, T., Yamada, N., Kawamura, M., Kozaki, K., Mori, N., Ishibashi, S., Shimano, H., Takaku, F., Yazaki, Y., Furuichi, Y., *et al.* (1991). Heterogeneous mutations in the human lipoprotein lipase gene in patients with familial lipoprotein lipase deficiency. J Clin Invest *88*, 1856-1864.

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

Hegele, R.A., Ban, M.R., Hsueh, N., Kennedy, B.A., Cao, H., Zou, G.Y., Anand, S., Yusuf, S., Huff, M.W., and Wang, J. (2009). A polygenic basis for four classical

Fredrickson hyperlipoproteinemia phenotypes that are characterized by hypertriglyceridemia. Hum Mol Genet *18*, 4189-4194.

Hegele, R.A., Connelly, P.W., Maguire, G.F., Huff, M.W., Leiter, L., Wolfe, B.M., Evans, A.J., and Little, J.A. (1991). An apolipoprotein CII mutation, CIILys19----Thr' identified in patients with hyperlipidemia. Dis Markers *9*, 73-80.

Henderson, H.E., Ma, Y., Hassan, M.F., Monsalve, M.V., Marais, A.D., Winkler, F., Gubernator, K., Peterson, J., Brunzell, J.D., and Hayden, M.R. (1991). Amino acid substitution (Ile194----Thr) in exon 5 of the lipoprotein lipase gene causes lipoprotein lipase deficiency in three unrelated probands. Support for a multicentric origin. J Clin Invest *87*, 2005-2011.

Ishimura-Oka, K., Faustinella, F., Kihara, S., Smith, L.C., Oka, K., and Chan, L. (1992). A missense mutation (Trp86----Arg) in exon 3 of the lipoprotein lipase gene: a cause of familial chylomicronemia. Am J Hum Genet *50*, 1275-1280.

Johansen, C.T., Kathiresan, S., and Hegele, R.A. (2011a). Genetic determinants of plasma triglycerides. J Lipid Res *52*, 189-206.

Johansen, C.T., Wang, J., and Hegele, R.A. (2011b). Bias due to selection of rare variants using frequency in controls. Nat Genet *43*, 394-395.

Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., *et al.* (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet *42*, 684-687.

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., *et al.* (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., *et al.* (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet *41*, 56-65.

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 Protoc *4*, 1073-1081.

Lee, J.H., Giannikopoulos, P., Duncan, S.A., Wang, J., Johansen, C.T., Brown, J.D., Plutzky, J., Hegele, R.A., Glimcher, L.H., and Lee, A.H. (2011). The transcription factor CREB-H regulates triglyceride metabolism. Nature Medicine, In press.

Lee, M.W., Chanda, D., Yang, J., Oh, H., Kim, S.S., Yoon, Y.S., Hong, S., Park, K.G., Lee, I.K., Choi, C.S., *et al.* (2010). Regulation of hepatic gluconeogenesis by an ER-bound transcription factor, CREBH. Cell Metab *11*, 331-339.

Ma, Y., Henderson, H.E., Murthy, V., Roederer, G., Monsalve, M.V., Clarke, L.A., Normand, T., Julien, P., Gagne, C., Lambert, M., *et al.* (1991). A mutation in the human lipoprotein lipase gene as the most common cause of familial chylomicronemia in French Canadians. N Engl J Med *324*, 1761-1766.

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., *et al.* (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753.

Marcais, C., Verges, B., Charriere, S., Pruneta, V., Merlin, M., Billon, S., Perrot, L., Drai, J., Sassolas, A., Pennacchio, L.A., *et al.* (2005). Apoa5 Q139X truncation predisposes to late-onset hyperchylomicronemia due to lipoprotein lipase impairment. J Clin Invest *115*, 2862-2869.

Neale, B.M., Rivas, M.A., Voight, B.F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S.M., Roeder, K., and Daly, M.J. (2011). Testing for an unusual distribution of rare variants. PLoS Genet *7*, e1001322.

Nejentsev, S., Walker, N., Riches, D., Egholm, M., and Todd, J.A. (2009). Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science *324*, 387-389.

Olivecrona, G., Ehrenborg, E., Semb, H., Makoveichuk, E., Lindberg, A., Hayden, M.R., Gin, P., Davies, B.S., Weinstein, M.M., Fong, L.G., *et al.* (2010). Mutation of conserved cysteines in the Ly6 domain of GPIHBP1 in familial chylomicronemia. J Lipid Res *51*, 1535-1545.

Orho-Melander, M., Melander, O., Guiducci, C., Perez-Martinez, P., Corella, D., Roos, C., Tewhey, R., Rieder, M.J., Hall, J., Abecasis, G., *et al.* (2008). Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. Diabetes *57*, 3112-3121.

Pennacchio, L.A., Olivier, M., Hubacek, J.A., Cohen, J.C., Cox, D.R., Fruchart, J.C., Krauss, R.M., and Rubin, E.M. (2001). An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. Science *294*, 169-173.

Peterfy, M., Ben-Zeev, O., Mao, H.Z., Weissglas-Volkov, D., Aouizerat, B.E., Pullinger, C.R., Frost, P.H., Kane, J.P., Malloy, M.J., Reue, K., *et al.* (2007). Mutations in LMF1 cause combined lipase deficiency and severe hypertriglyceridemia. Nat Genet *39*, 1483-1487.

Plomin, R., Haworth, C.M., and Davis, O.S. (2009). Common disorders are quantitative traits. Nat Rev Genet *10*, 872-878.

Price, A.L., Kryukov, G.V., de Bakker, P.I., Purcell, S.M., Staples, J., Wei, L.J., and Sunyaev, S.R. (2010). Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet *86*, 832-838.

Priore Oliva, C., Pisciotta, L., Li Volti, G., Sambataro, M.P., Cantafora, A., Bellocchio, A., Catapano, A., Tarugi, P., Bertolini, S., and Calandra, S. (2005). Inherited apolipoprotein A-V deficiency in severe hypertriglyceridemia. Arterioscler Thromb Vasc Biol *25*, 411-417.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.

Romeo, S., Pennacchio, L.A., Fu, Y., Boerwinkle, E., Tybjaerg-Hansen, A., Hobbs, H.H., and Cohen, J.C. (2007). Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nat Genet *39*, 513-516.

Sabatti, C., Service, S.K., Hartikainen, A.L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M., *et al.* (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet *41*, 35-46.

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., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707-713.

Wang, J., Ban, M.R., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., and Hegele, R.A. (2008). APOA5 genetic variants are markers for classic hyperlipoproteinemia phenotypes and hypertriglyceridemia. Nat Clin Pract Cardiovasc Med *5*, 730-737.

Wang, J., Cao, H., Ban, M.R., Kennedy, B.A., Zhu, S., Anand, S., Yusuf, S., Pollex, R.L., and Hegele, R.A. (2007). Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). Arterioscler Thromb Vasc Biol *27*, 2450-2455.

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., *et al.* (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet *40*, 161-169.

CHAPTER 4

AN INCREASED BURDEN OF COMMON AND RARE LIPID-ASSOCIATED RISK ALLELES CONTRIBUTES TO THE PHENOTYPIC SPECTRUM OF HYPERTRIGLYCERIDEMIA

This chapter is based on the following publication: **Johansen, CT,** Wang, J, Lanktree, MB, *et al*. (2011). An increased burden of common and rare lipid-associated risk alleles contributes to the phenotypic spectrum of hypertriglyceridemia. *ATVB*, 31(8): 1916-26.

4.1. Introduction

Plasma triglyceride (TG) concentration is an independent risk factor for cardiovascular disease (CVD). Both fasting and post-prandial plasma TG concentrations are determinants of CVD (Bansal et al., 2007, Cullen, 2000, Hokanson and Austin, 1996), increasing risk of both myocardial infarction and ischemic stroke (Freiberg et al., 2008, Nordestgaard et al., 2007). Thus, patients with hypertriglyceridemia (HTG) defined by plasma TG concentrations >95th percentile have significantly elevated CVD risk (Hegele and Pollex, 2009, Yuan et al., 2007), which is frequently exacerbated by concomitant pro-atherogenic co-morbidities such as obesity, metabolic syndrome, and type 2 diabetes (T2D) (Yuan et al., 2007).

Recently, the genetic architecture of HTG has become clearer. However, the genetic determinants underlying the phenotypic heterogeneity among HTG patients

remain less well defined. Phenotypic heterogeneity among HTG patients is defined by qualitative and quantitative biochemical differences in plasma lipoproteins that form the basis of the World Health Organization or Fredrickson hyperlipoproteinemia (HLP) phenotypes. The HLP phenotypes defined by HTG (HLP-HTG) include one monogenic pediatric phenotype called chylomicronemia (HLP type 1), and four polygenic familial phenotypes called combined hyperlipidemia (HLP type 2B), dysbetalipoproteinemia (HLP type 3), primary hypertriglyceridemia (HLP type 4) and mixed hyperlipidemia (HLP type 5). HLP type 1 is usually caused by homozygous loss-of-function mutations in genes such as *LPL* (Havel and Gordon, 1960), *APOC2* (Connelly et al., 1987), *APOA5* (Priore Oliva et al., 2005), *LMF1* (Peterfy et al., 2007), and *GPIHBP1* (Beigneux et al., 2009, Wang and Hegele, 2007). However, the genetic architecture underlying predisposition to the polygenic HLP-HTG is not known (Hegele, 2009).

Candidate gene studies and genome-wide analyses have established that genetic variants associated with unfavourable plasma TG concentration in populations are also associated with susceptibility to clinical HTG (Hegele et al., 2009, Johansen et al., 2010, Wang et al., 2008a, Wang et al., 2008b, Wang et al., 2007). Both common and rare TG-associated variants accumulate in HTG patients, and account for a large proportion of explained variation among HTG patients (Chapters 2-3). However, corresponding analyses have not yet been performed in the HLP-HTG phenotypes.

Earlier surveys of few isolated genetic determinants suggested to a first approximation that TG loci were associated with multiple different HLP phenotypes,

suggesting that a shared genetic architecture may underlie the biochemically-defined HLP-HTG phenotypes (Wang et al., 2008a, Wang et al., 2008b). Here, we provide a comprehensive analysis of the genetic etiology of the HLP-HTG phenotypes in a much larger sample of HTG patients and healthy controls to assess this hypothesis. Further, we include all GWAS-identified genetic determinants of lipid metabolism recently identified by the Global Lipids Genetic Consortium (GLGC) (Teslovich et al., 2010), to assess the contribution of common variants associated with concentrations of TG, high-density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C) to both HTG susceptibility and HLP-HTG phenotypic heterogeneity. Finally, we have integrated rare variants into the analyses of HLP phenotypes. We demonstrate a shared genetic architecture of common and rare TG-associated variants as the genetic foundation across HTG patients, and provide evidence that accumulation of pleiotropic TG-associated, HDL-C-associated and LDL-C-associated variants may together dictate susceptibility to the polygenic HLP phenotypes.

4.2. METHODS

4.2.1. Study subjects

This study was approved by the University of Western Ontario Institutional Review Board (protocol #07920E), and ethics boards at collaborating institutions. Study subjects provided informed consent for collection of clinical data, DNA extraction and analysis. In total, 504 HTG patients and 1213 population-based controls were studied. HTG patients were European subjects ascertained through tertiary referral lipid clinics in London, Canada, and Amsterdam, Netherlands. HTG patients were clinically diagnosed with one of four familial polygenic HLP-HTG phenotypes prior to ascertainment for this study and without genetic information (see Table 1.5 in Chapter 1); these patients are now referred to as having HLP-HTG case status. Population-based controls were European subjects ascertained through the Study of Health Assessment and Risk in Ethnic Groups (Anand et al., 2000) and the Myocardial Infarction Genetics Consortium (Kathiresan et al., 2009), as previously described. Control subjects with fasting plasma TG concentration >2.3 mmol/L were excluded from analyses due to potentially undiagnosed HTG. Biochemical analyses of lipoprotein traits were conducted separately in each cohort, as previously described (Anand et al., 2000, Hegele et al., 2009, Kathiresan et al., 2009). The study sample consisted of all subjects genotyped in Chapter 2 (most of which were sequenced in Chapter 3) with the addition of few recently ascertained subjects.

4.2.2. Genotyping and sequencing

Genotyping was conducted using Affymetrix Genome-Wide Human Single Nucleotide Polymorphism (SNP) Array 6.0, as previously described (Johansen et al., 2010). Imputation was conducted using HapMap2 European phased haplotypes (Li and Abecasis, 2006), only imputing SNPs with r^2>0.4. Genotypes were subsequently extracted for lipid-associated loci identified by the GLGC (Teslovich et al., 2010). Functional variants in *APOE* (rs7412 and rs429358) were genotyped using validated TaqMan assays on an ABI Prism 7900HT Sequence Detection System and automated software (Applied Biosystems, Foster City, CA). Rare variants in *APOA5*, *GCKR*, *LPL* and exons 26 and 29 of *APOB* were defined as nonsense or missense mutations with allele frequency <1%, as previously described in Chapter 3. The list of variants identified by this study is located in Chapter 3.

4.2.3. Statistical analysis

Data management, linkage disequilibrium (LD) calculations, association testing and risk score calculations were conducted as implemented in PLINK (Purcell et al., 2007). Association was tested using an additive multiple logistic regression model with clinical covariates age, sex, body mass index (BMI), T2D and 10 principal components of ancestry (Patterson et al., 2006, Price et al., 2006). Statistical significance for association was defined as concordant direction of effect and P<0.05, as all SNPs have *a priori* evidence for association with lipid metabolism.

Each subject's genetic risk index (or genetic risk score) was determined by summing risk (TG-raising, LDL-raising, and HDL-lowering) alleles at 32 TG-associated

SNPs, 47 HDL-C-associated SNPs, or 37 LDL-C-associated SNPs, as identified by the GLGC (Teslovich et al., 2010). Genetic risk indices were measured using either unweighted scores or weighted scores, where indicated. Unweighted scores were created by summing the raw number of risk alleles carried by each subject. Weighted risk scores were created by summing risk alleles after being multiplied by their population-based effect size calculated by the GLGC (Teslovich et al., 2010). Units were initially reported as mg/dl, but converted here to mmol/l; conversion factors for TG and HDL-C/LDL-C were 0.0113 and 0.0258, respectively. All SNPs were included in the risk score regardless of statistical significance in the replication phase. Where SNPs were associated with multiple lipid phenotypes, they were included in each risk score, although weighted by the appropriate lipid-specific effect estimate. Pleiotropy was assessed by comparing risk scores constructed using SNPs exclusively associated with HDL-C or LDL-C, or jointly associated with HDL-C or LDL-C plus plasma TG concentration (Teslovich et al., 2010). Frequency distributions were plotted by dividing the range of risk scores into equal bins, and plotting the relative number of HTG patients or controls in each bin. Forest plots were generated by comparing the number of HTG patients in each of 7 risk score bins to the median risk score bin or the lowest risk score bin, as indicated. Summary statistics, risk score comparisons, rare variant accumulation and explained variation calculations were conducted in SAS v9.2 (SAS Institute, Cary, NJ). Summary statistics are displayed as mean ± standard deviation. Significance of increasing risk score in forest plots was determined using Cochrane-Armitage test for trend. Comparison of risk scores among HLP-HTG phenotypes was assessed using one-way ANOVA and *post*

hoc pairwise comparisons using Tukey's test. In figures, values are mean ± standard error; means sharing letters are not statistically different.

Rare variant accumulation was compared between HLP-HTG phenotypes and controls with Fisher's exact test, using a two-sided $P<0.05$ as the threshold for statistical significance. Differences in number of rare variants in each HLP-HTG phenotype was assessed using a 2X4 contingency table (number of variants in carriers versus non-carriers for each phenotype); differences in distribution of rare variants among genes in each HLP-HTG phenotype was assessed using a 4X4 contingency table (number of variants in each gene for each phenotype).

A logistic regression model using case or control status as the dependent variable was used to calculate the proportion of variation explained by clinical and genetic variables. The proportion of variation explained by the coefficient of determination (R^2) in linear regression does not have a straightforward analogue in logistic regression. Therefore, we used a likelihood-based pseudo-R^2 metric for logistic regression as implemented in SAS-v9.2, which compares the increase in information provided by a model fitted with predictors over a null model. We used incremental changes in this metric to assess how well each variable improves model fit, thus comparing the relative contribution of variables between models. Here, to facilitate comprehension, we describe successive increases in model fit (R^2) with progressive introduction of each variable as the proportion of variation explained by each variable. These analyses only included subjects for whom both genome-wide genotyping and resequencing data were available,

which comprised 346 HTG patients and 205 controls. Forward modeling was used to systematically enter clinical variables (age, sex, BMI, and T2D), common variants at TG-associated loci (0, 1 or 2 TG-raising alleles as each locus), the number of rare variants identified in HTG-associated genes (*APOA5*, *GCKR*, *LPL*, and *APOB* as a composite score) (Johansen et al., 2010), and *APOE* genotype (copies of ε2/ε3/ε4 isoform) into the model. These analyses were conducted including all HTG patients, and subsequently using HLP phenotype subgroups, in comparison to the same group of healthy controls. This analysis serves only as a relative comparison of variables contributing to the model within our cohort, and should be considered only within the context of our study.

4.3. Results

4.3.1. Study subjects

Baseline clinical characteristics of the 504 HTG patients and 1213 population-based controls are shown in **Table 4.1**. Generally, HTG patients had less favourable clinical profiles than controls. HTG patients had increased BMI, increased plasma total cholesterol, decreased plasma HDL-C and increased prevalence of T2D. HLP type 5 had the least favourable clinical profile across all traits, including fasting plasma TG concentration >99th percentile.

4.3.2. Replication of lipid-associated loci

First, we sought to replicate the genetic determinants of plasma lipoprotein metabolism from the GLGC (Teslovich et al., 2010) in our HTG cohort. For TG-associated loci (**Table 4.2a-e**), we replicated many loci previously associated with HTG (Johansen et al., 2010), and additionally replicated recently identified loci *KLHL8* (OR=1.36; P=1.5X10^{-3}) and *CYP26A1* (OR=1.29; P=5.9X10^{-3}). Generally, there was a trend towards significance at additional loci, as 29/32 loci had directions of effect concordant with estimates from the GLGC, significantly more than would be expected by chance (P=1.15X10^{-6}). Most TG-associated loci showed different patterns of association with the HLP-HTG phenotypes (**Table 4.2 a-e**)**,** however there were again more concordant directions of effect than would be expected by chance (HLP 5, P=0.007; HLP 4, P=0.01; HLP 3, P=0.007; HLP 2B, P=0.0002). Taken together, these data extend the concept that

Table 4.1. Baseline clinical characteristics of HLP-HTG patients and controls.

	All HTG	HLP type 5	HLP type 4	HLP type 3	HLP type 2B	Controls
Number	504	180	128	37	159	1213
Females	32.4%	29.4%	22.7%	37.8%	42.1%	42.0%
Diabetes	26.4%	31.7%	29.7%	24.3%	18.2%	0.3%
Age (years)	51.7 ± 12.7	49.6 ± 11.7	53.0 ± 13.4	51.4 ± 12.7	53.2 ± 13.1	48.2 ± 10.8
BMI (kg/m^2)	30.0 ± 4.9	30.6 ± 4.9	30.5 ± 5.9	28.9 ± 2.5	29.3 ± 4.3	26.4 ± 4.4
TC (mmol/L)	8.2 ± 3.8	10.8 ± 5.1	5.1 ± 1	9.4 ± 1.8	7.9 ± 1.5	5.1 ± 0.9
HDL-C (mmol/L)	0.9 ± 0.3	0.8 ± 0.3	0.8 ± 0.3	1.0 ± 0.3	1.1 ± 0.3	1.4 ± 0.4
TG (mmol/L)	13.6 ± 18.2	28.3 ± 24.3	5.4 ± 1.9	7.0 ± 3.9	5.3 ± 1.9	1.1 ± 0.4

HTG, hypertriglyceridemia; HLP, hyperlipoproteinemia; BMI, body mass index; TC, plasma total cholesterol; HDL-C, plasma high-density lipoprotein cholesterol; LDL-C, plasma low-density lipoprotein cholesterol; TG, plasma triglycerides. HTG blood sampling was performed in the fasting state (12hr); control blood sampling was performed in fasting or semi-fasting (4hr) state. Values are mean ± standard deviation.

Table 4.2a. Replication of TG-associated loci in all HLP-HTG phenotypes.

CHR	Gene	SNP	Minor allele/ modeled allele	All HTG (n=504)	
				OR (95% CI)	P-value
11	*APOA5*	rs964184	G/G	3.43 (2.72-4.31)	$1.12X10^{-25}$
2	*GCKR*	rs1260326	T/T	1.64 (1.36-1.97)	$1.97X10^{-7}$
8	*LPL*	rs12678919	G/A	2.21 (1.52-3.22)	$3.5X10^{-5}$
8	*TRIB1*	rs2954029	T/A	1.50 (1.24-1.81)	$3.8X10^{-5}$
1	*ANGPTL3*	rs2131925	G/T	1.51 (1.23-1.85)	$1.0X10^{-4}$
7	*MLXIPL*	rs7811265	G/A	1.63 (1.25-2.13)	$3.3X10^{-4}$
4	*KLHL8*	rs442177	G/T	1.36 (1.13-1.64)	$1.5X10^{-3}$
10	*CYP26A1*	rs2068888	A/G	1.29 (1.08-1.55)	$5.9X10^{-3}$
19	*CILP2*	rs10401969	C/T	1.72 (1.16-2.54)	$6.8X10^{-3}$
2	*APOB*	rs1042034	C/T	1.28 (1.02-1.61)	0.032
5	*TIMD4*	rs1553318	G/C	1.21 (1.00-1.46)	0.054
11	*FADS1*	rs174546	T/T	1.20 (1.00-1.45)	0.054
16	*CETP*	rs7205804	A/G	1.20 (1.00-1.45)	0.056
2	*IRS1*	rs2943645	C/T	1.20 (0.99-1.46)	0.061
6	*HLA*	rs2247056	T/C	1.21 (0.98-1.50)	0.076
1	*GALNT2*	rs1321257	G/G	1.16 (0.96-1.40)	0.12
5	*MAP3K1*	rs9686661	T/T	1.19 (0.95-1.48)	0.12
15	*LIPC*	rs261342	G/G	0.84 (0.68-1.05)	0.13
15	*CAPN3*	rs2412710	A/A	1.50 (0.87-2.59)	0.14
2	*COBLL1*	rs10195252	C/T	1.13 (0.94-1.36)	0.19
12	*LRP1*	rs11613352	T/C	1.11 (0.89-1.40)	0.35
7	*TYW1B*	rs13238203	T/C	1.30 (0.65-2.60)	0.47
8	*NAT2*	rs1495743	G/G	1.07 (0.87-1.33)	0.52
3	*MSL2L1*	rs645040	G/T	1.06 (0.85-1.33)	0.58
16	*CTF1*	rs11649653	G/C	1.05 (0.87-1.27)	0.59
20	*PLTP*	rs4810479	C/C	1.06 (0.85-1.32)	0.60
19	*APOE*	rs439401	T/C	0.95 (0.74-1.21)	0.68
22	*PLA2G6*	rs5756931	C/T	1.04 (0.86-1.25)	0.70
12	*ZNF664*	rs12310367	G/A	0.97 (0.81-1.17)	0.77
15	*FRMD5*	rs2929282	T/T	1.06 (0.69-1.64)	0.79
10	*JMJD1C*	rs10761731	T/A	1.00 (0.83-1.20)	1.00
8	*PINX1*	rs11776767	C/C	1.00 (0.83-1.21)	1.00

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the TG-raising allele in the combined HTG cohort.

Table 4.2b. Replication of TG-associated loci in patients with HLP type 5.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 5 (n=180)	
				OR (95% CI)	P-value
11	*APOA5*	rs964184	G/G	5.65 (3.92-8.14)	$1.81X10^{-20}$
2	*GCKR*	rs1260326	T/T	1.60 (1.19-2.15)	$1.9X10^{-3}$
8	*LPL*	rs12678919	G/A	2.79 (1.43-5.44)	$2.7X10^{-3}$
8	*TRIB1*	rs2954029	T/A	1.43 (1.06-1.92)	0.019
1	*ANGPTL3*	rs2131925	G/T	1.85 (1.31-2.61)	$4.7X10^{-4}$
7	*MLXIPL*	rs7811265	G/A	1.70 (1.10-2.62)	0.017
4	*KLHL8*	rs442177	G/T	1.36 (1.02-1.82)	0.038
10	*CYP26A1*	rs2068888	A/G	1.44 (1.08-1.92)	0.013
19	*CILP2*	rs10401969	C/T	2.29 (1.18-4.43)	0.014
2	*APOB*	rs1042034	C/T	1.19 (0.84-1.70)	0.32
5	*TIMD4*	rs1553318	G/C	1.14 (0.85-1.54)	0.39
11	*FADS1*	rs174546	T/T	1.19 (0.88-1.59)	0.25
16	*CETP*	rs7205804	A/G	1.37 (1.01-1.85)	0.046
2	*IRS1*	rs2943645	C/T	1.20 (0.89-1.61)	0.24
6	*HLA*	rs2247056	T/C	1.29 (0.92-1.80)	0.13
1	*GALNT2*	rs1321257	G/G	1.27 (0.95-1.70)	0.11
5	*MAP3K1*	rs9686661	T/T	1.04 (0.73-1.47)	0.84
15	*LIPC*	rs261342	G/G	0.91 (0.64-1.28)	0.58
15	*CAPN3*	rs2412710	A/A	1.66 (0.74-3.72)	0.22
2	*COBLL1*	rs10195252	C/T	1.00 (0.75-1.33)	0.98
12	*LRP1*	rs11613352	T/C	1.51 (1.04-2.19)	0.030
7	*TYW1B*	rs13238203	T/C	0.75 (0.30-1.9)	0.54
8	*NAT2*	rs1495743	G/G	0.92 (0.65-1.31)	0.65
3	*MSL2L1*	rs645040	G/T	0.92 (0.66-1.27)	0.61
16	*CTF1*	rs11649653	G/C	0.86 (0.64-1.14)	0.29
20	*PLTP*	rs4810479	C/C	1.27 (0.91-1.79)	0.16
19	*APOE*	rs439401	T/C	0.91 (0.62-1.34)	0.64
22	*PLA2G6*	rs5756931	C/T	0.97 (0.73-1.30)	0.85
12	*ZNF664*	rs12310367	G/A	1.17 (0.86-1.59)	0.31
15	*FRMD5*	rs2929282	T/T	0.93 (0.46-1.89)	0.84
10	*JMJD1C*	rs10761731	T/A	1.01 (0.76-1.35)	0.94
8	*PINX1*	rs11776767	C/C	0.98 (0.73-1.33)	0.92

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the TG-raising allele in the combined HTG cohort.

Table 4.2c. Replication of TG-associated loci in patients with HLP type 4.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 4 (n=128)	
				OR (95% CI)	P-value
11	*APOA5*	rs964184	G/G	3.71 (2.56-5.39)	$4.9X10^{-12}$
2	*GCKR*	rs1260326	T/T	1.79 (1.29-2.48)	$5.3X10^{-4}$
8	*LPL*	rs12678919	G/A	5.78 (2.10-15.9)	$7.0X10^{-4}$
8	*TRIB1*	rs2954029	T/A	1.37 (0.99-1.89)	0.059
1	*ANGPTL3*	rs2131925	G/T	0.98 (0.70-1.39)	0.93
7	*MLXIPL*	rs7811265	G/A	2.09 (1.25-3.49)	$4.8X10^{-3}$
4	*KLHL8*	rs442177	G/T	1.19 (0.86-1.64)	0.29
10	*CYP26A1*	rs2068888	A/G	1.02 (0.75-1.40)	0.88
19	*CILP2*	rs10401969	C/T	1.53 (0.79-2.95)	0.20
2	*APOB*	rs1042034	C/T	0.88 (0.61-1.28)	0.50
5	*TIMD4*	rs1553318	G/C	1.22 (0.88-1.69)	0.24
11	*FADS1*	rs174546	T/T	1.33 (0.96-1.84)	0.088
16	*CETP*	rs7205804	A/G	1.19 (0.85-1.65)	0.31
2	*IRS1*	rs2943645	C/T	1.42 (1.00-2.00)	0.049
6	*HLA*	rs2247056	T/C	1.12 (0.77-1.62)	0.55
1	*GALNT2*	rs1321257	G/G	1.03 (0.74-1.43)	0.87
5	*MAP3K1*	rs9686661	T/T	0.97 (0.66-1.44)	0.90
15	*LIPC*	rs261342	G/G	0.86 (0.59-1.27)	0.45
15	*CAPN3*	rs2412710	A/A	0.90 (0.32-2.55)	0.84
2	*COBLL1*	rs10195252	C/T	1.42 (1.02-1.99)	0.039
12	*LRP1*	rs11613352	T/C	1.02 (0.69-1.50)	0.92
7	*TYW1B*	rs13238203	T/C	1.64 (0.40-6.69)	0.49
8	*NAT2*	rs1495743	G/G	1.11 (0.76-1.62)	0.61
3	*MSL2L1*	rs645040	G/T	1.31 (0.88-1.96)	0.18
16	*CTF1*	rs11649653	G/C	1.33 (0.95-1.86)	0.097
20	*PLTP*	rs4810479	C/C	0.96 (0.65-1.42)	0.85
19	*APOE*	rs439401	T/C	1.00 (0.66-1.54)	0.98
22	*PLA2G6*	rs5756931	C/T	1.06 (0.77-1.47)	0.71
12	*ZNF664*	rs12310367	G/A	0.91 (0.66-1.26)	0.58
15	*FRMD5*	rs2929282	T/T	0.68 (0.29-1.55)	0.35
10	*JMJD1C*	rs10761731	T/A	0.98 (0.71-1.35)	0.89
8	*PINX1*	rs11776767	C/C	0.95 (0.68-1.33)	0.77

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the TG-raising allele in the combined HTG cohort.

Table 4.2d. Replication of TG-associated loci in patients with HLP type 3.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 3 (n=37)	
				OR (95% CI)	P-value
11	*APOA5*	rs964184	G/G	5.02 (2.51-10.1)	$5.3X10^{-6}$
2	*GCKR*	rs1260326	T/T	0.66 (0.35-1.22)	0.18
8	*LPL*	rs12678919	G/A	1.34 (0.45-3.99)	0.60
8	*TRIB1*	rs2954029	T/A	1.60 (0.89-2.85)	0.12
1	*ANGPTL3*	rs2131925	G/T	1.16 (0.63-2.17)	0.63
7	*MLXIPL*	rs7811265	G/A	0.82 (0.40-1.68)	0.59
4	*KLHL8*	rs442177	G/T	1.86 (0.99-3.49)	0.052
10	*CYP26A1*	rs2068888	A/G	0.96 (0.53-1.74)	0.89
19	*CILP2*	rs10401969	C/T	2.94 (0.78-11.1)	0.11
2	*APOB*	rs1042034	C/T	1.92 (0.83-4.46)	0.13
5	*TIMD4*	rs1553318	G/C	1.07 (0.60-1.92)	0.82
11	*FADS1*	rs174546	T/T	1.08 (0.59-1.98)	0.80
16	*CETP*	rs7205804	A/G	1.04 (0.56-1.90)	0.91
2	*IRS1*	rs2943645	C/T	1.07 (0.59-1.95)	0.83
6	*HLA*	rs2247056	T/C	0.86 (0.45-1.66)	0.66
1	*GALNT2*	rs1321257	G/G	1.07 (0.59-1.93)	0.82
5	*MAP3K1*	rs9686661	T/T	1.68 (0.88-3.21)	0.12
15	*LIPC*	rs261342	G/G	1.26 (0.68-2.33)	0.46
15	*CAPN3*	rs2412710	A/A	0.83 (0.11-6.27)	0.86
2	*COBLL1*	rs10195252	C/T	1.89 (1.00-3.56)	0.049
12	*LRP1*	rs11613352	T/C	0.97 (0.49-1.92)	0.93
7	*TYW1B*	rs13238203	T/C	NC	NC
8	*NAT2*	rs1495743	G/G	2.16 (1.18-3.93)	0.012
3	*MSL2L1*	rs645040	G/T	1.22 (0.62-2.42)	0.56
16	*CTF1*	rs11649653	G/C	1.47 (0.81-2.64)	0.20
20	*PLTP*	rs4810479	C/C	1.15 (0.57-2.32)	0.70
19	*APOE*	rs439401	T/C	2.07 (0.96-4.43)	0.062
22	*PLA2G6*	rs5756931	C/T	0.77 (0.44-1.38)	0.38
12	*ZNF664*	rs12310367	G/A	1.45 (0.80-2.62)	0.22
15	*FRMD5*	rs2929282	T/T	NC	NC
10	*JMJD1C*	rs10761731	T/A	1.23 (0.67-2.24)	0.51
8	*PINX1*	rs11776767	C/C	1.38 (0.77-2.48)	0.28

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the TG-raising allele in the combined HTG cohort. [1]Variant is monomorphic in HTG cases.

Table 4.2e. Replication of TG-associated loci in patients with HLP type 2B.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 2B (n=159)	
				OR (95% CI)	P-value
11	*APOA5*	rs964184	G/G	2.57 (1.82-3.63)	$8.2X10^{-8}$
2	*GCKR*	rs1260326	T/T	1.87 (1.41-2.47)	$1.3X10^{-5}$
8	*LPL*	rs12678919	G/A	1.48 (0.90-2.43)	0.12
8	*TRIB1*	rs2954029	T/A	1.60 (1.20-2.12)	$1.2X10^{-3}$
1	*ANGPTL3*	rs2131925	G/T	1.63 (1.19-2.24)	$2.3X10^{-3}$
7	*MLXIPL*	rs7811265	G/A	1.42 (0.97-2.08)	0.073
4	*KLHL8*	rs442177	G/T	1.47 (1.11-1.96)	$7.6X10^{-3}$
10	*CYP26A1*	rs2068888	A/G	1.27 (0.96-1.68)	0.093
19	*CILP2*	rs10401969	C/T	2.07 (1.13-3.79)	0.018
2	*APOB*	rs1042034	C/T	1.60 (1.11-2.31)	0.01
5	*TIMD4*	rs1553318	G/C	1.38 (1.03-1.84)	0.032
11	*FADS1*	rs174546	T/T	1.16 (0.88-1.53)	0.29
16	*CETP*	rs7205804	A/G	1.17 (0.88-1.55)	0.28
2	*IRS1*	rs2943645	C/T	1.16 (0.87-1.53)	0.32
6	*HLA*	rs2247056	T/C	1.30 (0.95-1.77)	0.099
1	*GALNT2*	rs1321257	G/G	1.07 (0.81-1.42)	0.63
5	*MAP3K1*	rs9686661	T/T	1.40 (1.01-1.93)	0.041
15	*LIPC*	rs261342	G/G	0.8 (0.57-1.12)	0.19
15	*CAPN3*	rs2412710	A/A	1.37 (0.60-3.10)	0.46
2	*COBLL1*	rs10195252	C/T	0.95 (0.72-1.26)	0.73
12	*LRP1*	rs11613352	T/C	1.09 (0.78-1.53)	0.60
7	*TYW1B*	rs13238203	T/C	1.22 (0.41-3.62)	0.72
8	*NAT2*	rs1495743	G/G	0.93 (0.67-1.29)	0.65
3	*MSL2L1*	rs645040	G/T	1.23 (0.87-1.73)	0.24
16	*CTF1*	rs11649653	G/C	1.01 (0.77-1.33)	0.95
20	*PLTP*	rs4810479	C/C	0.95 (0.67-1.34)	0.77
19	*APOE*	rs439401	T/C	0.87 (0.60-1.24)	0.44
22	*PLA2G6*	rs5756931	C/T	1.00 (0.76-1.31)	0.98
12	*ZNF664*	rs12310367	G/A	0.85 (0.64-1.12)	0.24
15	*FRMD5*	rs2929282	T/T	1.38 (0.77-2.47)	0.28
10	*JMJD1C*	rs10761731	T/A	1.13 (0.85-1.49)	0.40
8	*PINX1*	rs11776767	C/C	1.14 (0.86-1.52)	0.35

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the TG-raising allele in the combined HTG cohort. [1]Variant is monomorphic in HTG cases.

common variants in TG-associated loci increase susceptibility to HTG and the HLP phenotypes.

For HDL-C-associated loci (**Table 4.3 a-e**), we replicated associations at four loci with HTG. The replicated loci were predominantly jointly associated with plasma TG as the lead trait, including *MLXIPL*, *TRIB1*, and *FADS1*, but we also replicated the HDL-C associated *IRS1* locus. Generally, there was a trend towards association at additional loci as 31/47 loci had directions of effect concordant with estimates from GLGC (P=0.01). Few loci were replicated among individual HLP-HTG types, although concordant directions of effect were over-represented among HLP types 4 and 5 (P=0.005 and P=0.05, respectively), but not HLP types 2B and 3 (P=0.11 and P=0.11, respectively). Together, these data suggest that common variants in some HDL-C-associated loci contribute to the phenotypic spectrum within HTG.

For LDL-C-associated loci (**Table 4.4 a-e**), we replicated associations at nine loci with HTG. The strongest association was at the *DNAH11* locus, while remaining loci had similarly large effects (OR=1.2-1.4) but with marginal statistical significance. Generally, there was a trend towards association at additional loci as 25/37 loci had directions of effect concordant with estimates from GLGC (P=0.01). Among HLP-HTG phenotypes, few loci were replicated in phenotypes other than HLP type 2B; however concordant directions of effect were over-represented among HLP types 2B, 4 and 5 (P=0.0009, P=0.04, P=0.02, respectively). Together, these data suggest that common variants in some LDL-associated loci also contribute to the phenotypic spectrum within HTG.

Table 4.3a. Replication of HDL-associated loci in patients with polygenic HTG.

CHR	Gene	SNP	Minor allele/ modeled allele	All HTG (n=504)	
				OR (95% CI)	P-value
7	*MLXIPL*	rs17145738	T/C	2.28 (1.63-3.19)	$1.5X10^{-6}$
8	*TRIB1*	rs10808546	T/C	1.46 (1.21-1.77)	$1.0X10^{-4}$
2	*IRS1*	rs1515100	C/A	1.28 (1.05-1.55)	0.015
11	*FADS1*	rs174601	T/T	1.25 (1.04-1.50)	0.019
19	*APOE*	rs4420638	G/G	0.62 (0.4-0.95)	0.029
16	*CETP*	rs3764261	A/C	1.23 (1.00-1.50)	0.051
6	*CITED2*	rs605066	C/C	1.20 (1.00-1.45)	0.052
1	*GALNT2*	rs4846914	G/G	1.21 (1.00-1.45)	0.052
9	*TTC39B*	rs643531	C/C	0.77 (0.59-1.01)	0.059
6	*LPA*	rs1084651	A/A	0.83 (0.64-1.07)	0.16
19	*ANGPTL4*	rs7255436	C/C	1.15 (0.94-1.40)	0.18
20	*PLTP*	rs6065906	C/C	1.17 (0.92-1.48)	0.20
1	*ZNF648*	rs1689800	G/G	1.13 (0.93-1.37)	0.21
12	*MVK*	rs7134594	C/C	0.90 (0.75-1.08)	0.24
12	*LRP1*	rs3741414	T/C	1.14 (0.91-1.43)	0.25
11	*UBASH3B*	rs7115089	G/C	1.12 (0.93-1.34)	0.25
20	*HNF4A*	rs1800961	T/T	0.75 (0.44-1.28)	0.29
7	*KLF14*	rs4731702	T/C	0.91 (0.76-1.09)	0.31
8	*PPP1R3B*	rs9987289	A/A	1.17 (0.84-1.64)	0.35
22	*UBE2L3*	rs181362	T/T	0.90 (0.71-1.14)	0.39
18	*LIPG*	rs7241918	G/G	0.90 (0.71-1.15)	0.40
12	*SBNO1*	rs4759375	T/C	1.16 (0.78-1.72)	0.47
17	*PGS1*	rs4082919	G/G	1.07 (0.89-1.28)	0.47
15	*LIPC*	rs1532085	A/G	1.07 (0.89-1.29)	0.49
4	*SLC39A8*	rs13107325	T/T	0.89 (0.64-1.25)	0.49
15	*LACTB*	rs2652834	A/A	0.93 (0.73-1.19)	0.56
18	*MC4R*	rs12967135	A/A	0.94 (0.76-1.18)	0.61
5	*ARL15*	rs6450176	A/A	0.95 (0.77-1.17)	0.61
2	*COBLL1*	rs12328675	C/T	1.07 (0.81-1.40)	0.64
11	*LRP4*	rs3136441	C/T	0.94 (0.73-1.23)	0.67
19	*LOC55908*	rs737337	C/C	1.06 (0.77-1.45)	0.72
16	*LCAT*	rs16942887	A/G	1.05 (0.80-1.38)	0.73
6	*C6orf106*	rs2814944	A/A	1.05 (0.80-1.38)	0.73
1	*PABPC4*	rs4660293	G/G	1.04 (0.84-1.28)	0.74
12	*SCARB1*	rs838880	C/T	1.02 (0.84-1.24)	0.82
11	*AMPD3*	rs2923084	G/G	1.02 (0.81-1.28)	0.86
17	*STARD3*	rs881844	C/C	1.02 (0.84-1.23)	0.86
16	*CMIP*	rs2925979	T/T	0.99 (0.81-1.21)	0.91
8	*TRPS1*	rs2293889	T/T	0.99 (0.83-1.19)	0.92
12	*PDE3A*	rs7134375	A/C	0.99 (0.83-1.19)	0.95
12	*ZNF664*	rs4765127	T/G	1.00 (0.82-1.20)	0.96
17	*ABCA8*	rs4148008	G/G	1.00 (0.82-1.21)	0.97
9	*ABCA1*	rs1883025	T/T	1.00 (0.81-1.22)	0.98
19	*LILRA3*	rs386000	C/G	1.00 (0.80-1.25)	1.00

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the HDL-lowering allele in the combined HTG cohort.

Table 4.3b. Replication of HDL-associated loci in patients with HLP type 5.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 5 (n=180)	
				OR (95% CI)	P-value
7	*MLXIPL*	rs17145738	T/C	1.93 (1.16-3.19)	0.011
8	*TRIB1*	rs10808546	T/C	1.32 (0.98-1.77)	0.07
2	*IRS1*	rs1515100	C/A	1.22 (0.90-1.65)	0.20
11	*FADS1*	rs174601	T/T	1.21 (0.91-1.62)	0.20
19	*APOE*	rs4420638	G/G	0.73 (0.38-1.38)	0.33
16	*CETP*	rs3764261	A/C	1.26 (0.91-1.73)	0.16
6	*CITED2*	rs605066	C/C	1.13 (0.85-1.50)	0.41
1	*GALNT2*	rs4846914	G/G	1.29 (0.97-1.73)	0.084
9	*TTC39B*	rs643531	C/C	1.05 (0.70-1.56)	0.82
6	*LPA*	rs1084651	A/A	0.82 (0.54-1.23)	0.33
19	*ANGPTL4*	rs7255436	C/C	0.93 (0.68-1.26)	0.63
20	*PLTP*	rs6065906	C/C	1.38 (0.96-1.97)	0.082
1	*ZNF648*	rs1689800	G/G	1.05 (0.77-1.43)	0.76
12	*MVK*	rs7134594	C/C	0.87 (0.66-1.16)	0.35
12	*LRP1*	rs3741414	T/C	1.44 (1.00-2.07)	0.051
11	*UBASH3B*	rs7115089	G/C	1.24 (0.92-1.67)	0.16
20	*HNF4A*	rs1800961	T/T	0.42 (0.14-1.21)	0.11
7	*KLF14*	rs4731702	T/C	1.00 (0.75-1.33)	0.99
8	*PPP1R3B*	rs9987289	A/A	1.09 (0.65-1.83)	0.76
22	*UBE2L3*	rs181362	T/T	0.85 (0.59-1.23)	0.38
18	*LIPG*	rs7241918	G/G	1.02 (0.70-1.48)	0.91
12	*SBNO1*	rs4759375	T/C	1.50 (0.75-3.01)	0.25
17	*PGS1*	rs4082919	G/G	1.17 (0.89-1.55)	0.26
15	*LIPC*	rs1532085	A/G	0.99 (0.74-1.32)	0.95
4	*SLC39A8*	rs13107325	T/T	0.60 (0.33-1.09)	0.093
15	*LACTB*	rs2652834	A/A	0.90 (0.60-1.33)	0.59
18	*MC4R*	rs12967135	A/A	0.91 (0.64-1.28)	0.58
5	*ARL15*	rs6450176	A/A	0.86 (0.61-1.21)	0.38
2	*COBLL1*	rs12328675	C/T	1.47 (0.93-2.34)	0.10
11	*LRP4*	rs3136441	C/T	1.26 (0.81-1.95)	0.30
19	*LOC55908*	rs737337	C/C	0.93 (0.55-1.56)	0.78
16	*LCAT*	rs16942887	A/G	1.10 (0.71-1.72)	0.66
6	*C6orf106*	rs2814944	A/A	1.11 (0.73-1.69)	0.63
1	*PABPC4*	rs4660293	G/G	1.18 (0.85-1.63)	0.32
12	*SCARB1*	rs838880	C/T	0.92 (0.68-1.23)	0.56
11	*AMPD3*	rs2923084	G/G	1.04 (0.72-1.49)	0.85
17	*STARD3*	rs881844	C/C	0.84 (0.61-1.14)	0.26
16	*CMIP*	rs2925979	T/T	1.12 (0.83-1.51)	0.47
8	*TRPS1*	rs2293889	T/T	0.99 (0.75-1.31)	0.94
12	*PDE3A*	rs7134375	A/C	0.88 (0.66-1.18)	0.40
12	*ZNF664*	rs4765127	T/G	1.16 (0.86-1.58)	0.34
17	*ABCA8*	rs4148008	G/G	0.92 (0.68-1.26)	0.61
9	*ABCA1*	rs1883025	T/T	0.86 (0.62-1.21)	0.39
19	*LILRA3*	rs386000	C/G	1.02 (0.72-1.44)	0.90

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the HDL-lowering allele in the combined HTG cohort.

Table 4.3c. Replication of HDL-associated loci in patients with HLP type 4.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 4 (n=128)	
				OR (95% CI)	P-value
7	*MLXIPL*	rs17145738	T/C	3.63 (1.75-7.54)	$5.4X10^{-4}$
8	*TRIB1*	rs10808546	T/C	1.43 (1.03-1.98)	0.032
2	*IRS1*	rs1515100	C/A	1.53 (1.07-2.18)	0.018
11	*FADS1*	rs174601	T/T	1.46 (1.05-2.01)	0.023
19	*APOE*	rs4420638	G/G	0.50 (0.22-1.14)	0.10
16	*CETP*	rs3764261	A/C	1.26 (0.89-1.80)	0.20
6	*CITED2*	rs605066	C/C	1.21 (0.87-1.66)	0.25
1	*GALNT2*	rs4846914	G/G	1.14 (0.82-1.58)	0.45
9	*TTC39B*	rs643531	C/C	0.68 (0.41-1.12)	0.13
6	*LPA*	rs1084651	A/A	1.03 (0.67-1.58)	0.91
19	*ANGPTL4*	rs7255436	C/C	1.20 (0.85-1.69)	0.31
20	*PLTP*	rs6065906	C/C	1.10 (0.72-1.67)	0.66
1	*ZNF648*	rs1689800	G/G	0.96 (0.68-1.36)	0.81
12	*MVK*	rs7134594	C/C	1.08 (0.79-1.48)	0.64
12	*LRP1*	rs3741414	T/C	0.99 (0.68-1.44)	0.97
11	*UBASH3B*	rs7115089	G/C	1.26 (0.90-1.76)	0.17
20	*HNF4A*	rs1800961	T/T	1.52 (0.70-3.29)	0.28
7	*KLF14*	rs4731702	T/C	0.80 (0.59-1.10)	0.17
8	*PPP1R3B*	rs9987289	A/A	1.03 (0.58-1.83)	0.92
22	*UBE2L3*	rs181362	T/T	1.15 (0.77-1.69)	0.50
18	*LIPG*	rs7241918	G/G	0.90 (0.58-1.37)	0.61
12	*SBNO1*	rs4759375	T/C	1.71 (0.78-3.77)	0.18
17	*PGS1*	rs4082919	G/G	1.22 (0.90-1.67)	0.20
15	*LIPC*	rs1532085	A/G	1.51 (1.08-2.11)	0.016
4	*SLC39A8*	rs13107325	T/T	1.15 (0.67-1.99)	0.61
15	*LACTB*	rs2652834	A/A	1.18 (0.79-1.76)	0.43
18	*MC4R*	rs12967135	A/A	1.11 (0.76-1.62)	0.59
5	*ARL15*	rs6450176	A/A	1.00 (0.70-1.42)	0.99
2	*COBLL1*	rs12328675	C/T	0.93 (0.60-1.45)	0.76
11	*LRP4*	rs3136441	C/T	0.58 (0.38-0.87)	$8.0X10^{-3}$
19	*LOC55908*	rs737337	C/C	0.64 (0.34-1.22)	0.17
16	*LCAT*	rs16942887	A/G	1.15 (0.71-1.88)	0.57
6	*C6orf106*	rs2814944	A/A	0.95 (0.57-1.56)	0.83
1	*PABPC4*	rs4660293	G/G	0.96 (0.67-1.39)	0.83
12	*SCARB1*	rs838880	C/T	1.01 (0.72-1.41)	0.97
11	*AMPD3*	rs2923084	G/G	1.05 (0.71-1.57)	0.81
17	*STARD3*	rs881844	C/C	1.26 (0.91-1.74)	0.17
16	*CMIP*	rs2925979	T/T	1.13 (0.80-1.59)	0.50
8	*TRPS1*	rs2293889	T/T	0.97 (0.71-1.32)	0.85
12	*PDE3A*	rs7134375	A/C	1.26 (0.91-1.76)	0.17
12	*ZNF664*	rs4765127	T/G	0.91 (0.66-1.26)	0.57
17	*ABCA8*	rs4148008	G/G	0.99 (0.70-1.39)	0.96
9	*ABCA1*	rs1883025	T/T	1.20 (0.85-1.69)	0.30
19	*LILRA3*	rs386000	C/G	1.11 (0.74-1.65)	0.62

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the HDL-lowering in the combined HTG cohort.

Table 4.3d. Replication of HDL-associated loci in patients with HLP type 3.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 3 (n=37)	
				OR (95% CI)	P-value
7	*MLXIPL*	rs17145738	T/C	1.38 (0.55-3.49)	0.49
8	*TRIB1*	rs10808546	T/C	1.56 (0.87-2.79)	0.14
2	*IRS1*	rs1515100	C/A	0.98 (0.54-1.79)	0.95
11	*FADS1*	rs174601	T/T	1.16 (0.64-2.12)	0.63
19	*APOE*	rs4420638	G/G	0.58 (0.13-2.52)	0.47
16	*CETP*	rs3764261	A/C	0.82 (0.44-1.52)	0.52
6	*CITED2*	rs605066	C/C	1.50 (0.84-2.67)	0.17
1	*GALNT2*	rs4846914	G/G	1.16 (0.64-2.09)	0.63
9	*TTC39B*	rs643531	C/C	0.49 (0.17-1.41)	0.19
6	*LPA*	rs1084651	A/A	0.57 (0.23-1.40)	0.22
19	*ANGPTL4*	rs7255436	C/C	1.32 (0.71-2.49)	0.38
20	*PLTP*	rs6065906	C/C	0.62 (0.25-1.56)	0.31
1	*ZNF648*	rs1689800	G/G	1.30 (0.71-2.39)	0.39
12	*MVK*	rs7134594	C/C	1.10 (0.63-1.91)	0.75
12	*LRP1*	rs3741414	T/C	1.12 (0.56-2.26)	0.75
11	*UBASH3B*	rs7115089	G/C	1.23 (0.70-2.17)	0.48
20	*HNF4A*	rs1800961	T/T	0.56 (0.07-4.33)	0.58
7	*KLF14*	rs4731702	T/C	0.76 (0.43-1.33)	0.34
8	*PPP1R3B*	rs9987289	A/A	2.59 (1.18-5.68)	0.018
22	*UBE2L3*	rs181362	T/T	0.57 (0.25-1.29)	0.18
18	*LIPG*	rs7241918	G/G	0.68 (0.29-1.57)	0.36
12	*SBNO1*	rs4759375	T/C	0.48 (0.18-1.27)	0.14
17	*PGS1*	rs4082919	G/G	0.82 (0.46-1.45)	0.49
15	*LIPC*	rs1532085	A/G	0.78 (0.45-1.35)	0.37
4	*SLC39A8*	rs13107325	T/T	0.75 (0.25-2.24)	0.60
15	*LACTB*	rs2652834	A/A	0.77 (0.35-1.68)	0.51
18	*MC4R*	rs12967135	A/A	1.85 (0.99-3.48)	0.054
5	*ARL15*	rs6450176	A/A	0.73 (0.36-1.49)	0.38
2	*COBLL1*	rs12328675	C/T	1.13 (0.47-2.68)	0.79
11	*LRP4*	rs3136441	C/T	1.08 (0.45-2.57)	0.87
19	*LOC55908*	rs737337	C/C	1.87 (0.86-4.07)	0.12
16	*LCAT*	rs16942887	A/G	0.67 (0.31-1.44)	0.30
6	*C6orf106*	rs2814944	A/A	0.83 (0.34-2.03)	0.68
1	*PABPC4*	rs4660293	G/G	0.56 (0.27-1.18)	0.13
12	*SCARB1*	rs838880	C/T	1.11 (0.62-1.99)	0.73
11	*AMPD3*	rs2923084	G/G	2.20 (1.17-4.12)	0.014
17	*STARD3*	rs881844	C/C	1.02 (0.55-1.89)	0.96
16	*CMIP*	rs2925979	T/T	0.62 (0.31-1.24)	0.17
8	*TRPS1*	rs2293889	T/T	1.21 (0.68-2.15)	0.51
12	*PDE3A*	rs7134375	A/C	0.98 (0.55-1.76)	0.95
12	*ZNF664*	rs4765127	T/G	1.44 (0.79-2.63)	0.24
17	*ABCA8*	rs4148008	G/G	1.58 (0.90-2.78)	0.11
9	*ABCA1*	rs1883025	T/T	0.61 (0.29-1.27)	0.18
19	*LILRA3*	rs386000	C/G	0.92 (0.47-1.79)	0.80

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the HDL-lowering allele in the combined HTG cohort.

Table 4.3e. Replication of HDL-associated loci in patients with HLP type 2B.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 2B (n=159)	
				OR (95% CI)	P-value
7	*MLXIPL*	rs17145738	T/C	2.17 (1.32-3.57)	$2.3X10^{-3}$
8	*TRIB1*	rs10808546	T/C	1.57 (1.18-2.09)	$1.9X10^{-3}$
2	*IRS1*	rs1515100	C/A	1.26 (0.95-1.68)	0.11
11	*FADS1*	rs174601	T/T	1.17 (0.89-1.54)	0.27
19	*APOE*	rs4420638	G/G	0.65 (0.34-1.23)	0.19
16	*CETP*	rs3764261	A/C	1.36 (1.00-1.85)	0.050
6	*CITED2*	rs605066	C/C	1.25 (0.95-1.64)	0.11
1	*GALNT2*	rs4846914	G/G	1.09 (0.83-1.45)	0.53
9	*TTC39B*	rs643531	C/C	0.78 (0.52-1.18)	0.25
6	*LPA*	rs1084651	A/A	0.81 (0.55-1.19)	0.28
19	*ANGPTL4*	rs7255436	C/C	1.25 (0.93-1.69)	0.14
20	*PLTP*	rs6065906	C/C	1.14 (0.80-1.64)	0.47
1	*ZNF648*	rs1689800	G/G	1.42 (1.07-1.90)	0.017
12	*MVK*	rs7134594	C/C	0.83 (0.63-1.09)	0.19
12	*LRP1*	rs3741414	T/C	1.15 (0.82-1.60)	0.42
11	*UBASH3B*	rs7115089	G/C	0.88 (0.67-1.15)	0.35
20	*HNF4A*	rs1800961	T/T	0.83 (0.38-1.78)	0.62
7	*KLF14*	rs4731702	T/C	0.98 (0.75-1.28)	0.83
8	*PPP1R3B*	rs9987289	A/A	0.86 (0.50-1.48)	0.57
22	*UBE2L3*	rs181362	T/T	0.80 (0.55-1.15)	0.23
18	*LIPG*	rs7241918	G/G	0.90 (0.63-1.29)	0.58
12	*SBNO1*	rs4759375	T/C	0.71 (0.42-1.19)	0.19
17	*PGS1*	rs4082919	G/G	1.00 (0.76-1.31)	1.00
15	*LIPC*	rs1532085	A/G	1.05 (0.80-1.38)	0.74
4	*SLC39A8*	rs13107325	T/T	1.00 (0.62-1.64)	0.99
15	*LACTB*	rs2652834	A/A	0.78 (0.53-1.14)	0.20
18	*MC4R*	rs12967135	A/A	0.86 (0.61-1.21)	0.38
5	*ARL15*	rs6450176	A/A	0.93 (0.68-1.26)	0.64
2	*COBLL1*	rs12328675	C/T	0.96 (0.65-1.41)	0.82
11	*LRP4*	rs3136441	C/T	1.02 (0.69-1.49)	0.94
19	*LOC55908*	rs737337	C/C	1.21 (0.78-1.87)	0.40
16	*LCAT*	rs16942887	A/G	1.06 (0.71-1.59)	0.76
6	*C6orf106*	rs2814944	A/A	1.06 (0.71-1.60)	0.76
1	*PABPC4*	rs4660293	G/G	0.96 (0.70-1.33)	0.82
12	*SCARB1*	rs838880	C/T	1.15 (0.86-1.55)	0.34
11	*AMPD3*	rs2923084	G/G	0.95 (0.67-1.35)	0.77
17	*STARD3*	rs881844	C/C	1.12 (0.84-1.48)	0.44
16	*CMIP*	rs2925979	T/T	1.02 (0.76-1.38)	0.89
8	*TRPS1*	rs2293889	T/T	1.10 (0.84-1.44)	0.50
12	*PDE3A*	rs7134375	A/C	0.87 (0.66-1.14)	0.31
12	*ZNF664*	rs4765127	T/G	0.88 (0.67-1.17)	0.38
17	*ABCA8*	rs4148008	G/G	0.96 (0.71-1.28)	0.77
9	*ABCA1*	rs1883025	T/T	1.15 (0.85-1.55)	0.36
19	*LILRA3*	rs386000	C/G	0.98 (0.71-1.35)	0.90

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the HDL-lowering allele in the combined HTG cohort.

Table 4.4a. Replication of LDL-associated loci in patients with polygenic HTG.

CHR	Gene	SNP	Minor allele/ modeled allele	All HTG (n=504)	
				OR (95% CI)	P-value
8	*TRIB1*	rs2954022	A/C	1.48 (1.23-1.8)	$5.3X10^{-5}$
1	*ANGPTL3*	rs3850634	G/T	1.49 (1.21-1.83)	$1.7X10^{-4}$
7	*DNAH11*	rs12670798	C/C	1.39 (1.13-1.70)	$1.7X10^{-3}$
2	*ABCG5/8*	rs4299376	G/G	1.33 (1.07-1.65)	0.011
19	*LDLR*	rs6511720	T/G	1.84 (1.12-3.03)	0.017
9	*ABO*	rs649129	T/T	1.27 (1.04-1.56)	0.021
5	*HMGCR*	rs12916	C/C	1.23 (1.02-1.48)	0.028
11	*FADS1*	rs174583	T/C	0.81 (0.68-0.98)	0.029
19	*APOE*	rs4420638	G/G	0.62 (0.4-0.95)	0.029
16	*CETP*	rs247616	T/C	1.24 (1.01-1.51)	0.041
1	*PCSK9*	rs2479409	G/G	1.21 (1.00-1.45)	0.045
5	*TIMD4*	rs6882076	T/C	1.21 (1.00-1.46)	0.051
1	*SORT1*	rs629301	G/T	1.18 (0.94-1.49)	0.16
7	*NPC1L1*	rs217386	A/G	1.13 (0.94-1.35)	0.19
1	*IRF2BP2*	rs514230	A/T	1.13 (0.94-1.35)	0.20
6	*LPA*	rs1564348	C/C	1.15 (0.90-1.47)	0.27
6	*HLA*	rs3177928	A/A	0.87 (0.68-1.13)	0.30
6	*MYLIP*	rs3757354	T/C	0.90 (0.72-1.13)	0.36
14	*NYNRIN*	rs2332328	T/T	0.92 (0.76-1.11)	0.36
8	*CYP7A1*	rs1030431	A/A	1.09 (0.90-1.32)	0.38
6	*HFE*	rs1800562	A/G	1.19 (0.81-1.75)	0.38
2	*APOB*	rs1367117	A/A	1.09 (0.90-1.32)	0.39
8	*PLEC1*	rs11136341	G/G	1.09 (0.90-1.31)	0.40
6	*FRK*	rs11153594	T/C	1.08 (0.90-1.29)	0.42
10	*GPAM*	rs1129555	A/A	1.08 (0.88-1.32)	0.45
12	*HNF1A*	rs1169288	C/C	1.07 (0.88-1.31)	0.50
17	*OSBPL7*	rs7225700	T/C	0.94 (0.78-1.14)	0.55
8	*PPP1R3B*	rs2126259	T/C	0.92 (0.67-1.26)	0.61
1	*LDLRAP1*	rs12027135	A/T	0.95 (0.79-1.14)	0.61
16	*HPR*	rs2000999	A/A	1.06 (0.85-1.32)	0.62
12	*BRAP*	rs11065987	G/A	0.96 (0.80-1.15)	0.65
1	*MOSC1*	rs2807834	T/G	0.96 (0.79-1.16)	0.66
11	*ST3GAL4*	rs11220462	A/A	0.97 (0.75-1.27)	0.84
20	*MAFB*	rs2902941	G/A	0.98 (0.81-1.19)	0.86
20	*TOP1*	rs909802	T/T	1.00 (0.84-1.20)	0.98

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the LDL-raising allele in the combined HTG cohort.

Table 4.4b. Replication of LDL-associated loci in patients with HLP type 5.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 5 (n=180)	
				OR (95% CI)	P-value
8	*TRIB1*	rs2954022	A/C	1.41 (1.05-1.89)	0.024
1	*ANGPTL3*	rs3850634	G/T	1.77 (1.26-2.49)	$1.1X10^{-3}$
7	*DNAH11*	rs12670798	C/C	1.26 (0.92-1.74)	0.15
2	*ABCG5/8*	rs4299376	G/G	1.18 (0.84-1.67)	0.34
19	*LDLR*	rs6511720	T/G	1.56 (0.75-3.24)	0.23
9	*ABO*	rs649129	T/T	1.34 (0.96-1.85)	0.082
5	*HMGCR*	rs12916	C/C	1.12 (0.84-1.50)	0.44
11	*FADS1*	rs174583	T/C	0.84 (0.62-1.12)	0.23
19	*APOE*	rs4420638	G/G	0.73 (0.38-1.38)	0.33
16	*CETP*	rs247616	T/C	1.23 (0.89-1.68)	0.21
1	*PCSK9*	rs2479409	G/G	1.14 (0.85-1.53)	0.38
5	*TIMD4*	rs6882076	T/C	1.17 (0.87-1.58)	0.30
1	*SORT1*	rs629301	G/T	0.93 (0.65-1.31)	0.66
7	*NPC1L1*	rs217386	A/G	1.02 (0.77-1.34)	0.91
1	*IRF2BP2*	rs514230	A/T	1.26 (0.94-1.67)	0.12
6	*LPA*	rs1564348	C/C	1.11 (0.76-1.63)	0.59
6	*HLA*	rs3177928	A/A	1.02 (0.69-1.50)	0.93
6	*MYLIP*	rs3757354	T/C	0.89 (0.64-1.26)	0.52
14	*NYNRIN*	rs2332328	T/T	0.9 (0.67-1.21)	0.48
8	*CYP7A1*	rs1030431	A/A	1.11 (0.82-1.50)	0.51
6	*HFE*	rs1800562	A/G	1.21 (0.66-2.23)	0.54
2	*APOB*	rs1367117	A/A	1.08 (0.80-1.47)	0.60
8	*PLEC1*	rs11136341	G/G	0.94 (0.70-1.27)	0.71
6	*FRK*	rs11153594	T/C	0.89 (0.67-1.18)	0.41
10	*GPAM*	rs1129555	A/A	0.92 (0.67-1.27)	0.62
12	*HNF1A*	rs1169288	C/C	0.99 (0.72-1.35)	0.94
17	*OSBPL7*	rs7225700	T/C	1.25 (0.92-1.69)	0.15
8	*PPP1R3B*	rs2126259	T/C	1.02 (0.61-1.68)	0.95
1	*LDLRAP1*	rs12027135	A/T	1.28 (0.95-1.72)	0.10
16	*HPR*	rs2000999	A/A	1.13 (0.81-1.58)	0.48
12	*BRAP*	rs11065987	G/A	0.96 (0.72-1.27)	0.76
1	*MOSC1*	rs2807834	T/G	1.16 (0.84-1.60)	0.38
11	*ST3GAL4*	rs11220462	A/A	0.85 (0.55-1.30)	0.44
20	*MAFB*	rs2902941	G/A	0.97 (0.72-1.31)	0.84
20	*TOP1*	rs909802	T/T	0.99 (0.74-1.32)	0.95

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the LDL-raising allele in the combined HTG cohort.

Table 4.4c. Replication of LDL-associated loci in patients with HLP type 4.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 4 (n=128)	
				OR (95% CI)	P-value
8	*TRIB1*	rs2954022	A/C	1.36 (0.98-1.88)	0.064
1	*ANGPTL3*	rs3850634	G/T	0.97 (0.69-1.37)	0.87
7	*DNAH11*	rs12670798	C/C	1.41 (0.98-2.02)	0.062
2	*ABCG5/8*	rs4299376	G/G	1.50 (1.04-2.17)	0.032
19	*LDLR*	rs6511720	T/G	4.35 (1.30-14.6)	0.017
9	*ABO*	rs649129	T/T	1.14 (0.79-1.65)	0.49
5	*HMGCR*	rs12916	C/C	1.27 (0.93-1.74)	0.14
11	*FADS1*	rs174583	T/C	0.72 (0.52-0.99)	0.046
19	*APOE*	rs4420638	G/G	0.50 (0.22-1.14)	0.10
16	*CETP*	rs247616	T/C	1.28 (0.90-1.83)	0.17
1	*PCSK9*	rs2479409	G/G	1.38 (1.00-1.90)	0.052
5	*TIMD4*	rs6882076	T/C	1.15 (0.83-1.59)	0.40
1	*SORT1*	rs629301	G/T	1.05 (0.71-1.55)	0.80
7	*NPC1L1*	rs217386	A/G	1.12 (0.83-1.53)	0.46
1	*IRF2BP2*	rs514230	A/T	1.08 (0.80-1.46)	0.63
6	*LPA*	rs1564348	C/C	1.15 (0.75-1.75)	0.53
6	*HLA*	rs3177928	A/A	0.79 (0.50-1.23)	0.30
6	*MYLIP*	rs3757354	T/C	0.65 (0.45-0.93)	0.017
14	*NYNRIN*	rs2332328	T/T	1.00 (0.73-1.38)	0.99
8	*CYP7A1*	rs1030431	A/A	1.46 (1.06-2.02)	0.02
6	*HFE*	rs1800562	A/G	1.54 (0.73-3.24)	0.25
2	*APOB*	rs1367117	A/A	0.88 (0.62-1.24)	0.45
8	*PLEC1*	rs11136341	G/G	1.19 (0.86-1.64)	0.29
6	*FRK*	rs11153594	T/C	1.27 (0.93-1.74)	0.14
10	*GPAM*	rs1129555	A/A	0.94 (0.66-1.34)	0.74
12	*HNF1A*	rs1169288	C/C	1.06 (0.76-1.50)	0.72
17	*OSBPL7*	rs7225700	T/C	1.01 (0.72-1.42)	0.95
8	*PPP1R3B*	rs2126259	T/C	1.04 (0.60-1.82)	0.88
1	*LDLRAP1*	rs12027135	A/T	0.91 (0.66-1.26)	0.57
16	*HPR*	rs2000999	A/A	0.94 (0.63-1.40)	0.74
12	*BRAP*	rs11065987	G/A	0.97 (0.71-1.32)	0.84
1	*MOSC1*	rs2807834	T/G	0.90 (0.64-1.25)	0.52
11	*ST3GAL4*	rs11220462	A/A	0.82 (0.51-1.34)	0.43
20	*MAFB*	rs2902941	G/A	0.80 (0.58-1.11)	0.19
20	*TOP1*	rs909802	T/T	0.91 (0.66-1.25)	0.56

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the LDL-raising allele in the combined HTG cohort.

Table 4.4d. Replication of LDL-associated loci in patients with HLP type 3.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 3 (n=37)	
				OR (95% CI)	P-value
8	*TRIB1*	rs2954022	A/C	1.59 (0.89-2.84)	0.12
1	*ANGPTL3*	rs3850634	G/T	1.24 (0.66-2.32)	0.50
7	*DNAH11*	rs12670798	C/C	1.30 (0.70-2.43)	0.41
2	*ABCG5/8*	rs4299376	G/G	1.14 (0.59-2.21)	0.69
19	*LDLR*	rs6511720	T/G	NC[1]	NC[1]
9	*ABO*	rs649129	T/T	1.26 (0.66-2.40)	0.49
5	*HMGCR*	rs12916	C/C	1.35 (0.78-2.34)	0.29
11	*FADS1*	rs174583	T/C	0.78 (0.43-1.41)	0.41
19	*APOE*	rs4420638	G/G	0.58 (0.13-2.52)	0.47
16	*CETP*	rs247616	T/C	0.79 (0.42-1.47)	0.46
1	*PCSK9*	rs2479409	G/G	1.32 (0.74-2.37)	0.35
5	*TIMD4*	rs6882076	T/C	0.88 (0.50-1.55)	0.66
1	*SORT1*	rs629301	G/T	1.33 (0.63-2.82)	0.45
7	*NPC1L1*	rs217386	A/G	0.62 (0.35-1.07)	0.087
1	*IRF2BP2*	rs514230	A/T	1.32 (0.74-2.35)	0.34
6	*LPA*	rs1564348	C/C	1.36 (0.67-2.76)	0.40
6	*HLA*	rs3177928	A/A	1.49 (0.72-3.07)	0.28
6	*MYLIP*	rs3757354	T/C	0.96 (0.48-1.91)	0.90
14	*NYNRIN*	rs2332328	T/T	1.50 (0.80-2.82)	0.21
8	*CYP7A1*	rs1030431	A/A	1.22 (0.68-2.18)	0.51
6	*HFE*	rs1800562	A/G	0.36 (0.15-0.85)	0.019
2	*APOB*	rs1367117	A/A	0.77 (0.41-1.46)	0.42
8	*PLEC1*	rs11136341	G/G	1.20 (0.66-2.17)	0.55
6	*FRK*	rs11153594	T/C	0.74 (0.43-1.29)	0.29
10	*GPAM*	rs1129555	A/A	1.50 (0.84-2.67)	0.17
12	*HNF1A*	rs1169288	C/C	1.30 (0.71-2.38)	0.39
17	*OSBPL7*	rs7225700	T/C	0.61 (0.35-1.07)	0.083
8	*PPP1R3B*	rs2126259	T/C	0.49 (0.23-1.05)	0.066
1	*LDLRAP1*	rs12027135	A/T	0.88 (0.50-1.55)	0.65
16	*HPR*	rs2000999	A/A	1.12 (0.57-2.17)	0.74
12	*BRAP*	rs11065987	G/A	1.16 (0.65-2.05)	0.62
1	*MOSC1*	rs2807834	T/G	0.99 (0.53-1.85)	0.98
11	*ST3GAL4*	rs11220462	A/A	0.77 (0.31-1.90)	0.57
20	*MAFB*	rs2902941	G/A	1.08 (0.58-2.01)	0.80
20	*TOP1*	rs909802	T/T	1.00 (0.57-1.76)	0.99

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the LDL-raising allele in the combined HTG cohort. [1]Variant is monomorphic in HTG cases.

Table 4.4e. Replication of LDL-associated loci in patients with HLP type 2B.

CHR	Gene	SNP	Minor allele/ modeled allele	HLP type 2B (n=159)	
				OR (95% CI)	P-value
8	*TRIB1*	rs2954022	A/C	1.58 (1.19-2.10)	$1.5X10^{-3}$
1	*ANGPTL3*	rs3850634	G/T	1.61 (1.17-2.20)	$3. X10^{-3}$
7	*DNAH11*	rs12670798	C/C	1.32 (0.97-1.78)	0.076
2	*ABCG5/8*	rs4299376	G/G	1.47 (1.07-2.02)	0.019
19	*LDLR*	rs6511720	T/G	1.46 (0.74-2.88)	0.27
9	*ABO*	rs649129	T/T	1.53 (1.13-2.07)	$5.7X10^{-3}$
5	*HMGCR*	rs12916	C/C	1.16 (0.88-1.52)	0.29
11	*FADS1*	rs174583	T/C	0.87 (0.66-1.15)	0.33
19	*APOE*	rs4420638	G/G	0.65 (0.34-1.23)	0.19
16	*CETP*	rs247616	T/C	1.36 (1.00-1.85)	0.049
1	*PCSK9*	rs2479409	G/G	1.22 (0.92-1.60)	0.17
5	*TIMD4*	rs6882076	T/C	1.59 (1.18-2.14)	$2.4X10^{-3}$
1	*SORT1*	rs629301	G/T	1.43 (0.99-2.06)	0.059
7	*NPC1L1*	rs217386	A/G	1.39 (1.06-1.82)	0.017
1	*IRF2BP2*	rs514230	A/T	1.03 (0.79-1.35)	0.81
6	*LPA*	rs1564348	C/C	1.16 (0.80-1.67)	0.43
6	*HLA*	rs3177928	A/A	0.79 (0.54-1.17)	0.24
6	*MYLIP*	rs3757354	T/C	1.30 (0.92-1.85)	0.14
14	*NYNRIN*	rs2332328	T/T	0.98 (0.74-1.30)	0.91
8	*CYP7A1*	rs1030431	A/A	0.92 (0.69-1.23)	0.57
6	*HFE*	rs1800562	A/G	1.46 (0.79-2.68)	0.22
2	*APOB*	rs1367117	A/A	1.38 (1.04-1.84)	0.026
8	*PLEC1*	rs11136341	G/G	1.18 (0.89-1.56)	0.26
6	*FRK*	rs11153594	T/C	1.21 (0.92-1.59)	0.18
10	*GPAM*	rs1129555	A/A	1.17 (0.87-1.58)	0.30
12	*HNF1A*	rs1169288	C/C	1.07 (0.79-1.45)	0.66
17	*OSBPL7*	rs7225700	T/C	0.71 (0.54-0.93)	0.014
8	*PPP1R3B*	rs2126259	T/C	1.04 (0.64-1.68)	0.88
1	*LDLRAP1*	rs12027135	A/T	0.89 (0.68-1.18)	0.42
16	*HPR*	rs2000999	A/A	1.00 (0.72-1.39)	1.00
12	*BRAP*	rs11065987	G/A	0.93 (0.71-1.23)	0.61
1	*MOSC1*	rs2807834	T/G	0.87 (0.65-1.16)	0.35
11	*ST3GAL4*	rs11220462	A/A	1.28 (0.89-1.85)	0.18
20	*MAFB*	rs2902941	G/A	1.05 (0.79-1.40)	0.73
20	*TOP1*	rs909802	T/T	1.05 (0.80-1.38)	0.72

CHR, chromosome; CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; Mb, megabases; NC, not calculated; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus; TG, plasma triglyceride. The modeled allele is the LDL-raising allele in the combined HTG cohort.

4.3.3. Genetic risk scores in HTG and the HLP-HTG phenotypes

Next, we constructed weighted allelic risk scores to assess the accumulation of lipid-associated risk alleles in the study sample (**Figure 4.1**). Consistent with our previous report in a smaller sample (Teslovich et al., 2010), weighted TG risk scores were significantly increased in HTG patients versus controls (**Figure 4.1, top**). The relative frequency distribution was clearly shifted towards increased scores in HTG patients versus controls: 0.982±0.004 mmol/L versus 0.896±0.003 mmol/L ($P=1.6X10^{-53}$); unweighted risk scores were similarly increased in HTG patients (**Figure 4.2**). Weighted TG risk scores were associated with increased HTG susceptibility ($P_{trend}<0.001$), as subjects in the highest risk score bin were 4.15 (95% CI: 2.84-6.09) times more likely to be HTG cases than healthy controls compared to the median risk score bin ($P=8.56X10^{-14}$). Subjects in the highest risk score bin were 23.0 (95% CI: 12.86-41.14) times more likely to be HTG cases than healthy controls compared with subjects in the lowest risk score bin as the referent group ($P=3.51X10^{-40}$).

Weighted HDL-C risk scores were also significantly increased in HTG patients versus controls (**Figure 4.1, middle**). The relative frequency distribution of HDL-C risk scores was modestly shifted towards less favorable scores in HTG patients versus controls, corresponding to a highly significant difference between HTG cases and controls of -0.449±0.002 mmol/L versus -0.431±0.001 mmol/L, respectively ($P=1.3X10^{-12}$). Incremental increases in HDL-C risk score also corresponded to increased HTG susceptibility ($P_{trend}<0.001$), as subjects in the highest risk score bin were 1.50 (95% CI: 1.02-2.12) times more likely to be HTG cases than healthy controls compared to the

Figure 4.1. Weighted allelic risk scores are increased in polygenic HTG patients versus controls. Left: relative frequency distribution of weighted risk scores in HTG patients and controls. Right: forest plot demonstrating incremental increases in HTG susceptibility or protection as weighted risk scores deviate from the median. P-values represent Cochrane-Armitage test for trend. CI, confidence interval; HTG, hypertriglyceridemia; TG, triglyceride.

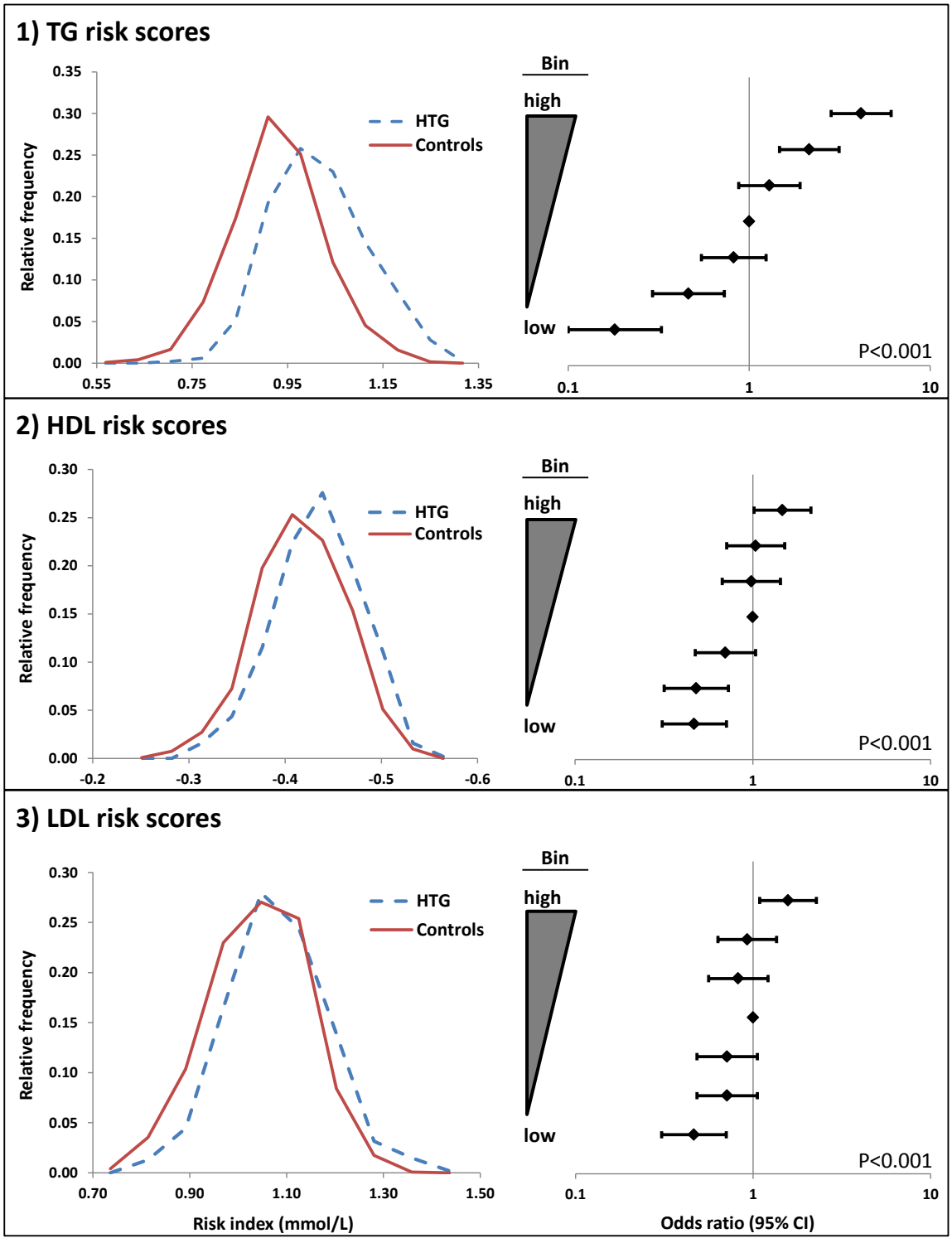
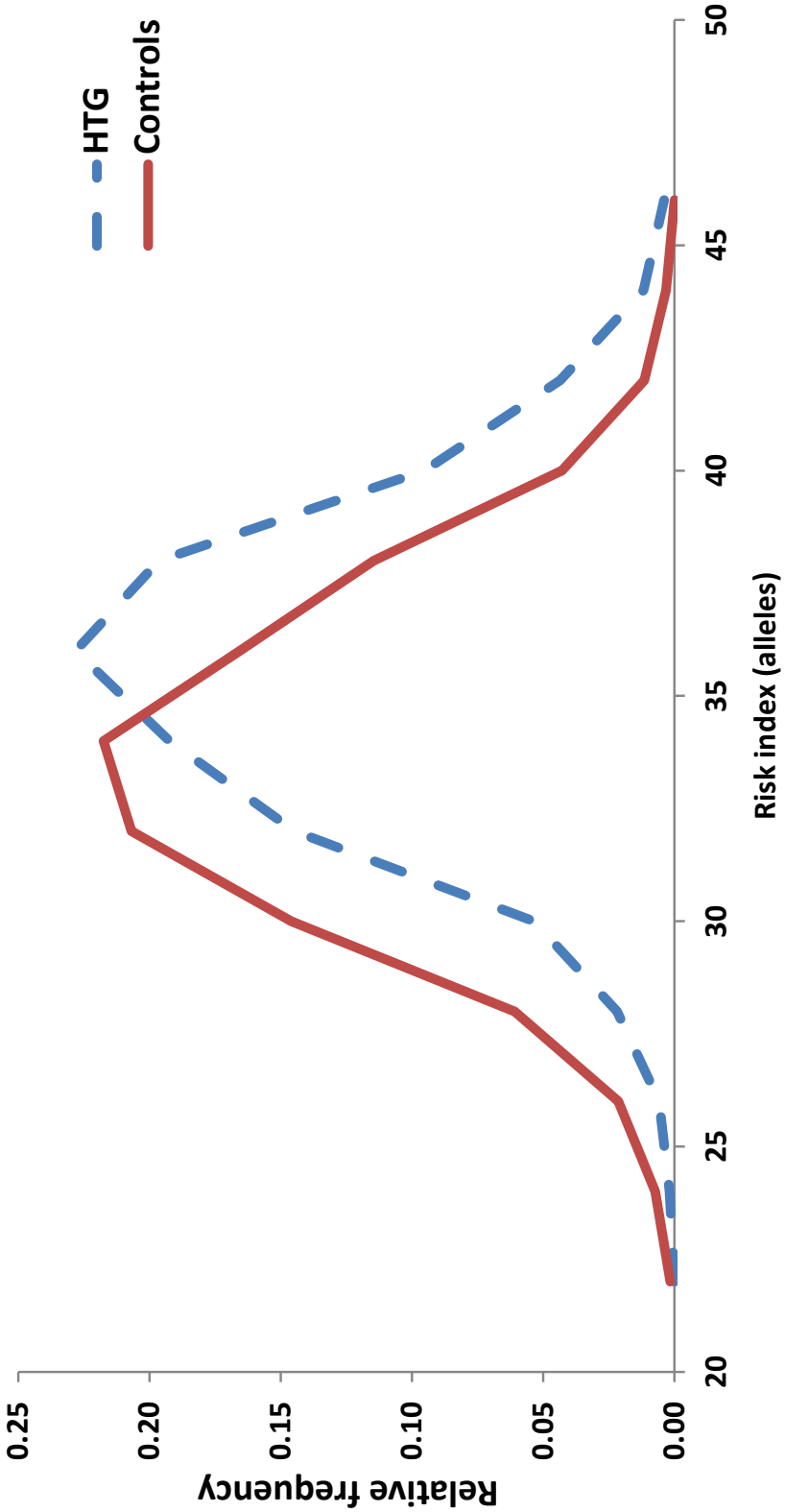

1) TG risk scores
Relative frequency
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.55
0.75
0.95
1.15
1.35
HTG
Controls
Bin
high
low
P<0.001
0.1
1
10
2) HDL risk scores
Relative frequency
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.2
-0.3
-0.4
-0.5
-0.6
HTG
Controls
Bin
high
low
P<0.001
0.1
1
10
3) LDL risk scores
Relative frequency
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.70
0.90
1.10
1.30
1.50
Risk index (mmol/L)
HTG
Controls
Bin
high
low
P<0.001
0.1
1
10
Odds ratio (95% CI)

Figure 4.2. Unweighted allelic TG scores are increased in polygenic HTG patients versus controls. Allelic risk scores were constructed as the unweighted sum of 32 TG-associated loci.


HTG
Controls
Relative frequency
0.25
0.20
0.15
0.10
0.05
0.00
20
25
30
35
40
45
50
Risk index (alleles)

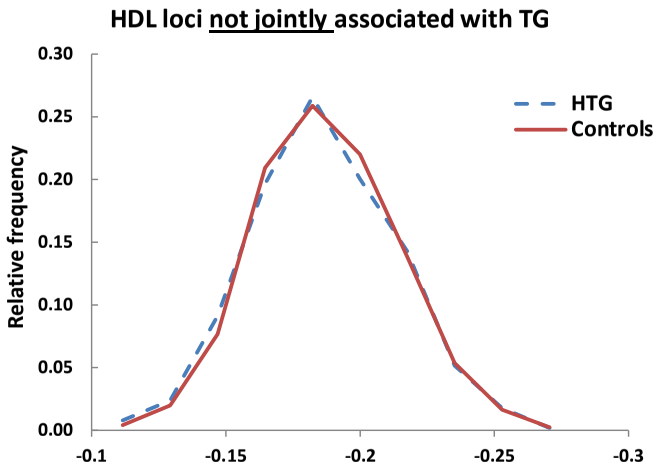
median risk score bin (P=0.05). Subjects in the highest risk score bin were 3.14 (95% CI: 2.09-4.72) times more likely to be HTG cases than healthy controls compared to the lowest risk score bin as the referent group ($P=2.87X10^{-8}$).

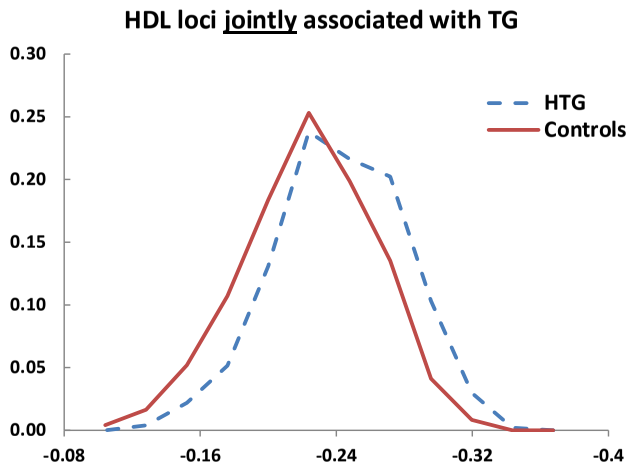
Weighted LDL-C risk scores were less dramatically increased in HTG patients versus controls (**Figure 4.1, bottom**). A modest shift towards higher weighted risk scores was observed in HTG patients, corresponding to a mean risk score of 1.04±0.005 mmol/L in HTG patients versus 1.00±0.003 mmol/L in controls ($P=6.9X10^{-12}$). Incremental increases in LDL-C risk score were marginally associated with increased HTG susceptibility ($P_{trend}<0.001$), as subjects in the highest risk score bin were 1.57 (95% CI: 1.09-2.27) times more likely to be HTG cases than healthy controls compared to the median risk score bin (P=0.02). Subjects in the highest risk score bin were 3.39 (95% CI: 2.25-5.11) times more likely to be HTG cases than healthy controls compared to the lowest risk score bin as the referent group ($P=3.03X10^{-9}$).

We hypothesized that the increased HDL-C and LDL-C risk scores observed in HTG patients were mediated by the pleiotropic effect of 15/47 (32%) HDL-C or 10/37 (27%) LDL-C variants that were jointly associated with plasma TG concentration (Teslovich et al., 2010). Accordingly, we constructed HDL-C and LDL-C risk scores composed of either (1) variants associated exclusively with HDL-C or LDL-C, or (2) pleiotropic variants jointly associated with plasma TG concentration in addition to HDL-C or LDL-C (**Figure 4.3**). For HDL-C, risk scores composed exclusively of HDL-C-associated variants were not different between HTG patients and healthy controls

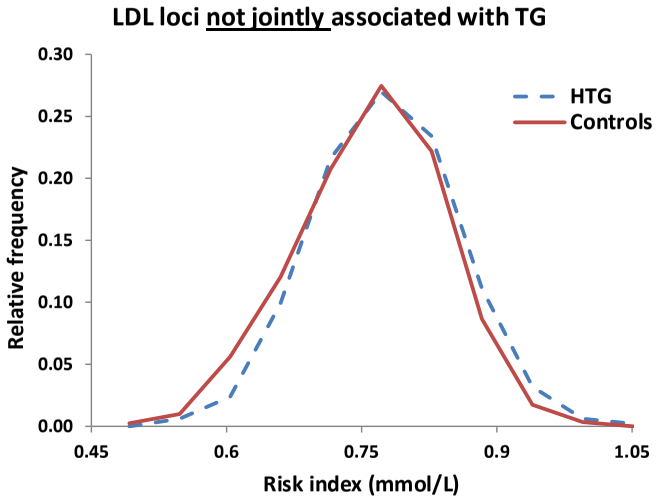
Figure 4.3. Increased weighted allelic HDL-C and LDL-C risk scores in HTG patients are driven primarily by loci jointly associated with plasma TG concentration. Weighted risk scores from Figure 4.1 were separated by loci associated with only HDL-C (top; left) or LDL-C (bottom; left), or loci jointly associated with plasma TG in addition to HDL-C (top; right) or LDL-C (bottom; right). Weighted allelic risk scores were constructed from the sum of HDL-C or LDL-C effect estimates at each locus, not effect estimates for associations with plasma TG.

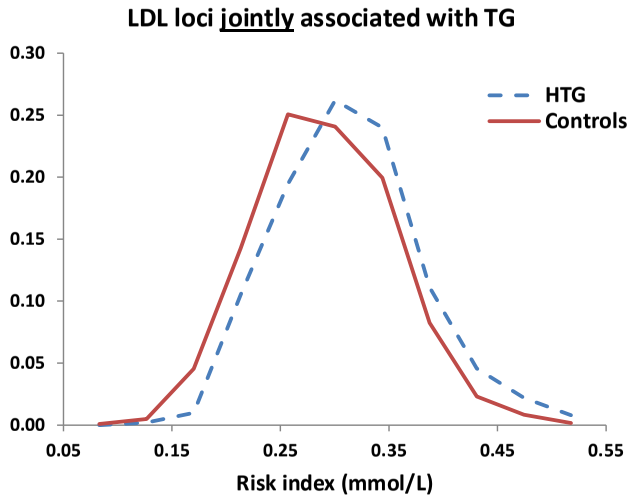
HDL risk scores





LDL risk scores





(P=0.72); however, risk scores composed of pleiotropic HDL-C-associated variants were increased in HTG patients versus controls ($P=1.7X10^{-18}$) **(Figure 4.3; top)**. For LDL-C, risk scores composed exclusively of LDL-C-associated variants were slightly increased in HTG patients versus controls (P=0.0001); however, risk scores composed of pleiotropic LDL-C-associated variants were greatly increased in HTG patients versus controls ($P=8.8X10^{-11}$) **(Figure 4.3; bottom**). We considered the possibility that the large effect sizes of many pleiotropic variants could have inflated the observed variant accumulation; however this is unlikely because parallel analyses using unweighted risk scores were similarly increased in HTG patients (**Figure 4.4**).

Finally, we tested whether the accumulation of lipid-associated risk alleles observed in HTG patients was consistent across the different HLP-HTG phenotypes (**Figure 4.5**). We hypothesized that weighted TG risk scores would be equal among HLP-HTG phenotypes, whereas weighted HDL-C and LDL-C risk scores might differ among HTG patients and phenotypes. For TG risk scores, all HLP-HTG phenotypes had risk scores elevated above controls, with HLP type 5 patients having risk scores further increased over other phenotypes, suggesting an accumulation of common variants with larger effects in these more severely affected patients. For HDL-C risk scores, only HLP types 4 and 5 had scores increased above controls, consistent with the excess of HDL-C-associated variants with directions of effect concordant with GLGC estimates. For LDL-C risk scores, HLP phenotypes 2B and 5 had risk scores increased above controls, also consistent with the excess of LDL-C-associated variants with directions of effect concordant with GLGC estimates. Although the greatest differences in risk scores

Figure 4.4. Unweighted allelic HDL-C and LDL-C risk scores are increased in polygenic HTG patients versus controls and are driven primarily by loci jointly associated with plasma TG concentration. Allelic risk scores were constructed as the unweighted sum of 47 HDL-C (top; left) or 37 LDL-C (bottom; left) associated loci. Unweighted risk scores were separated by loci associated only with HDL-C (top; middle) or LDL-C (bottom; middle), or loci jointly associated with plasma TG in addition to HDL-C (top; right) or LDL-C (bottom; right).

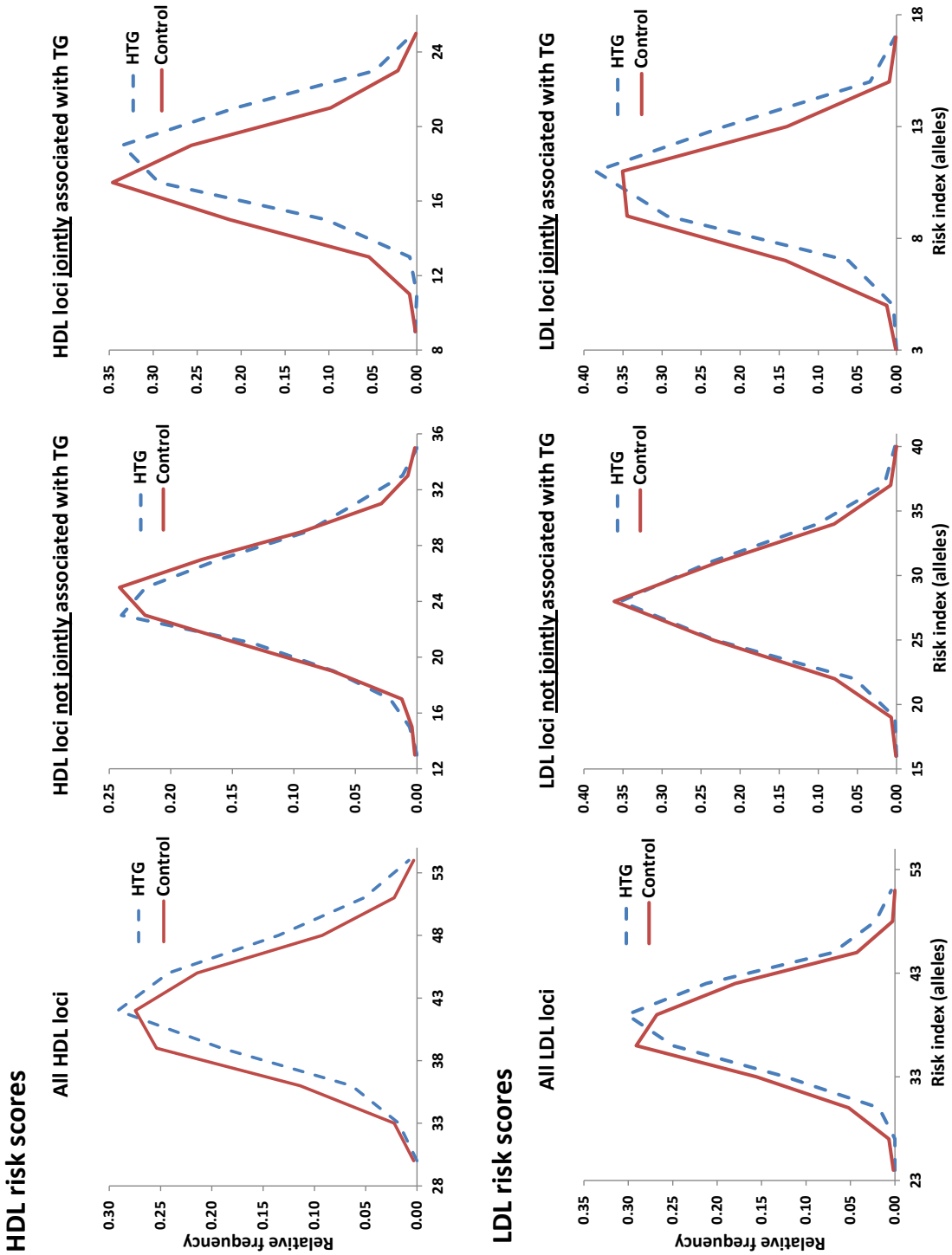
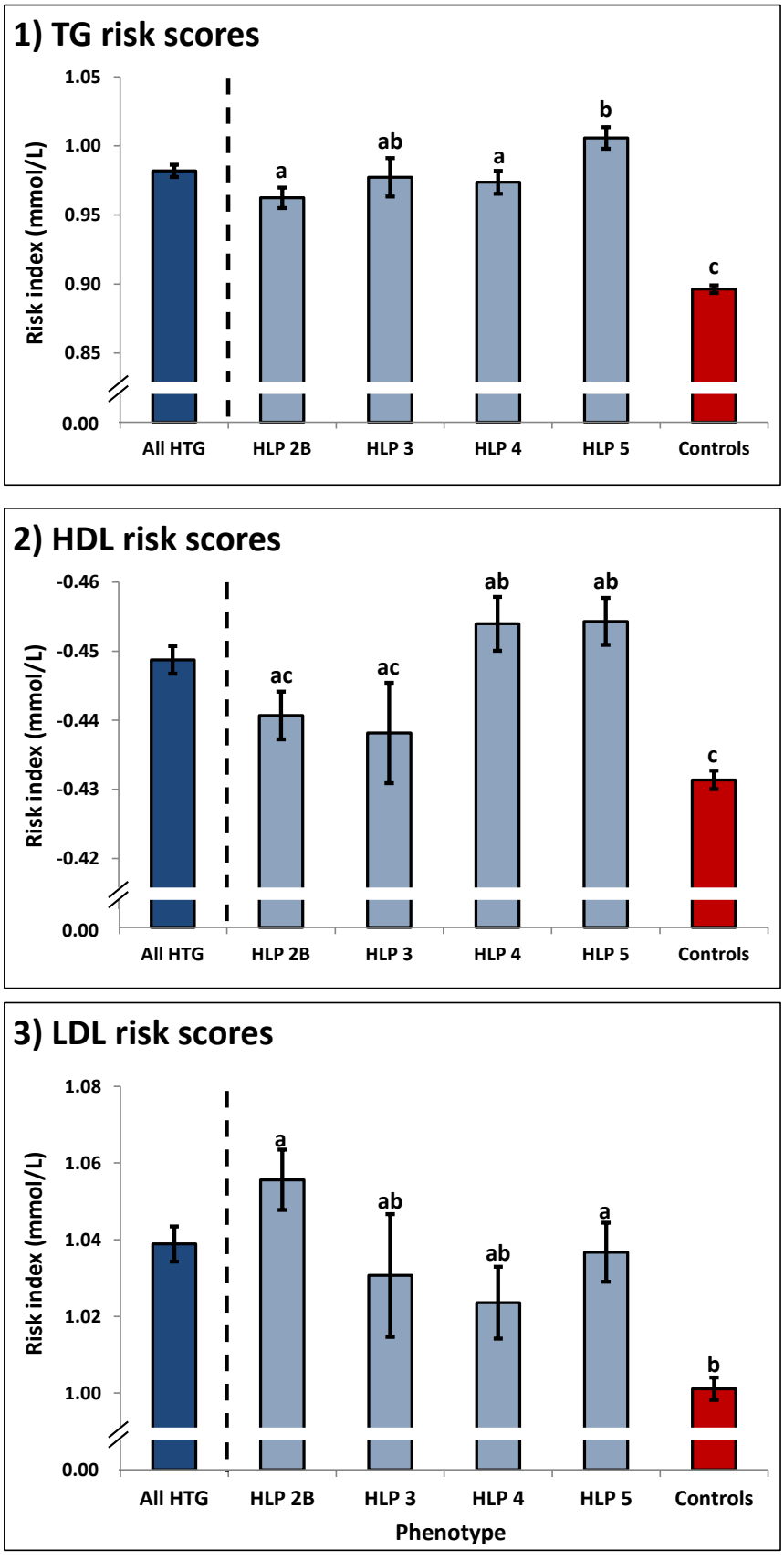

HDL risk scores
All HDL loci
HDL loci not jointly associated with TG
HDL loci jointly associated with TG
LDL risk scores
All LDL loci
LDL loci not jointly associated with TG
LDL loci jointly associated with TG
HTG
Control
Relative frequency
Risk index (alleles)

Figure 4.5. Weighted allelic risk scores differ among polygenic HLP-HTG phenotypes. Mean risk score in all HTG patients is provided as a reference only; it was not included in statistical comparisons. Values are mean ± standard error; means sharing letters are not statistically different.


1) TG risk scores
Risk index (mmol/L)
1.05
1.00
0.95
0.90
0.85
0.00
a
ab
a
b
c
All HTG
HLP 2B
HLP 3
HLP 4
HLP 5
Controls
2) HDL risk scores
Risk index (mmol/L)
-0.46
-0.45
-0.44
-0.43
-0.42
0.00
ac
ac
ab
ab
c
All HTG
HLP 2B
HLP 3
HLP 4
HLP 5
Controls
3) LDL risk scores
Risk index (mmol/L)
1.08
1.06
1.04
1.02
1.00
0.00
a
ab
ab
a
b
All HTG
HLP 2B
HLP 3
HLP 4
HLP 5
Controls
Phenotype

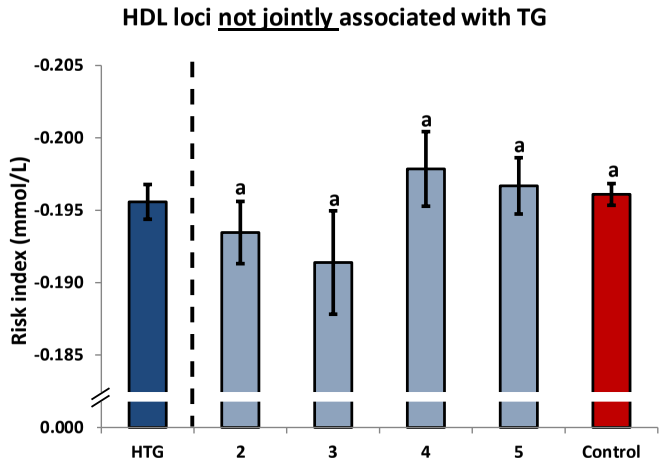
between HLP phenotypes and controls were attributed to pleiotropic variants, the contribution of both pleiotropic and lipid-specific variants likely contribute to the overall phenotypic heterogeneity among HLP-HTG phenotypes (**Figure 4.6**). These data collectively suggest that common variants associated with plasma lipid concentrations accumulate in HTG patients: TG-associated variants accumulate in all HLP-HTG phenotypes, whereas HDL-C and LDL-C associated variants preferentially accumulate in an HLP-HTG phenotype-dependent manner.

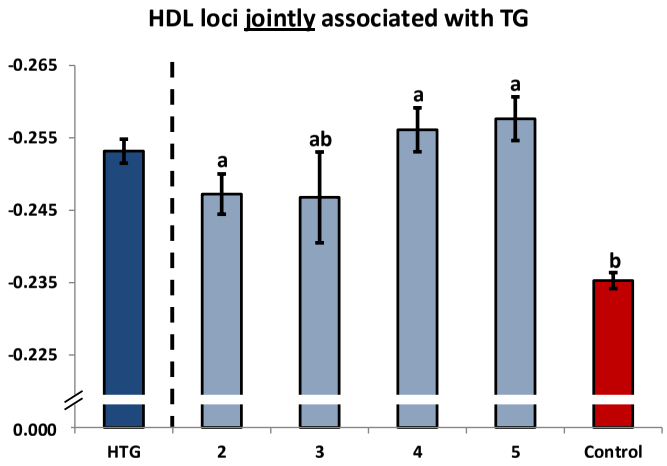
4.3.4. Rare variants in TG-associated genes among HLP-HTG phenotypes

We previously demonstrated that a significant excess of rare variants is present in GWAS-identified genes associated with HTG, including *APOA5*, *GCKR*, *LPL* and *APOB*, in HTG patients compared to controls (Johansen et al., 2010). Rare variants identified in this previous study were often found in multiple subjects, however we could not identify any trends between genes, variant or HLP-HTG phenotype. Here, we tested whether this excess of rare variants extended to the HLP phenotypes in subjects with available sequencing data. Carriers and non-carriers were compared in this analysis, as some subjects carried >1 rare variant. Indeed, a significant excess of rare variants was observed across all HLP-HTG phenotypes (**Figure 4.7**). Comparison of rare variants between each HLP-HTG phenotype and healthy controls using Fisher's exact test revealed that carriers were 1.8-2.6 times more likely to have HLP-HTG case status than control status. There were no differences in number of rare variant carriers or non-carriers among the 4 phenotypes (P=0.59), nor were there differences in distribution of rare variants among genes in each phenotype (P=0.69), as assessed using contingency table

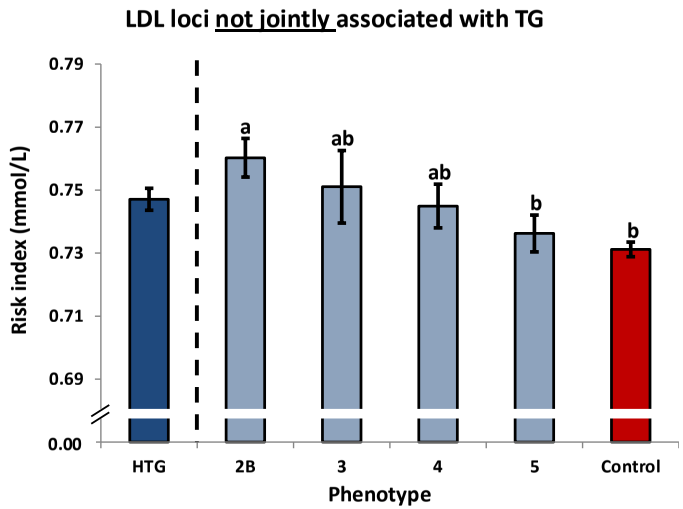
Figure 4.6. Weighted allelic HDL-C and LDL-C risk scores have contributions from loci exclusively associated with HDL-C or LDL-C, and pleiotropic loci jointly associated with plasma TG concentration, causing differences among the polygenic HLP-HTG phenotypes. Differences among weighted risk scores in HLP phenotypes and controls were assessed using ANOVA and *post hoc* pairwise comparisons using Tukey's test ($P<0.05$). Mean risk score in all HTG patients is provided as a reference only; it was not included in statistical comparisons. Values are mean ± standard error; means sharing letters are not statistically different.

HDL risk scores





LDL risk scores



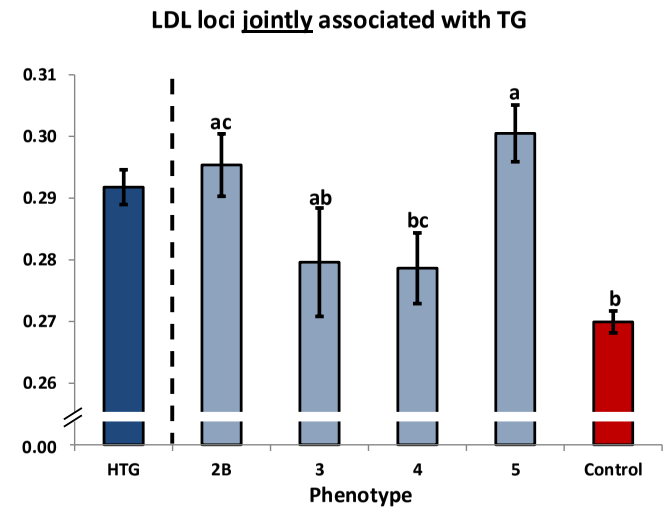
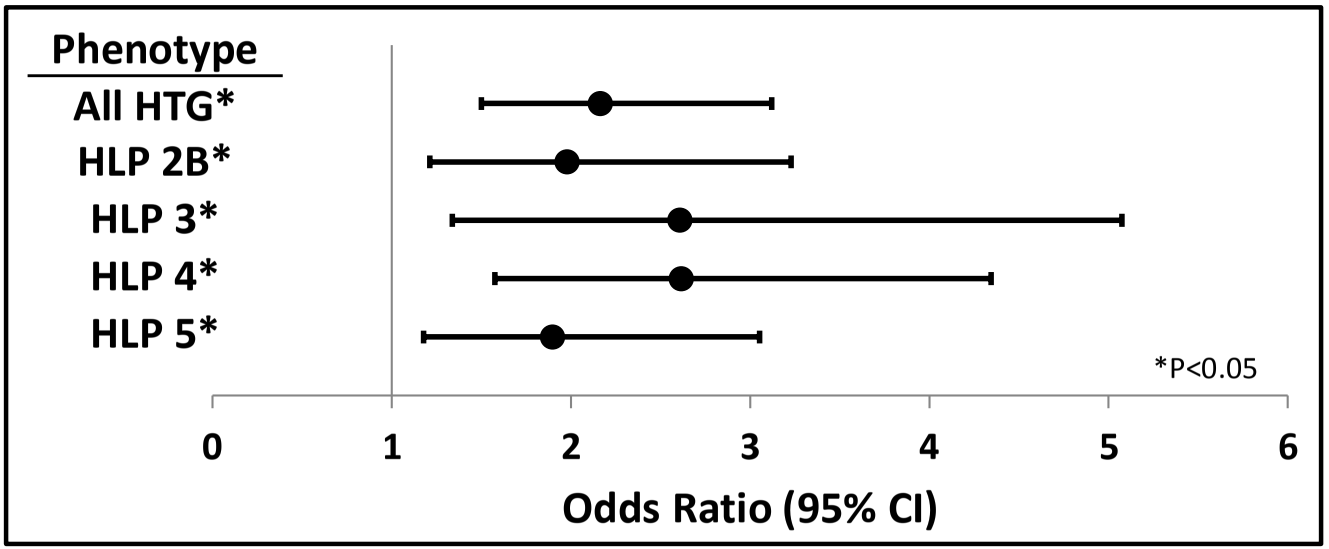


Figure 4.7. Excess of rare variants is similar among polygenic HLP-HTG phenotypes in terms of carrier number and gene distribution. Top: Rare variant accumulation was compared between each HLP-HTG phenotype and controls. Statistical significance was measured using Fisher's exact test using a 2-sided $p<0.05$. Bottom: Distribution of rare variants across genes and subjects. Carriers are subjects with ≥1 rare variant in ≥1 genes. CI, confidence interval; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia.



	Total Variants				Subjects (%)	
Phenotype	***APOA5***	***GCKR***	***LPL***	***APOB***	**Carriers**	**Non-carriers**
All HTG (n=438)	5	20	44	87	123 (28.1)	315 (71.9)
HLP 2B (n=133)	2	8	10	25	35 (26.3)	98 (73.7)
HLP 3 (n=50)	1	4	8	12	16 (32.0)	34 (68.0)
HLP 4 (n=106)	0	2	14	24	34 (32.1)	72 (67.9)
HLP 5 (n=149)	2	6	12	26	38 (25.5)	111 (74.5)
Control (n=327)	1	5	8	39	50 (15.3)	277 (84.7)

analysis. We also found no differences in TG risk scores among carriers and non-carriers of rare variants in HTG patients and controls (**Figure 4.8**). These data suggest that the distribution of rare variants in HTG-associated genes occurs non-preferentially across all HLP-HTG phenotypes in several TG-associated genes, on top of an accumulation of lipid-associated common variants.

4.3.5. Variation explained by clinical and genetic variables in HLP-HTG patients

Given the limitations associated with calculating explained variation from logistic regression models, we measured changes in model goodness-of-fit (pseudo-R^2) attributed to different clinical and genetic variables in a logistic regression model discriminating between HLP-HTG phenotypes and healthy controls as a surrogate metric for explained variation (**Figure 4.9; see Methods**). Within HTG patients, the total proportion of variation explained by the model was 42.0%, very similar to our previous estimates (Johansen et al., 2010). Clinical variables explained 21.2%, common variants explained 17.4%, and rare variants explained 1.4%. In this analysis, *APOE* genotype was also included due to its historical importance in HLP type 3, which explained an additional 2% of variation among HTG cases. Among the HLP-HTG phenotypes, the proportion of variation attributable to these variables differed significantly. In HLP type 3, *APOE* ε2 genotype explained 34.6% of variation; common variation from GWAS-identified genes only explained an additional 1.5% variation. Conversely, common genetic variants explained 20-24% of variation among remaining phenotypes. *APOA5*, *GCKR*, *MLXIPL* and *CILP2* explained the majority of this variation, although *LPL* explained additional variation in HLP type 4, *TRIB1* explained additional variation in HLP type 2B, and

Figure 4.8. TG risk scores are similar between carriers and non-carriers of rare variants. Differences among TG risk scores were assessed using ANOVA and *post hoc* pairwise comparisons using Tukey's test ($P<0.05$). Values are mean ± standard error; means sharing letters are not statistically different. HTG, hypertriglyceridemia.

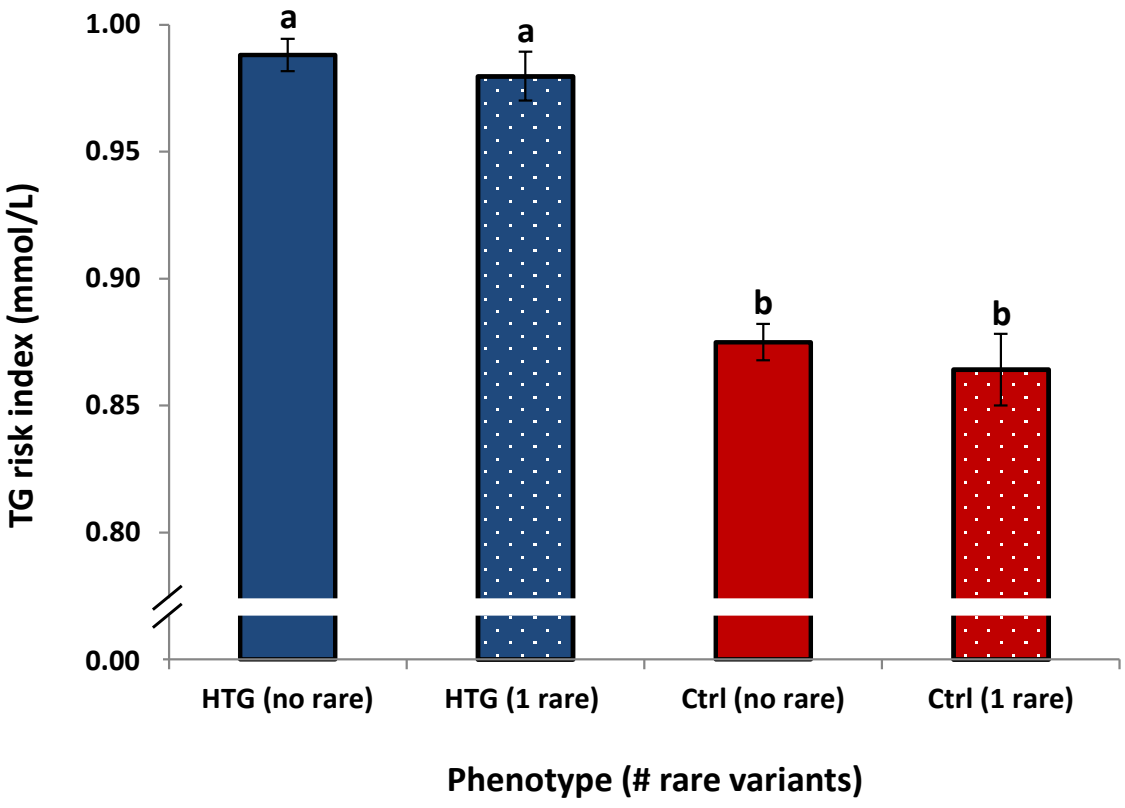
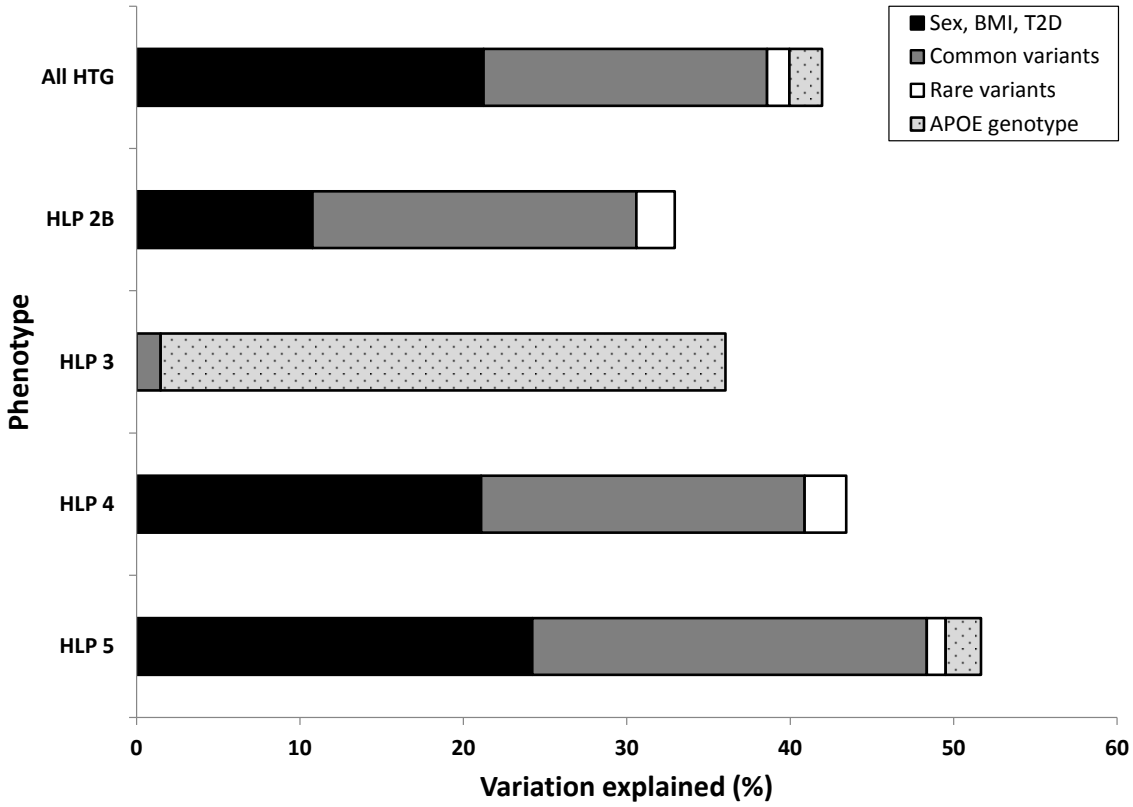

1.00
0.95
0.90
0.85
0.80
0.00
TG risk index (mmol/L)
a
a
b
b
HTG (no rare)
HTG (1 rare)
Ctrl (no rare)
Ctrl (1 rare)
Phenotype (# rare variants)

Figure 4.9. Comparison of variation attributable to clinical and genetic variables among polygenic HLP-HTG phenotypes within our cohort. The proportion of variation explained was calculated from changes in model fit (R^2) from multiple logistic regression entering clinical and genetic variables using forward modeling. Common variants included risk alleles (0, 1 or 2 alleles) from multiple TG-associated loci selected by the model. *APOE* genotype was coded as copies of the *APOE* E2 or E4 genotype. BMI, body mass index; HLP, hyperlipoproteinemia; HTG, hypertriglyceridemia; T2D, type 2 diabetes; TG, triglyceride.


Sex, BMI, T2D
Common variants
Rare variants
APOE genotype
All HTG
HLP 2B
HLP 3
HLP 4
HLP 5
Phenotype
0
10
20
30
40
50
60
Variation explained (%)

ANGPTL3 explained additional variation in HLP types 2B and 5. Rare genetic variants in the 4 genes also explained 1.2-2.6% of variation among HLP types 2B, 4 and 5. Finally, *APOE* genotype explained ~2.2% of variation among HLP type 5 patients, but did not explain additional variation among remaining phenotypes. Thus, the contribution of common and rare variants to the model was relatively similar among HLP phenotypes 2B, 4 and 5, whereas *APOE* genotype was the best predictor of HLP type 3 case status.

4.4. Discussion

The principal finding of this chapter is that many common lipid-associated variants in loci associated with TG, HDL-C and LDL-C underlie multiple biochemically-defined HLP-HTG phenotypes, and contribute to the phenotypic spectrum within HTG. Weighted TG risk scores were distinctly increased across all HLP-HTG phenotypes compared to controls, whereas weighted HDL-C and LDL-C risk scores were also increased, albeit less pronounced, in some HLP-HTG phenotypes compared to controls. Interestingly, decomposition of HDL-C and LDL-C risk scores revealed that pleiotropic variants (those jointly associated with TG) accounted for the greatest difference in HDL-C and LDL-C risk scores among HLP-HTG phenotypes. However, rare variants in 4 genes also accumulated equally among HLP-HTG phenotypes. Thus, HTG susceptibility and phenotypic heterogeneity appear to be influenced in part by the accumulation of common and rare lipid-associated variants. Taken together, these findings provide a more complete accounting of the genetic architecture of HTG susceptibility and the basis of the phenotypic heterogeneity in HTG.

Earlier preliminary studies provided a first approximation of the genetic architecture of HTG and the HLP-HTG phenotypes (Hegele et al., 2009, Wang et al., 2008a, Wang et al., 2008b), while the novelty of the current study lies in the comprehensive nature of our analysis, which included all recently identified lipid-associated variants from the GLGC (Teslovich et al., 2010). We have confirmed the hypothesis that a common genetic architecture of many common and rare TG-associated variants underlies predisposition to HTG (Johansen et al., 2011), and demonstrated that

pleiotropic variants and accumulation of additional lipid-associated variants may partly explain an origin for the phenotypic heterogeneity among HTG patients.

Our interpretation of these results is that a genetic burden of deleterious variants may predispose some subjects to HTG. Common and rare variants accumulate more often in HTG patients, suggesting that risk alleles may incrementally contribute to HTG susceptibility. However, it appears that no single variant or accumulation of variants is sufficient to cause HTG or any particular HLP-HTG phenotype. Despite seemingly strong phenotype-specific associations, risk alleles at associated loci are overrepresented in multiple HLP-HTG phenotypes compared to controls. Similarly, rare HTG-associated variants also accumulate comparably across the HLP-HTG phenotypes, in addition to an elevated background of common risk variants, again suggesting that rare variants are not sufficient for HTG causation. Ultimately, the significant overlap of risk alleles among HTG patients and controls suggests that – as with most complex traits – genetic variants only account for a portion of variation in HTG case status.

Our evolving model of polygenic HTG is that accumulation of common and rare TG-associated risk alleles gradually increases HTG susceptibility. Furthermore, the accumulation of pleiotropic risk alleles in addition to lipid-associated risk alleles may contribute to the phenotypic heterogeneity characteristic of the classical HLP-HTG phenotypes. However, we do not exclude the possibility that HTG susceptibility in some patients can be independent of risk allele accumulation. In our sample, we observe that HLP type 2B appears to be associated preferentially with common variants in pleiotropic

and exclusively LDL-C-associated loci; HLP type 4 appears to be associated with common variants in pleiotropic HDL-C-associated loci; and the most severe phenotype HLP type 5 is associated with large effect common TG-associated risk alleles, with additional contributions from both pleiotropic LDL-C- and HDL-C-associated loci. The contribution of risk alleles from *APOA5*, *GCKR*, *MLXIPL*, and *CILP2* appeared to contribute relatively equally among these phenotypes, with additional smaller contributions from other TG-associated variants in each HLP-HTG phenotype.

HLP type 3 is unique compared to other HLP-HTG phenotypes as it is predominantly explained by the classical APOE ε2/ε2 isoform (*APOE* E2/E2 genotype). However, even APOE ε2/ε2 is individually insufficient to cause HLP type 3. HLP type 3 is defined by the accumulation of intermediate density lipoproteins (IDL) visible as β-VLDL (or a broad beta band) on electrophoresis, resulting in a high ratio of VLDL cholesterol to TG. Many HLP type 3 patients have APOE ε2/ε2, although some HLP type 5 patients that are clinically defined by fasting chylomicronemia and pancreatitis in our sample also have APOE ε2/ε2 genotype; many healthy controls in the general population also have APOE ε2/ε2 genotype. Variants such as *APOA5* that are strongly associated with HLP type 3, in addition to other TG-raising risk alleles, likely contribute additionally to HLP type 3 susceptibility.

Our conclusions must be interpreted in the context of important limitations of our study. First, our sample size and statistical power were limited despite having the largest cohort of HTG patients reported to date. Accordingly, our sample was insufficient to

replicate most small effect variants identified by the GLGC (Teslovich et al., 2010). This was compounded by the need for subgroup analyses, which prevented replication of many variants in the HLP-HTG phenotypes, even those associated with HTG at $P<0.05$, but was unavoidable because of insufficient numbers. To compensate for this, we have used additional analyses including risk score analyses and estimates of directionality of effect to overcome this limitation.

Second, the analysis of GWAS lead SNPs is likely insufficient to demonstrate the full spectrum of genetic heterogeneity within the HLP-HTG phenotypes. We genotyped tagSNPs that serve as surrogates for functional variants at each locus, not truly querying the specific functional variants that increase HTG susceptibility. Direct genotyping of functional variants, such as the *APOA5* -1131T>C promoter variant and the S19W variant, which are both strongly associated with HTG (Hegele et al., 2009, Wang et al., 2008a, Wang et al., 2008b) but are each in relatively weak LD with the *APOA5* tag-SNP rs964184 (r^2=0.29 and r^2=0.34 in sample subsets) may improve the resolution of our risk scores, and thus ability to detect differences between phenotypes. This might be expected given that *APOE* genotype at functional variants rs7412 and rs429358 (encoding the APOE E2/E3/E4 isoforms) contributed significantly more information to HLP type 3 case status than the *APOE* tagSNP.

Third, only 4 HTG-associated genes were selected for resequencing. This limited view of rare variant accumulation may preclude our ability to detect genetic heterogeneity in other genes that could mediate phenotypic heterogeneity within the

HLP-HTG phenotypes. Furthermore, functional data is necessary to separate deleterious from benign variants, which may also introduce noise into this analysis preventing detection of heterogeneity. However, many variants in these genes were identified across multiple HLP-HTG phenotypes. For example, the loss-of-function variant *LPL* G188E (Emi et al., 1990) which completely attenuates LPL function was identified as a heterozygous mutation in HLP types 3, 4, and 5. This suggests that rare variants in these genes have a general effect on HTG susceptibility, without predisposing subjects to a specific HLP-HTG phenotype. Ultimately, the combination of low power arising from a limited sample size with an incomplete survey of functional genetic variants in all HTG-associated genes may hide some of the genetic heterogeneity initially expected to underlie the HLP-HTG phenotypes. However, our results based on GWAS tagSNPs and 4 HTG-associated genes provide evidence for a common overlapping genetic etiology for the phenotypic spectrum of HLP-HTG phenotypes.

Future studies will require exhaustive resequencing of TG-associated genes to identify additional genetic variants involved in HTG susceptibility. These will include both common functional variants that underlie GWAS signals and large effect rare variants that mediate significant HTG predisposition in their carriers. Together, these variants will likely explain additional variation in HTG case status and may provide a better explanation for the genetic basis of the phenotypic heterogeneity among HLP-HTG patients. Consideration of these determinants in the context of multiplicative or synergistic effects between genes, variants, and metabolic and environmental exposures, may contribute additionally to the unexplained variation among HTG patients.

Evaluation of remaining unexplored variables may eventually lead to improved detection of subjects with increased HTG susceptibility. For example, genetic risk scores composed of functional HTG-associated variants may facilitate presymptomatic identification of clinic patients at increased risk for development of HTG. Common variants in genes such as *APOA5*, *LPL*, *TRIB1* and *CILP2*, are strong determinants of HTG but also CVD, and could further be integrated into risk prediction algorithms such as the Framingham risk score as individual variants or genetic risk scores to help predict subjects at increased risk of CVD. The clinical utility of HTG-associated variants in the prediction of HTG or CVD susceptibility has not been evaluated, but could be used in the future to direct evidence-based treatments or lifestyle interventions at those with highest HTG susceptibility.

In summary, we demonstrated that a spectrum of common and rare genetic variants in lipid-associated loci underlies the biochemically-defined HLP phenotypes characterized by HTG. We confirm that an accumulation of common TG-associated risk alleles is observed in HTG patients, and provide new evidence supporting the accumulation of pleiotropic TG-, HDL-C and LDL-C-associated risk alleles as possible determinants of HTG phenotypic heterogeneity. The significant overlap of risk alleles among multiple HLP-HTG phenotypes indicates a shared, yet complex and heterogeneous genetic HTG susceptibility. Ongoing discovery of additional TG-associated alleles and elucidation of specific functional variants contributing to HTG pathophysiology will be required in the context of normal and perturbed metabolic and environmental conditions.

4.5. References

Anand, S.S., Yusuf, S., Vuksan, V., Devanesen, S., Teo, K.K., Montague, P.A., Kelemen, L., Yi, C., Lonn, E., Gerstein, H., *et al.* (2000). Differences in risk factors, atherosclerosis, and cardiovascular disease between ethnic groups in Canada: the Study of Health Assessment and Risk in Ethnic groups (SHARE). Lancet *356*, 279-284.

Bansal, S., Buring, J.E., Rifai, N., Mora, S., Sacks, F.M., and Ridker, P.M. (2007). Fasting compared with nonfasting triglycerides and risk of cardiovascular events in women. JAMA *298*, 309-316.

Beigneux, A.P., Franssen, R., Bensadoun, A., Gin, P., Melford, K., Peter, J., Walzem, R.L., Weinstein, M.M., Davies, B.S., Kuivenhoven, J.A., *et al.* (2009). Chylomicronemia with a mutant GPIHBP1 (Q115P) that cannot bind lipoprotein lipase. Arterioscler Thromb Vasc Biol *29*, 956-962.

Connelly, P.W., Maguire, G.F., and Little, J.A. (1987). Apolipoprotein CIISt. Michael. Familial apolipoprotein CII deficiency associated with premature vascular disease. J Clin Invest *80*, 1597-1606.

Cullen, P. (2000). Evidence that triglycerides are an independent coronary heart disease risk factor. Am J Cardiol *86*, 943-949.

Emi, M., Wilson, D.E., Iverius, P.H., Wu, L., Hata, A., Hegele, R., Williams, R.R., and Lalouel, J.M. (1990). Missense mutation (Gly----Glu188) of human lipoprotein lipase imparting functional deficiency. J Biol Chem *265*, 5910-5916.

Freiberg, J.J., Tybjaerg-Hansen, A., Jensen, J.S., and Nordestgaard, B.G. (2008). Nonfasting triglycerides and risk of ischemic stroke in the general population. JAMA *300*, 2142-2152.

Havel, R.J., and Gordon, R.S., Jr. (1960). Idiopathic hyperlipemia: metabolic studies in an affected family. J Clin Invest *39*, 1777-1790.

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

Hegele, R.A., Ban, M.R., Hsueh, N., Kennedy, B.A., Cao, H., Zou, G.Y., Anand, S., Yusuf, S., Huff, M.W., and Wang, J. (2009). A polygenic basis for four classical Fredrickson hyperlipoproteinemia phenotypes that are characterized by hypertriglyceridemia. Hum Mol Genet *18*, 4189-4194.

Hegele, R.A., and Pollex, R.L. (2009). Hypertriglyceridemia: phenomics and genomics. Mol Cell Biochem *326*, 35-43.

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.

Johansen, C.T., Kathiresan, S., and Hegele, R.A. (2011). Genetic determinants of plasma triglycerides. J Lipid Res *52*, 189-206.

Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., *et al.* (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet *42*, 684-687.

Kathiresan, S., Voight, B.F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P.M., Anand, S., Engert, J.C., Samani, N.J., Schunkert, H., *et al.* (2009). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet *41*, 334-341.

Li, Y., and Abecasis, G.R. (2006). Mach 1.0: rapid haplotype reconstruction and missing genotype inference. Am J Hum Genet *S79*, 2290.

Nordestgaard, B.G., Benn, M., Schnohr, P., and Tybjaerg-Hansen, A. (2007). Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. JAMA *298*, 299-308.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet *2*, e190.

Peterfy, M., Ben-Zeev, O., Mao, H.Z., Weissglas-Volkov, D., Aouizerat, B.E., Pullinger, C.R., Frost, P.H., Kane, J.P., Malloy, M.J., Reue, K., *et al.* (2007). Mutations in LMF1 cause combined lipase deficiency and severe hypertriglyceridemia. Nat Genet *39*, 1483-1487.

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.

Priore Oliva, C., Pisciotta, L., Li Volti, G., Sambataro, M.P., Cantafora, A., Bellocchio, A., Catapano, A., Tarugi, P., Bertolini, S., and Calandra, S. (2005). Inherited apolipoprotein A-V deficiency in severe hypertriglyceridemia. Arterioscler Thromb Vasc Biol *25*, 411-417.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.

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., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707-713.

Wang, J., Ban, M.R., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., and Hegele, R.A. (2008a). APOA5 genetic variants are markers for classic hyperlipoproteinemia phenotypes and hypertriglyceridemia. Nat Clin Pract Cardiovasc Med *5*, 730-737.

Wang, J., Ban, M.R., Zou, G.Y., Cao, H., Lin, T., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., *et al.* (2008b). Polygenic determinants of severe hypertriglyceridemia. Hum Mol Genet *17*, 2894-2899.

Wang, J., Cao, H., Ban, M.R., Kennedy, B.A., Zhu, S., Anand, S., Yusuf, S., Pollex, R.L., and Hegele, R.A. (2007). Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). Arterioscler Thromb Vasc Biol *27*, 2450-2455.

Wang, J., and Hegele, R.A. (2007). Homozygous missense mutation (G56R) in glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1 (GPI-HBP1) in two siblings with fasting chylomicronemia (MIM 144650). Lipids Health Dis *6*, 1-4.

Yuan, G., Al-Shali, K.Z., and Hegele, R.A. (2007). Hypertriglyceridemia: its etiology, effects and treatment. CMAJ *176*, 1113-1120.

CHAPTER 5

DISCUSSION AND CONCLUSIONS

The work in this chapter originates from sections of the following publications: (1) **Johansen CT**, Wang J, and Hegele RA. (2011). Bias due to selection of rare variants using frequency in controls. *Nat Genet* 43(5):394-395; (2) **Johansen CT**, Kathiresan S, and Hegele RA. (2011). Genetic determinants of plasma triglycerides. *J Lipid Res* 52(2): 189-206; (3) **Johansen CT** and Hegele RA. (2011). Genetic bases of hypertriglyceridemic phenotypes. *Curr Opin Lidiol* 22(4): 247-53; and (4) **Johansen CT** and Hegele RA. (2011). Allelic and phenotypic spectrum of plasma TG. *Biochim Biophys Acta*. Submitted.

5.1. The genetic architecture of HTG

Hypertriglyceridemia (HTG) is a complex, polygenic, adult-onset disorder defined by fasting plasma triglyceride (TG) concentrations >95th percentile (Yuan et al., 2007). Prior to the work in this thesis, the genetic architecture of HTG was poorly defined. Few common and rare genetic variants were associated with HTG, and the vast majority of phenotypic variation was unattributed (Hegele, 2009). However, the experiments described here have greatly enhanced our understanding of HTG. Our comprehensive evaluation of common and rare genetic variation in the largest HTG cohort assembled to date has expanded the number of possible genes implicated in HTG susceptibility to ~35, and provided a refined estimate of variation attributable to genetic and clinical variables.

Furthermore, we have provided a model for HTG susceptibility and phenotypic heterogeneity, speculating that all genes and variants that are associated with plasma TG concentration similarly increase HTG susceptibility. Ultimately, the research described in this thesis has contributed significantly to understanding the genetic determinants of HTG susceptibility, while providing hypothesis-generating proof-of-principle experiments applicable to the genetic architecture of complex diseases. This chapter will summarize the novelty of our experiments and conceptual advances.

5.1.1. Common genetic variation in HTG

We have described the first genome-wide association study (GWAS) ever conducted on an extreme complex lipid phenotype. Using an unbiased approach, we demonstrated that common variation in known and novel genes, including *APOA5*, *GCKR*, *LPL*, and *APOB*, is strongly associated with HTG. We demonstrated that *GCKR* is a novel gene involved in HTG susceptibility, despite no prior functional involvement in lipoprotein metabolism. Similarly, we demonstrated that *APOB* has a role in modulating HTG susceptibility, in addition to extreme phenotypes of high and low plasma low-density lipoprotein (LDL) concentration. Finally, we demonstrated that common variation in known loci including *LPL* and *APOA5* are the strongest HTG-associated variants.

Importantly, we have demonstrated that common genetic determinants of plasma TG concentration are also determinants of HTG susceptibility. We directly replicated associations with TG-associated loci including *MLXIPL*, *ANGPTL3*, *TRIB1*, and provided some evidence for association with *NCAN* and *FADS1-FADS2-FADS3*, implicating TG-

associated genes in HTG susceptibility. Not surprisingly, common variants with individually small effect sizes only detectable by epidemiologically-sized samples were not replicated in our cohort, however virtually every TG-associated locus identified by the Global Lipids Genetics Consortium (GLGC) that was estimated to increase plasma TG concentration had a concordant direction of effect on HTG susceptibility. Furthermore, our efforts demonstrated that TG-raising alleles associated with population-based plasma TG concentration accumulated in HTG patients. Interestingly, the effect sizes of single nucleotide polymorphisms (SNPs) found in common between studies of HTG and population-based TG concentration suggest that increases in plasma TG are proportional to increases in HTG susceptibility (**Figure 5.1; Table 5.1**). For example, each copy of the *APOA5* risk allele (rs964184) increases HTG susceptibility by 3.3-fold, whereas this allele is only estimated to increase TG concentration by ~0.2 mmol/L. However, common TG-associated variants only explain ~10% of variation within the population (Teslovich et al., 2010), whereas they explain ~21% of variation in HTG diagnoses, perhaps indicative of underlying gene-gene or gene-environment interactions that contribute to HTG susceptibility in patients. Ultimately, these experiments have firmly established that a genetic background of TG-raising alleles is one component of HTG susceptibility.

5.1.2. Rare genetic variation in HTG

We have demonstrated that rare variation in HTG-associated genes contributes significantly to HTG susceptibility. We demonstrated that GWAS-identified genes

Figure 5.1. The effect sizes of TG-associated risk alleles on plasma TG concentration is correlated with the effect on HTG susceptibility. Effect sizes of TG-associated loci on plasma TG concentration (Teslovich et al., 2010) were compared against effect sizes on HTG susceptibility . The dotted line represents a reference trendline between 0 effect on plasma TG or HTG susceptibility and the maximal effect of the APOA5 variant (~17 mg/dL; OR=3.3). The solid line represents a trendline fitted to all data points.

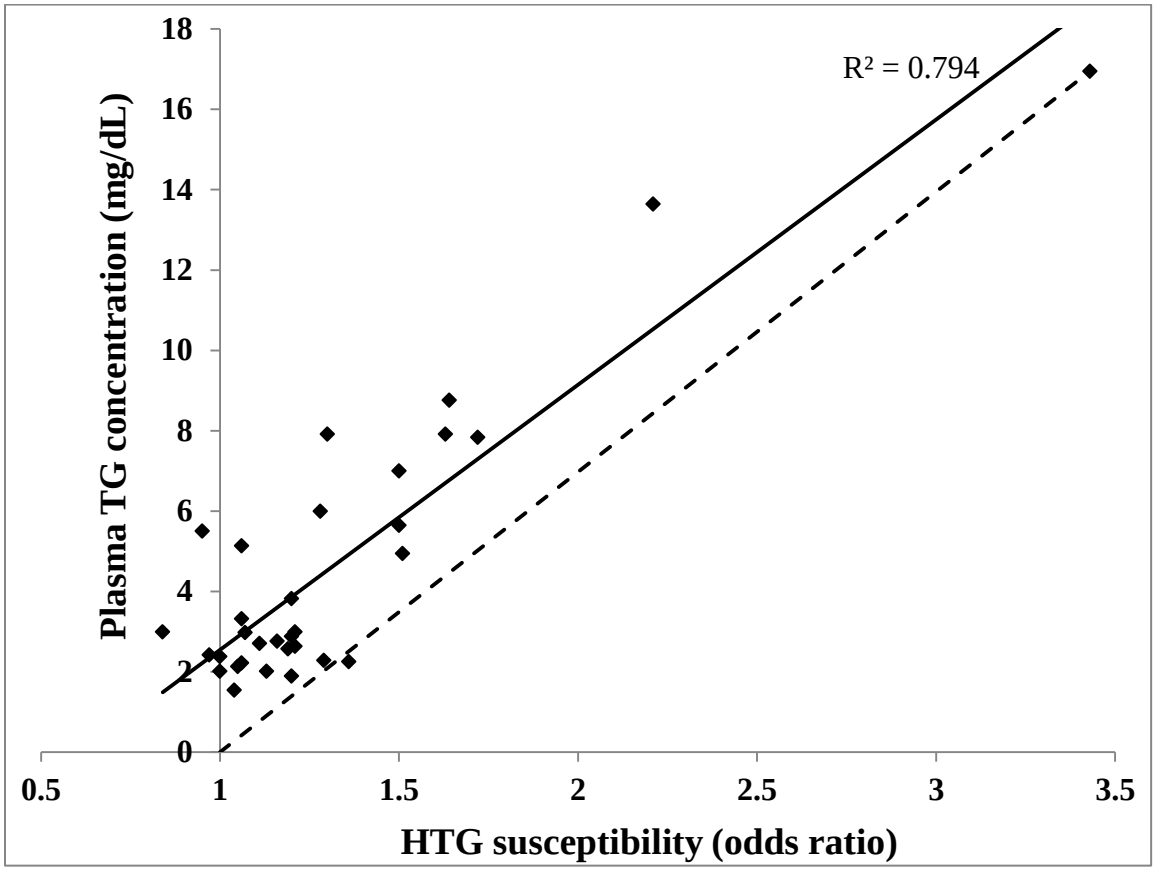

R² = 0.794
Plasma TG concentration (mg/dL)
HTG susceptibility (odds ratio)
0
2
4
6
8
10
12
14
16
18
0.5
1
1.5
2
2.5
3
3.5

Table 5.1. Effect of risk alleles at GWAS loci on HTG susceptibility and plasma TG concentration.

		HTG Susceptibility (Johansen et al., 2010)		Plasma TG concentration (Teslovich et al., 2010)	
Locus	**CHR**	**OR (95% CI)**	**P-value**	**mmol/L (mg/dL)**	**P-value**
APOA5	11	3.28 (2.61-4.14)	5.4×10^{-24}	0.19 (16.95)	7.0×10^{-240}
GCKR	2	1.75 (1.45-2.12)	6.5×10^{-9}	0.10 (8.76)	6.0×10^{-133}
LPL	8	3.13 (2.04-4.76)	2.0×10^{-7}	0.15 (13.64)	2.0×10^{-115}
APOB	2	1.67 (1.38-2.02)	2.0×10^{-7}	0.07 (5.99)	1.0×10^{-45}
MLXIPL	7	2.27 (1.61-3.23)	0.000003	0.09 (7.91)	9.0×10^{-59}
TRIB1	8	1.41 (1.16-1.69)	0.0004	0.06 (5.64)	3.0×10^{-55}
ANGPTL3	1	1.37 (1.12-1.69)	0.002	0.06 (4.94)	9.0×10^{-43}
NCAN/CILP2	19	1.41 (1.00-2.00)	0.05	0.09 (7.83)	2.0×10^{-29}
FADS1	11	1.20 (0.99-1.44)	0.07	0.04 (3.82)	5.0×10^{-24}
XKR6/PINX1	8	1.15 (0.95-1.39)	0.14	0.02 (2.01)	1.0×10^{-8}
PLTP	20	1.17 (0.94-1.47)	0.16	0.04 (3.32)	5.0×10^{-18}

CHR, chromosome; CI, confidence interval; GWAS, genome-wide association study; HTG, hypertriglyceridemia; OR, odds ratio; SNP, single nucleotide polymorphism. Effect sizes represent the additive effect of the risk allele at each locus. Association between each locus and HTG susceptibility was measured using logistic regression in a case-control study of 463 HTG patients and 1,197 healthy controls, as previously described (Johansen et al., 2010). Association between each locus and plasma TG concentration was measured using linear regression of plasma TG concentration in >100,000 healthy controls (Teslovich et al., 2010). Statistics are displayed for the lead SNP at each locus; data for identical SNPs was not available for both studies.

APOA5, *GCKR*, *LPL*, and *APOB* harbour together a significant excess of heterozygous rare variants in HTG patients versus healthy controls. Similarly, we demonstrated that non-GWAS candidate genes *APOC2*, *LMF1*, *GPIHBP1*, *CREB3L3*, and *ZHX3* also harbour together a significant excess of heterozygous rare variants in HTG patients versus healthy controls. Taken together, our data show that all genes known to cause Mendelian pediatric HTG phenotypes resulting from inherited homozygous (or compound heterozygous) mutations, including *LPL*, *APOC2*, *APOA5*, *LMF1*, and *GPIHBP1*, also harbour rare heterozygous loss-of-function variants that increase susceptibility to polygenic, adult-onset HTG. We have also established that genes recently implicated in TG metabolism by GWAS and other complementary experimental approaches, including *GCKR*, *CREB3L3*, and *APOB* also harbour rare variants associated with HTG susceptibility. We estimate that rare variation incrementally contributes ~1% of the unattributed variation in HTG diagnoses, however the size of this estimate is likely to increase with functional characterization using *in vitro* and *in vivo* models, as we predict that the true effect sizes of penetrant rare variants will approximate those of well appreciated markers of polygenic HTG such as the *APOA5* S19W common variant.

From a methodological perspective, we have contributed several conceptually novel proof-of-principle experiments. First, we demonstrated the proof-of-principle that GWAS-identified genes containing common variation associated with a complex disease also harbour rare variants that incrementally contribute to the unattributed variation in disease susceptibility. Second, we have demonstrated the utility of hypothesis-driven resequencing studies, implementing the first head-to-head comparison of gene-centric

permutation-based statistical analyses to implicate specific genes in disease pathophysiology. Third, we have demonstrated that genomic techniques complementary to GWAS will be important to fully elucidate the genetic determinants of complex phenotypes. Our resequencing experiments have additionally contributed to the development of consensus regarding methodological considerations for rare variant accumulation analyses, including experimental design, rare variant definition, and analysis approach. Ultimately, these experiments have provided a framework for the imminent tsunami of whole exome and whole genome resequencing studies involving complex phenotypes and diseases.

5.1.3. Phenotypic heterogeneity in HTG

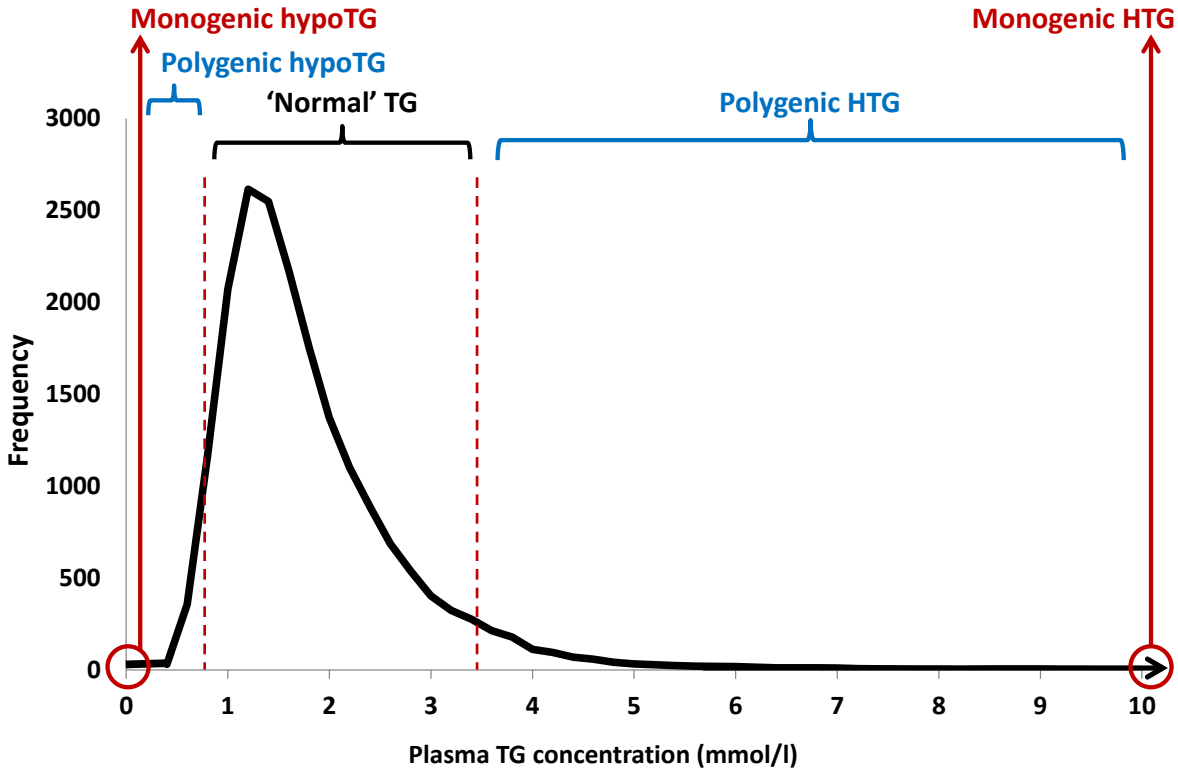
We have proposed that a shared genetic architecture of lipid-associated risk alleles underlies the biochemically-defined Fredrickson hyperlipoproteinemia (HLP) phenotypes. We conducted the most comprehensive analysis of genetic variation in the HLP-HTG phenotypes known to date, by including all common GWAS-identified variants associated with plasma lipid phenotypes by the GLGC, and integrating a comparison of rare variants in 4 HTG-associated genes among HLP-HTG phenotypes. We demonstrated that common and rare TG-associated variants accumulate equally among HLP-HTG phenotypes, whereas common high-density lipoprotein (HDL)- and LDL-associated variants accumulate in an HLP phenotype-dependent manner. Common and rare variants contributed equally to unattributed variation among HLP phenotypes 2B, 4 and 5, whereas HLP type 3 was defined primarily by the functional *APOE* E2/E2 alleles. Interestingly, we demonstrated that pleiotropic effects of genes jointly associated

with TG and HDL or LDL may contribute mechanistically to the phenotypic heterogeneity among HTG patients. Finally, we demonstrated that common TG-raising alleles accumulated equally among carriers and non-carriers of rare variants, suggesting that an overall accumulation of genetic variation contributes to HTG susceptibility. These analyses have greatly expanded to our understanding of the genetic architecture of HTG and the phenotypic heterogeneity therein.

5.1.4. Emerging model of HTG susceptibility

The genes and variants that contribute to the modulation of plasma TG concentration are better understood than those underlying most complex traits. We have demonstrated that there is a significant overlap between genes and variants that predispose subjects to one of a broad spectrum of TG-related phenotypes. This has led to our working model that the allelic spectrum of plasma TG, consisting of common and rare variation of either small or large effect that collectively determine a subject's specific plasma TG concentration, is intricately related to the phenotypic spectrum of plasma TG concentrations **(Figure 5.2.)**. Common variation may subtly improve or impair TG metabolism, whereas rare variation may have much more dramatic metabolic consequences. Importantly, it is unlikely that any single common or rare variant will be individually sufficient to cause HTG; additional gene-gene and gene-environment interactions are integral to lipoprotein metabolism and will likely be further required to produce a penetrant phenotype. Thus, our emerging opinion is that multiple common and rare variants in multiple genes with different functions contribute jointly to a given phenotype, such that extreme polygenic

Figure 5.2. The allelic and phenotypic spectrum of plasma TG concentration. A) Monogenic hypoTG is caused by rare homozygous loss-of-function (LOF) variants in genes required for lipoprotein secretion or inhibition of TG-rich lipoprotein catabolism, including *MTP*, *APOB*, and *ANGPTL3*. B) Polygenic hypoTG is likely caused by an absence of common deleterious variants (or an accumulation of common protective variants), and rare heterozygous variants with relatively large effects in *APOC3*, *ANGPTL3*, *ANGPTL4*, and *APOB*. C) Normal plasma TG concentrations likely arise from a balance of protective and deleterious common variants, with possible contributions of rare heterozygous small effect variants. D) Polygenic HTG is caused by an accumulation of common deleterious variants (or an absence of common protective variants) and rare heterozygous variants with relatively large effects in *APOA5*, *GCKR*, *LPL*, *APOB*, *APOE*, *CREBH*, *GPIHBP1*. E) Monogenic HTG is caused by rare homozygous LOF variants in genes essential for lipoprotein metabolism, including *LPL*, *APOC2*, *APOA5*, *LMF1*, and *GPIHBP1*. Dotted red lines illustrate 5^{th} and 95^{th} percentiles of plasma TG concentration. HTG, hypertriglyceridemia; hypoTG, hypotriglyceridemia; LOF, loss-of-function; TG, triglyceride.


Monogenic hypoTG
Monogenic HTG
Polygenic hypoTG
'Normal' TG
Polygenic HTG
3000
2500
2000
1500
1000
500
0
Frequency
0
1
2
3
4
5
6
7
8
9
10
Plasma TG concentration (mmol/l)

phenotypes of plasma TG manifest as a result of a genetic burden in TG-associated genes.

We speculate that elevated plasma TG concentration results from a genetic burden of deleterious TG-raising alleles (or absence of protective alleles) in TG-associated genes. Conversely, low plasma TG concentration likely indicates accumulation of protective TG-lowering alleles (or absence of deleterious alleles) in TG-associated genes. Subjects near the middle of the plasma TG concentration distribution likely have a balance of protective and deleterious small effect common variation, while those at either extreme of plasma TG have an accumulation of deleterious or protective alleles plus incremental contributions of large effect heterozygous rare variants. Patients with the most extreme phenotypes may express these phenotypes as the result of highly penetrant homozygous loss-of-function mutations.

Furthermore, the accumulation of pleiotropic TG and other lipid-associated risk alleles may contribute to phenotypic heterogeneity among HTG patients defined by the classical Fredrickson HLP phenotypes. We suggest that genetic susceptibility to HTG requires a minimal burden (‘quorum’) of common and rare variants predisposing to HTG. The additional ‘stacking’ of risk alleles on top of the background ‘quorum’ of HTG drives the phenotypic spectrum of HTG. HLP type 2B appears to be associated preferentially with common variants in pleiotropic and exclusively LDL cholesterol-associated loci; HLP type 4 appears to be associated with common variants in pleiotropic HDL cholesterol-associated loci; HLP type 3 is classically explained by the *APOE* E2/E2

isoform in addition to other TG-raising risk alleles; and the most severe phenotype HLP type 5 is associated with large effect common TG-associated risk alleles, with additional contributions from both pleiotropic LDL- and HDL-associated loci. However, there is ultimately significant overlap between lipid-associated variants and multiple HLP-HTG patients, such that the biochemically-defined HLP-HTG phenotypes are virtually indistinguishable at the level of genetic variation.

5.2. Methodological considerations for rare variant association studies

High-throughput resequencing technologies and rare variant accumulation studies are relatively novel genetic analyses that have not yet become commonplace. Methodological considerations affecting standards for appropriate design, quality control, statistical analysis and interpretation are not yet established. The experiments described in this dissertation involving rare variant accumulation are among the first of their kind. Accordingly, we have contributed to development of methodological consensus and established an analysis framework for rare variant data.

One important issue that directly influences design of re-sequencing studies is how best to define rare variants. In our resequencing study of genes identified by GWAS of HTG, we defined rare variants using a frequency threshold of <1% in healthy controls to eliminate variants with a low probability of deleterious effects. This identified a significant excess of rare nonsynonymous variants in TG-associated genes gleaned from GWAS of patients with HTG. However, contributing to the dialogue of how best to define a "rare variant" were two recent analyses using simulated data, which suggested that defining rare variants based on control-only frequency might over-estimate the extent of accumulation in cases (Lemire, 2011, Pearson, 2011). We addressed this issue with real data, using *post hoc* analyses to demonstrate that the proposed selection criteria, either defining rare variants by their frequency in the combined cohort or by their frequency in either cases or controls, do not affect our results or interpretation (**Table 5.2**) (Johansen et al., 2011). We believe these differences are probably due to certain

Table 5.2. Rare variant accumulation in genes identified by GWAS of HTG is robust regardless of rare variant definition or analysis strategy.

		Carrier test[1]			Allelic test (diploid genomes)[2]			Allelic test (total variant loci)[3]		
Selection	Rare (n)	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
"Controls"	80	2.16	1.50-3.12	$2.59X10^{-5}$	2.40	1.72-3.34	$6.31X10^{-8}$	2.16	1.58-2.95	$4.48X10^{-7}$
"Either"	81	1.91	1.35-2.72	$2.44X10^{-4}$	2.21	1.61-3.02	$3.68X10^{-7}$	1.99	1.48-2.67	$2.47X10^{-6}$
"Combined"	79	1.95	1.33-2.86	$6.11X10^{-4}$	2.20	1.55-3.12	$5.45X10^{-6}$	2.02	1.45-2.82	$1.86X10^{-5}$

Rare variant selection criteria and statistical tests follow concepts addressed in the correspondence by Lemire (Lemire, 2011). All statistical comparisons between HTG patients and controls were conducted using Fisher's exact test, considering nominal significance as a two-tailed $P<0.05$. [1]Carrier test compares carriers of ≥1 rare variant between 438 patient and 327 controls; [2]Alleleic test (diploid genomes) compares the number of variant alleles in 876 patient chromosomes and 654 control chromosomes; [3]Allelic test (total variant loci) compares the number of variant alleles at 70080 patient loci and 53320 control loci.

properties of simulated and real biological datasets, in addition to essential hypotheses underlying our study design.

Both simulations appear to be sensitive to the underlying frequency distribution of simulated variants. For instance, over-estimation of variant counts and allele frequencies introduces selection bias into permutation-based analyses when variant frequencies are recalculated after each permutation. Variants with multiple occurrences are more prone to frequency changes, since more variants are available for permutation among cases and controls, leading to variant inclusion when their frequencies are <1% in controls but exclusion when their frequencies are >1% in controls. Higher frequency variants are more likely to introduce such bias, whereas singleton variants are immune to such bias. Indeed, Pearson's analysis shows minimal bias is introduced when the variant frequency distribution is nearly identical to our sample, while increasing variant frequencies systematically introduce more bias (Pearson, 2011). Similarly, Lemire's analysis shows that selection bias is exaggerated with increasing numbers of rare variants, since more variants are available for permutation, and smaller sample sizes, where simulated variant frequencies are more liable to deviate from 'true' population frequency (Lemire, 2011). These instances of selection bias are only problematic when a restriction is placed upon rare variant inclusion in cases or controls alone; whereas a combined frequency threshold imposing no restriction upon either group protects against potential inflation of test statistics.

Despite our particular definition of rare variants, our findings resist such selection bias because the majority of variants in our combined sample have frequencies <1%. We identified 94 variants across our entire sample; 80 were designated as rare with frequencies <1% in controls. Only 2 variants were affected by alternate selection criteria: 1) *LPL* N291S (2.7% in cases and 0.8% in controls) would have been removed from our analysis using a frequency <1% in cases and controls *combined*; and 2) *GCKR* R540Q (0.9% in cases and 1.2% in controls) would have been additionally included in our analysis using the definition of a frequency <1% in *either* cases or controls, although it would have been excluded under a combined frequency threshold <1%. Furthermore, most variants in our cohort were singletons: 44 (47%) in patients but only 10 (11%) in normolipidemic controls. Thus, we observe a true excess of rare variants seen in our cases, in contrast to the distribution of some simulated frequency ratios from either Pearson or Lemire.

While a frequency threshold in the combined sample is statistically robust to rare variant selection bias, a further consideration is the biological relevance of different rare variant definitions in the context of the specific study design, since not all definitions would necessarily lend themselves to appropriate hypothesis testing. Our approach to rare variant accumulation was appropriate for our study design. First, we hypothesized that rare large-effect variants involved in HTG pathophysiology would essentially be absent from controls. We also assumed that an excess of protective rare variants would only be expected in controls drawn from a low extreme of plasma TG concentration when a gene product can affect both extreme phenotypes (e.g. "supernormal" controls <5th

percentile of TG concentration); accordingly controls were drawn from the middle of the distribution of TG concentration. Thus, the upper bound on controls could only exclude or minimize benign rare variants that accumulated both in cases and controls, but were not pathogenic in the high extreme. Second, we surmised that variants with modest effects on HTG should have been more frequent in our cohort (Manolio et al., 2009). Excluding variants with frequency >1% in the total sample would have selected against variants with modest effects on HTG (Plomin et al., 2009). For instance, the *LPL* N291S variant would have been so excluded, despite its proven functional compromise (Reymer et al., 1995). These specific examples show potential pitfalls of study designs with doctrinaire adherence to any particular definition of a rare variant.

In re-sequencing studies, power to detect associations between complex phenotypes and rare variants – individually or even cumulatively within a particular gene – is limited. Factors such as sample size, case-control ratio, ethnicity and gene size can all influence the number of identified variants. Therefore, methodologies that reduce noise from benign variants and maximize the ability to detect potentially causal variants are required. Aggregation of rare variants across multiple loci is suggested as one method to improve statistical power in re-sequencing studies (Bansal et al., 2010). Future analyses may additionally require: 1) greatly increased sample sizes to more accurately reflect 'true' variant frequencies in each group separately and in the combined dataset, preferably by increasing the number of controls especially when cases are limited; 2) selecting controls from the middle of the distribution of a quantitative trait to minimize the chance of including protective variants (depending on the underlying biology of the

gene product); and 3) reducing noise contributed by benign variants by using weighted effect sizes, such as inverse allele frequencies or by using *in silico* predictions of deleterious effect. Population-scale re-sequencing efforts, including the 1000 Genomes Project (Durbin et al., 2010), plus *in vitro* and *in vivo* characterization of variants in GWAS-identified genes, will also help define variant frequency and effect. We anticipate that no single study design or definition, including the arguably arbitrary threshold frequency of 1%, will apply uniformly across studies of rare variant accumulation.

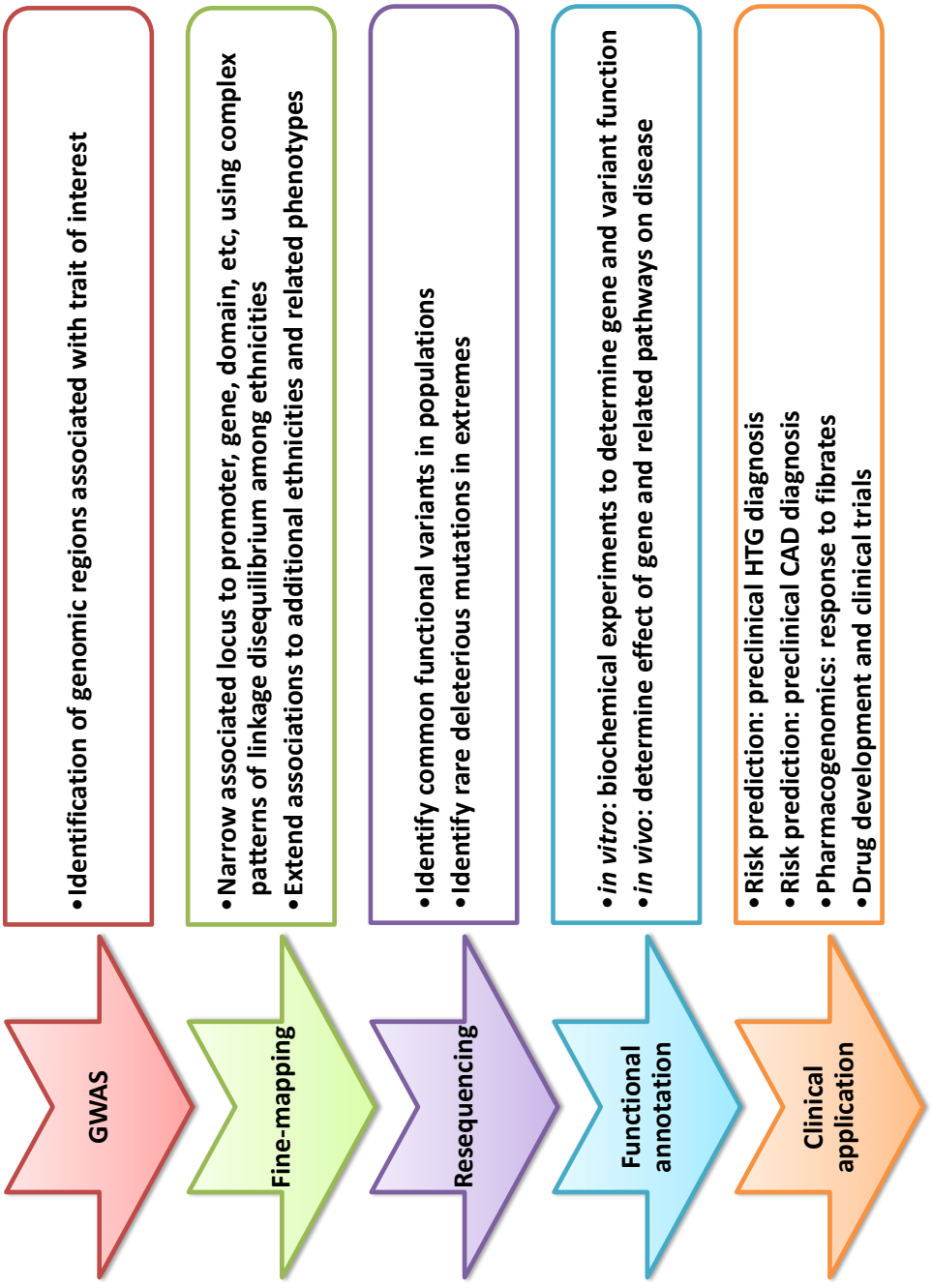
5.3. Future directions in the analysis of complex disease phenotypes

Therefore, experiments described in this dissertation have greatly contributed to our understanding of the genetic architecture of HTG. Furthermore, they have provided insight into the allelic and phenotypic spectrum of plasma TG. However, genetic variation identified to date – especially from GWAS - does not completely explain the unattributed variation in plasma TG concentration or HTG susceptibility. Accordingly, future investigations must seek to identify remaining TG-associated loci and specific functional common and rare variants that modulate of plasma TG **(Figure 5.3)**. This section describes future directions that are necessary to explain the unattributed variation within complex phenotypes including the phenotypic spectrum of plasma TG and HTG susceptibility.

5.3.1. Replication and fine mapping of GWAS loci

Replication of GWAS-identified loci in additional cohorts is required to confirm true associations and establish a priority for fine-mapping and subsequent functional characterization of genes at newly associated loci. The statistical methods used in GWAS require very large samples to generate sufficient power to robustly identify associated loci, without which false positive associations may occur. Replication of associations in independent cohorts validates the authenticity of associations which may be artifacts of chance association, population stratification or population-specific interactions. Replication in non-European samples is also necessary to confirm the involvement of loci in other ethnicities, where genomic patterns of linkage disequilibrium (LD) are not equivalent to European populations.

Figure 5.3. Sequence of experimental approaches required to achieve clinical benefit from current TG-associated loci.


GWAS
• Identification of genomic regions associated with trait of interest
Fine-mapping
• Narrow associated locus to promoter, gene, domain, etc, using complex patterns of linkage disequilibrium among ethnicities
• Extend associations to additional ethnicities and related phenotypes
Resequencing
• Identify common functional variants in populations
• Identify rare deleterious mutations in extremes
Functional annotation
• *in vitro*: biochemical experiments to determine gene and variant function
• *in vivo*: determine effect of gene and related pathways on disease
Clinical application
• Risk prediction: preclinical HTG diagnosis
• Risk prediction: preclinical CAD diagnosis
• Pharmacogenomics: response to fibrates
• Drug development and clinical trials

Following confirmation of true positive associations, fine mapping of GWAS loci is required to identify the precise genomic elements and functional variants that mediate the association. The experimental design and statistical analysis is virtually identical to GWAS, although using commercially-available high-density targeted microarrays, to effectively narrow the region of association to a more specific region within GWAS-identified loci. This process serves to focus the initial GWAS-associated locus to a region ideally containing a single candidate gene or regulatory element that could serve as a functional candidate for the association. Following identification of a functional candidate gene or locus, deep resequencing is required to identify the specific functional variants that underlie each GWAS association. Eventually, cost-effective high-throughput sequencing technologies will facilitate simultaneous rapid identification of all functional common and rare variation in GWAS-associated loci, expediting the process of fine-mapping and identification of causal genes and variants.

It is likely that one or more functional variants may underlie each GWAS association. As such, the strongest association at any given locus may not represent the only functional variant that is causally associated with a complex phenotype or disease. Thus, subsequent functional analyses are required to establish all functional variants, which will likely explain a significant portion of the unattributed variation in complex phenotypes and disease susceptibility.

5.3.2. Attributing causality to GWAS loci

The identification of specific functional variants mediating GWAS associations will permit functional characterization of the biochemical mechanisms by which genomic elements contribute to metabolism. Systematic elucidation of newly identified genes will reveal their contribution to either novel or existing biological pathways, which will expand our understanding of biological networks that contribute to disease susceptibility; thus providing novel targets for interventions. The development of a novel treatment capable of lowing plasma TG concentration that is traceable to a GWAS-identified pathway will be a monumental success for the GWAS approach.

An emerging framework capable of attributing causality to GWAS-identified loci has only been attempted for few loci to date. For instance, the LDL-associated *SORT1* locus is perhaps the best characterized of novel GWAS-identified loci (Musunuru et al., 2010b). Beginning with expression quantitative trait locus mapping, the functional variant at this locus was shown to modulate expression of several genes, but most notably *SORT1* (encoding the protein product sortilin). A few additional studies have implicated SORT1 in modulation of very low-density lipoprotein (VLDL) secretion, although controversy stills surrounds the true function of SORT1 and its involvement in lipoprotein metabolism (Dube et al., 2011). Strategies for high-throughput functional annotation will soon be required to effectively interrogate remaining GWAS-identified loci. Analysis strategies may include luciferase assays to map regulatory elements in loci associated with transcript concentrations (Jarinova et al., 2009), and murine adenoviral-based over-expression and knockdown of gene products suspected in lipoprotein

metabolism (Burkhardt et al., 2010). However, such discoveries could take many years to be validated, even with a novel straightforward mechanism without inherent biological complexities.

5.3.3. Use of complementary experimental approaches

Although GWAS is an established genomic technique capable of identifying loci strongly associated with complex phenotypes, important genes in pathophysiology are not always identified by GWAS. There are several possible reasons for this: 1) a locus encodes a crucial protein, but there is no structural or functionally relevant genetic variation in humans; 2) only very rare functional variants exist at a locus, and these have no consistent LD relationship with the common genetic variation that is genotyped on microarrays; 3) common functional genetic variation at a locus affects TG metabolism, but is not included on the genotyping microarray nor is it in strong LD with surrogate markers on microarrays; 4) common functional variation has too small an effect size to be significantly associated; 5) GWAS studies that evaluate an additive mode of inheritance may miss associations with variants that act through simple dominant or recessive models, or non-additive gene-gene or gene-environment interactions, or non-Mendelian mechanisms including inheritance of mitochondrial DNA; or 6) common variation is present at a biologically important locus, detectable and included on genotyping microarrays, but it has no functional consequence. Thus, non-GWAS methodologies such as linkage and family-based studies, or analysis of human homologs of animal models may help with gene discovery.

5.3.4. Gene-gene and gene-environment interactions

Genetic variation in concert with gene-gene and gene-environment interactions will likely explain additional unattributed variation in complex phenotypes. Gene-gene interactions (or epistasis) could lead to additive, multiplicative or synergistic effects that together modulate metabolic pathways leading to effects greater than GWAS estimates for individual variants alone (Lanktree and Hegele, 2009). Gene-environment interactions could explain how the same environmental exposure may contribute disproportionately to intermediate traits and disease susceptibility among individuals (Corella and Ordovas, 2009). Together, the analysis of gene-gene and gene-environment interactions could explain a significant portion of missing heritability in complex phenotypes (Manolio et al., 2009). However, the statistical methods necessary to analyze gene-gene and gene-environment interactions are currently very complex and not sufficiently developed for high-throughput genetic analyses (Cordell, 2009). Preliminary analyses of gene-gene and gene-environment interactions may include: 1) stratification of subjects by genetic variants or environmental/metabolic exposures; 2) controlled exposure to environmental variables; 3) inclusion of interaction terms in regression models; or 4) assessment of covariates in multivariate regression models. However, these suggested analyses are highly sensitive to power and multiple testing which severely limits their effectiveness in genome-scale experimental designs. Future detailed collection of patient data pertaining environmental exposures may also be useful for association studies, as heterogeneity among exposures could confound statistical analyses.

5.4. Clinical utility of genetic variation in complex diseases

The legacy of genetic association studies, including both GWAS and rare variant accumulation studies, will be gauged by the number of genetic loci that help to elucidate and functionally annotate disease-associated metabolic pathways. However, the greatly expanded number of TG-associated genes may challenge our ability to rapidly study their functions. Taking full advantage of the new opportunities that have arisen through the GWAS discoveries will require analogous technological advances allowing fine-mapping of association peaks, deep re-sequencing for functional variants and finally high throughput, robust functional validation at all stages: *in vitro*, *in vivo* model systems and ultimately translational human research, including clinical trials. A more complete understanding of the genes and variants that modulate plasma TG will enable development of markers for risk prediction, diagnosis, and response to therapies, and might help specify new directions for therapeutic interventions. This section will describe the potential clinical utility of genetic information gleaned from genetic association studies.

5.4.1. Genetic risk prediction

Genetic information may be useful to estimate a patient's genetic burden and predict disease susceptibility. A widely proposed idea is that genetic variables may be used to improve existing risk prediction algorithms for coronary artery disease (CAD) such as the Framingham risk score (Wilson et al., 1998). Genetic information could be entered into risk prediction equations using a single genetic variant, a genetic risk score composed of risk alleles for a single trait, or risk alleles associated with multiple independent traits.

Generally, integration of a single common variant into risk prediction equations has not yet been shown to improve discrimination, although risk stratification into low, moderate and high risk groups has shown modest improvement (Johansen and Hegele, 2009). Perhaps most successfully, a comprehensive lipid risk score composed as a weighted sum of multiple lipid-associated risk alleles from early GWAS was strongly associated with cardiovascular disease (CVD) risk, and significantly improved risk stratification in patients (Kathiresan et al., 2008). The modest improvement of risk prediction provided by genetics, while biologically important, appears to be limited due to small effect sizes. Therefore, future analyses considering the combined impact of genetic variation, including common and rare variants, may improve the performance of genetic information in risk prediction algorithms (Kraft and Hunter, 2009).

Genetic information could also permit early identification of subjects at risk for developing CVD disease, allowing prompt early lifestyle modification or evidence-based pharmacological intervention to reduce risk of clinical end points. Genetic variation in *PCSK9* that results in decreased lifelong exposure to high circulating LDL-cholesterol concentrations substantially improves CAD risk (Cohen et al., 2006, Huang et al., 2009), suggesting that lifetime lowering of CVD risk factors has the potential to greatly improve outcomes. We have shown that there is a significant difference in mean score of risk alleles in HTG patients compared to population-based controls; thus TG risk scoring could help identify subjects at risk for developing HTG and who might benefit from early intervention. However, while few HTG patients have very low risk allele scores and few healthy individuals have very high risk allele scores, there is substantial overlap toward

the centres of the distributions of risk allele scores. Emerging methodologies and future analyses that include known common functional variants and rare deleterious or protective alleles may help produce more discriminating TG risk scores.

5.4.2. Pharmacogenomics

Genetic information could also help realize the concept of personalized medicine by identifying patient-specific genetic variants for use in personalized therapy. For instance, fibrates and niacin are the primary pharmacotherapies for HTG. Inter-individual variation in plasma TG response has been reported to differ according to genotype of common *APOA5* SNPs (Lai et al., 2007) and to the presence of rare loss-of-function variants of candidate genes in TG metabolism, including *LPL*, *APOC2*, and *APOA5* (Wang et al., 2007). A complete genetic profile related to TG metabolism and possibly plasma TG responsiveness, including common and rare variants in TG metabolizing genes or drug metabolizing genes, could be used to inform such personalized treatment.

5.4.3. Mendelian randomization

Genetic information could also be used to investigate whether a causal relationship exists between TG-associated loci and CVD risk. The identification of causally associated functional variants with CAD susceptibility would identify genes that are ideal pharmacological targets. Mendelian randomization (MR) is an emerging methodology that utilizes common genetic variation as a surrogate for exposure to different dosages of an endogenous disease risk factor (Davey Smith and Ebrahim, 2003). It predicts that if a genetic variant is associated with both a biomarker and a clinical endpoint, then the

biomarker would be more likely to play a functional etiological role in the clinical outcome (**Figure 5.4**). MR is suggested to be robust and immune to confounding variables encountered in observational studies and to be capable of providing support for causality of an intermediate trait on a disease outcome. However, an important assumption underlying MR is that pleiotropy and gene-gene or gene-environment interactions are not involved in the system. The validity of MR in such scenarios will require thorough evaluation.

However, MR may prove effective at implicating plasma TG concentration in direct CVD susceptibility. For instance, one recent study has used MR to infer causality of plasma TG concentration on CAD susceptibility using genetic variation in the *APOA5* locus, specifically the -1131T>C promoter variant, to demonstrate an association with plasma TG concentration that was proportional to increased CAD risk (Sarwar et al., 2010). Similar analyses would prove useful to confirm whether additional TG associated loci are causally associated with CAD. The GLGC recently identified 6 TG-associated loci associated with CAD at a stringent threshold of significance ($P<0.001$), including *APOA5*, *LPL*, *TRIB1*, *CILP2*, *NAT2*, and *IRS1*, while 13 TG-associated loci were associated with CAD at a more modest threshold ($P<0.05$). Interestingly, among the 13 TG- and CAD-associated loci, 7 were independently associated with TG, while 6 were jointly associated with either HDL or LDL cholesterol **(Table 5.3)**. By comparison, among the 14 LDL- and CAD-associated loci, 9 were independently associated with LDL, while 5 were jointly associated with either HDL or TG. The fact that there are

Figure 5.4. Mendelian randomization and causal inference. Genetic variation is used as a surrogate exposure to biomarker or environmental variable (e.g. change in plasma TG concentration) to measure its effect on a clinical outcome (CAD). Genetic variation that is associated with a biomarker, and proportionally associated with increased disease susceptibility based on what is known about the relationship between biomarker and clinical endpoint, infers that a causal association exists between biomarker and clinical outcome.

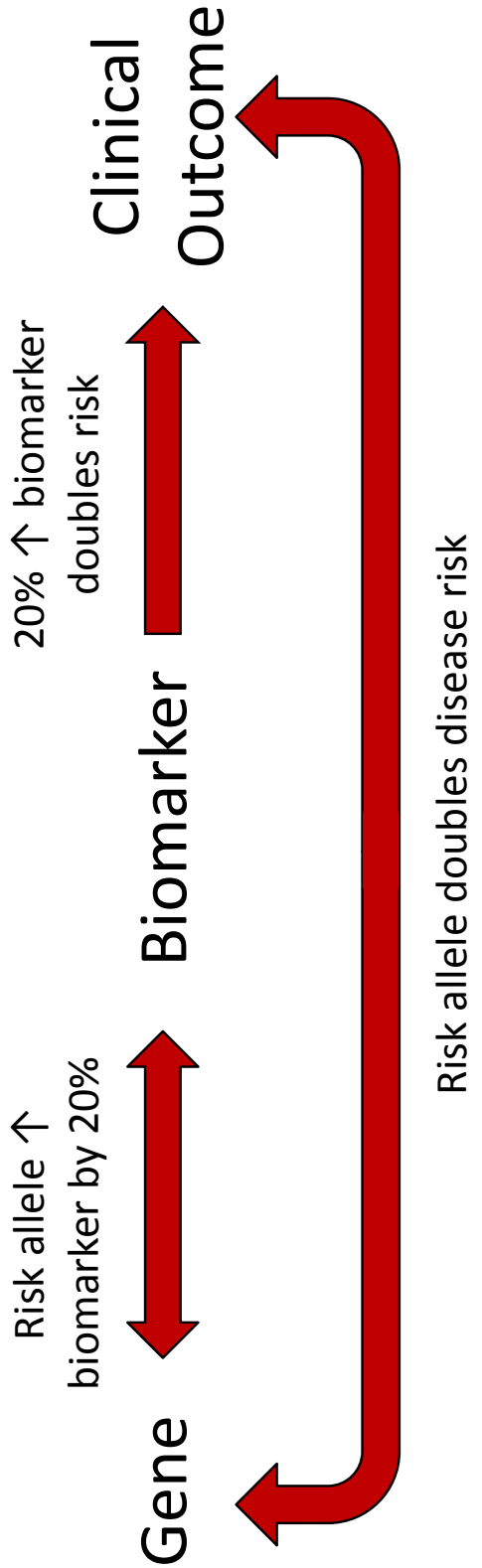

Gene
Biomarker
Clinical Outcome
Risk allele ↑ biomarker by 20%
20% ↑ biomarker doubles risk
Risk allele doubles disease risk

Table 5.3. Comparison of lipid-associated loci that increase susceptibility to CAD.

Traits	**P<0.001**	**0.001<P<0.05**	**P<0.05**
LDL only	5	4	9
HDL only	1	3	4
TG only	1	5	6
LDL + HDL	1	0	1
LDL + TG	1	0	1
HDL + TG	2	1	3
LDL + HDL + TG	2	1	3
Total	**P<0.001**	**0.001<P<0.05**	**P<0.05**
LDL	9	5	14
HDL	6	5	11
TG	6	7	13

HDL, high-density lipoprotein; LDL, low-density lipoprotein; TG, triglycerides. Associations between TG-associated loci are shown in Table 1. Results were obtained from Teslovich et al .These tallies include 1 LDL-associated locus (*BRAP*; $P=1X10^{-6}$) and 1 TG-associated locus (*PINX1*; P=0.02) that have apparently paradoxical effects on CAD risk, where lipid raising alleles do not correspond to increased CAD risk.

comparable numbers of loci for TG and LDL that are independently associated with CAD appears to challenge the widely propagated idea that LDL-associated loci are more consistently associated with CAD than are TG-associated loci (or HDL-associated loci for that matter). As additional GWAS data accrue, it will become necessary to consider and evaluate lipid-associated loci at more modest levels of significance ($0.05<P<0.001$).

5.4.4. Pharmacological interventions

The ultimate goal of any GWAS is identification of novel genomic elements involved in biological pathways that mediate disease susceptibility. Genes associated with either disease risk factors or endpoints represent novel targets for pharmacological interventions. Most importantly, genetic variation that reduces gene function yet has a favourable effect on metabolism will be most useful for inhibition using small molecule inhibitors or siRNA-based therapeutics. For instance, the study of patients with monogenic hypoTG syndromes has helped to develop pharmacological inhibitors of key genes and gene products in TG metabolism, such as lomitapide, which is a small molecule inhibitor of MTP (Samaha et al., 2008), and mipomersen as an RNA interference-based inhibitor of *APOB* (Raal et al., 2010). Future molecular inhibition of genes such as *ANGPTL3* or *APOC3* may also reduce circulating TG-rich lipoproteins, potentially without unwanted effects such as hepatosteatosis (Musunuru et al., 2010a).

5.5. Conclusions

Until recently, the genetic architecture of HTG was poorly understood. Mendelian HTG was associated with rare homozygous loss-of-function mutation in *LPL*, *APOC2*, *APOA5*, *LMF1*, and *GPIHBP1*; whereas polygenic HTG had only confirmed associations with

common functional variants in *APOA5* and *LPL*. Accordingly, we sought to comprehensively elucidate the genetic determinants of polygenic HTG using modern high-throughput genomic technologies including GWAS and high-throughput resequencing. Today, the genetic architecture of HTG is among complex disease traits most thoroughly characterized at the genetic level, including common and rare variants in >30 genes; a significant portion of this characterization has been conducted by our laboratory. We have 1) established the importance of common and rare variation in HTG susceptibility; 2) identified novel genes involved in HTG susceptibility; and 3) proposed a model for HTG susceptibility and phenotypic heterogeneity. Together with our understanding of Mendelian disorder of extreme high and low plasma TG, we have greatly contributed to a working model for the allelic and phenotypic spectrum of plasma TG. However, these analyses have only explained a modest portion of variation in plasma TG concentration. Future analyses including fine-mapping, resequencing, functional characterization of genes and variants, and higher order gene-gene and gene-environment interactions are necessary to fully explain the variation in plasma TG concentration and polygenic HTG diagnoses. Ultimately, elucidation of remaining determinants of HTG will serve to identify genes and pathways that could be targeted for pharmacological modulation in the hopes of lowering plasma TG concentration and improving CVD risk in HTG patients.

5.6. References

Bansal, V., Libiger, O., Torkamani, A., and Schork, N.J. (2010). Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet *11*, 773-785.

Burkhardt, R., Toh, S.A., Lagor, W.R., Birkeland, A., Levin, M., Li, X., Robblee, M., Fedorov, V.D., Yamamoto, M., Satoh, T., *et al.* (2010). Trib1 is a lipid- and myocardial infarction-associated gene that regulates hepatic lipogenesis and VLDL production in mice. J Clin Invest *120*, 4410-4414.

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

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

Corella, D., and Ordovas, J.M. (2009). Nutrigenomics in cardiovascular medicine. Circ Cardiovasc Genet *2*, 637-651.

Davey Smith, G., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol *32*, 1-22.

Dube, J.B., Johansen, C.T., and Hegele, R.A. (2011). Sortilin: An unusual suspect in cholesterol metabolism: From GWAS identification to in vivo biochemical analyses, sortilin has been identified as a novel mediator of human lipoprotein metabolism. Bioessays *33*, 430-437.

Durbin, R.M., Abecasis, G.R., Altshuler, D.L., Auton, A., Brooks, L.D., 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.

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

Huang, C.-C., Fornage, M., Lloyld-Jones, D.M., Wei, G.S., Boerwinkle, E., and Liu, K. (2009). Longitudinal Association of PCSK9 Sequence Variations with LDL-Cholesterol Levels: The Coronary Artery Risk Development in Young Adults (CARDIA) Study. Circ Cardiovasc Genet, CIRCGENETICS.108.828467.

Jarinova, O., Stewart, A.F., Roberts, R., Wells, G., Lau, P., Naing, T., Buerki, C., McLean, B.W., Cook, R.C., Parker, J.S., *et al.* (2009). Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. Arterioscler Thromb Vasc Biol *29*, 1671-1677.

Johansen, C.T., and Hegele, R.A. (2009). Predictive genetic testing for coronary artery disease. Crit Rev Clin Lab Sci *46*, 343-360.

Johansen, C.T., Wang, J., and Hegele, R.A. (2011). Bias due to selection of rare variants using frequency in controls. Nat Genet *43*, 394-395.

Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., *et al.* (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet *42*, 684-687.

Kathiresan, S., Melander, O., Anevski, D., Guiducci, C., Burtt, N.P., Roos, C., Hirschhorn, J.N., Berglund, G., Hedblad, B., Groop, L., *et al.* (2008). Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med *358*, 1240-1249.

Kraft, P., and Hunter, D.J. (2009). Genetic risk prediction--are we there yet? N Engl J Med *360*, 1701-1703.

Lai, C.Q., Arnett, D.K., Corella, D., Straka, R.J., Tsai, M.Y., Peacock, J.M., Adiconis, X., Parnell, L.D., Hixson, J.E., Province, M.A., *et al.* (2007). Fenofibrate effect on triglyceride and postprandial response of apolipoprotein A5 variants: the GOLDN study. Arterioscler Thromb Vasc Biol *27*, 1417-1425.

Lanktree, M.B., and Hegele, R.A. (2009). Gene-gene and gene-environment interactions: new insights into the prevention, detection and management of coronary artery disease. Genome Med *1*, 28.

Lemire, M. (2011). Defining rare variants by their frequencies in controls may increase type I error. Nat Genet *43*, 391-392.

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., *et al.* (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753.

Musunuru, K., Pirruccello, J.P., Do, R., Peloso, G.M., Guiducci, C., Sougnez, C., Garimella, K.V., Fisher, S., Abreu, J., Barry, A.J., *et al.* (2010a). Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. N Engl J Med *363*, 2220-2227.

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K.V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M., *et al.* (2010b). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature *466*, 714-719.

Pearson, R.D. (2011). Bias due to selection of rare variants using frequency in controls. Nat Genet *43*, 392-393.

Plomin, R., Haworth, C.M., and Davis, O.S. (2009). Common disorders are quantitative traits. Nat Rev Genet *10*, 872-878.

Raal, F.J., Santos, R.D., Blom, D.J., Marais, A.D., Charng, M.J., Cromwell, W.C., Lachmann, R.H., Gaudet, D., Tan, J.L., Chasan-Taber, S., *et al.* (2010). Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous familial hypercholesterolaemia: a randomised, double-blind, placebo-controlled trial. Lancet *375*, 998-1006.

Reymer, P.W., Gagne, E., Groenemeyer, B.E., Zhang, H., Forsyth, I., Jansen, H., Seidell, J.C., Kromhout, D., Lie, K.E., Kastelein, J., *et al.* (1995). A lipoprotein lipase mutation (Asn291Ser) is associated with reduced HDL cholesterol levels in premature atherosclerosis. Nat Genet *10*, 28-34.

Samaha, F.F., McKenney, J., Bloedon, L.T., Sasiela, W.J., and Rader, D.J. (2008). Inhibition of microsomal triglyceride transfer protein alone or with ezetimibe in patients with moderate hypercholesterolemia. Nat Clin Pract Cardiovasc Med *5*, 497-505.

Sarwar, N., Sandhu, M.S., Ricketts, S.L., Butterworth, A.S., Di Angelantonio, E., Boekholdt, S.M., Ouwehand, W., Watkins, H., Samani, N.J., Saleheen, D., *et al.* (2010). Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. Lancet *375*, 1634-1639.

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., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707-713.

Wang, J., Cao, H., Ban, M.R., Kennedy, B.A., Zhu, S., Anand, S., Yusuf, S., Pollex, R.L., and Hegele, R.A. (2007). Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). Arterioscler Thromb Vasc Biol *27*, 2450-2455.

Wilson, P.W., D'Agostino, R.B., Levy, D., Belanger, A.M., Silbershatz, H., and Kannel, W.B. (1998). Prediction of coronary heart disease using risk factor categories. Circulation *97*, 1837-1847.

Yuan, G., Al-Shali, K.Z., and Hegele, R.A. (2007). Hypertriglyceridemia: its etiology, effects and treatment. CMAJ *176*, 1113-1120.

APPENDICES

A-1. University of Western Ontario ethics approval

Office of Research Ethics

The University of Western Ontario
Room 4180 Support Services Building, London, ON, Canada N6A 5C1
Telephone: (519) 661-3036 Fax: (519) 850-2466 Email: ethics@uwo.ca
Website: www.uwo.ca/research/ethics

Use of Human Subjects - Ethics Approval Notice

Principal Investigator: Dr. R.A. Hegele

Review Number: 07920E **Revision Number:** 9

Review Date: November 30, 2009 **Review Level:** Expedited

Protocol Title: Candidate gene sequencing, genetic and genomic analysis for identification of new genetic determinants of intermediate traits of atherosclerosis, dyslipidemia, diabetes, obesity, hypertension, lipodystrophy and other rare metabolic or cardiovascular disorders in the human population.

Department and Institution: Vascular Biology Group, Robarts Research Institute

Sponsor: CANADIAN DIABETES ASSOCIATION & GENOME CANADA

Ethics Approval Date: November 30, 2009 **Expiry Date:** December 31, 2015

Documents Reviewed and Approved: Revised number of study participants and study end date.

Documents Received for Information:

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the HSREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of monitor, telephone number). Expedited review of minor change(s) in ongoing studies will be considered. Subjects must receive a copy of the signed information/consent documentation.

Investigators must promptly also report to the HSREB:

a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to this office for approval.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

Chair of HSREB: Dr. Joseph Gilbert

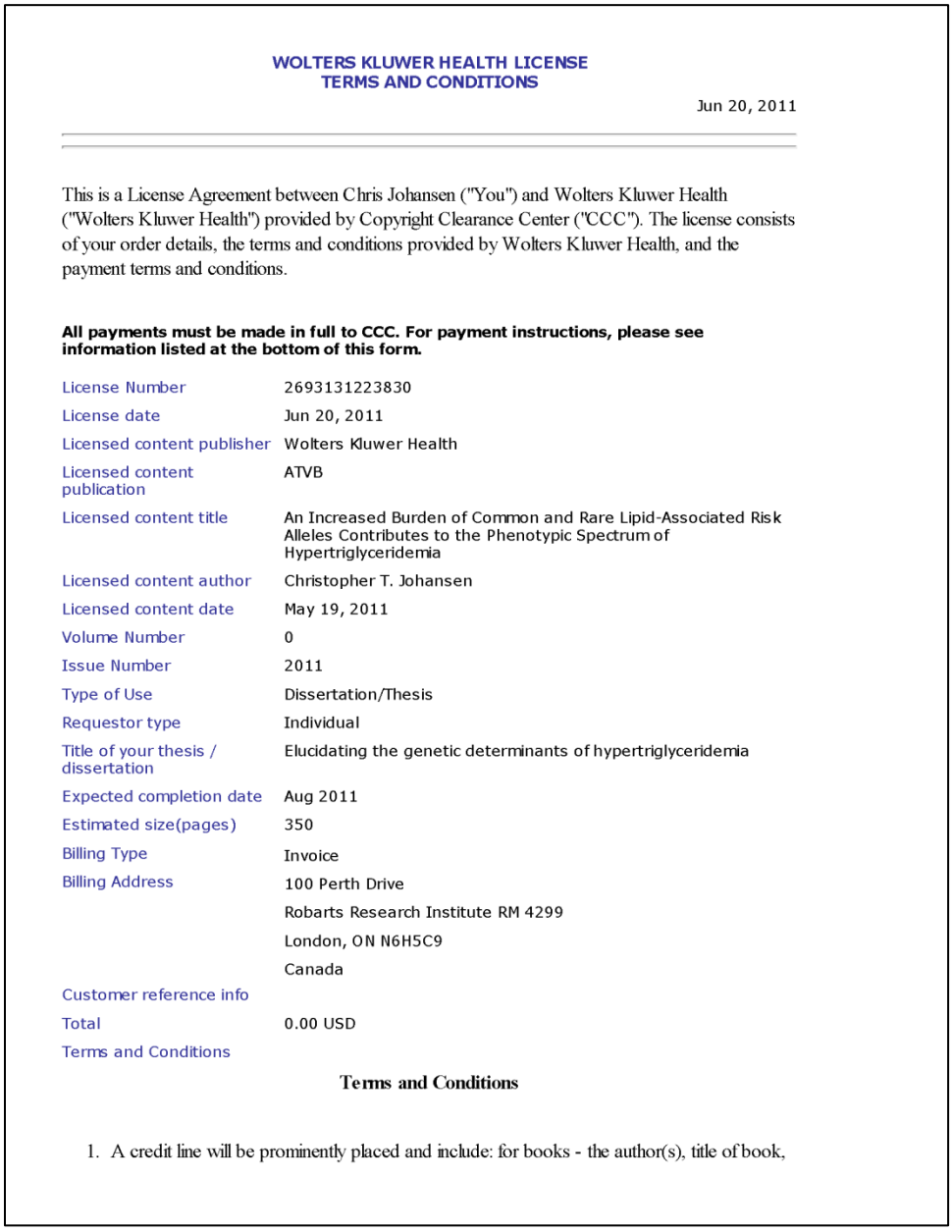
Ethics Officer to Contact for Further Information			
☐ Janice Sutherland (jsutherl@uwo.ca)	☐ Elizabeth Wambolt (ewambolt@uwo.ca)	☑ Grace Kelly (grace.kelly@uwo.ca)	☐ Denise Grafton (dgrafton@uwo.ca)

This is an official document. Please retain the original in your files.

cc: ORE File
LHRI

A-2. Copyright permissions

1. **Johansen CT**, Wang J, Lanktree MB, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Cao H, Hegele RA. (2011). An increased burden of common and rare lipid-associated risk alleles contributes to the phenotypic spectrum of hypertriglyceridemia. *Arterioscler Thromb Vasc Biol.*



WOLTERS KLUWER HEALTH LICENSE TERMS AND CONDITIONS

Jun 20, 2011

This is a License Agreement between Chris Johansen ("You") and Wolters Kluwer Health ("Wolters Kluwer Health") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Wolters Kluwer Health, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2693131223830
License date	Jun 20, 2011
Licensed content publisher	Wolters Kluwer Health
Licensed content publication	ATVB
Licensed content title	An Increased Burden of Common and Rare Lipid-Associated Risk Alleles Contributes to the Phenotypic Spectrum of Hypertriglyceridemia
Licensed content author	Christopher T. Johansen
Licensed content date	May 19, 2011
Volume Number	0
Issue Number	2011
Type of Use	Dissertation/Thesis
Requestor type	Individual
Title of your thesis / dissertation	Elucidating the genetic determinants of hypertriglyceridemia
Expected completion date	Aug 2011
Estimated size(pages)	350
Billing Type	Invoice
Billing Address	100 Perth Drive Robarts Research Institute RM 4299 London, ON N6H5C9 Canada
Customer reference info	
Total	0.00 USD
Terms and Conditions	

Terms and Conditions

1. A credit line will be prominently placed and include: for books - the author(s), title of book,

editor, copyright holder, year of publication; For journals - the author(s), title of article, title of journal, volume number, issue number and inclusive pages.

2. The requestor warrants that the material shall not be used in any manner which may be considered derogatory to the title, content, or authors of the material, or to Wolters Kluwer/Lippincott, Williams & Wilkins.
3. Permission is granted for one time use only as specified in your correspondence. Rights herein do not apply to future reproductions, editions, revisions, or other derivative works. Once term has expired, permission to renew must be made in writing.
4. Permission granted is non-exclusive, and is valid throughout the world in the English language and the languages specified in your original request.
5. Wolters Kluwer Health/ Lippincott, Williams & Wilkins, cannot supply the requestor with the original artwork or a "clean copy."
6. The requestor agrees to secure written permission from the author (for book material only).
7. Permission is valid if the borrowed material is original to a LWW imprint (Lippincott-Raven Publishers, Williams & Wilkins, Lea & Febiger, Harwal, Igaku-Shoin, Rapid Science, Little Brown & Company, Harper & Row Medical, American Journal of Nursing Co, and Urban & Schwarzenberg - English Language).
8. If you opt not to use the material requested above, please notify Rightslink within 90 days of the original invoice date.
9. Other Terms and Conditions:

v1.0

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK11006885. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact Rightslink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

2. **Johansen CT**, Wang J, and Hegele RA. (2011). Bias due to selection of rare variants using frequency in controls. *Nat Genet* **43**(5):394-5. PMID: 21522174

   No permission required:
   http://www.nature.com/reprints/permission-requests.html

3. **Johansen CT** and Hegele RA. (2011). Genetic bases of hypertriglyceridemic phenotypes. *Curr Opin Lipidol.* In press. PMID: 21519249.

WOLTERS KLUWER HEALTH LICENSE TERMS AND CONDITIONS

Au

This is a License Agreement between Chris Johansen ("You") and Wolters Kluwer Heal ("Wolters Kluwer Health") provided by Copyright Clearance Center ("CCC"). The licer of your order details, the terms and conditions provided by Wolters Kluwer Health, and payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2730260196804
License date	Aug 15, 2011
Licensed content publisher	Wolters Kluwer Health
Licensed content publication	Current Opinion in Lipidology
Licensed content title	Genetic bases of hypertriglyceridemic phenotypes.
Licensed content author	Johansen, Christopher; Hegele, Robert
Licensed content date	Jan 1, 2011
Volume Number	22
Issue Number	4
Type of Use	Dissertation/Thesis
Requestor type	Individual
Title of your thesis / dissertation	Elucidating the genetic determinants of hypertriglycerid
Expected completion date	Aug 2011
Estimated size(pages)	350
Billing Type	Invoice
Billing Address	100 Perth Drive Robarts Research Institute RM 4299 London, ON N6H5C9 Canada
Customer reference info	
Total	0.00 USD
Terms and Conditions	

4. **Johansen CT**, Kathiresan S, and Hegele RA. (2011). Genetic determinants of plasma triglycerides. *J Lipid Res*. **52**(2): 189-206. PMID: 21041806.

   No permission required:
   http://www.jlr.org/site/misc/Copyright_Permission.xhtml

5. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, **Johansen CT**, et al. (2010). Biological, clinical and population relevance of 95 loci mapped for serum lipid concentrations. *Nature* **466**(7307): 707-713. PMID: 20686565.

   No permission required:
   http://www.nature.com/reprints/permission-requests.html

6. **Johansen CT**, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, and Hegele RA. (2010). Mutation skew in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* **42**(8): 684-687. PMID: 20657596.

   No permission required:
   http://www.nature.com/reprints/permission-requests.html

7. **Johansen CT**, Lanktree MB, and Hegele RA. (2010) Translating genomic analyses into improved management of coronary artery disease. *Future Cardiol* **6**(4): 507-521. PMID: 20608823.

Permission to use Future Medicine Ltd copyright material

Request from:

- Contact name: Chris Johansen
- Publisher/company name:
- Address:
- Telephone/e-mail: christophertjohansen@gmail.com

Request details:

- Request to use the following content: Extracts - Future Cardiology, 6(4): 507-521. PMID: 20608823
- In the following publication: PhD thesis
- In what media (print/electronic/print & electronic): Print & Electronic
- In the following languages: All

We, Future Medicine Ltd, grant permission to reuse the material specified above within the publication specified above.

Notes and conditions:

1. This permission is granted free of charge, for one-time use only.
2. Future Medicine Ltd grant the publisher non-exclusive world rights to publish the content in the publication/website specified above.
3. Future Medicine Ltd retains copyright ownership of the content.
4. Permission is granted on a one-time basis only. Separate permission is required for any further use or edition.
5. The publisher will make due acknowledgement of the original publication wherever they republish the content: citing the author, content title, publication name and Future Medicine Ltd as the original publisher.
6. The publisher will not amend, abridge, or otherwise change the content without authorization from Future Medicine Ltd.
7. Permission does not include any copyrighted material from other sources that may be incorporated within the content.
8. Failure to comply with the conditions above will result in immediate revocation of the permission here granted.

Date: 21/06/2011.....

Future Medicine Ltd, Unitec House, 2 Albert Place, London, N3 1QB, UK
T: +44 (0)20 8371 6080 **F:** +44 (0) 20 8371 6099 **E:** info@futuremedicine.com
www.futuremedicine.com

CHRISTOPHER T. JOHANSEN

EDUCATION

2008-2015 **MD/PhD Program**
CIHR Fellow in Vascular Biology
Department of Biochemistry, Schulich School of Medicine and Dentistry
Robarts Research Institute, University of Western Ontario, London, Canada

2003-2008 **B.Sc. Honours Specialization in Genetics and Biochemistry**
Department of Biology, University of Western Ontario, London, Canada
Internship 2006-2007: Agriculture and Agri-Food Canada

AWARDS AND DISTINCTION

2011-2014: CIHR Scriver Family MD/PhD Studentship Award ($21,000 + $1,000)
2011: CIHR Nomination to attend Lindau Conference of Nobel Laureates
2010-2011: CIHR Banting and Best Canada Graduate Scholarship ($30,000 + $5,000)
2010-2012: Heart and Stroke Foundation of Canada DRA ($21,000 + $1,000; declined)
2010-2011: Ontario Graduate Scholarship ($15,000; declined)
2010: Department of Medicine Research Day Best Poster Award ($500)
2009-2010: CIHR Banting and Best Canada Graduate Scholarship ($17,500)
2008-2011: Western Graduate Research Scholarship ($7,000/year)
2008-2010: CIHR and HSFC Strategic Training Plan in Vascular Biology ($5000/year)
2008-2010: Heart and Stroke Foundation of Ontario Program Grant ($4000/year)
2008-2009: Ontario Graduate Scholarship ($15,000)
2009: Canadian Lipoprotein Conference Oral Presentation Award ($1000)
2008: Gold Medal for Honours Specialization in Genetics/Biochemistry (92.7%)
2008: NSERC Undergraduate Student Research Award ($4,500)
2007: Ontario Genomics Institute Summer Research Fellowship ($4,500)
2003-2008: Dean's Honor Roll
2003: University of Western Ontario Scholarship of Distinction ($1000)

RESEARCH POSITIONS

2008-Pres. **MD/PhD Supervisor: Dr. Robert A. Hegele, Cardiovascular Genetics**
Robarts Research Institute, University of Western Ontario, London, Canada

- Project focused on elucidating the genetic determinants of hypertriglyceridemia susceptibility and phenotypic heterogeneity, including contributions of common, rare and structural variation, using modern high-throughput genomic technologies
- Expert knowledge regarding the allelic and phenotypic spectrum of plasma triglycerides and other complex lipid/cardiovascular phenotypes.
- Genome-wide association analyses using specialized software packages including PLINK, EIGENSTRAT, MACH, METAL, Perl, SAS v9.2.
- Rare variation accumulation studies using Sanger and high-throughput sequencing data using software packages including PLINKSEQ.

- Structural variation analysis using Affymetrix microarray-based genotyping and software packages including Genotyping Console, Partek, and PLINK.

2006-2008 **B.Sc Thesis Supervisor: Dr. Krzysztof Szczyglowski, Plant Genetics**
Agriculture and Agri-Food Canada, London, ON, Canada

- Integrated zinc-finger nuclease technology into the laboratory, establishing a protocol for targeted mutagenesis and genomic manipulation in *Lotus japonicus*
- Characterization of histidine kinase receptors using cloning, in vitro mutagenesis techniques, promoter-reporter constructs, and *in vivo* bacterial and yeast complementation assays
- Map-based cloning to link defined mutant nodulation phenotypes with discrete positions in the *Lotus japonicus* genome

MENTORSHIP OF MASTERS/UNDERGRADUATE STUDENTS

2011: Joseph B. Dube "Elucidating the genetic determinants of cognitive impairment"
2011: Omar Abdel-Razek "Copy number variation in hypertriglyceridemia"
2010: Justine Denomme "Lipoprotein(a) and cardiovascular risk"
2009: Neil Nsueh "Pharmacogenomics of fibrate response in hypertriglyceridemia"
2008: Tim Lin "Pharmacogenomics of statin myopathy"

PEER REVIEWED PUBLICATIONS

1. Lee JH, Giannikopoulos P, Duncan SA, Wang J, **Johansen CT**, Brown JD, Plutzky J, Hegele RA, Glimcher LH, Lee AH. (2011). The transcription factor cyclic AMP-responsive element-binding protein H regulates triglyceride metabolism. *Nat Med*. **17**(7): 812-815. PMID: 21666694.
2. **Johansen CT**, Wang J, Lanktree MB, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Cao H, Hegele RA. (2011). An increased burden of common and rare lipid-associated risk alleles contributes to the phenotypic spectrum of hypertriglyceridemia. *Arterioscler Thromb Vasc Biol*. **31**(8): 1916-26. PMID: 21597005.
3. **Johansen CT**, Wang J, and Hegele RA. (2011). Bias due to selection of rare variants using frequency in controls. *Nat Genet* **43**(5):394-5. PMID: 21522174
4. **Johansen CT** and Hegele RA. (2011). Genetic bases of hypertriglyceridemic phenotypes. *Curr Opin Lipidol*. **22**(4): 247-53. PMID: 21519249.
5. Dubé JB, **Johansen CT**, Hegele RA. (2011). Sortilin: An unusual suspect in cholesterol metabolism: From GWAS identification to in vivo biochemical analyses, sortilin has been identified as a novel mediator of human lipoprotein metabolism. *Bioessays* **33**(6):430-7. PMID: 21462369.
6. **Johansen CT**, Kathiresan S, and Hegele RA. (2011). Genetic determinants of plasma triglycerides. *J Lipid Res*. **52**(2): 189-206. PMID: 21041806.

7. Lanktree MB, **Johansen CT**, Joy TR, Hegele RA. (2010). A translational view of the genetics of lipodystrophy and ectopic fat deposition. *Prog Mol Biol Transl Sci* **94**: 159-96. PMID: 21036325.
8. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, **Johansen CT**, et al. (2010). Biological, clinical and population relevance of 95 loci mapped for serum lipid concentrations. *Nature* **466**(7307): 707-713. PMID: 20686565.
9. **Johansen CT**, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, and Hegele RA. (2010). Mutation skew in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* **42**(8): 684-687. PMID: 20657596.
10. **Johansen CT**, Lanktree MB, and Hegele RA. (2010) Translating genomic analyses into improved management of coronary artery disease. *Future Cardiol* **6**(4): 507-521. PMID: 20608823.
11. Lanktree MB, **Johansen CT**, Anand SS, Yusuf S, Hegele RA. (2010). Genetic variation in hyaluronan metabolism loci is associated with plasma plasminogen activator inhibitor-1 concentration. *Blood* **116**(12): 2160-2163. PMID: 20558613.
12. **Johansen CT**, Gallinger Z, Young TK, Bjerregaard P, and Hegele RA. (2010). ATGL haplotypes are associated with increased plasma triglycerides in the Greenland Inuit. *Int J Circumpolar Health.* **69**(1):3-12. PMID: 20167152.
13. **Johansen, CT** and Hegele, RA. (2009) Genetic testing for Coronary Artery Disease. *Crit Rev Clin Lab Sci* **46**(5-6): 343-360. PMID: 19958218.
14. Kosuta S, Held M, Hossain MS, Morieri G, Macgillivary A, **Johansen C**, Antolín-Llovera M, Parniske M, Oldroyd GE, Allan Downie J, Karas B, Szczyglowski K. (2011). Lotus japonicus symRK-14 uncouples the cortical and epidermal symbiotic program. *Plant J*. In press. PMID: 21595760
15. Karas B, Amyot L, **Johansen C**, Sato S, Tabata Y, Kawaguchi M, and Szczyglowski K. (2009) Functional conservation of *Lotus japonicus* and *Arabidopsis* basic helix-loop-helix proteins reveal new players in root hair development. *Plant Physiol* **151**(3): 1175-1185. PMID: 19675148.

PARTICIPATION IN PEER REVIEW PROCESS

Science (3), Nature Genetics (4), Nature Reviews Cardiology (1), Circulation (1), PLoS Genetics (1), Journal of Clinical Investigation (1), ATVB (1), Human Molecular Genetics (1), Journal of the American College of Cardiology (1), Other (7)

PREVIOUS WORK/LEADERSHIP/VOLUNTEER EXPERIENCE

2006-2007: Nursery Worker, Agriculture and Agri-Food Canada
2002-2006: Shift Supervisor, Starbucks Coffee Company
2006-2007: London Regional Mental Health Hospital Volunteer
Volunteered as a tutor at the institution's youth education centre
2004-2005: Soph, University of Western Ontario
Student leader during University Orientation Week
2002-2003: Nursery Worker, Weall & Cullen Nurseries

1998-2003: Squadron Commander, Royal Canadian Air Cadets

ORAL AND POSTER ABSTRACT PUBLICATIONS

1. **Johansen CT**. (2011). Genetic architecture of plasma triglycerides and hypertriglyceridemia. Oral presentation at UWO Department of Medicine Research Day, London, Ontario.
2. **Johansen CT**, Wang J, McIntyre AD, Martins RA, Ban MR, Lanktree MB, Huff MW, Péterfy M, Mehrabian M, Lusis AJ, Kathiresan S, Anand SS, Yusuf S, Lee AH, Glimcher LH, Cao H and Hegele RA. (2011). Excess of rare variants in non-GWAS candidate genes in patients with HTG. Oral presentation at Arteriosclerosis, Thrombosis and Vascular Biology Annual Meeting, Chicago, Illinois, USA.
3. **Johansen, CT**, Lanktree, MB, Wang, J, Ban, MR, Martins, RA, Kennedy, BA, Hassell, RG, Visser, ME, Schwartz, SM, Voight, BF, Elosua, R, Salomaa, V, O'Donnell, CJ, Dallinga-Thie, GM, Anand, SS, Yusuf, S, Huff, MW, Kathiresan, S, and Hegele, RA. (2010). Genome-wide association study of patients with severe hypertriglyceridemia. Oral presentation at the Canadian Lipoprotein Conference 2010, Niagara-On-The-Lake, Ontario, Canada.
4. **Johansen, CT**, Lanktree, MB, Wang, J, Ban, MR, Martins, RA, Kennedy, BA, Hassell, RG, Visser, ME, Schwartz, SM, Voight, BF, Elosua, R, Salomaa, V, O'Donnell, CJ, Dallinga-Thie, GM, Anand, SS, Yusuf, S, Huff, MW, Kathiresan, S, and Hegele, RA. (2010). Genome-wide association study of patients with severe hypertriglyceridemia. Oral presentation at the American Society of Human Genetics 2010, Washington, District of Columbia, USA.
5. **Johansen, CT**, Wang, J, Lanktree, MB, Ban, MR, Martins, RA, Kennedy, BA, Hassell, RG, Danielson, R, Rogan, PK, Anand, SS, Yusuf, S, Huff, MW, and Hegele, RA. (2010). Mutation skew in triglyceride-metabolizing genes associated with hypertriglyceridemia. Poster presented at the Gordon Research Conference in Lipid and Lipoprotein Metabolism, Waterville Valley, New Hamshire, USA.
6. Lanktree, MB, **Johansen, CT**, Anand, SS, Yusuf, S, Hegele, RA. (2010) Lipid associated genes drive the accumulation of metabolic risk alleles in subjects with metabolic syndrome. Poster presented at the Gordon Research Conference in Lipid and Lipoprotein Metabolism, Waterville Valley, New Hamshire, USA.
7. **Johansen, CT**, Lanktree, MB, Wang, J, Ban, MR, Martins, RA, Kennedy, BA, Hassell, RG, Visser, ME, Schwartz, SM, Voight, BF, Elosua, R, Salomaa, V, O'Donnell, CJ, Dallinga-Thie, GM, Anand, SS, Yusuf, S, Huff, MW, Kathiresan, S, and Hegele, RA. (2010). Genome-wide association study of patients with severe hypertriglyceridemia. Poster presented at Arteriosclerosis, Thrombosis and Vascular Biology Annual Meeting, San Francisco, California, and 5th Annual Canadian Genetic Epidemiology and Statistical Genetics Meeting, Toronto, Ontario.
8. Lanktree MB, **Johansen CT**, Yusuf S, Anand SS, Hegele RA (2010). Genetic variation in hyaluronan metabolism genes is associated with plasma plasminogen activator inhibitor-1. Oral presentation at Arteriosclerosis, Thrombosis and Vascular Biology Annual Meeting, San Francisco, California, and 5th Annual Canadian Genetic Epidemiology and Statistical Genetics Meeting, Toronto, Ontario.
9. **Johansen, CT**, Lanktree, MB, Anand, SS, Yusuf, S, and Hegele, RA. (2009). Common genetic determinants of normolipidemic plasma triglyceride concentrations

are over-represented in patients with severe hypertriglyceridemia. Oral presentation at the Canadian Lipoprotein Conference 2009, Caesar's Hotel and Casino, Windsor, Ontario, Canada.

10. Phillips, MS, Brown, AMK, Dubé, MP, **Johansen, CT**, Hegele, RA, and Tardif, JC. (2009). Statin-induced myotoxicity adverse drug reactions: pharmacogenomic evaluation of SLCO1B1 and Atogen-1 in Statin Myotoxic Cohorts. Poster presented at the 59th Annual Meeting of the American Society of Human Genetics 2009, Honolulu, Hawaii, United States.
11. **Johansen, C.T.**, Lanktree, M.B., Anand, S.A., Yusuf, S., and Hegele, R.A. Elucidating the genetic determinants of severe hypertriglyceridemia." Oral presentation at the Genome Canada Project Team Meeting, Toronto, ON, Canada. September 2009.
12. **Johansen, C.**T., Phillips, M.S., Dubé, M.P., Wang, J., Lin, T., Kennedy, B.A., Ban, M.R., Tardif, J.C., and Hegele, R.A. "SLCO1B1 is not associated with statin-induced myopathy in patients from tertiary referral lipid clinics." Oral presentation at the MD/PhD Research Day, University of Western Ontario, London, ON, Canada. May 2009.
13. **Johansen, CT**, Phillips, MS, Dubé, MP, Wang, J, Lin, T, Kennedy, BA, Ban, MR, Tardif, JC, and Hegele, RA. SLCO1B1 is not associated with statin-induced myopathy in patients from tertiary referral lipid clinics. Poster presented at Arteriosclerosis, Thrombosis, and Vascular Biology Annual Conference 2009, Washington, D.C., United States, and the 2nd Annual Canadian Human Genetics and 4th Annual Canadian Genetic Epidemiology and Statistical Genetics Meetings 2009, Harrison Hot Springs, British Colombia, Canada.
14. **Johansen, CT**, Gallinger, ZR, Wang, J, Ban, MR, Young, TK, Bjerregaard, P, and Hegele, RA. ATGL haplotypes are associated with plasma triglyceride concentrations in the Greenland Inuit. Poster presented at Arteriosclerosis, Thrombosis, and Vascular Biology Annual Conference 2009, Washington, D.C., United States.
15. Held, M, **Johansen, C**, Perry, J, Wang, T, and Szczyglowski, K. 2009. Characterization of the *Lotus japonicus* cytokinin receptor gene family and their role in symbiosis. Oral presentation at the 2009 Model Legume Congress, Asilomar, California, USA.
16. Kosuta, S, Karas, B, MacGillivary, A, Held, M, **Johansen, C**, and Szczyglowski, K. 2007. An evolutionarily-conserved GDPC motif in the extracellular domain of symRK is required for arbuscular mycorrhiza and root nodule symbioses. Poster presented at the XIII International Congress on Molecular Plant-Microbe Interaction, pp. 111, Sorrento, Italy.
17. Karas, B, Murray, J, Ross, L, Nowakowski, K, **Johansen, C**, Sato, S, Tabata, S, and Szczyglowski, K. **2007**. *Lotus japonicus LjVRH1/LjSRH1* encodes a cellulose synthase-like D protein required for root hair morphogenesis and root nodule symbiosis. Poster presented at the XIII International Congress on Molecular Plant-Microbe Interaction, pp. 154, Sorrento, Italy.
18. Kosuta, S, MacGillivary, A, Held, M, **Johansen, C**, Karas, B, and Szczyglowski, K. 2007. *symRK-14*, a new allele with attitude. Invited oral presentation at the XX North American Symbiotic Nitrogen Fixation Conference, Milwaukee, Wisconsin, USA.

19. **Johansen, C**, Voytas, D, and Szczyglowski, K. 2007. Zinc-Finger Scissors for Plant Genome Modification. Oral and poster presentation at the Ontario Genomics Institute Summer Research Fellowship Culmination Event, Toronto, Ontario, Canada.