




2016

Mining Genomic Variants And Causal Pathways Linking Hdl And Triglycerides To Coronary Disease

Sumeet Anand Khetarpal
University of Pennsylvania, sumeetak@mail.med.upenn.edu

Follow this and additional works at: <https://repository.upenn.edu/edissertations>

 Part of the [Genetics Commons](#), [Medicine and Health Sciences Commons](#), and the [Molecular Biology Commons](#)

Recommended Citation

Khetarpal, Sumeet Anand, "Mining Genomic Variants And Causal Pathways Linking Hdl And Triglycerides To Coronary Disease" (2016). *Publicly Accessible Penn Dissertations*. 2387.
<https://repository.upenn.edu/edissertations/2387>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/2387>
For more information, please contact repository@pobox.upenn.edu.

Mining Genomic Variants And Causal Pathways Linking Hdl And Triglycerides To Coronary Disease

Abstract

Blood lipids are important biomarkers of risk of coronary heart disease (CHD), the leading cause of death in the world. Myriad data support a causal role of low-density lipoprotein cholesterol (LDL-C) in increasing risk of CHD. Long-standing epidemiology suggests that high-density lipoprotein cholesterol (HDL-C) may protect from disease while high triglycerides (TGs) increase CHD risk. However, the causality of HDL-C and TG to CHD remains controversial. New genetic methodologies have allowed a better look into causal pathways underlying relationships between these traits and disease. Using a combination of approaches for interrogating rare genetic variation in humans, we investigated how HDL and TG may relate to CHD. First, through sequencing and exome-wide genotyping of subjects with extremely high HDL-C, we identified the first homozygote for a loss-of-function (LOF) variant in *SCARB1*, which encodes scavenger receptor class BI (SR-BI), a hepatic receptor for HDL-C. Despite markedly elevated HDL-C, carriers of this variant had an increased risk of CHD. These findings suggest that HDL functionality in driving cholesterol removal through SR-BI (the reverse cholesterol transport hypothesis) is protective from CHD in humans. Next, we functionally examined one of the first novel loci from genome-wide association studies (GWAS) for HDL-C, *GALNT2*. Through discovery of humans with genetic *GALNT2* LOF and additional studies in rodents and nonhuman primates, we showed that *GALNT2* LOF lowers HDL-C across mammals. We also identify one physiological mechanism linking *GALNT2* to HDL-C through its enzymatic function. Thirdly, we studied the mechanism of protection of the *APOC3* A43T variant recently reported to lower TGs and CHD risk from exome sequencing. Studies in human carriers and animal models suggest that A43T accelerates renal clearance of circulating ApoC-III, thus hindering its function in delaying TG-rich lipoprotein turnover. These data establish ApoC-III clearance mechanisms as potential therapeutic targets for TG lowering. Finally, we adapted a targeted sequencing approach to increase discovery of causal rare coding and noncoding variants at candidate loci influencing HDL-C and TGs. Collectively, this work provides a sampling of approaches for leveraging the spectrum of genomic methods to identify clinically relevant variants impacting HDL, TG and CHD risk.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Cell & Molecular Biology

First Advisor

Daniel J. Rader

Keywords

cholesterol, heart disease, human genetics, lipids, triglycerides

Subject Categories

Genetics | Medicine and Health Sciences | Molecular Biology

MINING GENOMIC VARIANTS AND CAUSAL PATHWAYS LINKING HDL AND
TRIGLYCERIDES TO CORONARY DISEASE

Sumeet Anand Khetarpal

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

Daniel J. Rader

Professor, Departments of Genetics and Medicine

Graduate Group Chairperson

Daniel S. Kessler

Associate Professor, Department of Cell and Developmental Biology

Dissertation Committee

Zoltan P. Arany, Associate Professor, Department of Medicine

Maja Bucan, Professor, Department of Genetics

Benjamin A. Garcia, Professor, Department of Biochemistry and Biophysics

Mitchell A. Lazar, Professor, Departments of Genetics and Medicine

MINING GENOMIC VARIANTS AND CAUSAL PATHWAYS LINKING HDL AND
TRIGLYCERIDES TO CORONARY DISEASE

COPYRIGHT

2016

Sumeet Anand Khetarpal

ACKNOWLEDGMENT

Over the course of my dissertation work, I have been fortunate to receive the support, collaboration, mentorship, and friendship of literally hundreds of people, both near and far. Firstly, I would like to acknowledge the mentorship and support of Daniel Rader, my thesis advisor, who has guided my outlook on science and medicine for more than nine years and given me innumerable opportunities to pursue the training and questions that have excited me immensely. I also wish to acknowledge the thoughtful insight, critiques, and dedication of my thesis committee, consisting of Mitchell Lazar (chair), Maja Bucan, Benjamin Garcia, Zoltan Arany, and Morris Birnbaum (former member).

The entirety of the Rader laboratory and its friends have provided invaluable guidance, assistance, feedback and reassurance during my training, and I wish to thank all of the former and present Rader lab members. I would especially like to acknowledge the guidance and collaboration of Marina Cuchel, Nicholas Hand, Jeffrey Billheimer, John Millar, Athanasia Skoura, Xin Bi, Donna Conlon, Ali Javaheri, Sylvia Nürnberg, Cecilia Vitali, Stephanie DerOhannessian, Dawn Marchadier, Linda Carmichael, Debra Cromley, Edwige Edouard, Susannah Elwyn, Amrith Rodrigues, Maosen Sun, Teo Tran, Kevin Trindade, Aisha Wilson, Paolo Zanoni, William Hancock-Cerutti, Jennie Lin, Amritha Varshini, Wei Zhao, James McParland, Marjorie Risman, Mary McCoy, and Cathy Warford throughout my training. It has also been a distinct privilege to receive the mentorship of Michael Phillips and Sissel Lund-Katz, who have stewarded my scientific progress since my undergraduate years with patience and unwavering enthusiasm. In addition, I would particularly like to recognize the constant mentorship and wisdom of former Rader lab members William Lagor, Robert Bauer, Andrew Edmondson, Amit Khera, Arman Qamar and Kashif Jafri, who have guided my

scientific and personal growth throughout my medical and graduate training thus far and who will continue to be mentors for me as I move forward.

I have been very fortunate to enjoy close collaborations with many scientists and scholars at Penn and across the country and world during my training. I would especially like to acknowledge Muredach Reilly, Kiran Musunuru, Danish Saleheen, Paul Babb, Benjamin Voight, Casey Brown, YoSon Park, Walter Englander, Leland Mayne, Palani Chetty, and David Nguyen for several rewarding endeavors together at Penn. I also wish to thank Sekar Kathiresan, Gina Peloso, Nathan Stitzel, and Pradeep Natarajan of the Broad Institute / MGH, Nathan Yates and Xuemei Zeng of the University of Pittsburgh, Katrine Schjoldager and Henrik Clausen of the University of Copenhagen, as well as Todd Kirchgessner of Bristol-Myers Squibb for invaluable collaborations and learning together.

My training would not be possible without the support of the Perelman School of Medicine, the Cell and Molecular Biology graduate program, and especially the Medical Scientist Training Program. In particular, I would like to acknowledge Lawrence Brass and Maggie Krall for their patience and guidance. I would also like to acknowledge the support of the NIH, which has provided a predoctoral fellowship (F30 NRSA) for my training and has supported the majority of my projects. I acknowledge the support of the Foundation Leducq and the NIH-NHLBI Gene Therapy Resource Program.

Finally, I thank all of the members of my family for their endless and selfless support during all of the phases of my education.

ABSTRACT

MINING GENOMIC VARIANTS AND CAUSAL PATHWAYS LINKING HDL AND TRIGLYCERIDES TO CORONARY DISEASE

Sumeet A. Khetarpal

Daniel J. Rader

Blood lipids are important biomarkers of risk of coronary heart disease (CHD), the leading cause of death in the world. Myriad data support a causal role of low-density lipoprotein cholesterol (LDL-C) in increasing risk of CHD. Long-standing epidemiology suggests that high-density lipoprotein cholesterol (HDL-C) may protect from disease while high triglycerides (TGs) increase CHD risk. However, the causality of HDL-C and TG to CHD remains controversial. New genetic methodologies have allowed a better look into causal pathways underlying relationships between these traits and disease. Using a combination of approaches for interrogating rare genetic variation in humans, we investigated how HDL and TG may relate to CHD. First, through sequencing and exome-wide genotyping of subjects with extremely high HDL-C, we identified the first homozygote for a loss-of-function (LOF) variant in *SCARB1*, which encodes scavenger receptor class BI (SR-BI), a hepatic receptor for HDL-C. Despite markedly elevated HDL-C, carriers of this variant had an increased risk of CHD. These findings suggest that HDL functionality in driving cholesterol removal through SR-BI (the reverse cholesterol transport hypothesis) is protective from CHD in humans. Next, we functionally examined one of the first novel loci from genome-wide association studies (GWAS) for HDL-C, *GALNT2*. Through discovery of humans with genetic *GALNT2* LOF and additional studies in rodents and nonhuman primates, we showed that *GALNT2* LOF lowers HDL-C across mammals. We also identify one physiological mechanism

linking GALNT2 to HDL-C through its enzymatic function. Thirdly, we studied the mechanism of protection of the APOC3 A43T variant recently reported to lower TGs and CHD risk from exome sequencing. Studies in human carriers and animal models suggest that A43T accelerates renal clearance of circulating ApoC-III, thus hindering its function in delaying TG-rich lipoprotein turnover. These data establish ApoC-III clearance mechanisms as potential therapeutic targets for TG lowering. Finally, we adapted a targeted sequencing approach to increase discovery of causal rare coding and noncoding variants at candidate loci influencing HDL-C and TGs. Collectively, this work provides a sampling of approaches for leveraging the spectrum of genomic methods to identify clinically relevant variants impacting HDL, TG and CHD risk.

TABLE OF CONTENTS

Acknowledgement	iii
Abstract	v
Table of Contents	vii
List of Tables	xii
List of Illustrations.....	xiv
CHAPTER 1 – INTRODUCTION.....	1
Heart Disease – Epidemiology and Pathophysiology	1
Blood Lipids and CHD Risk – the Evidence for Causality	5
Low-density lipoproteins	5
High-density lipoproteins.....	9
Triglyceride-rich lipoproteins	13
ApoC-III	15
ApoA-V	16
Angiopietin-like Proteins 3 and 4	16
Next-generation Genomic Approaches to Studying Plasma Lipids	18
Genome-wide association studies for plasma lipids	20
Upstream issues—from locus to causal variant.....	21
Downstream issues—from variants to causal genes and biological mechanisms.....	24
New biology inspired by common variant studies.....	26
<i>SORT1</i> (sortilin)	27
<i>TRIB1</i> (tribbles-1).....	30
<i>GALNT2</i> (polypeptide N-acetylgalactosaminyltransferase 2)	31

Looking ahead – leveraging low-frequency variants, Mendelian randomization, and sequencing for rare causal variants.....	33
Overall Summary and Dissertation Goals.....	39

CHAPTER 2 – RARE VARIANT IN SCAVENGER RECEPTOR BI RAISES HDL

CHOLESTEROL AND INCREASES RISK OF CORONARY HEART DISEASE 45

Introduction.....	45
Materials and Methods	47
Results.....	61

Identification of <i>SCARB1</i> P376L Homozygote and Association with High HDL-C	61
---	----

Conservation of SR-BI Proline 376 Across Species and Paralogues	62
--	----

<i>SCARB1</i> P376L results in complete loss-of-function of SR-BI through abolishing post-translational processing and cell-surface localization	63
--	----

HDL-related phenotypes of <i>SCARB1</i> P376L homozygote and heterozygotes	65
---	----

<i>SCARB1</i> P376L is associated with increased risk of CHD in humans	67
---	----

Discussion	68
------------------	----

CHAPTER 3 – LOSS-OF-FUNCTION OF GALNT2 LOWERS HIGH-DENSITY

LIPOPROTEINS IN HUMANS, NONHUMAN PRIMATES, AND RODENTS 88

Introduction.....	88
-------------------	----

Materials and Methods	89
Results.....	107
Humans homozygous for LOF mutations in <i>GALNT2</i> have low HDL-C	107
Reduced HDL-C in rodent and nonhuman primate models of <i>GALNT2</i> deficiency	109
Impact of <i>GALNT2</i> deficiency in mice on HDL-C and TG turnover..	111
Knockdown of hepatic <i>GALNT2</i> in non-human primates reduces HDL-C levels.....	111
Quantitative differential O-glycoproteomics confirm ANGPTL3 and identify PLTP as O-glycoprotein targets controlled by GalNAc-T2.....	112
<i>GALNT2</i> SNPs associated with lower HDL-C are associated with reduced <i>GALNT2</i> expression in human liver	116
Discussion	117

CHAPTER 4 – APOC3 A43T VARIANT CONFERS VARIANT CONFERS LOSS-OF-FUNCTION THROUGH DISRUPTING APOC-III BINDING AND ACCELERATING TRIGLYCERIDE-RICH LIPOPROTEIN CLEARANCE.....	153
Introduction.....	153
Materials and Methods	154
Results.....	170
Human <i>APOC3</i> A43T heterozygotes have reduced TRLs and reduced ApoC-III.....	170
<i>APOC3</i> A43T confers improved TRL metabolism in mice in vivo....	174

A43T increases clearance and renal excretion of circulating ApoC-III	177
Discussion	180
CHAPTER 5 – MULTIPLEXED TARGETED RESEQUENCING USING MOLECULAR INVERSION PROBES IDENTIFIES RARE NONCODING VARIANTS ASSOCIATED WITH HIGH HDL CHOLESTEROL LEVELS.....	212
Introduction	212
Materials and Methods	215
Results.....	221
Candidate regions for targeted sequencing	221
Variants identified by MIP sequencing.....	222
Association of single variants from targeted sequencing with extremely high HDL-C	223
Replication of HDL-C associations from GWAS through MIP sequencing	224
Gene-burden associations across target regions with extremely high HDL-C	224
Discussion	225
CHAPTER 6 – SUMMARY AND FUTURE DIRECTIONS.....	252
Summary	252
Future Directions	255
Role of <i>SCARB1</i> in HDL metabolism and CHD risk in humans.....	255
Biology of <i>GALNT2</i> and lipoprotein metabolism	258

Therapeutically relevant mechanisms of <i>APOC3</i> loss-of-function ...	262
Resequencing the noncoding genome for rare causal variation.....	266
BIBLIOGRAPHY	270

LIST OF TABLES

Table 2.1 †: Association of <i>SCARB1</i> P376L with HDL-C in High vs. Low HDL-C Cohorts.	82
Table 2.2: Association of <i>SCARB1</i> P376L with plasma lipid traits in Global Lipids Genetics Consortium Exome Array genotyping.	83
Table 2.3: Characteristics of <i>SCARB1</i> P376L carriers and controls recruited for deep phenotyping.	84
Table 2.4: Meta-analysis of association of <i>SCARB1</i> P376L variant with CHD.	86
Table 3.1: Glycopeptides identified from ANGPTL3, ApoC-III, and PLTP from Human Controls and GALNT2 p.Gln289* Homozygote.	143
Table 3.2: Differential Glycoproteomes of <i>Galnt2</i> KO vs. WT Rats.	145
Table 3.3: Differential Glycoproteomes of <i>Galnt2</i> KO vs. WT Mice.	148
Table 4.1: Discovery and carrier frequency of <i>APOC3</i> LoF variants through exome chip genotyping.	203
Table 4.2: Predicted impact of <i>APOC3</i> LoF variants on <i>APOC3</i> mRNA splicing. ...	205
Table 4.3: Plasma lipids in murine models of WT vs. A43T <i>APOC3</i> expression.	210
Table 5.1: Indexed adapter oligo sequences.	234
Table 5.2: Sequencing primers.	238
Table 5.3: Variant Filtering Criteria.	239
Table 5.4: Characteristics of participants for MIP targeted sequencing.	240
Table 5.5: Variants identified by MIP sequencing of high HDL-C participants.	241
Table 5.6: Single variant associations with high HDL-C.	243

Table 5.7: Replication of GWAS-significant HDL-C associations with MIP sequencing.	248
Table 5.8: Gene-burden associations of identified variants with high HDL-C	250

† All table legends include a brief description of contributions of the dissertation author (S.A.K.) to the data and analyses presented in the respective table. Contributions appear in italic font at the end of the legends.

LIST OF ILLUSTRATIONS

Figure 1.1: Chylomicron Metabolism.....	41
Figure 1.2: VLDL and LDL metabolism.	42
Figure 1.3: HDL metabolism.....	43
Figure 2.1†: SR-BI Protein Sequence Alignment Across Species.	71
Figure 2.2: <i>SCARB1</i> P376L abrogates SR-BI function in transfected COS7 cells..	72
Figure 2.3. Gene expression in control and <i>SCARB1</i> P376L iPSC-derived hepatocyte- like cells (HLCs).....	74
Figure 2.4: <i>SCARB1</i> P376L is a null variant <i>in vitro</i> and <i>in vivo</i>	75
Figure 2.5: Hepatic <i>SCARB1</i> expression and impact on selective cholesterol uptake from HDL in mice.	77
Figure 2.6: HDL composition and functionality in <i>SCARB1</i> P376L homozygote, heterozygous carriers, and controls.	78
Figure 2.7: Platelet aggregation and cholesterol content.	79
Figure 2.8: Impact of <i>SCARB1</i> P376L on adrenal and gonadic steroidogenesis....	80
Figure 2.9: Carotid intima media thickness (cIMT) in <i>SCARB1</i> P376L carriers vs. controls.....	81
Figure 3.1: Identification and lipid phenotype of human homozygotes with <i>GALNT2</i> loss- of-function coding variants.	122
Figure 3.2: <i>GALNT2</i> expression in mammalian models of <i>GALNT2</i> deficiency.....	124
Figure 3.3: Plasma HDL cholesterol in mammalian models of <i>GALNT2</i> loss-of-function..	126

Figure 3.4: Non-HDL cholesterol in <i>GALNT2</i> loss-of-function models.	128
Figure 3.5: HDL and TG clearance in <i>Galnt2</i> KO mice.	130
Figure 3.6: Differential glycoproteomics of tissues from <i>Galnt2</i> WT vs. KO rodents identifies PLTP as a candidate target of GalNAc-T2.	132
Figure 3.7: Overview of differential O-glycoproteomes from <i>GALNT2</i> deficient humans and rodents.	134
Figure 3.8: Reconstitution of hepatic <i>Galnt2</i> expression in <i>Galnt2</i> KO mice restores HDL-C and PLTP activity.	136
Figure 3.9: Common variants in <i>GALNT2</i> associated with HDL-C confer allelic imbalance of <i>GALNT2</i> expression in human liver.	138
Figure 3.10: Investigation of AAV-mediated <i>Galnt2</i> overexpression and knockdown.	140
Figure 3.11: Relationship of the previously reported D314A variant with HDL-C.	142
Figure 4.1: Plasma lipids in carriers of all <i>APOC3</i> LoF variants vs. noncarriers in CGI BioBank and HHDL cohorts.	182
Figure 4.2: Plasma ApoC-III levels in carriers of <i>APOC3</i> LoF variants vs. controls from CGI BioBank and HHDL cohorts.	184
Figure 4.3: Human <i>APOC3</i> A43T carriers exhibit reduced ApoC-III levels due to lower mutant protein.	186
Figure 4.4: Additional lipids measurements in carriers of the <i>APOC3</i> A43T variant vs. noncarrier controls in the CGI BioBank and HHDL cohorts.	188
Figure 4.5: Dose-optimization studies of WT <i>APOC3</i> AAV in mice.	190
Figure 4.6: Mice expressing <i>APOC3</i> A43T have reduced TRLs and circulating ApoC-III <i>in vivo</i>	191

Figure 4.7: VLDL TG and ApoB secretion in mice expressing WT vs. A43T <i>APOC3</i>	193
Figure 4.8: Hepatic AAV vector levels in mice expressing WT vs. A43T <i>APOC3</i> AAVs.	194
Figure 4.9: A43T promotes circulating ApoC-III clearance and augments renal uptake by perturbing binding to lipoproteins.	195
Figure 4.10: Urinary clearance of ¹²⁵ I-labeled WT vs. A43T ApoC-III in mice.	197
Figure 4.11: Binding of WT vs. A43T ApoC-III to lipid and human lipoprotein fractions.	199
Figure 4.12: Impact of A43T on LPL inhibition by ApoC-III.	201
Figure 5.1: Diagram of MIP target capture and sequencing reaction.	231
Figure 5.2: Candidate gene regions for MIP targeted sequencing.	232
Figure 5.3: Concordance of variants identified from MIP sequencing with Exome Chip genotyping.	233
Figure 6.1: Proposed mechanism of disruption of folding of SR-BI due to <i>SCARB1</i> P376L variant.	267
Figure 6.2: Possible role of C-terminal O-glycosylation of PLTP in mediating phospholipid binding capacity and transfer to HDLs.	268
Figure 6.3: Impact of <i>APOC3</i> A43T variant on lipid binding affinity of ApoC-III.	269

† Illustration legends for all figures in Chapters 2 through 5 include a brief description of contributions of the dissertation author (S.A.K.) to the data and analyses presented in the respective figure. Where relevant, contributions of collaborating colleagues are also indicated. Contributions appear in italic font at the end of the legends.

CHAPTER 1 - INTRODUCTION

Heart Disease – Epidemiology and Pathophysiology

Coronary heart disease (CHD), also called coronary artery disease (CAD) and ischemic heart disease (IHD), is a chronic, multifactorial disorder involving the pathological narrowing, stiffening, and remodeling of the coronary arteries, the blood vessels supplying the heart muscle or myocardium with oxygen [1,2]. CHD is the predominant form of cardiovascular diseases that together are the leading cause of mortality worldwide [3,4]. In the United States in 2013, the American Heart Association estimated that cardiovascular diseases accounted for 402,851 male deaths and 398,086 female deaths, or approximately one in four deaths overall [5]. More than 50% of these deaths were due to CHD. The prevalence of cardiovascular diseases in the United States is also profound, with an estimated 85.6 million individuals having one or more cardiovascular diseases, or 35% of the population. While the incidence and mortality from cardiovascular diseases has been decreasing in developed countries since the mid-1990s, disease burden has sustained in developing countries, and globally, cardiovascular diseases accounted for more than 17.3 million deaths in 2013 (approximately 31% of all deaths worldwide), a statistic that is projected to increase to more than 23.6 million deaths by 2030 [5,6]. In addition to significant mortality and morbidity, cardiovascular diseases impose a significant direct and indirect cost worldwide, with an estimated total cost of disease of 863 billion dollars in 2010 [5]. Thus, cardiovascular diseases as a whole and CHD in particular constitute a pressing burden of morbidity, mortality and financial strain across societies.

CHD is caused by the pathological narrowing of the lumen of the coronary arteries due to the thickening, stiffening, and remodeling of the arterial wall, a process

called atherosclerosis [1,2,7]. Atherosclerosis occurs due to the deposition of fat in the intima (innermost layer of artery adjacent a layer of endothelial cells in direct contact with blood flow) of arteries. This fat includes oxidized phospholipids, cholesterol and triglycerides carried on low-density lipoproteins (LDLs), one class of circulating particles in the blood comprised of several protein and lipid components. Both circulating free radicals, and enzymes in the intima such as lipoprotein phospholipase A2, contribute to the oxidization of phospholipids on the surface of LDLs, rendering them prone to aggregation. This in turn enhances their susceptibility to phagocytosis by macrophages in the arterial wall [1,8]. Accumulation of modified lipoproteins in macrophages causes a profound phenotypic change in the cells [9,10], which are often termed foam cells due to a change their appearance after internalization of fatty deposits. Foam cell formation after uptake of oxidized proatherogenic lipoproteins is the initial step in lesion formation in the arterial wall and the first step in atherogenesis [7,10,11].

Augmenting this lesion (also termed plaque) formation is the subsequent proliferation of vascular smooth muscle cells, which secrete collagen and other extracellular matrix components, and T cells, which induce inflammation in the vessel [2]. The foam cells eventually die and release cellular lipid contents in the intima. This process perpetuates further inflammation through the release of inflammatory cytokines from these dying cells [12]. Concurrently, smooth muscle cells block the proliferating plaques from access to flow from the luminal surface of the vessels, and thus promote formation and proliferation of a “necrotic core” of inflammatory cells, cell debris, and fat-laden lesions. As a plaque develops in the arterial wall, the lumen of the coronary artery diminishes in volume. This may reduce blood flow and cause inadequate oxygenation of

the myocardium supplied by such an artery during times of increased demand such as exercise; this manifests in chest pain and is known as stable angina [1,8,12].

The initially walled-off atherosclerotic plaque may also rupture and the endothelium lining the lumen may erode thus leading to release of the atherosclerotic plaque into the coronary lumen [2]. Partial obstruction of blood flow through the adjacent arterial lumen may also diminish oxygenation of the myocardium and may cause chest pain at rest as well; this is termed unstable angina. Obstruction of blood flow due to plaque rupture that leads to myocardial ischemia initially may then result in death of the myocardial cells, a process termed myocardial infarction or MI [13].

Depending on the site and extent of vessel obstruction, a MI may have profound consequences for the function of the heart, which can lead to a patient's death from several mechanisms [2]. Acutely, a MI may severely reduce or completely inactivate the heart muscle from maintaining forward blood flow due to death of the contractile myocardial cells, which results in systemic loss of perfusion of tissues. Additionally, myocardial scarring of dead tissue can cause abnormal electrical impulse conduction throughout the heart, which can cause the sudden disorganized movement of the heart muscle known as ventricular fibrillation; this also leads to inability of the heart to pump blood and perfuse the body and is known as sudden cardiac death. Both of these acute consequences are predominant causes of mortality in the setting of an MI. Additional consequences include long-term impairment in cardiac contractility and subsequent compensatory remodeling of the heart to attempt to maintain forward blood flow; this disease is known as ischemic cardiomyopathy and is the prevailing cause of heart failure, another major cardiovascular disorder with profound morbidity and mortality worldwide. Heart valve dysfunction due to myocardial contractile and anatomical

changes after an MI, as well as risk of future arrhythmias are also long-term consequences of an MI from CHD [2].

The etiology of CHD is complex with many recognized risk factors [14,15]. These include high blood levels of atherogenic lipoproteins, particularly LDL-cholesterol (LDL-C) and triglyceride-rich lipoproteins (TRLs) whose importance is becoming increasingly apparent, low levels of high-density lipoprotein cholesterol (HDL-C), hypertension, obesity, diabetes mellitus and insulin resistance, smoking, lack of physical activity and consumption of unhealthy diet, the latter two of which may exacerbate several of the other above mentioned contributors. Among these risk factors, high LDL-C levels are a predominant cause of CHD with substantial evidence from epidemiology and human genetics, an established pathophysiology, and significant pharmacological success in reducing LDL-C levels to decrease incidence, morbidity and mortality from disease. Yet despite the effectiveness of targeting LDL-C to lower CHD risk over the last four decades, disease burden remains high and not all patients can benefit maximally from LDL-C lowering alone. In addition, some patients develop CHD or remain susceptible to complications from pre-existing CHD despite optimally low LDL-C concentrations; this signals the need for additional orthogonal therapeutic targets for risk and mortality reduction from disease. In this regard, efforts to target the other plasma lipid classes associated with cardiovascular disease risk (discussed in the next section) have been at the forefront of research endeavors in this field, and human genetics has played a key role in helping to refine our understanding of how these traits contribute to disease risk.

Blood Lipids and CHD Risk – the Evidence for Causality

Plasma lipoproteins are circulating complexes of proteins and lipids (cholesterol, triglycerides, phospholipids) that mediate the delivery of their lipid cargo as nutrients and substrates for biosynthetic and signaling processes throughout the body [16]. Lipoproteins have traditionally been categorized by their density, lipid and protein content, and putative roles in physiology and disease. Broadly speaking, lipoproteins can be grouped into two categories: those containing apolipoprotein B (ApoB) as a critical constitutional protein vs. those that contain apolipoprotein A-I as their key structural protein. Using this categorization, the former group includes the LDLs, the very low-density lipoproteins (VLDLs) and the chylomicrons as well as remnant particles of these classes. The latter group containing ApoA-I includes the HDLs. The blood concentrations of each of these lipoprotein classes have been studied for more than seven decades and have been related to risk or protection from CHD and other vascular diseases through large epidemiological studies [17]. Because these traits are easily measurable and have a polygenic etiology of inheritance, they have also served as invaluable tools for understanding the architecture of complex trait genetics [18]. The potential role and causality of each of these traits to risk of CHD has been a central topic of translational medical discovery of the early twenty-first century and certainly not a field without controversy, as will be discussed below.

Low-density lipoproteins

LDLs are the lipoproteins with the most profound evidence for a causal relationship to CHD. The defining protein component of LDL, ApoB, serves as a scaffold for cholesterol, cholesterol esters (CEs), triglycerides (TGs) and phospholipids [19,20].

LDL is generated from the lipolytic processing and remodeling of VLDLs in the circulation by the action of extracellular lipases: enzymes on the capillary luminal surface of tissues that hydrolyze TGs and phospholipids of lipoproteins to provide nutrient sources for underlying tissues and thereby alter the structure and composition of the lipoproteins. Circulating LDLs deliver CEs to peripheral tissues and are eventually cleared from the circulation by the LDL receptor in the liver. Internalization of LDL by the LDL receptor occurs through receptor-mediated endocytosis and yields catabolic products that ultimately may be recycled for storage or re-secretion. Importantly, LDLs may also traverse the arterial wall, where oxidation of their surface phospholipids promotes macrophage activation, LDL uptake, and foam cell formation in the intima, as mentioned above [20].

Evidence for a role of LDL-C in promoting atherogenesis and CHD has developed since the 1910s from the discovery of cholesterol in arterial lesions from humans and cholesterol-fed rabbits [21-23]. From those early studies to the subsequent elucidation of the cholesterol and lipoprotein metabolic pathways, and finally to the therapeutics to lower LDL-C levels that are now the mainstay of treatment and prevention for cardiovascular diseases, human genetics has proven pivotal. The identification of families with an autosomal dominant inheritance pattern of extremely high cholesterol levels and MIs at an early age fueled a subsequent career of study by Drs. Michael Brown and Joseph Goldstein of the University of Texas Southwestern Medical Center that uncovered the *LDLR* gene and receptor-mediated uptake pathway of LDL clearance by the liver and showed that these patients who suffered from Familial Hypercholesterolemia (FH) inherited loss-of-function (LOF) mutations in the LDL receptor gene *LDLR* [20,24,25]. The understanding of this Mendelian disorder of LDL

metabolism in the 1970s complemented earlier evidence showing LDL-C as a major risk factor in MI patients [26,27], which together solidified the evidence supporting the notion that LDL-C lowering could reduce CHD and MI risk. These findings and the serendipitous discovery of statins [28], inhibitors of the rate-limiting cholesterol biosynthetic enzyme 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase. Statins function by reducing hepatic cholesterol synthesis, which induces a compensatory upregulation of LDL receptor-mediated clearance from the circulation. Thus, the discovery of the first major efficacious drug and pathway for lowering LDL-C and disease risk was driven by human genetics, and was motivated by a desire to elucidate the cause of FH.

Subsequent studies of additional Mendelian disorders of LDL metabolism that confer increased CHD risk have affirmed the importance of the hepatic LDL clearance pathway as a critical phenomenon and target for lowering LDL-C and disease. These efforts have uncovered mutations in *APOB* [29], the gene encoding the key LDL structural protein ApoB that interacts with the LDL receptor for uptake; *ARH* [29,30], a gene encoding an adapter for the LDL receptor that when mutated results in an autosomal recessive hypercholesterolemia; and *PCSK9* [20,31], a gene encoding a secreted protein that reduces cell-surface LDL receptor levels by promoting lysosomal turnover, and thus inhibits LDL uptake via this pathway. This last example of *PCSK9* is particularly notable because it offers an impressive demonstration of the translation of genetics discovery to the development of novel medicines for reducing CHD risk in less than a decade. Through positional cloning in a large family and additional sequencing of probands with autosomal dominant hypercholesterolemia absent of *APOB* or *LDLR* mutations, two variants in the *PCSK9* gene were identified that segregated with

extremely high LDL-C levels [32]. Subsequent studies in humans confirmed that these variants were likely gain-of-function mutations that increased PCSK9 function [33-35], and additional work in mice demonstrated that indeed PCSK9 overexpression increased LDL-C levels in vivo [36,37]. Following this work, sequencing of human subjects with extremely low LDL-C levels in the Dallas Heart Study identified *PCSK9* mutations that were demonstrated to confer LOF [38]. Caucasians heterozygous for a low-frequency missense variant [3% minor allele frequency (MAF)] demonstrated a 15% reduction in LDL-C levels and 46% lower incidence of CHD adverse events, and African Americans heterozygous for either of two nonsense variants demonstrated even more profound LDL-C and disease risk reductions [38,39]. This and other studies [40,41] not only further established the causal and direct relationship of LDL-C levels to CHD risk through an example of genetic gain- and LOF in *PCSK9* conferring opposing LDL-C phenotypes and disease risk, but they also paved the way for the development of therapies targeting PCSK9 [42-50]. In 2015, two PCSK9 inhibitory monoclonal antibodies were approved for the treatment of hypercholesterolemia in addition to statins in patients on maximally tolerated statin therapy who require additional LDL-C lowering, as well as for heterozygous and homozygous FH patients [51]. This marked the success of a bench-to-bedside journey that started only 12 years earlier with the initial report of PCSK9 gain-of-function mutations with hypercholesterolemia.

PCSK9 inhibitors, statins, and other LDL-C lowering therapies currently approved largely function through enhancing hepatic LDL clearance from the blood. Thus, human genetics has not only helped unveil a key pathway of LDL regulation but also has conclusively demonstrated their causality to the development of atherosclerosis and has helped facilitate targeted therapies that have been efficacious for disease reduction. As

will be discussed below, additional investigations of both low-frequency and common genetic variants have also supported a causal role of smaller heritable variations in LDL-C with direct effects on CHD risk. While LDL-C lowering through current therapies has been remarkably efficacious, it has not eliminated risk of CHD in all patients, and many patients are still at risk of MI even with optimal LDL-C lowering through one or more of these medications. Additionally, many patients are intolerant to their adverse effects (e.g. statins) or not able to access them due to still limited use or high cost (e.g. PCSK9 inhibitors). This has signaled the need for additional therapies targeting orthogonal pathways to those of LDL-C focused drugs. In this vein, HDL-C and TRLs have remained candidates for possible therapeutic intervention.

High-density lipoproteins

HDLs are formed through the secretion of ApoA-I from the liver and small intestinal enterocytes [52,53]. ApoA-I in the circulation quickly acquires phospholipids and free cholesterol from cells via ATP-binding cassette transporter proteins on the cell surface (particularly ABCA1 and ABCG1), which efflux lipids to lipid-poor ApoA-I acceptors. This allows the nascent particles to mature into HDLs. Esterification of the cholesterol of these HDLs with fatty acyl chains by the lecithin:cholesterol acyltransferase (LCAT) enzyme to CEs further matures and stabilizes these particles. HDLs traverse the blood and may acquire free cholesterol from cells via ABC-transporter mediated efflux. Subsequent LCAT activity facilitates this flux by converting the free cholesterol to CE and thereby promoting a concentration gradient for the mobility of the free cholesterol. CEs on HDLs exchange for TGs from ApoB-containing lipoproteins via facilitation by cholesteryl ester transfer protein (CETP). Additionally, phospholipids on

HDLs may redistribute to ApoB-containing lipoproteins through the actions of phospholipid transfer protein (PLTP) in the circulation [19,52,53]. HDLs also have a diverse and dynamic proteome that changes with their metabolism [54]. Like LDLs, HDLs are susceptible to lipolysis by extracellular lipases, which promote their catabolism largely through the kidneys. HDL-C is also selectively taken up by the liver and other tissues involved in steroid hormone biosynthesis via the scavenger receptor class BI (SR-BI) through a process distinct from receptor-mediated endocytosis that retains the CE-reduced HDLs in the circulation where they can re-mature through the subsequent re-accumulation of lipids by the above described processes.

In contrast to case for LDL, epidemiological studies have strongly supported an inverse relationship between HDL-C levels and CHD risk. The earliest studies of separation of lipoprotein classes from plasma of MI patients demonstrated that a majority of them harbored low HDL-C levels in addition to their elevated LDL [26]. The Framingham Heart Study and other large community-based prospective studies affirmed this relationship and showed that it was independent of the relationship of LDL-C or TGs to vascular risk [17,55,56]. Mechanistically, HDLs are thought to promote atheroprotection through several putative mechanisms. Of these, perhaps the most attractive is the reverse cholesterol transport (RCT) hypothesis [57,58]: the notion that HDL can accept cholesterol from arterial plaque macrophages via ABC transporter-mediated efflux and deliver this cholesterol after esterification to the liver via SR-BI-mediated selective cholesterol uptake. The CE in the liver is then recycled for bile acid synthesis and later excreted via the bile, thus providing a unidirectional sink for cholesterol removal from the plaques and ultimately the body. Animal model investigations studying the disruption of this pathway of macrophage-to-feces RCT have

strongly supported its potential role in reducing atherosclerosis in vivo. Additional potentially anti-atherogenic functions of HDL include its ability to reduce inflammation, oxidative stress, and promote endothelial cell function [59].

Despite the epidemiology and animal physiology studies that link higher HDL-C to lower CHD risk, neither human genetics, nor pharmacological interventions that specifically target HDL-C have supported a causal role thus far. Mendelian disorders of HDL metabolism have offered unclear evidence [60]. *APOA1* mutations that confer ApoA-I deficiency or structural alterations and low-to-absent HDL-C levels are extremely rare, but the preponderance of evidence has not supported a definitive increase in CHD risk, especially for the structural mutations. Likewise, both homozygous LOF of *ABCA1* (Tangier Disease) and *LCAT* (Fish Eye Disease and Familial LCAT Deficiency) cause profound reductions or complete loss of circulating HDL-C with unsubstantiated evidence of increased atherosclerotic risk, in large part because of the rarity of patients and possibly also because of variable effects for the other lipoproteins in the few identified patients. Consistent with these findings, additional methods to assess causality of more common genetic variants that impact HDL-C to CHD risk (Mendelian randomization; discussed below) have also failed to support a direct link between HDL-C itself and CHD or MI risk [19].

In addition to weak support from these genetic cases of extreme HDL-C deficiency, efforts to raise HDL-C specifically have not translated to reduced CHD risk. Multiple trials of CETP inhibitors that robustly increased HDL-C have met with either no benefit or adverse effects on cardiovascular mortality (though the latter finding was later ascribed to off-target effects) [19,61-64]. Similarly, two large trials of nicotinic acid (niacin), a potential elevator of HDL-C used for decades, have shown that when added

to statins, it failed to show any additional benefit in lowering adverse cardiovascular events in at-risk patients [65,66].

While the exact reasons for why each of these medications failed to reduce CHD risk or adverse events despite potent HDL-C elevation will be difficult to determine, one prevailing notion is that these therapies raised HDL-C levels and did not necessarily improve the functionality of the HDL particles in mediating RCT or other atheroprotective roles. In the last five years, there has been increasing evidence from both cross-sectional and prospective studies that HDL cholesterol efflux capacity, one metric of its functionality in potentially mediating delivery of CE to the liver through the RCT pathway, is a more reliable measure of CHD risk than the static measure of HDL-C alone [19,53]. Indeed, the cholesterol efflux capacity of HDL has been demonstrated to be a significant predictor of incident cardiovascular events even after adjusting for HDL-C concentrations. Pharmacological studies in limited numbers of subjects involving the infusion of recombinant HDL that promotes cholesterol efflux showed preliminary efficacy in reducing atherosclerotic plaques [19,53,67-70]. Larger studies evaluating these and other developing therapies targeting HDL's ability to promote RCT will be more suited to test the hypothesis that HDL functionality is a causal contributor to CHD risk. Likewise, human genetics efforts to find variants that specifically alter the RCT pathway at different nodes will be able to assess the lifelong and heritable contribution of this phenomenon to disease risk and potentially rectify the current controversy surrounding the direct relevance of HDL to vascular disease beyond serving as a biomarker.

Triglyceride-rich lipoproteins

Independently of LDL-C, plasma TRL levels have recently emerged as an important and independent risk factor for development of CHD [71]. TRLs are produced from two sites in the body: the liver, which secretes VLDLs, and the small intestinal enterocytes, which secrete TG-rich chylomicrons into the lymph in the setting of dietary consumption. In both cases, assembly of TRLs involves the lipidation of ApoB through the secretory pathway in part by the actions of microsomal triglyceride transfer protein (MTP) in the respective cell type to yield a buoyant, lipid-laden particle that delivers fatty acid nutrients to peripheral tissues. Other than the site of origin, chylomicrons and VLDL differ also in the nature of their ApoB (a shorter form of the protein, ApoB48, is incorporated into chylomicrons due to RNA editing that occurs in the enterocytes but not hepatocytes in certain organisms including humans), their size and TG content (chylomicrons typically are larger, less dense and more enriched in TGs), and the exact milieu of lipids and proteins on the TRL surfaces. Circulating TRLs are modified by extracellular lipases that hydrolyze the TGs to fatty acids that are taken up by nearby tissues [19,71,72]. In this regard, lipoprotein lipase (LPL) is a critical regulator of TRL metabolism with a complex regulation, as detailed below. The actions of LPL and other lipases remodel the TRL lipid and protein content to yield smaller, denser particles such as chylomicron remnants from chylomicrons and intermediate density lipoproteins (IDLs) and LDLs from VLDLs. Remnants are cleared from the circulation by hepatic receptors such as the LDL receptor and related family members via receptor-mediated endocytosis utilizing apolipoprotein E (ApoE) and other surface proteins as ligands [73].

While recognized for decades as a correlative biomarker of risk [74], recent prospective studies have demonstrated that elevated TGs are independent predictors of CAD risk. Furthermore, TGs are strongly associated with incident CHD even in statin-treated

patients with low LDL-C levels. The ACCORD study found that among dyslipidemic patients with elevated TG and low HDL-C, the addition of TG-lowering fenofibrate therapy reduced incidence of cardiovascular events [75]. Moreover, a large meta-analysis in 2012 by Boekholdt and colleagues demonstrated that TRLs beyond LDL-C, including non-HDL-C, were the strongest predictors of vascular risk for patients already maximally treated on statins [76]. Mechanistically, TRLs are cholesterol-rich particles that can cause plaque accumulation similarly to LDL-C but also directly damage the vascular endothelium and augment inflammation, promote foam cell formation, and cause lipotoxicity to vascular smooth muscle cells [71].

Human genetics has also causally implicated TRLs and specifically the LPL pathway with cardiovascular risk. Several studies of a gain-of-function variant, *LPL* S447X, have demonstrated an association with lower TG levels and reduced risk of CHD [77-83]. Using newer methods to interrogate common and low-frequency variants such as genome-wide association studies (GWAS) and Mendelian randomization (both discussed below), single-nucleotide polymorphisms (SNPs) at the *LPL* gene locus and loci of *LPL* pathway regulators have also been implicated in regulating fasting TG levels and CHD [84-88]. Additional approaches such as population based exome or candidate gene sequencing in participants at the phenotypic extremes of lipid traits (also described below) have uncovered rare variants in *LPL* pathway genes associated with TGs and CHD or MI risk that have collectively posited that activation of the *LPL* pathway is atheroprotective in a manner that may be independent and synergistic with LDL-C lowering [87,89-92]. These studies have identified a few key regulators, namely ApoC-III, ApoA-V, and the angiopoietin-like proteins ANGPTL3 and ANGPTL4 as particularly critical to influencing disease risk through impacting TRL levels.

ApoC-III

ApoC-III may act at multiple nodes influencing TRL homeostasis. A small (8.8 kDa) apolipoprotein, ApoC-III is expressed primarily in the liver and circulates on HDLs and TRLs [93,94]. ApoC-III undergoes O-glycosylation and exists in nonglycosylated (ApoC-III-0), monosialylated (ApoC-III-1) or disialylated (ApoC-III-2) forms, though this modification does not appear to impact plasma TG or ApoC-III binding to lipoproteins [95-97]. Like ApoE, ApoC-III exchanges between lipoproteins. Several studies have demonstrated that ApoC-III may inhibit LPL activity [98]. Further insight gained from studies of human *APOC3* and mouse *Apoc3* transgenic mice and *Apoc3* knockout (*Apoc3* KO) mice suggested that ApoC-III may compete for lipoprotein binding with ApoE and delay the catabolism of TRL remnants by the liver [99,100]. Additionally, ApoC-III may promote VLDL secretion [101]. Preclinical studies utilizing an antisense oligonucleotide (ASO) to silence *Apoc3* expression in the liver have shown efficacy in lowering plasma TG in rodents and nonhuman primates as well as in human clinical studies [102-104].

In the past decade, human genetics studies have identified *APOC3* variants that provide insight into its role in TRL metabolism and CHD. In 2008, genotyping and sequencing in an Amish population identified the R19X nonsense mutation in *APOC3*, which was associated with plasma TG and reduced coronary artery calcification (CAC), a surrogate marker of atherosclerosis [105]. More recently, two large exome sequencing studies identified this and three other rare coding variants in *APOC3* associated with reduced TG levels and collectively associated with a highly significant 40% reduction in risk of CHD [106] [91]. These data strongly support that genetic deficiency of ApoC-III reduces not only TRL levels but also risk of CHD. While many variants in *APOC3* have been identified, most have not been mechanistically investigated to understand how they impact TG metabolism.

ApoA-V

Like ApoC-III, ApoA-V exchanges between HDLs and TRLs and is primarily secreted from the liver. A 39 kDa protein, ApoA-V has a very low concentration in human plasma (~150 ng/ml) [107]. Despite this, ApoA-V plays a crucial role in TG metabolism. *Apoa5* KO mice demonstrate profound hypertriglyceridemia (hyperTG) while human *APOA5* transgenic mice have significantly lower plasma TG than controls [108]. ApoA-V has been shown to enhance LPL activity and may do so by facilitating proximity between TRLs and LPL, in part through interaction with glycosylphosphatidylinositol-anchored HDL binding protein 1 (*GPIHBP1*), a required chaperone for LPL on capillary luminal surface. A wealth of human genetic data from common, low-frequency, and rare (MAF < 1%) variants has suggested that LOF of ApoA-V augments risk of CAD and MI [73,109]. Most notably in this regard was a recent exome sequencing study in ~4,700 early MI cases and ~5,000 healthy controls that identified several rare coding variants in *APOA5* associated together with increased frequency of MI, a finding that was second only to that for *LDLR* variants in the same cohort [92]. As with the *APOC3* mutations, the mechanisms of action of these *APOA5* variants as they relate to TRL metabolism and CHD risk remain largely unknown.

Angiopoietin-like Proteins 3 and 4

The angiopoietin-like (ANGPTL) proteins ANGPTL3 and ANGPTL4 are additional regulators of LPL activity and TRL metabolism that have generated considerable interest as putative therapeutic targets [110]. Like ApoC-III, ANGPTL3 and ANGPTL4 are thought to inhibit LPL activity, leading to elevated plasma TG levels, though their respective mechanisms conferring LPL inhibition may be distinct [110-115]. Both common and rare variants in *ANGPTL3* are associated with plasma TG levels, and rare *ANGPTL3* LOF

mutations underlie the Mendelian condition familial combined hypolipidemia, characterized by very low plasma TG in addition to other lipid classes [115-118]. While common variants at the *ANGPTL4* locus are associated primarily with HDL-C levels, rare coding mutations are robustly associated with reduced plasma TG [115-117,119]. Prior evidence linking *ANGPTL3* and *ANGPTL4* LOF to CHD risk have been smaller and/or inconsistent [120-123], but two recent large exome sequencing studies have suggested a robust relationship between *ANGPTL4* LOF, lower TGs and decreased CHD [89,90].

While epidemiology and recent human genetics efforts have positioned TRLs as possibly the next 'holy grail' for lipid lowering therapies to reduce CHD risk, so far, existing TG-lowering therapies have not been applied appropriately to test the question of whether reducing TRLs specifically in hyperTG patients would yield a clinical benefit for MI risk [124]. Now with the discovery of a key pathway and several putative therapeutic targets, rational therapies may be designed that mimic the beneficial effects of protective variants in the implicated genes. Such efforts will require a deep study of the mechanisms of action of the identified variants for these LPL pathway genes that integrates the findings in humans with experimental models before new treatment strategies can be devised.

In summary, the plasma lipoproteins all exhibit a complex biology with multiple nodes of regulation that have been studied from the vantage points of animal physiology, human genetics, and in some cases pharmacological interventions in humans. Regarding the causality of these traits to vascular disease risk, while the case for LDL-C lowering has and will remain salient, the potential value in modulating HDLs and TRLs has been more tenuous. What is clear for all the lipoproteins is that human genetic discoveries and developing approaches will undoubtedly inform strategies to better determine which facets of these traits are causally related to disease risk and worthy of therapeutic pursuits. As the last decade has

brought myriad new genomic paradigms for the study of rare and common genetic variation for complex traits such as lipids, I will now review the key methodologies as they have been applied to the study of plasma lipids and relevant discoveries that have informed the goals of this dissertation.

Next-generation Genomic Approaches to Studying Plasma Lipids

Complex traits, including continuously distributed traits such as blood pressure, height, body-mass-index, plasma lipids, and others, are subject to immense genetic variation and environmental influence [18,125]. This property has made pinpointing the exact genetic contributors for a given trait challenging. Given that many of these traits, as well as diseases with polygenic inheritance, have significant clinical implications or burden, there remains a major interest in comprehensively identifying all the genes that contribute to these phenotypes, the magnitude of their contribution, and the mechanisms of their effects. Recent human genetics advances focused on genotyping common variants have provided a route to better interrogate the heritability of such traits. Such GWAS determine genotypes at thousands to millions of common single nucleotide polymorphisms (SNPs) across the genome in individuals measured for a trait or disease of interest, and test the association of the allele at each SNP with a quantitative trait or disease status [126-128]. These studies have been applied to numerous complex traits and have uncovered hundreds of novel loci significantly associated with diseases and their risk factors.

While useful as starting points for uncovering new loci influencing a trait, GWAS themselves always require further validation of causal genes and variants. This is because GWAS interrogate many common variants with usually small effect sizes

[129,130]. Thus, the actual contribution of all genetic variation at a given locus to a trait cannot be ascertained from the effect of the associated common variant alone. Moreover, since most common variants that are genotyped on SNP genotyping arrays are noncoding in nature, it is often difficult to assign an associated variant at a locus to a single gene without further experimental follow-up [129,130]. However, when GWAS is coupled with candidate gene sequencing at the extremes of a phenotypic distribution, it can be a powerful tool for establishing a novel gene influencing a trait as well as providing a directionality of the relationship between a gene's expression or function and the trait. A recent example is the identification of variants at or near the *GCKR* gene with plasma triglyceride (TG) levels. Initially reported in the first GWAS for plasma lipids, common non-coding variants at this locus were robustly and consistently associated with plasma TG, as well as with fasting glucose levels, in subsequent follow-up studies and meta-analyses [131,132]. Subsequently, several groups sought to identify functional coding variants in *GCKR* that influenced its role in TG and glucose metabolism [133,134]. One study, which involved sequencing *GCKR* in individuals from the ClinSeq medical resequencing project, identified 19 missense variants, many of which were rare in frequency and deleterious to protein function *in vitro*, and demonstrated association of these variants to TG in that cohort [135]. Similar approaches are underway to establish causality and identify functional variants in novel loci for plasma lipids and other complex traits identified through GWAS.

Genome-wide association studies for plasma lipids

Plasma high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) and triglycerides (TG) have been among the most commonly interrogated complex traits since the advent of GWAS in the mid-2000s. Because of their clinical utility, ease of measurement, and high degree of heritability, they have been commonly studied in multi-ethnic populations. The first large GWAS for lipids were published in 2008 and involved almost 9,000 individuals [131,132,136]. These studies identified both known loci previously shown to regulate lipid metabolism as well as 8 novel loci associated with at least one lipid trait. These included the *SORT1/CELSR2/PSRC1* gene locus and the *NCAN/CILP2/PBX4* gene locus for LDL-C, the *GALNT2* gene locus and *MVK/MMAB* gene locus for HDL-C, and the *TBL2*, *MLXIPL*, *TRIB1*, and *ANGPTL3* gene loci for TG. Several of the lead SNPs for these loci were also associated with additional lipid traits. For example, SNPs near the *GALNT2* gene associated with decreased HDL-C were also associated with increased plasma TG. Shortly after these loci were initially reported, a meta-analysis in almost 20,000 individuals, including association testing in the participants of the Framingham Heart Study and additional European cohorts, replicated the previously reported associations for these loci and identified 11 new loci associated with at least one lipid trait [137]. Since these first reports, follow-up studies and meta-analyses subsequently validated these loci and identified additional ones [86,88,117].

Of note are two large meta-analyses from the Global Lipids Genetics Consortium (GLGC), a collaborative organization led by the groups who reported the initial 2008 study. In 2010, the GLGC reported a GWAS for total cholesterol (TC), HDL-C, LDL-C, and TG in more than 100,000 European individuals [86]. This study yielded 95 genetic

loci associated with at least one plasma lipid trait, 59 of which were novel associations. The results included replications in a smaller sample of East Asian, South Asian and African American cohorts, as well as sex-specific analyses and testing of the lipid-associated SNPs with incidence of coronary artery disease (CAD).

A second study from the consortium, published in late 2013, expanded this effort to include more than 188,000 individuals and identified an additional 62 novel loci, 30 of which had never been previously connected to lipid metabolism [117]. As part of this effort, the investigators selected 65 loci to interrogate on a cardiovascular trait- and metabolic trait-centric array (Metabochip) in a fine-mapping effort for the common variants [117]. Additionally, to assess the potential pleiotropic effects of the lipid-associated SNPs on other cardiometabolic traits, they tested the association of these SNPs with fasting glucose, systolic and diastolic blood pressure, BMI, waist-to-hip ratio, and incidence of type 2 diabetes and CAD in subsets of the 188,000 subjects. They found that several lipid-associated SNPs were also associated with one or more of these traits as well. In total, the authors note 240 genes within 1 kilobase (kb) of the lead SNPs for the 62 novel loci, highlighting the challenging task of identifying the causal gene(s) at these loci.

Upstream issues—from locus to causal variant

Despite the wealth of novel loci for complex traits, few of the causal functional variants that biologically underlie the genetic associations have been conclusively identified. This is in part due to the limitation of SNPs that can be interrogated on GWAS genotyping arrays and the need for indirect imputation of SNPs in linkage disequilibrium with the genotyped SNPs [130,138,139]. While this approach often allows identification

of a causal 'region' containing SNPs associated with the trait, it is usually insufficient to identify which SNP at a given region drives the association and may underestimate the extent of variation at a locus contributing to a trait. For common noncoding variants that are within (or in LD with SNPs within) defined promoter or known regulatory regions of nearby genes, one may assess the effects of variants in these known regulatory regions on gene expression through gene reporter assays, binding affinity for specific transcription factors, and related functional approaches. Such strategies have been successfully applied to a limited number of lipid-associated SNPs, such as those at the LDL cholesterol-associated chromosome 1p13 locus (containing *SORT1*) (discussed below) [140].

Another approach to refining the subset of potentially functional SNPs at a locus is to test the association of a denser region of SNPs from a trait-associated LD block in ethnically diverse populations from that of the initial association [141,142]. Because the sizes and compositions of SNP haplotype blocks may be ethnicity-specific, genotyping SNPs corresponding to a haplotype across diverse ethnicities may filter out nonfunctional SNPs part of the haplotype of the originally tested population. Most of the initial GWASs for lipids have been done using populations of European ancestry. Thus, testing the association of lipid-associated SNPs in non-European populations may allow the filtering of some SNPs at novel lipid loci. Using this approach, Deo et al. studied approx. 4,400 subjects of the Jackson Heart Study and identified African American-specific SNPs at the *LPL* locus that masked the protective effect of the gain-of-function coding variant, S447X, in LD with the noncoding variant [143]. This approach is limited by the need to amass sufficient numbers of subjects in non-European cohorts to compare to original European cohorts so that power does not confound interpretation of

multiethnic association comparisons. This concern is even more important when considering populations that have undergone extensive migration and diversification such as those of African ancestry.

The Encyclopedia of DNA Elements (ENCODE) project offers new potential to comprehensively focus the search for causal variants at the many novel loci[144-146]. This collaborative project used chromatin immunoprecipitation-sequencing (ChIP-Seq) for many transcription factors across cell lines spanning several tissue types to annotate DNA regulatory elements functionally important to transcription at a genome-wide level. Overlaying trait-associated SNPs with important transcriptional regulatory regions informed from the ENCODE project offers one systematic means to localize the subset of potentially causal variants at each locus. An analogous method of prioritizing SNPs is to overlay them with regions with chromatin marks or regions of DNase I hypersensitivity, suggesting open chromatin and active transcription. Maurano et al. utilized the latter approach to catalogue the distribution of more than 5,600 noncoding GWAS-significant SNPs for various traits in terms of their presence in regions of active transcription (DNase I hypersensitivity) [147]. This work demonstrated a preponderance of trait- or disease-associated SNPs in regions of open chromatin that were developmentally- or tissue type-specific in their regulatory nature. An analogous study demonstrated the cell-type specificity of chromatin marks and showed that GWAS-associated SNPs were consistently proximal to such marks [148,149]. They also showed that marks such as histone 3 lysine 4 trimethylation (H3K4me3) were highly cell-type specific and overlapped with 37 GWAS-associated SNPs for LDL-C in liver. Such integrative studies utilizing the breadth of genome-wide regulatory DNA annotation data currently available provide possible paradigms for further honing in on functional SNPs.

Downstream issues—from variants to causal genes and biological mechanisms

An equally if not more important problem in the post-GWAS era for traits such as plasma lipids is the correct identification of the causal gene at the locus harboring trait-associated SNPs and the elucidation of the biological mechanisms involved. A few properties of the trait-associated loci assist in identifying causal genes: lack of multiple genes at a given locus, proximity within or very near a gene (as opposed to within regulatory elements many thousands of kb up or downstream of genes), presence of an expression quantitative trait locus (eQTL) for a SNP with expression of a particular gene at the locus (cis-acting eQTL), and biological plausibility of a particular gene at the locus regulating the associated trait. However, these properties are rarely met and are not in any case definitive proof of the causal gene, and thus the identification of the causal gene influencing a trait can be especially difficult to ascertain for novel loci.

A major focus in identifying causal genes at novel lipid loci has been to better characterize SNPs at these loci for potential eQTLs in tissues of relevance to the trait. For lipid traits, these studies have largely focused on the liver as the tissue underlying allele-specific expression of causal genes, because of the predominant role that hepatocytes play in regulating lipoprotein metabolism. Additional tissues of potential interest to interrogate for lipid traits include enterocytes, adipocytes, skeletal myocytes, as well as endothelial cells and macrophages, which mediate the interplay of lipoproteins and their role in atherosclerosis. The results of the lipid GWAS suggested that of the 157 loci associated with lipids, 47 have significant eQTLs for at least one gene proximal to the associated SNPs in liver, subcutaneous, or omental fat [117]. Methods for measuring allele-specific expression of candidate genes from tissues of subjects heterozygous for trait-associated SNPs have allowed better attempts to define new eQTLs from existing

SNPs. Ongoing efforts to utilize catalogues of human tissue for which gene expression and SNP genotypes have been interrogated, such as the Genotype Tissue Expression consortium efforts, will undoubtedly help in this regard [150,151].

Common variant association studies for lipids have also uncovered trans-eQTLs, SNPs influencing the expression of genes indirectly through affecting expression of their transcriptional regulators. The study of *KLF14*, a HDL-C and type 2 diabetes candidate gene, offers a striking example. SNPs at the locus containing *KLF14* were associated with these traits in GWAS and were also shown in initial studies to be maternally restricted cis-eQTLs for *KLF14* expression in adipose tissue [152]. Subsequent work in 2011 showed that these same SNPs were strongly correlated with expression of several genes in adipose tissue in trans. This study also demonstrated that many of those genes whose expression was strongly associated with the *KLF14* noncoding SNP allele were themselves correlated significantly with signatures of metabolic diseases including plasma lipids, BMI, and measures of insulin resistance [153]. Taking a more global approach, a recent eQTL meta-analysis in approx. 8000 subjects identified several SNPs existing as cis-eQTLs at one locus while also acting as trans-eQTLs at other loci, including one at the 3' UTR of the *FADS1* gene which altered the expression of the *LDLR* gene in trans [154]. Such studies highlight both the complexity and the value gained from careful study of casual variants as determined through gene expression measurements.

Because it is not always possible to identify eQTLs from a given spectrum of SNPs at a locus or because eQTLs may be identified at loci for genes with unknown biological function, additional experimental approaches may allow further identification of causal genes. One approach has been the in vivo overexpression or knockdown of

candidate genes at a locus in animal models. The initial GLGC GWAS reported such work for three candidate genes influencing HDL-C: *GALNT2*, *PPP1R3B* and *TTC39B* [86]. In accompanying work, hepatic *SORT1* (but not *PSRC1*) overexpression in hypercholesterolemic mice demonstrated substantial reduction in LDL-C levels, directionally consistent with the eQTL data in human liver and providing strong evidence that the causal gene at the LDL-associated chromosome 1p13 locus was *SORT1* (see below for more discussion) [140]. An ongoing effort in our laboratory has utilized this approach of somatic overexpression and knockdown in mice to systematically interrogate gene(s) at many of the novel loci reported in the 2010 GLGC GWAS for effects on plasma lipids in vivo. Others are utilizing complementary approaches in cellular models through overexpression and siRNA knockdown [155,156] or genome editing using transcription activator-like effector nuclease (TALEN) and clustered regularly interspaced palindromic repeat DNAs (CRISPR)/Cas9 nuclease genome-editing methodologies [129,157,158].

New biology inspired by common variant studies

Among the multiple novel lipid loci borne from recent GWAS, several have revealed new biological paradigms for the genetic and physiological regulation of lipoprotein metabolism. Additionally, synergistic common and low-frequency (MAF between 0.01-0.05) variant-focused approaches that have emerged in the post-GWAS era have allowed for further studies of loci that were lipid candidate genes before the emergence of GWAS. In an effort to illustrate examples of how follow-up studies may be performed for such loci (an approach taken in this dissertation), I discuss three of the six original novel lipid loci from GWAS: *SORT1*, *TRIB1*, and *GALNT2*. Functional studies of

these genes benefited from localization of their tissue sites of action with regard to lipoprotein metabolism (in the case of these three, the liver), development of tools for somatic gene manipulation in liver, and in the case of *GALNT2*, identification of functional coding variation through sequencing at the phenotypic extremes.

SORT1 (sortilin)

SNPs at the chromosome 1p13 locus were among the first associated with both CAD and LDL-C levels in GWAS, dating back to the original Wellcome Trust Case Control Consortium GWAS for CAD in 2007 [159]. Subsequent studies have validated the associations between this locus and LDL-C and CAD in large European and non-European cohorts. Homozygosity for the minor allele of the lead SNP at this locus (approximately 25% minor allele frequency in Europeans) was associated with a 16 mg/dL decrease in plasma LDL-C and 40% reduction in risk of MI [131]. The locus contains several genes in close proximity to the lead SNPs, including *SORT1*, *CELSR2*, and *PSRC1*, none of which were previously implicated in LDL metabolism.

Initial work aimed at identifying both the casual variant and the causal gene at this locus underlying the association with lipids and protection from disease [140]. The lead SNPs for association with plasma lipids occur in a noncoding region between *PSRC1* and *CELSR2*, but this region is also close in proximity to *SORT1*. Adding further complexity, expression of all three genes is elevated by the minor allele of the SNP at the locus in eQTL analyses of human liver samples but not in subcutaneous or omental adipose samples. Through fine mapping, generation of bacterial artificial chromosome constructs expressing different alleles of the SNP haplotype, luciferase reporter assays, and additional functional studies, Musunuru et al. showed that the minor allele of the

causal variant within a cis-regulatory region, rs12740374, created a de novo CEBP transcription factor binding site that caused CEBP-dependent upregulation of expression of the nearby genes [140].

In the same study, as noted above, adeno-associated virus (AAV) vectors were used to effect liver-directed gene transfer and demonstrated that hepatic *SORT1* overexpression, but not *PSRC1* overexpression, markedly reduced plasma LDL-C levels in hypercholesterolemic mice. This finding was directionally consistent with the human genetics data showing that the minor allele at rs12740374 substantially increased *SORT1* expression and reduced plasma LDL-C. By contrast, the lack of effect on LDL-C by *PSRC1* overexpression suggested that this gene was probably not the causal gene at the locus underlying the LDL-C association.

SORT1 encodes sortilin, a multi-ligand cell-surface receptor and intracellular trafficking chaperone with endolysosomal sorting motifs. It was hypothesized that it may promote trafficking and presecretory degradation of VLDL in hepatocytes (a process previously described but without a molecular mechanism [160]) as a mechanism to reduce VLDL secretion and eventually circulating LDL-C levels. Indeed, overexpression of *SORT1* was shown to reduce VLDL TG and ApoB secretion in mice [140].

However, studies of *SORT1* knock-out (KO) mouse models have highlighted its complexity in VLDL and LDL metabolism. Soon after the initial report by Musunuru et al., Kjolby, et al reported a *SORT1* KO mouse, which when crossed to an *LDLR*-deficient background exhibited lower LDL-C and reduced VLDL secretion [161]. The mouse model used in these studies was a *SORT1* deletion mutation resulting from disruption of exon 14, resulting in some aberrant protein being expressed [162,163]. Yet another *SORT1* KO model, demonstrating complete lack of *SORT1* expression [164], was

subsequently also shown to have reduced VLDL secretion [165]. Thus, both hepatic overexpression and complete deficiency of *SORT1* reduce VLDL secretion in mice. This paradoxical finding has yet to be fully resolved.

Sortilin traffics to the plasma membrane and has several known extracellular ligands that it binds and chaperones to the lysosome [166]. LDL was shown to have high affinity for sortilin by surface plasmon resonance and to be taken up and degraded at a higher rate by liver cells overexpressing sortilin [165]. Moreover, hepatic overexpression of sortilin resulted in faster catabolism of LDL *in vivo* in mice, and this also occurred in *LDLR* deficient mice, indicating that the effect was independent of the LDL receptor. In contrast, sortilin deficient mice had slower catabolism of LDL and this was also true in *LDLR* deficient mice [165]. These studies demonstrated that sortilin binds LDL directly and is a physiologically relevant cell-surface LDL receptor in the liver, suggesting a complex orchestration of LDL metabolism by this protein. Recent work has also indicated that sortilin may facilitate hepatic PCSK9 secretion [167] and also may regulate vascular disease in extrahepatic tissues through multiple mechanisms [168-170]. Returning to the human genetics, increased expression of sortilin in the liver might be expected to both reduce VLDL production and increase LDL catabolism, perhaps accounting for the very strong association of the minor allele (causing elevated sortilin expression) with lower levels of LDL-C. While much more work on the cell biology and physiology is needed, it is safe to say that GWAS first put sortilin on the map as a molecule relevant to LDL metabolism in humans.

TRIB1 (tribbles-1)

TRIB1 was also among the initial loci implicated in influencing plasma lipids in the first GWAS for plasma lipids [131,132]. Lead SNPs are located on chromosome 8p24 approximately 40 kb away from the *TRIB1* gene. SNPs at this locus are among the handful of SNPs associated with all major plasma lipid traits (TG, HDL-C, LDL-C); the minor allele at the lead SNP is associated with the atheroprotective lipid profile of reduced TG, reduced LDL-C and increased HDL-C [86]. Indeed, in addition to being associated with a favorable lipid profile, these SNPs have been associated with significantly reduced risk of CAD in both GWAS of large European populations and in gene-centric array association studies [86,88,171,172]. Additionally, these SNPs have been associated with related metabolic phenotypes such as plasma adiponectin levels and serum transaminases [173,174].

TRIB1 encodes a protein called tribbles-1 that is a cytosolic scaffolding protein with a poorly defined molecular function in the liver [175,176]. Additionally, no known liver eQTLs exist for the SNPs at this locus. However, given the high level of expression of *TRIB1* in the liver, the association of *TRIB1* SNPs with serum transaminases, and the prominent role of the liver in lipoprotein metabolism, the liver has been the focus of much work on relating *TRIB1* to plasma lipid levels. Using a design similar to the initial in vivo validation of *SORT1*, *TRIB1* was somatically overexpressed in the liver in mice and produced a reduction in plasma TG and cholesterol through impaired VLDL secretion [177]. In contrast, *TRIB1* KO mice demonstrated elevated plasma TG and cholesterol and increased VLDL secretion. Differences in plasma lipids in the overexpression and KO models were accompanied by reciprocal differences in hepatic expression of lipogenic genes, and subsequent studies in a liver-specific Trib1 KO model have

demonstrated that *TRIB1* is a key regulator hepatic lipogenesis in vivo through modulation of *CEBP α* turnover [178].

Additionally, *TRIB1* may influence lipid metabolism through extrahepatic roles. One recent study of mice with *TRIB1* deficiency in the spleen and hematopoietic cells including macrophages demonstrated increased lipolysis in these mice [179]. Consequently, these mice exhibited elevated serum free fatty acids, lower adipose mass, elevated cholesterol, TG, fasting glucose, and insulin resistance on a high-fat diet relative to wild-type controls. The results of this study and the previous one focused on hepatic *TRIB1* function suggest that this gene may contribute to metabolic phenotypes through actions in multiple tissues.

GALNT2 (polypeptide N-acetylgalactosaminyltransferase 2)

SNPs within intron 1 of *GALNT2* at chromosome 1q42 have been consistently associated with HDL-C and TG in multiple GWAS [86,88,117,131,132]. The lead SNP at this locus, rs4846914, has a minor allele frequency of 40% in Europeans and is associated with decreased HDL-C and increased TG. In addition to fasting lipids, this SNP has been associated with rare Fredrickson hyperlipoproteinemias in a smaller case-control study of 386 cases and 242 controls [180]. In addition to lipid traits, SNPs at this locus have also been related to several non-lipid traits in a recent 'phenome'-wide association study, including serum calcium levels and circulating white blood cell count, among other traits [181]. *GALNT2* encodes an O-glycosyltransferase enzyme that, like *SORT1* and *TRIB1*, had never been implicated previously in lipoprotein metabolism.

Given the presence of the lipid-associated SNPs within an intron of *GALNT2* and lack of any genes for several 100 kb either upstream or downstream, *GALNT2* has

regularly been considered to be the causal gene driving the lipid association signal at this locus [130]. Allelic imbalance and eQTL studies using human liver samples from subjects heterozygous for the GWAS SNPs have both shown that SNP alleles associated with higher HDL-C from the multiple GWAS studies were also correlated with increased hepatic *GALNT2* transcript levels [182,183].

This directionality conflicts with initial physiological studies of *GALNT2* expression or activity in animals and humans. Somatic AAV-mediated overexpression of mouse *GALNT2* in the livers of mice resulted in decreased plasma HDL-C and increased TG, while AAV-mediated short hairpin RNA (shRNA) knockdown had reciprocal effects [86]. Additionally, sequencing of the gene in individuals with elevated HDL-C identified humans heterozygous for a missense variant D314A in *GALNT2* that was reported to have reduced catalytic activity in vitro [184]. Compared with non-carrier controls, these individuals had improved postprandial TG clearance and reduced sialylation of ApoC-III, suggesting that the missense variant confers loss-of-function. Indeed, ApoC-III was found independently to be a target of *GALNT2* in liver cells in vitro; a number of other plausible candidate proteins have also been identified as *GALNT2* targets [185,186]. The true directional relationship of *GALNT2* expression, function, and glycosyltransferase activity and repertoire of targets with respect to lipoprotein metabolism has remained unclear and has been one of the foci of efforts in this dissertation.

Looking ahead – leveraging low-frequency variants, Mendelian randomization, and sequencing for rare causal variants

As the field of lipid genetics moves beyond GWAS to focusing on identification of causal variants, causal loci, and biological mechanisms underlying novel genes, additional tools and approaches have emerged to aid these efforts. These methodologies include searching for low-frequency and rare coding variants through exome-wide and targeted genotyping arrays, exome and targeted gene sequencing, and most recently whole-genome sequencing. In the last decade, these tools have developed and matured from limited applications in small sample sizes to large population-based or case-control study designs to uncover new variants or new associations between variants and lipid traits and disease risk. As a central goal of this dissertation work is to demonstrate through examples the utility of these approaches as complementary tools for understanding causal variation influencing CHD risk, I will now briefly review these developing methods and major findings to date.

One focus has been the attempt to identify low-frequency variants of strong effect sizes on plasma lipids that also associate with disease risk. This effort has been fueled in part by the discovery of functional low-frequency coding variants in PCSK9, as described above [32,38,39]. To increase the throughput of identifying functional coding variants of low-frequency at a cost more manageable than currently possible for exome sequencing, an exome-wide SNP array focused on coding SNPs (the “exome chip”) has been developed [187]. Peloso et al. recently used this array to genotype >200,000 low-frequency and rare coding sequence variants in >56,000 individuals of European and African American ancestry [188]. This study identified four new low-frequency variants in novel genes influencing HDL-C and/or TG: ANGPTL8, PAFAH1B2, COL18A1, and

PCSK7. Overexpression of one of these, PCSK7, in mice resulted in elevated TG and reduced HDL-C levels. Despite their robust lipid association, the coding SNPs in these genes were not associated with changes in risk of CHD in this study. Another recent study applied the exome array to genotype >80,000 coding variants in >5,600 individuals of European ancestry and subsequently followed up on 10 loci with genome-wide associations for a coding variant with a lipid trait in an additional cohort. This work identified a novel candidate gene, TM6SF2, at a known locus on chromosome 19 containing *NCAN/CILP2/PBX4*. The coding variant, Glu167Lys, was common (MAF 0.087) and associated with lower total cholesterol. Adenoviral overexpression of this gene in mice raised total cholesterol while adenoviral knockdown with shRNA reduced plasma cholesterol, suggesting a physiological role for this newly identified gene in regulating plasma cholesterol levels. This variant was also associated with reduced risk of MI, thus providing another example of the utility of the exome array coding variant genotyping-focused approach for elucidating variants of relevance to both intermediary lipid traits and disease risk [189].

Currently, low-frequency variants influencing lipids, whether coding or noncoding, have also allowed preliminary interrogation of the causality of these lipid biomarkers with risk of disease. These causality studies have been informally called Mendelian randomization studies and ascertain the effects of randomized collections of variants assigned at meiosis and independent assortment [190,191]. Using variants with robust associations with an intermediary phenotype (e.g. plasma lipid type), one can interrogate the deviation of the predicted effect of a variant on a secondary trait (e.g. disease risk) based on the impact the variant has on the intermediary trait. These approaches allow study of the causality of variants without traditional limitations of observational

epidemiological studies such as confounding and reverse causation (the onset of disease before alteration in intermediary biomarker) [192]. These studies rely on lack of pleiotropic effects of the variants of interest, which can rarely be measured comprehensively, thus adding a caveat to interpretation of results of these efforts. Mendelian randomization using variants associated with HDL-C at the *LCAT*, *CETP*, *APOA1*, *ABCA1*, *LIPC*, and *LIPG* loci have been used to test the causal relationship between HDL-C and CHD. These investigations have largely failed to support a strong causal relationship between HDL-C and risk of CHD [193-196]. Whether it will be possible to perform analogous studies to investigate the causal relationship of functional aspects of HDL, such as cholesterol efflux capacity, as opposed to HDL-C levels, will depend on the identification and ability to test low-frequency genetic variants specifically influencing such traits.

Similarly to these HDL-C-focused efforts, a recent Mendelian randomization study was carried out to test the association of TG-regulating loci with risk of CHD [197]. Using a modified approach, the investigators tested 185 SNPs for association and effect sizes on all lipid traits in the >188,000 individuals of the most recent GLGC cohort. Then, focusing on SNPs with large effect sizes on TG but not on LDL-C, they tested the relationship of the effect size of a SNP on TG with that on CAD risk in >86,000 subjects in the CARDIoGRAM study. This approach demonstrated highly significant causal relationships between SNPs at multiple TG loci with CAD.

Candidate gene and whole exome sequencing permit the identification of very rare and private mutations that may be causal for lipid traits. An example of the power of candidate gene sequencing is the work by Johansen, Hegele and colleagues, who resequenced candidate genes for hyperTG identified from GWAS in a case-control

design. They found a significant burden of rare missense or nonsense coding variants in these candidate genes in HTG cases relative to controls [109]. Motazacker et al recently reported a targeted sequencing study of 195 lipid candidate genes and 78 non-lipid genes in a small cohort of subjects at extremes of the HDL-C distribution. The authors showed an increased frequency of rare nonsynonymous variants in lipid genes, but not non-lipid genes, in individuals with extremely high or extremely low HDL-C levels relative to general population and normal HDL-C participants [198]. The diagnostic yield of coding sequence-only targeted sequencing efforts for HDL-C from this study was relatively low, and similar results have recently been reported for the sequencing of key LDL-C genes in hypercholesterolemic patients [199]. This suggests that perhaps the ability to ascertain causal variants at the lipid phenotypic extremes may be augmented by the inclusion of key noncoding regions likely to harbor functional variants. Indeed, the development of approaches to simultaneously and efficiently capture both the coding and noncoding genome for lipid genes while enriching for extreme phenotypes to increase variant burden yield is one goal of the current dissertation work.

Exome sequencing of individuals at extremes of lipid traits or even whole populations has revealed additional insights, especially for genes conferring risk by recessive modes of inheritance. Exome sequencing in a family with familial combined hypolipidemia identified loss-of-function mutations in *ANGPTL3* as a cause of this inherited lipid disorder [118]. Another recent example by the National Heart Blood and Lung Institute GO Exome Sequencing Project involved sequencing in three family members with autosomal recessive hypercholesterolemia lacking mutations in known candidate genes. Exome sequencing of these family members revealed a novel splice-site mutation in the *LIPA* gene that segregated with the hypercholesterolemia diagnosis

and with accumulation of hepatic cholesterol, the latter of which had been reported previously in patients with other mutations in *LIPA* [200]. Using a case-control design, another recent study sequenced the exomes of 2500 subjects with either extremely high or extremely low LDL-C and followed up by additional sequencing in another >1300 subjects and exome array genotyping in >55000 subjects. This work identified significant associations of rare or low-frequency variants in known LDL-C modifying genes *PCSK9*, *LDLR*, and *APOB*, but also a novel gene, *PNPLA5*. The authors noted that the effect sizes of the low-frequency variants on LDL-C were greater than those found for LDL-C associated common SNPs from GWAS, suggesting that low-frequency focused approaches may be useful for identifying more functionally significant SNPs for certain traits [201]. Demonstrating the medical utility of population-based exome sequencing, an effort from the Myocardial Infarction Genetics Consortium investigators demonstrated through exome sequencing of ~7,300 CHD cases and ~14,700 controls across ethnicities that inactivating mutations in *NPC1L1*, a gene encoding an enterocyte transporter protein critical for cholesterol absorption, were associated with protection from CHD risk [202]. Soon after this study was published, a large clinical trial of ezetimibe, an LDL-C drug which targets the product of *NPC1L1*, was released showing that this medication improved cardiovascular outcomes and offered additional LDL-C lowering when added to statin therapy [203]. This example highlights the potential of rare variant genetics uncovered from exome sequencing to inform and relate to ongoing pharmacological interventions.

As mentioned earlier, several key results from sequencing efforts aimed at identifying variants underlying TG levels are likely to further shape the development of future TRL-focused therapies. These include the discovery of *APOC3* LOF coding

variants that were associated with marked decreases in TG levels, minimal effects on LDL-C but 40% reductions in CHD or ischemic vascular disease risk. Two seminal studies in 2014 identified the same LOF variants in low TG individuals through different approaches. The first of these implemented exome sequencing in an initial population cohort of ~3,700 individuals followed by replication sequencing and exome chip genotyping in ultimately >110,000 subjects (stratified as CHD cases vs. healthy controls) [106], and the second performing targeted sequencing of *APOC3* codons in a large population-based cohort of ~75,700 individuals from the Copenhagen City Heart Study [91]. These complementary investigations demonstrate the utility of the varied approaches in arriving at a similar conclusion, namely that *APOC3* is a critical node in the relationship between TGs and CHD risk and that intervention to lower ApoC-III levels or function may have clinical benefit. Similar parallels were observed in the even more recent reports of *ANGPTL4* LOF variants associated with reduced TGs and CHD risk. Here, one group of investigators (the Myocardial Infarction Genetics and CARDIoGRAM Exome consortia) again combined sequencing in a smaller discovery cohort with genotyping in a larger cohort of CHD cases vs. controls to relate LPL and *ANGPTL4* variants with TGs and CHD risk[90]. A second study also implemented exome sequencing in a large BioBank initiative of phenotyped patients from a single community health system to relate *ANGPTL4* variants with CHD and related measures[89]. Both studies found an association between LOF variants in *ANGPTL4*, low TGs, and protection from CHD. Notably, the second study also demonstrated preliminary efficacy of a monoclonal antibody targeting *ANGPTL4* in mice and a nonhuman primate, further highlighting the potential for translation of the findings from current genomics discoveries to novel therapies.

Overall Summary and Dissertation Goals

The last decade has yielded much excitement and also some controversy regarding the study of plasma lipid traits, still a prevailing and modifiable set of risk factors for CHD and MI. Novel therapies and insight have reaffirmed our ability to drastically lower disease risk through reduction of LDL-C levels, but at the same time, the role and future of efforts to target HDL-C and TG remain uncertain. Likewise, the advent of genomics has allowed the unprecedented study of coding and noncoding heritable variation in literally hundreds of thousands of individuals, thus providing new opportunities to better understand the mechanisms by which these traits impact disease, if at all. Moreover, the current ability to dive deeply into the coding genome may allow a better understanding of TG- and HDL-modifying variants that protect from disease in a manner that could influence the development of novel therapies based on these genetic mechanisms.

As the realm for investigation in this current era of lipid genetics is inconceivably vast, a goal of this dissertation was to take a sophisticated sampling of the ways in which the modern genomics toolkit can be applied in concert to address key topics relevant to the relationship of HDL, TG, and CHD. Namely, I sought to address the following questions:

- 1) How does genetic variation that impacts HDL functionality influence CHD risk? – Here, my colleagues and I utilized the sequencing and genotyping approaches above coupled with functional biology for the identification and interrogation of a human genetic variant in SR-BI causally disrupting the HDL reverse cholesterol transport pathway to assess how this may impact HDL metabolism and CHD risk in humans (Chapter 2).

- 2) What is the mechanism and directional relationship underlying a novel GWAS-locus for HDL-C and TG to these traits and how does this novel association refine biological understanding of these traits? – Here, I combined interrogation in multiple model systems and exome sequencing in humans to study loss-of-function of *GALNT2* across mammalian species and defined plausible mechanisms relating inactivation of this gene to low HDL levels, thus outlining one strategy to approach novel GWAS loci for complex traits (Chapter 3).
- 3) What is the mechanism of protection from high TGs and CHD for coding variants in *APOC3*, a critical regulator of TRL metabolism? Here I interrogated a missense variant in *APOC3* that was shown in exome and targeted sequencing studies to protect from CHD by unknown mechanisms. Through study in humans, animal models, and biochemical approaches, I outline a mechanism of protection via ApoC-III lowering that may inform therapeutic efforts targeting this pathway, and outline how structure-function studies can be applied to the evaluation of the many coding variants arising from population-based exome sequencing for complex diseases (Chapter 4).
- 4) How can we improve the detection of causal noncoding variation underlying phenotypic extremes? Here I apply a novel sequencing methodology to perform one of the first targeted sequencing studies for noncoding variants at HDL-C candidate genes in humans with extremely high HDL levels. I demonstrate a proof-of-principle approach that validates previous variants and discovers novel regulatory mutations and highlights the potential for larger systematic efforts to capture rare noncoding variation influencing complex traits (Chapter 5).

Figure 1.1

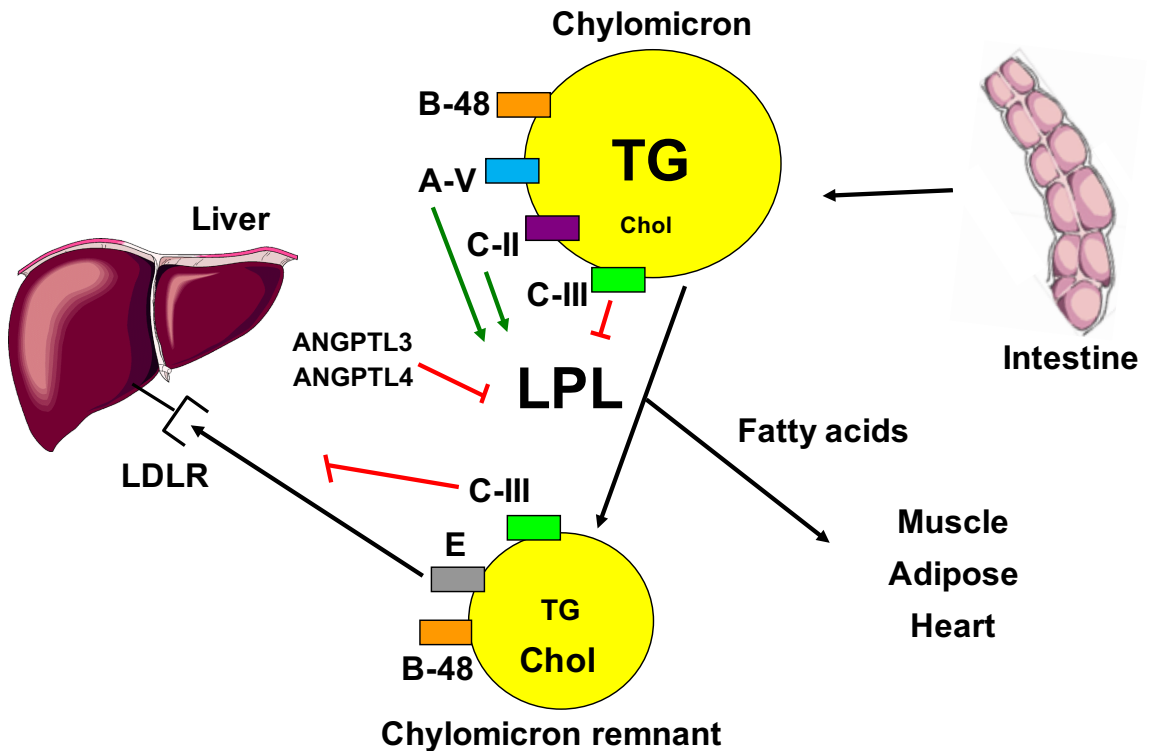


Figure 1.1: Chylomicron Metabolism. Chylomicrons are made in the small intestinal enterocytes after dietary intake and contain the apolipoproteins B48, A-V, C-II, and C-III. ApoB48 is the key structural component of chylomicrons and remnants. ApoC-II is a required cofactor for lipoprotein lipase (LPL) activity, and ApoA-V also promotes LPL function. ApoC-III and the angiopoietin-like proteins ANGPTL3 and ANGPTL4 inhibit LPL activity. LPL hydrolyzes TGs on chylomicrons to yield fatty acids which can be taken up by peripheral tissues such as the muscles, heart, and adipose. LPL-mediated remodeling of chylomicrons yields remnant particles which can acquire ApoE, a critical ligand for the LDL receptor family members that mediate remnant particle clearance by the liver. ApoC-III on remnant particles can inhibit their clearance by this pathway.

Figure 1.2

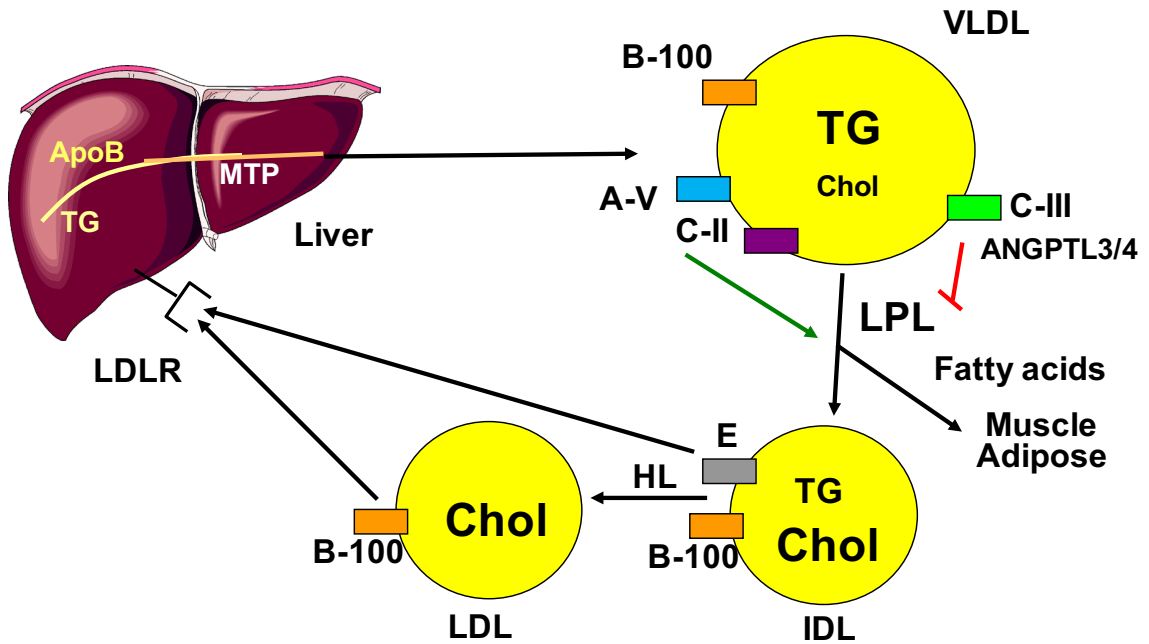


Figure 1.2: VLDL and LDL metabolism. VLDL is synthesized in the liver through the lipidation of ApoB100 with TGs. This process is in part mediated by the microsomal triglyceride transfer protein (MTP). Secretion of VLDL yields a TG-rich particle containing ApoB100, ApoA-V, ApoC-II and ApoC-III. ApoA-V and ApoC-II activate LPL-mediated lipolysis of VLDL-TGs, while ApoC-III and the circulating ANGPTL3 and ANGPTL4 proteins inhibit this process. LPL and hepatic lipase (HL) mediated remodeling of VLDLs yield smaller, denser lipoproteins, the intermediate density lipoproteins (IDLs) and LDLs. These particles can be cleared from the circulation through LDLR and related receptors in the liver through ApoE or ApoB as ligands for receptor-mediated endocytosis.

Figure 1.3

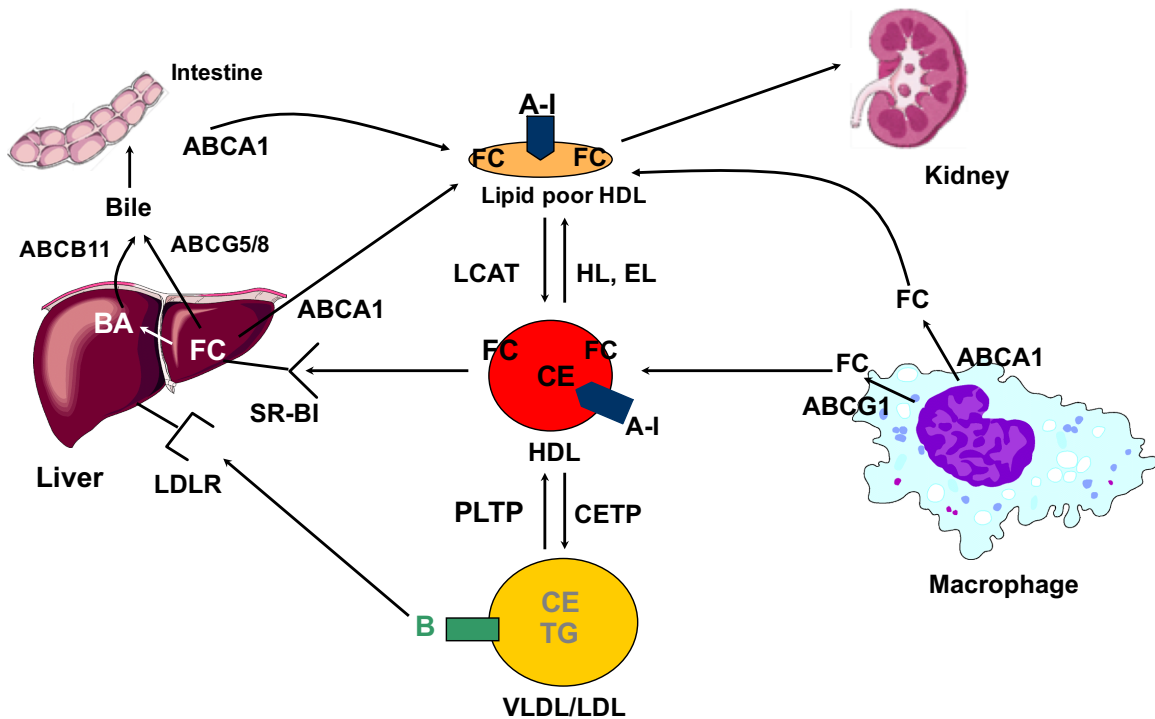


Figure 1.3: HDL metabolism. HDL is initially synthesized through the hepatic or intestinal secretion of ApoA-I, which then quickly acquires free cholesterol (FC) and some phospholipids to become lipid-poor HDLs. These particles are cleared rapidly from the kidney unless they mature. Maturation of HDL occurs through the further acquisition of FC and esterification of the FC using phospholipid-derived fatty acids to yield cholesteryl esters (CE); this process is mediated by the enzyme lecithin:cholesterol acyltransferase (LCAT) and results in a mature, spherical HDL particle. Lipid-poor HDLs and mature HDLs can also acquire FC from macrophages and other cells in peripheral tissues through efflux by ATP-binding cassette transporters ABCA1 and ABCG1 on the cell surface. ABCA1 effluxes FC to lipid-poor HDLs while ABCG1 effluxes FC to mature spherical HDLs. In both cases, LCAT esterifies the FC to CE and promotes further remodeling of the particles. HDL CEs and phospholipids can be transferred to ApoB-containing lipoproteins (such as VLDL and LDL) through the actions of cholesteryl ester

transfer protein (CETP) and phospholipid transfer protein (PLTP), respectively. HDL CEs are also cleared from the circulation by the scavenger receptor BI (SR-BI) on the surface of liver cells and other tissues. SR-BI mediates the selective uptake of HDL CEs to yield CE depleted lipid-poor particles which can re-mature through the above mentioned processes or be cleared by the kidney. CEs taken up by the liver can be hydrolyzed to FCs and incorporated into bile acids (BA) and effluxed to the bile through ABCB11 or FCs directly can be incorporated into bile through efflux by ABCG5 and ABCG8. Biliary excretion of cholesterol originally present in peripheral tissues such as arterial plaque macrophages through the delivery by lipoproteins to the liver and selective uptake is known as reverse cholesterol transport and is thought to be one of the key atheroprotective processes mediated by HDL.

CHAPTER 2 – RARE VARIANT IN SCAVENGER RECEPTOR BI RAISES HDL CHOLESTEROL AND INCREASES RISK OF CORONARY HEART DISEASE

Introduction

The strong inverse association between plasma high density lipoprotein (HDL) cholesterol concentrations and coronary heart disease (CHD) led to intense interest in a potential causal relationship between HDL metabolism and CHD. Recent clinical trials with the HDL cholesterol raising drugs niacin and CETP inhibitors have been disappointing, suggesting the possibility that higher HDL-C levels are not causally protective against CHD [61,64]. Recent studies of human genetic variants that are associated with HDL-C levels have likewise generally failed to show association with CHD [193,194]. Most notably, a loss-of-function variant in the gene *LIPG* encoding the enzyme endothelial lipase that in the heterozygous state clearly raises HDL-C by ~5 mg/dL was found to have no association with CHD [204]. Additionally, early studies of genetic variants in the CETP and LIPC genes suggested that some genetic variants could be associated with increased HDL-C and an increased risk of CHD[205,206]. Thus, human genetics can provide insight into the complex relationship between HDL and CHD.

The scavenger receptor class BI (SR-BI) was discovered to be an HDL receptor 2 decades ago by Krieger and colleagues [207]. SR-BI promotes the selective uptake of HDL cholesteryl esters into cells (particularly hepatocytes and steroidogenic cells)[207,208]. In mice, overexpression of SR-BI in the liver markedly reduces HDL-C levels [209-212], and genetic deletion of SR-BI results in markedly increased HDL-C levels [213-215]. Remarkably, effects of these genetic manipulations in mice have

effects on atherosclerosis opposite to the direction predicted by the human epidemiology: overexpression reduces atherosclerosis despite the lower HDL-C [216-218] and gene deletion increases atherosclerosis despite the higher HDL-C levels [219-222]. One potential explanation relates to the flux of cholesterol from macrophages through the reverse cholesterol transport (RCT) pathway: SR-BI overexpression increases macrophage RCT and knockout reduces macrophage RCT [223]. The relevance of these observations in mice to humans has been of great interest.

SR-BI is encoded by the gene *SCARB1*, and common gene variants at the *SCARB1* locus have been genome-wide significantly associated with plasma HDL-C levels [86,117]. A rare missense variant in *SCARB1* was reported in one extended family to be associated in the heterozygous state with modestly elevated HDL-C levels, but non-invasive imaging failed to show any relationship with atherosclerosis [224]. Thus, human genetics has not yet clarified the relationship between SR-BI function, HDL-C levels, and cardiovascular disease in humans.

Through targeted sequencing of a number of individuals with extremely high levels of HDL-C, we identified a homozygote for a rare missense variant in *SCARB1* - P376L. We established that this mutation causes virtually complete loss of function of SR-BI due to impaired processing and cell-surface presentation. This variant was seen much more frequently in other subjects with very high HDL-C compared with those with low HDL-C; furthermore, after genotyping > 100,000 subjects we found it to be definitively associated with higher HDL-C levels. In a separate genotyping experiment, we found this variant to be significantly associated with *increased* risk of CHD. Thus, this series of studies identifies the first *SCARB1* 'knockout' human, establishes the first example in which a coding gene variant raises HDL-C (without affecting other plasma

lipids) but *increases* risk of CHD, and extends our understanding of SR-BI as a protein that not only modulates HDL metabolism but also cardiovascular risk in humans.

Materials and Methods

Subject Identification and Ascertainment

Individuals with high HDL-C levels above the 95th percentile for age and sex were systematically recruited. For this study, 341 individuals of European ancestry with HDL-C > 95th percentile were selected for targeted sequencing, as were 407 individuals of European ancestry with low HDL-C levels below the 25th percentile for age and sex. Briefly, using customized hybrid capture (Agilent) baits, we targeted and enriched for sequencing the exons of the ~990 genes located within 300 kb of each of the 95 loci with significant associations ($P < 5 \times 10^{-8}$) with plasma lipid levels identified by the Global Lipids Genetics Consortium as of 2010 [86]. We performed next-generation sequencing using Illumina Genome Analyzers at the Broad Institute as described previously [225]. Base pairs were called and sequencing reads were aligned to the human genome reference GRCh37 (hg19), and sequencing metrics were processed using the Picard pipeline with an output of Binary Alignment Map (BAM) files.

The Genome Analysis Toolkit Unified Genotyper was used to genotype all variant sites, calculate initial quality control metrics, and filter based on these values, and variants were annotated using SnpEFF [226]. Variants were indexed by individuals bearing them in a Variant Call Format (VCF) file. Genotypes for each individual were computed at each site and tabulated.

Additionally, *SCARB1* mutation carriers were identified by genotyping subjects with high vs. low HDL-C using the Exome Chip (HumanExome BeadChip v1.0, Illumina,

Inc., San Diego, CA). The Exome Chip contains >240,000 coding SNPs derived from all mutations found >2 times across >1 dataset among 23 separate datasets comprising a total of >12,000 individual exome and whole genome sequences. The P376L variant is included. In total, 353 high HDL-C males (HDL-C \geq 70 mg/dL), 747 high HDL-C females (HDL-C \geq 80 mg/dL), 1106 low HDL-C males (HDL-C \leq 40 mg/dL), and 573 low HDL-C females (HDL-C \leq 50 mg/dL) were genotyped using the Exome Chip. Subject samples were drawn from previous research studies conducted in our laboratory. After removing subjects from analysis for which demographic covariates were not available, there remained 805 subjects with high HDL-C and 989 subjects with low HDL-C analyzed for association of *SCARB1* P376L allele frequency with HDL-C.

To confirm the presence of the P376L mutation in carriers, a nearby region of 995 bp of *SCARB1* was amplified by PCR using the following primers: forward - TGGTTTGGTTGGTCAGTGGCG, reverse - AGGGCTGCCTCCAGCTCACAT; and the following PCR conditions: 95°C, 2 min; (95°C, 30 s; 62.6°C, 30 s; 72°C; 100.0 s) repeat 29 times; 72°C, 5 min; 4°C forever. PCR products were purified using a QIAquick PCR Purification Kit (Qiagen, Germantown, MD, USA) and sequenced by Sanger sequencing [227] by Genewiz Inc (South Plainfield, NJ, USA). Sequences were compared with *SCARB1* NCBI reference sequence NG_028199.1 using Sequencher (Gene Codes, Ann Arbor, MI, USA).

In silico analysis

To predict the effect of the P376L mutation on SR-BI protein structure and function we used the publicly available algorithms Condel and Raptor X [228-231]. Condel generated a “consensus deleteriousness” score, obtained by combining 3

prediction algorithms (PolyPhen-2, MutationAssessor, and SIFT). Raptor X was used to predict effects on local secondary structure of the SR-BI due to the P376L variant. Protein alignment data were generated by MacVector (MacVector, Inc., Cary, NC) using a Gonnet similarity matrix with open gap penalty = 10 and extended gap penalty = 0.1.

DNA cloning and adeno-associated virus generation

The coding sequence of SR-BI (CCDS 9259.1) was obtained from Thermo Scientific (cDNA clone MGC:120767 IMAGE:7939577) in a pCR4-TOPO cloning vector (Invitrogen). The P376L and the P297S mutations were inserted in the sequence by site directed mutagenesis using the Quickchange II site directed mutagenesis kit (Agilent technologies) and following primers: P376L-F 5'-cctggacatccacctggtcacgggaatcc-3'; P376L-R 5'-ggattcccgtgaccaggtggatgtccagg-3'; P297S-F 5'-gggtgttgaaggcatctccacctatcgcttcgt-3'; P297S-R 5'-acgaagcgataggtggagatgccttcaaacc-3'. The wild type and the mutated coding sequences were then amplified by PCR and subcloned into a pcDNA3.1/V5-His TOPO expression plasmid (Invitrogen) according to manufacturer`s instructions. For *in vivo* studies, *SCARB1* WT and P376L cDNA sequences were amplified by PCR using the following primers containing KpnI and NotI sites, and then subcloned by KpnI/NotI double digestion followed by ligation into adeno-associated virus serotype 8 vector (AAV8) containing the liver-specific thyroxine-binding globulin (TBG) promoter provided by the University of Pennsylvania Vector Core [232-237]. Virus production, purification and quantification were carried out by the core facility.

Radioactive labeling of HDL

Total human HDL ($1.063 < d < 1.21$ g/ml) and HDL₃ ($1.125 < d < 1.21$ g/ml) were obtained by sequential ultracentrifugation as described before [238]. HDL was labeled with ¹²⁵I by a modification of the iodine monochloride method for the binding experiment at 4 °C and they were labeled with ¹²⁵I-tyramine-cellobiose (¹²⁵I-TCB) for all the other experiments [239]. To further label ¹²⁵I-TC-HDL with ³H-cholesteryl-ether (³H-CE), 500 µCi of ³H-CE resuspended in 50 µL of ethanol were added to a solution containing heat-inactivated human lipoprotein deficient serum (200 mg protein) and iodinated HDL (6 mg protein). The solution was incubated overnight at 37 °C with gentle shaking, followed by reisolation of the dual-labeled HDL by sequential ultracentrifugation.

Cell culture and in vitro assays

COS7 cells were cultured in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum at 37°C in a humidified 5% CO₂ incubator and were passaged using trypsin. For selective cholesterol uptake COS7 cells were plated at a density of 3×10^4 cells/cm² in 6 well plates. One day following plating (Day 1), cells were transfected using Lipofectamine 2000 (4 µg DNA/well, 3:1 Lipofectamine:DNA ratio) with a pcDNA3.1/V5-His-TOPO plasmid encoding either *SCARB1* WT, *SCARB1* P376L, or *SCARB1* P297S cDNAs. One group received a mock transfection with plasmid containing no cDNA. Each experimental group was tested in quintuplicate. On Day 2, culture medium was changed to DMEM 0.5% BSA containing 20 µg/mL of ³H-cholesteryl hexadecyl ether/¹²⁵I-TCB-labeled HDL₃ and incubated for 3 hours at 37 °C. In one well per plate a 40-fold excess of nonradiolabeled (“cold”) HDL was added to a final

concentration of 800 µg/mL. After incubation cells were washed with PBS and lipids were extracted in two consecutive steps using a 3:2 hexane-isopropanol mixture. After drying, proteins were solubilized using 2 mL of 0.1N NaOH. The hexane/isopropanol fraction was dried under nitrogen, resuspended in 600 µL of toluene, and counted separately in scintillation and gamma counters. The NaOH fraction was counted in the gamma counter and measured for protein concentration using a Lowry protein assay [240]. Counts from wells with 40-fold excess of cold HDL (non receptor-specific counts) were subtracted from counts from other wells to calculate the receptor-specific activity internalized. Data analysis was performed as previously described [241], with the following difference: ¹²⁵I activity (counts per minute) in the NaOH fraction were considered as cell-surface associated lipoproteins (it has been shown previously that when performing this experiment under similar conditions, the fraction of cell associated counts that is trichloroacetic acid-soluble is approximately 5%) [242]. A negligible amount of ¹²⁵I counts were found in the hexane-isopropanol fraction. SR-BI protein expression was determined by western blotting. Human β-actin was used as a loading control.

The binding of ¹²⁵I-HDL₃ at 4°C to COS7 cells was measured using a modification of the method from Nieland *et al* [243]. Briefly, COS7 cells were seeded at a concentration of 3x10⁴ cells/cm² in 12 well plates. The next day, cells were transfected using Lipofectamine 2000 (1.6 µg DNA/well, 3:1 Lipofectamine:DNA ratio) with a pcDNA3.1/V5-His-TOPO plasmid encoding either *SCARB1* WT, *SCARB1* P376L, or *SCARB1* P297S cDNAs. An additional plate of cells was transfected with a plasmid encoding GFP (pAAV-CB-EGFP), which was used as a control. On the following day, cell plates were precooled for 30 minutes on ice, washed with cold DMEM and then

exposed for 2 hours to the assay medium (DMEM, 1% P/S, 0,5%BSA + lipoproteins) at 4 °C. After the incubation, cells were washed twice with ice-cold PBS containing 2 mg/mL of bovine serum albumin (BSA) and a third time with PBS without BSA. Cells were then lysed in 0.1N NaOH. Lysates were used for ¹²⁵I counting and for measuring protein content using a Lowry assay. In parallel to this, cells were seeded in 10 cm plates, transfected as described above and lysed after 24 hours in RIPA buffer to check SR-BI protein expression.

Cell surface biotinylation assay

COS7 cells were seeded and transfected as described above in 175 cm² flasks. 24 hours after transfection, cell-surface-associated proteins were isolated using the Pierce Cell Surface Protein Isolation Kit (Pierce Biotechnology Inc., Rockford, IL) according to the manufacturer's instructions. After lysis, protein concentration was determined by bicinchoninic acid (BCA) assay and the same amount of proteins was loaded on NeutrAvidin Agarose beads to isolate biotinylated proteins. After multiple washing steps, proteins were eluted from the beads and loaded on a 10% Bis-Tris polyacrylamide gels for western blotting. Blots for β -actin and Na⁺/K⁺-ATPase were used as intracellular and surface-associated controls, respectively.

Induced pluripotent stem cell differentiation to hepatocyte-like cells

We generated induced pluripotent stem cells (iPSC) from peripheral blood mononuclear cells (PBMCs) and differentiated them into hepatocyte-like cells as described elsewhere [244,245]. After complete differentiation of iPSCs to hepatocyte-like cells (approx. 20-22

days from initiation of differentiation), we performed selective cholesterol uptake as described above.

Endoglycosidase H sensitivity assays

For Endoglycosidase H sensitivity assays, 100 µg of liver homogenates from mice transduced with human *SCARB1* or Null AAV8 vectors were treated with Endoglycosidase H (New England Biolabs, Ipswich, MA), Sialidase A, or a combination of Sialidase A and PNGase F (Prozyme, Hayward, CA) according to manufacturer's instructions. The digestion products were then loaded on a 10% Bis-Tris polyacrylamide gel for western blotting. Non-treated homogenates were used as controls. Endoglycosidase H sensitivity assays were also performed using lysates from transfected COS7 cells and from iPSC-derived hepatocytes. In these experiments, cell lysate in RIPA buffer plus Complete Protease Inhibitor Cocktail (Roche) was used.

AAV-mediated overexpression of SCARB1 WT and P376L in Scarb1 KO mice

AAV serotype 8 vectors expressing *SCARB1* WT or P376L were expressed in *Scarb1* KO mice, which were subsequently studied for effects on lipoprotein metabolism. Male *Scarb1* KO mice (6 per group) were weighed, bled through retro-orbital bleeding and injected with 10^{12} genome copies (GC) of AAV-*SCARB1*-WT, AAV-*SCARB1*-P376L, or Null vector via intraperitoneal injection. Mice were fasted for 5 hours, weighed, and bled again at 12 days after injection. Plasma lipid levels were determined as described above. At 2 weeks after injection, HDL kinetics using radiolabeled human HDL was measured [246]. Briefly, mice were injected via tail-vein injection with a mixture of ^{125}I -TCB-HDL (3×10^6 cpm/mouse) and ^3H -cholesteryl-ether (CE)-HDL (3×10^6 cpm/mouse)

and bled at 2 minutes post-injection and 1, 3, 6, 9, and 24 hours. Plasma ^3H and ^{125}I activity at each timepoint were determined using scintillation counting and gamma counting, respectively. At 24 hours post-injection of radiolabeled HDLs, mice were weighed, anesthetized, terminally bled, and sacrificed. Livers were lysed in phosphate buffered saline (PBS) and ^3H and ^{125}I were counted in the lysates to determine liver-associated counts. Liver-associated counts were then expressed as μg of HDL component/mg of liver to allow a direct comparison between liver associated ^3H and ^{125}I counts. Selective cholesterol uptake was calculated as the difference between liver-associated ^3H counts and liver associated ^{125}I counts expressed as μg of HDL component/mg of liver. SR-BI protein expression levels in liver homogenates were determined by western blot followed by densitometric analysis performed with ImageJ (U.S. National Institutes of Health, Bethesda, MD) [247].

Subject selection and study visit

Carriers of the P376L variant identified through next-generation sequencing were recruited through a comprehensive recall protocol approved by the institutional review board of the Perelman School of Medicine at the University of Pennsylvania, Philadelphia. Additionally, control subjects were chosen from a database of previous participants in studies conducted in our laboratory. All subjects recruited gave informed consent. Two groups of controls were selected: those having HDL-C levels between the 25th and the 75th percentile for age and sex (normal HDL-C control group) and those with HDL-C levels greater than the 75th percentile for age and sex but confirmed to lack the P376L variant (high HDL-C control group). The two control groups were selected to match on aggregate the age, gender and race of the carrier group. Study visits were

performed in the Clinical and Translational Research Center (CTRC) facility at the Perelman School of Medicine at the University of Pennsylvania. Venous blood after overnight fasting was drawn from each subject to measure a comprehensive metabolic panel, complete blood count, standard urinalysis, reticulocyte count, T- and B-cell counts, anti-nuclear antibody screen, anti double-stranded DNA antibody screen, and anti-neutrophil cytoplasmic antibody screen, which were measured by the William Pepper Laboratory of the Hospital of the University of Pennsylvania. A comprehensive lipid panel (total cholesterol [TC], VLDL cholesterol [VLDL-C], LDL cholesterol [LDL-C], HDL-C, triglycerides [TG], lipoprotein(a) [Lp(a)], and apolipoproteins A-I, A-II, C-II, C-III, and E [ApoA-1, ApoA-II, ApoC-II, ApoC-III, ApoE]) was also performed by the Lipid Research Laboratory of the Hospital of the University of Pennsylvania per standard protocols as described previously [248]. HDL-C and LDL-C levels were measured both directly and after precipitation with phosphotungstic acid (PTA). Lipoproteins were separated by fast protein liquid chromatography on a Superose 6 column as described previously [249].

Blood was also collected in BD Vacutainer CPT Cell Preparation tubes (BD, Franklin Lakes, NJ, USA) for PBMC isolation and storage and in sodium citrate tubes for platelet isolation and testing. Platelet aggregation studies (PAS) were performed at the Translational Core Laboratory (TCL) using a photometric aggregometer (Biodata Corp, Horsham, PA).

To induce platelet aggregation the following stimuli were used at the concentrations reported in brackets: arachidonic acid (200, 250, 300, 400, 500 mg/ml), collagen (0.04, 0.08, 0.12, 0.16, 0.2 mg/ml), ADP (0.625, 1.25, 2.5, 5, 10 mM), TRAP (1, 2, 3, 4, 5 μ M). Arachidonic acid, collagen and ADP were obtained from Biodata Corp.

(Horsham, PA) while TRAP was obtained from Sigma Chemical Corp (St. Louis, MO). To correct for possibly unreported intake of drugs with antiaggregant action, subjects whose platelets failed to reach 95% aggregation with 500 mg/ml of arachidonic acid were excluded from the analysis. Additional blood was also drawn and frozen for later batch measurement of adrenocorticotrophic hormone (ACTH), cortisol, estradiol, progesterone, luteinizing hormone (LH), follicle-stimulating hormone (FSH), and testosterone also by the TCL using standard radioimmunoassay methods ("Coat-A-Count") from SIEMENS Healthcare Diagnostics. Subjects taking exogenous steroids were excluded from the analysis. Study subjects were also given the option of performing a 24-hour urine collection according to standard methods the day before their visit or on a later date. Collected urine was frozen in single-use aliquots for batch measurement of cortisol by the Translational Core Laboratory as per standard protocols. The acquisition of carotid intima-media thickness (IMT) ultrasound images was performed according to a standardized protocol, adopted from the Atherosclerosis Risk in Communities (ARIC) study[250] and as per American Society of Echocardiography and Society for Vascular Medicine Guidelines for IMT analysis [251]. The scanner used was a Siemens Sequoia (Mountain View, CA, USA) with a 9 Linear probe and a custom designed carotid IMT preset. One heterozygous subject who was unable to travel to the study site had blood drawn locally. These samples were shipped to the study site for the lipid panel and autoimmune tests; the subject also sent the results of a recent comprehensive metabolic panel and complete blood count. Data from clinical phenotyping studies were collected and managed using REDCap (Research Electronic Data Capture) electronic data capture tools hosted at the Perelman School of Medicine at the University of Pennsylvania [252].

Platelet separation and cholesterol measurement

Approximately 4.5 mL of blood was drawn from each subject into tubes containing 25 g sodium citrate, 8 g citric acid and 500 mL H₂O₂. Tubes were centrifuged using a tabletop centrifuge at 200xg for 15 min. at room temperature. Platelet-rich plasma (PRP) was collected as supernatant from this spin and was centrifuged at 900xg for 5 min. The pellet from this spin was resuspended in 8 mL 1xPBS and centrifuged again at 900xg for 5 min. Platelet pellets were collected and resuspended in 200 µL 1xPBS and stored at -20 °C. For measurement of cholesterol, platelet extracts were dried using a centrifugal evaporator (Genevac). D7-cholesterol (Cat. No. 700041; Avanti Polar Lipids, Inc.) was added to each sample as an internal standard. Lipids were extracted by addition of 1 mL chloroform:methanol (2:1) at 4 °C for 2 hours. The nonpolar phase from the platelet extracts was collected and dried through centrifugal evaporation along with cholesterol standards (1.0-100.0 nmol) containing equivalent amounts of D7-cholesterol internal standard. All samples were derivatized using pentafluorobenzoyl chloride and extracted with petroleum ether before measurement by gas chromatography-mass spectrometry using negative chemical ionization. Peak areas for cholesterol and D7-cholesterol from platelet extracts were compared to that from the standard curve generated from cholesterol standards to yield cholesterol content in each sample. For platelet protein measurement, aqueous extracts from chloroform:methanol extraction of platelet samples were spun to remove excess methanol and dried using centrifugal evaporation and then resuspended in 200 µL RIPA buffer containing 0.1 N NaOH and heated at 55°C for approximately 24 hrs with periodic vortexing. Protein concentrations from these extracts were measured by a Lowry assay[240] (Pierce).

Cholesterol measurements for each sample were then normalized to the corresponding protein concentrations.

Fractionation of plasma lipoproteins from study subjects

VLDL, LDL, and HDL subfractions were isolated from EDTA plasma by single step, isopycnic non-denaturing density gradient ultracentrifugation based on a modification of the method developed by Chapman et al. as previously described[253,254]. Using this procedure, five subfractions of LDL (LDL₁, d 1.019-1.023 g/ml; LDL₂, d 1.023-1.029 g/ml; LDL₃, d 1.029-1.039 g/mL; LDL₄, d 1.039-1.050 g/ml; and LDL₅, d 1.050-1.063 g/ml) were isolated, followed by two subfractions of large, light HDL₂ (HDL_{2b}, d 1.063–1.087 g/ml and HDL_{2a}, d 1.088–1.110 g/ml), and three subfractions of small, dense HDL₃ (HDL_{3a}, d 1.110–1.129 g/ml; HDL_{3b}, d 1.129–1.154 g/ml; and HDL_{3c}, d 1.154–1.170 g/ml). Total HDL was prepared as a mixture of HDL_{2b}, 2a, 3a, 3b and 3c subfractions at their equivalent plasma concentrations.

Protein quantification and chemical composition of HDL

Overall chemical composition of HDL subfractions was assessed using commercially available enzymatic kits (total protein: Thermo Scientific, Villebon-sur-Yvette, France; total cholesterol, free cholesterol, phospholipids: DiaSys, Holzheim, Germany; triglycerides: Biomérieux, Marcy l'Etoile, France). Cholesteryl ester (CE) concentration was calculated by multiplying the difference between total and free cholesterol concentrations by 1.67. Total lipoprotein mass was calculated as the sum of total protein, CE, FC, PL and TG and expressed as plasma concentrations (mg/dl). HDL apolipoproteins (ApoA-I, ApoA-II, ApoC-III) were quantified using automated enzymatic

methods (Konelab, Thermo Scientific, Waltham, MA, USA).

Cellular cholesterol efflux capacity of HDL

The cholesterol efflux capacity of total HDL was assessed in a human THP-1 monocytic cell line (ATCC, Manassas, VA, USA) as previously reported[255]. THP-1 monocytes were cultured at 37°C and 5% CO₂ in RPMI 1640 media with 10% FBS, 2 mM glutamine, 100 µg/ml penicillin, and 100 µg/ml streptomycin (complete media) using standard cell culture procedures and differentiated into macrophage-like cells using 50 ng/ml phorbol 12-myristate 13-acetate (PMA). Cells were loaded for 24 h with [³H] cholesterol-labeled acetylated LDL (acLDL, 1 µCi/mL) in serum-free RPMI 1640 culture medium supplemented with 50 mM glucose, 2 mM glutamine, 0.2% BSA, 100 µg/ml penicillin and 100 µg/ml streptomycin (serum-free media). After equilibration in serum-free media, cells were incubated for 4 h with HDL from subjects. Efflux capacity was normalized to HDL-phospholipid content because PL has been shown to represent the key component in determining cholesterol efflux capacity of HDL (15 µg/ml HDL-PL in serum-free media, total volume 250 µL)[256]. Cholesterol efflux capacity was measured as the percent of radioactivity counts in the media over counts in the cell lysate, after adjustment for non-specific diffusion to non-HDL containing media.

Statistical analysis

Data analysis was conducted using Excel (Microsoft Corp.) and GraphPad Prism 6.0 (GraphPad Software Inc.). *In vitro* and *in vivo* data were compared by Student's unpaired T-test. A p-value of less than 0.05 was considered statistically significant. The same test was employed to compare HDL subclasses and composition between controls

and heterozygotes. For data regarding clinical samples, Chi-square testing (for categorical variables) and one-way ANOVA (for continuous variables) were used to determine the effect of *SCARB1* P376L carrier status on a number of variables. A P value less than 0.05 was considered statistically significant. Tukey's multiple comparison test ($\alpha=0.05$) was used following ANOVA where appropriate to determine which groups were responsible for statistically significant differences among groups. Data are reported as mean \pm SD. The association of P376L and CHD was tested among CHD cases and controls of European ancestry assembled from the MICAD Exome Consortium and the CHD Exome+ Consortium. The summary OR of CHD for P376L carriers was calculated using a Mantel-Haenszel fixed-effects meta-analysis without continuity correction, a method that is well suited to low (and even zero) counts and resultant ORs. The association of the P376L variant with HDL-C, LDL-C and TG in the Global Lipids Genetics Consortium Exome Chip Genotyping study was assessed through meta-analysis of single variant score statistics in up to 301,025 individuals [257].

Results

Identification of SCARB1 P376L Homozygote and Association with High HDL-C

We hypothesized that humans with extremely high levels of HDL-C may harbor loss-of-function variants in *SCARB1* and undertook a targeted resequencing discovery experiment in 328 participants with very high HDL-C (> 95th percentile, mean HDL-C 106.8 mg/dL) and a control group of 398 subjects with low HDL-C < 25th percentile, mean HDL-C 30.4 mg/dL). In this cohort, we sequenced the exons of ~990 genes located within 300 kb of each of the 95 loci with significant associations ($P < 5 \times 10^{-8}$) with plasma lipid levels identified by the Global Lipids Genetics Consortium as of 2010 [86]. Among the high HDL-C subjects, we identified a homozygote for *SCARB1* P376L (g.125284671 G>A, c.1127C>T, p.P376L, rs74830677), a 67-year-old female with an HDL-C of 152 mg/dL, and confirmed this finding by Sanger sequencing. This subject harbored no mutations in other high HDL-C genes such as *CETP* and *LIPG*. In addition to this homozygote, four P376L heterozygotes were identified by targeted sequencing in the high HDL-C group; no heterozygotes were found in the low HDL-C group ($P=0.008$, Fisher exact test).

To identify additional P376L carriers, we genotyped an expanded cohort of very high vs. low HDL-C subjects. Among 524 additional subjects with very high HDL-C (mean HDL-C 95.0 mg/dL) we identified 11 heterozygotes for P376L; whereas among 758 subjects with low HDL-C (mean HDL-C 33.5 mg/dL) we identified 3 heterozygotes. In total, our combined sequencing and genotyping for discovery of the P376L variant showed that this variant is significantly over-represented in subjects with high HDL-C

[minor allele frequency (MAF) = 0.010 in high HDL-C vs. 0.0013 in low HDL-C controls, $P=0.000127$, Fisher exact test, **Table 2.1**].

Because this variant is present on the Exome Array, we expanded our analysis to the Global Lipid Genetics Consortium Exome Array data in > 300,000 individuals. The P376L variant was very rare in this population (MAF ~ 0.0003). It was significantly associated with higher HDL-C levels with a relatively large effect size (beta = 8.4 mg/dL; $p = 1.4 \times 10^{-15}$). Notably, this variant was not associated with plasma levels of low density lipoprotein-cholesterol (LDL-C) or triglycerides (TGs) (**Table 2.2**). Thus, we conclude that *SCARB1* P376L is associated specifically with elevated HDL-C levels.

Conservation of SR-BI Proline 376 Across Species and Paralogues

The proline at position 376 of SR-BI is highly conserved down through *Drosophila* and *Anopheles*, as well as in the closely related human genes CD36 and LIMP2 and is in a sequence of highly invariant residues (**Figure 2.1**). The “consensus deleteriousness” score according to Condel was 0.906 (on a scale from 0 to 1, with 1 indicating maximum deleteriousness). Raptor X, a secondary structure prediction program, predicts that substitution of proline 376 with leucine increases the probability of this region to be in a beta-sheet confirmation from 36% to 61%. This finding is in agreement with the suggestion from the recently reported structure of another scavenger receptor LIMP-2 [258] that the region containing P376 in SR-BI is a disordered linker sequence joining beta strands 15 and 16 on the extracellular loop proximal to the C-terminal transmembrane domain.

SCARB1 P376L results in complete loss-of-function of SR-BI through abolishing post-translational processing and cell-surface localization

Given the profound HDL phenotype of the P376L carriers, we sought to understand the impact of the variant on SR-BI function. We first tested the functional impact of the P376L variant on HDL-CE selective uptake in transfected COS7 cells. Studies with transfected COS7 cells revealed that the P376L variant was defective in promoting selective CE uptake (**Figure 2.2 A**) despite similar protein expression in total cell lysates (**Figure 2.2 B**). Notably, the P376L variant had less activity in this assay than one previously reported P297S variant [224] (**Figure 2.2 A**). In studies of ¹²⁵I-labeled HDL₃ binding at 4 °C, the P376L variant abrogated cell-surface binding (**Figure 2.2 C**), despite equal SR-BI protein levels in total cell lysates among groups (**Figure 2.2 D**).

We next generated induced pluripotent stem cells (iPSCs) using peripheral blood mononuclear cells from the P376L homozygote and a noncarrier control and differentiated these cells into hepatocyte-like cells (HLCs) to study HDL metabolism in the setting of endogenous cellular *SCARB1* expression. HLCs differentiated through this protocol recapitulate phenotypes of cultured primary hepatocytes such as albumin and VLDL secretion [259-262]. The cell lines from the control donor and the P376L homozygous subject demonstrated expression of hepatocyte-specific genes *ALB* (albumin) and *AFP* (alpha-fetoprotein) and exhibited comparable *SCARB1* gene expression (**Figure 2.3**). Compared to control iPSC-hepatocyte lines, those from the P376L homozygote demonstrated a profound reduction in selective cholesterol uptake from HDL *in vitro* (**Figure 2.4 A**).

To evaluate the physiological impact of the P376L variant on HDL-C levels and catabolism *in vivo*, we used adeno-associated virus (AAV) vectors to direct hepatic overexpression of WT SR-BI or the P376L variant in *Scarb1* KO mice. The two groups of mice demonstrated similar hepatic expression levels of *SCARB1* mRNA (**Figure 2.5 A**) and SR-BI protein (**Figure 2.5 B**). Mice expressing WT *SCARB1* demonstrated a robust 73% decrease in HDL-C. In contrast, mice expressing the P376L variant had no reduction in HDL-C; their HDL-C levels were comparable to those in the control AAV-null injected mice (**Figure 2.4 B**). While the clearance of ¹²⁵I-HDL protein was not different among the three groups, the clearance of ³H-HDL-CE was much slower in mice expressing the P376L variant compared with those expressing WT SR-BI and was comparable to that in the control mice (**Figure 2.4 C-D**). Selective HDL-CE clearance from plasma was increased by WT SR-BI but was undetectable in the P376L-expressing mice (**Figure 2.4 E & Figure 5 C**), as was hepatic uptake of ³H-CE at 24 hours (**Figure 2.5 D**). This indicates that the P376L sequence variant results in complete loss of the canonical function of SR-BI--namely, selective uptake of HDL cholesteryl ester.

We hypothesized that the markedly reduced HDL-CE uptake could be due to aberrant processing of the P376L SR-BI protein, leading to impaired cell-surface localization. To test this, we isolated cell-surface proteins from COS7 cells transfected with WT and P376L SR-BI using biotinylation and found markedly reduced cell surface SR-BI in the P376L transfected cell lysates after streptavidin cell-surface protein pull-down (**Figure 2.2 E**). Given that SR-BI undergoes N-glycosylation in the endoplasmic reticulum concomitant with proper folding, we hypothesized that altered post-translational modification may underlie its reduced cell-surface localization [256,263,264]. We measured the molecular weights of SR-BI forms after

endoglycosidase-H (Endo-H) treatment of transfected COS7 (**Figure 2.2 F**) and iPSC-derived HLC lysates as well as mouse liver lysates expressing WT or mutant SR-BI (**Figure 2.4 F-G**). Higher molecular weight forms represent N-glycosylation modified Endo-H resistant and partially sensitive forms at the cell surface after modification by alpha-mannosidase II in the Golgi apparatus [263]. In the iPSC-derived differentiated hepatocyte-like cells from the P376L homozygote (**Figure 2.4 F**), we found much less total cellular SR-BI in the mutant cell lines relative to that of WT cells, despite comparable *SCARB1* gene expression (**Figure 2.3 C**). After Endo-H treatment, the SR-BI from *SCARB1* WT cell and liver lysates across models was predominantly the partially sensitive form, along with small amounts of the fully resistant form. In contrast, the SR-BI from cell and tissue lysates across P376L-expressing groups was all the immature, fully Endo-H sensitive form (**Figure 2.2 F & Figure 2.4 F-G**). Together these data are consistent with a model in which the P376L sequence variant alters the endogenous post-translational N-glycosylation of SR-BI to prevent either transit from the ER to the Golgi or further post-translational modifications in the Golgi, ultimately resulting in reduced cell surface expression.

HDL-related phenotypes of SCARB1 P376L homozygote and heterozygotes

We next recruited the P376L homozygote, eight heterozygous carriers, and both high HDL-C and normal HDL-C non-carrier controls for deep phenotyping of HDL metabolism and related traits. All of the P376L study participants were of European ancestry, almost exclusively of Ashkenazi Jewish descent. Clinical characteristics and lipid profiles of the subjects are reported in **Table 2.3**. FPLC (fast protein liquid chromatography) analysis of plasma lipoproteins confirmed the increase in large HDL

particles in the homozygote (**Figure 2.6 A**). Cholesterol and apoA-I levels in HDL were significantly increased in the homozygote and heterozygotes compared with controls, but HDL apoA-II levels were not elevated (**Table 2.3 & Figure 2.6 B**). There were no differences between P376L carriers and controls in the absolute amount of HDL free cholesterol or the ratio of free-to-esterified cholesterol in their HDL (**Figure 2.6 C**). P376L heterozygotes had a 2.8-fold increase and the homozygote a 6.1-fold increase in large HDL-2b particles compared with noncarrier controls (**Figure 2.6 D**). There was a higher amount of apoA-I (**Figure 2.6 E**) and apoC-III (**Figure 2.6 F**) in large HDL particles in the homozygote and heterozygous carriers. Cholesterol efflux capacity was similar in carriers and controls (**Figure 2.6 G**).

Previous *in vitro*, mouse and human genetics studies have suggested that SR-BI in platelets is necessary for proper platelet activity and thrombosis [208,224,265-267]. To test the effects of the P376L variant on platelet activity, we isolated platelets from carriers and controls and performed light transmission aggregometry (LTA) after stimulation with known platelet activators arachidonic acid, collagen, ADP and TRAP over a range of doses. We found only a slight decrease in ADP-induced maximal aggregation in platelets isolated from the P376L homozygote relative to heterozygotes and control subjects at a dose of 5 μ M (**Figure 2.7 A**). No differences in platelet activity in response to other stimulants were observed between the groups. We also extracted lipids from platelets and measured platelet cholesterol content among groups. We observed that platelet cholesterol increased in a genotype dose-dependent manner from controls (mean 122 nmol/mg protein) to heterozygotes (mean 139 nmol/mg protein) to the homozygote subject (244 nmol/mg protein) (**Figure 2.7 B**). However, the difference between normal HDL-C controls and heterozygotes was not significant, and

these differences were reduced when the values normalized to plasma total cholesterol levels, suggesting that elevated platelet cholesterol in carriers reflects increased plasma HDL-C levels rather than a platelet SR-BI specific function (**Figure 2.7 C**). There was also no difference in total circulating platelet levels among groups (data not shown).

SR-BI also takes up HDL-cholesteryl esters in adrenal glands and reproductive tissues for steroid hormone production in mice and humans [208,220,268], so we evaluated the impact of SCARB1 loss-of-function on steroid hormones in our recruited participants. We found no difference in morning serum cortisol, ACTH and 24-hr urinary cortisol-to-creatinine (**Figure 2.8**) across participants, moderately higher testosterone in male P376L heterozygotes relative to normal HDL-C controls, but no differences across groups in FSH and LH.

SCARB1 P376L is associated with increased risk of CHD in humans

Despite a profound increase in HDL-C, SR-BI deficiency in mice causes accelerated atherosclerosis [219-222]. The relationship of reduced SR-BI function to atherosclerotic cardiovascular disease in humans has not been established. The P376L homozygous subject did not have clinical CHD, but her carotid intimal-medial thickness (cIMT) was 0.789 mm (left-right average), which is > 75th percentile for females of her age; in addition, she had detectable plaque throughout the left internal carotid artery and at the bifurcation of her right internal carotid artery. cIMT measurements were not significantly different in the P376L heterozygotes compared with both groups of controls (**Figure 2.9**), but due to small sample size the statistical power is limited.

To achieve greater statistical power to address this question, we performed a meta-analysis of large Exome Array genotyping studies of CHD cases and healthy

controls to determine the relationship of the P376L variant with risk of CHD (**Table 2.4**). Among 16 sample sets from two consortia (The MICAD Exome Consortium and the CHD Exome+ Consortium), we tested the association between P376L carrier status and CHD in 137,995 individuals. Across 49,846 CHD cases and 88,149 CHD controls, we found that P376L carriers had significantly higher risk of CHD compared with non-carriers (Odds Ratio for disease among carriers = 1.79; P = 0.018, **Table 2.4**) Thus, carriers of this *SCARB1* P376L variant have significantly increased HDL-C levels and significantly increased risk of CHD.

Discussion

Studies of mice have provided important insights into the effects of SR-BI on HDL metabolism, reverse cholesterol transport (RCT), and atherosclerosis. These studies revealed that overexpression of SR-BI reduces HDL-C [209-212] and reduces [216-218], whereas gene deletion of SR-BI increases HDL-C [213-215] and accelerates atherosclerosis [219-222]. The clinical relevance of these findings has remained uncertain, however. Studies of injected labeled HDL-CE in humans suggested that the majority of the HDL-CE was transported to the liver via CETP-mediated exchange to apoB-containing lipoproteins rather than via direct uptake from HDL by the liver [269], bringing into question the importance of hepatic SR-BI in human physiology. Common genetic variants near the *SCARB1* locus were found to be significantly associated with plasma HDL-C levels, suggesting that SR-BI may play a role in HDL metabolism in humans [86,117]. A family with a rare *SCARB1* variant P297S was reported in which the heterozygous carriers of the variant had modestly elevated HDL-C levels [224]. However, the variant retains substantial SR-BI activity, no homozygotes were identified,

the apparent effect on HDL-C was modest, and there was insufficient power to address its effects on atherosclerosis.

Through sequencing of subjects with extremely high plasma levels of HDL-C, we identified a homozygote for a P376L variant in SR-BI. Our complementary approaches consistently demonstrated that this variant confers virtually complete loss-of-function of SR-BI. Our results demonstrate many similarities in the consequences of SR-BI deficiency on HDL composition between mice and humans, including a shift toward large, buoyant HDL particles and a significant increase in apoA-I but not apoA-II in plasma and HDL [214,270,271]. The homozygote is a woman who had two healthy children without fertility issues or delivery complications, suggesting that in humans, SR-BI deficiency may not impair reproductive function in the same manner as it does in mice [220,272]. In mice, SR-BI-mediated CE uptake from HDL is a critical process underlying steroid hormone synthesis in adrenal and gonadal tissues, and SR-BI deficiency alters adrenal cholesterol content, impairs adrenal glucocorticoid response under stress, and can lead to fasting-induced hypoglycemia [208,273,274]. We did not observe any differences in fasting glucose, serum cortisol, ACTH, or female gonadal hormones in P376L heterozygous subjects versus controls and we saw only a modest increase in testosterone in male P376L heterozygotes relative to noncarriers. We postulate that differences in expression or capacity for upregulation of apoB-containing lipoprotein receptors relative to SR-BI between mouse models and humans in steroidogenic tissues may account at least partially for the lack of recapitulation of some of the phenotypes of SR-BI deficiency in mice. We also observed no differences in platelet levels, cholesterol content, and activation from the P376L carriers, despite reports of thrombocytopenia and altered platelet activity in *Scarb1* KO mice [224]. These results suggest a relatively

different contribution of SR-BI to platelet function between mice and humans. Interestingly, the phenotypes of human *SCARB1* P376L homozygote (elevated HDL-C and large HDL particles but relatively normal steroidogenesis, reproductive viability, and platelet function) are comparable to those observed in mice lacking the PDZ domain containing 1 (PDZK1), an adaptor protein for SR-BI [275].

Perhaps the most important finding of our study is that despite the elevation in HDL-C, P376L carriers exhibit increased risk of CHD, consistent with the results in *Scarb1* KO mice. Our results are consistent with a growing theme in HDL biology that steady-state concentrations of HDL-C are not causally protective against CHD and that HDL function and cholesterol flux may be more important than absolute levels. Using an in vivo assay of macrophage RCT, we previously showed that *Scarb1* KO mice have impaired macrophage RCT even though they have elevated HDL-C levels [223]. Our results suggest that reduced hepatic SR-BI function in humans causes impaired RCT, leading to increased risk of CHD despite elevation in HDL-C levels. However, SR-BI is also expressed in vascular cell types, including endothelial cells, vascular smooth muscle cells, and macrophages, where it could have protective effects against atherosclerosis as well [276,277]. Our results are consistent with the previously suggested concept [277] that upregulation or enhancement of SR-BI could be a novel therapeutic approach to reducing CHD risk in the general population.

Figure 2.1

<i>Homo sapiens</i>	L	N	A	D	P	V	L	A	E	A	V	T	G	L	H	P	N	Q	E	A	-	H	S	L	F	L	D	I	H	P	V	T	G	I	P	M	N	C	S	V	K	L	Q	L	S	L	Y	M	K	S	V	A	G	I	G	Q	T	G	K	I
<i>Pan troglodytes</i>	L	N	A	D	P	V	L	A	E	A	V	T	G	L	H	P	N	Q	E	A	-	H	S	L	F	L	D	I	H	P	V	T	G	I	P	M	N	C	S	V	K	L	Q	L	S	L	Y	M	K	S	V	A	G	I	G	Q	T	G	K	I
<i>Macaca mulatta</i>	L	N	A	D	P	V	L	A	E	A	V	T	G	L	H	P	N	Q	E	A	-	H	S	L	F	L	D	I	H	P	V	T	G	I	P	M	N	C	S	V	K	L	Q	L	S	L	Y	M	K	S	V	A	G	I	G	Q	T	G	K	I
<i>Canis lupus familiaris</i>	Y	N	A	D	P	M	L	A	E	A	V	L	G	L	H	P	N	Q	E	E	-	H	S	L	F	L	D	I	H	P	V	T	G	I	P	M	N	C	S	V	K	L	Q	L	S	L	Y	I	K	A	I	K	G	I	G	Q	T	G	K	I
<i>Bos taurus</i>	Y	N	A	D	P	V	L	A	E	A	V	S	G	L	H	P	N	P	K	E	-	H	S	L	F	L	D	I	H	P	V	T	G	I	P	M	N	C	S	V	K	L	Q	L	S	L	F	V	K	S	V	K	G	I	G	Q	T	G	N	I
<i>Mus musculus</i>	Y	N	A	D	P	V	L	S	E	A	V	L	G	L	N	P	N	P	K	E	-	H	S	L	F	L	D	I	H	P	V	T	G	I	P	M	N	C	S	V	K	M	Q	L	S	L	Y	I	K	S	V	K	G	I	G	Q	T	G	K	I
<i>Rattus norvegicus</i>	Y	N	A	D	P	V	L	S	E	A	V	L	G	L	N	P	D	P	K	E	-	H	S	L	F	L	D	I	H	P	V	T	G	I	P	M	N	C	S	V	K	M	Q	L	S	L	Y	I	K	S	V	K	G	V	G	Q	T	G	K	I
<i>Gallus gallus</i>	Y	N	A	D	P	S	L	V	D	A	V	E	G	L	H	P	S	R	E	E	-	H	A	L	F	L	D	V	H	P	V	T	G	I	P	M	N	C	S	I	K	L	Q	L	N	Q	Y	M	K	Q	V	S	G	I	L	Q	T	G	K	I
<i>Salmo salar</i>	Y	N	A	D	P	V	L	L	D	Y	V	Q	G	L	Q	P	T	E	D	E	-	H	G	L	F	I	D	I	H	P	E	T	G	V	F	L	N	V	S	I	R	L	Q	L	N	L	Y	M	K	K	V	S	G	I	T	E	T	G	K	I
<i>Danio rerio</i>	Y	N	A	D	P	Q	L	L	D	T	V	S	G	L	S	P	S	E	D	E	-	H	G	L	F	I	D	I	H	P	E	T	G	V	F	V	N	V	S	V	R	L	Q	L	N	L	F	L	K	K	V	S	G	I	S	E	T	G	N	I
<i>Drosophila melanogaster</i>	Y	L	A	D	E	S	L	R	T	Q	V	E	G	I	S	P	P	M	K	E	K	H	Q	F	F	D	V	Q	P	K	M	G	T	T	L	R	V	R	A	R	I	Q	I	N	L	A	V	S	Q	V	F	D	I	K	Q	V	A	N	F	
<i>Anopheles gambiae</i>	Y	M	A	D	Q	T	L	R	T	A	V	E	G	I	S	P	P	E	K	D	K	H	Q	L	F	I	D	V	Q	P	D	M	G	T	A	L	R	A	R	A	R	I	Q	I	N	L	A	V	S	Q	V	D	I	K	Q	V	A	N	F	
<i>hCD36</i>	L	Y	A	S	P	D	V	S	E	P	I	D	G	L	N	P	N	E	E	E	-	H	R	T	Y	L	D	I	E	P	I	T	G	F	T	L	Q	F	A	K	R	L	Q	V	N	L	L	V	K	P	S	E	K	I	Q	V	L	K	N	L
<i>hLIMP2</i>	Y	Q	A	D	E	R	F	V	S	A	I	E	G	M	H	P	N	Q	E	D	-	H	E	T	F	V	D	I	N	P	L	T	G	I	I	L	K	A	A	K	R	F	Q	I	N	I	Y	V	K	K	L	D	D	F	V	E	T	G	D	I

Figure 2.1: SR-BI Protein Sequence Alignment Across Species. Amino acid sequence alignment of SR-BI across 12 species and human SR-BI paralogues CD36 and LIMP-2. Shown is the ~60 residue sequence alignment adjacent human SR-BI residue Pro376 (indicated by red box). Dark grey shading indicates full conservation of a given residue across indicated species. Light grey shading indicates a different but conservative amino acid for the given species compared to the others listed. *Contributions: This data was generated by a collaborator Paolo Zanoni and analyzed in part by SAK.*

Figure 2.2

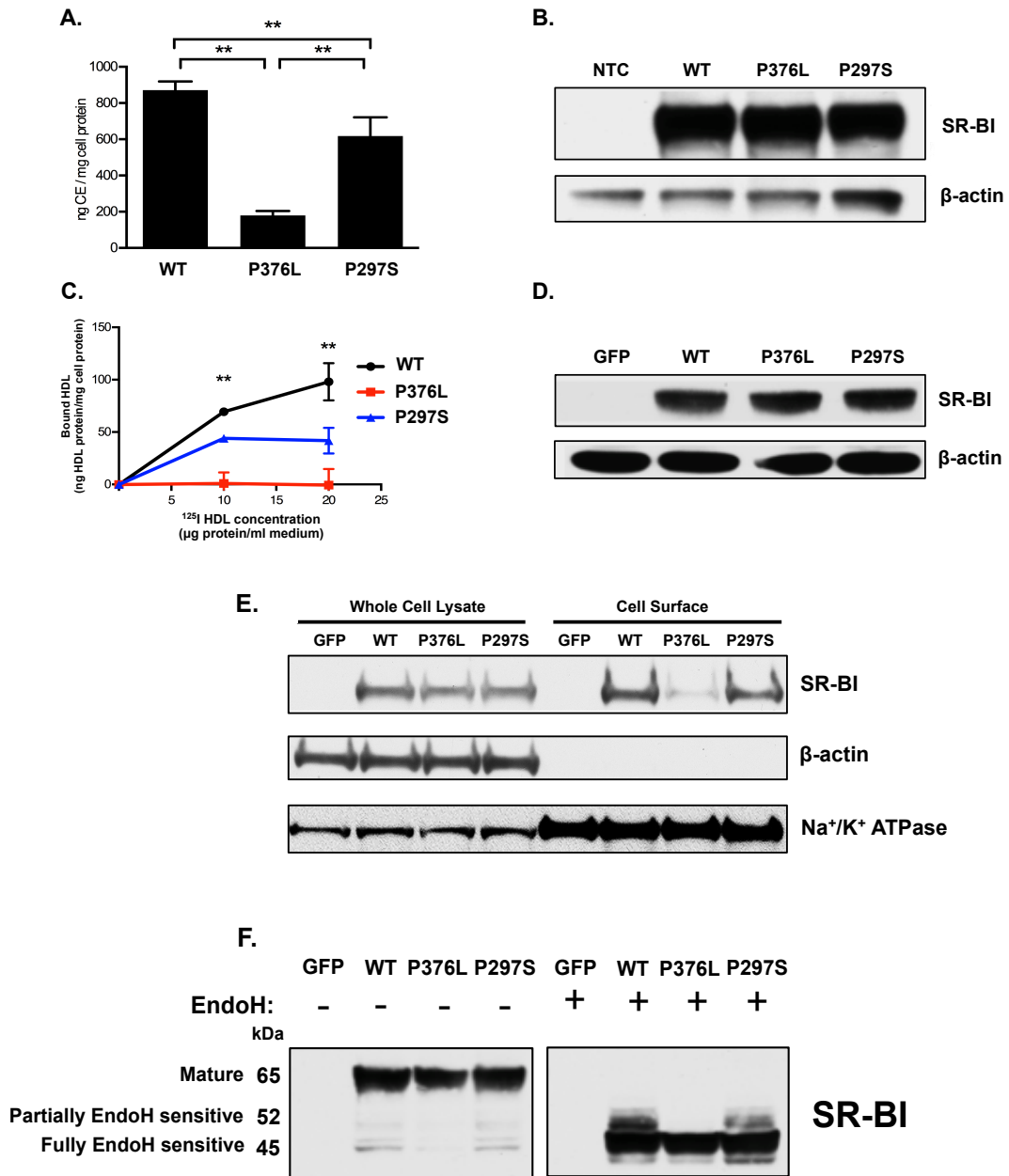


Figure 2.2: *SCARB1* P376L abrogates SR-BI function in transfected COS7 cells. A. Selective cholesterol uptake in COS7 cells expressing SR-BI WT vs. P376L. Cells were transfected with plasmids expressing WT, P376L or P297S forms of SR-BI and

incubated at 37 °C for 3 hrs with $^{125}\text{I}/^3\text{H}$ -labeled HDL₃ to determine HDL cholesterol ester (CE) specific uptake. Data represent the mean of a quintuplicate determination after subtraction of the determinations done with the addition of a 40-fold excess of cold HDL.

B. Western blot showing SR-BI expression levels in whole cell lysates from COS7 cells transfected for selective cholesterol uptake experiment in (A). **C.** Binding of HDL to SR-BI at 4 °C in transfected COS7 cells. Transfected cells were exposed to ^{125}I -labeled HDL₃ for 2 hrs at 4 °C to measure HDL binding. Radioactive counts from cells were then quantified to determine the amount of cell-associated ^{125}I -HDL₃. Data points represent the mean +/- S.D. ** P<0.01, one-way ANOVA. **D.** Western blot showing SR-BI expression levels in whole cell lysates from COS7 cells transfected in panel (C). **E.** Immunoblot of SR-BI after cell-surface biotinylation in transfected COS7 cells. Cells were transfected with GFP, SR-BI WT, P376L or P297S plasmids and biotinylated before collection of whole cell lysate (left) or incubation with NeutrAvidin beads and elution of cell-surface localized proteins (right). Whole cell lysates (lanes 1-4) and cell-surface proteins (lanes 5-8) were separated by SDS-PAGE and immunoblotted for human SR-BI. Actin and Na/K-ATPase were used respectively as intracellular and surface-associated controls. **F.** Endo-H sensitivity of SR-BI from transfected COS7 cells. Cells were transfected with plasmids encoding GFP or different forms of SR-BI (WT, P376L, P297S) and cell lysates were treated with Endo-H to release complex N-linked glycans and molecular forms of SR-BI were monitored by immunoblotting. For A & C, data represent mean Error bars indicate mean values ± S.D. ** P<0.01, Student's unpaired T-test. *Contributions: Data in panels A-D generated by Paolo Zanoni; data in panels E-F generated by SAK and collaborator; all data analyzed by SAK.*

Figure 2.3

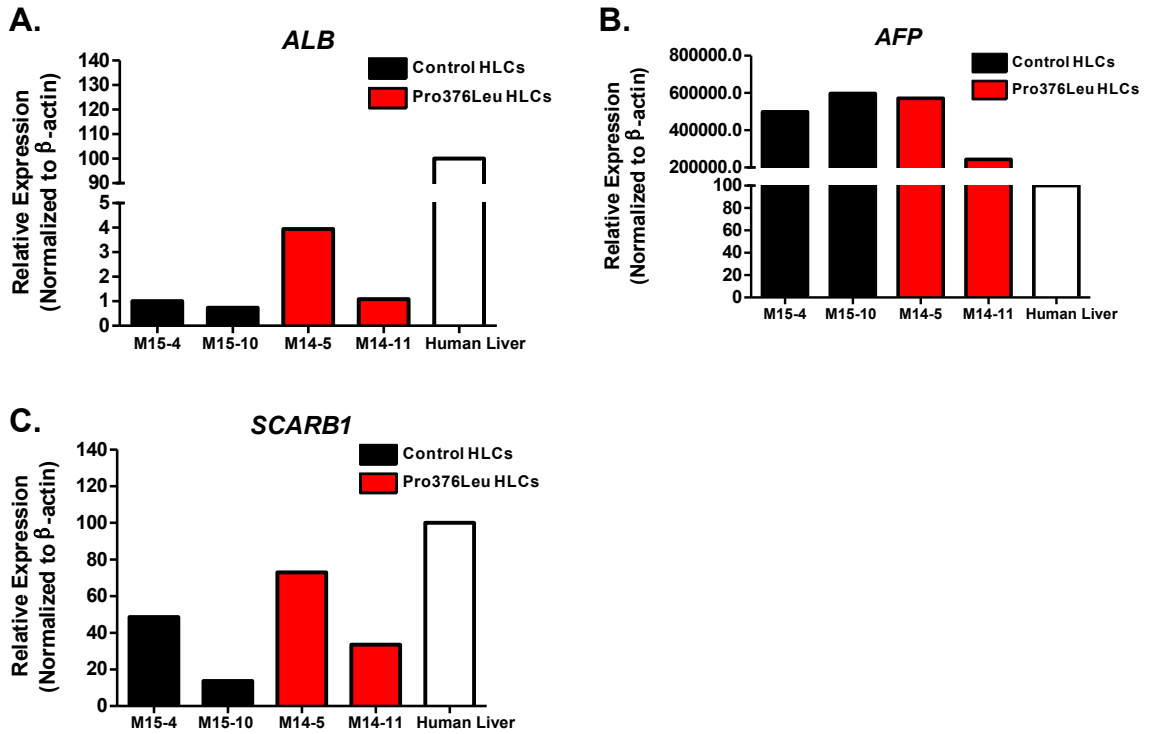


Figure 2.3. Gene expression in control and *SCARB1* P376L iPSC-derived hepatocyte-like cells (HLCs). **A.** *ALB* Gene expression by quantitative RT-PCR of mRNA from control and P376L mutant iPSC-derived HLCs. Cells were differentiated 21-25 days before experiments and RNA isolation for gene expression analysis. **B.** *AFP* gene expression in iPSC-derived HLCs. **C.** *SCARB1* gene expression in iPSC-derived HLCs. *Contributions: All data in this figure was generated and analyzed by SAK.*

Figure 2.4

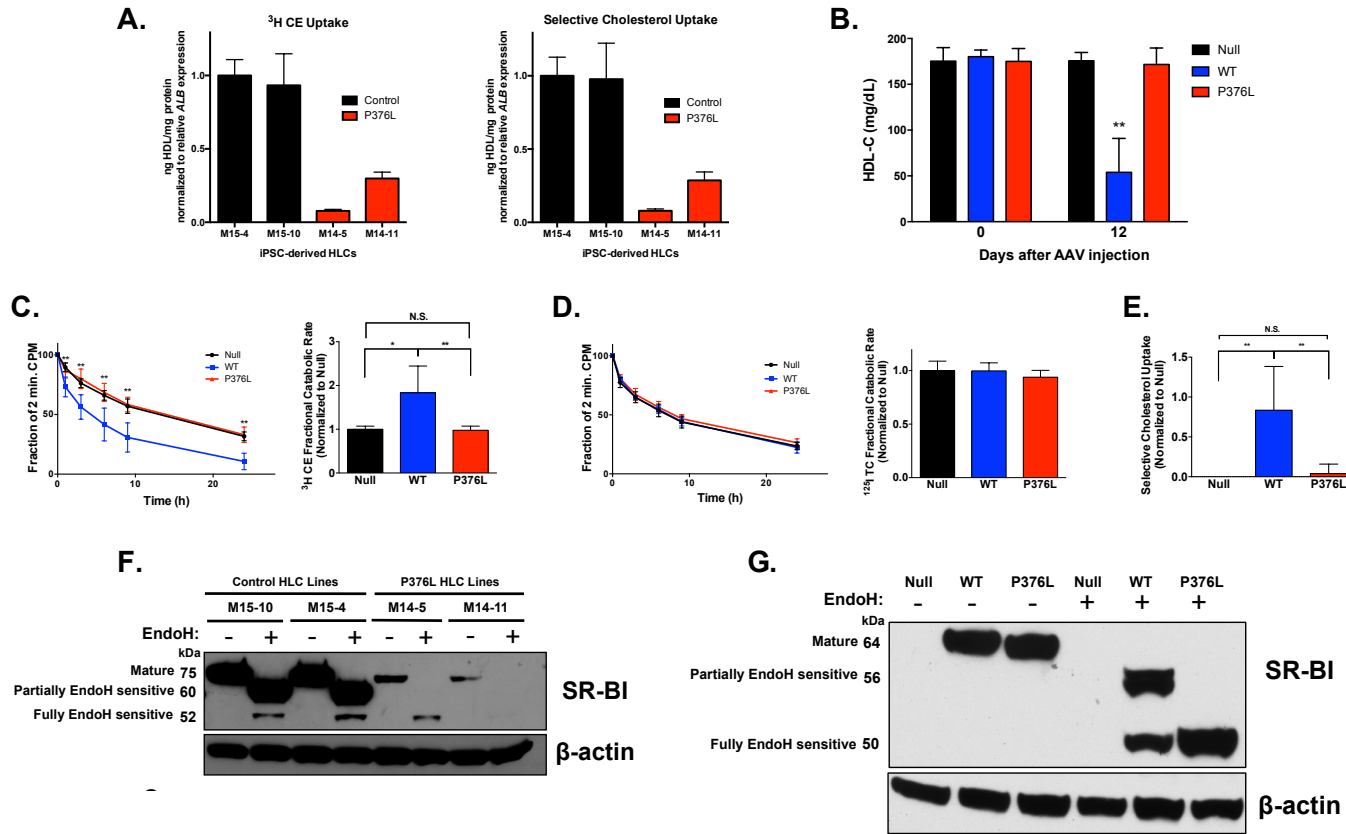


Figure 2.4: *SCARB1* P376L is a null variant *in vitro* and *in vivo*. A. ^3H cholesterol ether (CE) uptake (left panel) and selective cholesterol uptake from HDL (right panel) in iPSC-derived hepatocyte-like cells (HLCs) from the P376L homozygote vs. a noncarrier control. Cells were incubated with ^3H cholesteryl ether (CE)/ ^{125}I tyramine cellobiose (TC) dual-labeled human HDL. All values are

normalized to relative *ALB* gene expression in each cell line. All data represents mean values for wells of respective cell lines \pm S.D.

B. Plasma HDL cholesterol before and 12 days after AAV administration to *Scarb1* knockout mice. **C.** ^3H cholesterol ether (CE) clearance (left panel) and fractional catabolic rate (right panel) from plasma of *Scarb1* knockout mice injected with Null or SR-BI AAVs after administration of ^3H -CE/ ^{125}I -TC dual-labeled human HDL. **D.** ^{125}I tyramine cellobiose (TC) clearance (left panel) and fractional catabolic rate (right panel) from plasma after administration of dual-labeled HDL. **E.** Selective cholesterol uptake in mice expressing Null, SR-BI WT, or P376L measured by relative differences in ^3H and ^{125}I fractional catabolic rates. **F.** Sensitivity to endoglycosidase-H (Endo-H) in P376L homozygous vs. noncarrier iPSC-derived HLCs. Cell lysates of each genotype were treated with Endo-H to remove complex N-linked glycans from mature forms of proteins and then immunoblotted for SR-BI. Molecular weights of different forms of SR-BI after Endo-H treatment are given on the left. **G.** SR-BI Endo-H sensitivity from liver lysates from mice expressing Null, SR-BI WT, or SR-BI P376L AAV. Lysates were treated with Endo-H followed by immunoblotting for SR-BI. Molecular weights of different forms of SR-BI after Endo-H treatment are given on the left. For A, data represent mean values for wells of respective cell lines \pm S.D. For B-E, data represent mean values \pm S.D. for each of the 3 groups. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ANOVA (B, C – plasma clearance), Unpaired T-test (E). *Contributions: All data in this figure generated by SAK; data in panels C-E generated by SAK and Paolo Zanoni; all data analyzed by SAK and collaborator.*

Figure 2.5

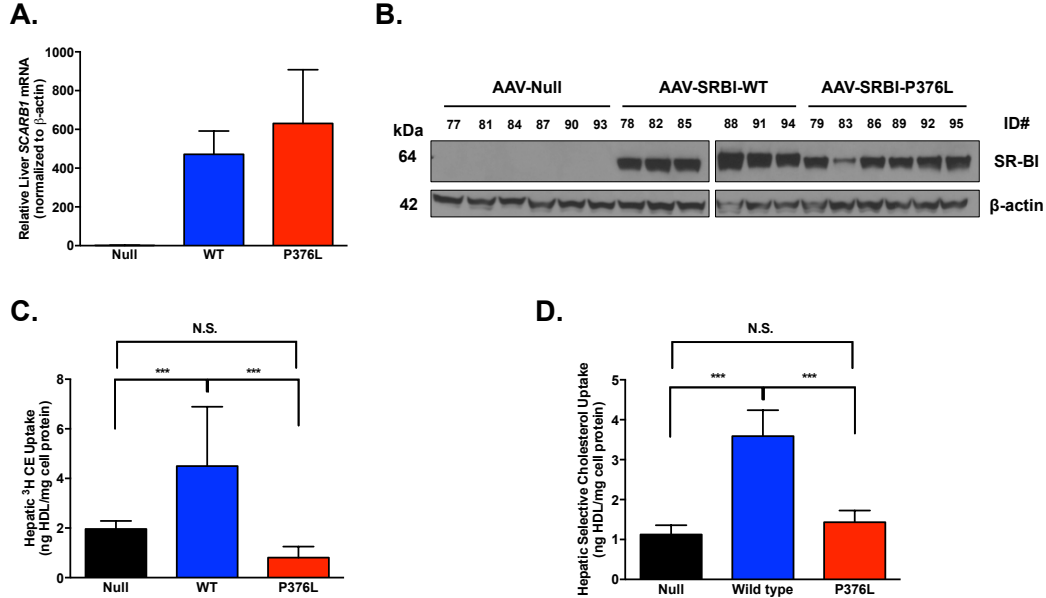


Figure 2.5: Hepatic *SCARB1* expression and impact on selective cholesterol uptake from HDL in mice. **A.** Human *SCARB1* transcript expression levels measured by quantitative RT-PCR from livers of mice expressing Null or SR-BI AAV vectors. Gene expression was measured from total hepatic RNA after reverse transcription and normalized to expression levels of actin. **B.** SR-BI immunoblot (right) from livers of *Scarb1* KO mice expressing Null, SR-BI WT, and SR-BI P376L 2 weeks after AAV administration. **C.** Liver ^3H CE uptake from dual-labeled HDL administration in mice expressing Null or SR-BI AAVs. **D.** Hepatic selective cholesterol uptake measured by relative difference of hepatic ^3H CE and ^{125}I TC uptake in livers of mice expressing Null or SR-BI AAVs after dual-labeled HDL administration. All data represent mean values \pm S.D. for each of the 3 groups. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, Unpaired T-test. *Contributions: Data in panels A, C, and D generated and analyzed by SAK; data in panel B generated by Paolo Zanoni.*

Figure 2.6

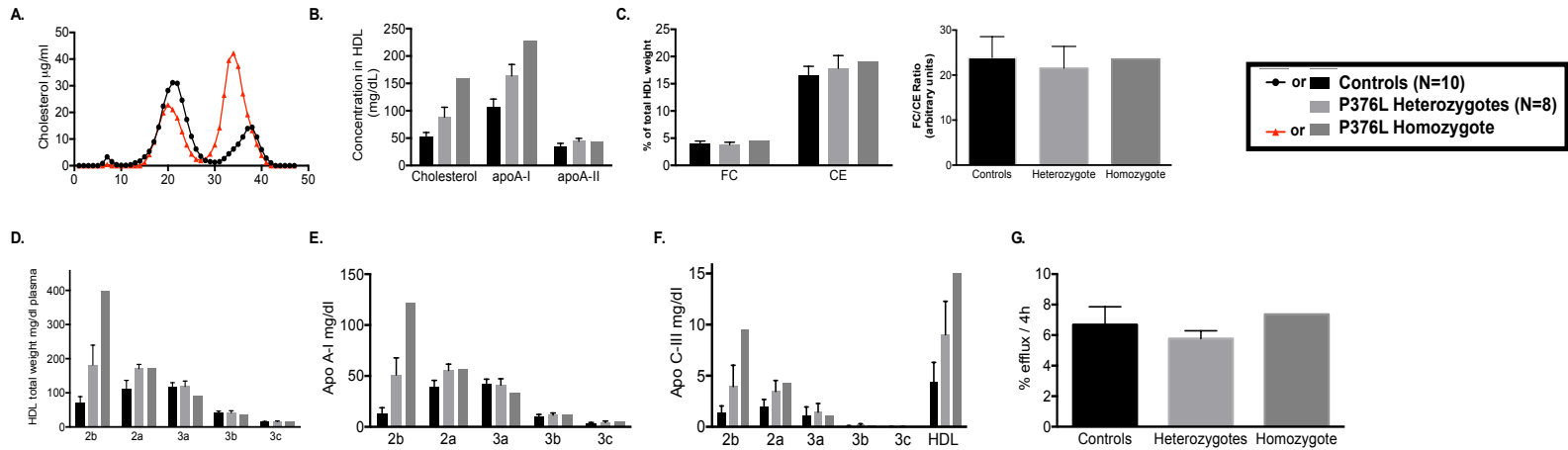


Figure 2.6: HDL composition and functionality in *SCARB1* P376L homozygote, heterozygous carriers, and controls. **A.** FPLC fractionation of plasma lipoproteins from the P376L homozygote subject (red) and from a control with normal HDL-C. **B.** Cholesterol, apoA-I and apoA-II content in total HDL. **C.** Free cholesterol (FC) and esterified cholesterol (CE) in total HDL (left panel) and the FC/CE ratio in total HDL (right panel). **D.** HDL subclass concentrations after separation by density gradient ultracentrifugation. **E.** ApoA-I content in the same HDL subclasses. **F.** ApoC-III content in the same HDL subclasses. **G.** Cholesterol efflux capacity from THP-1 macrophages. All data are reported as mean values \pm S.D. *Contributions: All data generated by collaborators (Paolo Zanoni, Will Hancock-Cerutti, and Anatol Kontush) and analyzed by SAK and collaborators.*

Figure 2.7

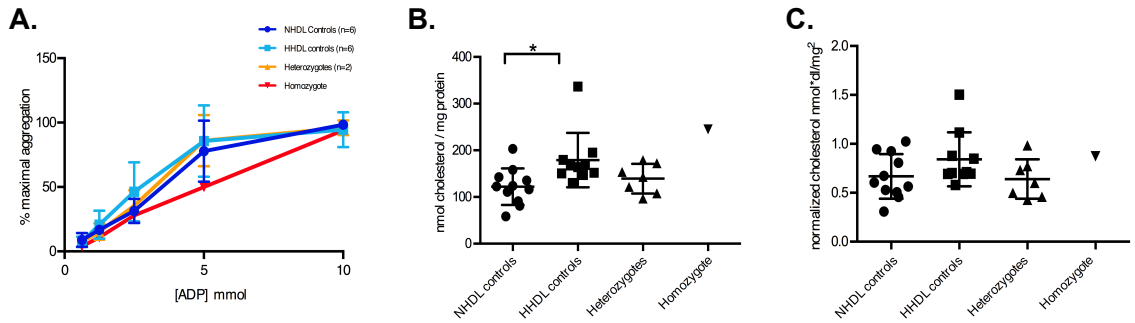


Figure 2.7: Platelet aggregation and cholesterol content. A. Platelet aggregation measured by light transmission aggregometry after stimulation with increasing doses of ADP. Data represent the percentage maximal aggregation. **B.** Platelet cholesterol content measured by LC/MS. **C.** Platelet cholesterol content after normalization for plasma total cholesterol levels. Bars represent mean values \pm S.D.. * $P < 0.05$, one-way ANOVA followed by Tukey's multiple comparisons test. *Contributions: All data generated by core facility and analyzed by SAK and Paolo Zanoni.*

Figure 2.8

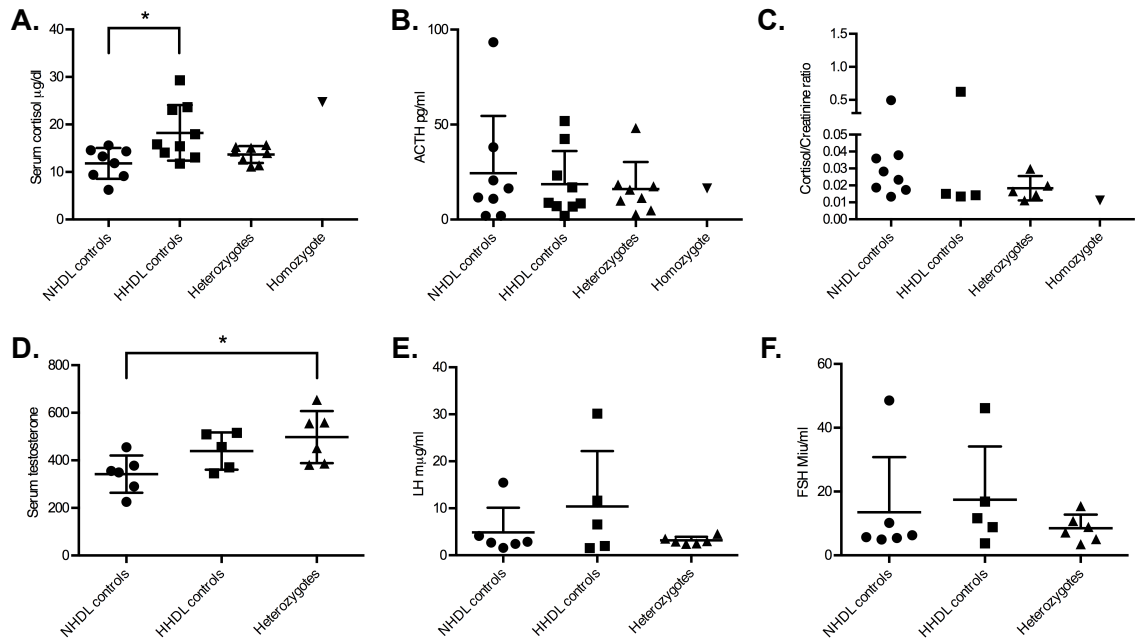


Figure 2.8: Impact of *SCARB1* P376L on adrenal and gonadal steroidogenesis. A.

Morning serum cortisol in carriers vs. controls. **B.** Morning plasma ACTH. **C.** Cortisol /

creatinine ratio in 24 hr urine. **D.** Serum testosterone in males. **E.** Serum LH in males. **F.**

Serum FSH in males. Bars represent mean values ± S.D.. * P<0.05, one-way ANOVA

followed by Tukey's multiple comparisons test. *Contributions: All data generated by core*

facilities and analyzed by SAK and Paolo Zanoni.

Figure 2.9

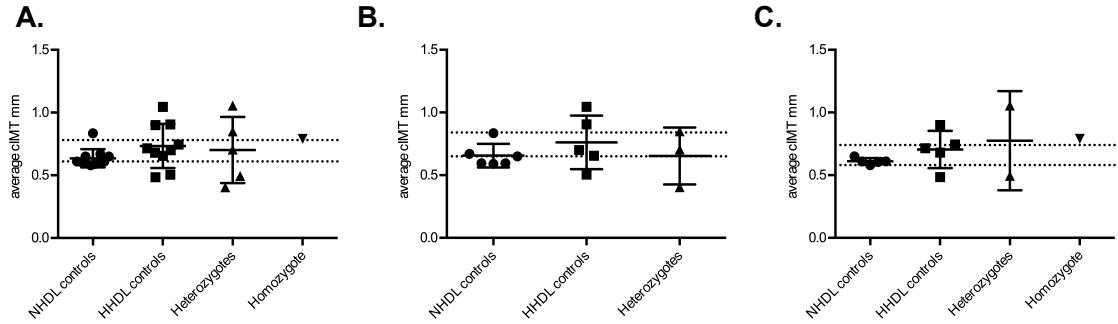


Figure 2.9: Carotid intima media thickness (cIMT) in *SCARB1* P376L carriers vs. controls. A. cIMT for all subjects (male and female combined). **B & C.** cIMT results for males and females, respectively. Dotted lines represent the 25th and the 75th percentile values from the ARIC study. Bars represent mean values \pm S.D.. All data shows the average left / right cIMT. *Contributions: All data generated by clinical laboratories and analyzed by SAK, Paolo Zanoni, and Emil Mohler.*

Table 2.1

Table 2.1: Association of <i>SCARB1</i> P376L with HDL-C in High vs. Low HDL-C Cohorts									
Discovery Cohort	High HDL-C (>95th percentile)				Low HDL-C (<25th percentile)				Association of P376L Carriers with Higher HDL-C (Fisher Exact Test)
	<u>No. of Participants</u>				<u>No. of Participants</u>				
	Total	Noncarriers	Heterozygotes	Homozygotes	Total	Noncarriers	Heterozygotes	Homozygotes	
Targeted Sequencing of <i>SCARB1</i>	328	323	4	1	398	398	0	0	P=0.008398
Exome Array Genotyping	524	513	11	0	758	755	3	0	P=0.005296
Combined	852	836	15	1	1156	1153	3	0	P=0.000127

Table 2.1: Association of *SCARB1* P376L with HDL-C in High vs. Low HDL-C Cohorts. Carriers of the P376L variant were ascertained from the Penn High HDL Study through 2 approaches, targeted sequencing of the *SCARB1* gene in a total of 726 subjects (328 high HDL-C and 398 low HDL-C subjects) and genotyping on the Exome Array (Illumina) in an additional 1282 subjects (524 high HDL-C subjects and 758 low HDL-C subjects). The association of the P376L variant with the high HDL-C cohort from both approaches individually and combined together was tested using a Fisher exact test. *Contributions: Sequencing and genotyping completed by core facilities; data QC and analysis performed by SAK.*

Table 2.2

Table 2.2: Association of SCARB1 P376L with Plasma Lipid Traits in Global Lipids Genetics Consortium Exome Array Genotyping				
Trait	Number of Subject included	Minor Allele Frequency	Beta (SE) in Standard Deviation Units	P-value (Score Test)
HDL-C	301025	0.00033	+0.57 (0.071)	1.41 x 10⁻¹⁵
LDL-C	280551	0.00033	+0.065 (0.074)	0.381
TG	290277	0.00034	-0.052 (0.072)	0.474

Table 2.2: Association of SCARB1 P376L with plasma lipid traits in Global Lipids Genetics Consortium Exome Array genotyping. The relationship between the frequency of P376L carriers and plasma lipid traits was measured in the Global Lipids Genetics Consortium cohort by genotyping of the variant on the Exome Array and using the Score test. *Contributions: Genotyping data and analysis performed by Gina Peloso, Sekar Kathiresan, and Global Lipids Genetics Consortium members.*

Table 2.3

Table 2.3: Characteristics of SCARB1 P376L Carriers and Controls Recruited for Deep Phenotyping						
Measure	1: Normal HDL-C controls	2: High HDL-C controls	3: P376L Heterozygotes	P376L Homozygote	ANOVA/Chi-square	Tukey's multiple comparison
Number of Subjects	11	10	8	1	-	-
Age (years)¹	61.6 (9.7)	64.2 (12.5)	67.5 (15.3)	65	n.s.	-
Sex (M/F)	6/5	5/5	6/2	0/1	n.s.**	-
BMI - Kg/m²	26.4 (2)	22.9 (1.3)	25.6 (3.9)	21	*	1/2 ²
TC (mg/dl)	185.8 (22.3)	215.8 (29.9)	228 (33.2)	280	*	1/3
Glucose (mg/dl)	93.5 (2.9)	91.6 (7.0)	98.8 (5.3)	86	n.s.	-
LDL-C (mg/dl)	109.1 (17.3)	97.4 (21.6)	116.6 (27.1)	109	n.s.	-
HDL-C (PTA) (mg/dl)	51 (11.4)	110.1 (19.8)	86.9 (19.9)	152	*	1/2, 1/3, 2/3
TG (mg/dl)	121.2 (35)	71.5 (32.3)	99.5 (23.7)	57	*	1/2
Alcohol >1/day (n)	4	4	2	0	n.s.**	-
VLDL-C (mg/dl)	26.9 (8.8)	19 (6.2)	23.1 (9.2)	13	n.s.	-
Lp(a) (mg/dl)	22.3 (18.8)	19 (22.7)	15.9 (21.2)	17	n.s.	-
apoA-I (mg/dl)	172.2 (33.3)	241.7 (41.2)	229.6 (36.1)	327	*	1/2, 1/3
apoA-II (mg/dl)	40.5 (7)	49.5 (11.5)	46.6 (5.5)	45	n.s.	-
apoB (mg/dl)	99.7 (13.4)	82.8 (17.1)	95.9 (18.2)	92	n.s.	-
apoC-II (mg/dl)	4.32 (1.55)	6.09 (2.69)	4.49 (2.17)	5.3	n.s.	-
apoC-III (mg/dl)	11.4 (4.3)	15.5 (6.9)	13.7 (2.7)	16.1	n.s.	-
apoE (mg/dl)	4.52 (0.89)	6.03 (1.86)	4.94 (1.12)	6.4	*	1/2

Table 2.3: Characteristics of SCARB1 P376L carriers and controls recruited for deep phenotyping. Demographic, plasma lipid and apolipoprotein traits measured from 1 P376L homozygote, 8 heterozygotes, and noncarrier controls from subjects identified

from sequencing or genotyping of the Penn HHDL Study cohort for deep phenotyping.

Lipid measurements were performed using an autoanalyzer from plasma.

¹Where applicable, data is presented as mean values \pm S.D.

²Numbers correspond to groups for comparison. Group 1 – Normal HDL-C controls,

Group 2 – High HDL-C controls, Group 3 – SCARB1 P376L heterozygotes

Contributions: Data measurements performed by core laboratory and analyzed by SAK and Paolo Zanoni.

Table 2.4

Table 2.4: Meta-analysis of Association of <i>SCARB1</i> P376L Variant with CHD							
Consortium	Study Cohort	Case P376L Carriers	Control P376L Carriers	Total CHD Cases	Total Controls	Frequency of Cases	Frequency of Controls
MICAD Exome Consortium	BioVU	6	10	4587	16546	0.0013	0.0006
	BHF	1	0	2833	5912	0.0004	0
	GoDARTS-CAD	1	0	1568	2772	0.0006	0
	MHI	0	4	2483	8085	0	0.0005
	North German	0	1	4464	2886	0	0.0004
	Ottawa	0	1	1024	2267	0	0.0004
	PAS	1	1	728	808	0.0014	0.0012
	Penn	3	0	683	156	0.0044	0
	South German	4	0	5255	2921	0.0008	0
	WHI-EA	8	29	2860	14929	0.0028	0.0019
CHD Exome+ Consortium	CCHS	1	1	2020	6087	0.0003	0.0001
	CIHDS/CGPS	4	3	8079	10367	0.0003	0.0001
	EPIC-CVD	4	2	9810	10970	0.0002	0.0001
	MORGAM	0	0	2153	2118	0	0
	PROSPER	1	0	640	638	0.0008	0
	WOSCOPS	0	0	659	687	0	0
Total		34	52	49846	88149	0.00068	0.00059
Meta-analysis Summary							
OR		1.79					
P-value		0.018					

Table 2.4: Meta-analysis of association of *SCARB1* P376L variant with CHD. CHD cases and healthy controls across the MICAD Exome Consortium and CHD Exome+ Consortium were genotyped for the *SCARB1* P376L variant using the Exome Array. The

association of the P376L variant with CHD cases was determined using a Mantel-Haenszel fixed-effects meta-analysis. *Contributions: Genotyping data generated and QCed by Nathan Stitzel, Sekar Kathiresan, John Danesh, CHD Exome+ Consortium members and CARDIoGRAM Exome Consortium members; analysis performed by SAK and Nathan Stitzel.*

CHAPTER 3 – LOSS-OF-FUNCTION OF *GALNT2* LOWERS HIGH-DENSITY LIPOPROTEINS IN HUMANS, NONHUMAN PRIMATES, AND RODENTS

Introduction

Single nucleotide polymorphisms (SNPs) in the first intron of *GALNT2* were among the first common variants associated with high-density lipoprotein cholesterol (HDL-C) and triglycerides (TG) through genome-wide association studies (GWAS) for plasma lipids [131,132,137]. We previously showed that *Galnt2* overexpression reduced and knockdown increased HDL-C in mice, offering further support that *GALNT2* was indeed the causal gene at the locus for this trait [86]. *GALNT2* encodes the enzyme UDP-N-Acetyl-D-galactosamine:polypeptide N-Acetylgalactosaminyl-transferase 2 (GalNAc-T2), which initiates O-glycosylation of proteins through addition of GalNAc to specific serine or threonine residues [278,279]. The biological functions of GalNAc-glycosylation are not fully understood but include regulation of protein processing by proprotein convertases [280], protein multimerization [281] and ectodomain shedding [282]. Among the proposed targets of GalNAc-T2 that may affect lipoprotein metabolism are angiopoietin-like 3 (ANGPTL3) [186] and apolipoprotein C-III (ApoC-III) [184,185], both liver-secreted regulators of TG and HDL-C.

Despite the early association of this locus with plasma lipids from GWAS, the mechanisms underlying how *GALNT2* regulates lipids have remained unknown and the directional relationship between *GALNT2* expression and HDL-C has been controversial. Our *Galnt2* overexpression and knockdown work suggested that *GALNT2* is a negative regulator of HDL-C levels since overexpression reduced and knockdown increased HDL-C in mice [86]. This directional relationship was supported by another study identifying

heterozygous carriers of a putative loss-of-function (LOF) coding variant in *GALNT2*, p.Asp314Ala (D314A) [184]. Contrasting these two studies, a recent evaluation of the GWAS SNPs in *GALNT2* showed that alleles associated with increased HDL-C correlated with increased hepatic *GALNT2* expression [183]. Given these discordant findings, here we investigated the mechanisms and directionality relating *GALNT2* to plasma lipids across humans and mammalian models of *GALNT2* loss-of-function.

Materials and Methods

Linkage analysis and whole exome sequencing

Identification of the p.Gln289* variant: Genomic DNA was extracted from EDTA blood probes by standard methods and genotyped with the Affymetrix Mapping array 6.0 (Affymetrix, Santa Clara, CA, USA). Analysis did not reveal pathogenic deletions or duplications. Mendelian segregation was calculated using PedCheck software and was confirmed in all instances. DNA from individual II:2 was enriched using the SureSelect Human All Exon Kit, which targets approximately 50 Mb of human genome (Agilent, Santa Clara, Ca, USA) and paired-end sequenced on a SOLiD 5500 xl instrument (Life Sciences, Carlsbad, CA, U.S.A.). Image analysis and base calling was performed using the SOLiD instrument control software with default parameters. Read alignment was performed with LifeScope 2.5 using the default parameters with human genome assembly hg19 (GRCh37) as reference. Single-nucleotide variants and small insertions and deletions (indels) were detected using LifeScope, GATK 2 and samtools/bcftools [283,284]. Variant annotation was performed using Annovar, integrating data from a variety of public databases [285,286]. Additionally, variants were compared to an in-house database containing more than 750 sequenced exomes to identify further

common variants which are not present in public databases. Finally, the variants were validated by PCR and Sanger sequencing according to the standard protocols to exclude technical artifacts and to test for segregation.

Identification of the p.Phe104Ser variant: DNA was extracted from blood samples (subjects I.1, I.2, II.1, II.2, II.3, II.4 and II.5 for family II) using a standard protocol. A genome-widescreen was performed with microarrays using the Illumina 6K panel at the P₃S platform (Pitié-Salpêtrière Hospital, France). By haplotype reconstruction, a homozygosity region was identified for subject II.2 and II.5 in the 1q41-q42 chromosomal region.

Exons from 3 µg of genomic DNA from patient II.2 (Family II) were captured using the SureSelect kit following the manufacturer's protocols (Agilent). The whole-exome DNA library was sequenced on the Hiseq2000 Illumina Genome Analyzer (Illumina). Sequence reads were mapped to the reference human genome using whole genomic sequence, refgene.txt file and reflink.txt (UCSC hg19) using the BWA v0.5.9 algorithm (<http://samtools.sourceforge.net>). Variant detection was performed with the SAMTools v0.1.17 software (<http://samtools.sourceforge.net>) and filtered variants were annotated according to snp132.txt.

Measurements of plasma ApoC-III and plasma lipids from human subjects

Human plasma, either crude plasma (1 µl) or C18 microtip purified as previously described [185], was separated on an 18% Tris-Glycine protein gel (Novex) for 90 min at 150 Volts on ice. Subsequently separated proteins were blotted onto a nitrocellulose membrane for 2 hours at 100 Volts and non-specific binding was blocked by incubating the membrane in 5% fat-free milk in TBS-T. The membranes were incubated overnight

with polyclonal rabbit anti-human ApoC-III (abcam ab21032) and subsequently in polyclonal swine anti-rabbit IgG-HRP (Dako) 1:4000 in TBS-T for 1 hour at RT. Protein bands were visualized by enhanced chemiluminescence (ECL) using a pierce ECL Western Blotting substrate (Thermo Scientific, Denmark) and a LAS 4000 (General Electric, US).

Total plasma cholesterol, LDL-C, and HDL-C were measured by Cobas® Modular analyzer series in samples from proband 1 (p.Gln289* homozygote) and age- and sex-matched noncarriers and with plasma from proband 2 (p.Phe104Ser homozygote) and relatives.

Recombinant GalNAc-T2 p.Phe104Ser activity assay

Recombinant soluble GalNAc-T2 WT and the p.Phe104Ser variant was expressed and purified using NiNTA agarose beads as previously described [287].

Galnt2^{-/-} rat

Galnt2^{-/-} rats were created at SAGE labs [288] operating under approved animal protocols. Zinc-finger nuclease (ZFN) targeting constructs for *Galnt2* were designed with the following binding site: AACAAAGTTCAACCAGGTGgagagtGACAAGCTACGCA. Sprague Dawley rats were used for microinjection and breeding and were housed in standard IVC (Individually Ventilated Cages) and racks and maintained under a 12-hour light/dark cycle with ad libitum access to food and water. ZFN technology was used to generate a 7 bp deletion in GalNAc-T2 exon 3 that resulted in a frame shift and a stop codon within the coding region of exon 3 disrupting a region in the *Galnt2* gene corresponding to its catalytic domain. Genotyping was accomplished with forward

primers 5'-GCGACACTCTCTGGTCTCCT -3' together with reverse primer 5'-AAGTACAAGGCCCATGCAAC-3'. WT rat DNA produces a 374 bp band, whereas a *Galnt2* knockout allele generates 161+213 bp bands. Briefly, *Galnt2* ZFNs were injected into fertilized single-cell embryos from Sprague Dawley rats. Upon confirmation of successful disruption of *Galnt2*, embryos were implanted to mother rats that generated progeny that were intercrossed to yield homozygous KOs. *Galnt2*^{-/-} rats demonstrated reduced fertility but were otherwise viable.

Plasma lipids in Galnt2 KO rat

Rats were fasted for 12-16 hours and anaesthetized using isoflurane before blood was drawn from the retro-orbital plexus. Cholesterol and triglyceride was measured by enzymatic methods in total plasma samples, fractions with $d < 1.063$ g/l and fractions with $d > 1.063$ g/l isolated by fixed-density ultracentrifugation using Beckman Optima MAX-E ultracentrifuge (Beckman Coulter) (100,000 rpm and 10°C for 4 hours).

RNA-Seq from rat liver

Total RNA was extracted from rat WT and *Galnt2* KO livers (N=3 per group) using RNeasy® kit (Qiagen). RNA integrity and quality was determined using Bioanalyzer instrumentation (Agilent Technologies). RNA sequencing was performed by the Beijing Genetics Institute (BGI). Briefly, libraries were constructed using the Illumina Truseq RNA Sample Preparation Kit. High quality total RNA was DNase I treated, mRNA isolated by magnetic bead Oligo (dT) enrichment, followed by mRNA fragmentation, cDNA synthesis, single nucleotide A (adenine) addition, and adapter ligation. Agarose gel purified fragments (~160 bp) were subjected to PCR amplification and quality control

(QC) before and subjected to next generation sequencing using Illumina HiSeq 2000 System (Illumina, USA). Total reads per sample were around 60 million with an average of 75% total mapped reads and an averaged 50% perfect match reads. Gene coverage ranged from 90-100% for 60% and 0-10% for 8%.

Bioinformatics analysis of whole exome sequencing reads were analyzed by alignment to the rat (*Rattus norvegicus*) reference genome (RGSC 5.0/rn5) using the CLC Genomic Workbench (Qiagen, Aarhus, Denmark). RPKM values were used for gene expression detection and mutant identification was done by aligning reads for *Galnt2* exon2.

GALNT2^{-/-} mice (Complete Galnt2 KO mice)

The *Galnt2* null allele harbors a deletion of exons 4-6, which contains the region essential for catalytic activity. Heterozygous *Galnt2* mice on a C57Bl/6 background were obtained from Merck Research Laboratories and intercrossed to generate *Galnt2^{-/-}* mice. All KO mice described herein were derived from heterozygous crosses, as both male and female KO mice exhibited impaired fertility and were unable to produce viable progeny. WT, heterozygous, and KO mice were distinguished by PCR-based genotyping of tail DNA, using two independent reactions. The first reaction utilizes primers that flank exons 4-6 of mouse *GALNT2* (F: 5'-GTACGTGAGACAGGCCTAAGG-3' R: 5'-CAAGCTTCATTTAGGACCAAGC-3'), amplifying a 2000 bp band in WT mice, a 334 bp band in KO mice, and both bands in heterozygous mice. PCR cycling parameters were as follows: 1 cycle of 95°C for 30 seconds; 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 2.5 min; and 1 cycle of 72°C for 10 min. The second reaction utilizes primers that flank either side of the 5' deletion site upstream of exon 4 (F: 5'-

AAGAGTCCACGTGAGTTCTGC-3' R: 5'-GGCTCTCACAGTATCAAAGG-3'), amplifying a 341 bp band in WT and heterozygous mice, and no band in KO mice due to loss of the reverse primer annealing site. PCR cycling conditions were identical to those described above except for an abbreviated extension time of 1 minute. Mice were subject to alternating 12-hour cycles of light and dark and given access to standard chow diet *ad libitum* unless otherwise indicated. Mice selected for experiments were age and gender-matched littermates, and genotypes were intermixed in cages to avoid undue cage effects.

Galnt2^{fl/fl} mice

The *Galnt2* conditional allele harbors *loxP* sites flanking exons 4-6. *Galnt2^{fl/+}* mice were obtained from Merck Research Laboratories and intercrossed to generate *Galnt2^{fl/fl}* mice, which were further intercrossed with one another. Conditional KO mice did not display any defects in fertility or viability. Mice were distinguished by PCR genotyping using primers that flank the 5' *loxP* site upstream of exon 4 (same as second primer set described above), which amplify a 341 bp band in *Galnt2^{+/+}* mice, a 502 bp band in *Galnt2^{fl/fl}* mice, and both bands in *Galnt2^{fl/+}* mice.

Adeno-associated virus serotype 8 (AAV8) vector encoding Cre recombinase under control of the liver-specific thyroxine-binding globulin (TBG) promoter was used to achieve post-natal deletion of hepatic *Galnt2* in *Galnt2^{fl/fl}* mice. Empty AAV8 particles (Null) served as a control. Viruses were provided by the University of Pennsylvania Vector Core. 10-12 week old gender-matched *Galnt2^{fl/fl}* littermates were fasted for 4 hours, pre-bled from the retro-orbital plexus, and intraperitoneally injected with 5×10^{11}

viral genomes/mouse. Bleeds were performed at 2, 4, 6 and 8 weeks post-injection for lipid analysis. Experimental groups were intermixed in cages to avoid cage effects.

Plasma lipid measurements in mice

Mice were fasted for 4 hours and anesthetized by isoflurane inhalation prior to retro-orbital bleed. Plasma was separated from whole blood by centrifugation at 10,000 rpm for 7 min at 4°C. Plasma levels of TC, HDL-C, TG, and alanine aminotransferase (ALT) were determined using a Cobas Mira autoanalyzer (Roche Diagnostic Systems) and Wako Chemicals reagents.

RNA isolation and real-time PCR analysis

Liver pieces (~50 mg) were homogenized in 1 mL Trizol reagent (Invitrogen) using a TissueLyser bead mill (Qiagen). Total RNA was isolated according to the manufacturer's instructions. RNA samples were subjected to DNase digestion using the DNA-free kit (Ambion) to remove residual genomic DNA contamination. 1 µg of DNase-treated RNA was reverse transcribed using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems) according to the manufacturer's instructions. TaqMan gene expression assays (Applied Biosystems) were used to detect mouse *Galnt2* (Mm00519804_g1) and *Actb* (4352341E). Reactions contained 0.5 µL 20x assay mix, 5 µL 2x TaqMan reaction mix, and 4.5 µL cDNA (diluted 2:13) and were performed in duplicate. Quantitative RT-PCR was performed on an Applied Biosystems 7900HT real-time PCR system. Relative expression differences were calculated by the delta delta Ct method using β -actin as the housekeeping gene.

Postprandial TG clearance studies

Mice were fasted overnight and administered 15 μ L olive oil (Sigma) per gram body weight by oral gavage. Retro-orbital bleeds were taken immediately prior to gavage and at 3, 5 and 7 hours post-gavage. Plasma samples at each time point were individually assayed for TG content on 96-well microtiter plates using the Infinity TG kit.

Radiolabeled HDL clearance studies

Human HDL ($d = 1.063\text{--}1.21$ g/ml) was isolated by conventional ultracentrifugation methods and dual-labeled with ^{125}I -tyramine-cellobiose (TCB), which labels primarily apoA-I, and ^3H -cholesteryl ether, a tracer of the cholesteryl ester core of HDL, as previously described [239]. Tyramine cellobiose (TCB) was prepared and provided by William Cain, University of Delaware. Both tracers are non-hydrolyzable allowing tissue uptake to be accurately assessed. After injection, blood samples were obtained by retro-orbital bleeding at 2 min and at 1, 3, 6, 9, and 24 hr. At the 24 hr timepoint, fasted animals were sacrificed, the vascular system was flushed with 15 ml ice-cold PBS, the liver, spleen, adrenals, and kidneys were harvested and radioactivity counted. The fractional catabolic rate (FCR) of apoA-I and HDL-cholesteryl ester was determined by fitting the corresponding tracer data to a single pool model containing an extravascular exchange using data normalized to the 2 min time point. Selective uptake was calculated as the difference between the HDL-cholesteryl ester and apoA-I FCRs.

Cynomolgus monkeys

Potent and efficacious *GALNT2* targeting siRNAs were identified following an in vitro screen of 20 different siRNA sequences conducted in a cyno hepatocyte cell line. The

best *GALNT2* siRNA (Sense (5'-3'): GCGUGGUGAUCACGUUUCAUU; Antisense (5'-3'): UGAAACGUGAUCACCACGCUG) and the luciferase siRNA (Sense (5'-3'): UAAGGCUAUGAAGAGAUACUU; Antisense (5'-3'): GUAUCUCUUCAUAGCCCUAUU) were chemically modified for stability, after which they were formulated in lipid nanoparticles (Tekmira Pharmaceuticals, Vancouver, CA) for specific delivery to the liver.

A mixed cohort of male and female cynomolgus monkeys (*Macaca fascicularis*) with body weights between 4-6 kg were obtained from the Bioculture Group (Mauritius) and pair-housed. Animals were acclimated to a feeding schedule of standard primate chow (Harlan Teklad Global 2050, Frederick, MD) and transferred to single housing in standard non-human primate cages one week prior to the start of the study. Water was provided ad libitum.

For the studies described here, plasma HDL-C and TG levels were measured several days prior to the start of the study and these lipid measurements were used to randomize the cynos into 3 treatment groups: saline (N=4), anti-luciferase siRNA (2.5 mg/kg, N=4), or anti-*GALNT2* siRNA (2.5 mg/kg, N=4). Injection of siRNAs were done as a single administration following which liver biopsies were taken on Days 3, 14, and 28 post-injection to assess *GALNT2* knockdown. Blood was also collected for plasma measurements on Days 0 (immediately preceding administration of siRNA), 3, 7, 14, 21 and 28 of dosing. All cynos were fasted daily from 3:45 pm to 8:30 am for the duration of the experiment.

Plasma lipid measurements in cynomolgus monkeys

Plasma lipids were determined using a OLYMPUS AU2700 chemistry analyzer (Olympus CO Ltd., Tokyo, Japan) and kits for total cholesterol, 2nd generation direct

LDL-C, 3rd generation direct HDL-C, and triglycerides from Roche Diagnostics (Indianapolis, IN).

Hepatic gene expression in cynomolgus monkeys

For quantitative RT-PCR, cyno liver biopsy samples (2.5 mg) were stabilized in RNAlater (Qiagen) prior to RNA isolation. Tissues were homogenized in Trizol reagent (Invitrogen) using a TissueLyser bead mill (Qiagen) and Total RNA was purified using the QIAGEN RNeasy Mini Kit according to the manufacturer's protocol. Purified RNA samples were normalized and converted to cDNA according to the manufacturer's protocol using the High-Capacity cDNA Archive Kit (Life Technologies). Real-time PCR was conducted on the Applied Biosystems 7900HT instrument. TaqMan gene expression assays (Life Technologies) were used to detect *GALNT2* (Hs00954065_m1) and 18S ribosomal RNA expression (Hs99999901_s1). Standard curves for *GALNT2* and 18s rRNA were generated by serially diluting cDNA from the saline treated animals. All measurements were performed in duplicate and expression levels were averaged within treatment groups.

Glycoproteomics sample preparation

For liver O-glycoproteomes, WT and *Galnt2* KO rats and mice animals were sacrificed and livers were perfused and dissected free. 100 mg rat or mouse liver (N=3 per group, pooled together by group) tissue was cut into small pieces using a scalpel and further homogenized in 5x volumes of 1% Rapigest in 50 mM AmBic using an Ultra-turrax blender keeping the sample on ice. Tissue homogenate suspension was sonicated at 50-60% Amplitude 3 times for 10 seconds and non-solubilised tissue debris was pelleted

at 10,000g for 15 min at 4°C. Solubilized liver proteins were heated at 80 °C for 5 min to further activate the RapiGest.

For plasma O-glycoproteomes from rat and mouse samples, 200 µL EDTA plasma was collected from WT or *Galnt2* KO rats (N=3 per group, pooled together by group), diluted in Acetate buffer treated with 0.1 U/mL Neuraminidase (Sigma) for 1 hour at 37 °C. The neuraminidase treated plasma was diluted by 20x in PNA binding buffer and passed over a 0.8 mL PNA-agarose column. After binding and washing PNA bound proteins are eluted using 0.1% Rapigest in 50 mM AmBic and heat.

Both solubilized liver proteins and plasma glycoprotein-enriched samples were reduced (5 mM DTT, 60°C for 45 min) and alkylated (10 mM IAA, RT in the dark for 30 min) and proteins digested using trypsin (1:100) (Roche) in 50 mM AmBic at 37°C, shaking overnight. Prior to purification using SepPak C18 columns, samples were acidified and precipitated and Rapigest was removed by centrifugation. Eluted peptides were treated with neuraminidase 0.1U/mL in Acetate buffer pH 5.0 for 1 hr 37 C to remove sialic acid. *Quantitative glycoproteomics.* Tryptic digests of isogenic pairs of WT and KO rat or mouse liver or plasma were labeled with light (L) and medium (M) isotopomeric dimethyl labels. Briefly, cleared neuraminidase treated acidified digests were loaded in equal amounts (peptide concentration) onto equilibrated SepPak C18 cartridges (Waters) followed by 3 x CV 0.1% trifluoroacetic acid wash. Digests were labeled on-column by adding 5 CV 30 mM NaBH₃CN and 0.2% formaldehyde in 50 mM sodium phosphate buffer pH 7.5 (L-labelling) or 30 mM NaBH₃CN and 0.2% D-formaldehyde in 50 mM sodium phosphate buffer pH 7.5 (M-labeling). Columns were washed using 3 CV 0.1% formic acid and eluted with 0.5 mL 50% MeOH in 0.1% formic

acid. Labeled Galb1-3GalNAc glycopeptides were separated from non-glycosylated peptides using a long PNA-agarose LWAC as previously described [289].

Human plasma O-glycoproteomes were essentially generated as described above for rat and mouse samples, except for the human plasma, the labeling step was omitted. 500 μ L plasma from proband #1 and 500 μ L from an age- and sex-matched control pool of 8 subjects were prepared as described above and run in two separate LWAC steps.

Isoelectric focusing

Lectin weak affinity chromatography fractions from total cell lysate digests were screened by preliminary liquid chromatography-mass spectrometry (LC-MS) for glycopeptide content, and fractions most enriched in glycopeptides were pooled together, dried by vacuum centrifugation, reconstituted in IPG rehydration buffer, and submitted to IEF fractionation [290]. Isoelectric focusing was performed by a 3100 OFFGEL fractionator (Agilent) using pH 3-10 strips (GE Healthcare) 12 fractions were collected and desalted by custom Stage Tips (C18 sorbent from Empore 3 M) and submitted to LC-MS and higher energy collisional dissociation/electron transfer dissociation tandem MS (HCD/ETD-MS/MS) as described below.

Mass spectrometry

EASY-nLC 1000 UHPLC (Thermo Scientific) interfaced via nanoSpray Flex ion source to an LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Scientific) was used for glycopeptide analysis. Briefly, a precursor MS1 scan (m/z 350–1,700) of intact mixture was acquired in the Orbitrap at a nominal resolution setting of 30,000, followed by

Orbitrap HCD-MS2 and ETD-MS2 (m/z of 100–2,000) of the five most abundant multiply charged precursors in the MS1 spectrum; a minimum MS1 signal threshold of 50,000 was used for triggering data-dependent fragmentation events; MS2 spectra were acquired at a resolution of 7,500 for HCD MS2 and 15,000 for ETD MS2. Activation times were 30 and 200 ms for HCD and ETD fragmentation, respectively.

Data analysis

Data processing was performed using Proteome Discoverer 1.4 software (Thermo Scientific). All spectra were searched with Sequest HT node applying sequentially full and semi-specific enzymatic cleavage. In all cases the precursor mass tolerance was set to 6 ppm and fragment ion mass tolerance to 50 mmu. Carbamidomethylation on cysteine residues was used as a fixed modification. Methionine oxidation and HexNAc attachment to serine, threonine and tyrosine were used as variable modifications for ETD MS2. All HCD MS2 were pre-processed as described [291] and searched under the same conditions mentioned above using only methionine oxidation as variable modification. All spectra were searched against a concatenated forward/reverse human-specific database (UniProt, January 2013, containing 20,232 canonical entries and another 251 common contaminants) using a target false discovery rate (FDR) of 1 %. FDR was calculated using target decoy PSM validator node, a part of the Proteome Discoverer workflow. The resulting list was filtered to include only peptides with glycosylation as a modification. This resulted in a final glycoprotein list identified by at least one unique glycopeptide. ETD MS2 data were used for unambiguous site assignment. HCD MS2 data were used for unambiguous site assignment only if the number of GalNAc residues was equal to the number of potential sites on the peptide.

Glycopeptide M/L ratios were determined using dimethyl 2plex method. The M/L ratio of the detected precursor ions doublet was calculated using sequentially the Event Detector Node and the Precursor Ion Node followed by manual validation of the precursor ions without reported quantitative values (singlets) within a selected pool as it is described previously [291].

In vitro glycosylation assays of PLTP peptides

Human, rat and mouse C-terminal PLTP peptides were designed and synthesized (NeoBioScience). GalNAc-transferase assays using soluble GalNAc transferases purified after recombinant expressed and monitored by MALDI-TOF mass spectrometry, were performed as previously described (Schjoldager et al., 2010; [185]).

PLTP activity measurement

PLTP activity measurements from rat plasma samples were performed using the Roar assay (Roar Biochemical, Inc). Fluorescent substrate measurements were completed using the Biotek Synergy 2 plate reader (BioTek). PLTP activity assays from mouse and human plasma samples were performed using a ³H-phosphatidylcholine (PC) substrate as previously described [292,293] and measured the transfer of ³H-PC to total human HDL over 2 hours at 37 °C. Assays were performed in triplicate and average values for each technical replicate were plotted.

Immunoblotting

All plasma samples subjected to SDS-PAGE for immunoblotting were measured using NuPAGE reagents for 1-dimensional SDS-PAGE (Life Technologies). ApoC-III

immunoblots from plasma samples were performed by separating 1 µl of plasma on 10% Bis-Tris NuPAGE gels using MES buffer, followed by transferring separated proteins to nitrocellulose membranes, blocking membranes for 2 hrs at room temperature with 5% fat-free milk in PBS (0.05% tween 20). Membranes were then incubated for 2 hrs at room temperature with a rabbit polyclonal antibody against murine apoC-III (sc-50378, Santa Cruz) at a dilution of 1:2000, followed by 3 x 15 min. washes with PBS / Tween solution and incubation with goat anti-rabbit IgG HRP conjugate (sc-2030, Santa Cruz) for 30 min. at room temperature. Membranes were again washed for 3 x 15 min. and proteins visualized by Luminata Crescendo chemiluminescent reagent (Millipore). PLTP immunoblots from plasma were performed in the same manner, but SDS-PAGE was completed using 4-12% Bis-Tris gels and MOPs running buffer. The primary antibody used for PLTP immunoblots was ab85562 (abcam), which was used at a dilution of 1:2000 for 2 hrs at room temperature and subsequently goat anti-rabbit IgG HRP (Santa Cruz) was used as a secondary antibody.

Lipase activity assays

Phospholipase activity from plasma samples was assessed by measuring activity from post-heparin plasma against an emulsion of thioPEG substrate (Avanti) and chromogenic agent 5, 5'-dithiobis (2-nitrobenzoic acid) (DTNB) (Sigma). 20 µl of 1:8 diluted post-heparin plasma is used as the enzyme source and the release of the chromogenic reagent is measured by a kinetic assay at 37 °C for 30 min. The data is calculated as the rate of appearance of the chromogenic byproduct of the DTNB-ThioPEG emulsion over time. All assays were performed in triplicate.

Statistics for results of animal studies. Data represent mean \pm standard deviation. Statistical analysis was performed by two-tailed unpaired Student's *t*-test or one-way ANOVA with Tukey's post-test as appropriate. Statistical significance was defined as $P < 0.05$.

RNA sequencing of human liver samples and allelic imbalance analysis

Allele-specific expression (ASE) analysis used data from two cohorts: 40 human liver samples processed at the University of Pennsylvania (UPenn data) and 32 human liver samples processed as a part of the Genotype Tissue Expression (GTEx) consortium (release v4; GTEx data) [150]. UPenn data was genotyped on the Illumina human 610 quad beadchips (GPL8887) at the Northwestern University Center for Genetic Medicine Genomics Core Facility according to the manufacturer's protocols [294]. One sample was removed due to call rate $<10\%$. Of the total 620,901 markers, 583,073 SNPs were used for analysis after stringent quality control including Fisher's exact tests for significant deviation from Hardy-Weinberg equilibrium at $P < 1 \times 10^{-3}$ (8,300 SNPs) and genotyping rate >0.9 (29,705 SNPs). GTEx data was genotyped on the Illumina Human Omni 5.0 Quad and Infinium ExomeChip at the Broad Institute. Detailed information on genotype quality control was recently described [150]. Briefly, using the software PLINK [295], the following quality control measures were applied: i) removal of SNPs mapping to more than one genomic location; ii) SNPs with <0.9 genotyping rate; iii) samples with call rate <0.95 ; iv) samples with inconsistent sex information; v) samples with excess heterozygosity or cryptic relatedness; vi) HWE $P < 1 \times 10^{-6}$; vii) SNPs with predicted missingness $P < 1 \times 10^{-8}$; viii) testing for batch effects $P < 1 \times 10^{-3}$, and ix) SNPs with genotyping rate <0.95 .

Phasing and imputation were performed using SHAPEIT (v2.r790) [296] and IMPUTE2 (v2.3.1) [297]. SNPs were aligned to the 1000 Genomes Project Phase I freeze haplotypes by chromosomal position (GRCh37/hg19) and imputed using these references(https://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated_SHAPEIT2_16-06-14.html). SHAPEIT was run for each chromosome, with default parameters. IMPUTE2 was run on non-overlapping 5Mb fragments, with an estimated effective population size of 20,000. Phased and imputed genotypes were further filtered for best-guessed genotypes at posterior probability >0.9 and HWE<1×10⁻⁶.

Total RNA was extracted from each liver biopsy by homogenization in TRIzol, followed by RNAeasy clean up (Qiagen). RNA quality was assessed by Agilent Bioanalyzer (minimum RIN = 7). Sequencing libraries were constructed using Illumina's TruSeq stranded mRNA library prep kit. Sequencing was performed using HiSeq2500 at the University of Pennsylvania Next Generation Sequencing Core. All libraries were multiplexed into one large pool and sequenced across multiple lanes. RNA sequencing reads were aligned to the human genome (GRCh37/hg19) using STAR (v2.3.1z4) [298], using the following options: --outSAMattributes Standard --outFilterMultimapNmax 10 --outFilterMismatchNmax 10 --outFilterMismatchNoverLmax 0.3 --alignIntronMin 21 --alignIntronMax 0 --alignMatesGapMax 0 --alignSJoverhangMin 5 --alignSJBoverhangMin 3 --runThreadN 12 --readFilesCommand zcat --. Reads with alignment quality scores less than 10 were removed using SAMtools [284]. Coding SNPs were identified from the imputed genotypes and reference mapping bias was controlled using WASP (<http://dx.doi.org/10.1101/011221>).

Reads derived from PCR duplication during library preparation were identified by identifying read pairs with identical start and end alignment coordinates; one read pair

was randomly retained. Reads overlapping each coding SNP were extracted from each BAM file. Only properly paired reads mapping to the correct strand were retained. Overlapping read pairs were only counted once. Heterozygous coding SNPs were retained for analysis if the genotype probabilities estimated by IMPUTE2 were > 0.9. We then extracted allele-specific read depth at each coding SNP, assigning reads to either the 'A' or 'G' allele of rs4846914 using the phased haplotypes from each individual. Imputed, coding SNPs with more than 90% of their RNAseq reads mapping to any particular allele were filtered out, under the assumption that the genotype was improperly imputed (i.e., the SNP is homozygous). Allele-specific read depths were summed across all coding SNPs in the gene. Only individuals with more than 30 reads meeting these criteria were included in downstream analyses.

Within each individual, we tested for deviation from the null hypothesis that both alleles are expressed at equal levels using a simple binomial test. To combine information across individuals (i), we modeled the probability each read is derived from allele A (p_i), by logistic regression:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i$$

Where X_i indicates if the individual i is heterozygous ($X=1$) or homozygous ($X=0$) at rs4846914. The model was run using the R glm function. To assess the significance of the effect of genotype, we computed a Z-score and, from that, a two tailed p-value by comparison to the normal distribution.

Rare variant association studies from exome sequencing

A linear regression was performed separately for each variant with HDL-C after adjusting for age, sex, MI status, cohort, and first 4 principal components of ancestry as covariates, as was described previously [299].

Study approval

All animal protocols for rat studies were approved by the SAGE laboratories at Institutional Animal Care and Use Committee (IACUC). Mouse studies were approved by the University of Pennsylvania Institutional Animal Care and Use Committee (IACUC). Studies in cyno were performed according to American Association for Accreditation of Laboratory Animal Care guidelines and protocols were approved by the Bristol-Myers Squibb Animal Care and Use Committee. The study of the family of variant 2 was approved by the Ethic Committees of the Universities of Bonn and of Erlangen-Nürnberg in Germany. Informed consent of all examined persons or of their guardians was obtained.

Results

Humans homozygous for LOF mutations in GALNT2 have low HDL-C

Through exome sequencing and mapping in two independent studies, we identified two unrelated individuals homozygous for different mutations in *GALNT2*. First, using linkage analysis and whole exome sequencing in families in search of causes of heritable neurological traits, we identified a female of Moroccan origin homozygous for a T>C mutation at position Chr1:230338973 in exon 3 of *GALNT2*, resulting in a p.Phe104Ser mutation (**Figure 3.1 A-C**). This mutation was absent from the 1000

Genomes, NCBI dbSNP, NHLBI Exome Sequencing Project, and Exome Aggregation Consortium (ExAC) variant databases and was predicted deleterious by Align GVGD, SIFT, Polyphen2, MutationTaster and SNP&GO software. We tested recombinant secreted GalNAc-T2 with the p.Phe104Ser variant (mutant) or no variant (WT) for their ability to glycosylate peptide substrates corresponding to two targets of GalNAc-T2, ApoC-III and immunoglobulin A1 (IgA1). While the WT GalNAc-T2 was highly active with both substrates, the p.Phe104Ser GalNAc-T2 mutant was completely inactive with ApoC-III and almost inactive with the IgA1 substrate.

Another exome sequencing study for recessive heritable causes of intellectual disability [300] identified a nonsense mutation in *GALNT2* (NM_004481.4:c.865C>T; p.Gln289*) in a male proband of Pakistani origin. This mutation within the catalytic domain leads to a non-functional protein (**Figure 3.1 A**) [301]. The variant was confirmed with Sanger sequencing in the proband and family carriers (**Figure 3.1 B-C**). Like the p.Phe104Ser variant, this variant was also absent from existing variant databases.

To further validate the LOF mutations *in vivo*, we analyzed plasma of the probands and sampled relatives for ApoC-III glycoforms by SDS-PAGE immunoblotting (**Figure 3.1 D**). While ApoC-III normally circulates as a mono- or di-sialylated protein with a molecular weight of ~9.5 kDa, we found only a band for non-glycosylated ApoC-III in the plasma of the two probands. In plasma from the heterozygous family members, we identified the normal bands with a distribution of mono- and di-sialylated isoforms, suggesting that heterozygosity for both *GALNT2* LOF mutations retain sufficient enzyme activity to glycosylate ApoC-III *in vivo*. Notably, we found that both probands had reduced levels of HDL-C (<5th percentile for age and gender) (**Figure 3.1 E**). We also

found moderately reduced plasma TG in both probands and a moderately reduced LDL-C in the homozygote for the p.Phe104Ser variant (**Figure 3.1 E**).

Reduced HDL-C in rodent and nonhuman primate models of GALNT2 deficiency

To further address the effect of GalNAc-T2 LOF on plasma lipids, we investigated *GALNT2* deficiency *in vivo* in multiple mammalian models. We generated *Galnt2* KO mice and rats, which harbor deletions in exons corresponding to the catalytic domain of the enzyme (**Figure 3.2 A-B**). *Galnt2*-deficient mice had near complete loss of hepatic *Galnt2* mRNA levels (**Figure 3.2 C-E**), with no observed effects on expression of other GalNAc-transferase genes (**Figure 3.2 G-H**). In both chow-fed and Western diet-fed conditions, complete *Galnt2* KO mice demonstrated reduced HDL-C relative to WT littermates (20-22% reduction in KOs, $P < 0.05$; **Figure 3.3 A**). Because rodents lack cholesteryl ester transfer protein (CETP), a critical regulator of plasma lipids in humans, we tested whether the relationship of *Galnt2* to HDL-C was maintained after expression of CETP in *Galnt2* KO and WT mice through administering AAV vectors expressing CETP. After 4 weeks of CETP expression, *Galnt2* KO mice had 40% lower HDL-C levels compared to WT littermates ($P < 0.01$, **Figure 3.3 A**), with similar plasma CETP levels in both groups (**Figure 3.3 B**). *Galnt2* KO mice also demonstrated a robust elevation in fasting TG (>50% increased TG in *Galnt2* KOs on chow diet relative to WT, $P < 0.05$), which was also observed upon 3 weeks of feeding a Western-type diet and after 4 weeks of AAV-mediated CETP expression ($P < 0.05$ for both; **Figure 3.3 C**). Importantly, in chow fed mice, we observed no intermediate effect of heterozygous *Galnt2* deficiency on HDL-C or TG (**Figure 3.3 A**).

Given the importance of the liver in HDL-C metabolism, we also generated liver-specific models of *GALNT2* deficiency to test if the association between *Galnt2* and plasma lipids is mediated by hepatic expression. Liver-specific *Galnt2* KO mice (*Galnt2* LSKOs) were generated by injecting *Galnt2^{fl/fl}* mice with an AAV vector expressing Cre recombinase under the regulation of a liver-specific promoter, as well as by breeding *Galnt2^{fl/fl}* mice to Albumin-Cre transgenic mice. *Galnt2* LSKOs showed near-complete reduction of liver *Galnt2* mRNA levels (**Figure 3.2 D**). LSKO mice demonstrated 15-22% reduced HDL-C relative to WT controls ($P < 0.01$ for both; **Figure 3.3 D**). While the HDL-C reduction in the LSKOs was similar to that of complete KOs, plasma TG in LSKOs were similar to WT controls (**Figure 3.3 E**).

Similarly to the results in mice, *Galnt2* KO rats had near complete loss of hepatic *Galnt2* mRNA levels (**Figure 3.2 E**) and 30% lower plasma HDL-C levels on a chow diet than WT littermates ($P < 0.01$; **Figure 3.3 F**). There was a similar reduction in a smaller cohort of WT vs. KO rats (3 per group) fed a Western diet for 6 weeks ($P < 0.05$; **Figure 3.3 F**). But unlike the complete KO mice, *Galnt2* KO rats demonstrated reduced plasma TG while fed a chow diet. There were no significant differences in plasma TG after 6 weeks of Western diet feeding (3 per group) (**Figure 3.3 G**). This may highlight a potentially complex, species-specific regulation of lipoprotein metabolism by *GALNT2*.

We also explored the impact of *GALNT2* deficiency on non-HDL cholesterol (nonHDL-C) levels in our rodent models. *Galnt2* KO mice fed a chow diet, a Western diet, or administered CETP AAV all demonstrated no differences in plasma nonHDL-C (**Figure 3.4 A**), with similar results when measured by fast-protein liquid chromatography (FPLC) of plasma lipoproteins (**Figure 3.4 B-C**). Similarly, liver-specific KO mice also demonstrated no significant differences in nonHDL-C (**Figure 3.4 D-E**), while KO rats on

chow diet demonstrated a modest reduction in nonHDL-C on a chow diet with no differences in nonHDL-C after 6 weeks of Western diet feeding (**Figure 3.4 F**). Collectively, these findings suggest a minimal impact of *GALNT2* deficiency on nonHDL-C levels, consistent with the lack of association of this trait with the *GALNT2* locus in the prior human GWAS.

Impact of GALNT2 deficiency in mice on HDL-C and TG turnover

Using the LSKO model, we performed clearance studies of HDL radiolabeled with both ^3H cholesteryl ether and ^{125}I -ApoA-I. *Galnt2* LSKO mice demonstrated a moderate increase in the clearance of ^3H cholesteryl ether but not in ^{125}I -ApoA-I relative to controls (**Figure 3.5 A-B**), consistent with enhanced HDL cholesterol ester-selective catabolism. Given the previous implication of ApoC-III and ANGPTL3 as putative GalNAc-T2 targets, we also explored postprandial TG clearance in the complete KO and liver-specific KO mice (**Figure 3.5 C-D**). We observed delayed TG clearance in the complete KOs but not the liver-specific KO mice. This further supports the notion that the regulation of plasma TG levels by *Galnt2* in mice may be extra-hepatic. Despite the observation of delayed lipolysis *ex vivo* in the complete KO mice, we observed no differences in post-heparin plasma TG or phospholipase activities in these models (**Figure 3.5 E-H**)

Knockdown of hepatic GALNT2 in non-human primates reduces HDL-C levels

We also studied the effects of hepatic *GALNT2* silencing on plasma lipids in a nonhuman primate model, the cynomolgus monkey (cyno) using siRNA. Hepatic *GALNT2* mRNA knockdown mediated by a single dose of siRNA was substantial and

persisted for up to 4 weeks (**Figure 3.2 F**). *GALNT2* knockdown in cyno resulted in >30% reduction of HDL-C relative to controls as early as 3 days after siRNA treatment (**Figure 3.3 G**). Knockdown of *GALNT2* in cyno resulted in a trend towards elevated fasting TG that was 2-fold higher in the knockdown group 7 days after siRNA treatment ($P<0.05$; **Figure 3.3 H**). Additionally, we found a reduction in nonHDL-C in the anti-*GALNT2* siRNA treated monkeys (**Figure 3.4 G**), possibly due to the elevation in nonHDL-C in that control group itself; there were no differences in nonHDL-C between the anti-*GALNT2* siRNA treated monkeys and saline-treated controls.

Quantitative differential O-glycoproteomics confirm ANGPTL3 and identify PLTP as O-glycoprotein targets controlled by GalNAc-T2

To understand mechanisms underlying the effect of *GALNT2* deficiency on plasma lipids we observed in the human subjects and animal models, we employed an unbiased quantitative O-glycoproteomics strategy to identify protein targets with altered glycosylation [302]. We performed differential quantification of plasma (human, rats and mice) and liver (rats and mice) tissue O-glycopeptides from *GALNT2* LOF models (**Figure 3.6 A**), revealing multiple glycosites in each model present in WT samples but >10-fold reduced or absent in KO samples (**Figure 3.7 A-C & Table 3.1-3.3**).

From the human O-glycoproteome, we confirmed ApoC-III (Thr74) as a GalNAc-T2 dependent substrate. Notably, ApoC-III was not found in the same analysis of rat and mouse plasma and liver samples (**Table 3.1**), in agreement with our finding that the rodent sequences around the glycosites are different and not specifically served by GalNAc-T2 but redundantly by multiple isoforms [185]. This was confirmed by immunoblotting of ApoC-III from rat and mouse plasma showing no difference in SDS-

PAGE migration and levels with only one major band detectable in both WT and KO samples (**Figure 3.7 D-E**). We also confirmed that O-glycosylation of ANGPTL3 (Thr226) was selectively lost in the human *GALNT2*-deficient O-glycoproteome (**Table 3.1**), in agreement with our previous *in vitro* studies [185]. The sequence around the Thr glycosylation site in ANGPTL3 is conserved in rodents and *in vitro* analysis of rodent ANGPTL3 confirmed that only GalNAc-T2 was active with this substrate, however we did not detect the corresponding glycopeptide in our glycoproteomics analysis of rodent plasma and liver (**Table 3.2**).

Our differential O-glycoproteomics identified phospholipid transfer protein (PLTP) as O-glycosylated in the plasma and liver of WT mice but absent in KOs (**Table 3.3**). The identified O-glycosites were located in the C-terminus of PLTP, a region largely conserved among humans, mice, and rats (**Figure 3.6 B**). To validate PLTP as a non-redundant target of GalNAc-T2, we tested the *in vitro* specificity of the two main GalNAc-Ts expressed in liver, GalNAc-T1 and GalNAc-T2, with peptide substrates corresponding to the C-terminus of human, rat and mouse PLTP (**Figure 3.6 C**). We found that GalNAc-T2 was more active than GalNAc-T1 with all three peptides and in particular with the rodent peptides (**Figure 3.6 C**). These data support PLTP as a conserved preferential target of GalNAc-T2.

PLTP transfers phospholipids from apoB-containing lipoproteins to HDLs in circulation, and is a known regulator of plasma HDL metabolism [303]. Mice lacking PLTP activity exhibit reduced HDL-C [304] and increased clearance of HDL-C *in vivo* [305], a finding resembling that of the *Galnt2* KO rodents. We hypothesized that GalNAc-T2 may modulate HDL metabolism at least in part through promoting PLTP function. We compared PLTP activity in plasma from PLTP-deficient mice, rats, and humans relative

to controls. *Galnt2* KO mice and rats both showed >20% lower plasma PLTP activity relative to WT littermates ($P < 0.01$ for each; **Figure 3.6 D-E**), with no apparent difference in total plasma PLTP protein was observed by immunoblotting (**Figure 3.6 D-E**), suggesting that the reduction of PLTP activity in the plasma of *Galnt2* KO rodents is not due to reduced plasma PLTP protein levels but rather reduced function. This finding is consistent with prior mutagenesis and deletion studies implicating the final 30 residues of PLTP as critical for lipid transfer activity *in vitro* [306]. We measured plasma PLTP activity from the p.Gln289* and p.Phe104Ser homozygotes and family carriers and controls (**Figure 3.6 F**) and found that plasma from both probands had reduced PLTP activity relative to the mean of the p.Phe104Ser heterozygotes or noncarriers. Furthermore, when we measured a 'specific activity' for plasma PLTP, we found a reduction in the two homozygotes relative to controls, with the p.Gln289* homozygous subject's activity >60% lower and the p.Phe104Ser homozygous subject's activity >20% lower than the mean of the noncarrier control subjects' activity (**Figure 3.6 G**). These data suggest that *GALNT2* is required for facilitating PLTP function and maintaining HDL-C levels in rodents and humans. It has been shown that PLTP deficiency in hyperlipidemic mice models attenuated both VLDL secretion and atherosclerosis [307]. Thus we measured VLDL-TG secretion in the *Galnt2* WT vs. complete KO mice and found a ~15% reduction in the rate of secretion in the KOs ($P < 0.05$; **Figure 3.5 I**). These data further support the notion of reduced PLTP function in the setting of *GALNT2* deficiency.

We asked if restoration of *Galnt2* expression was sufficient to rescue the HDL-C and PLTP activity reduction *in vivo*. We reconstituted murine *Galnt2* in the liver of *Galnt2* KO mice using AAV at a dose of 1×10^{11} genome copies (GC) of virus per mouse, which

was 10-fold lower than the dose used in our prior overexpression study of *Galnt2* in WT mice [86]. We compared hepatic *Galnt2* restoration at this dose in *Galnt2* KO mice to WT mice and *Galnt2* KO mice receiving the same dose of Null virus and observed robust restoration of hepatic *Galnt2* in the rescue group at 28 days after AAV administration (**Figure 3.8 A**). Restoration of hepatic *Galnt2* expression markedly raised the reduced levels of HDL-C in the KO mice back to the levels in WT controls receiving Null AAV (**Figure 3.8 B**). Similarly, we monitored plasma phospholipid levels, which are low in the setting of PLTP deficiency, and found that while KO mice had lower phospholipids relative to WT initially, the KO mice with hepatic *Galnt2* reconstitution showed elevated phospholipids over 4 weeks after dosing (**Figure 3.8 C**). We also measured plasma TGs and found that levels were variable across the different timepoints but hepatic *Galnt2* rescue in the KO mice caused a moderate reduction in TG levels by 2 weeks post-injection (**Figure 3.8 D**). We observed similar trends in cholesterol and triglycerides in the lipoprotein fractions from FPLC separation of pooled plasma of rescued mice as we observed above (**Figure 3.8 E-H**).

Since *Galnt2* hepatic reconstitution rescued plasma HDL-C levels in the *Galnt2* KO mice, we next assessed the impact of hepatic *Galnt2* restoration on plasma PLTP activity. We measured plasma PLTP-mediated transfer of phospholipids to HDL using plasma from the 3 groups of mice described above both before and after AAV administration. Prior to AAV dosing, plasma PLTP activity in the KO groups of mice was approximately 33% lower than that of WT mice ($P < 0.0001$; **Figure 3.8 I**). 28 days after *Galnt2* reconstitution in KO mice, their plasma PLTP activity had risen to the level of WT mice (**Figure 3.8 J**). At both timepoints, total plasma PLTP levels were comparable

(**Figure 3.8 I-J, bottom**). These data support the notion that hepatic *Galnt2* is both necessary and sufficient to potentiate full plasma PLTP activity in mice *in vivo*.

In addition to PLTP, we identified other specific targets for glycosylation by GalNAc-T2 that may contribute to the differences observed in the rodent phenotypes. These include a non-redundant site in the C-terminus of ApoE (Thr305), a critical ligand for hepatic lipoprotein receptors, in both rat liver and plasma. Additionally, we identified a non-redundant T2-specific glycosite in the N-terminus of hepatic lipase (HL) (Thr38) in mouse liver and plasma. HL is a hepatocyte-expressed TG lipase and phospholipase that hydrolyzes lipids on TG-rich lipoproteins after secretion from liver cells onto the cell surface in a manner dependent on HDL [308]. Mouse liver glycoproteomics also identified two GalNAc-T2 specific glycosites in the hepatic lipoprotein receptor low-density lipoprotein related receptor 1 (LRP1) (Thr27 and Thr3937). LRP1 is a hepatic receptor mediating the uptake of ApoE-containing TG-rich lipoprotein-remnant particles.

GALNT2 SNPs associated with lower HDL-C are associated with reduced GALNT2 expression in human liver

We next asked if the common SNPs in *GALNT2* associated with lipids in GWAS impacted *GALNT2* expression in direction consistent with that observed in the LOF models. To answer this, we sequenced the RNA transcriptome from human liver samples and quantified allele-specific *GALNT2* expression while accounting for genotype at rs4846914, the lead SNP in *GALNT2* associated with HDL-C and TG (**Figure 3.9 A**). In a recent GWAS meta-analysis, each copy of the G allele for this SNP was associated with a mean reduction in HDL-C of 0.48 standard deviation (S.D.) units and a 0.40 S.D. unit increase in TG per copy (**Figure 3.9 B**). We compared *GALNT2*

mRNA transcript levels arising from the G allele vs. the A allele for the rs4846914 SNP for each liver RNA transcriptome. From analysis of rs4846914 G/A heterozygotes, we observed an 8% reduction in *GALNT2* expression from the G allele relative to the A allele (**Figure 3.9 C**). The heterozygotes thus demonstrated a modest but significant allele-specific expression ($P < 2 \times 10^{-16}$; **Figure 3.9 D**), with the G allele (associated with lower HDL-C) having lower expression than the A allele (associated with higher HDL-C), consistent with the directionality from LOF variant carriers and animal models. These results are also consistent with a prior eQTL study of plasma lipid GWAS SNP effects on nearest-gene expression levels in human liver [182], and a recent fine mapping of the GWAS-associated *GALNT2* noncoding SNPs [183]. We note that the directionality found here of reduced *GALNT2* expression by alleles associated with higher TG in GWAS contrasts with the finding in the two *GALNT2* LOF variant carriers of relatively low TG; this highlights the complexities of assessing plasma TGs in humans and the need to identify additional homozygotes for LOF variants for additional assessment.

Discussion

Here we aim to better reconcile the contrasting directionalities from our current work and these previous efforts by further exploring the prior conclusions. We first examined our conclusions from our 2010 report (Teslovich et al., 2010) that showed that *Galnt2* overexpression reduced HDL-C levels while shRNA-mediated knockdown of hepatic *Galnt2* raised HDL-C in mice. We subsequently found that at the AAV dose of 1×10^{12} GC/mouse used previously, overexpression of *Galnt2* reproducibly decreased HDL-C by 25-30% 14 days after administration (**Figure 3.10 A**). This dose also caused a >100,000-fold overexpression above endogenous *Galnt2* levels of WT mice treated with

Null virus (**Figure 3.10 B**). A dose escalation of *Galnt2* AAV in WT mice showed that the HDL-C reduction was not observed at AAV doses lower than 3×10^{11} GC/mouse despite high overexpression of *Galnt2*. We also questioned whether the HDL-C reduction was due to *Galnt2* enzymatic activity. To test this, we engineered an AAV variant that expressed a mutant form of *GALNT2* harboring the p.Glu334Gln (E334Q) variant, a highly conserved residue in the catalytic domain critical to glycosyltransferase activity [309]. Glycosyltransferase activity assays testing this mutant showed that it conferred no activity relative to WT GalNAc-T2 (**Figure 3.10 C**). Comparing WT and E334Q mutant *Galnt2* AAVs at the high dose of 1×10^{12} GC/mouse showed that both equally reduced HDL-C by 30% after 14 days (**Figure 3.10 D**). These data support the conclusion that *Galnt2* overexpression by AAV at the doses we used previously causes a transgene-specific but glycosyltransferase activity-independent reduction in HDL-C levels.

Our prior work also suggested that AAV shRNA-mediated hepatic *Galnt2* knockdown raises HDL-C levels, which contrasts with the findings from the current work in multiple LOF models. Subsequent studies by others have shown that high doses of AAV shRNA expression in the liver can mediate hepatotoxicity due to oversaturation of microRNA (miRNA) export pathways from the nucleus and reduction of essential miRNAs [310]. Importantly, these findings were shown to be shRNA sequence-specific but not target gene-specific. Since our 2010 report, we observed that anti-*Galnt2* AAV shRNA treatment at the high dose of 1×10^{12} genome copies/mouse caused an elevation in HDL-C (**Figure 3.10 E**) but also hepatic dysfunction as measured by plasma alanine aminotransferase (ALT) levels (**Figure 3.10 F**). In our prior AAV shRNA experiments at this dose, we also found reduced levels of miR-122a (**Figure 3.10 G**), which constitutes a majority of miRNAs in the liver [311]. As several other hepatic miRNAs have been

shown to modulate HDL-C levels [312-314], we postulate that our results from AAV shRNA experiments may be influenced by additional hepatocellular changes as a result of the degree of shRNA overexpression. Indeed, experiments with other AAV shRNAs targeting *Galnt2* at a lower dose of 2.5×10^{11} GC/mouse caused no elevation in HDL-C and also no increase in plasma aminotransferases (**Figure 3.10 H-K**). We posit that the germline *Galnt2*-deficient models we study here offer a 'cleaner' view of the phenotype of heritable LOF, which we demonstrate is consistent across multiple species.

The 2011 study by Holleboom and colleagues [184] reported two probands with elevated HDL-C who were found to be heterozygotes for a nonsynonymous missense variant in *GALNT2*, D314A. This report suggested that this variant modestly reduces catalytic activity of GalNAc-T2 as measured by an *in vitro* enzymatic assay. However, a subsequent identification of this variant in a cohort of 2,000 Danes and functional testing showed the variant to confer normal enzymatic activity [315]. Because the 2011 report identified the variant in 2 high HDL-C probands and an additional 6 heterozygotes with normal HDL-C levels, we wondered if identification of additional carriers of this variant and potentially homozygotes would help us better determine the relationship of this variant to HDL-C levels and any allele-dosage effect of the variant on this trait. We searched for carriers of the D314A variant in 2 ethnically distinct exome sequencing cohorts in which HDL-C levels were measured (**Figure 3.11 A**). In both cohorts, we identified numerous heterozygous carriers of this *GALNT2* variant and one homozygote. We tested the association of this variant with HDL-C levels in both cohorts and found no significant relationship in either cohort; indeed, the D314A homozygote had low HDL-C levels (23 mg/dl). For both cohorts, we identified highly significant associations of known variants at other HDL-C loci, supporting the validity of these findings. These data

suggest that in larger populations across ethnicities, the D314A variant in *GALNT2* is not associated with higher HDL-C levels and that the initial report was limited by power considerations.

Our data demonstrate that ApoC-III is a human species-specific non-redundant substrate for GalNAc-T2, but the role of O-glycosylation in modulating its lipoprotein modulatory functions remains unclear. Prior work showed that lack of ApoC-III O-glycosylation at Thr74 did not impair protein secretion or binding affinity for lipoproteins *in vitro* [97]. Moreover, human subjects lacking ApoC-III glycosylation due to inheritance of a missense variant at the glycosylation site Thr74 demonstrate normal plasma TG levels and ApoC-III incorporation into lipoproteins [96], suggesting that the glycosylation of ApoC-III alone may not substantially impact plasma lipoprotein levels. Indeed, our own measurement of *in vitro* LPL activity inhibition by total plasma ApoC-III (containing mostly glycosylated isoforms) and nonglycosylated ApoC-III showed equally potent inhibition of lipolysis of a radiolabeled TG substrate in the presence of recombinant LPL (**Figure 3.11 B**).

Integrating our findings, the data presented here support a working model (**Figure 3.11 C**) of the regulation of lipoprotein metabolism by *GALNT2* by which it regulates at least three substrates, ANGPTL3, ApoC-III, and PLTP, in ways that may have differing impacts on these target proteins with regard to their respective processing and functions. These modifications may be species-specific as in the case of ApoC-III and it is likely that additional targets underlie the complex phenotypic differences among *GALNT2* LOF models, which remain our area of continued study.

To date, few of the novel lipid loci identified by GWAS have been shown to physiologically influence plasma lipids *in vivo*. Strategies for studying function of these

novel loci in animal models have largely involved somatic overexpression and knockdown in murine models as well as generation of whole-body or tissue-specific KO mice. These approaches meet with the challenges of 1) the difficulty in pinpointing the causal gene at a newly associated locus harboring multiple genes, 2) the ability to appropriately ascertain the relevant tissue(s) in which expression of the candidate gene regulates the associated phenotype, and 3) the limitations of relating findings in murine models to humans due to the complex physiological differences between rodent and human lipoprotein metabolism. This is an exceptionally difficult task when the candidate gene is involved in partly redundant glycosylation of a multitude of proteins. Through our cross-species studies of *GALNT2*, we provide an example of how studies of LOF mutations in rare individuals combined with multiple animal model systems might be leveraged to validate and determine the relationship of GWAS candidate genes to complex traits.

Figure 3.1

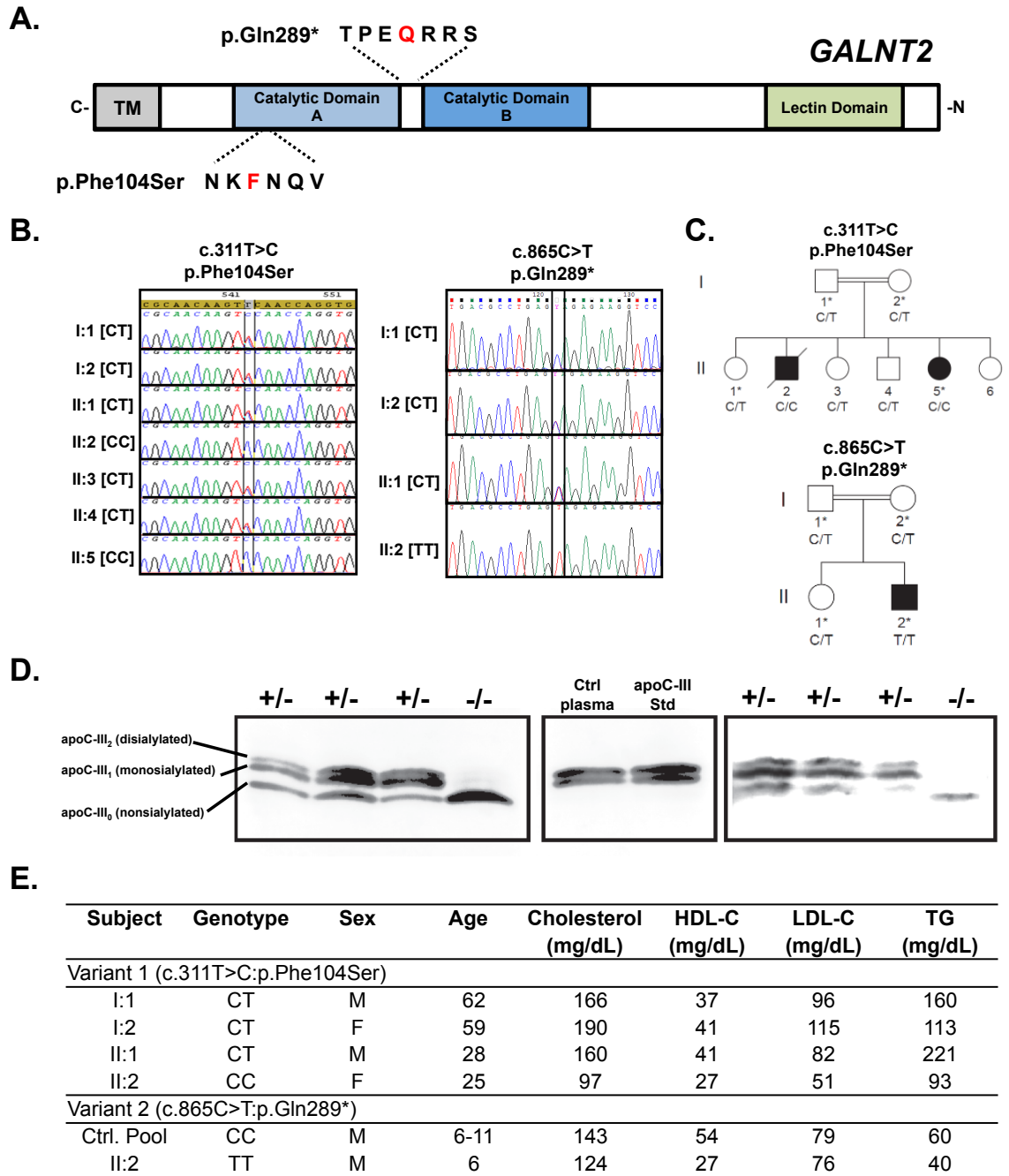


Figure 3.1: Identification and lipid phenotype of human homozygotes with *GALNT2* loss-of-function coding variants. A. Schematic *GALNT2* protein showing

position of the two identified variants, p.Phe104Ser and p.Gln289. **B.** Sanger sequencing of identified variants, c.C311T>C: p.Phe104Ser, and c.C865T: p.Gln289*. **C.** Pedigree of family members of probands for variants in **A-B**. Asterisk denotes individuals analyzed by ApoC-III immunoblot. **D.** Immunoblot of plasma ApoC-III from the probands and family controls for the *GALNT2* variants. Migration positions of the 3 major ApoC-III isoforms, nonsialylated ApoC-III (ApoC-III₀) lacking any O-glycan modification; monosialylated ApoC-III (ApoC-III₁) containing N-acetylgalactosamine, galactose and 1 terminal sialic acid; and disialylated ApoC-III containing N-acetylgalactosamine, galactose, and 2 terminal sialic acids (ApoC-III₂), are indicated on the left. **E.** Plasma lipids of *GALNT2* variant carriers. *Contributions: All data generated by Katrine Schjoldager and in part analyzed in final form by SAK and Katrine Schjoldager.*

Figure 3.2

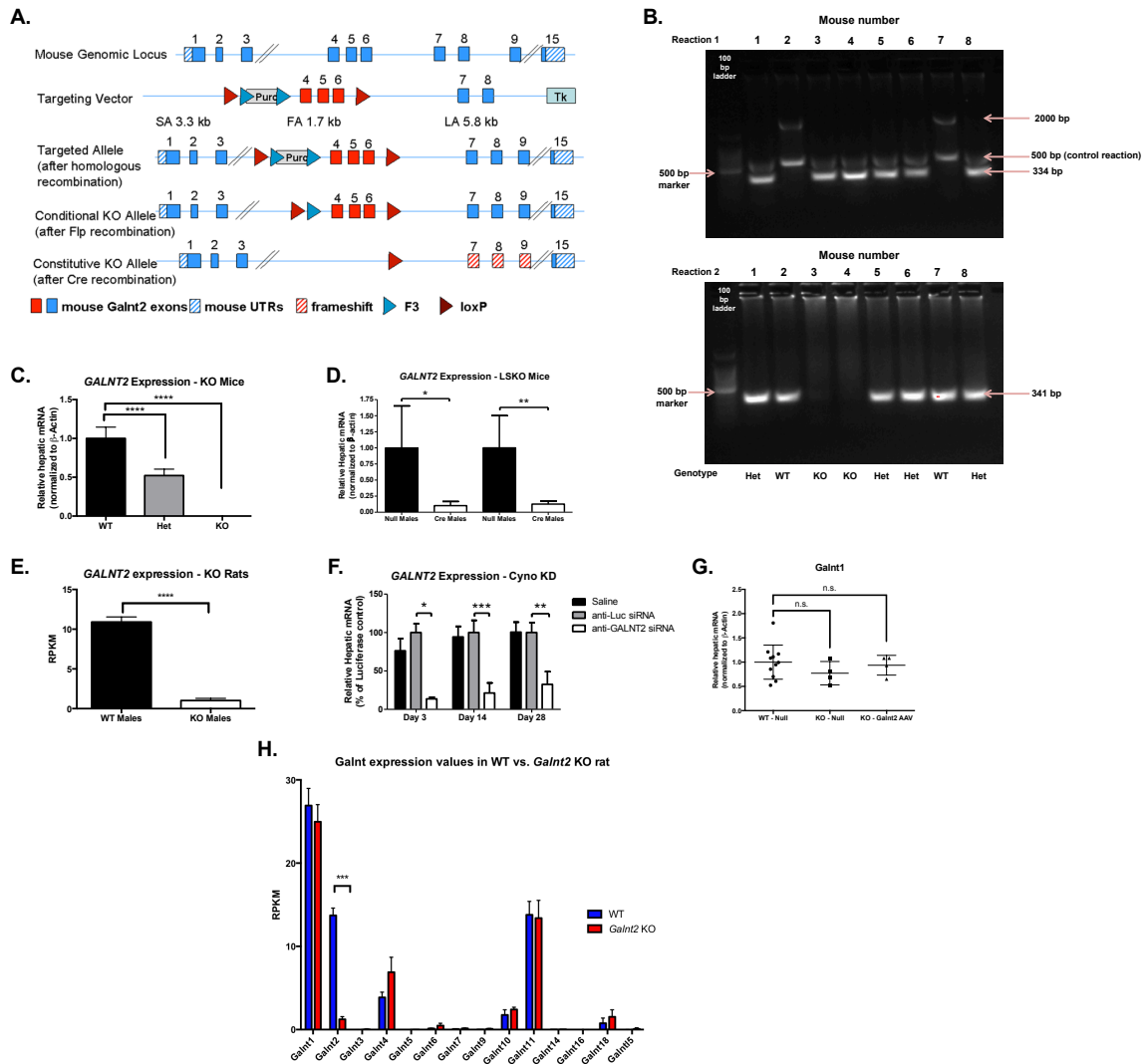


Figure 3.2: GALNT2 expression in mammalian models of GALNT2 deficiency. A.

Summary of *Galnt2* targeting strategy to excise exons 4–6 of mouse *Galnt2* on chromosome 8. Conditional KO mice were generated and bred to constitutive Cre-deleter mice to generate mice heterozygous for the deletion, which were bred to generate WT, heterozygous, and KO mice for experiments. **B.** Representative results from genotyping tail DNA from *Galnt2* WT, heterozygous and KO mice generated using

the strategy in (A). Two PCR reactions were used to identify the deletion of *Galnt2* exons 4-6 with primers that annealed uniquely to the *Galnt2* gene at flanking sites. **C.** Hepatic *Galnt2* mRNA expression in *Galnt2* WT, heterozygous, and complete KO mice as measured by quantitative RT-PCR. Cycle number of amplification (Cts) were normalized to that of actin. **D.** *Galnt2* mRNA expression in the livers of *Galnt2* LSKO mice (*Galnt2^{fl/fl}* mice treated with AAV-TBG-Cre vs. Null AAV-treated floxed control mice). Cts were normalized to those of actin. **E.** Quantification of *Galnt2* transcripts in the livers of *Galnt2* WT vs. KO rats by RNA sequencing. **F.** *GALNT2* mRNA expression in livers of cyno treated with saline, anti-luciferase siRNA or anti-*GALNT2* siRNA, quantitated using RT-PCR at 3, 14, and 28 days after a single administration of siRNA. *GALNT2* mRNA was normalized using 18s rRNA. **G.** Liver gene expression of *Galnt1*, the only other major GalNAc transferase expressed in liver, from WT mice administered Null AAV, *Galnt2* KO mice administered Null AAV, and *Galnt2* KO mice with reconstituted hepatic *Galnt2* by AAV. Expression was measured relative to that of actin from the same samples. **H.** RNA sequencing results for glycosyltransferase gene mRNA transcripts from livers of the *Galnt2* KO rats vs. WT controls (3 per group). For each glycosyltransferase gene, data is plotted as reads per kilobase of transcript per million mapped reads (RPKM) for the WT vs. *Galnt2* KO rats. Data is presented as mean values \pm S.D., with error bars showing S.D.. * P<0.05, ** P<0.01, *** P<0.001, Student's unpaired T-test. *Contributions: Data in panels A-D and G performed and analyzed by SAK. Data in panel E-F and H performed and analyzed by Katrine Schjoldager.*

Figure 3.3

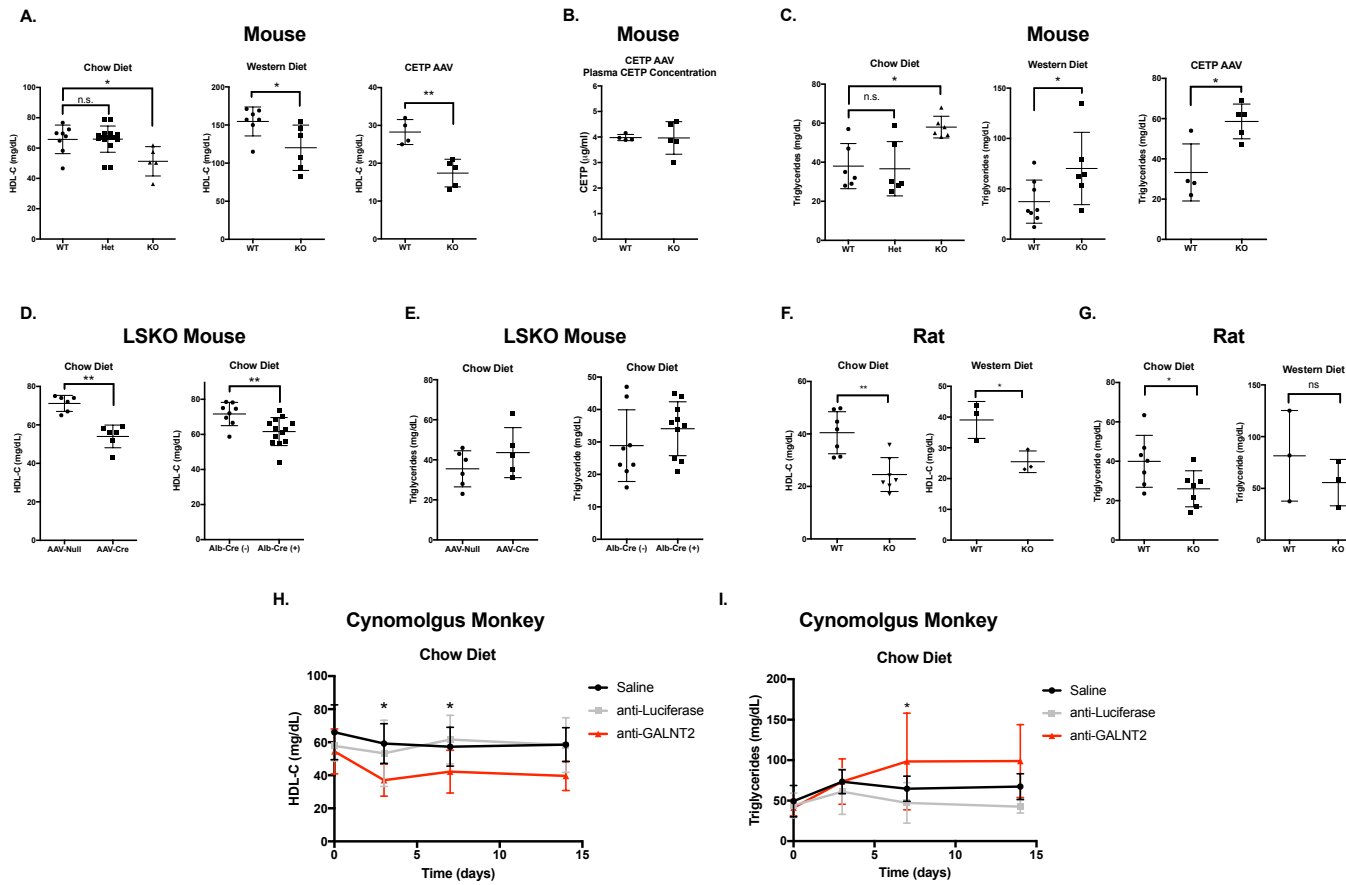


Figure 3.3: Plasma HDL cholesterol in mammalian models of *GALNT2* loss-of-function. A. Plasma HDL-C after a 4 hour fast from *Galnt2* WT, heterozygous (Het), and KO mice fed a chow diet (left), a chow diet after 4 weeks or administration of human CETP

AAV (center), or a Western-type diet for 3 weeks (right). **B.** Plasma CETP concentration after 4 weeks in mice from (A, right). **C.** Plasma triglycerides in mice from (A). **D.** Plasma HDL-C from *Galnt2* WT vs. liver-specific KO mice generated by AAV-Cre delivery in *Galnt2^{fl/fl}* mice (left) or crossing *Galnt2^{fl/fl}* with Alb-Cre transgenic mice (right). **E.** Plasma triglycerides in mice from (D). **F.** Plasma HDL-C after fasting for 12-16 hours from *Galnt2* WT vs. KO rats fed a chow diet (left) or 6 weeks of Western-type diet (right). **G.** Plasma triglycerides in rats from (F). **H.** Plasma HDL-C after overnight fasting in cynomolgus monkeys (cyno) treated with saline, anti-luciferase siRNA or anti-GALNT2 siRNA. **I.** Plasma triglycerides in cyno monkeys from (H). Data is presented as mean values \pm S.D.. * P<0.05, ** P<0.01, Student's unpaired T-test. *Contributions: Data in panels A-E performed and analyzed by SAK. Data in F-I generated by Katrine Schjoldager and Todd Kirchgessner and analyzed by SAK.*

Figure 3.4

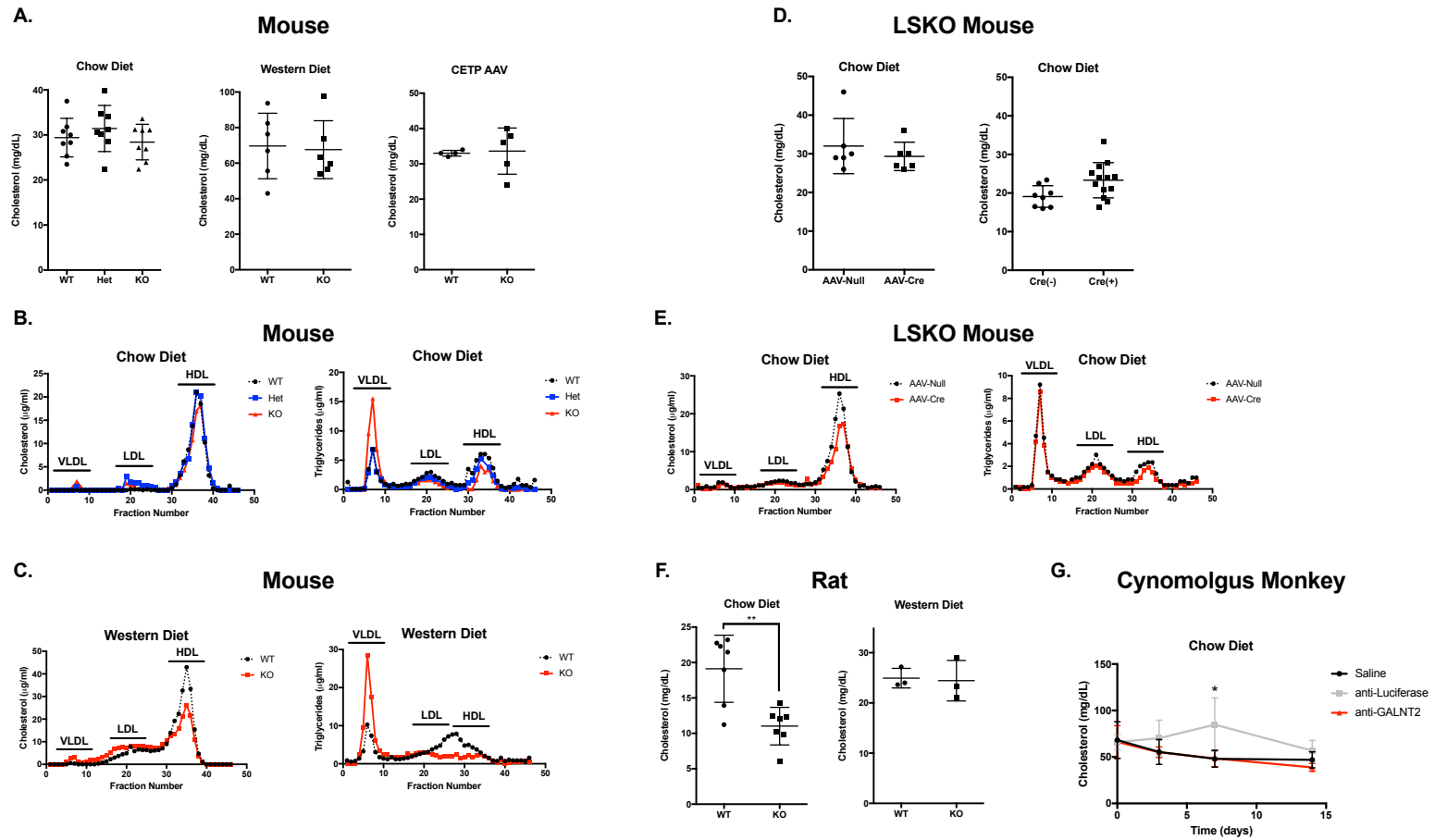


Figure 3.4: Non-HDL cholesterol in *GALNT2* loss-of-function models. A. Plasma non-HDL cholesterol (NonHDL-C) in WT, heterozygous (het), and complete KO mice fed a chow diet (left), WT vs. KO mice fed a Western diet for 3 weeks (center), and WT

vs. KO mice after 4 weeks of AAV CETP administration (right). All measurements were made from plasma from mice after 4 hours of fasting. **B.** Cholesterol (left) and triglycerides in FPLC-separated fractions from plasma from WT, Het, and KO mice fed a chow diet (4 hour fasted plasma was used for FPLC measurements). **C.** Cholesterol (left) and triglycerides (right) from FPLC-separated fractions of plasma of WT vs. KO mice fed a Western diet for 3 weeks. **D.** Plasma NonHDL-C in *Galnt2* liver-specific KO mice, generated by either administration of AAV-TBG-Cre to *Galnt2* floxed mice (left) or crossing Albumin-Cre transgenic mice with *Galnt2* floxed mice (right). Measurements were made on plasma from mice after 4 hours of fasting. **E.** Cholesterol (left) and triglycerides (right) from FPLC-separated plasma from *Galnt2* liver-specific KO mice. **F.** Plasma NonHDL-C in *Galnt2* WT vs. KO rats fed a chow diet or Western diet for 6 weeks. Measurements were from plasma of rats after 12-16 hours of fasting. **G.** LDL cholesterol in cynomolgus monkeys treated with saline, anti-luciferase control siRNA, or anti-GALNT2 siRNA for 14 days. All measurements indicate mean values \pm S.D. where applicable, with error bars showing S.D.. *P<0.05, **P<0.01, Student's unpaired T-test. For cyno measurements, comparisons were between the anti-luciferase and anti-GALNT2 groups. *Contributions: Data in panels A-E generated and analyzed by SAK; data in panels F-G generated by Katrine Schjoldager and Todd Kirchgessner and analyzed by SAK.*

Figure 3.5

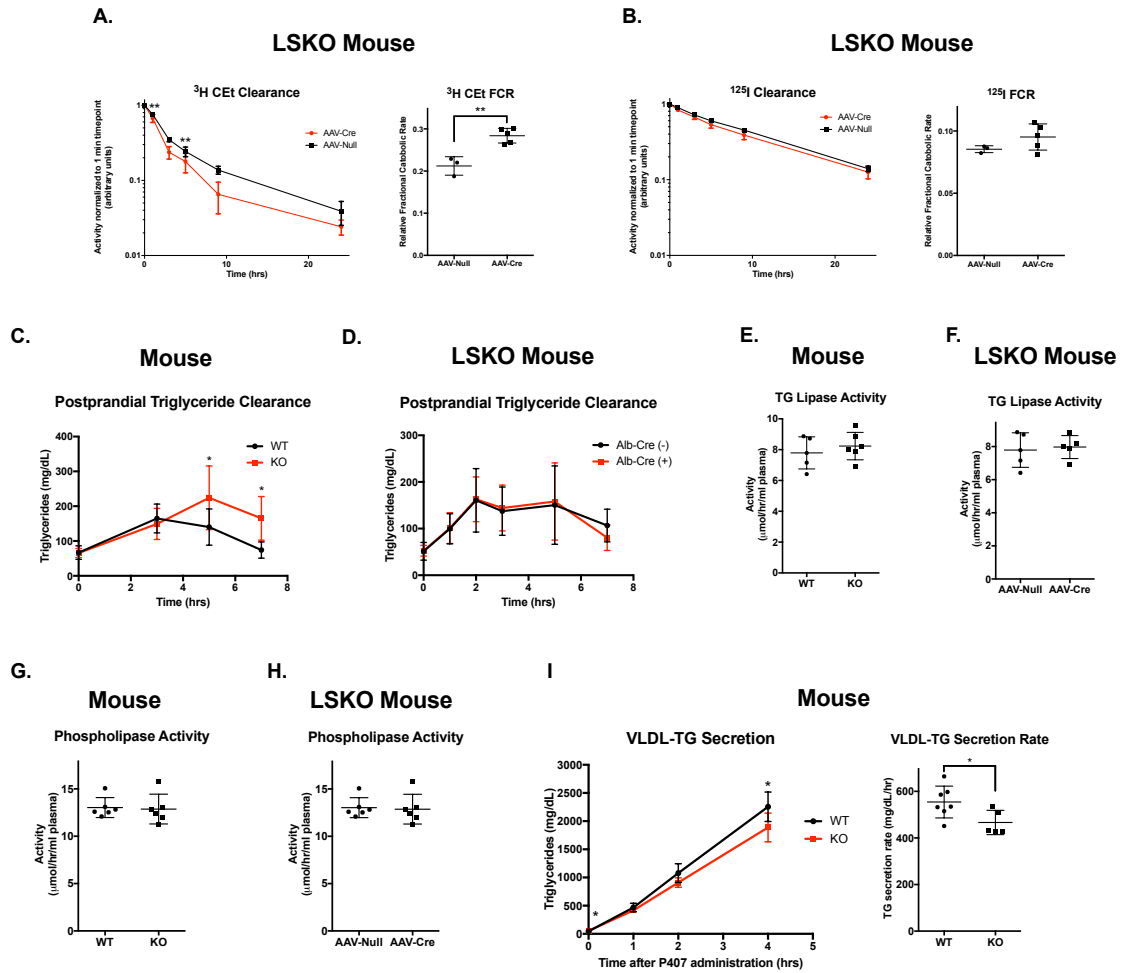


Figure 3.5: HDL and TG clearance in *Galnt2* KO mice. A. Plasma clearance of ³H-HDL-cholesteryl ester in *Galnt2* liver-specific KO mice vs. controls over 24 hours. Dual-labeled HDL was administered to the mice intravenously and blood was sampled for measurement of activity at the indicated timepoints. Clearance curves are shown on the left and fractional catabolic rates calculated through multiexponential modeling are shown on the right. **B.** Plasma clearance of ¹²⁵I-HDL-ApoA-I in *Galnt2* liver-specific KO mice vs. controls over 24 hrs. Clearance curves are shown on the left and fractional catabolic rates calculated through multiexponential modeling are shown on the right. **C.**

Postprandial TG clearance after oral olive oil gavage in *Galnt2* WT vs. complete KO mice. **D.** Postprandial TG clearance after olive oil gavage in WT (Cre-) vs. *Galnt2* liver-specific KO mice (Cre+). **E.** TG lipase activity from post-heparin plasma of *Galnt2* WT vs. complete KO mice. **F.** TG lipase activity from post-heparin plasma of *Galnt2* WT (AAV-Null) vs. liver-specific KO (AAV-Cre) mice. **G.** Phospholipase activity from post-heparin plasma of *Galnt2* WT and complete KO mice. **H.** Phospholipase activity from post-heparin plasma of *Galnt2* WT (AAV-Null) vs. liver-specific KO (AAV-Cre) mice. **I.** VLDL-TG secretion into plasma of WT vs. complete KO mice. X-axis indicates time after pluronic P407 administration by IP injection. Plasma TG were measured at each timepoint. Plasma TG levels over time are plotted on the left, and secretion rates as measured by the slopes of the curves are shown on the right. Data is presented as mean values \pm S.D., with error bars showing S.D.. *P<0.05, ** P<0.01, Student's unpaired T-test. *Contributions: All data in this figure generated and analyzed by SAK.*

Figure 3.6

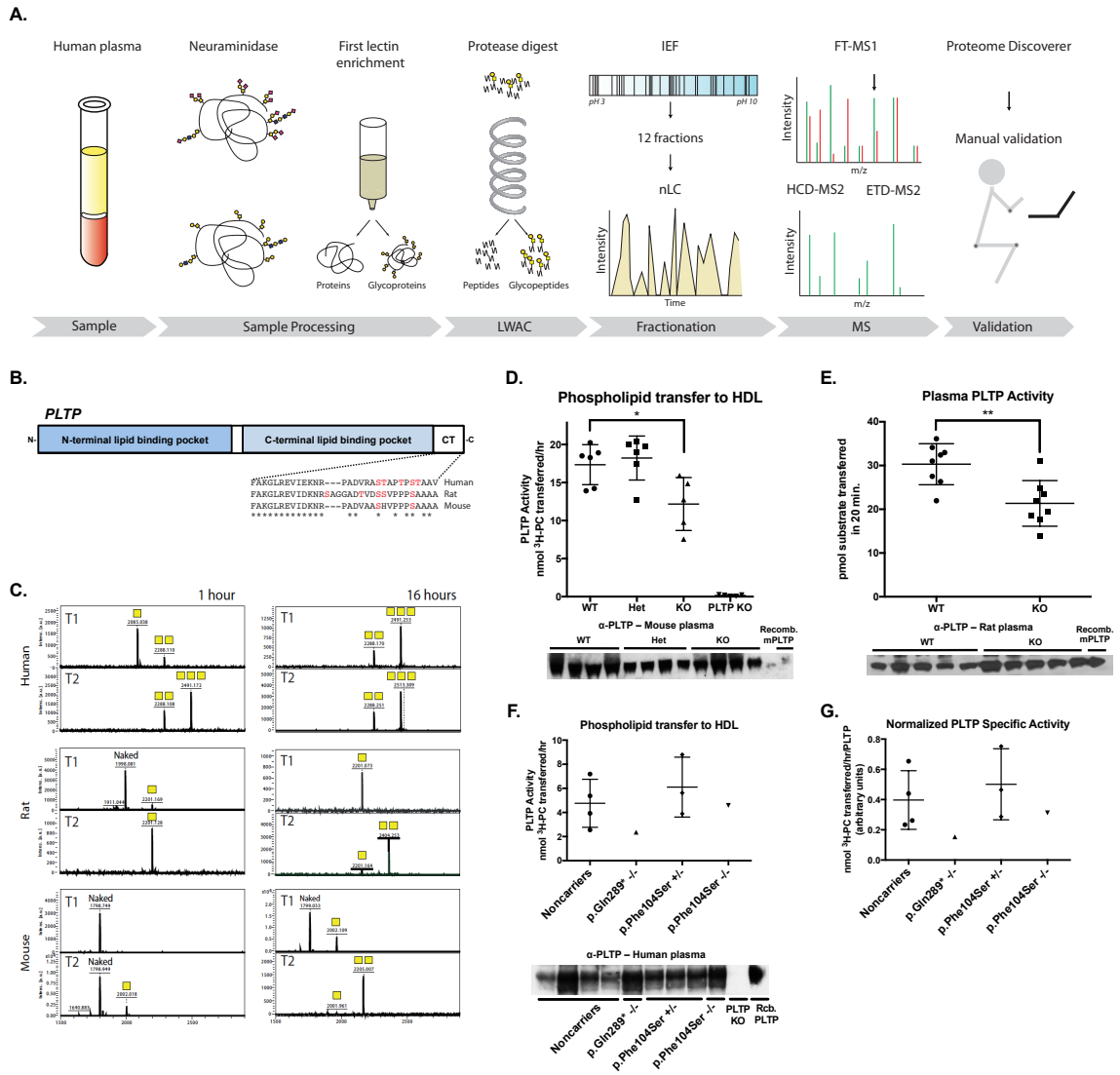


Figure 3.6: Differential glycoproteomics of tissues from *Galnt2* WT vs. KO rodents identifies PLTP as a candidate target of GalNAc-T2. A. Schematic of differential glycoproteomics approach to identify GalNAc-T2 targets from humans and rodent models. **B.** Phospholipid transfer protein (PLTP) schematic and cross-species alignment showing the C-terminal HDL-binding region. Asterisk shows conserved residues among mice, rats, and humans. Potential O-glycosylated residues (Ser and Thr) are indicated in

red. **C.** *In vitro* glycosylation assay of PLTP C-terminal peptides by recombinant human GalNAc-T1 and -T2 over 1 hour and 16 hours. Yellow squares denote the number of GalNAc residues on each peptide. **D.** (Top) Plasma PLTP activity from *Galnt2* WT, *Galnt2* heterozygous (Het), *Galnt2* KO (KO), and PLTP-deficient (PLTP KO) mice. (Bottom) Immunoblot of plasma PLTP from mice from activity assay above. **E.** (Top) Plasma PLTP activity from *Galnt2* WT vs. KO rats. (Bottom) Immunoblot of plasma PLTP from rats from activity assay above. **F.** (Top) Plasma PLTP activity from human control subjects and homozygotes for *GALNT2* nonsynonymous variants. (Bottom) Immunoblot of plasma PLTP from human participants from PLTP activity assay above. **G.** Plasma PLTP specific activity from human participants in (F). Specific activity was measured as the plasma activity from (F Top) normalized to the densitometric intensity for the immunoblot from (F bottom). Where applicable, data is presented as mean values \pm S.D.. ** P<0.01, ***P<0.001, Student's unpaired T-test. *Contributions: Data in C generated and analyzed by Katrine Schjoldager; data in panels D-G performed and analyzed by SAK using reagents and samples from SAK and provided by Katrine Schjoldager.*

Figure 3.7

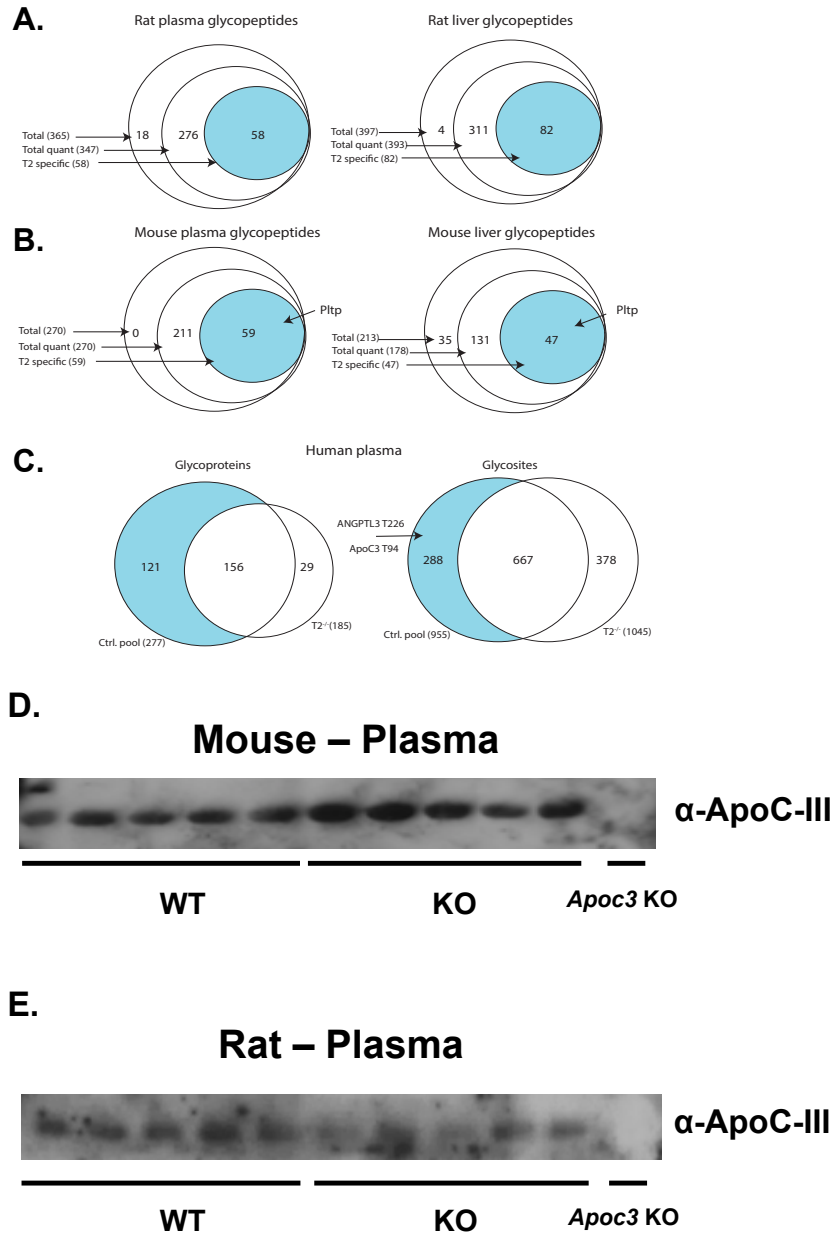


Figure 3.7: Overview of differential O-glycoproteomes from *GALNT2* deficient humans and rodents. A. & B. Quantitative differential O-glycoproteomes of rat (A) and mouse (B) plasma and liver showing the total number unambiguously identified glycopeptides, quantitated glycopeptides and isoform-specific glycopeptides and sites.

C. Venn diagrams showing the distribution of O-glycoproteins and O-glycosites in plasma from *GALNT2* p.Gln289* proband and a pool of nonaffected sex- and age-matched individuals. Glycoproteins and sites absent from the plasma of p.Gln289* proband constitute candidates for GalNAc-T2 specific activity. **D. & E.** Immunoblot of ApoC-III from plasma of *Galnt2* WT vs. KO mice (D) and rats (E) after separation of plasma proteins by SDS-PAGE. *Contributions: Data in panels A-C generated and analyzed by Katrine Schjoldager using samples from Katrine Schjoldager and some prepared by SAK; data in panels D-E generated and analyzed by SAK.*

Figure 3.8

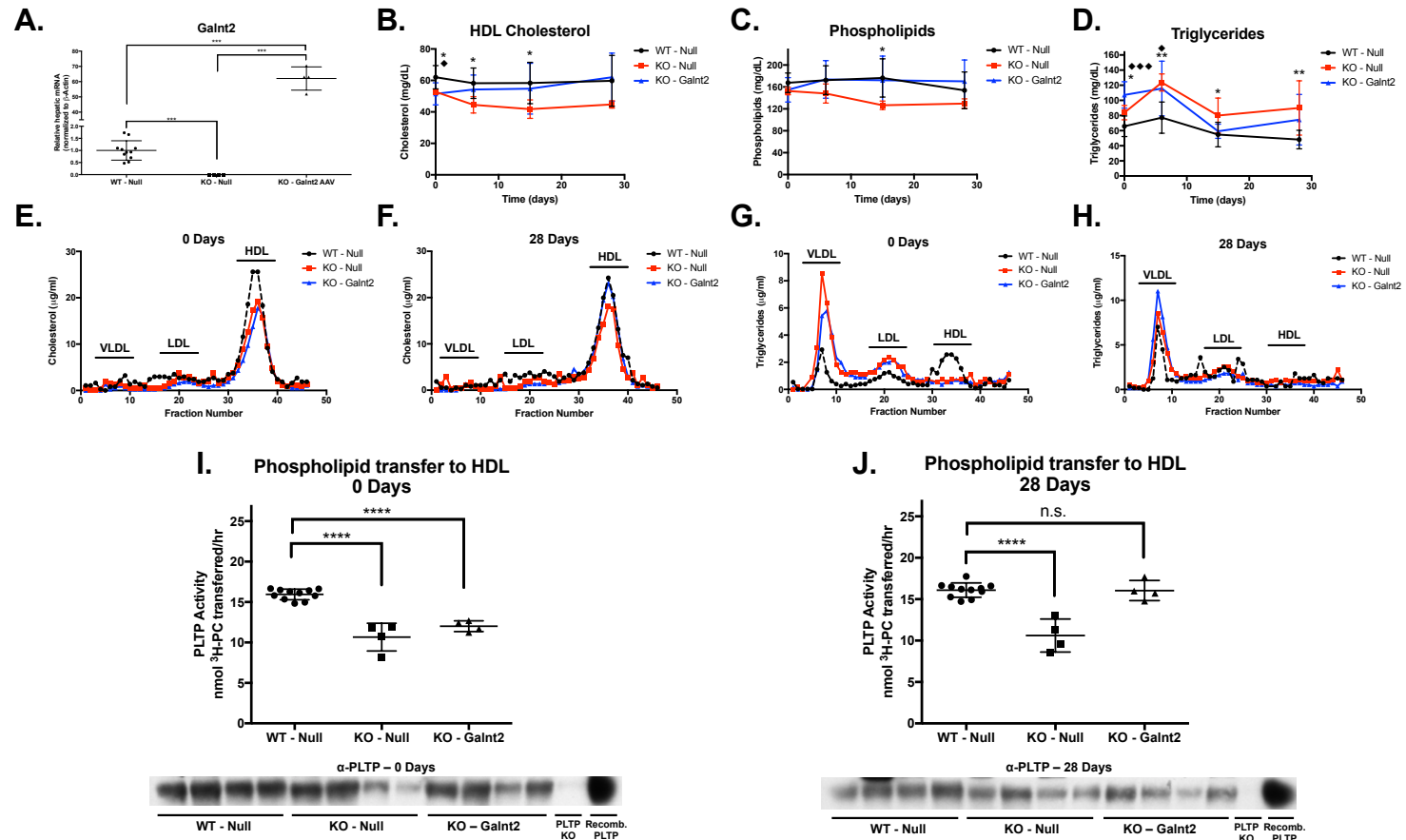


Figure 3.8: Reconstitution of hepatic *Galnt2* expression in *Galnt2* KO mice restores HDL-C and PLTP activity. A. Hepatic *Galnt2* expression in WT mice transduced with Null AAV (WT – Null), *Galnt2* KO mice transduced with Null AAV (KO – Null), and KO

mice transduced with *Galnt2* AAV (KO – *Galnt2*). ***P<0.001, Student's unpaired T-test. **B.** Plasma HDL-C in mice from (A). **C.** Plasma phospholipids from mice in (A). **D.** Plasma Triglycerides from mice in (A). For B-D, significance was determined at each timepoint. *P<0.05, **P<0.01, Student's unpaired T-test of comparison of WT – Null to KO – Null. ◆ P<0.05, ◆◆◆ P<0.001, Student's unpaired T-test of comparison of WT – Null to KO – *Galnt2*. **E.** Day 0 cholesterol in fractions after FPLC separation from pooled plasma from each of the groups in (A). **F.** Day 28 cholesterol from FPLC fractions of pooled plasma from groups in (A). **G.** Day 0 triglycerides in FPLC fractions from pooled plasma from groups in (A). **H.** Day 28 triglycerides in FPLC fractions from pooled plasma from groups in (A). **I.** (Top) Plasma PLTP activity from Day 0 plasma from mice in (A). (Bottom) Immunoblot of plasma PLTP from PLTP activity assay above. **J.** (Top) Plasma PLTP activity from Day 28 plasma from mice in (A). (Bottom) Immunoblot of plasma PLTP from PLTP activity assay above. For I-J, ****P<0.0001, Student's unpaired T-test for comparisons between the indicated groups. Data is presented as mean values ± S.D.. *Contributions: All data in this figure generated and analyzed by SAK.*

Figure 3.9

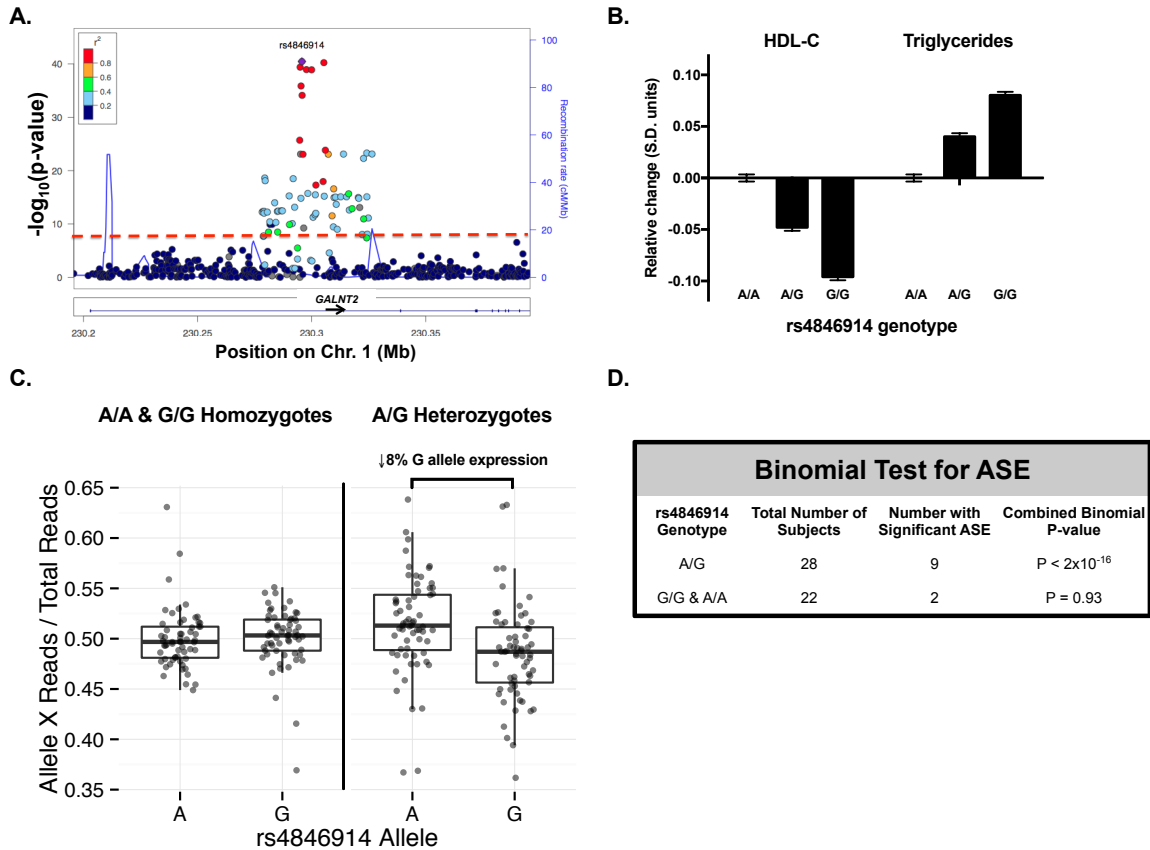


Figure 3.9: Common variants in *GALNT2* associated with HDL-C confer allelic imbalance of *GALNT2* expression in human liver. A. Manhattan plot of association of *GALNT2* SNPs with HDL-C from the Global Lipids Genetics Consortium (GLGC) GWAS. Colors indicate the amount of linkage disequilibrium between plotted SNPs. Purple colored circle indicates lead SNP rs4846914 in *GALNT2* intron 1. **B.** Relative change in HDL-C and TG (in S.E. units) per copy of rs4846914 allele from the GLGC GWAS. **C.** Allele-specific expression (ASE) of the rs4846914 SNP on *GALNT2* expression from 72 human liver samples as measured by RNA-Seq. The fraction of *GALNT2* reads arising from either the A or G allele for the rs4846914 SNP relative to total *GALNT2* reads is

plotted. Triangles indicate samples with significant ASE while circles indicate nonsignificant ASE. **D.** One-sided binomial test used to assess ASE for the distribution of *GALNT2* transcripts arising from rs4846914 A vs. G allele in samples with A/A or A/G genotype for each sample. Data shows mean \pm S.D.. *Contributions: Data in panels A-B generated previously by consortia (publicly available) and analyzed here by SAK; data in panels C-D generated and analyzed by SAK, Casey Brown, and YoSon Park.*

Figure 3.10

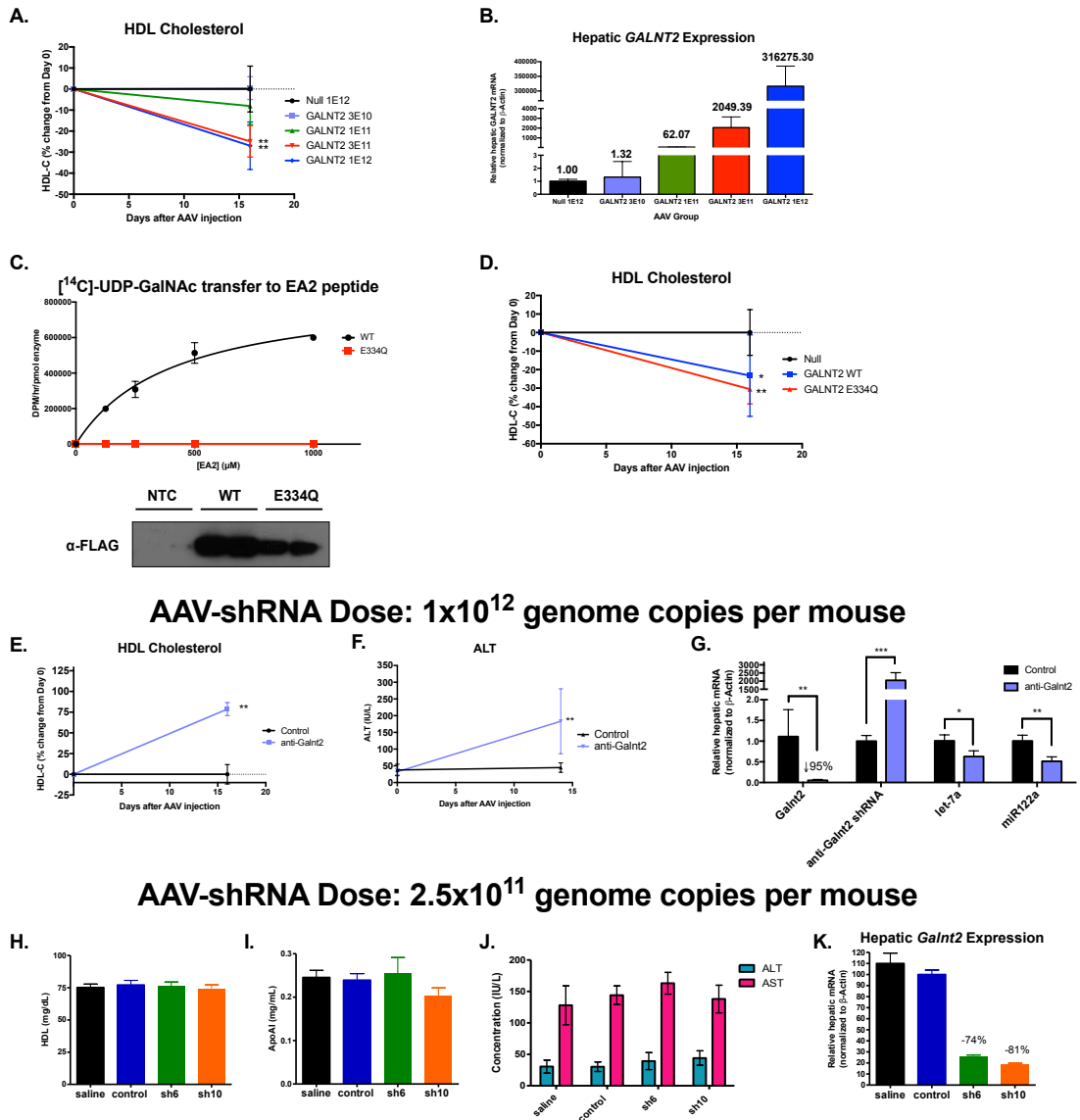


Figure 3.10: Investigation of AAV-mediated *Galnt2* overexpression and knockdown. **A.** Dose-titration experiment of AAV-*Galnt2* at indicated doses and impact on HDL-C. Measurements indicate percent changes in HDL-C relative to Day 0 (pre-injection) HDL-C levels. ** $P < 0.01$, indicated group vs. Null 1E12 group, Student's unpaired t-test for Day 14 data. **B.** Hepatic *Galnt2* expression in mice from (A). Values

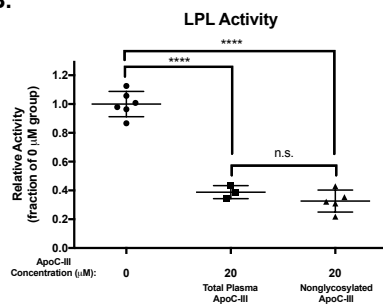
above bars indicate relative fold-change compared to Null 1E12 group. **C.** *In vitro* glycosyltransferase activity of WT vs. E334Q mutant GalNAc-T2. Recombinant FLAG-tagged enzymes were made through transfection in COS7 cells and immunoprecipitated. Immunoprecipitates were used in an assay to measure transfer of [¹⁴C]-UDP-GalNAc to a peptide substrate over 1 hour at 37 °C. Immunoblot of equal volumes of immunoprecipitates is shown below kinetic assay curves. **D.** Impact of WT vs. E334Q *Galnt2* AAV overexpression on HDL-C (change from baseline after 14 days). *P<0.05, **P<0.01, indicated group vs. Null 1E12, Student's unpaired T-test. **E.** Plasma HDL-C in mice treated with either an AAV-shRNA with no murine target (Control) vs. that of mice treated with an anti-*Galnt2* AAV shRNA at 0 and 14 days after AAV administration. This experiment is a representative experiment of 7 in total that evaluated these anti-*Galnt2* AAV-shRNAs at a dose of 1x10¹² genome copies (GC) per mouse. Similar results were obtained using another shRNA against *Galnt2* at this dose and at a dose of 3x10¹¹ GC/mouse. **F.** Plasma ALT levels in mice from (A). **G.** Liver gene expression from mice in (A). **H.** Plasma HDL-C at 14 days after treatment of mice with saline, a control AAV-shRNA, or 2 different AAV-shRNAs with sequences targeting *Galnt2*. This study, performed by Merck, was done at a dose of 2.5x10¹¹ GC/mouse. **I.** Plasma ApoA-I levels measured by autoanalyzer at 14 days after AAV administration from mice in (D). **J.** Plasma ALT and AST levels at 14 days after AAV administration from mice in (D). **K.** Hepatic *Galnt2* expression in mice from (D). Data is presented as mean values ± S.D., where error bars show S.D.. *P<0.05, ** P<0.01, Student's unpaired T-test. *Contributions: Data in all panels generated by SAK, Andrew Edmondson and Merck Research Laboratories and analyzed by SAK.*

Figure 3.11

A.

Myocardial Infarction Genetics Exome Sequencing (MIExSeq) Study												
Trait	Position	rsID	Gene	Ref. allele	Alt. allele	Amino Acid Change	No. of Participants	No. of Carriers	MAF	Beta	S.E.	P-value
HDL-C	Chr1:230386238	rs78164071	GALNT2	A	C	D314A	9381	73	0.004	-1.61	1.58	0.308
HDL-C	Chr18:47109955	rs77960347	LIPG	A	G	N396S	9381	208	0.01	2.2	0.95	0.02
HDL-C	Chr12:125284671	rs74830677	SCARB1	G	A	P376L	9381	8	0.0004	11.8	4.91	0.02
Pakistan Risk of Myocardial Infarction Study (PROMIS)												
Trait	Position	rsID	Gene	Ref. allele	Alt. allele	Amino Acid Change	No. of Participants	No. of Carriers	MAF	Beta	S.E.	P-value
HDL-C	Chr1:230386238	rs78164071	GALNT2	A	C	D314A	6864	25	0.00182	-0.4262	1.954	0.827
HDL-C	Chr16:56996288	rs708272	CETP	G	A	intronic (c.118+279G>A)	6864	4878	0.466	1.683	0.167	1.11x10 ⁻²³

B.



C.

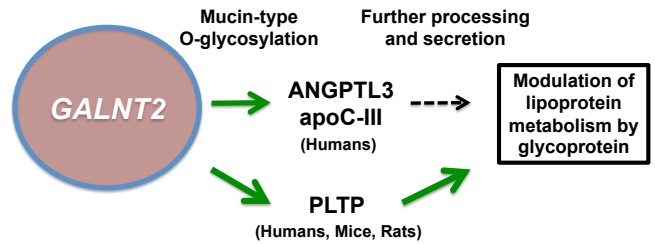


Figure 3.11: Relationship of the previously reported D314A variant with HDL-C. A.

Association of *GALNT2* D314A with HDL-C in the MIExSeq rare variant association study (top) and the PROMIS study (bottom). For each association test, representative variants significantly associated with HDL-C for each respective study are given below the results for the *GALNT2* D314A variant. Each study was 100% powered to detect a 1 S.D. change in HDL-C with significance level $\alpha=0.05$.

B. *In vitro* LPL activity in the presence of 0 μM ApoC-III, 20 μM total plasma ApoC-III, and 20 μM nonglycosylated ApoC-III. Data is shown as mean values \pm S.D.. **** $P<0.0001$, Student's unpaired T-test.

C. Model for proposed targets of *GALNT2* that modulate lipoprotein metabolism. Green lines show activating roles of GalNAc-T2 mediated O-glycosylation. Black dotted line indicates yet unclear regulatory effect of the glycosylation of indicated targets.

Contributions: Data generated and analyzed by SAK, Gina Peloso, and Wei Zhao.

Table 3.1

Table 3.1: Glycopeptides identified from ANGPTL3, ApoC-III, and PLTP from Human Controls and GALNT2 p.Gln289* Homozygote				
Human Control Plasma Glycopeptides Identified from Lectin Weak-Affinity Glycoproteomics				
Gene	Accessions	Protein Name (UniProt)	Sequence	Unambiguous Glycan Addition
ANGPTL3	Q9Y5C1	Angiotensin-converting enzyme 2	TSIQEPTEISLSSKPR	S10(Hex1HexNAc1); S12(Hex1HexNAc1)
			RTSIQEPTEISLSSKPR	S11(Hex1HexNAc1); S13(Hex1HexNAc1)
			RTSIQEPTEISLSSKPR	S11(Hex1HexNAc1); S13(Hex1HexNAc1)
			SRIDQDNSSFDSLSPK	S14(Hex1HexNAc1)
			TTPFLQLNEIR	T2(Hex1HexNAc1)
			TSIQEPTEISLSSKPR	T7(Hex1HexNAc1)
			TSIQEPTEISLSSKPR	T7(Hex1HexNAc1)
			TSIQEPTEISLSSK	T7(Hex1HexNAc1)
			TSIQEPTEISLSSKPR	T7(Hex1HexNAc1)
			TSIQEPTEISLSSKPR	T7(Hex1HexNAc1)
			TSIQEPTEISLSSKPR	T7(Hex1HexNAc1)
			RTSIQEPTEISLSSKPR	T8(Hex1HexNAc1)
			RTSIQEPTEISLSSKPR	T8(Hex1HexNAc1)
RTSIQEPTEISLSSKPR	T8(Hex1HexNAc1)			
APOC3	P02656	Apolipoprotein C-III	WDLDPVPRPTSAVAA	T10(Hex1HexNAc1)
			PEVRPTSAVAA	T6(Hex1HexNAc1)
			PEVRPTSAVAA	T6(Hex1HexNAc1)
			DPEVRPTSAVAA	T7(Hex1HexNAc1)
			LDPEVRPTSAVAA	T8(Hex1HexNAc1)
			LDPEVRPTSAVAA	T8(Hex1HexNAc1)
			LDPEVRPTSAVAA	T8(Hex1HexNAc1)
			LDPEVRPTSAVAA	T8(Hex1HexNAc1)
PLTP	P55058	Phospholipid transfer protein	ASTAPTPTAAV	T3(Hex1HexNAc1)
			ASTAPTPTAAV	T6(Hex1HexNAc1)
GALNT2 p.Gln289* Homozygote Plasma Glycopeptides Identified from Lectin Weak-Affinity Glycoproteomics				
Gene	Accessions	Protein Name (UniProt)	Sequence	Unambiguous Glycan Addition
ANGPTL3	Q9Y5C1	Angiotensin-converting enzyme 2	RTSIQEPTEISLSSKPR	S11(Hex1HexNAc1); S13(Hex1HexNAc1)
			TSIQEPTEISLSSKPR	S12(Hex1HexNAc1)
			RTSIQEPTEISLSSKPR	S13(Hex1HexNAc1)
			SRIDQDNSSFDSLSPK	S14(Hex1HexNAc1)
			TSIQEPTEISLSSKPR	T7(Hex1HexNAc1)
			TSIQEPTEISLSSKPR	T7(Hex1HexNAc1)
			RTSIQEPTEISLSSKPR	T8(Hex1HexNAc1)
APOC3	P02656	Apolipoprotein C-III	PEVRPTSAVAA	S7(Hex1HexNAc1)
			PEVRPTSAVAA	T6(Hex1HexNAc1)
PLTP	P55058	Phospholipid transfer protein	ASTAPTPTAAV	T3(Hex1HexNAc1)
			ASTAPTPTAAV	T3(Hex1HexNAc1)
			ASTAPTPTAAV	T3(Hex1HexNAc1)
			ASTAPTPTAAV	T3(Hex1HexNAc1); T6(Hex1HexNAc1)
			ASTAPTPTAAV	T6(Hex1HexNAc1)

Table 3.1: Glycopeptides identified from ANGPTL3, ApoC-III, and PLTP from Human Controls and GALNT2 p.Gln289* Homozygote. Weak lectin-affinity glycoproteomics was used to identify glycopeptides in ANGPTL3, ApoC-III, and PLTP from of a control pool of plasma and plasma from the participant homozygous for the p.Gln289* variant. Peptide sequences and glycan modifications for each glycopeptide identified are provided for each of the samples generated after alignment to a human proteome database and glycan site assignment as described in the Materials and Methods. *Contributions: All data in this table generated and analyzed by Katrine Schjoldager.*

Table 3.2

Table 3.2: Differential Glycoproteomes of <i>Galnt2</i> KO vs. WT Rats					
Differential Plasma Glycoproteomes					
Gene	Accessions	Protein Name (UniProt)	Sequence	Unambiguous Glycan Addition	Fold-difference KO vs. WT for Glycopeptide [log₁₀ (fold difference)]
<i>Adamtsl4</i>	Q4FZU4	ADAMTS-like protein 4	qPHPFsPVTR	S6(Hex1HexNAc1)	-5.00
<i>A1i3</i>	P14046	Alpha-1-inhibitor 3	gmYEslPVVAVk	S5(Hex1HexNAc1)	-5.00
<i>A1i3</i>	P14046	Alpha-1-inhibitor 3	eDNSIHWERPQkPtK	T14(Hex1HexNAc1)	-5.00
<i>Abca5</i>	Q8CF82	ATP-binding cassette sub-family A member 5	vSDIELSPMDkSILSNLILG YtPVTNTTSSVmQRVSTD HLPDVLVTEEYASEK	T22(Hex1HexNAc1)	-5.00
<i>F10</i>	Q63207	Coagulation factor X	vGPISelPRLTHPPY	T4(Hex1HexNAc1);T7(Hex1HexNAc1)	-5.00
<i>Fetub</i>	Q9QX79	Fetuin-B	nTAPTSPSITAPR	S6(Hex1HexNAc1)	-5.00
<i>Fetub</i>	Q9QX79	Fetuin-B	tAPISSPSITAPR	T4(Hex1HexNAc1)	-5.00
<i>Fetub</i>	Q9QX79	Fetuin-B	nTAPTSPSITAPR	T5(Hex1HexNAc1)	-5.00
<i>Fetub</i>	Q9QX79	Fetuin-B	nTAPISSPSITAPR	T5(HexNAc)	-5.00
<i>Fetub</i>	Q9QX79	Fetuin-B	nTAPTSPSITAPR	T5(HexNAc)	-5.00
<i>Fga</i>	P06399	Fibrinogen alpha chain	msPVPDLVPGSFk	S2(Hex1HexNAc1)	-5.00
<i>Fga</i>	P06399	Fibrinogen alpha chain	mHPELGsFYDSR	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	Q68FP1	Gelsolin	vSETRPsTmVVEHPEFLk	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	Q68FP1	Gelsolin	vSETRPsTmVVEHPEFLk	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	Q68FP1	Gelsolin	vSETRPsTmVVEHPEFLk	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	Q68FP1	Gelsolin	vSETRPsTmVVEHPEFLk	S7(Hex1HexNAc1)	-5.00
<i>Hrg</i>	Q99PS8	Histidine-rich glycoprotein	ISPIncDASKPLAEk	T4(Hex1HexNAc1)	-5.00
<i>Ighd</i>	P01883	Ig delta chain C region (Fragment)	hYIEAtkPTATk	T6(Hex1HexNAc1)	-5.00
<i>Ighd</i>	P01883	Ig delta chain C region (Fragment)	nHYIEAtkPTATk	T7(Hex1HexNAc1)	-5.00
<i>Il1r2</i>	P43303	Interleukin-1 receptor type 2	vsESPVTSEk	S2(Hex1HexNAc1)	-5.00
<i>Csf1</i>	Q8JZQ0	Macrophage colony-stimulating factor 1	tEGSSLLPsDLPLR	S9(Hex1HexNAc1)	-5.00
<i>Mug1</i>	Q03627	Murinoglobulin-1	sAPAVESELSPR	S1(HexNAc)	-5.00
<i>Pam</i>	P14925	Peptidyl-glycine alpha-amidating monooxygenase	fHQLESILRPAESR	T7(Hex1HexNAc1)	-5.00
<i>Serping1</i>	Q6P734	Plasma protease C1 inhibitor	dSVPERDGsR	S9(Hex1HexNAc1)	-5.00
<i>Pigr</i>	P15083	Polymeric immunoglobulin receptor	gSPHINPtDANAR	T8(Hex1HexNAc1)	-5.00
<i>F2</i>	P18292	Prothrombin	tTDAEFHTFFDER	T1(Hex1HexNAc1)	-5.00
<i>Alb</i>	P02770	Serum albumin	kQtALAELVk	T3(Hex1HexNAc1)	-5.00
<i>Spep</i>	Q63638	Striated muscle-specific serine/threonine-protein kinase	vASPPPGAsEkRVPSAR	S9(HexNAc)	-5.00

Differential Liver Glycoproteomes					
Gene	Accessions	Protein Name (UniProt)	Sequence	Unambiguous Glycan Addition (ETD)	Fold-difference KO vs. WT for Glycopeptide [log10 (fold difference)]
<i>Aebp1</i>	A2RUV9	Adipocyte enhancer-binding protein 1	FSTVAPLETPER	S2(Hex1HexNAc1); T3(Hex1HexNAc1)	-5.00
<i>Ahsg</i>	P24090	Alpha-2-HS-glycoprotein	FSPVASVESASGEVLHSPK	S17(Hex1HexNAc1)	-5.00
<i>Anpep</i>	P15684	Aminopeptidase N	DLSVIPVINR	S3(Hex1HexNAc1)	-5.00
<i>ApoE</i>	P02650	Apolipoprotein E	IQASVATNSIASTTVPLENQ	T13(Hex1HexNAc1)	-5.00
<i>St6gal1</i>	P13721	Beta-galactoside alpha-2,6-sialyltransferase 1	MGSASQVVFNSK	S3(Hex1HexNAc1)	-5.00
<i>St6gal1</i>	P13721	Beta-galactoside alpha-2,6-sialyltransferase 1	MGSASQVVFNSK	S5(Hex1HexNAc1)	-5.00
<i>St6gal1</i>	P13721	Beta-galactoside alpha-2,6-sialyltransferase 1	EDIPILSYHRVTAK	T12(Hex1HexNAc1)	-5.00
<i>Adar</i>	P55266	Double-stranded RNA-specific adenosine deaminase	DSGQPEELSNCPME EDPEKPAESQPPSS SATSL	S9(HexNAc); S23(HexNAc)	-5.00
<i>Man1b1</i>	B2GUY0	Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase	AEASIKPLFLASR	S4(Hex1HexNAc1)	-5.00
<i>Man1b1</i>	B2GUY0	Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase	AEASIKPLF	S4(Hex1HexNAc1)	-5.00
<i>Fetub</i>	Q9QX79	Fetuin-B	TAPTSSPSITAPR	T4(Hex1HexNAc1)	-5.00
<i>Fetub</i>	Q9QX79	Fetuin-B	NTAPTSSPSITAPR	T5(Hex1HexNAc1)	-5.00
<i>Fetub</i>	Q9QX79	Fetuin-B	NTAPTSSPSITAPR	T5(Hex1HexNAc1)	-5.00
<i>Fgb</i>	P14480	Fibrinogen beta chain	ATTDSKVDLSIAR	S5(Hex1HexNAc1)	-5.00
<i>Fstl1</i>	Q62632	Follistatin-related protein 1	SVSPSASPVVCYQANR	S3(Hex1HexNAc1)	-5.00
<i>Gsn</i>	Q68FP1	Gelsolin	VSETRPSTMVVEHPEFLK	T8(Hex1HexNAc1)	-5.00
<i>Golm4</i>	Q5BJK8	Golgi integral membrane protein 4	AAHLLDGRPQAEIEH STKAATNFR	S16(Hex1HexNAc1)	-5.00
<i>Mug1</i>	Q03626-2	Isoform 2 of Murinoglobulin-1	GGEFEMMPLGVNKS PLPK	S14(Hex1HexNAc1)	-5.00
<i>Fgg</i>	P02680-2	Isoform Gamma-A of Fibrinogen gamma chain	LSIGDGGQHHMGGSKVQVGM	S14(Hex1HexNAc1)	-5.00
<i>Kng1</i>	P08934	Kininogen-1	STTVSPSYIAR	T3(Hex1HexNAc1)	-5.00
<i>Mxi1</i>	O09015	Max-interacting protein 1	MDSIGSTISSDR	S9(HexNAc);S10(HexNAc)	-5.00
<i>Mme</i>	P07861	Nephrilysin	DLQNLMSWRFIMDL VSSLSRNYK	S16(HexNAc);S17(Hex1HexNAc1);S19(Hex1HexNAc1)	-5.00
<i>Nid1</i>	P08460	Nidogen-1 (Fragment)	GYDPHPNVPRTLAP SY	T11(Hex1HexNAc1)	-5.00
<i>Nucb1</i>	Q63083	Nucleobindin-1	AAPHQEDNQATETP DTGLYYHRY	T11(Hex1HexNAc1)	-5.00
<i>Nucb2</i>	Q9JI85	Nucleobindin-2	KFQQGIAPSGPAGEL K	S9(Hex1HexNAc1)	-5.00
<i>Pigr</i>	P15083	Polymeric immunoglobulin receptor	GSPHINPTDANAR	T8(Hex1HexNAc1)	-5.00
<i>Pigr</i>	P15083	Polymeric immunoglobulin receptor	GSPHINPTDANAR	T8(Hex1HexNAc1)	-5.00
<i>Pcolce</i>	O08628	Procollagen C-endopeptidase enhancer 1	VQPVEKPEGSPATQ ATPVAPDAPSITCPK	S10(Hex1HexNAc1)	-5.00

<i>Lman1</i>	Q62902	Protein ERGIC-53	VSGVQHPGSAGVYE TTQHFMDIK	T15(Hex1HexNAc1)	-5.00
<i>Smo</i>	P97698	Smoothened homolog	SAGGSARRNAPVTS PPPLLSHCGR	S5(HexNAc);T13(Hex1HexNAc1); S14(HexNAc)	-5.00
<i>Slc9a1</i>	P26431	Sodium/hydrogen exchanger 1	LNPTASTIRGSEPPR ER	T4(Hex1HexNAc1);S11(Hex1He xNAc1)	-5.00
<i>Knt2</i>	P08932	T-kininogen 2	AGPAPDHQAEASTV T	T13(Hex1HexNAc1)	-5.00
<i>Knt2</i>	P08932	T-kininogen 2	AGPAPDHQAEASTV TP	T13(Hex1HexNAc1)	-5.00
<i>Knt2</i>	P08932	T-kininogen 2	AGPAPDHQAEASTV TP	T13(Hex1HexNAc1)	-5.00
<i>Fn1</i>	P04937	Fibronectin	GAAEPSDGTGHT YNQYTQR	S6(Hex1HexNAc1)	-5.00

Table 3.2: Differential Glycoproteomes of *Galnt2* KO vs. WT Rats. Quantitative differential lectin-affinity glycoproteomics was used to identify glycopeptides differentially observed in plasma (top) and livers (bottom) from *Galnt2* KO vs. WT rats. Peptide sequences and glycan modifications for each glycopeptide identified are provided for each of the samples generated after alignment to a rat proteome database and glycan site assignment as described in the Materials and Methods. Fold-differences in glycopeptide levels in WT vs. KO models were determined by relative quantification of light vs. medium chain dimethyl isotopomeric labels in the WT vs. KO samples, respectively. *Contributions: All data in this table generated and analyzed by Katrine Schjoldager.*

Table 3.3

Table 3.3: Differential Glycoproteomes of *Galnt2* KO vs. WT Mice

Differential Plasma Glycoproteomes					
Gene	Accessions	Protein Name (UniProt)	Sequence	Unambiguous Glycan Addition	Fold-difference KO vs. WT for Glycopeptide [log ₁₀ (fold difference)]
<i>B4galnt1</i>	Q09200	Beta-1,4 N-acetylgalactosaminyltransferase 1	YRYPGSLDQSQVAK	S6(HexNAc)	-5.00
<i>Bmp3</i>	Q8BHE5	Bone morphogenetic protein 3	ATGGSPSPDLRPHDK	S5(Hex1HexNAc1)	-5.00
<i>Bmp3</i>	Q8BHE5	Bone morphogenetic protein 3	ATGGSPSPDLRPHDK	ND	-5.00
<i>Ccdc97</i>	Q9DBT3	Coiled-coil domain-containing protein 97	EEFTSRMHQR	T4(HexNAc); S5(HexNAc)	-5.00
<i>Cd55</i>	Q61475	Complement decay-accelerating factor, GPI-anchored	VSATQHVPVTK	T4(Hex1HexNAc1); T10(Hex1HexNAc1)	-5.00
<i>Cd55</i>	Q61475	Complement decay-accelerating factor, GPI-anchored	VSATQHVPVTK	T4(Hex1HexNAc1); T10(Hex1HexNAc1)	-5.00
<i>Ecm1</i>	Q61508	Extracellular matrix protein 1	VDHSVTLHDPPLFEEQR	T6(Hex1HexNAc1); S7(Hex1HexNAc1)	-5.00
<i>Ecm1</i>	Q61508	Extracellular matrix protein 1	NVALVAGDTGNATGLGEQGPTTR	T21(HexNAc)	-5.00
<i>Ecm1</i>	Q61508	Extracellular matrix protein 1	EVKPHLAGQLPPEPR	T6(Hex1HexNAc1)	-5.00
<i>Extl1</i>	Q9JKV7	Exostosin-like 1	ITSTHASACLWDR	ND	-5.00
<i>F10</i>	O88947	Coagulation factor X	VGPTAETPR	T4(HexNAc)	-5.00
<i>F10</i>	O88947	Coagulation factor X	VGPTAETPR	T4(HexNAc)	-5.00
<i>Fam20c</i>	Q5MJS3	Extracellular serine/threonine protein kinase Fam20C	VPPPGPSGDGSLAK	S7(Hex1HexNAc1)	-5.00
<i>Fam20c</i>	Q5MJS3	Extracellular serine/threonine protein kinase Fam20C	VPPPGPSGDGSLAK	S7(Hex1HexNAc1)	-5.00
<i>Fetub</i>	Q9QXC1	Fetuin-B	NTAPTSSPSVTAPR	T2(HexNAc)	-5.00
<i>Fn1</i>	P11276	Fibronectin	LGVRSQGGEAPR	S6(Hex1HexNAc1)	-5.00
<i>Ghr</i>	P16882	Growth hormone receptor	INPSLGTSSSGKPR	S4(Hex1HexNAc1)	-5.00
<i>Gbbp11</i>	Q6NZP2	Vasculin-like protein 1	GHDGMSQRSGGSSTGNHR	T14(HexNAc)	-5.00
<i>Grn</i>	P28798	Granulins	CVSPTGHTLLK	T5(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1); T8(Hex1HexNAc1)	-5.00

<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(HexNAc); T8(HexNAc)	-5.00
<i>Heg1</i>	E9Q7X6	Protein HEG homolog 1	TSSDHTDHGYVPSTFTK	T6(Hex1HexNAc1)	-5.00
<i>Heg1</i>	E9Q7X6	Protein HEG homolog 1	ISDSPSQAQPK	S2(Hex1HexNAc1); S4(Hex1HexNAc1)	-5.00
<i>lgfbp5</i>	Q07079	Insulin-like growth factor-binding protein 5	FVGGAEHTAHR	T8(HexNAc)	-5.00
<i>Igh-3</i>	P01867	Ig gamma-2B chain C region	KLEPSGPISTINPCPPCK	T10(Hex1HexNAc1)	-5.00
<i>Itih1</i>	Q61702	Inter-alpha-trypsin inhibitor heavy chain H1	RTFVLSAIQPSPTAHPIDSK	T13(Hex1HexNAc1)	-5.00
<i>Itih1</i>	Q61702	Inter-alpha-trypsin inhibitor heavy chain H1	RTFVLSAIQPSPTAHPIDSK	S11(Hex1HexNAc1); S19(Hex1HexNAc1)	-5.00
<i>Itih1</i>	Q61702	Inter-alpha-trypsin inhibitor heavy chain H1	TFVLSAIQPSPTAHPIDSK	S10(Hex1HexNAc1)	-5.00
<i>Itih1</i>	Q61702	Inter-alpha-trypsin inhibitor heavy chain H1	TFVLSAIQPSPTAHPIDSK	S10(HexNAc); T12(Hex1HexNAc1)	-5.00
<i>Itih1</i>	Q61702	Inter-alpha-trypsin inhibitor heavy chain H1	RTFVLSAIQPSPTAHPIDSK	S11(Hex1HexNAc1)	-5.00
<i>Itih1</i>	Q61702	Inter-alpha-trypsin inhibitor heavy chain H1	RTFVLSAIQPSPTAHPIDSK	S11(Hex1HexNAc1); S19(Hex1HexNAc1)	-5.00
<i>Klhl1</i>	Q9J174	Kelch-like protein 1	TDTWTMVAPLSMPR	T5(HexNAc)	-5.00
<i>Kng1</i>	O08677	Kininogen-1	TESLASSSEYSTTSTQMQR	S11(Hex1HexNAc1); T12(HexNAc); T13(HexNAc); T15(HexNAc)	-5.00
<i>Kng1</i>	O08677	Kininogen-1	IAPSKAPILK	S4(Hex1HexNAc1)	-5.00
<i>Kng1</i>	O08677	Kininogen-1	TEGPTLTTPR	T5(HexNAc)	-5.00
<i>Kng1</i>	O08677	Kininogen-1	TEGPTLTTPR	T1(HexNAc); T5(Hex1HexNAc1)	-5.00
<i>Lipc</i>	P27656	Hepatic triacylglycerol lipase	SLGATEASKPLK	T5(HexNAc)	-5.00
<i>Lipc</i>	P27656	Hepatic triacylglycerol lipase	SLGATEASKPLK	T5(Hex1HexNAc1); S8(Hex1HexNAc1)	-5.00
<i>Madcam1</i>	Q61826	Mucosal addressin cell adhesion molecule 1	TLSEGGPCRPK	S4(Hex1HexNAc1)	-5.00
<i>Nid1</i>	P10493	Nidogen-1	GVPTERTR	T4(Hex1HexNAc1); T7(Hex1HexNAc1)	-5.00

<i>Pam</i>	P97467	Peptidyl-glycine alpha-amidating monooxygenase	LESTLRPAESR	T4(Hex1HexNAc1); S10(Hex1HexNAc1)	-5.00
<i>Pi16</i>	Q9ET66	Peptidase inhibitor 16	SSATSVSPK	S5(Hex1HexNAc1)	-5.00
<i>Pltp</i>	P55065	Phospholipid transfer protein	NRPADVAASHVPPSAAAA	S9(Hex1HexNAc1)	-5.00
<i>Plxnb1</i>	Q8CJH3	Plexin-B1	APSTASDVLPGAKPSR	T4(Hex1HexNAc1); S15(Hex1HexNAc1)	-5.00
<i>Prg4</i>	Q9JM99	Proteoglycan 4	TSDKDVEPTSTTPK	T9(Hex1HexNAc1); T11(Hex1HexNAc1)	-5.00
<i>Prg4</i>	Q9JM99	Proteoglycan 4	TSDKDVEPTSTTPK	T9(Hex1HexNAc1); T11(Hex1HexNAc1)	-5.00
<i>Ptprg</i>	Q05909	Receptor-type tyrosine-protein phosphatase gamma	TAAEGGHQTIPGR	T9(HexNAc)	-5.00
<i>Ptprg</i>	Q05909	Receptor-type tyrosine-protein phosphatase gamma	TAAEGGHQTIPGRR	T9(Hex1HexNAc1)	-5.00
<i>Ptprg</i>	Q05909	Receptor-type tyrosine-protein phosphatase gamma	TAAEGGHQTIPGRR	T1(Hex1HexNAc1); T9(Hex1HexNAc1)	-5.00
<i>Qsox1</i>	Q8BND5	Sulfhydryl oxidase 1	IAPTVMKFAFR	T4(Hex1HexNAc1)	-5.00
<i>Qsox1</i>	Q8BND5	Sulfhydryl oxidase 1	DMQENAPGQQHLSKR	S13(Hex1HexNAc1)	-5.00
<i>Rgma</i>	Q6PCX7	Repulsive guidance molecule A	VRTLPPAGDSQER	T3(Hex1HexNAc1)	-5.00
<i>Rgma</i>	Q6PCX7	Repulsive guidance molecule A	SDSPEICHYEK	S1(Hex1HexNAc1)	-5.00
<i>Sdc1</i>	P18828	Syndecan-1	QTPSTWK	T2(Hex1HexNAc1)	-5.00
<i>Serpinf2</i>	Q61247	Alpha-2-antiplasmin	ADFHGGKTFGPDLK	T8(Hex1HexNAc1)	-5.00
<i>Serpinf2</i>	Q61247	Alpha-2-antiplasmin	TFGPDLK	T1(Hex1HexNAc1)	-5.00
<i>Shc1</i>	P98083	SHC-transforming protein 1	LSGGGGRTR	T9(HexNAc)	-5.00
<i>Sparcl1</i>	P70663	SPARC-like protein 1	AVSTEPTDAAVVPR	T7(Hex1HexNAc1)	-5.00
<i>St5</i>	Q924W7	Suppression of tumorigenicity 5 protein	NSSITHGTGGTK	T5(HexNAc); T8(HexNAc)	-5.00
<i>Sybu</i>	Q8BHS8	Syntabulin	IGMVGEGSIQSAR	S8(HexNAc)	-5.00
<i>Tgoln2</i>	Q62314	Trans-Golgi network integral membrane protein 2	DQDKTTLAAVSSK	T5(Hex1HexNAc1); S12(Hex1HexNAc1)	-5.00
<i>Vtn</i>	P29788	Vitronectin	TDGTLKPTAFLDPEEQPSTPAPK	T8(Hex1HexNAc1); T19(HexNAc)	-5.00
<i>Vwce</i>	Q3U515	von Willebrand factor C and EGF domain-containing protein	QQSPHGELAK	S4(Hex1HexNAc1)	-5.00
Differential Liver Glycoproteomes					

Gene	Accessions	Protein Name (UniProt)	Sequence	Unambiguous Glycan Addition	Fold-difference KO vs. WT for Glycopeptide [log10 (fold difference)]
<i>Itih1</i>	Q61702	Inter-alpha-trypsin inhibitor heavy chain H1	TFVLSAIQPSPTAHPIDSK	S5(Hex1HexNAc1);S10(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(HexNAc);T8(HexNAc)	-5.00
<i>Plexnb1</i>	Q8CJH3	Plexin-B1	STASDVLPGAKPSR	T2(Hex1HexNAc1);S13(Hex1HexNAc1)	-5.00
<i>Ptprg</i>	Q05909	Receptor-type tyrosine-protein phosphatase gamma	EKSEATHTAESDRTPAPTSSPHR	T6(Hex1HexNAc1);S22(Hex1HexNAc1);S23(Hex1HexNAc1)	-5.00
<i>Itih1</i>	Q61702	Inter-alpha-trypsin inhibitor heavy chain H1	TFVLSAIQPSPTAHPIDSK	S10(Hex1HexNAc1)	-5.00
<i>Acaa2</i>	Q8BWT1	3-ketoacyl-CoA thiolase, mitochondrial	VSGGAIALGHPLGGSGSR	S2(HexNAc)	-5.00
<i>Gsn</i>	P13020	Gelsolin	ARPSTMVVEHPEFLK	S4(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	SEARPSTMVVEHPEFLK	S6(Hex1HexNAc1)	-5.00
<i>Slc38a10</i>	Q51012	Putative sodium-coupled neutral amino acid transporter 10	GRPAPSQDMGQHLPGEVK	S6(Hex1HexNAc1)	-5.00
<i>Col18a1</i>	P39061	Collagen alpha-1(XVIII) chain	AGQGGSPATPAVPIPLVAPAASPMK	S6(Hex1HexNAc1)	-5.00
<i>Slc38a10</i>	Q51012	Putative sodium-coupled neutral amino acid transporter 10	GRPAPSQDMGQHLPGEVK	S6(Hex1HexNAc1)	-5.00
<i>Col18a1</i>	P39061	Collagen alpha-1(XVIII) chain	AGQGGSPATPAVPIPL	S6(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1)	-5.00
<i>Gsn</i>	P13020	Gelsolin	VSEARPSTMVVEHPEFLK	S7(Hex1HexNAc1)	-5.00
<i>Traf3</i>	Q60803	TNF receptor-associated factor 3	VSLQNESVEKKNK	S8(Hex1HexNAc1)	-5.00
<i>Tgoln2</i>	Q62314	Trans-Golgi network integral membrane protein 2	TTLAAVSSK	T1(Hex1HexNAc1)	-5.00
<i>Ebp</i>	P70245	3-beta-hydroxysteroid-Delta(8),Delta(7)-isomerase	TTNTVPLHPYWPR	T1(Hex1HexNAc1)	-5.00
<i>Klhl1</i>	Q9J174	Kelch-like protein 1	TWTVLPPMSTHR	T10(HexNAc)	-5.00

<i>Ctsd</i>	P18242	Cathepsin D	TTEPVSELLK	T2(Hex1HexNAc1)	-5.00
<i>Tgoln2</i>	Q62314	Trans-Golgi network integral membrane protein 2	TTRPTDGQGQK	T2(Hex1HexNAc1)	-5.00
<i>Fetub</i>	Q9QXC1	Fetuin-B	NTAPTSSPSVTAPR	T5(Hex1HexNAc1)	-5.00
<i>Gpc4</i>	P51655	Glypican-4	EQRPTTAAGTSLDR	T5(Hex1HexNAc1)	-5.00
<i>Fetub</i>	Q9QXC1	Fetuin-B	NTAPTSSPSVTAPR	T5(Hex1HexNAc1)	-5.00
<i>Itih1</i>	Q61702	Inter-alpha-trypsin inhibitor heavy chain H1	IQPSPTAHPIDSK	T6(Hex1HexNAc1)	-5.00
<i>Itih1</i>	Q61702	Inter-alpha-trypsin inhibitor heavy chain H1	AIQPSPTAHPIDSK	T7(Hex1HexNAc1)	-5.00
<i>Gpc4</i>	P51655	Glypican-4	HPEQRPTTAAGTSLDR	T7(Hex1HexNAc1)	-5.00
<i>Kng1</i>	O08677	Kininogen-1	AETEQQPTHGHGWLHEK	T8(Hex1HexNAc1)	-5.00
<i>Chga</i>	P26339	Chromogranin-A	AGPEEVPTAASSSHFHAGYK	T8(Hex1HexNAc1)	-5.00
<i>Slc38a10</i>	Q51012	Putative sodium-coupled neutral amino acid transporter 10	AAPPEVPKSPEK	S9(Hex1HexNAc1)	-5.00
<i>Slco1a4</i>	Q9EP96	Solute carrier organic anion transporter family member 1A4	TQTLKPTQDPTECVK	T7(Hex1HexNAc1)	-5.00
<i>Kng1</i>	O08677	Kininogen-1	IGCVHAISTDSDLEPVLK	T9(Hex1HexNAc1)	-5.00
<i>Lrp1</i>	Q91ZX7	Prolow-density lipoprotein receptor-related protein 1	MDAPKTCSPK	T6(Hex1HexNAc1)	-5.00

Table 3.3: Differential Glycoproteomes of *Galnt2* KO vs. WT Mice. Quantitative differential lectin-affinity glycoproteomics was used to identify glycopeptides differentially observed in plasma (top) and livers (bottom) from *Galnt2* KO vs. WT mice. Fold-differences in glycopeptide levels in WT vs. KO models were determined by relative quantification of light vs. medium chain dimethyl isotopomeric labels in the WT vs. KO samples, respectively. *Contributions: All data in this table generated by Katrine Schjoldager with samples processed by SAK.*

CHAPTER 4 – *APOC3* A43T VARIANT CONFERS VARIANT CONFERS LOSS-OF-FUNCTION THROUGH DISRUPTING APOC-III BINDING AND ACCELERATING TRIGLYCERIDE-RICH LIPOPROTEIN CLEARANCE

Introduction

ApoC-III is a small apolipoprotein (~8.8 kDa) secreted from the liver and small intestine that circulates on both TRLs such as VLDL and chylomicrons in addition to high-density lipoproteins (HDLs). In biochemical studies and animal models, ApoC-III has been shown to increase plasma TGs by both directly inhibiting the activity of lipoprotein lipase (LPL) on TRLs and by inhibiting the clearance of TRLs by the liver, possibly through competitive interactions with ApoE, a critical ligand on TRLs for receptor-mediated clearance [98-100,316,317]. In a murine model of delayed ApoB clearance through LDL receptor deficiency, human ApoC-III overexpression increased circulating TRLs and exacerbated the development of atherosclerosis [318], offering early support for the notion that ApoC-III may be positive mediator of CHD risk.

Genetic discoveries of the last decade have highlighted the potential importance of ApoC-III to CHD risk in humans. Common noncoding polymorphisms in *APOC3* are associated with plasma TGs, HDL and non-HDL cholesterol and CHD risk through genome-wide association studies (GWAS) [117]. In 2008, a GWAS in the Lancaster Amish identified a rare truncation variant in *APOC3*, Arg19Ter (R19X) that was associated with reduced levels of serum ApoC-III, improved fasting and postprandial TGs, and reduced coronary artery calcification, a surrogate measure of atherosclerosis [105]. More recently, two large sequencing efforts identified multiple rare coding variants in *APOC3* associated in aggregate with ~40% reduced TGs and an equal reduction in risk of CHD [91,106]. One of these efforts, utilizing whole-exome sequencing for

discovery followed by exome-wide genotyping in >110,000 individuals [106], identified four *APOC3* nonsynonymous variants that constituted the protective association, R19X, two splice-site variants IVS2+1 G>A and IVS3+1 G>T, and a missense variant A43T (A23T of the mature secreted protein lacking the 20 residue signal peptide) and also showed that these variants were associated with lower ApoC-III levels in carriers vs. controls, thus positing them as LoF variants. The other study performed targeted *APOC3* exon sequencing and genotyping and confirmed three of these four variants as protective from TG levels and risk of ischemic heart and vascular diseases [91]. Importantly, in both studies, *APOC3* variants were not associated with LDL-C concentrations, suggesting a protective impact of the mutations is independent of LDL-C lowering.

These genetic data and other subsequent studies posit that reducing circulating ApoC-III concentrations may be a novel therapeutic approach to reduce CHD risk potentially synergistic with and orthogonal to existing LDL-C focused treatments. Indeed, at least one ApoC-III-targeting therapy, an antisense oligonucleotide to *APOC3*, has already passed preclinical screening [103] and is currently under clinical development and has shown efficacy in TG lowering in patients with extreme hypertriglyceridemia [102,104]. We sought to identify mechanisms of lowering ApoC-III that would be efficacious and safe, and thus hypothesized that studying the mechanism of protection by the identified *APOC3* LoF variants may offer insights for therapeutic discovery.

Materials and Methods

Ethics Statement

All human participants for this study and all analyses performed were completed following the Declaration of Helsinki and were approved by the Institutional Review Board of the Perelman School of Medicine at the University of Pennsylvania and all participants provided informed consent.

Human Subject Identification and Ascertainment

Human participants were recruited from two independent cohorts for ascertainment of *APOC3* loss-of-function variants and subsequent phenotyping. The first cohort, the Penn High HDL Study (HHDL), is a cross-sectional study of heritable factors contributing to elevated levels of HDL-C. Participants were recruited with HDL-C levels greater than the 95th percentile for age and gender by physician referrals or through the Hospital of the University of Pennsylvania clinical lipid laboratory. Relatives of high HDL-C probands were also invited to participate in the study. Participants were consented for additional plasma lipid analysis and research genetic studies including targeted sequencing of candidate genes, exome sequencing, and SNP array genotyping under a protocol approved by the Institutional Review Board (IRB) of the University of Pennsylvania. Over 2,500 participants have been recruited for the HHDL study, with recruitment ongoing. The second cohort, the Complex Genomics Initiative and BioBank cohort (CGI BioBank), is a collection of banked plasma samples from participants who were treated at the University of Pennsylvania Health System and were consented for genetic analysis and biochemical studies of collected plasma for the study of the heritable risk factors and novel biomarkers related to CHD risk. Approximately 10,000 participants have been recruited so far, with recruitment of additional participants ongoing through IRB approved protocols of the University of Pennsylvania.

For both studies, genomic DNA was isolated from whole blood from participants and prepared for genotyping using the Exome Chip SNP genotyping array (Human Exome BeadChip v1.0, Illumina, Inc., San Diego, CA, USA). The Exome Chip contains more than 240,000 SNP markers mostly corresponding to coding regions across the genome that are derived from variants found two or more times across more than one data set among 23 unique data sets of >12,000 exome or whole genome sequences. This included the four *APOC3* coding variants studied here, R19X (rs76353203), IVS2+1 G>A (rs138326449), IVS3+1 G>T (rs140621530) and A43T (rs147210663). After removing subjects from analysis for which genotyping quality control measures failed or demographic covariates were not available, a total of 2,736 participants from the HHDL study and 5,244 subjects from the CGI BioBank cohort were genotyped. *APOC3* variant genotypes for participants with identified variants from Exome Chip genotyping were confirmed by Sanger sequencing. A 674 bp fragment of the *APOC3* gene sequence was amplified by PCR from 200 ng of genomic DNA from identified variant carriers using the following oligonucleotide primers: F – CTCCTTCTGGCAGACCCAGCTAAGG, R – CCTAGGACTGCTCCGGGGAGAAAG. PCR products were purified using a QIAquick PCR purification Kit (Qiagen, Germantown, MD, USA) and Sanger sequenced by Genewiz (South Plainfield, NJ, USA). Sequence chromatograms were aligned to the *APOC3* reference from NCBI using Sequencher (Gene Codes, Ann Arbor, MI, USA).

Lipid Phenotyping of Human Participants

Carriers of any of the four *APOC3* coding variants along with noncarriers from the same identification cohorts matched for age, gender, and ethnicity were selected for further lipid analysis from collected plasma samples. Plasma total cholesterol (TC), HDL-C, TG,

ApoC-III, ApoA-I, ApoB, and ApoE were measured by an Axcel clinical autoanalyzer (Alfa Wassermann Diagnostic Technologies, West Caldwell, NJ, USA). Non-HDL cholesterol (NonHDL-C) was calculated by the difference of the TC and HDL-C measurements. All measurements were made from overnight fasted plasma samples.

Selective Reaction Monitoring of ApoC-III from Human Plasma

Fasting plasma from A43T carriers (N=17) and noncarriers (N=24) was processed for measurement of the relative amounts of WT vs. mutant protein using multiple reaction monitoring assays. AspN protease was used to digest 0.15 μ l of plasma for each sample. A total of 0.3 μ l was ultimately used for each reaction. Synthetic deuterium-labeled peptides corresponding to 20-mers harboring the Ala or Thr43 residues (residues 25-44 of the full-length ApoC-III protein) were added to each plasma sample for selective monitoring. The following mass-to-charge (m/z) ratios were used for monitoring protease-digested ApoC-III peptides and transitions on a TSQ Quantiva triple-stage quadrupole mass spectrometer (Thermo Fisher Scientific) with UltiMate 3000 Nano Liquid Chromatography (LC) Systems: WT light (unlabeled) – 557.8 \rightarrow 547.9, 577.0, 614.7; WT heavy (deuterium-labeled) – 559.5 \rightarrow 547.9, 577.0, 617.0; mutant light (unlabeled) – 565.3 \rightarrow 557.9, 587.0, 625.7; mutant heavy (deuterium-labeled) – 567.0 \rightarrow 557.9, 587.0, 627.0. Ratios of light to heavy peptides for WT and mutant 20-mer peptides were measured for each sample and compared to that measured for a standard curve of combinations of WT and mutant full-length proteins at known molar ratios and subject to the same AspN digest conditions as the plasma samples. From this, the mutant:WT ratio for each sample was calculated.

In silico Analysis of Splicing Effects of APOC3 Variants

Impact of the four *APOC3* coding variants on *APOC3* exon splicing was assessed using the Automated Splice Site and Exon Definition Analyses (ASSEDA) server (MutationForecaster, <https://www.mutationforecaster.com/learn.php#asseda> [319]).

Cloning of human APOC3 cDNA and Adeno-associated Virus Generation

The cDNA of the human *APOC3* gene (NM_000040) was obtained from OriGene (Rockville, MD, USA) and cloned in the pcDNA3.1/V5-His TOPO expression plasmid (Thermo Fisher Scientific, Waltham, MA, USA). The A43T variant was introduced into the pcDNA3.1/V5-His plasmid containing the WT *APOC3* cDNA using the Quickchange II site-directed mutagenesis kit (Agilent technologies, Wilmington, DE, USA) using the following primers: F – GCACGCCACCAAGACCACCAAGGATGCACTGAGCAG, R – CTGCTCAGTGCATCCTTGGTGGTCTTGGTGGCGTGC, where the bolded and underlined nucleotide indicates the position of the variant in the primer sequences. The *APOC3* WT and A43T variant cDNA sequences were further amplified by PCR to introduce KpnI and Sall restriction sites using the following primers: F – ACGCGTGGTACCATGCAGCCCCGGGTACTCCTTG, R – CCGCCCCGGGTCGACTCAGGCAGCCACGGCTGAAG. Amplified PCR products corresponding to the WT and A43T variant cDNA sequences of *APOC3* were digested with KpnI and Sall and ligated with an adeno-associated virus (AAV) serotype 8 vector plasmid containing the liver-specific thyroxine binding globulin (TBG) promoter provided by the Vector Core of the University of Pennsylvania [232-234,320]. The ligation of the AAV plasmids and integrity of sequences were confirmed by Sanger sequencing and restriction digest and agarose gel electrophoresis analysis. AAV production, purification,

and titering from the cloned *APOC3* WT and A43T vector plasmids was performed by the Vector Core.

¹²⁵I Radiolabeling of ApoC-III

Mature forms of WT and A43T variant ApoC-III (79 residues for each and lacking the signal peptide) were produced by solid-state peptide synthesis by Pierce Custom Peptides (Thermo Fisher Scientific). Purified peptides were confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry to be of >95% purity. Proteins were solubilized in 10 mM ammonium bicarbonate buffer, pH 7.4 to a concentration of approximately 0.2-0.4 mg/ml.

Solubilized proteins were iodinated with ¹²⁵I directly using the iodine monochloride (ICI) method [321] or through incorporation with ¹²⁵I tyramine cellobiose (TCB) [322]. For direct ¹²⁵I labeling of ApoC-III by the ICI method, approximately 0.5 ml of ApoC-III in 10 mM ammonium bicarbonate buffer (0.2-0.4 mg/ml) was iodinated with 1 mCi of ¹²⁵I, 300 μ l of 1 M glycine, and 150 μ l of 1.84 M NaCl / 2.84 μ M ICI solution, vortexed and applied to a PG-10 desalting column (Amersham Biosciences) that was pre-equilibrated with 0.15 M NaCl / 1 mM EDTA solution. Iodinated proteins were eluted in a final volume of 2.5 ml in NaCl/EDTA solution and dialyzed against PBS before measurement of protein concentration by BCA assay and ¹²⁵I activity by gamma counting.

For iodination of TCB, 180 μ l of 0.4 M sodium phosphate, Pierce iodination beads (Thermo Fisher Scientific), and 1 mCi of ¹²⁵I were combined and then added with 20 μ l of 0.2 μ M TCB (a kind gift from Dr. David Usher and Dr. William Cain, University of Delaware, Newark, DE, USA). This mixture was incubated at room temperature for 30

minutes with shaking every five minutes, and then 10 μ l of 0.1 M NaI and 20 μ l of 0.1 M NaHSO₃ were added to stop the reaction. Next, 40 μ l of 0.2 μ M cyanuric chloride was added to activate the ¹²⁵I-TCB. The activated ¹²⁵I-TCB was then added to the ApoC-III protein samples (solubilized in ammonium bicarbonate buffer at a concentration of 0.2-0.4 mg/ml) and incubated at room temperature for three hours. Finally, the reaction was fractionated over a PD-10 desalting column to remove unbound iodine, and fractions containing iodinated protein were pooled together and dialyzed against PBS for further use.

Animal Models

C57BL/6 wild-type (WT) mice were purchased from The Jackson Laboratory (Cat# 000664). Additionally, *Apoc3* KO mice on a C57BL/6 background (Cat# 002057) were recovered from cryo-preservation from The Jackson Laboratory and bred at the University of Pennsylvania. *Apoc3* KO mice were bred with *Apobec1*-deficient, human *APOB* transgenic mice that either were wild-type (LahB WT), heterozygous (LahB Het), or deficient (LahB KO) for the murine *Ldlr* gene. Mice were maintained in monitored small animal facility at the University of Pennsylvania under IACUC-approved protocols, were fed ad libitum with a standard chow diet or Western diet containing 0.21% cholesterol from OpenSource Diets (Cat# D12079B, Research Diets, New Brunswick, NJ, USA) for the indicated periods of time. All mice were provided access to water ad libitum and were maintained with a 12 hour on/off light cycle with lights off from 7:00pm to 7:00am daily. All blood samples from mice were collected by retro-orbital bleeding from mice anesthetized with isoflurane using EDTA-coated glass tubes under approved protocols.

For *APOC3* AAV expression experiments in mouse models, mice were initially fasted for four hours, bled, and administered a control AAV serotype 8 vector lacking a transgene (Null), *APOC3* WT AAV (WT), or *APOC3* A43T (A43T), all at a dose of 3×10^{11} genome copies of virus per mouse as determined by digital PCR based titering by the Vector Core. AAVs were diluted in sterile phosphate-buffered saline (PBS) in a sterile chemical hood and drawn in individual insulin syringes for administration by intraperitoneal injection. For some experiments, CETP AAV, which was generated as described previously, was co-administered with Null or *APOC3* AAVs at a dose of 3×10^{10} genome copies per mouse; for these experiments, appropriate amounts of CETP AAV vector was mixed with Null or *APOC3* AAV vectors and PBS for administration by intraperitoneal injection. Subsequent blood collections were performed after fasting the mice for four hours and under isoflurane anesthesia at the indicated timepoints. Blood was separated to obtain plasma by centrifugation in a microcentrifuge at 4 °C at 10,000 RPM for seven minutes. All mouse experiments were performed with six mice per group unless otherwise noted.

Lipid and human apolipoprotein measurements were performed on mouse plasma using an Axcel autoanalyzer. In addition, for some experiments, plasma was pooled by experimental group freshly after collection and 150 μ l of plasma was separated by fast-protein liquid chromatography (FPLC) on a Superose 6 gel-filtration column (GE Healthcare Life Sciences, Pittsburgh, PA, USA) into fractions of 0.5 ml volume each. Total cholesterol and TG were measured from FPLC-separated fractions using Infinity Liquid Stable cholesterol and triglyceride reagents (Thermo Scientific, Waltham, MA, USA) in 96-well microplates using a Synergy Multi-Mode Microplate Reader (BioTek, Winooski, VT, USA).

Oral Fat Tolerance Tests

Oral fat tolerance tests (OFTTs) were performed in *APOC3* WT, A43T, and Null AAV treated mice four weeks following AAV administration. For these experiments, mice were fasted for 12 hours and weighed. Olive oil (Sigma, St. Louis, MO, USA) was prepared in 1.0 ml syringes at a volume of 10 μ l x fasting body weight (in grams) for each mouse. Mice were bled prior to olive oil administration to measure pre-gavage plasma TGs after fasting and then were gavaged orally with olive oil and bled subsequently at 1, 3, 5, and 7 hours after gavage. Plasma TGs at each timepoint were measured by colorimetric assays in 96-well microplates as described above.

TRL Clearance Studies

For ^3H -TRL clearance studies in mouse, human TRLs were obtained from pooled human nonfasted plasma from the Hospital of the University of Pennsylvania by density-gradient ultracentrifugation in a Beckman XL-90 Ultracentrifuge (Beckman Coulter, Brea, CA, USA) using a 70.1 Ti rotor (Beckman Coulter) at a speed of 40,000 RPM for 18 hours. Ultracentrifugation of the plasma enabled a white, cream-like TRL layer ($d < 1.006$) to float, and this was collected and subjected to another ultracentrifugation at a density of 1.006. The top cream-like layer was collected and measured for total protein concentration by bicinchoninic assay (Thermo Fisher Scientific, Waltham, MA, USA). Approximately 3 mg of TRL protein was used for labeling with 0.5 mCi of ^3H -Triolein (Perkin Elmer, Waltham, MA, USA). 0.5 mCi of ^3H -Triolein in toluene was dried under a nitrogen gas evaporator and resolubilized in 150-200 μ l ethanol and then added dropwise to a mixture of TRLs (3 mg protein) and lipoprotein-deficient human serum ($d > 1.220$, 100 mg protein per 3 mg TRL protein) in a glass tube. The mixture was then

incubated at 37 °C with gentle shaking for 12-14 hours. Then the mixture was subjected to ultracentrifugation again by addition of d=1.006 KBr solution and the cream-like top layer (approx. 2.5-3.0 ml) was collected, dialyzed against PBS, and fractionated by FPLC to measure TG and ³H activity in the TRL fractions (5-12) relative to unincorporated ³H activity. Dialyzed ³H-TRLs were then administered by intravenous tail-vein injection into mice expressing Null, WT, or A43T AAVs. Mice were bled at 1, 8, 15, 30, 60, and 120 minutes after ³H-TRL administration. ³H activity in 10 µl of plasma from each timepoint was measured by scintillation counting and relative ³H activity remaining in circulation was calculated by normalizing the activity from each timepoint by that of the 1 minute timepoint for each mouse.

VLDL-TG, ApoC-III, and ApoB Secretion Studies

Hepatic VLDL secretion was measured in mice expressing Null or APOC3 AAVs using ³⁵S metabolic labeling as previously described [323]. Briefly, mice expressing Null or APOC3 AAVs at four weeks after AAV administration were fasted for four hours and then administered the surfactant Poloxamer P407 (25 mg in 0.4 ml PBS) by intraperitoneal injection to inhibit peripheral lipolysis. Fifteen minutes after P407 administration, each mouse was administered ³⁵S-labeled Cysteine / Methionine (0.5 mCi in 0.2 ml PBS, Perkin Elmer, Waltham, MA, USA) by intravenous tail-vein injection. Mice were bled at 0, 30, 60, 90, and 120 minutes after radioisotope administration for plasma collection. Plasma TGs were measured at each timepoint with the Infinity Triglycerides reagent (Thermo Scientific) in 96-well microplates. 60 and 120 minute plasma samples were also used for measuring the secretion rate of newly synthesized, ³⁵S-labeled ApoB and ApoC-III. For ApoB measurements, 2 µl of each plasma sample

from these timepoints was subjected to SDS-PAGE under reducing conditions using NuPAGE 3-8% Tris-Acetate gels (Thermo Fisher Scientific). Gels were then fixed in 25% methanol / 20% acetic acid solution for 30 minutes and then with AutoFluor intensifier solution (National Diagnostics, Atlanta, GA, USA) for 30 minutes. Gels were then dried using a vacuum pump for four hours at 65 °C and then exposed to autoradiography film for one week at -80 °C and films were then developed. Visible bands on the film corresponding to ApoB-100 and ApoB-48 were marked and positions correlated on the dried gels, and then those bands were excised from the gels, solubilized with Solvable reagent (0.5 ml, Perkin Elmer) at 55 °C for three hours and then ³⁵S activity measured by liquid scintillation counting. Activity of each band was normalized to the activity measured after trichloroacetic acid precipitation of proteins from 5 µl of plasma from the same plasma sample used for the ApoB SDS-PAGE. Normalized ³⁵S activity from the 60 and 120 minute timepoints was used to calculate the relative ApoB-100 and ApoB-48 secretion rates. For ApoC-III secretion measurements, the procedure followed was the same as written above for measuring ApoB secretion, except NuPAGE 4-12% Bis-Tris gels were used for SDS-PAGE instead of 3-8% Tris-Acetate gels.

ApoC-III Kinetic Studies

For ¹²⁵I ApoC-III clearance studies, iodinated ApoC-III [specific activity of approx. 300,000-500,000 counts per minute (CPM) / µg protein] was mixed with plasma from recipient mice *ex vivo* at a ratio of 1 µg ¹²⁵I-labeled protein / 200 µl plasma. Iodinated ApoC-III and mouse plasma mixtures were incubated at 37 °C for one hour were then administered by intravenous tail-vein injection in mice expressing either WT or A43T

variant *APOC3* by AAV. Mice expressing WT *APOC3* received ^{125}I -WT ApoC-III protein and those expressing A43T *APOC3* received ^{125}I -A43T ApoC-III. Mice were bled 1 minute, 15 minutes, 30 minutes, 1 hour, 3 hours, 5 hours, 9 hours, and 24 hours after radioisotope administration and sacrificed at the 24 hour timepoint. ^{125}I activity was measured from 10 μl of each plasma sample by counting on a Packard Cobra II Auto-Gamma counter. Relative activity at each timepoint was measured as the fraction of activity from the one minute timepoint for each mouse respectively. For experiments with ^{125}I -TCB labeled ApoC-III, the above procedure was followed, and in addition liver and kidneys were collected from mice at the 24 hour timepoint after perfusion with 10 ml of cold PBS. 30 mg of each tissue was homogenized in PBS and measured for ^{125}I activity. Tissue ^{125}I activity was normalized to the one minute plasma ^{125}I activity for each mouse respectively. For experiments measuring urinary clearance of ^{125}I -ApoC-III, mice were administered ^{125}I -labeled proteins in the manner above, but mice were bled at 1 minute, 3 hours, 6 hours, and 24 hours after administration only. For these experiments, mice were housed singly in rodent metabolic cages (Tecniplast, West Chester, PA, USA) and urine was collected from 0-6 hours, 6-12 hours, and 12-24 hours during the course of the study. Plasma and urine samples (20 μl of each sample) were counted for total ^{125}I activity by gamma counting. In addition, proteins were separated by TCA precipitation from 20 μl of each sample and the TCA soluble (degraded) and TCA pellet (intact protein) fractions were counted for ^{125}I activity.

Liver Gene Expression

For gene expression measurements, approximately 20 ng of liver tissues from each mouse was snap-frozen at $-80\text{ }^{\circ}\text{C}$. Total RNA was extracted from tissues using TRIzol

reagent (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. Synthesis of cDNA was completed using 1 µg of RNA following extraction using the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). Quantitative real-time reverse transcriptase PCR (RT-PCR) was performed on the synthesized cDNA using a QuantStudio 7 Real-Time PCR System using the following Taqman primer sets (Thermo Fisher Scientific): human APOC3 - Hs00163644_m1, mouse Apoc3 - Mm00445670_m1, mouse Actb (actin) - Mm00607939_s1. Relative expression of human APOC3 and mouse Apoc3 in the liver was analyzed for each mouse by normalizing to actin expression from the same sample.

Hepatic AAV vector DNA content in each mouse was also assessed by quantitative real-time RT-PCR. Total hepatic DNA was isolated from 25 mg of liver for each mouse using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). Isolated liver DNA was used for quantitative RT-PCR with Taqman primers targeting the rabbit beta-globin poly-A sequence of the AAV vector backbone. Cycle numbers from this assay were compared to those from a standard curve of serial dilutions of a linearized AAV plasmid of known quantity.

Immunoblotting

For immunoblotting of ApoC-III from mouse liver samples, 30 µg of liver was separated on 10% Bis-Tris NuPAGE gels using MES buffer by one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Separated proteins were transferred to nitrocellulose membranes and membranes were then blocked with 5% fat-free milk in PBS (0.05% tween 20) for three hours. Membranes were then incubated for two hours at room temperature with a rabbit polyclonal antibody against human ApoC-III

(33A-R1a, Academy Biomedical, Houston, TX, USA) at a dilution of 1:2000, followed by 3 x 15 minute washes with PBS / Tween solution and incubation with goat anti-rabbit IgG HRP conjugate (sc-2030, Santa Cruz) for 30 min. at room temperature. Membranes were again washed for 3 x 15 minutes and proteins visualized by Luminata Crescendo chemiluminescent reagent (Millipore). Mouse β -actin was used as a loading control and was detected from the same membranes of separated hepatic proteins after incubation with mouse anti-actin primary antibody (sc-81178, Santa Cruz Biotechnology, Santa Cruz, CA, USA) at a dilution of 1:200 for two hours at room temperature, washed, and then incubated with goat anti-mouse IgG HRP (sc-2302, Santa Cruz Biotechnology) at a dilution of 1:5000 for 30 minutes at room temperature.

ApoC-III-Lipoprotein Binding Measurements

Human TRLs and HDL₃ were isolated from pooled plasma from the Hospital of the University of the Pennsylvania by density-gradient ultracentrifugation. Isolated lipoproteins were dialyzed against PBS and protein concentration measured by BCA assay. Radio-iodinated WT or A43T ApoC-III, generated as described above, was also solubilized in 10 mM ammonium bicarbonate buffer and dialyzed against PBS. For binding studies of ApoC-III with plasma lipoproteins, 1 μ g of ¹²⁵I-ApoC-III was combined with 200 μ l of pooled donor plasma and incubated at 37 °C for one, five, or 24 hours with gentle shaking. Immediately after incubation, mixtures were fractionated on a Superose 6 gel-filtration column. FPLC fractions were measured for TG and cholesterol in 0.1 ml of each fraction by colorimetric assays in a 96-well microplate and for ¹²⁵I activity by gamma counting. The percentage of total activity administered through the FPLC column in each FPLC fraction was calculated, and the sum of total area for the lipoprotein

fractions vs. free protein fractions was calculated and compared across replicate experiments. For binding of ApoC-III to TRLs or HDL₃, the procedure above was followed, with the exception that 200 µg of TRL protein or 100 µg of HDL₃ protein was used for each experiment in combination with 1 µg of ¹²⁵I-ApoC-III.

Binding of ¹²⁵I-ApoC-III to apolipoprotein-free lipid emulsions was also measured. Large lipid emulsions were prepared as described previously[324]. Briefly, a mixture of 300 mg triolein (Sigma Aldrich) and 18 mg egg-phosphatidylcholine (Sigma Aldrich) were dried under a nitrogen gas evaporator and then sonicated after addition of 5 ml glycerol (Thermo Fisher Scientific) using a Branson 450 microtip sonicator for five minutes. Concentrated emulsions were allowed to clear overnight. 15 µl of this concentrated emulsion was combined with 90 µl distilled water, 15 µl of 1.0 M Tris pH 8.0, 15 µl of 15% BSA solution, and 15 µl of 3.0 M NaCl to give a final volume of 150 µl of working emulsions. 1 µg of ¹²⁵I-ApoC-III was combined with 150 µl of working emulsions and incubated at 37 °C for one hour and then mixtures were fractionated by FPLC and analyzed for ¹²⁵I-activity in the emulsion-containing fractions vs. unbound protein fractions as detailed above. Preliminary FPLC fractionation experiments with emulsions alone showed that they eluted in fractions 5-15 of the 79 fractions obtained from gel-filtration analysis, which mirrored the fractionation pattern of human TRLs on the same Superose 6 column under the same conditions.

LPL Activity Assays

In vitro LPL activity in the presence of WT vs. A43T variant ApoC-III was measured by two independent methods. First, activity was measured against a large lipid emulsion made as described above but with the addition of ³H-Triolein (Perkin Elmer; 99 µg per

300 mg nonradioactive triolein) prior to glycerol addition and sonication as described previously [324]. Working emulsions were made by the method above, except that 75 μ l of water was used and in addition 15 μ l of heat-inactivated human serum was added to provide a source of ApoC-II. Working emulsion substrate (150 μ l) was combined with conditioned media (100 μ l) from COS7 cells expressing human LPL through adenoviral transduction, as described previously [324]. 50 μ l of WT vs. A43T ApoC-III (final reaction concentration of 20 μ M for each) or culture medium without ApoC-III was added to give a final reaction volume of 300 μ l. Reactions were incubated at 37 °C for 30 minutes with gentle shaking prior to addition of 3.25 ml of methanol:chloroform:heptane solvent (1.41:1.25:1.00) to stop the reactions and subsequent addition of 1.05 ml of pH 10.0 Buffer (Thermo Fisher Scientific). Stopped reactions were then spun at 2,000 RPM for 20 minutes to complete the extraction, and the upper phase containing liberated fatty acids was used for scintillation counting (0.5 ml per sample). The relative amount of hydrolysis of 3 H-Triolein to 3 H-oleic acid was calculated for each sample and compared across groups. A second LPL activity assay was measured using an Intralipid TG substrate in the presence or absence of WT vs. A43T ApoC-III, as described before [113]. Briefly, 10% Intralipid substrate (Santa Cruz; 1.2 nM per reaction) and WT or A43T ApoC-III (range of 0-58 μ M per reaction) were combined in 96-well microplates to a volume of 40 μ l and incubated at 30 °C for 15 minutes, followed by the addition of purified LPL (Sigma Aldrich; 40 nM per reaction) to give a final reaction volume of 60 μ l. Reactions were incubated at 30 C for 30 minutes and then 100 μ l of Wako NEFA-HR(2) R1 reagent (Wako Chemicals) was added along with Orlistat to inhibit further lipolysis (Santa Cruz; 20 μ M final reaction concentration), and the reactions were incubated for another 15 minutes at 30 °C. Then 50 μ l of Wako NEFA-HR(2) R2 reagent was added

and the reactions were incubated for another five minutes at 30 °C and absorbance was measured at 550 nm wavelength from each sample. All LPL assays were performed with three independent replicate experiments.

Statistical Analysis

All discrete data shown are represented as mean standard deviation, with error bars showing standard deviation. Statistical comparisons between two groups were performed using a two-tailed Student's t-test or one-way ANOVA as appropriate. Comparisons between three experimental groups were performed using two-way ANOVA. For measurement of mutant:WT ApoC-III ratios in A43T heterozygous carriers, the mean of the ratios among carriers was compared to an expected theoretical mean of 1.0 using a one-sample t-test. Statistical significance was defined as $P < 0.05$ for all analyses.

Results

Human APOC3 A43T heterozygotes have reduced TRLs and reduced ApoC-III

We thus searched for carriers of any of the four LoF coding variants by genotyping for the variants in two human cohorts using the Exome Chip. Given that prior studies have reported both an inverse relationship of ApoC-III levels to HDL-C levels and that *APOC3* LoF carriers were observed to have higher levels of HDL-C than matched noncarriers, we questioned whether we could enrich for the discovery of *APOC3* LoF variants in participants with high HDL-C levels. We genotyped the variants in a cohort of 2,736 individuals with HDL-C >95th percentile for age and gender (HHDL Cohort; **Table 4.1**). In this cohort, we identified a total of 36 carriers (all heterozygous) for any of the

four LoF variants; this constitutes a carrier frequency of 1/76 for any LoF variant. We also genotyped *APOC3* variants in an additional cohort of 5,744 participants from the Complex Genomics Initiative BioBank of the Hospital of the University of Pennsylvania (CGI BioBank), a cohort not enriched for high HDL-C levels. In this cohort, we identified an additional 28 *APOC3* LoF variant carriers, with a cohort-wide carrier frequency of ~1/205 (**Table 4.1**). We compared the carrier frequency in the HHDL cohort to the CGI BioBank cohort and the previously reported carrier frequencies from the NHLBI Exome Sequencing Project [106], all of which were not enriched for high HDL-C participants. Comparing the carrier frequencies in our HHDL Cohort with those from these normal HDL-C cohorts, we found a significant enrichment of *APOC3* LoF variants with high HDL-C relative to a normally distributed population ($P < 0.0001$ vs. CGI BioBank, $P = 0.0028$ vs. NHLBI ESP Cohort, $P < 0.0001$ vs. NHLBI ESP CHD Case-Control Cohort, chi-square test; **Table 4.1**).

We measured fasting plasma lipids and apolipoproteins in *APOC3* LoF variant carriers and age- and gender-matched noncarriers from the HHDL and CGI BioBank cohorts. Consistent with previous reports, LoF variant carriers exhibited marked reductions in TG levels ($P < 0.0001$ for both cohorts; **Figure 4.1 A**), and increased HDL-C and ApoA-I ($P < 0.05$ in the CGI BioBank Cohort, $P < 0.0001$ in the HHDL Cohort; **Figure 4.1 B-C**). We also found modest reductions in nonHDL-C and ApoB in the LoF variant carriers in CGI BioBank Cohort but not in the HHDL Cohort ($P < 0.05$, **Figure 4.1 D-E**).

In aggregate, carriers of *APOC3* LoF variants displayed reduced ApoC-III levels in plasma relative to noncarriers ($P < 0.0001$ for both cohorts; **Figure 4.2 A**). While previous studies grouped these four *APOC3* LoF variants together, we hypothesized that they may differ in their mechanisms conferring a reduction in ApoC-III levels. Since three

of the four variants (R19X, IVS2+1 G>A, IVS3+1 G>T) occur at intron-exon junctions of the *APOC3* gene, we assessed their predicted impact on RNA processing and intron splicing of *APOC3* using information-theory based predictions about the impact of each variant on the thermodynamic requirements for RNA processing [319,325] (**Table 4.2**). We found that the R19X variant is predicted to increase binding of heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1), a ubiquitously expressed regulator of nascent mRNA processing (**Table 4.2**). Additionally, both splice site variants IVS2+1 G>A and IVS3+1 G>T are predicted to abolish splice-donor sites (**Table 4.2**), causing retention of the adjacent introns. Since these three variants are thus predicted to disrupt *APOC3* gene expression, we assessed their combined impact on ApoC-III levels and found that heterozygotes for these variants had markedly reduced ApoC-III levels relative to noncarriers (**Figure 4.2 B**). Notably, their ApoC-III levels were less than half the levels of noncarriers ($P < 0.01$ for CGI BioBank Cohort & $P < 0.05$ for HHDL Cohort), suggesting that partial *APOC3* loss-of-function may further promote reduction in circulating ApoC-III.

The A43T missense variant is not predicted to impact binding of RNA processing proteins hnRNP H and hnRNP A1 and is not predicted to alter any splice-donor or splice-acceptor sites in the nascent *APOC3* transcript (**Table 4.2**). Because the A43T variant differs from the other three variants in that it is not predicted to disrupt expression of the protein, we studied the mechanism by which it confers LoF to *APOC3* further as it could uncover a different facet of regulation of ApoC-III's metabolism and functions. We measured plasma lipids and ApoC-III levels in carriers of A43T vs. matched controls from each of our identification cohorts and found that, like the other LoF variant carriers, A43T heterozygotes displayed decreased fasting TGs (**Figure 4.3 A**), increased HDL-C

(**Figure 4.3 B**), and no differences in nonHDL-C or ApoB levels (**Figure 4.4**). In addition, A43T carriers displayed approx. 50% reduced levels of plasma ApoC-III (**Figure 4.3 C**).

Since A43T is not predicted to impact splicing or expression of APOC3 mRNA, we questioned the cause of the lower circulating ApoC-III in the heterozygous carriers of A43T. We reasoned that the A43T allele might result in a mutant form of ApoC-III that is expressed and secreted into the circulation of the carriers. To determine if a mutant form of ApoC-III exists in the plasma of A43T heterozygotes, we developed a multiple reaction monitoring assay to simultaneously measure peptides corresponding to the wild-type (WT) or A43T mutant forms of ApoC-III from the plasma of A43T heterozygotes by liquid chromatography and tandem mass spectrometry. After proteolysis of the plasma samples by AspN protease, a 20 amino acid peptide corresponding to either the WT form of ApoC-III containing alanine in original residue position 43 (**Figure 4.3 D**), and an analogous peptide with threonine corresponding to the mutant form (**Figure 4.3 E**), were monitored for fragments of expected m/z ratios and the relative amount of mutant to WT peptides were quantified for each sample. From this measurement, we found that A43T carrier plasma exhibited a detectable amount of A43T mutant ApoC-III peptides while noncarriers for ApoC-III variants demonstrated no observable mutant peptides (**Figure 4.3 F**). Notably, the ratio of mutant to WT ApoC-III, as assessed by ratio of the relative amounts of the mutant to WT peptides used for monitoring, was 0.14:1, which was lower than an expected ratio of 1:1. This suggests a significant imbalance in the amount of mutant ApoC-III in the circulation relative to WT ($P < 0.0001$, one-sample T-test; **Figure 4.3 F**). Using the observed ratios to determine the amount of WT vs. mutant ApoC-III in the plasma of the carriers, we found that A43T carriers also

demonstrate a trend towards reduced levels of the WT ApoC-III relative to that expected from the levels observed in noncarriers (**Figure 4.3 G**).

APOC3 A43T confers improved TRL metabolism in mice in vivo

Given the presence of a mutant ApoC-III protein in the plasma of A43T carriers, we concluded that A43T alleles encode for a secreted ApoC-III form and sought to further understand how this variant may confer LoF. We generated murine models to compare WT to A43T human ApoC-III *in vivo* through expression of adeno-associated virus (AAV) vectors expressing either form of the human *APOC3* transgene. AAV-mediated WT *APOC3* expression caused a dose-dependent elevation in circulating human ApoC-III, as well as plasma TGs in WT mice relative to controls administered a Null vector lacking a transgene (**Figure 4.5**). Because an AAV dose of 3×10^{11} genome copies (GC)/mouse conferred a robust increase in circulating ApoC-III to approx. 10-15 mg/dL, mirroring the range of normal ApoC-III concentration in humans, and also increased fasting TGs by two weeks of administration, we used this dose to compare the effect of Null, WT, and A43T *APOC3* AAVs on plasma lipids and ApoC-III levels in multiple murine models, including WT mice, *Apoc3* KO mice, and multiple humanized hyperlipidemic models (**Table 4.3**). In all models, WT *APOC3* expression raised while A43T expression failed to elevate fasting plasma TGs over five weeks after AAV administration (**Figure 4.6 A-C & Table 4.3**). In WT mice co-expressing cholesteryl ester transfer protein (CETP), a key mediator of lipid exchange between TRLs and HDL in humans but absent in rodents, A43T expression conferred markedly lower fasting TGs (**Figure 4.6 C**), higher HDL-C (**Figure 4.6 D**), and lower nonHDL-C (**Figure 4.6 E**), which was also observed after fractionation of plasma into lipoprotein fractions by fast-protein liquid chromatography (**Figure 4.6 F-G**). The effects of A43T on plasma lipids

through AAV-mediated expression in this ‘humanized’ mouse model thus recapitulate the protective lipoprotein profile of the human A43T variant carriers.

Given the roles of ApoC-III in delaying postprandial TG clearance through both inhibiting LPL activity on circulating TRLs and blunting hepatic TRL remnant clearance, we sought to understand if A43T impairs postprandial TG metabolism as well. We tested the *in vivo* activity of LPL mediated TG clearance in mice expressing Null, WT, or A43T *APOC3* by measuring TGs after intragastric administration of olive oil. Mice expressing WT *APOC3* demonstrated a robust increase in plasma TGs and delayed postprandial TG clearance relative to mice treated with Null virus (**Figure 4.6 H**). Mice expressing A43T exhibited no increase in postprandial TGs relative to Null treated mice, thus showing a significant increase in TG clearance relative to WT expressing mice. To test the impact of post-absorptive TG clearance, we administered ³H-triolein labeled TRLs in mice expressing Null, WT, or A43T *APOC3*. Similarly to the results for postprandial TG clearance, we observed a faster TRL-TG clearance of labeled TRLs in A43T expressing mice relative to WT expressing mice (**Figure 4.6 I**). Some prior reports have suggested that ApoC-III may also increase circulating TGs by promoting hepatic VLDL-TG secretion by facilitating the fusion of TG droplets on nascent VLDL particles[326,327]; we measured VLDL-TG secretion into blood in mice after administration of Poloxamer P407 [328], a competitive inhibitor of LPL, and found no differences in TG, ApoB100 or ApoB48 secretion rates in mice expressing A43T vs. WT or Null virus (**Figure 4.7**). Collectively, these data support a role for A43T in weakening ApoC-III’s effects on the *in vivo* metabolism of postprandial TRLs.

Because we observed a reduced amount of A43T mutant ApoC-III in the plasma of human A43T carriers, we next asked the impact of the A43T variant on circulating

ApoC-III levels in mice expressing A43T via AAV. Compared to mice expressing WT *APOC3*, mice expressing A43T demonstrated a <50% steady state levels of ApoC-III in fasting plasma (**Figure 4.6 J**). Notably, ApoC-III was present in the plasma of A43T expressing mice and was relatively constant from two to five weeks after AAV administration, while no detectable ApoC-III was observed in mice given Null virus. We next asked whether the cause of the reduced ApoC-III levels in A43T expressing mice was related to reduced expression of *APOC3* in mice treated with this AAV. We noted no differences in total hepatic copies of *APOC3* AAV in mice treated with WT vs. the mutant virus (**Figure 4.8**), suggesting that viral delivery to the liver did not differ between the viruses. Hepatic levels of human *APOC3* mRNA were also comparable in the mice expressing WT vs. A43T (**Figure 4.6 K**), as were relative levels of total hepatic ApoC-III protein as measured by immunoblotting (**Figure 4.6 L**). To determine if A43T impacted ApoC-III secretion from the liver, we administered ³⁵S labeled cysteine and methionine intravenously to mice expressing Null, WT, or A43T *APOC3* AAVs and measured newly synthesized, metabolically labeled ApoC-III secreted from the liver. Because ApoC-III exchanges between TRLs and HDLs and can be cleared by hepatic receptors when on metabolized TRL remnants, we blunted TRL clearance during this measurement by inhibiting LPL activity with Poloxamer P407. We compared the activity of labeled ApoC-III in plasma up to two hours after metabolic labeling and observed increased ApoC-III secretion in mice expressing either WT or A43T relative to Null treated mice (**Figure 4.6 M**), consistent with increased total ApoC-III production due to the transgene. Notably, we observed no difference in the secretion rate of ApoC-III from mice expressing A43T vs. WT ApoC-III (**Figure 4.6 N**). These findings are thus consistent with those from human

carriers and demonstrate that A43T confers expression of a mutant form of ApoC-III that is secreted and appears in the circulation.

A43T increases clearance and renal excretion of circulating ApoC-III

Since ApoC-III secretion did not differ between mice expressing A43T vs. WT ApoC-III, we asked if the cause of the lower steady-state ApoC-III conferred by A43T expression could be due to more rapid catabolism of the mutant protein. This notion is supported by the predicted helical structure of ApoC-III, which suggests that the alanine 43 residue (residue 23 of the mature ApoC-III protein after signal peptide cleavage) exists on the hydrophobic face of a predicted alpha helix that could mediate binding to lipids[327]. We tested the impact of A43T on the circulating stability of ApoC-III *in vivo* by measuring the clearance rate of ¹²⁵I tyramine cellobiose (TCB) labeled WT or A43T ApoC-III. Mice expressing WT or A43T *APOC3* were administered labeled ApoC-III of the same form pre-incubated with plasma from the same mice to allow initial binding to lipoproteins, and then this mixture was administered to mice intravenously. ApoC-III containing the A43T mutant exhibited a marked increase in clearance from the circulation over 24 hours after administration relative to that of WT (**Figure 4.9 A**), with a >3-fold increase in the total fractional catabolic rate in plasma (**Figure 4.9 B**). These results were also observed in mice administered ¹²⁵I labeled ApoC-III without the tyramine cellobiose modification, suggesting that the observed clearance and relative differences were not a result of conformational changes caused by the tyramine cellobiose moiety (**Figure 4.10 A-B**).

ApoC-III clearance is thought to occur largely through two pathways: receptor-mediated remnant lipoprotein clearance by the liver and filtration of unbound ApoC-III by the kidney [329-333]. However, the relative contributions of these two clearance

pathways and how they may contribute to the functional pool of circulating ApoC-III that promotes TRL retention remains unclear. We compared the uptake of ¹²⁵I TCB-labeled WT to A43T ApoC-III in the two main tissues mediating ApoC-III catabolism, the liver and kidney, at the end of the 24 hours of clearance and found that the A43T mutant protein was cleared ~40% less by the liver and ~40% more by the kidney than the WT (**Figure 4.9 C-D**). Additional clearance studies with radio-iodinated ApoC-III lacking the trapped TC moiety allowed the measurement of urinary ApoC-III clearance and showed that the A43T mutant protein was degraded and cleared into the urine at a higher rate than the WT protein (**Figure 4.9 E & Figure 4.10 C-E**). Indeed, the TCA precipitable activity of the mutant ApoC-III in the plasma constituted a lower fraction of the total activity at the one minute, three hour and six hour timepoints compared to that of the WT, suggesting that the mutant protein is degraded more rapidly and then filtered by the kidney for urinary excretion (**Figure 4.10 F-G**).

Since that we observed faster total ApoC-III clearance and increased renal uptake with decreased hepatic uptake for the A43T mutant, we hypothesized that the A43T mutant may impact the relative pools of lipoprotein-bound vs. free ApoC-III. To test this, we fractionated plasma from mice administered ¹²⁵I TCB labeled WT vs. A43T ApoC-III one minute after administration of the ApoC-III / plasma mixture and measured the relative amount of ApoC-III bound to lipoprotein fractions vs. free. For WT ApoC-III, approximately 40% of the total radiolabeled protein was bound to either VLDL/TRLs or HDL fractions, with the remainder of the protein being unbound. In contrast, TRLs from mice given the A43T mutant lacked the radiolabeled ApoC-III entirely, and HDL-bound ApoC-III was also reduced. A ~35% increase in unbound ApoC-III was observed from the fractionated plasma from this group (**Figure 4.9 F**). We further tested whether A43T

also blunted binding of ApoC-III to human lipoproteins by measuring the association of radiolabeled ApoC-III with human plasma lipoproteins *in vitro*. Consistent with the *ex vivo* findings from mouse plasma, ~50% of total WT ApoC-III bound to TRL and HDL fractions after incubations for one, five, and 24 hours at 37 °C (**Figure 4.9 G & Figure 4.11 A-C**). In contrast, the A43T mutant demonstrated reduced binding to TRL fractions and HDL with ~45% increased activity in the unbound fractions. We further explored the impact of A43T on lipoprotein binding by testing the individual association of WT vs. mutant ApoC-III with isolated human lipoproteins. A43T exhibited similar 50-60% reductions in binding to TRLs (**Figure 4.9 H & Figure 4.11 D**) and HDLs (**Figure 4.9 I & Figure 4.11 E**), with increased free protein appearing in unbound fractions. Similar results were obtained in experiments in which VLDL and HDL were mixed with radiolabeled WT or A43T mutant ApoC-III and lipoprotein-bound ApoC-III content was measured after ultracentrifugation (data not shown). These results support the notion that A43T impairs the binding capacity of ApoC-III for lipoproteins, thus promoting its dissociation from lipoproteins *in vivo* and resulting in increased free ApoC-III clearance via the renal clearance pathway.

The mechanism of A43T-mediated disruption of ApoC-III's ability to bind lipoproteins may have implications for the development of targeted therapies to block ApoC-III function in the circulation. We thus wanted to better understand how A43T impairs ApoC-III:lipoprotein association, and the potential implications this might have on ApoC-III's ability to impair TG clearance. To this end, we first asked if the binding impairment by A43T was a result of a protein:lipid vs. protein:protein interaction. We generated large TRL-like emulsions composed of TG and phospholipids but devoid of any apolipoproteins and tested the ability of ¹²⁵I labeled ApoC-III to bind these emulsions

compared to binding of isolated TRLs and similar emulsions containing the same lipid content in the presence of albumin. For all measurements, A43T mutant ApoC-III exhibited approximately the same 50% reduction in binding to emulsions or TRLs with commensurate increases in the amount of unbound ApoC-III (**Figure 4.11 F-G**). These data were suggestive of an ApoC-III:lipid interaction mediating the association of ApoC-III to lipoproteins that is reduced by the A43T mutant. Consistent with this, we found that A43T ApoC-III exhibited a ~five-fold dissociation constant (K_D) in binding to an immobilized surface of phospholipids *in vitro* by surface plasmon resonance (**Figure 4.9 J**). Consistent with the lipid-binding deficits driving the LoF conferred by A43T, we observed that, for equal concentrations of ApoC-III, the A43T mutant was less capable of inhibiting LPL activity *in vitro* against large TG-rich emulsions (**Figure 4.12 A-B**) and an Intralipid TG substrate *in vitro* relative to WT (2.7-fold increased IC_{50} , $P < 0.001$, **Figure 4.12 C-D**).

Discussion

Our data thus unveil one mechanism of LoF of *APOC3* as revealed by the study of a naturally occurring missense variant, A43T. We show that, though the lipid phenotype and steady state ApoC-III levels conferred by this variant are similar to other *APOC3* LoF variants, A43T confers ApoC-III LoF and thus protection from dyslipidemia and CHD risk by a different mechanism. Notably, the increased ApoC-III turnover and renal clearance of the mutant ApoC-III as a consequence of the A43T variant predominately enhances the lipolytic pathway of TG turnover *in vivo*. This contrasts with prior studies suggesting that ApoC-III may primarily inhibit TRL turnover at the level of delaying remnant particle clearance rather than inhibiting the LPL pathway [99,100,333-

335]. Our results here suggest that the A43T variant may be most relevant in rebalancing the pool of lipoprotein-bound to free ApoC-III in the circulation, which may thus promote lipolysis of TRL-TGs in the periphery as the mutant ApoC-III is no longer retained on the particles due to defective binding. The impact of this variant, and more broadly, the promotion of ApoC-III dissociation from TRLs in the peripheral circulation on the relative ApoC-III content of the remnant particles that interface with hepatic receptors thus remains an important area for continued investigation.

Based on our findings, therapeutic discovery related to promoting disruption of ApoC-III binding to lipoprotein particles may be efficacious in reducing ApoC-III levels by promoting renal clearance, a mechanism which appears safe both from human and murine model studies of this variant. Prior studies of murine models of human *APOC3* and *CETP* co-expression demonstrated a synergistic proatherogenic effect resembling the human familial combined hyperlipidemia phenotype [318]; our data here from the humanized murine models expressing *CETP* suggest that the protective effect of reduced ApoC-III binding to lipoproteins is possibly enhanced in the setting of CETP-mediated lipid exchange between TRLs and HDL. This supports the notion that such a strategy to reduced ApoC-III function would be efficacious in humans. Therapeutic leveraging of this pathway based on the findings from *APOC3* A43T to lower TGs could mirror the approach to targeting PCSK9 to reduce LDL-C [336,337], which emerged from human genetics studies of both LoF and gain-of-function nonsynonymous coding variants [32,38,39], which were shown to be protective from CHD. As supported by that example, the success and safety of developing treatments for ApoC-III lowering may benefit from further comparisons between a putative drug in development with the metabolic signatures of LoF variant carriers and appropriate animal models.

Figure 4.1

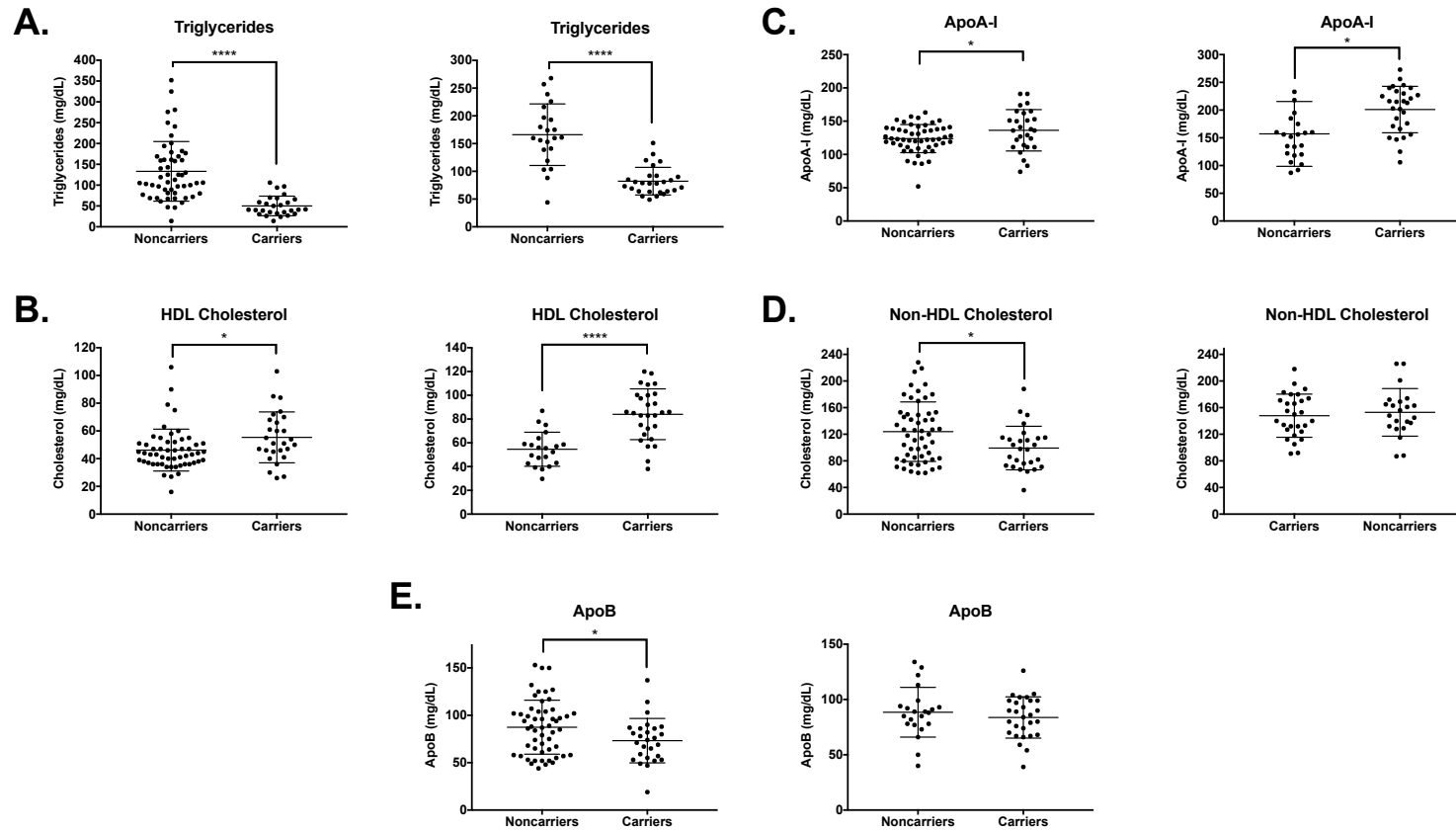


Figure 4.1: Plasma lipids in carriers of all *APOC3* LoF variants vs. noncarriers in CGI BioBank and HHDL cohorts. A. Triglycerides (TGs) in overnight fasted plasma of carriers of any of the four *APOC3* LoF variants vs. matched controls. Carriers were

identified in two cohorts, the CGI BioBank cohort (left) and the HHDL cohort (right). Controls were matched for age and gender for each cohort. **B.** HDL cholesterol (HDL-C) from fasting plasma of *APOC3* LoF variant carriers vs. controls shown in A. **C.** ApoA-I concentration in fasting plasma from *APOC3* LoF variant carriers vs. controls shown in A. **D.** Non-HDL cholesterol concentration in *APOC3* LoF variant carriers vs. controls. **E.** ApoB concentration in *APOC3* LoF variant carriers vs. controls. All measurements show mean \pm S.D. where appropriate. For A-C and G, * $P < 0.05$, **** $P < 0.0001$, Student's unpaired T-test. *Contributions: All data in this figure generated and analyzed by SAK.*

Figure 4.2

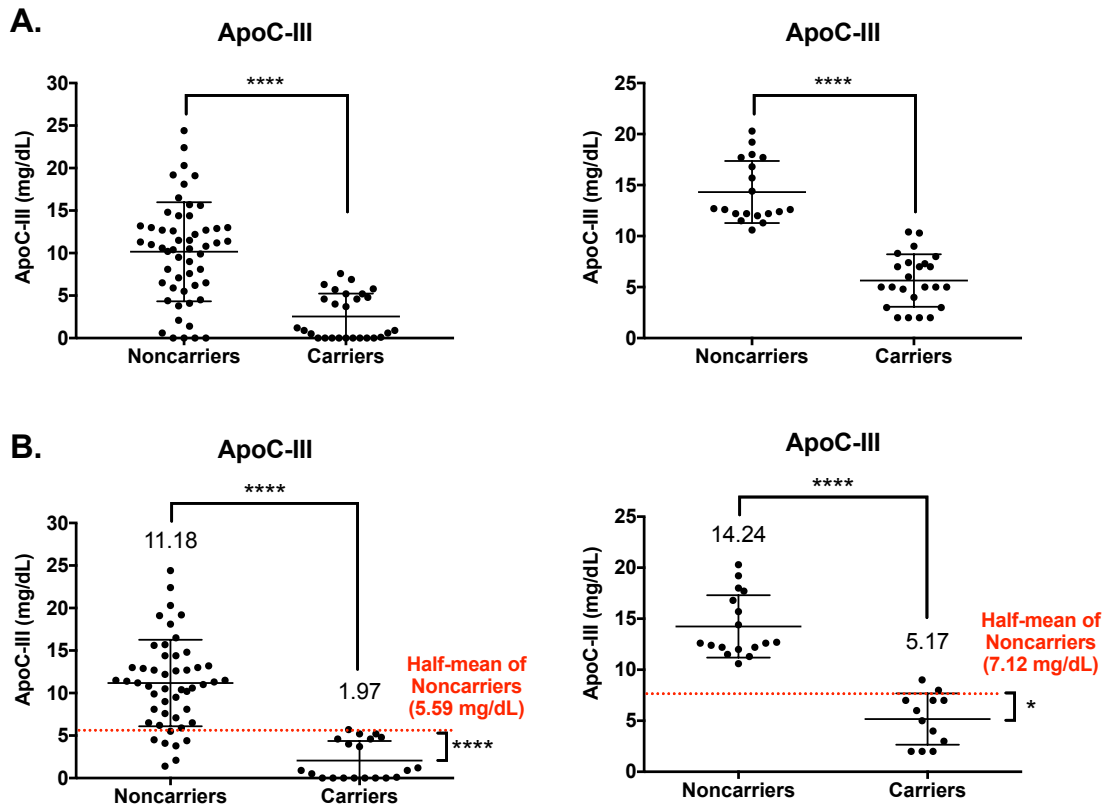


Figure 4.2: Plasma ApoC-III levels in carriers of *APOC3* LoF variants vs. controls from CGI BioBank and HHDL cohorts. A. Plasma ApoC-III levels in overnight fasted plasma from carriers of any of the four *APOC3* LoF variants vs. noncarriers. Measurements were made in a sample of carriers and age- and gender-matched controls from the CGI BioBank cohort (left) and HHDL cohort (right). **B.** ApoC-III concentrations in carriers of the three LoF variants predicted to abolish splicing or expression (R19X, IVS2+1 G>A, IVS3+1 G>T) vs. controls in the CGI BioBank cohort (left) and HHDL cohort (right). Mean total concentrations for each group are indicated above the data points. Half of the mean levels observed in the noncarriers for each cohort are indicated by the red dotted line. All measurements show mean \pm S.D. where

appropriate. For comparisons of total ApoC-III concentrations between carriers and controls in A and B, ****P<0.0001, Student's unpaired T-test. For comparisons of the concentrations of carriers of R19X, IVS2+1 G>A and IVS3+1 G>T to the value of half of the mean total concentration of the noncarriers for each cohort in B, ****P<0.0001, one-sample T-test with an expected value of the half-mean of the noncarriers for each respective cohort. *Contributions: All data in this figure generated and analyzed by SAK.*

Figure 4.3

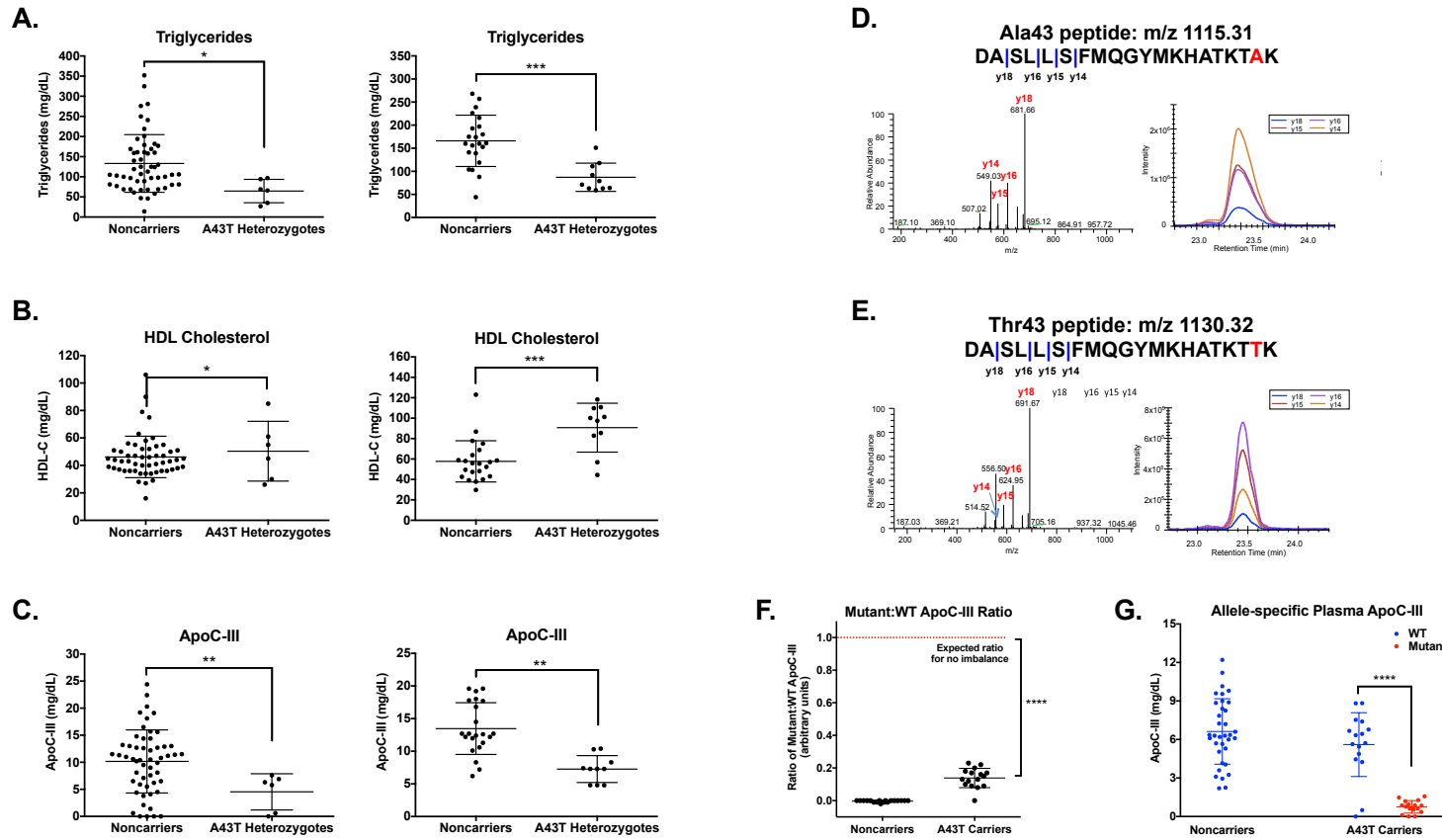


Figure 4.3: Human *APOC3* A43T carriers exhibit reduced ApoC-III levels due to lower mutant protein. A. Triglycerides (TGs) in overnight fasted plasma of carriers of A43T vs. matched controls. Carriers were identified in two cohorts, the CGI BioBank cohort

(left) and the HHDL cohort (right). Controls were matched for age and gender for each cohort. **B.** HDL cholesterol (HDL-C) from fasting plasma of A43T carriers vs. controls shown in A. **C.** ApoC-III concentration in fasting plasma from A43T carriers vs. controls shown in A. **D.** (Top) Peptide sequence for AspN-generated 20 amino acid peptide corresponding to WT ApoC-III. Position of the variant amino acid between WT and mutant peptides is shown in red. (Bottom) Mass spectra of fragment ions generated and observed from peptide above after fragmentation in triple quadrupole mass spectrometer (left), and fragment ion intensities for respective retention time (right). For each peptide, four fragment ions were monitored that contained the variant amino acid. **E.** (Top) Peptide sequence and (Bottom) mass spectra of fragment ions for AspN-generated 20 amino acid peptide corresponding to mutant ApoC-III. Position of the variant amino acid between WT and mutant peptides is shown in red. **F.** Ratio of mutant:WT ApoC-III as determined by fitting relative intensities of mutant:WT fragment ions for the parent peptides shown in D-E from plasma samples of A43T carriers (left) vs. noncarriers (right). The mutant:WT ratio of the group of A43T carriers was compared to an expected ratio of 1:1 for no imbalance by a one-sample T-test with expected mean of 1.0. **G.** Allele-specific ApoC-III concentration in plasma samples of noncarriers and A43T carriers as measured in F. Mutant:WT ratios from A43T carriers in F were multiplied by total ApoC-III concentrations for each participant to determine mutant and WT ApoC-III concentrations. For noncarriers, total ApoC-III concentration was divided by two to estimate allele-specific ApoC-III levels. All measurements show mean \pm S.D. where appropriate. For A-C and G, *P<0.05, **P<0.01, ***P<0.001, Student's unpaired T-test. *Contributions: Data in panels A-C generated and analyzed by SAK; data in panels D-G generated and analyzed by SAK, Xuemei Zeng, and Nathan Yates using samples processed by SAK.*

Figure 4.4

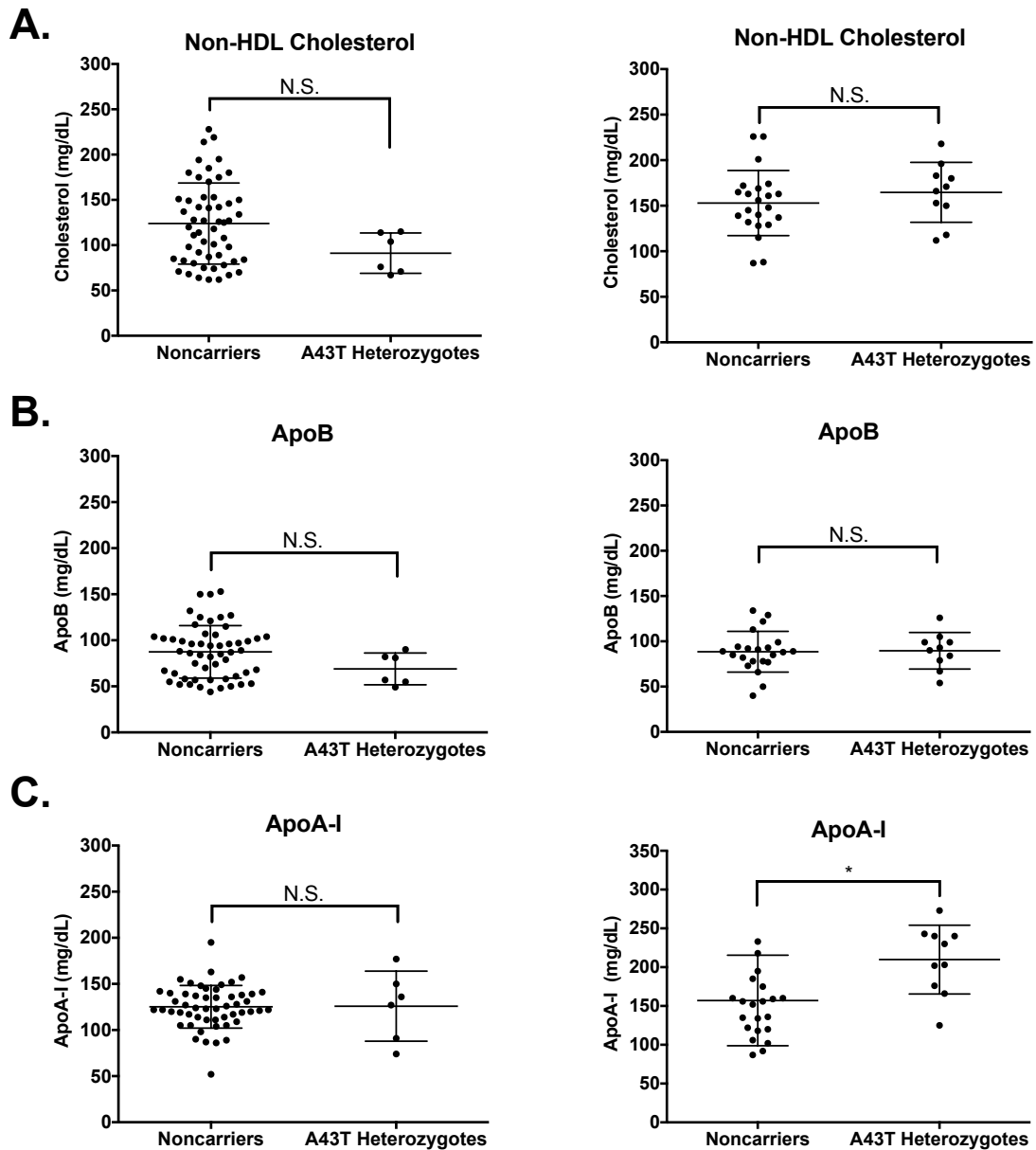


Figure 4.4: Additional lipids measurements in carriers of the *APOC3* A43T variant vs. noncarrier controls in the CGI BioBank and HHDL cohorts. **A.** Non-HDL cholesterol in overnight fasted plasma of A43T carriers vs. age- and gender-matched controls from the CGI BioBank cohort (left) and HHDL cohort (right). **B** ApoB

concentrations in A43T variant carriers vs. controls from the two cohorts in A. **C.** ApoA-I concentrations in A43T carriers vs. controls from the two cohorts in A. All measurements show mean \pm S.D. where appropriate. *P<0.05, Student's unpaired T-test. *Contributions:* *All data in this figure generated and analyzed by SAK.*

Figure 4.5

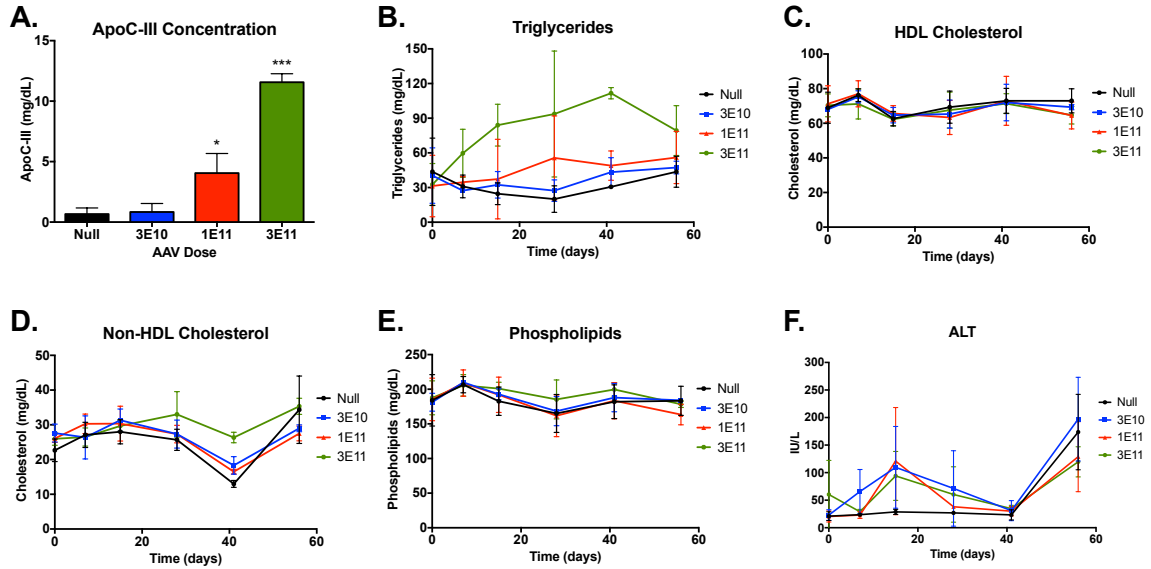


Figure 4.5: Dose-optimization studies of WT *APOC3* AAV in mice. **A.** Plasma human ApoC-III concentration in C57BL/6 WT mice administered the indicated doses (genome copies of vector per mouse) of WT *APOC3* AAV by intraperitoneal injection (N=3 mice per group). Measurements were made eight weeks after AAV administration. **B.** Plasma TGs from four hour fasted plasma from mice in A over eight weeks. **C.** Plasma HDL-C from mice in A. **D.** Plasma Non-HDL-C in mice from A. **E.** Plasma phospholipids from mice in A. **F.** Plasma alanine aminotransferase (ALT) levels in mice from A. All measurements show mean \pm S.D. where appropriate. *P<0.05, ***P<0.001, Student's unpaired T-test compared to Null group. *Contributions: All data generated and analyzed by SAK.*

Figure 4.6

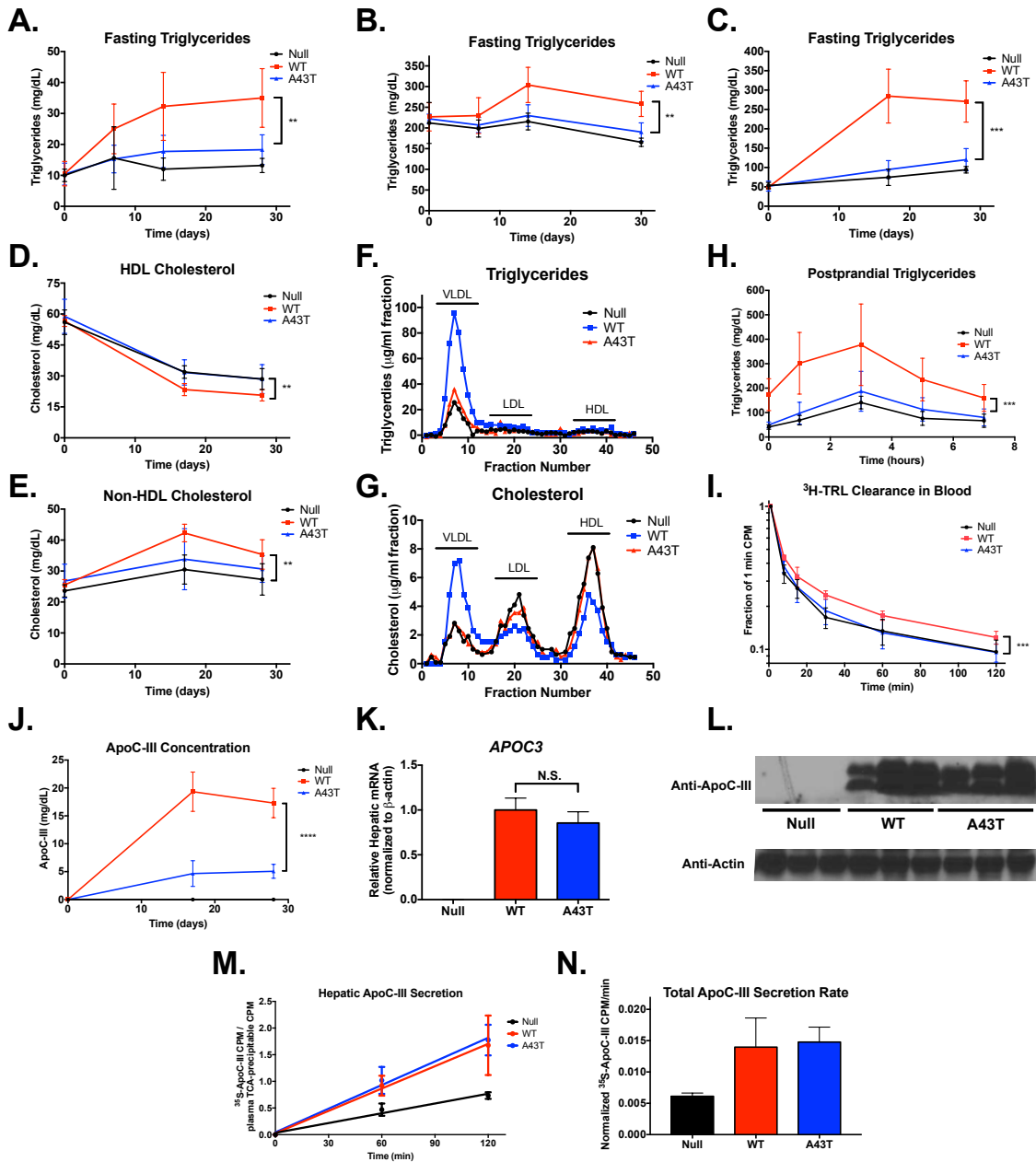


Figure 4.6: Mice expressing *APOC3* A43T have reduced TRLs and circulating ApoC-III *in vivo*. A. Fasting plasma TGs in *Apoc3* KO mice expressing Null, WT *APOC3*, or A43T *APOC3* after AAV-mediated expression in the liver for the indicated

timepoints. Plasma was drawn after 4 hours of fasting. **B.** Fasting plasma TGs in human APOB transgenic, Apobec1 KO (LahB WT) mice expressing the indicated transgenes. **C.** Fasting TGs in WT mice expressing the indicated transgenes and co-expressing human CETP by AAV. **D.** Plasma HDL-C in the mice from C. **E.** Non HDL-C in the mice from C. **F.** TGs in FPLC-separated plasma fractions from Day 28 plasma from mice in C. Lipoprotein fractions are indicated above fraction numbers. **G.** Cholesterol in FPLC fractions from F. **H.** Postprandial TGs in *Apoc3* KO mice expressing indicated transgenes by AAV and co-expressing CETP following olive oil gavage. **I.** Plasma ^3H activity (expressed as fraction of activity in 1 minute plasma samples) in WT mice expressing indicated transgenes after intravenous administration of ^3H triolein-labeled human TRLs. **J.** Plasma ApoC-III concentration in fasting plasma from mice from C. **K.** Hepatic *APOC3* mRNA levels (normalized to that of actin) in mice from C. **L.** Total hepatic ApoC-III immunoblots from liver lysates of mice from C after sacrificing 28 days after AAV administration. Blots for actin are shown below for each corresponding ApoC-III blot as a loading control. **M.** Plasma ^{35}S total ApoC-III activity (normalized to activity in 2 μl of plasma after TCA precipitation) for mice expressing the indicated transgenes 35 days after AAV administration. **N.** Secretion rates as measured by the slope of the curves in M. All measurements show mean \pm S.D. where appropriate. ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, Two-way ANOVA, WT vs. A43T groups. *Contributions: All data in this figure generated and analyzed by SAK.*

Figure 4.7

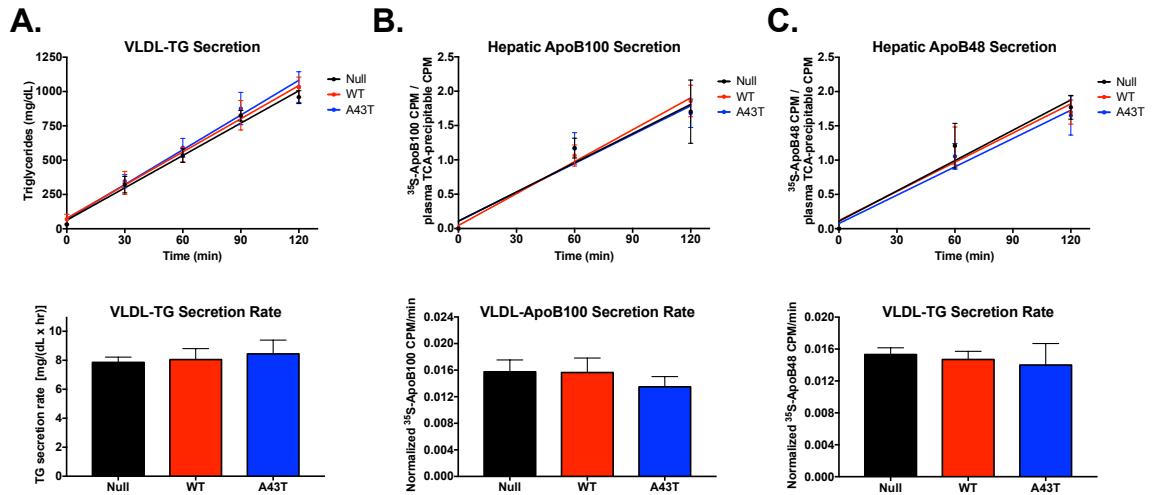


Figure 4.7: VLDL TG and ApoB secretion in mice expressing WT vs. A43T APOC3.

A. (top) Plasma TG levels over time after Poloxamer P407 and ^{35}S administration in WT mice expressing Null or APOC3 AAVs for four weeks. Curves were fitted to a line in GraphPad Prism. (bottom) Secretion rates as measure by the slope of the curves in A.

B. (top) Plasma ^{35}S -ApoB100 activity (normalized to activity in 2 μl of plasma after TCA precipitation) for mice expressing the indicated transgenes four weeks after AAV administration. ApoB100 bands were excised from gels after SDS-PAGE and activity measured by scintillation counting and normalized to the TCA precipitable activity from plasma from the same plasma samples. (bottom) Secretion rates as measure by the slope of the curves in B.

C. Plasma ^{35}S -ApoB48 activity from mice in A and B. ApoB48 bands were excised after SDS-PAGE, measured for activity and normalized to TCA precipitable activity from the same plasma samples. (bottom) Secretion rates as measured by the slope of the curves in C. All measurements show mean \pm S.D. where appropriate. *Contributions: All data generated and analyzed by SAK.*

Figure 4.8

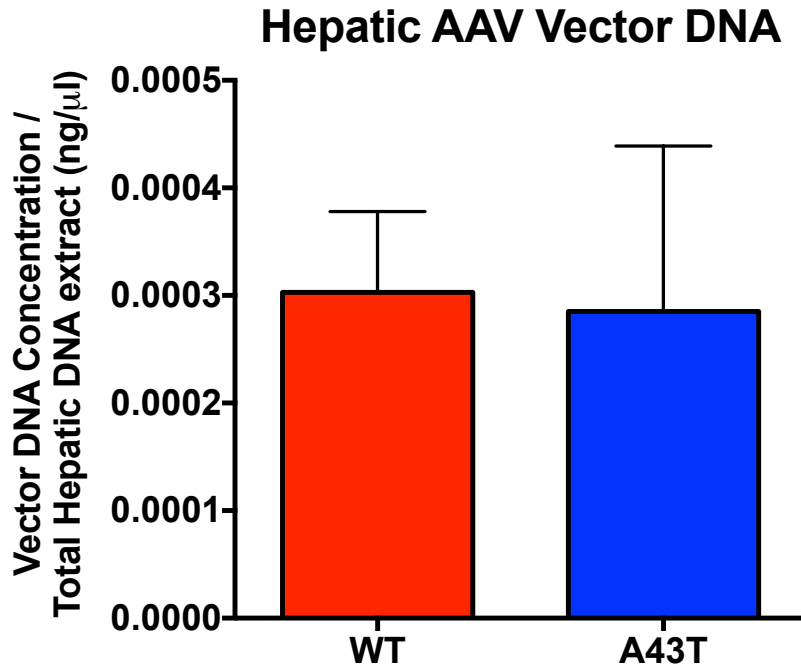


Figure 4.8: Hepatic AAV vector levels in mice expressing WT vs. A43T APOC3 AAVs. AAV vector DNA copy number was measured in the livers of mice expressing WT vs. A43T APOC3 AAV. Total DNA was isolated from 25 mg of liver from each mouse and used to measure AAV copy number by quantitative RT-PCR for the rabbit beta-globin polyA sequence. Measurements show mean \pm S.D. where appropriate. *Contributions: All data generated and analyzed by SAK.*

Figure 4.9

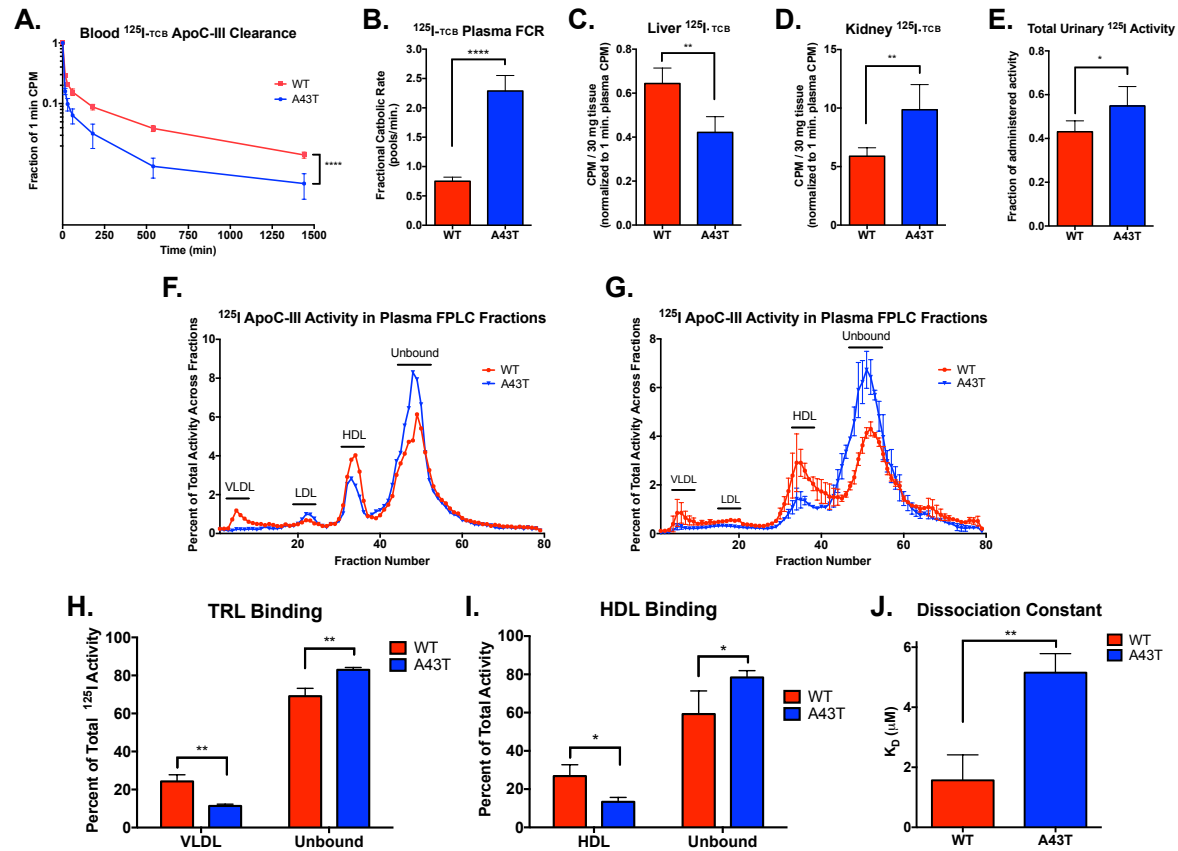


Figure 4.9: A43T promotes circulating ApoC-III clearance and augments renal uptake by perturbing binding to lipoproteins.

A. Plasma ^{125}I tyramine cellobiose (TCB) labeled WT or A43T ApoC-III clearance in plasma in LahB WT mice over 24 hours. Mice

expressing WT *APOC3* by AAV were administered WT labeled ApoC-III and those expressing A43T were administered the radiolabeled A43T ApoC-III. Normalized ^{125}I activity relative to 1 minute plasma activity is shown. **B.** Fractional catabolic rate of plasma ^{125}I activity, as measured by WinSAAM multi-exponential modeling program. Catabolic rates were calculated for each mouse individually. **C.** Hepatic ^{125}I TCB activity in 30 mg of tissue for mice in A. Activity was normalized to 1 minute plasma activity. **D.** Kidney ^{125}I TCB activity in 30 mg of tissue from mice in A. **E.** Total ^{125}I activity measured in urine collected from WT vs. A43T expressing mice at the end of 24 hours of collection following intravenous ^{125}I ApoC-III administration. **F.** ^{125}I activity in FPLC fractions of 1 minute plasma from mice in A after pooling by indicated group. Activity is expressed as fraction of total activity in plasma pool prior to FPLC separation. **G.** ^{125}I activity in FPLC fractions after incubation of ^{125}I labeled WT vs. A43T ApoC-III (1 μg) with human plasma (200 μl) for one hour at 37 °C. Results show mean of three replicate experiments. **H.** Relative ^{125}I activity in VLDL fractions vs. unbound protein fractions after incubation of ^{125}I labeled WT vs. A43T ApoC-III (1 μg) with isolated human VLDL (200 μg protein) for one hour at 37 °C. **I.** Relative ^{125}I activity in HDL fractions vs. unbound protein fractions after incubation of ^{125}I labeled WT vs. A43T ApoC-III (1 μg) with isolated human HDL (200 μg protein) for one hour at 37 °C. **J.** Dissociation constant from measurement of K_{on} and K_{off} constants from binding of WT or A43T ApoC-III to dimyristoylphosphatidylcholine surfaces by surface plasmon resonance. For H-J, bars indicate mean of results from three replicate experiments. All measurements show mean \pm S.D. where appropriate. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, Student's unpaired T-test; **** $P < 0.0001$, 2-way ANOVA (A), **** $P < 0.0001$, Student's unpaired T-test (B). *Contributions: All data generated and analyzed by SAK.*

Figure 4.10

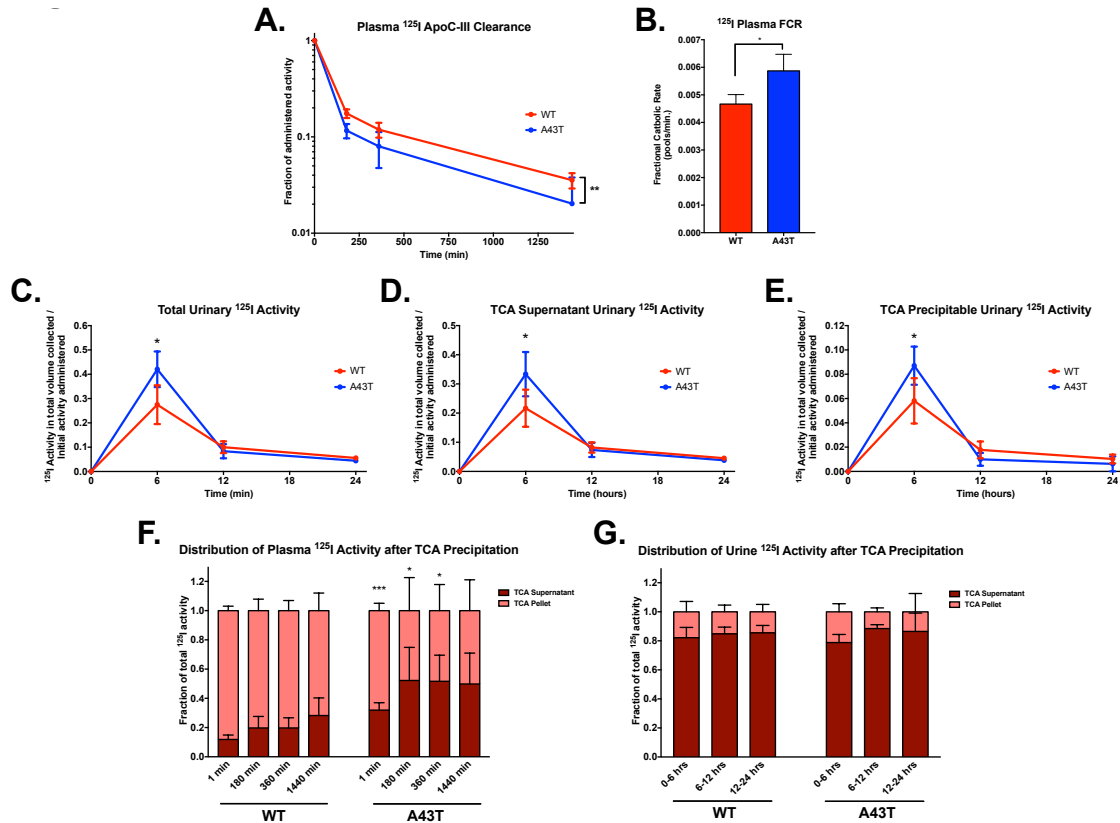


Figure 4.10: Urinary clearance of ^{125}I -labeled WT vs. A43T ApoC-III in mice. A. Plasma ^{125}I clearance curves over 24 hours for radiolabeled WT vs. A43T ApoC-III in WT mice expressing either WT or A43T *APOC3* AAVs (N=4 mice per group). Mice expressing WT *APOC3* AAV received WT iodinated ApoC-III by intravenous administration and A43T *APOC3*-expressing mice received iodinated A43T ApoC-III protein. **B.** Fractional catabolic rate of plasma ^{125}I activity, as measured by WinSAAM multi-exponential modeling program. Catabolic rates were calculated for each mouse individually. **C.** Total urinary ^{125}I activity at the indicated time intervals over the course of the ^{125}I ApoC-III clearance study. Urine was collected individually at the indicated

timepoints for each mouse and ^{125}I activity was measured and removed from metabolic cages prior to the next collection period up through 24 hours. **D.** ^{125}I activity in TCA soluble fraction of urine collected from each mouse at the indicated time intervals. 20 μl of urine was subjected to TCA precipitation and TCA soluble fractions were collected and re-measured for ^{125}I activity. The fraction of the 20 μl of urinary ^{125}I activity appearing in the TCA soluble fraction was used to calculate the proportion of TCA soluble ^{125}I activity in the total urine collected for each mouse for the indicated time intervals. **E.** ^{125}I activity in the TCA precipitable fraction of the urine collected from each mouse at the indicated time intervals. The TCA precipitable activity from the urine used in D was measured for ^{125}I activity and the fraction of total urinary ^{125}I activity prior to TCA precipitation that appeared in the TCA precipitate from 20 μl urine was used to calculate the TCA precipitable ^{125}I activity from the total urine collected for each mouse at each time interval. **F.** Fraction of plasma ^{125}I activity in TCA supernatant vs. pellet for each timepoint from 20 μl of plasma from mice administered ^{125}I -labeled WT vs. A43T ApoC-III. **G.** Fraction of urinary ^{125}I activity in TCA supernatant vs. pellet for each time interval of urine collection for mice administered ^{125}I labeled WT vs. A43T ApoC-III. TCA soluble vs. precipitable ^{125}I activity was measured from 20 μl of urine for each mouse for each time interval as indicated in D-E. All measurements show mean \pm S.D. where appropriate. For A, **P<0.01, two-way ANOVA of WT vs. A43T clearance curves. For B-F, *P<0.05, ***P<0.001, Student's unpaired T-test. For F, comparisons are of the fraction of total activity in TCA precipitable fraction for WT vs. A43T for a given time interval.

Contributions: All data generated and analyzed by SAK.

Figure 4.11

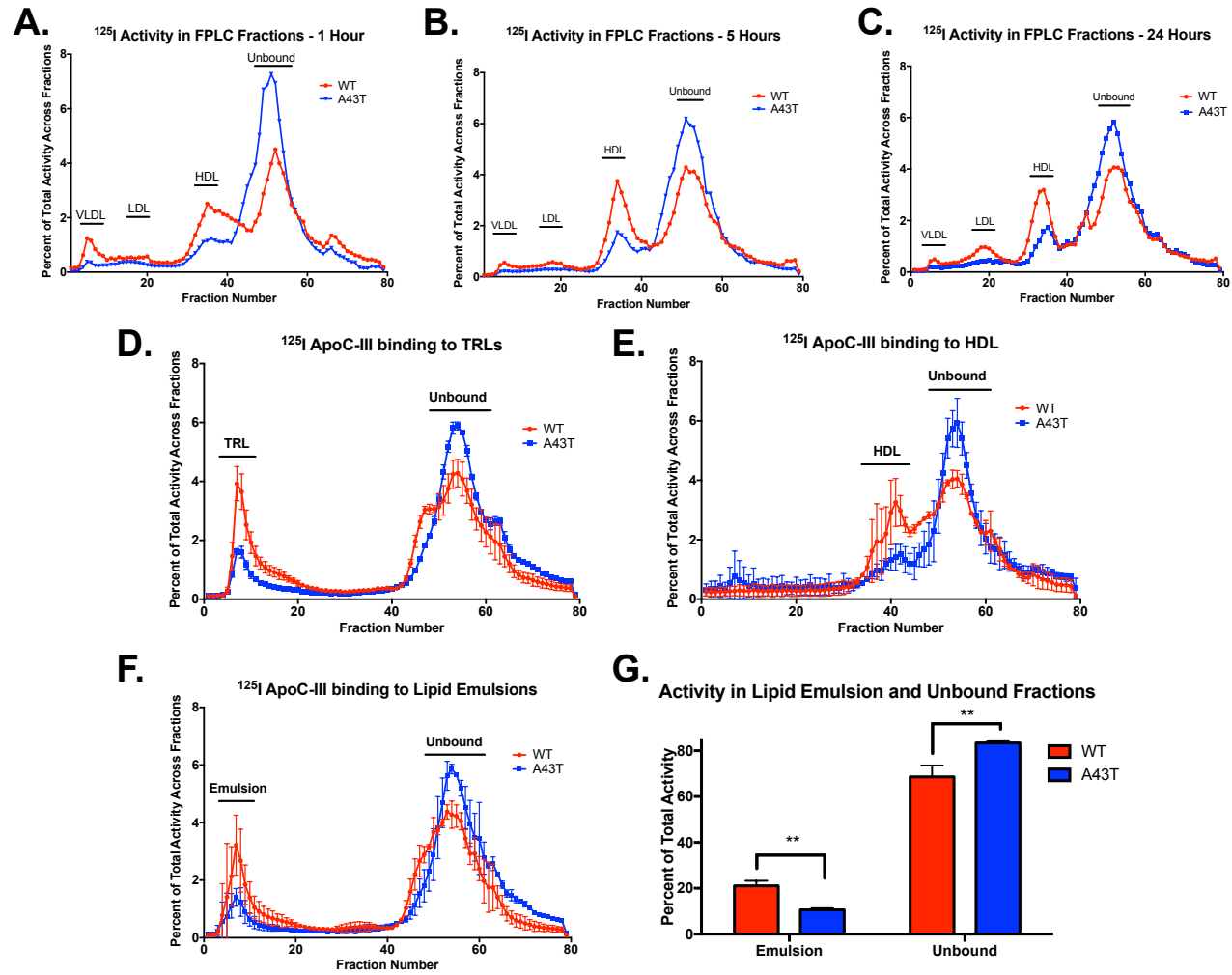


Figure 4.11: Binding of WT vs. A43T ApoC-III to lipid and human lipoprotein fractions. **A.** ^{125}I activity from FPLC-separated fractions of human plasma after incubation of ^{125}I -labeled WT vs. A43T ApoC-III with human plasma (1 μg ApoC-III and 200 μl plasma) for 1 hour at 37 $^{\circ}\text{C}$. Activity is expressed as the fraction of total ^{125}I activity in the ApoC-III:plasma mixture prior to FPLC separation. **B.** ^{125}I activity from FPLC-separated fractions of human plasma incubated with ^{125}I -labeled WT vs. A43T ApoC-III for 5 hours at 37 $^{\circ}\text{C}$. **C.** ^{125}I activity from FPLC-separated fractions from human plasma incubated with ^{125}I -labeled WT vs. A43T ApoC-III for 24 hours at 37 $^{\circ}\text{C}$. **D.** ^{125}I activity from FPLC-separated fractions of human TRLs incubated with ^{125}I -labeled WT vs. A43T ApoC-III (1 μg ApoC-III and 100 μg TRL protein) for 1 hour at 37 $^{\circ}\text{C}$. Graphs show mean \pm S.D. of three replicate experiments. **E.** ^{125}I activity from FPLC-separated fractions of human HDL₃ incubated with ^{125}I -labeled WT vs. A43T ApoC-III (1 μg ApoC-III and 200 μg HDL₃ protein) for 1 hour at 37 $^{\circ}\text{C}$. Graphs show \pm S.D. of three replicate experiments. **F.** ^{125}I activity from FPLC-separated fractions of large lipid emulsions incubated with ^{125}I -labeled WT vs. A43T ApoC-III for 1 hour at 37 $^{\circ}\text{C}$. Graphs show \pm S.D. of three replicate experiments. **G.** Relative ^{125}I activity in lipid emulsion fractions vs. unbound protein fractions from experiments in F. Bars indicate sum of relative amount of activity in fractions from three replicate experiments. All measurements show mean \pm S.D. where appropriate. For A-C and G, $**P < 0.01$, Student's unpaired T-test. *Contributions: All data generated and analyzed by SAK.*

Figure 4.12

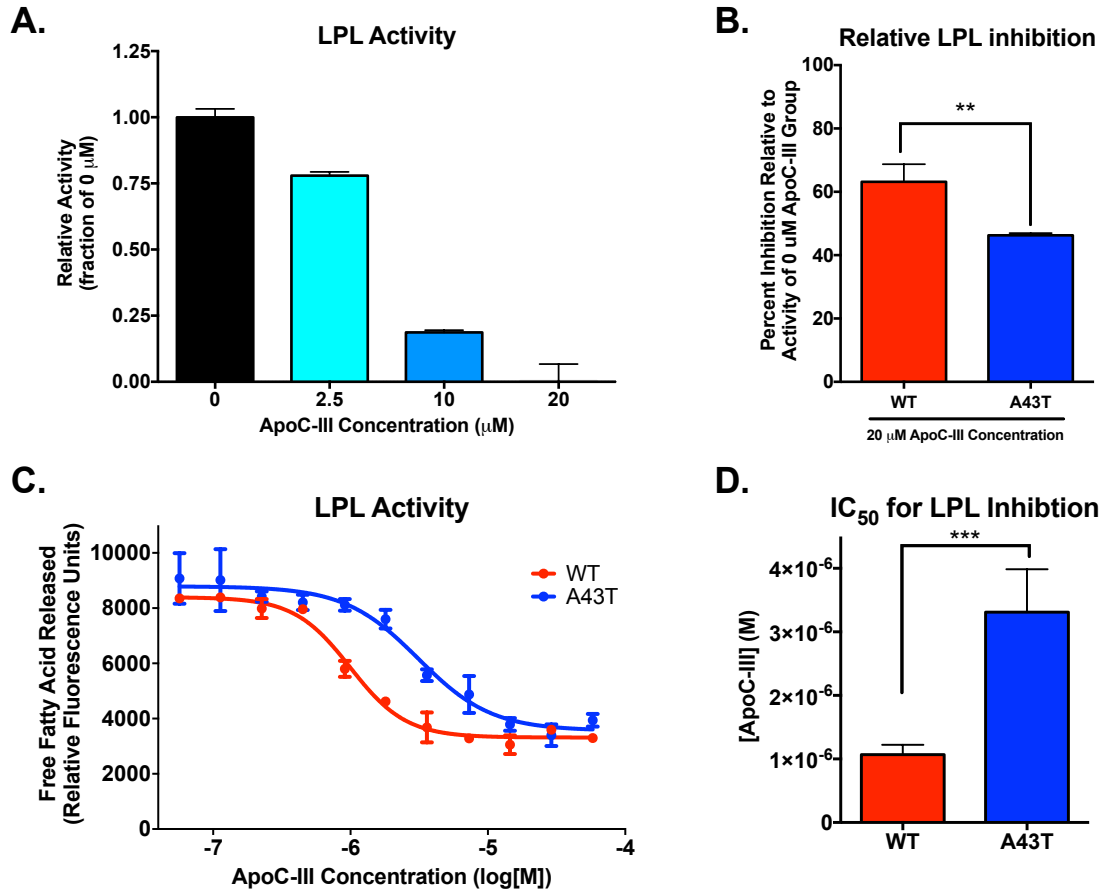


Figure 4.12: Impact of A43T on LPL inhibition by ApoC-III. **A.** Relative LPL activity against ^3H -triolein labeled large lipid emulsion substrate with increasing reaction concentrations of WT ApoC-III. LPL activity was measured as the amount of ^3H -oleic acid released over 30 minutes of incubation at 37 °C. Data is shown as the fraction of ^3H -oleic acid released relative to that from the 0 μM ApoC-III group. Each bar represents triplicate reactions. **B.** Relative LPL inhibition of WT vs. A43T ApoC-III using the assay in A with an ApoC-III concentration of 20 μM . LPL inhibition is graphed as the relative difference in LPL activity between a 0 μM ApoC-III group and each of the 20 μM ApoC-III

groups (WT and A43T). All reactions took place for 30 minutes at 37 °C and were performed in triplicate. **C.** Relative LPL activity against an Intralipid emulsion substrate with concentrations of WT or A43T ApoC-III as indicated. LPL activity was measured as the amount of free fatty acids released for each reaction as measured by relative light units after addition of colorimetric reagent to reaction incubations. Each reaction was performed in triplicate. **D.** IC₅₀ concentration of ApoC-III for WT vs. A43T from the activity curves in C. Each bar represents triplicate reactions. All measurements show mean ± S.D. where appropriate. **P<0.01, ***P<0.001, Student's unpaired T-test. *Contributions: Data in panels A-B generated and analyzed by SAK; data in panels C-D generated by a James Landro and analyzed by SAK.*

Table 4.1

Table 4.1: Discovery and Carrier Frequency of <i>APOC3</i> LoF Variants through Exome Chip Genotyping									
Discovery Cohort	Total Number of Participants	APOC3 LoF Variant Carriers Identified				Total Number of APOC3 LoF Variant Carriers Identified	APOC3 LoF Allele Frequency	Carrier Frequency	Association of LoF Carrier Frequency in HHDL vs. Other Cohorts (Chi-square test)
		R19X	IVS2+1 G>A	A43 T	IVS3+1 G>T				
HHDL Cohort	2,736	6	14	14	2	36	0.0066 (~1/152)	0.0132 (~1/76)	
CGI BioBank Cohort	5,744	4	17	6	1	28	0.0024 (~1/410)	0.0049 (~1/205)	P < 0.0001
NHLBI ESP Participants	6,496	3	19	15	6	43	0.0033 (~1/302)	0.0066 (~1/151)	P = 0.0028
NHLBI ESP APOC3 CHD Case-Control Genotyping Cohort	110,970					498	0.0022 (~1/446)	0.0045 (~1/223)	P < 0.0001

Table 4.1: Discovery and carrier frequency of *APOC3* LoF variants through exome chip genotyping. The indicated numbers of participants for the HHDL cohort (HDL-C >95th percentile for age and gender) and CGI BioBank cohort (normally distributed HDL-C cohort) were genotyped using the Exome Chip for the four *APOC3* LoF variants, with the indicated number of carriers for each variant identified. For comparison, the carriers of the *APOC3* LoF variants were also determined for the National Heart Lung and Blood Institute Exome Sequencing Project (NHLBI ESP) participants (<http://evs.gs.washington.edu/EVS/>) and the previously reported NHLBI ESP CHD case-control genotyping participants[106]. The aggregate variant allele frequency for any of the four *APOC3* LoF variants was determined based on the total number of variant alleles identified relative to the total number of alleles in the cohort. The carrier

frequency was determined to be twice the LoF allele frequency for each cohort. The carrier frequencies for each of the normally distributed HDL cohorts (CGI BioBank cohort and the two NHLBI ESP participant groups) were compared to that of the HHDL cohort by a Chi-square test for which the P-value for significantly different carrier frequencies is given. *Contributions: Genotyping data generated by core facilities and analyzed by SAK; final analysis and statistical testing performed by SAK.*

Table 4.2

Table 4.2: Predicted Impact of APOC3 LoF Variants on mRNA processing and splicing							
R19X (g.116701353C>T; rs76353203)							
Genomic Coordinates	Position relative to natural site	Initial(Ri)	Final(Ri)	Δ Ri	Fold change	% Binding (Final/Initial)	Predicted Change
hnRNP H binding prediction							
116701372	19	0.9	0.9	0	1	100	No change
116701378	25	4.4	4.4	0	1	100	No change
116701379	26	4.2	4.2	0	1	100	No change
116701386	33	4.8	4.8	0	1	100	No change
hnRNP A1 binding prediction							
116701318	-35	1.5	1.5	0	1	100	No change
116701321	-32	1.5	1.5	0	1	100	No change
116701332	-21	0.1	0.1	0	1	100	No change
116701362	9	2.4	2.4	0	1	100	No change
116701371	18	2.2	2.2	0	1	100	No change
116701376	23	4.7	4.7	0	1	100	No change
116701377	24	0.3	0.3	0	1	100	No change
116701383	30	4.9	4.9	0	1	100	No change
116701388	35	0.5	0.5	0	1	100	No change
116701398	45	4.5	4.5	0	1	100	No change
116701399	46	1.2	1.2	0	1	100	No change
116701352	-1	-4.5	4.7	9.2	26.1	2610.9	Increased
Genome-wide splice donor prediction							
116701311	-42	0.3	0.3	0	1	100	No change
116701354	1	2.5	3.8	1.3	2.5	249.1	Increased
116701365	12	2.5	2.5	0	1	100	No change
116701387	34	1.3	1.3	0	1	100	No change
Gene-specific splice-donor prediction							
116701311	-42	1.8	1.8	0	1	100	No change
116701354	1	5.2	6	0.7	1.7	165.8	Increased
116701365	12	4.2	4.2	0	1	100	No change
116701387	34	3	3	0	1	100	No change

Splice branch-point prediction							
116701396	43	0.4	0.4	0	1	100	No change
Genome-wide splice-acceptor prediction							
116701334	-19	0	0	0	1	100	No change
Gene-specific splice-acceptor prediction							
116701334	-19	3.3	3.3	0	1	100	No change
116701335	-18	0	0	0	1	100	No change
116701343	-10	1.2	1.2	0	1	100	No change
116701358	5	1.3	1.5	0.2	1.1	114	Increased
116701357	4	-0.5	0.1	0.6	1.1	107.8	Increased
IVS2+1 G>A (g.116701354 G>A; rs138326449)							
Genomic Coordinates	Position relative to natural site	Initial(Ri)	Final(Ri)	ΔRi	Fold change	% Binding (Final/Initial)	Predicted Change
hnRNP H binding prediction							
116701372	18	0.9	0.9	0	1	100	No change
116701378	24	4.4	4.4	0	1	100	No change
116701379	25	4.2	4.2	0	1	100	No change
116701386	32	4.8	4.8	0	1	100	No change
hnRNP A1 binding prediction							
116701318	-36	1.5	1.5	0	1	100	No change
116701321	-33	1.5	1.5	0	1	100	No change
116701332	-22	0.1	0.1	0	1	100	No change
116701362	8	2.4	2.4	0	1	100	No change
116701371	17	2.2	2.2	0	1	100	No change
116701376	22	4.7	4.7	0	1	100	No change
116701377	23	0.3	0.3	0	1	100	No change
116701383	29	4.9	4.9	0	1	100	No change
116701388	34	0.5	0.5	0	1	100	No change
116701398	44	4.5	4.5	0	1	100	No change
116701399	45	1.2	1.2	0	1	100	No change
Genome-wide splice donor prediction							
116701311	-43	0.3	0.3	0	1	100	No change
116701365	11	2.5	2.5	0	1	100	No change
116701387	33	1.3	1.3	0	1	100	No change
116701354	0	2.5	-16.1	-18.6	-5.8	NA	Decreased

Gene-specific splice-donor prediction							
116701311	-43	1.8	1.8	0	1	100	No change
116701365	11	4.2	4.2	0	1	100	No change
116701387	33	3	3	0	1	100	No change
116701354	0	5.2	-7.6	-12.8	-38	NA	Decreased
Splice branch-point prediction							
116701396	42	0.4	0.4	0	1	100	No change
Genome-wide splice-acceptor prediction							
116701334	-20	0	0	0	1	100	No change
Gene-specific splice-acceptor prediction							
116701334	-20	3.3	3.3	0	1	100	No change
116701335	-19	0	0	0	1	100	No change
116701343	-11	1.2	1.2	0	1	100	No change
116701358	4	1.3	1.6	0.3	1.2	123.3	Increased
A43T (g.116701560 G>A; rs147210663)							
Genomic Coordinates	Position relative to natural site	Initial(Ri)	Final(Ri)	ΔRi	Fold change	% Binding (Final/Initial)	Predicted Change
hnRNP H binding prediction							
116701531	-29	1	1	0	1	100	No change
116701580	20	1.2	1.2	0	1	100	No change
hnRNP A1 binding prediction							
116701530	-30	4.9	4.9	0	1	100	No change
116701539	-21	4.6	4.6	0	1	100	No change
116701563	3	4.6	4.6	0	1	100	No change
116701572	12	2.4	2.4	0	1	100	No change
116701584	24	5.2	5.2	0	1	100	No change
116701585	25	0.1	0.1	0	1	100	No change
116701593	33	3	3	0	1	100	No change
Genome-wide splice donor prediction							
116701596	36	2.2	2.2	0	1	100	No change
Gene-specific splice-donor prediction							
116701596	36	4.2	4.2	0	1	100	No change
Splice branch-point prediction							
116701521	-39	1.3	1.3	0	1	100	No change
116701555	-5	0.9	0.9	0	1	100	No change

116701564	4	0.4	0.9	0.5	1.4	140	Increased
116701575	15	1.1	1.1	0	1	100	No change
Genome-wide splice-acceptor prediction							
116701532	-28	6	6	0	1	100	No change
Gene-specific splice-acceptor prediction							
116701532	-28	9.7	9.7	0	1	100	No change
116701607	47	0.3	0.3	0	1	100	No change
IVS3+1 G>T (g.116701613G>T; rs140621530)							
Genomic Coordinates	Position relative to natural site	Initial(Ri)	Final(Ri)	Δ Ri	Fold change	% Binding (Final/Initial)	Predicted Change
hnRNP H binding prediction							
116701580	-33	1.2	1.2	0	1	100	No change
116701606	-7	3	-1.9	-4.9	-8.2	NA	Decreased
hnRNP A1 binding prediction							
116701563	-50	4.6	4.6	0	1	100	No change
116701572	-41	2.4	2.4	0	1	100	No change
116701584	-29	5.2	5.2	0	1	100	No change
116701585	-28	0.1	0.1	0	1	100	No change
116701593	-20	3	3	0	1	100	No change
116701610	-3	3.9	3.1	-0.9	-1.8	54.8	Decreased
Genome-wide splice donor prediction							
116701596	-17	2.2	2.2	0	1	100	No change
116701613	0	3	-15.6	-18.6	-8.1	NA	Decreased
Gene-specific splice-donor prediction							
116701596	-17	4.2	4.2	0	1	100	No change
116701613	0	4.5	-3.3	-7.8	-22.9	NA	Decreased
Splice branch-point prediction							
116701564	-49	0.4	0.4	0	1	100	No change
116701575	-38	1.1	1.1	0	1	100	No change
116701637	24	1	1	0	1	100	No change
Genome-wide splice-acceptor prediction							
116701650	37	4.3	4.3	0	1	100	No change
Gene-specific splice-acceptor prediction							
116701607	-6	0.3	0.3	0	1	100	No change
116701646	33	1.7	1.7	0	1	100	No change

116701650	37	9.4	9.4	0	1	100	No change
116701653	40	0.7	0.7	0	1	100	No change
116701638	25	-0.3	0.6	0.9	1.5	146.9	Increased

Table 4.2: Predicted impact of APOC3 LoF variants on APOC3 mRNA splicing.

Each of the four APOC3 LoF variants were assessed for impact on interaction with mRNA processing machinery and on the relative energetic changes to splice-donor and splice-acceptor site usage using the Automated Splice Site and Exon Definition Analyses information theory based prediction tool (<http://splice.uwo.ca/>). Genomic Coordinates indicate genomic position on chromosome 11 for which a particular prediction is made. Position relative to the natural site indicates the position on the (+) strand of of the variant assessed relative to the Genomic Coordinates. Initial and Final Ri indicate the information content of the base in the WT or variant form, respectively, and ΔRi indicates the difference between Final Ri and Initial Ri. Fold change indicates the difference in binding affinity for a site in the variant vs. WT form based on the ΔRi . % Binding indicates the percentage increase or decrease in binding propensity for a given binding factor. Predicted change indicates the overall predicted effect for the given measure based on the ΔRi and predicted fold change for a given position. *Contributions:* All data in this table analyzed by SAK and Paolo Zanoni using in silico tool.

Table 4.3

Table 4.3: Plasma Lipids in Murine Models of WT vs. A43T APOC3 Expression																		
Measure:	Total Cholesterol (mg/dL)			HDL-C (mg/dL)			NonHDL-C (mg/dL)			TG (mg/dL)			ApoC-III (mg/dL)			ApoB (mg/dL)		
Group:	Null	WT	A43T	Null	WT	A43T	Null	WT	A43T	Null	WT	A43T	Null	WT	A43T	Null	WT	A43T
C57BL/6 WT	95.8 (5.4)	97.0 (3.4)	94.3 (8.4)	76.2 (5.7)	73.8 (4.3)	73.5 (8.3)	19.7 (1.0)	23.2 (2.9)	20.9 (1.3)	46.0 (10.2)	104.8 (21.0)	58.2 (15.6)**	0.0 (0.0)	14.8 (3.0)	6.0 (1.4) ****			
Apoc3 KO	86.4 (12.3)	101.0 (9.7)	90.6 (4.4)	65.3 (11.6)	74.9 (7.2)	70.7 (6.4)	21.1 (2.0)	26.1 (3.5)	19.9 (2.1)**	33.0 (2.4)	76.4 (15.5)	43.6 (7.3)**	0.0 (0.0)	14.2 (1.2)	3.8 (3.2)***			
LahB WT	157.7 (11.4)	169.8 (8.6)	166.3 (14.8)	83.8 (7.0)	76.7 (6.1)	75.2 (6.1)	73.9 (5.3)	93.1 (5.6)	91.1 (9.0)	165.5 (10.1)	260.6 (33.3)	190.2 (22.2)**	0.0 (0.0)	13.7 (2.6)	6.4 (1.9)***			
LahB Het	291.3 (26.1)	268.5 (28.3)	281.2 (7.4)	108.7 (7.3)	113.4 (10.0)	114.0 (3.8)	182.6 (18.9)	155.1 (19.4)	167.2 (7.9)	268.3 (29.1)	335.8 (73.0)	257.4 (5.5)*	0.0 (0.0)	24.0 (13.4)	5.4 (4.5)*	171.4 (12.4)	147.1 (19.0)	157.9 (7.4)
C57BL/6 WT + CETP	69.0 (20.2)	58.7 (4.9)	58.7 (2.1)	43.7 (19.4)	21.9 (2.9)	32.9 (2.4) ****	25.3 (5.7)	36.7 (2.5)	25.8 (1.2)****	61.3 (10.2)	252.0 (18.7)	90.7 (14.6) ****	0.0 (0.0)	16.4 (1.8)	4.5 (1.0) ****			
Apoc3 KO + CETP	70.8 (15.5)	63.8 (5.9)	61.6 (3.0)	43.4 (14.3)	26.3 (4.3)	33.6 (2.9)*	27.5 (5.3)	37.5 (6.5)	27.8 (2.5)**	42.3 (9.2)	192.4 (35.8)	55.0 (14.2) ****	0.0 (0.0)	16.7 (2.2)	2.4 (1.6) ****			
LahB KO + CETP	376.0 (139.4)	575.6 (90.2)	457.6 (133.6)	119.7 (38.3)	139.4 (8.6)	140.1 (12.2)	256.3 (103.5)	436.2 (92.2)	317.5 (127.6)*	124.2 (60.2)	242.0 (74.0)	144.4 (64.2)*	0.0 (0.0)	16.5 (1.3)	3.9 (3.3) ****	121.0 (85.9)	210.2 (59.2)	146.8 (85.8)*

Table 4.3: Plasma lipids in murine models of WT vs. A43T APOC3 expression. Mean and S.D. values [shown as mean(S.D.)] are given for each group of mice expressing Null, WT APOC3, or A43T APOC3 AAVs by intraperitoneal administration at a dose of 3×10^{11} genome copies (GC)/mouse. Measurements show data from plasma drawn after four hours of fasting at four weeks after AAV administration. Each experiment utilized N=6 male mice per group at approximately 10-12 weeks of age. LahB WT mice are *Apobec1* KO; human *APOB* transgenic; *Ldlr* WT mice. LahB Het mice are *Apobec1* KO; human *APOB* transgenic; *Ldlr* heterozygous mice. LahB KO mice are *Apobec1* KO; human *APOB* transgenic; *Ldlr* KO mice. Groups with “+CETP” were co-transduced with 3×10^{10}

GC/mouse of human *CETP* AAV vector along with the indicated Null or *APOC3* AAV vectors. Bold sets of data indicate significant results ($P < 0.05$) by two-way ANOVA. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, comparison of WT and A43T groups by Tukey's post-test after two-way ANOVA. *Contributions: All data included in this table generated and analyzed by SAK.*

CHAPTER 5 – MULTIPLEXED TARGETED RESEQUENCING USING MOLECULAR INVERSION PROBES IDENTIFIES RARE NONCODING VARIANTS ASSOCIATED WITH HIGH HDL CHOLESTEROL LEVELS

Introduction

Coronary heart disease (CHD) is the leading cause of death in Western societies, and despite a substantial environmental contribution, risk of disease is highly heritable. Blood cholesterol levels are among the heritable biomarkers of disease risk and protection. High-density lipoprotein cholesterol (HDL-C) have been inversely associated with risk of CHD for several decades [338-343], supporting this measure as a biomarker for assessing disease risk and prevention. While epidemiological findings supporting an inverse association of HDL-C with CHD have been consistent, the direct role that HDL and HDL-C plays in modulating CHD risk has been highly controversial, with doubt concerning the validity of HDL-C as a causal intermediate of CHD risk protection increasing over the last decade [53,344]. This doubt has arisen in part from multiple failed pharmacological attempts to raise HDL-C specifically to lower CHD risk [65,66,345,346]. Human genetics discoveries of the last decade have also cast doubt on the causality of HDL-C in mediating CHD risk protection. Mendelian randomization studies of low-frequency coding variants in candidate loci associated with HDL-C elevation have not translated to reduction in the incidence of CHD or myocardial infarction (MI) [193,194,204,347,348]. Recently, we also identified a nonsynonymous coding variant in one HDL candidate gene, *SCARB1*, which was associated with a marked elevation in HDL-C levels but increased risk of CHD in humans [349]. The disparate conclusions about the directionality of HDL-C to CHD from these collective

data have motivated efforts to identify genes and pathways involving HDL function that causally impact CHD.

Unbiased genetics discoveries have provided the opportunity to identify novel loci and pathways involved in HDL metabolism for further investigation into their causality to CHD. Genome-wide association studies (GWAS) for plasma lipid traits have uncovered at least 70 loci to date significantly associated with HDL-C levels through testing of common variants (minor allele frequency or MAF > 0.05) on genome-wide genotyping arrays [117]. However, pinpointing the causal variants and genes in these loci is challenging. Current efforts to resolve this have included fine mapping of the identified loci to determine causal variants, but even if successful, these methods are limited in that they focus on common single nucleotide polymorphisms (SNPs) with generally small effect sizes. Given that common SNPs are estimated to explain only ~1.6% of the total heritability of HDL-C levels [117], the missing heritability is likely contained in low frequency and rare variants that have been largely underexplored.

One strategy to uncover the missing heritability and identify putatively causal variants at HDL-C candidate loci is targeted resequencing of genes at these loci. Indeed, this approach has been applied to the follow-up of initial association studies for myriad complex traits, including additional lipid traits LDL-C [350-352] and TG [109,350,353-355], obesity [356-361], Alzheimer's disease [85,362-365], and many others. Nearly all such efforts have focused on sequencing coding regions of candidate genes, with the goal of identifying protein-altering variants that may have a profound functional impact. However, given that the majority of GWAS-implicated variants are in the noncoding genome [366,367], and many lie in "gene deserts", the contribution of rare noncoding variants to these traits is underexplored and may be critical determinants of the missing

heritability underlying many complex traits including HDL-C and CHD. Furthermore, the identification of rare, causal, noncoding variants with strong effect sizes on HDL-C may help further to reveal causal genes at GWAS loci and delineate specific mechanisms of HDL-C regulation that could directly relate to CHD risk. One limitation hampering targeted sequencing efforts for the noncoding genome is the relatively poor annotation of functional elements most likely to harbor variants of significance, and difficulty in readily ascertaining the functional consequences and directionality of noncoding variants identified to the associated trait. A related complication is that targeted sequencing efforts are costly and scale with the size of the genomic targets, so methods have largely been developed for reliably amplifying and sequencing coding regions of genes. Thus, there is a need for efficiently capturing the noncoding genome for the identification of rare, noncoding variants in manner that could be applied to sufficient numbers of participants to identify relationships to complex traits such as HDL-C.

Here, we sought to identify the feasibility of targeting the noncoding genome for candidate gene loci underlying HDL-C that could be extended to larger numbers of samples. We performed targeted resequencing of seven HDL-C loci including both coding and noncoding regions in a cohort of 1,532 subjects with either extremely high vs. low HDL-C. We adapt a recently reported target capture method involving molecular inversion probes (MIPs) [368,369] for amplifying genomic targets and show for the first time the ability to capture noncoding regions of the genome by this method. Our results validate previously reported coding and noncoding SNP associations with HDL-C, identify gene-level associations in these seven regions with this trait, and also show the promise of large-scale targeted resequencing of noncoding regions for complex traits.

Materials and Methods

Ethics statement

All human participants of this study and all analyses performed were completed following the Declaration of Helsinki and were approved by the Institutional Review Board of the Perelman School of Medicine at the University of Pennsylvania and all participants provided informed consent.

Subject Selection and Ascertainment

1532 participants mostly of European ancestry, with either extremely high HDL-C (>95th percentile for age and sex), or low HDL-C (20 mg/dL or higher to 25th percentile for age and sex) were recruited for targeted sequencing. Participants were recruited as part of the University of Pennsylvania High HDL Study (HHDL), a cross-sectional study of genetic factors contributing to elevated HDL-C levels. Individuals with elevated HDL-C (>90th percentile for age and gender) were identified by physician referrals or through the Hospital of the University of Pennsylvania clinical laboratory. Plasma lipids for all subjects were measured after fasting by a clinical autoanalyzer (Hitachi). HDL-C percentiles for inclusion were calculated for individuals of European ancestry from the Framingham Heart Study Offspring cohort adjusted for age and sex.

MIP Design

Molecular inversion probes (MIPs) were designed according to the method and pipeline previously described by O’Roak et al [368]. Briefly, MIPs capturing chosen targets were all designed using a common 30 bp linker sequence flanked by an extension arm of 16-20 bp and a ligation sequence of 20-24 bp, with a total MIP length of 70 bp. The unique

arms of the MIPs that anneal to the target sequence by complementary base pairing were designed to amplify a specific 112 bp target region by gap-filling and circularization. After MIP capture, a PCR amplification reaction using Nextera-like (Illumina) sequencing adaptor-containing primers (Illumina) allowed amplification with the primers annealing to the 30 bp common linker sequence (**Figure 5.1**). Given prior demonstration of significant variability in MIP capture efficiency due to properties of annealing arm base pairing with sequences adjacent to individual targets, an initial set of 549 MIPs were designed to cover all of the proposed target, and a pilot-phase MIP sequencing study was performed to evaluate per sample and per MIP coverage depth in an initial set of 95 DNA samples. Based on the coverage from this run, MIPs demonstrating less than 10-fold coverage per base for more than 50% of sequenced samples were redesigned and redesigned MIPs were used to replace poorly performing MIPs in a repeat pilot-phase sequencing run of the same samples. From this second and final pilot-phase sequencing run, 569 MIPs were included in total to capture the targeted regions. These 569 MIPs were used for subsequent sequencing of genomic DNA samples from the 1,532 participants.

MIPs were designed to capture the coding sequences (exons) of the following genes (GRC37/hg19): *GALNT2* (Chr1:230338882-230415202), *APOA5* (Chr11:116660886-116663095), *APOC3* (Chr11:116700650-116703573), *CCDC92* (Chr12:124421729-124428847), *ZNF664* (Chr12:124488089-124497396), *SCARB1* (Chr12:125267297-125348261), *CETP* (Chr16:56995891-57017572), and *LIPG* (Chr18:47088681-47110124) (**Supplementary File 1: MIP Sequences and Coordinates**). The total coding sequence designed to be captured by the 569 MIPs corresponding to these regions was 29,288 bp. In addition, noncoding sequences proximate to these loci were chosen if they were previously shown to harbor variants

significantly associated with HDL-C ($P < 5E-08$; GLGC+MetaboChip GWAS), and also were found to overlap DNase I hypersensitivity sites in HepG2 cells (human hepatocellular carcinoma) from the ENCODE project [370] or enhancers in HepG2 cells from the Epigenome-Roadmap project [371]. Regions with 250 bp flanking the positions harboring these elements were selected for MIP design. The total noncoding target across the loci for which MIPs were designed to capture was 16,696 bp. MIP oligonucleotides were purchased from Eurofins Genomics with high-purity salt-free purification and at a scale of 50 nmol per oligonucleotide. Lyophilized MIPs were hydrated with 1x TE buffer to a concentration of 100 μ M and stored at -20 °C.

MIP Capture and Amplification of Targeted Sequences

MIP oligonucleotides were used to capture targets from genomic DNA derived from whole blood from the participants in a manner described previously. Hydrated MIP oligonucleotides were pooled together and phosphorylated with T4 polynucleotide kinase (NEB) at 37 °C for 45 min, followed by heat inactivation at 65 °C for 20 min. Phosphorylated MIPs were used to capture genomic targets by combining with genomic DNA from each participant (100 ng of each individual sample; ratio of 800:1 of each MIP copy to haploid genome copy) using NEB Hemo KlenTaq (NEB) and Ampligase for 24 hrs at 60 °C in a thermocycler. 96 samples were individually captured in one reaction by individually plating reactions in a 96-well thermocycler plate. A total of 16 plates of 96 samples apiece were processed. Reactions were digested with Exonuclease I and Exonuclease III (NEB) after incubations for 45 min at 37 °C and then 2 min at 95 °C. Digested MIP capture reactions were PCR amplified using primers with barcoded adapter sequences (**Table 5.1**). In order to sequence all 1532 samples from a single

multiplexed pool, a dual-barcoding strategy similar to that of Illumina's Nextera protocol was employed. To provide unique combinations of forward and reverse primers for all 1536 samples (1532 individual subjects plus 4 ddH₂O controls) across the 16 plates, a common forward barcoded adapter primer was used for each plate, and 96 unique reverse barcoded adapter primers were used for each of the 96 samples within a plate. PCR reactions to ligate adapters and barcode MIP capture reactions were completed with iProof master mix reagent (BioRad). Barcoded and PCR amplified MIP capture reactions were then pooled together at equal volumes, purified using AMPure magnetic bead purification (Agencourt) at 0.9-fold the total volume of the pooled reaction, and visualized on agarose gels. Purified, pooled capture reactions were then sequenced in paired-end mode (150 bp X 150 bp) on Illumina MiSeq and HiSeq2500 sequencers using standard Nextera sequencing reagents plus a custom pool of Nextera-like sequencing primers (**Table 5.2**). All MIP oligonucleotides, adapters, PCR primers and sequencing primers were synthesized by Eurofins MWG Operon.

Sequencing and Data Analysis

Initial MIP sequences were obtained as paired-end FASTQ reads from three separate sequencing runs (one lane of sequences from a single MiSeq run and two lanes of HiSeq2500 RapidRun from two independent runs). De-multiplexing was performed using CASAVA v1.8.2's bcl2fastq conversion script (Illumina), and all reads were processed using Trimmomatic v0.32 to remove adapter artifacts, sequencing artifacts, and low quality bases [372]. Sequences were aligned to the UCSC hg19 human genome build on a per-sample and per-sequencing run basis using BWA v0.7.8 (MEM algorithm; [373]), and the resulting alignment files were compressed and sorted using SAMtools v0.1.19

[374]. The variant calling was conducted utilizing Genome Analysis Toolkit v3.5 (GATK; [283]), and pre-processing of each sample's lane-specific alignment files was performed in accordance with the GATK's 'Best Practices' workflow [375,376]. This workflow featured duplicate read removal using Picard v1.141 (Picard website: <http://broadinstitute.github.io/picard>), and run-specific insertion-deletion (indel) realignment and base recalibration using GATK and hg19 "gold-standard" variant catalogs (dbSNP v138 database: <http://www.ncbi.nlm.nih.gov/SNP/>, [377,378]). Run-specific alignments were then merged for each sample, and subjected to a second round of indel realignment and base recalibration with GATK. Preliminary sample-level variants were called using GATK's HaplotypeCaller tool in gVCF mode at base-pair, and all known variants were annotated with dbSNP v138 identities. Sample-level variant callsets were then combined and joint-genotyped with GATK. SNPs and indels called at this stage were evaluated using metrics collected by Picard and GATK, and then hard-filtered on the basis of variant-class-specific criteria (**Table 5.3**) in order to flag potential false positives. Filtered variant calls were then further annotated with RefSeq gene coordinates for human genome build hg19 (RefSeq database: <http://www.ncbi.nlm.nih.gov/books/NBK21091/>) using VCFtools v0.1.9.0 [379] to include features such as full gene lengths, protein coding sequences, exon and intron boundaries, and five- and three-prime untranslated regions (UTRs). The final annotated and filtered SNP and indel call sets were then combined, and the union of these variants was input into the association testing pipeline. Variant statistics, including MAF, were computed across the sequencing cohort. Some extremely rare variants were removed from the analysis due to lack of validation during filtering.

Association Testing and Statistics

Single variant association results for targeted sequencing analysis were computed using the Score test statistic with HDL-C levels treated as a dichotomous phenotype of high (> 95th percentile for age and gender) or low (<25th percentile for age and gender). Ashkenazi ethnicity ascertained by self-identification was included as a covariate in the regression analysis. Single variant associations were considered statistically significant if P-values for associations were below the Bonferroni corrected P-value, and associations with $P < 0.01$ were considered nominally significant. Gene-burden tests were computed after grouping variants by gene region and performing a logistic Wald test after grouping samples by HDL-C levels treated as a dichotomous phenotype. Additionally, a separate grouping of noncoding variants across the targeted regions was compared in high vs. low HDL-C participants using HDL-C as a dichotomous variable using the SKAT-O test. For all gene- and region-wide burden testing, $P < 0.05$ was considered significant.

Validation by Exome Array Genotyping

Genomic DNA from 1,139 of the 1,532 participants were also subject to genotyping using the Exome Array (HumanExome BeadChip v1.0, Illumina, Inc., San Diego, CA). The Exome Chip contains >240,000 coding SNPs derived from all mutations found >2 times across >1 dataset among 23 separate datasets comprising a total of >12,000 individual exome and whole genome sequences. In total, 704 high HDL-C participants and 458 low HDL-C participants were genotyped using the Exome Array.

Tools used

VCFtools v0.1.9.0 [379].

bgzip; tabix v0.2.6; bcftools v1.3.1; samtools v0.1.19. [284].

Ensembl Variant Effect Predictor [380].

Ensembl API and SNP Effect Predictor [381].

Ensembl 2016 [382].

Ensembl REST API [383].

LOFTEE (LOFTEE website: <https://github.com/konradjk/loftee>)

dbNSFP v2.9.1. [374].

EPACTS v3.2.6 (EPACTS website: <http://csg.sph.umich.edu/kang/epacts/home>)

R v3.2.5 (R Core Team 2015).

Results

Candidate regions for targeted sequencing

We sought to test the hypothesis that noncoding variants contribute to the allelic spectrum of extremely high HDL-C in a manner similar to coding variants, and that multiplexed targeted sequencing, previously applied only in sequencing coding regions of candidate loci, could also identify noncoding variants making this contribution. To that end, we performed a targeted resequencing study of HDL-C candidate gene regions in 1,532 participants with either extremely high HDL-C (plasma HDL-C >95th percentile for age and sex, 797 participants) vs. low HDL-C controls (plasma HDL-C between 20 mg/dL and 25th percentile for age and sex, 735 participants; **Table 5.4**).

For this effort, we chose seven candidate loci for targeted sequencing of coding and noncoding regions in our cohorts (**Figure 5.2**). Among these regions, four loci, *APOC3* (Chr11:116700624-116703787), *SCARB1* (Chr12:125262174-125348519), *CETP* (Chr16:56995835-57017756) and *LIPG* (Chr18:47088401-47119279) have known

roles in HDL metabolism for which loss-of-function has been shown to elevate HDL-C in humans [384]. In addition to these loci, we also explored the hypothesis that rare noncoding variants may underlie GWAS-implicated loci for HDL-C levels and thus included two novel HDL-C GWAS loci, *GALNT2* (chr1:230202942-230417876) and the *CCDC92-ZNF664* region (Chr12:124,420,955-124499986), in our targeted sequencing. The directional relationship of *GALNT2* SNPs to HDL-C levels has not been conclusively established, but previous discoveries of coding variants in *GALNT2* have suggested that loss-of-function may segregate with high HDL-C levels [184,385,386]. Similarly, the contribution to HDL-C levels of rare variants at the *CCDC92-ZNF664* locus, including coding variants in both genes, remains completely unexplored. We evaluated these loci from GWAS for rare coding and noncoding variants to better determine the directional relationship of these genes with HDL-C beyond the initial common variant associations.

Variants identified by MIP sequencing

We performed multiplexed capture of the genomic targets using 569 MIPs in 1,532 participants and sequenced all samples together after dual-index barcoding through Illumina HiSeq2500 sequencing runs. Multiplexed sequencing of our designated genomic targets, which included 29,288 bp of coding sequence and 16,686 bp of noncoding sequence, across the 1,532 samples resulted in a median sequencing coverage of 110-fold per base for an individual HiSeq2500 sequencing platform. We observed a high uniformity of target coverage per MIP across the subjects in our cohort, with approximately 489 MIPs (86% of total) demonstrating coverage of >10-fold depth in each sequenced participant.

To validate the variants identified from our MIP sequencing, we genotyped 1,162 of the 1,532 participants (704 high HDL-C individuals and 458 low HDL-C individuals) on the probe-based Illumina Exome Array [387]. Among the variants genotyped on this array, 51 were within our target regions. We observed a high concordance rate in variant discovery between MIP sequencing and genotyping results, with 48 of 51 SNPs (94.1%) on the Exome Array called with 100% concordance across all participants, and 1088 of 1162 participants (93.6%) demonstrating 100% concordance of all genotyped SNPs (**Figure 5.3**). Including these variants present on the Exome array, we identified 2,657 individual SNPs and 2,342 insertions or deletions (indels) from targeted sequencing. Among these variants, 4,371 were novel discoveries without and previous annotation (**Table 5.5**). We tested the association of the individual SNPs and indels identified with high vs. low HDL levels, and observed 28 variants with significantly increased frequency in the high HDL-C participants ($P < 2.5E-05$, Score test, **Table 5.5**).

Association of single variants from targeted sequencing with extremely high HDL-C

To assess the ability of our MIP sequencing approach to identify noncoding variants that may influence HDL-C levels, we assessed variants associated with elevated HDL-C with an association of $P < 0.01$ using the Score test statistic (**Table 5.5 & Table 5.6**). We identified 276 such SNPs and indels, which included 75 noncoding variants (variant positions occurring outside of exons), 32 rare or low frequency variants ($MAF < 0.05$), and 17 novel variants. In total, we identified 15 rare or low-frequency, novel, noncoding variants across our regions targeted that were nominally associated with high HDL-C from MIP sequencing.

We sought to further explore the biological mechanisms underlying the novel associations of noncoding variants associated with HDL-C from MIP sequencing. We chose a subset of variants significantly associated with HDL-C from our MIP sequencing that were either novel variants or previously identified but for which HDL-C associations and functional impact remained unknown to further assess their functional consequences. In selecting noncoding variants, we focused on those present in the proximal promoter or 5' untranslated regions (UTRs) of candidate genes, intronic variants, and 3' UTR variants.

Replication of HDL-C associations from GWAS through MIP sequencing

In addition to rare, noncoding variants identified from MIP sequencing, our targeting of noncoding regions of candidate gene loci allowed us to also identify common noncoding variants previously associated with HDL-C through GWAS. We identified 28 SNPs across our targeted regions that had been previously associated with HDL-C measured as a continuous trait in GWAS for plasma lipids [117]. We validated all 30 SNPs as significantly associated with elevated HDL-C from our targeted sequencing. All 30 SNPs displayed a consistent MAF and directional relationship of the minor allele to HDL-C levels with the previous associations (**Table 5.7**).

Gene-burden associations across target regions with extremely high HDL-C

Finally, we tested the hypothesis that the genomic regions we targeted harbor variants that in aggregate contribute to the relationship of these genes with HDL-C levels. First, we performed variant aggregation over the coding regions of the individual gene targets and tested the association of each region's variant burden with HDL-C after filtering

coding variants based on their predicted impact on protein function. We grouped coding variants into the following categories: disruptive (nonsense, splice-donor, splice-acceptor, and frameshift variants), and missense variants. We further classified missense variants by their *in silico* predicted deleteriousness and grouped variants as either hypothesized to confer deleteriousness by at least one of five variant scoring tools vs. those predicted to be deleterious by all five predictions (see Materials and Methods). When we performed gene-burden association analysis of high vs. low HDL-C participants, we found that the *CETP* gene region demonstrated a significant burden of variants associated with HDL-C (Collapsing test; **Table 5.8**). This association was observed with inclusion of disruptive variants only as well as inclusion of missense variants predicted by one tool or all five tools to be deleterious to protein function. In addition, we observed that *GALNT2*, *APOA5*, and *CCDC92* coding variants that were either deleterious or missense variants predicted by at least one bioinformatics tool to confer deleteriousness also demonstrated a significant gene-level variant burden with HDL-C (**Table 5.8**). Additionally, we asked if noncoding variants across these targets collectively contributed to HDL-C and thus performed an aggregation of all noncoding variants across our eight targeted regions and tested the association of the burden of these variants in high vs. low HDL-C participants. We found a significant association of noncoding variants with HDL-C ($P=0.05$, SKAT-O test; **Table 5.8**).

Discussion

Translating novel loci from complex trait GWAS efforts of the last decade to bona fide causal variants and genes has been a major challenge for human genetics. Part of this difficulty lies in the small effect sizes of GWAS SNPs, and thus resequencing of

candidate genes at GWAS loci at the phenotypic extremes of complex traits has been one predominant approach to identify rare variants with larger effects. To date, this approach has been applied to coding regions of GWAS candidate genes. However, coding variants account for a small fraction (< 20%) of all variants in genetic association studies for complex traits [366], underscoring the need for further study of the noncoding genome for rare, putatively causal variants. Here, we present an inexpensive and scalable targeted sequencing approach for identifying rare noncoding variants in candidate genes influencing HDL-C, a complex trait with 72 associated loci from GWAS [117]. Our proof-of-principle resequencing study of seven candidate gene regions in 797 extremely high HDL-C vs. 735 low HDL-C participants revealed more than 2,700 variants in noncoding regions of targets, including rare, novel noncoding variants associated with HDL-C and validation of prior GWAS-implicated tag SNPs. As such, our findings provide the first example of a multiplexed targeted resequencing study of noncoding variants at the phenotypic extremes of a complex trait.

Our targeted resequencing study of subjects with HDL-C extremes successfully identified multiple rare, novel noncoding variants in target regions that were significantly enriched in either cases or controls. While the functional contributions of all of the newly identified variants remain unknown, several of these variants overlap important regulatory elements that promote gene expression or transcript stability, including gene promoters, 5' UTRs and 3' UTRs, and other epigenetic signatures of gene transcription and transcription factor binding. The identification here of 13 rare, novel noncoding variants associated with HDL-C contrasts with a recent targeted resequencing study of the coding regions of genes at 95 loci previously implicated in GWAS for plasma lipids in a smaller cohort of extremely high vs. low HDL-C [350]. This study involving the

sequencing of exons in 939 candidate genes in 684 high or low HDL-C participants with replication in 6,424 individuals, failed to identify any novel coding nonsynonymous variants associated with HDL-C. The authors argue that their results highlight the importance of noncoding variants, either rare or common, at these GWAS-implicated loci as causal contributors to HDL-C and the other lipid traits measured, a notion supported by our findings here.

Several variants identified from our study lie in gene regions of known candidate genes for which loss-of-function variants have been previously shown to raise HDL-C levels in humans. In particular, several noncoding variants significantly enriched in high HDL-C participants were observed in *CETP* and *LIPG*. *CETP* is a critical mediator of exchange of triglycerides from proatherogenic low-density and very low-density lipoproteins to HDL in exchange for HDL cholesterol [19,387]. Higher *CETP* plasma activity is associated with reduced HDL-C levels [19,388,389], while pharmacological inhibition of *CETP* raises HDL-C in humans [388,390,391]. Importantly, several common coding and noncoding variants in *CETP* that putatively reduce *CETP* levels or expression have been identified in individuals with high HDL-C levels [196,385,392-398]. Our discovery of novel *CETP* noncoding variants individually associated with high HDL-C levels, as well as the finding of a significant gene-level burden of *CETP* variants with HDL-C levels, further supports the notion that *CETP* is a critical genetic mediator of HDL-C in humans. In addition to *CETP*, *LIPG* was among the targets demonstrating several rare noncoding variants significantly increased in frequency in high HDL-C subjects through MIP sequencing. Our group and others identified *LIPG* as a gene encoding endothelial lipase (EL) and showed that it physiologically mediated catabolism of circulating HDLs through its lipolytic enzymatic activity [399-401]. We also identified

both rare coding and noncoding loss-of-function variants in *LIPG* that were significantly associated with high HDL-C levels in humans [402-404]. Here, through an expanded targeted sequencing effort of the *LIPG* noncoding region, we identified an additional novel noncoding variant in the *LIPG* promoter significantly associated with elevated HDL-C, thus furthering the notion that *LIPG* regulatory variants also contribute to HDL-C likely through disruption of *LIPG* gene expression and plasma EL levels.

In addition to the novel variants associated with HDL-C through MIP sequencing, we also verified several previously implicated variants in our cohort. We identified six rare (MAF < 0.01) or low frequency (MAF < 0.05) nonsynonymous coding variants across our selected targeted that were significantly associated with HDL-C levels with directionalities consistent with previous reports. These included the *CETP* Ala390Pro and Arg468Gln variants, the *APOC3* Ala43Thr variant, and the *LIPG* Asn396Ser variants. We further demonstrated that 28 SNPs associated with HDL-C from GWAS were significantly associated with HDL-C cohort status in our MIP high vs. low HDL-C sequencing study, with consistent directionality and overall MAF to the previously reported results. Though the latest GWAS for HDL-C from the Global Lipids Genetics Consortium involved 188,577 participants [117], we were able to replicate these associations in less than 1/100th a cohort size through our case-control design. Collectively, these findings support the utility of candidate gene and noncoding locus resequencing at the extremes of a continuous trait distribution to enrich for trait-associated alleles, which may allow ascertainment of genetic associations in smaller populations than historical sizes for complex trait GWAS, such as understudied ethnicities and population isolates.

Our study also has methodological implications for future targeted resequencing efforts. To date, MIP-sequencing has been applied to targeted sequencing of coding regions of candidate genes with a sample preparation cost of less than \$1 per participant [368,405,406]. Our application of MIPs to noncoding regions of HDL-C candidate genes represents one of the first applications of this methodology for regulatory DNA regions. Our sequencing efforts were completed at a comparable cost to the prior applications, with similar target coverage depths for both the coding and noncoding targets. Additionally, our approach allowed us to multiplex all 1,532 individually barcoded capture libraries for sequencing in a single lane of an Illumina HiSeq2500 sequencing run with a median base coverage per participant of 110-fold, a robust depth for novel and rare variant identification at a cost of approximately \$1,800 for sequencing. Thus, our study supports the putative application of MIP-based sequencing of noncoding regions for much larger targeted resequencing efforts at a relatively low per-sample cost.

Our study has limitations, which highlight areas for further investigation. Our population of high and low HDL-C participants was largely of European ancestry, thus limiting the ability to extrapolate the variants discovered to other populations. Additionally, our total cohort size of 1,532 participants may be too small to sufficiently identify the spectrum of very rare variants that may underlie extremely high HDL-C levels. An additional limitation of our work is the selection of the noncoding regions for targeting by MIPs and yet untested MIP capture efficiency across any region of the genome.

In conclusion, our MIP-based targeted sequencing approach has pioneered the capture of noncoding regions for the discovery of rare, noncoding variants associated with HDL-C in a cohort of extremely high vs. low HDL-C participants. Our results support

the promise of future, larger candidate loci resequencing efforts for the discovery of putatively causal noncoding variants. These efforts, coupled with appropriate functional investigation of identified variants for impact on gene regulation, may substantially help in refining the causal genes at loci implicated from GWAS studies and also help further explain the missing heritability underlying complex traits such as HDL-C. Though efforts to better identify the spectrum of noncoding variants underlying complex traits have initiated, including denser genotyping of noncoding variants [407] and whole-genome sequencing [408], these approaches remain expensive and not readily applicable to the study of large populations or large case-control designs.

Figure 5.1

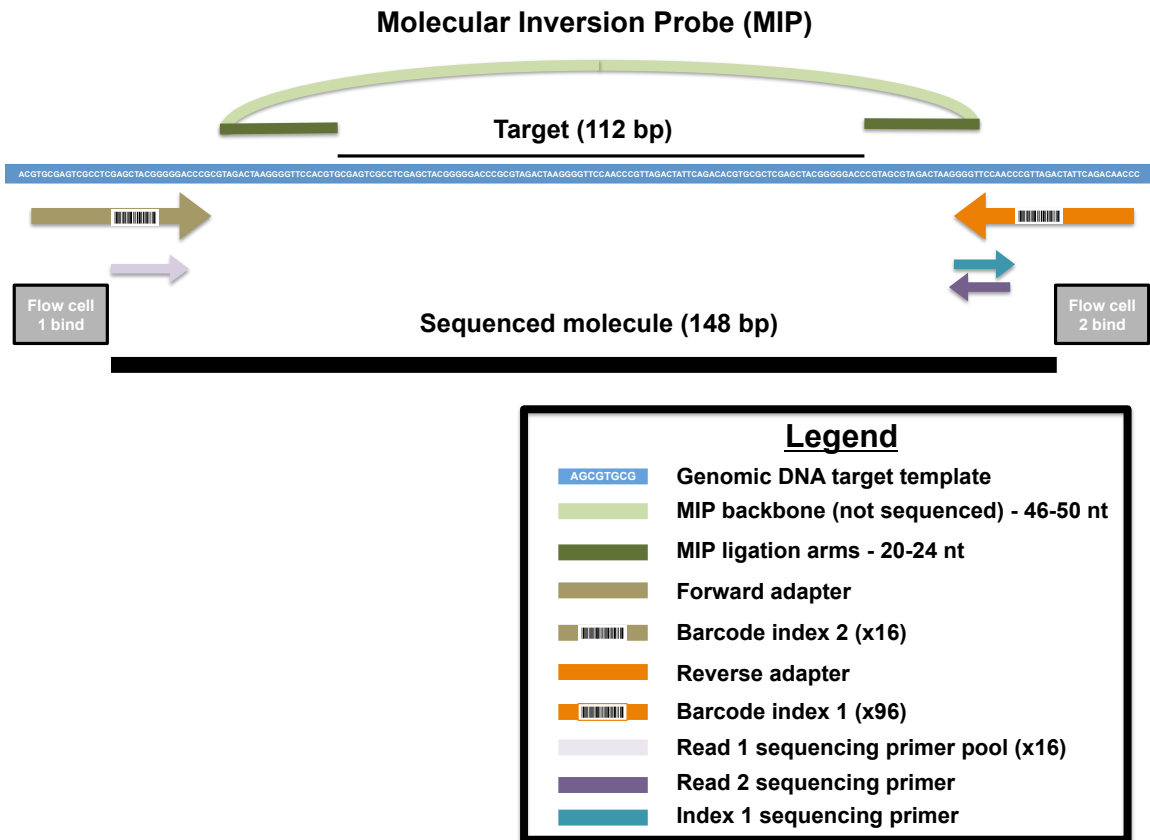


Figure 5.1: Diagram of MIP target capture and sequencing reaction. MIPs were designed to capture target sequences of 112 bp from genomic DNA of participants isolated from whole blood, as described in the Materials and Methods. Adapter ligation involving a dual-index barcoding strategy allowed the efficient barcoding of all 1,532 MIP capture libraries and simultaneous multiplexing of all samples together after PCR amplification for the Illumina sequencing reactions, which were performed on HiSeq2500 instruments with 150x150 bp paired end sequencing reads. *Contributions: Figure generated by SAK and Paul Babb.*

Figure 5.2

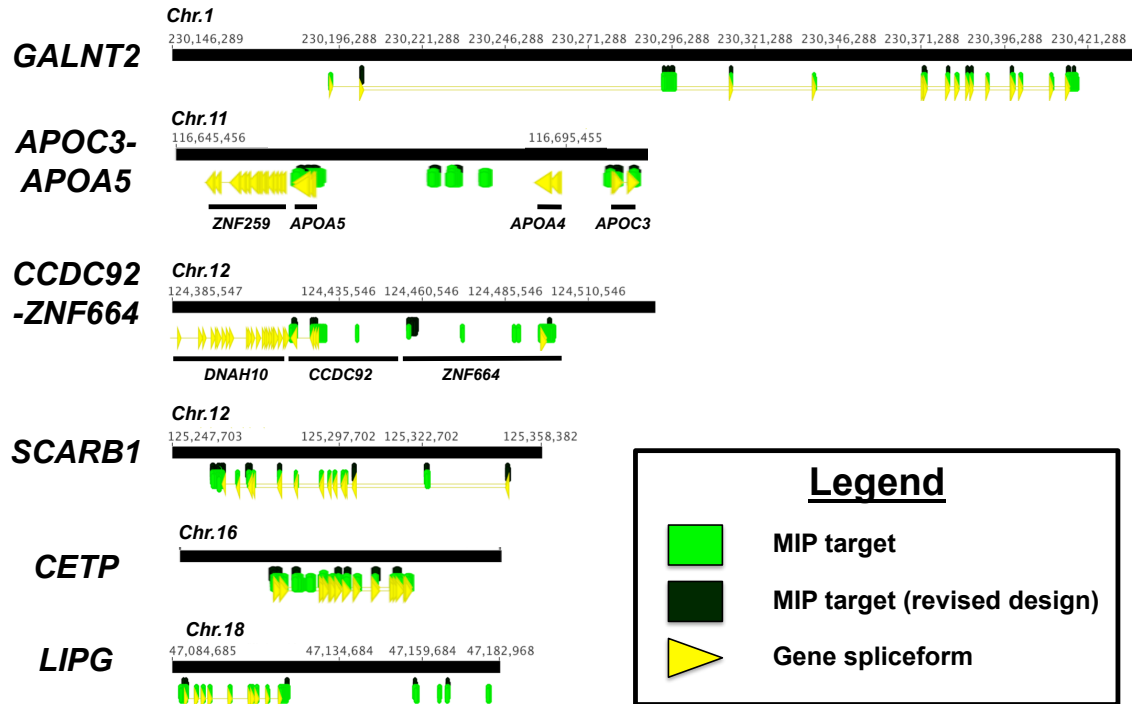


Figure 5.2: Candidate gene regions for MIP targeted sequencing. All coordinates correspond to genomic build GRC37/hg19. *Contributions: Figure generated by SAK and Paul Babb.*

Figure 5.3

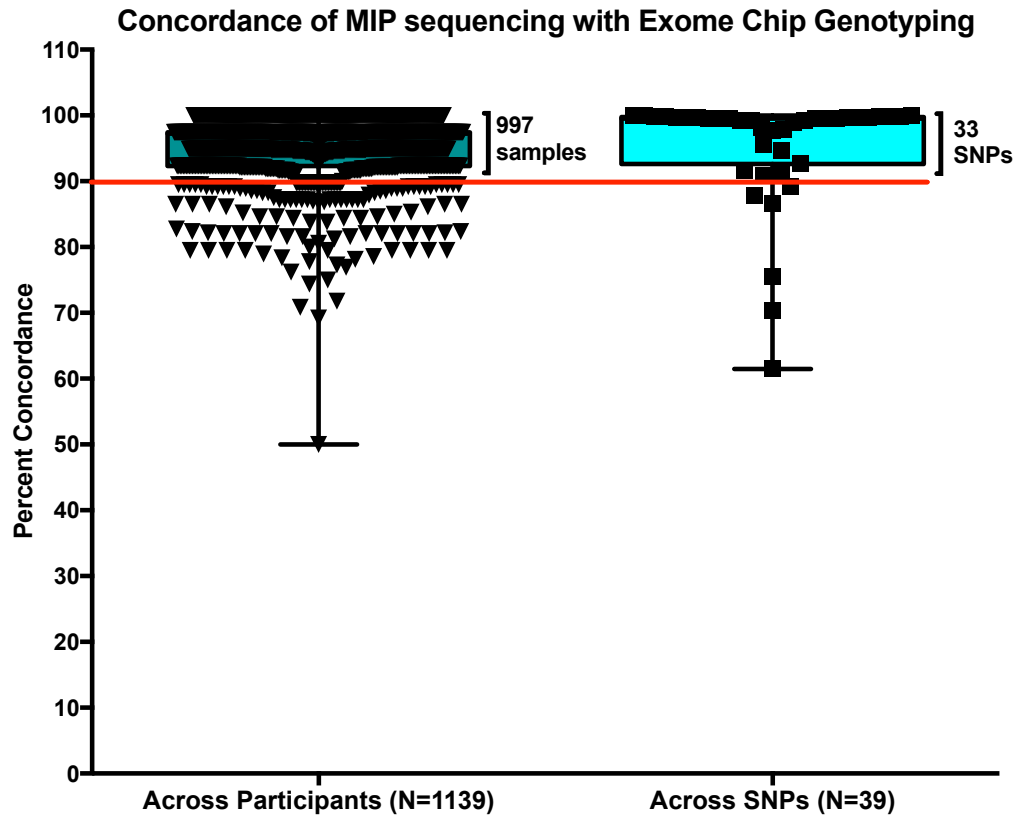


Figure 5.3: Concordance of variants identified from MIP sequencing with Exome Chip genotyping. Single nucleotide variants identified in the targeted regions by MIP-sequencing were compared to the discovery of those variants by genotyping on the Exome Chip in a subset of 1,139 participants who were included in both variant discovery efforts. 39 SNPs that were included in the Exome Chip were found to overlap the targeted regions by MIPs. Box plot on the left shows the percentage of total SNPs that were found by both discovery methods for each individual. Plot on right shows the percentage of individuals for which a given SNP was found to be concordant across the two discovery methods. *Contributions: Data generated by SAK and analyzed by SAK and Paul Babb; genotyping data generated by core facility.*

Table 5.1

Table 5.1: Indexed Adapter Oligo Sequences		
Forward Indexed Primers		
Name	Sequence	Pool According to DNA Plate
SLXA_PE_MIPBC_F OR	AATGATACGGCGACCACCGAGATCTACACATACGAGATCC GTAATCGGGAAGCTGAAG	1
FOR_001	AATGATACGGCGACCACCGAGATCTACACAAAGATACTCCG TAATCGGGAAGCTGAAG	2
FOR_002	AATGATACGGCGACCACCGAGATCTACACAGTCTCTCCG TAATCGGGAAGCTGAAG	3
FOR_003	AATGATACGGCGACCACCGAGATCTACACCAAGGTGATCC GTAATCGGGAAGCTGAAG	4
FOR_004	AATGATACGGCGACCACCGAGATCTACACCAGTCTGGTCC GTAATCGGGAAGCTGAAG	5
FOR_005	AATGATACGGCGACCACCGAGATCTACACGACACTACTCC GTAATCGGGAAGCTGAAG	6
FOR_006	AATGATACGGCGACCACCGAGATCTACACGTGCTTATTCCG TAATCGGGAAGCTGAAG	7
FOR_007	AATGATACGGCGACCACCGAGATCTACACTGGAGCTGTCC GTAATCGGGAAGCTGAAG	8
FOR_008	AATGATACGGCGACCACCGAGATCTACACTGGCACCTTCC GTAATCGGGAAGCTGAAG	9
FOR_009	AATGATACGGCGACCACCGAGATCTACACTACGGTGTCC GTAATCGGGAAGCTGAAG	10
FOR_010	AATGATACGGCGACCACCGAGATCTACACTTTCATAGTCCG TAATCGGGAAGCTGAAG	11
FOR_011	AATGATACGGCGACCACCGAGATCTACACGTTTCAGATCCG TAATCGGGAAGCTGAAG	12
FOR_012	AATGATACGGCGACCACCGAGATCTACACTCAAAAAGTCCG TAATCGGGAAGCTGAAG	13
FOR_013	AATGATACGGCGACCACCGAGATCTACACTATGCTGCTCCG TAATCGGGAAGCTGAAG	14
FOR_014	AATGATACGGCGACCACCGAGATCTACACTAGCTGCCTCC GTAATCGGGAAGCTGAAG	15
FOR_015	AATGATACGGCGACCACCGAGATCTACACAGCGAATGTCC GTAATCGGGAAGCTGAAG	16
Reverse Indexed Primers		
Name	Sequence	Position according to schematic for each plate
SLXA_PE_MIPBC2_ REV_001	CAAGCAGAAGACGGCATAACGAGATGTTAAGACACACGCAC GATCCGACGGTAGTGT	Position 1
SLXA_PE_MIPBC2_ REV_002	CAAGCAGAAGACGGCATAACGAGATTCTAAGTTACACGCAC GATCCGACGGTAGTGT	Position 2
SLXA_PE_MIPBC2_ REV_003	CAAGCAGAAGACGGCATAACGAGATTGTACATTACACGCAC GATCCGACGGTAGTGT	Position 3
SLXA_PE_MIPBC2_ REV_004	CAAGCAGAAGACGGCATAACGAGATAATACCGCACACGCAC GATCCGACGGTAGTGT	Position 4
SLXA_PE_MIPBC2_ REV_005	CAAGCAGAAGACGGCATAACGAGATTGTCTATGACACGCAC GATCCGACGGTAGTGT	Position 5
SLXA_PE_MIPBC2_ REV_006	CAAGCAGAAGACGGCATAACGAGATTCACTGTTACACGCAC GATCCGACGGTAGTGT	Position 6
SLXA_PE_MIPBC2_ REV_007	CAAGCAGAAGACGGCATAACGAGATTTGTATTACACGCAC GATCCGACGGTAGTGT	Position 7
SLXA_PE_MIPBC2_ REV_008	CAAGCAGAAGACGGCATAACGAGATAGACTAACACACGCAC GATCCGACGGTAGTGT	Position 8
SLXA_PE_MIPBC2_ REV_009	CAAGCAGAAGACGGCATAACGAGATATTGTCAAACACGCAC GATCCGACGGTAGTGT	Position 9
SLXA_PE_MIPBC2_ REV_010	CAAGCAGAAGACGGCATAACGAGATGTTCTACAACACGCAC GATCCGACGGTAGTGT	Position 10
SLXA_PE_MIPBC2_ REV_011	CAAGCAGAAGACGGCATAACGAGATTTTCATTTCGACACGCAC GATCCGACGGTAGTGT	Position 11

SLXA_PE_MIPBC2_ REV_012	CAAGCAGAAGACGGCATAACGAGATAAGCGTCAACACGCAC GATCCGACGGTAGTGT	Position 12
SLXA_PE_MIPBC2_ REV_013	CAAGCAGAAGACGGCATAACGAGATCGTAGGATACACGCAC GATCCGACGGTAGTGT	Position 13
SLXA_PE_MIPBC2_ REV_014	CAAGCAGAAGACGGCATAACGAGATAGTTAGCGACACGCAC GATCCGACGGTAGTGT	Position 14
SLXA_PE_MIPBC2_ REV_015	CAAGCAGAAGACGGCATAACGAGATGAACGTCTACACGCAC GATCCGACGGTAGTGT	Position 15
SLXA_PE_MIPBC2_ REV_016	CAAGCAGAAGACGGCATAACGAGATAGAACACTACACGCAC GATCCGACGGTAGTGT	Position 16
SLXA_PE_MIPBC2_ REV_017	CAAGCAGAAGACGGCATAACGAGATTGTTGGCAACACGCAC GATCCGACGGTAGTGT	Position 17
SLXA_PE_MIPBC2_ REV_018	CAAGCAGAAGACGGCATAACGAGATGGCTAATAACACGCAC GATCCGACGGTAGTGT	Position 18
SLXA_PE_MIPBC2_ REV_019	CAAGCAGAAGACGGCATAACGAGATGCTAGAATACACGCAC GATCCGACGGTAGTGT	Position 19
SLXA_PE_MIPBC2_ REV_020	CAAGCAGAAGACGGCATAACGAGATTGTAATAACACGCAC GATCCGACGGTAGTGT	Position 20
SLXA_PE_MIPBC2_ REV_021	CAAGCAGAAGACGGCATAACGAGATGTACACACACGCAC GATCCGACGGTAGTGT	Position 21
SLXA_PE_MIPBC2_ REV_022	CAAGCAGAAGACGGCATAACGAGATGTACACGAACACGCAC GATCCGACGGTAGTGT	Position 22
SLXA_PE_MIPBC2_ REV_023	CAAGCAGAAGACGGCATAACGAGATACTGTAAGACACGCAC GATCCGACGGTAGTGT	Position 23
SLXA_PE_MIPBC2_ REV_024	CAAGCAGAAGACGGCATAACGAGATGATCATGAACACGCAC GATCCGACGGTAGTGT	Position 24
SLXA_PE_MIPBC2_ REV_025	CAAGCAGAAGACGGCATAACGAGATTGCTAACGACACGCAC GATCCGACGGTAGTGT	Position 25
SLXA_PE_MIPBC2_ REV_026	CAAGCAGAAGACGGCATAACGAGATTAATGCTCACACGCAC GATCCGACGGTAGTGT	Position 26
SLXA_PE_MIPBC2_ REV_027	CAAGCAGAAGACGGCATAACGAGATCTTGTGAACACGCAC GATCCGACGGTAGTGT	Position 27
SLXA_PE_MIPBC2_ REV_028	CAAGCAGAAGACGGCATAACGAGATGATACAAGACACGCAC GATCCGACGGTAGTGT	Position 28
SLXA_PE_MIPBC2_ REV_029	CAAGCAGAAGACGGCATAACGAGATATACAGCGACACGCAC GATCCGACGGTAGTGT	Position 29
SLXA_PE_MIPBC2_ REV_030	CAAGCAGAAGACGGCATAACGAGATCTGGATAGACACGCAC GATCCGACGGTAGTGT	Position 30
SLXA_PE_MIPBC2_ REV_031	CAAGCAGAAGACGGCATAACGAGATTCATCAGCACACGCAC GATCCGACGGTAGTGT	Position 31
SLXA_PE_MIPBC2_ REV_032	CAAGCAGAAGACGGCATAACGAGATTTATACCGACACGCAC GATCCGACGGTAGTGT	Position 32
SLXA_PE_MIPBC2_ REV_033	CAAGCAGAAGACGGCATAACGAGATGCTGATTCACACGCAC GATCCGACGGTAGTGT	Position 33
SLXA_PE_MIPBC2_ REV_034	CAAGCAGAAGACGGCATAACGAGATATAAGGCCACACGCAC GATCCGACGGTAGTGT	Position 34
SLXA_PE_MIPBC2_ REV_035	CAAGCAGAAGACGGCATAACGAGATAAGACTAGACACGCAC GATCCGACGGTAGTGT	Position 35
SLXA_PE_MIPBC2_ REV_036	CAAGCAGAAGACGGCATAACGAGATTAGCAACAACACGCAC GATCCGACGGTAGTGT	Position 36
SLXA_PE_MIPBC2_ REV_037	CAAGCAGAAGACGGCATAACGAGATCAAGTAATACACGCAC GATCCGACGGTAGTGT	Position 37
SLXA_PE_MIPBC2_ REV_038	CAAGCAGAAGACGGCATAACGAGATTCATGATGACACGCAC GATCCGACGGTAGTGT	Position 38
SLXA_PE_MIPBC2_ REV_039	CAAGCAGAAGACGGCATAACGAGATTAGCTCTTACACGCAC GATCCGACGGTAGTGT	Position 39
SLXA_PE_MIPBC2_ REV_040	CAAGCAGAAGACGGCATAACGAGATACTACATGACACGCAC GATCCGACGGTAGTGT	Position 40
SLXA_PE_MIPBC2_ REV_041	CAAGCAGAAGACGGCATAACGAGATTGTGCCATACACGCAC GATCCGACGGTAGTGT	Position 41
SLXA_PE_MIPBC2_ REV_042	CAAGCAGAAGACGGCATAACGAGATGACTTCCAACACGCAC GATCCGACGGTAGTGT	Position 42
SLXA_PE_MIPBC2_ REV_043	CAAGCAGAAGACGGCATAACGAGATTATTCTCGACACGCAC GATCCGACGGTAGTGT	Position 43
SLXA_PE_MIPBC2_ REV_044	CAAGCAGAAGACGGCATAACGAGATCACTAGATACACGCAC GATCCGACGGTAGTGT	Position 44

SLXA_PE_MIPBC2_ REV_045	CAAGCAGAAGACGGCATAACGAGATCTTCGCATACACGCAC GATCCGACGGTAGTGT	Position 45
SLXA_PE_MIPBC2_ REV_046	CAAGCAGAAGACGGCATAACGAGATCTAGCGATACACGCAC GATCCGACGGTAGTGT	Position 46
SLXA_PE_MIPBC2_ REV_047	CAAGCAGAAGACGGCATAACGAGATTACGTTAAACACGCAC GATCCGACGGTAGTGT	Position 47
SLXA_PE_MIPBC2_ REV_048	CAAGCAGAAGACGGCATAACGAGATAAGTAGTCACACGCAC GATCCGACGGTAGTGT	Position 48
SLXA_PE_MIPBC2_ REV_049	CAAGCAGAAGACGGCATAACGAGATGACGTATCACACGCAC GATCCGACGGTAGTGT	Position 49
SLXA_PE_MIPBC2_ REV_050	CAAGCAGAAGACGGCATAACGAGATACGTTTCAGACACGCAC GATCCGACGGTAGTGT	Position 50
SLXA_PE_MIPBC2_ REV_051	CAAGCAGAAGACGGCATAACGAGATAGCTATATACACGCAC GATCCGACGGTAGTGT	Position 51
SLXA_PE_MIPBC2_ REV_052	CAAGCAGAAGACGGCATAACGAGATAGTGCAGAAACACGCAC GATCCGACGGTAGTGT	Position 52
SLXA_PE_MIPBC2_ REV_053	CAAGCAGAAGACGGCATAACGAGATCAATGGTAACACGCAC GATCCGACGGTAGTGT	Position 53
SLXA_PE_MIPBC2_ REV_054	CAAGCAGAAGACGGCATAACGAGATATCCGATACACGCAC GATCCGACGGTAGTGT	Position 54
SLXA_PE_MIPBC2_ REV_055	CAAGCAGAAGACGGCATAACGAGATAGCTGTTGACACGCAC GATCCGACGGTAGTGT	Position 55
SLXA_PE_MIPBC2_ REV_056	CAAGCAGAAGACGGCATAACGAGATTTCCGACTCACACGCAC GATCCGACGGTAGTGT	Position 56
SLXA_PE_MIPBC2_ REV_057	CAAGCAGAAGACGGCATAACGAGATCGAATGGTACACGCAC GATCCGACGGTAGTGT	Position 57
SLXA_PE_MIPBC2_ REV_058	CAAGCAGAAGACGGCATAACGAGATTAGTGTACACACGCAC GATCCGACGGTAGTGT	Position 58
SLXA_PE_MIPBC2_ REV_059	CAAGCAGAAGACGGCATAACGAGATACGCATGAACACGCAC GATCCGACGGTAGTGT	Position 59
SLXA_PE_MIPBC2_ REV_060	CAAGCAGAAGACGGCATAACGAGATTTCTGTAACACGCAC GATCCGACGGTAGTGT	Position 60
SLXA_PE_MIPBC2_ REV_061	CAAGCAGAAGACGGCATAACGAGATTTCTGTAACACGCAC GATCCGACGGTAGTGT	Position 61
SLXA_PE_MIPBC2_ REV_062	CAAGCAGAAGACGGCATAACGAGATTGGAAGCAACACGCAC GATCCGACGGTAGTGT	Position 62
SLXA_PE_MIPBC2_ REV_063	CAAGCAGAAGACGGCATAACGAGATGATGCTCAACACGCAC GATCCGACGGTAGTGT	Position 63
SLXA_PE_MIPBC2_ REV_064	CAAGCAGAAGACGGCATAACGAGATAGCCTGAAACACGCAC GATCCGACGGTAGTGT	Position 64
SLXA_PE_MIPBC2_ REV_065	CAAGCAGAAGACGGCATAACGAGATCACCAGTAACACGCAC GATCCGACGGTAGTGT	Position 65
SLXA_PE_MIPBC2_ REV_066	CAAGCAGAAGACGGCATAACGAGATGTTACACCACACGCAC GATCCGACGGTAGTGT	Position 66
SLXA_PE_MIPBC2_ REV_067	CAAGCAGAAGACGGCATAACGAGATGCGTTAGTACACGCAC GATCCGACGGTAGTGT	Position 67
SLXA_PE_MIPBC2_ REV_068	CAAGCAGAAGACGGCATAACGAGATTGACGTTACACGCAC GATCCGACGGTAGTGT	Position 68
SLXA_PE_MIPBC2_ REV_069	CAAGCAGAAGACGGCATAACGAGATTAGGCGTAACACGCAC GATCCGACGGTAGTGT	Position 69
SLXA_PE_MIPBC2_ REV_070	CAAGCAGAAGACGGCATAACGAGATAAGTGAGCACACGCAC GATCCGACGGTAGTGT	Position 70
SLXA_PE_MIPBC2_ REV_071	CAAGCAGAAGACGGCATAACGAGATCCGGAATAACACGCAC GATCCGACGGTAGTGT	Position 71
SLXA_PE_MIPBC2_ REV_072	CAAGCAGAAGACGGCATAACGAGATAATAGGCAACACGCAC GATCCGACGGTAGTGT	Position 72
SLXA_PE_MIPBC2_ REV_073	CAAGCAGAAGACGGCATAACGAGATTTTCAGGTCACACGCAC GATCCGACGGTAGTGT	Position 73
SLXA_PE_MIPBC2_ REV_074	CAAGCAGAAGACGGCATAACGAGATTAATCGCTACACGCAC GATCCGACGGTAGTGT	Position 74
SLXA_PE_MIPBC2_ REV_075	CAAGCAGAAGACGGCATAACGAGATCTTAGGTGACACGCAC GATCCGACGGTAGTGT	Position 75
SLXA_PE_MIPBC2_ REV_076	CAAGCAGAAGACGGCATAACGAGATTTGATCTCACACGCAC GATCCGACGGTAGTGT	Position 76
SLXA_PE_MIPBC2_ REV_077	CAAGCAGAAGACGGCATAACGAGATTCTTACTGACACGCAC GATCCGACGGTAGTGT	Position 77

SLXA_PE_MIPBC2_ REV_078	CAAGCAGAAGACGGCATAACGAGATGAGTACACACACGCAC GATCCGACGGTAGTGT	Position 78
SLXA_PE_MIPBC2_ REV_079	CAAGCAGAAGACGGCATAACGAGATACGTTACAACACGCAC GATCCGACGGTAGTGT	Position 79
SLXA_PE_MIPBC2_ REV_080	CAAGCAGAAGACGGCATAACGAGATTGGTAAGCACACGCAC GATCCGACGGTAGTGT	Position 80
SLXA_PE_MIPBC2_ REV_081	CAAGCAGAAGACGGCATAACGAGATGATAGCACACACGCAC GATCCGACGGTAGTGT	Position 81
SLXA_PE_MIPBC2_ REV_082	CAAGCAGAAGACGGCATAACGAGATGACTTAGGACACGCAC GATCCGACGGTAGTGT	Position 82
SLXA_PE_MIPBC2_ REV_083	CAAGCAGAAGACGGCATAACGAGATAGTCGTTAACACGCAC GATCCGACGGTAGTGT	Position 83
SLXA_PE_MIPBC2_ REV_084	CAAGCAGAAGACGGCATAACGAGATCGTGATTAACACGCAC GATCCGACGGTAGTGT	Position 84
SLXA_PE_MIPBC2_ REV_085	CAAGCAGAAGACGGCATAACGAGATTCAGTAGAACACGCAC GATCCGACGGTAGTGT	Position 85
SLXA_PE_MIPBC2_ REV_086	CAAGCAGAAGACGGCATAACGAGATATGTGTCTACACGCAC GATCCGACGGTAGTGT	Position 86
SLXA_PE_MIPBC2_ REV_087	CAAGCAGAAGACGGCATAACGAGATCAGAATATACACGCAC GATCCGACGGTAGTGT	Position 87
SLXA_PE_MIPBC2_ REV_088	CAAGCAGAAGACGGCATAACGAGATTAAGTCTGACACGCAC GATCCGACGGTAGTGT	Position 88
SLXA_PE_MIPBC2_ REV_089	CAAGCAGAAGACGGCATAACGAGATATCGACATACACGCAC GATCCGACGGTAGTGT	Position 89
SLXA_PE_MIPBC2_ REV_090	CAAGCAGAAGACGGCATAACGAGATACAGCTGTACACGCAC GATCCGACGGTAGTGT	Position 90
SLXA_PE_MIPBC2_ REV_091	CAAGCAGAAGACGGCATAACGAGATATCACAAAGACACGCAC GATCCGACGGTAGTGT	Position 91
SLXA_PE_MIPBC2_ REV_092	CAAGCAGAAGACGGCATAACGAGATGTGCAATTACACGCAC GATCCGACGGTAGTGT	Position 92
SLXA_PE_MIPBC2_ REV_093	CAAGCAGAAGACGGCATAACGAGATTGCCGATTACACGCAC GATCCGACGGTAGTGT	Position 93
SLXA_PE_MIPBC2_ REV_094	CAAGCAGAAGACGGCATAACGAGATACTAATGCACACGCAC GATCCGACGGTAGTGT	Position 94
SLXA_PE_MIPBC2_ REV_095	CAAGCAGAAGACGGCATAACGAGATCTAGTAGGACACGCAC GATCCGACGGTAGTGT	Position 95
SLXA_PE_MIPBC2_ REV_096	CAAGCAGAAGACGGCATAACGAGATTTAAGCCAACACGCAC GATCCGACGGTAGTGT	Position 96

Table 5.1: Indexed adapter oligo sequences. 16 forward barcoded adapter oligos (one for each plate of 96 DNA samples comprising genomic DNA used for this study) were used along with 96 individual reverse barcoded adapter oligos in the PCR amplification step to allow for individual barcoding of all 1,532 samples across the experiment. Reverse barcode positions refer to positions on the 96 well plate of DNA samples that were maintained for the MIP capture reactions and subsequent amplifications. *Contributions: primers generated by commercial facility after being designed by SAK and Ben Voight.*

Table 5.2

Table 5.2: Sequencing Primers	
Name	Sequence
Sequencing Forward Primers to sequence Barcode Index 2 (16 oligos pooled together)	
SLXA_PE_MIPBC_FOR-SeqRead1	CATACGAGATCCGTAATCGGGAAGCTGAAG
FOR_001-SeqRead1	CAAAGATACTCCGTAATCGGGAAGCTGAAG
FOR_002-SeqRead1	CACAGTCTCTCCGTAATCGGGAAGCTGAAG
FOR_003-SeqRead1	CCAAGGTGATCCGTAATCGGGAAGCTGAAG
FOR_004-SeqRead1	CCAGTCTGGTCCGTAATCGGGAAGCTGAAG
FOR_005-SeqRead1	CGACACTACTCCGTAATCGGGAAGCTGAAG
FOR_006-SeqRead1	CGTGCTTATTCCGTAATCGGGAAGCTGAAG
FOR_007-SeqRead1	CTGGAGCTGTCCGTAATCGGGAAGCTGAAG
FOR_008-SeqRead1	CTGGCACCTTCCGTAATCGGGAAGCTGAAG
FOR_009-SeqRead1	CTTACGGTGTCCGTAATCGGGAAGCTGAAG
FOR_010-SeqRead1	CTTTCATAGTCCGTAATCGGGAAGCTGAAG
FOR_011-SeqRead1	CGTTTCAGATCCGTAATCGGGAAGCTGAAG
FOR_012-SeqRead1	CTCAAAAAGTCCGTAATCGGGAAGCTGAAG
FOR_013-SeqRead1	CTATGCTGCTCCGTAATCGGGAAGCTGAAG
FOR_014-SeqRead1	CTAGCTGCCTCCGTAATCGGGAAGCTGAAG
FOR_015-SeqRead1	CAGCGAATGTCCGTAATCGGGAAGCTGAAG
Sequencing Reverse Primer to sequence Barcode Index 2	
MIPBC_SEQ_REV	ACACGCACGATCCGACGGTAGTGT
Sequencing Index Primer to sequence Barcode Index 2	
MIPBC_SEQ_IND	ACACTACCGTCGGATCGTGCGTGT

Table 5.2: Sequencing primers. 16 individual sequencing primers, each corresponding to the 16 respective forward indexed adapters, were used in an equimolar pool along with a common reverse sequencing primer and sequencing index primer to sequence through the barcode index 2 that was ligated from the 96 individual reverse indexed adapters for the Illumina HiSeq2500 sequencing reactions. *Contributions: Primers designed by SAK and Paul Babb.*

Table 5.3

Table 5.3: Variant Filtering Criteria	
Variant type	Filtering criteria
SNPs	Quality by Depth < 2.0, Fisher Strand Test > 60.0, Mapping Quality < 30.0, Strand Odds Ratio Test > 4.0, Mapping Quality Rank Sum Test < -12.5, Read Position Rank Sum Test < -8.0
INDELs	Quality by Depth < 2.0, Read Position Rank Sum Test < -20.0, Inbreeding Coefficient < -0.8, Fisher Strand Test > 200.0, Strand Odds Ratio Test > 10.0

Table 5.3: Variant Filtering Criteria. Criteria for inclusion of SNP and insertion-deletion (INDEL) variants using the GATK pipeline for variant annotation during analysis of sequencing run data, as described in the Materials and Methods. *Contributions: Analysis relevant to this table performed by Paul Babb.*

Table 5.4

Table 5.4: Characteristics of Participants for Targeted Sequencing								
	High HDL Cohort			Low HDL Cohort			High vs. Low HDL Cohort (T test)	
	All (N=789)	Males (N=228)	Females (N=561)	All (N=743)	Males (N=454)	Women (N=289)		
Age (SD)	58 (13)	59 (15)	58 (12)	Age	55 (13)	56 (12)	53 (15)	P<0.0001
Caucasian %	86.2	89.9	84.7	Caucasian	61.5	65	56.1	N/A
Ashkenazi %	7.9	8.3	7.7	Ashkenazi	2.6	3.5	1	N/A
Black %	4.6	2.2	5.5	Black	27.5	23.3	33.9	N/A
Total Cholesterol (mg/dL)	240 (42)	227 (40)	245 (42)	Total Cholesterol	177 (72)	172 (74)	185 (68)	P<0.0001
HDL-C (mg/dL)	107 (21)	94 (19)	112 (19)	HDL-C	32 (11)	31 (12)	34 (8)	P<0.0001
LDL-C (mg/dL)	127 (60)	127 (40)	127 (71)	LDL-C	100 (59)	96 (58)	105 (61)	N.S.
TG (mg/dL)	77 (34)	78 (37)	77 (32)	TG	266 (566)	270 (537)	259 (610)	P<0.0001

Table 5.4: Characteristics of participants for MIP targeted sequencing. Participants were recruited from the Penn High HDL Study as previously described. All lipid measurements were performed on plasma collected after participants fasted overnight. Comparisons of absolute measurements were performed using a Student's unpaired T-test of all High HDL Cohort participants vs. all Low HDL Cohort participants. All absolute data is reported as mean \pm S.D.. *Contributions: Data generated by clinical laboratory and analyzed by SAK.*

Table 5.5

Table 5.5: Variants identified by MIP targeted sequencing of high HDL-C participants						
All variants identified (4999)						
Region	Total number of variants (SNPs & Indels)	Known		Novel		Variants associated with HDL-C (P<2.5x10E-05, Score test)
<i>GALNT2</i>	958	127		831		0
<i>APOC5-APOC3</i>	709	77		632		0
<i>SBNO1</i>	83	5		78		0
<i>CCDC92-ZNF664</i>	900	107		793		2
<i>SCARB1</i>	813	90		723		0
<i>CETP</i>	633	95		538		22
<i>LIPG</i>	903	127		776		4
<i>Total</i>	4999	628		4371		28
Variants nominally associated with high HDL-C (P<0.01, Score Test)						
Region	Coding	Noncoding	Common (MAF>0.05)	Rare & Low-frequency (MAF<0.05)	Known	Novel
<i>GALNT2</i>	0	10	8	2	8	2
<i>APOC5-APOC3</i>	2	8	8	2	7	3
<i>SBNO1</i>	0	0	0	0	0	0
<i>CCDC92-ZNF664</i>	2	14	11	5	14	2
<i>SCARB1</i>	2	3	0	5	3	2
<i>CETP</i>	6	25	21	10	28	3
<i>LIPG</i>	5	15	12	8	15	5
<i>Total</i>	17	75	60	32	75	17

Table 5.5: Variants identified by MIP sequencing of high HDL-C participants. Single nucleotide variants (SNPs) and insertion-deletion variants (INDELs) were assessed for each gene region (GRC37/hg19),. Known variants were those for which an rsID existed or were able to be ascertained in publically available variant databases including dbSNP,

1000 Genomes, the NHLBI Exome Variant Server and the Exome Aggregation Consortium (ExAC) database. Minor alleles of identified variants were compared for frequency in the high vs. low HDL cohort by the Score test statistic. Noncoding variants included any variants that were not present in protein-coding regions of the gene regions, including splice-site, intronic, 5'UTR, 3'UTR and intergenic variants.

Contributions: All data in this table generated from bench experiments performed by SAK and analyses of sequencing data performed by a Paul Babb and then by SAK. Final data analyses and QC performed by SAK.

Table 5.6

Table 5.6: Variants nominally associated with high HDL-C (P<0.01, Score Test)								
Region	Chromosome	Position	Variant	rsID (if known variant)	Variant call rate ¹	MAF	Score statistic	P-value (Score test)
GALNT2	1	230293530	C/T	rs4631704	0.97781	0.4723	2.5884	0.009643
	1	230294185	A/T	rs910502	0.99478	0.11975	2.9236	0.0034603
	1	230294989	C/A	rs1555290	0.98303	0.1328	2.7426	0.0060964
	1	230295245	C/T	rs12065546	0.96932	0.13737	2.9757	0.0029235
	1	230295307	C/G	rs17315646	0.99413	0.48424	2.7679	0.0056412
	1	230295691	G/A	rs4846914	0.98368	0.47545	2.7894	0.0052813
	1	230296153	C/T	rs10864726	0.98368	0.48142	3.3998	0.00067442
	1	230296492	G/A	rs4846916	0.60444	0.18143	3.3858	0.00070963
	1	230416051	C/T		0.43277	0.021116	3.3749	0.00073839
	1	230417044	AC/A		0.99608	0.0022936	-2.5816	0.0098341
APOC5-APOC3	11	116662331	G/T	rs12287066	0.99413	0.082075	-3.1879	0.0014332
	11	116662407	G/C	rs3135506	0.87663	0.078183	-3.4676	0.00052513
	11	116662579	C/T	rs651821	0.98042	0.096538	3.4607	0.00053868
	11	116663707	G/A	rs662799	0.99217	0.096053	2.8851	0.0039128
	11	116678249	CAA/CAAA		0.76893	0.25806	3.4891	0.00048469
	11	116680918	G/A	rs1263172	0.96802	0.38065	2.7578	0.0058186
	11	116681008	G/A	rs1263173	0.95104	0.45504	2.8793	0.0039852
	11	116684516	C/A		0.98499	0.00066269	3.2042	0.0013543
	11	116684562	C/G	rs2727791	0.98303	0.40604	2.9501	0.0031764
	11	116703697	A/G		0.9876	0.0019828	2.7368	0.0062034
CCDC92-ZNF664	12	124421453	T/C	rs9863	0.99608	0.40957	3.5326	0.00041144

	12	124427306	T/A	rs11057401	0.99543	0.33967	4.5426	5.56E-06
	12	124428162	T/A	rs4930725	0.99739	0.36257	3.8431	0.00012147
	12	124428331	T/C	rs4930726	0.9752	0.37383	3.9729	7.10E-05
	12	124428459	GCACC/G		0.90927	0.00035894	3.1688	0.0015308
	12	124428577	AT/A	rs143840819	0.97063	0.0090787	-2.8329	0.0046133
	12	124429279	G/A	rs3186071	0.98695	0.32474	4.0264	5.66E-05
	12	124430612	G/A	rs4765305	0.95496	0.48154	3.7265	0.00019412
	12	124430812	G/A	rs4765335	0.98825	0.39861	3.5877	0.00033365
	12	124431049	G/A	rs11835839	0.96214	0.41689	4.4826	7.37E-06
	12	124431109	GT/G		0.97715	0.000334	3.1966	0.0013907
	12	124472571	T/C	rs4765562	0.98433	0.39025	3.1208	0.0018036
	12	124496316	A/G	rs1054852	0.95561	0.42418	2.7152	0.0066242
	12	124497177	A/G	rs377752073	0.99347	0.040736	-2.5847	0.0097451
	12	124499129	T/TG	rs34578644	0.99804	0.015697	-2.6971	0.0069944
	12	124499839	C/T	rs3768	0.99282	0.22485	3.8327	0.00012677
SCARB1	12	125260864	TCC/T		0.99413	0.003283	2.676	0.0074497
	12	125263061	C/T	rs3825140	0.50522	0.0096899	3.0476	0.0023067
	12	125266613	A/G		0.062663	0.026042	2.6595	0.0078265
	12	125271101	A/C	rs2293439	0.99086	0.006917	2.7355	0.0062292
	12	125298877	G/A	rs5889	0.99347	0.0098555	3.7806	0.00015646
CETP	16	56995236	C/A	rs1800775	0.87337	0.32175	7.307	2.73E-13
	16	56995814	G/A	rs34498052	0.94321	0.0010381	4.2945	1.75E-05
	16	56995908	T/A	rs34119551	0.98564	0.00033113	3.201	0.0013696
	16	56996104	C/A	rs17231534	0.98303	0.072709	3.2765	0.0010511

16	56996158	T/C	rs3816117	0.98433	0.47513	9.7264	2.33E-22
16	56996211	G/A	rs711752	0.97911	0.42967	7.9663	1.63E-15
16	56996288	G/A	rs708272	0.97977	0.44204	7.8132	5.58E-15
16	56997136	G/A		0.36292	0.0017986	3.3039	0.00095353
16	56997137	G/A		0.36162	0.0018051	3.3026	0.00095806
16	56998918	A/G	rs12720926	0.92102	0.36464	8.1117	4.99E-16
16	56999184	AT/A	rs35585922	0.99608	0.08519	3.6898	0.00022441
16	56999258	A/C	rs7203984	0.96345	0.23103	-8.5873	8.90E-18
16	56999328	C/T	rs11508026	0.97846	0.3956	9.7472	1.90E-22
16	57001216	C/T	rs4784741	0.9654	0.26775	7.7259	1.11E-14
16	57001254	T/TCACA	rs12720908	0.96606	0.19527	-7.8463	4.29E-15
16	57001274	AC/A	rs200751500	0.87337	0.13229	6.1576	7.39E-10
16	57001438	G/A	rs12444012	0.23825	0.49315	4.4254	9.63E-06
16	57001547	A/G	rs142093596	0.38316	0.0017036	3.0925	0.0019848
16	57004889	G/A	rs7205804	0.96345	0.35637	7.3489	2.00E-13
16	57005300	G/A		0.97258	0.0026846	-2.6028	0.0092458
16	57005301	C/T	rs1532625	0.96997	0.35834	8.4272	3.54E-17
16	57005883	G/A	rs374409989	0.8577	0.0022831	4.5547	5.24E-06
16	57007353	C/T	rs5883	0.97454	0.073007	4.9778	6.43E-07
16	57007446	T/G	rs11076176	0.98303	0.1836	-7.0854	1.39E-12
16	57015084	C/G	rs7192120	0.99086	0.0098814	-2.6659	0.0076776
16	57015091	G/C	rs5880	0.99021	0.034608	-4.7289	2.26E-06
16	57016092	G/A	rs5882	0.99151	0.37261	-4.38	1.19E-05
16	57017319	G/A	rs1800777	0.99151	0.024358	-5.3247	1.01E-07

	16	57017474	G/A	rs289741	0.91971	0.3577	-4.7505	2.03E-06
	16	57017662	G/A	rs1801706	0.98172	0.17121	5.0081	5.50E-07
	16	57017796	G/A	rs289743	0.93016	0.22561	-2.988	0.0028081
LIPG	18	47088548	CT/C		0.98368	0.0023225	3.1977	0.0013855
	18	47088690	C/A	rs874565	0.98433	0.020889	-2.8211	0.0047863
	18	47088741	C/T	rs874566	0.99347	0.02661	-3.7617	0.00016877
	18	47088830	C/G	rs35816125	0.99086	0.29249	2.679	0.007384
	18	47093864	C/T	rs2000813	0.99347	0.26905	3.2313	0.0012322
	18	47096016	G/A	rs1320700	0.95692	0.27729	4.276	1.90E-05
	18	47109955	A/G	rs77960347	0.9765	0.01738	3.4682	0.0005239
	18	47110088	C/T		0.96475	0.0023681	-2.6056	0.0091717
	18	47113196	G/T		0.99935	0.00032658	3.195	0.0013981
	18	47117376	C/T	rs3744840	0.99543	0.043934	3.1834	0.0014553
	18	47157400	T/C	rs2000825	0.9876	0.14243	3.3521	0.00080214
	18	47157439	C/T	rs2000826	0.98172	0.34973	3.6989	0.00021654
	18	47158186	T/C	rs10438978	0.99608	0.19397	5.0649	4.09E-07
	18	47158234	C/T	rs9304381	0.99608	0.17857	4.7841	1.72E-06
	18	47158307	AC/A		0.022846	0.071429	-2.7203	0.0065219
	18	47164926	T/C	rs6507937	0.98042	0.13848	3.3464	0.00081856
	18	47167214	T/C	rs4939883	0.99739	0.20844	5.0238	5.07E-07
	18	47167344	G/C		0.98042	0.00033289	3.1924	0.0014108
	18	47167407	T/C	rs4939884	0.98107	0.13373	3.3699	0.00075194
	18	47179516	G/A	rs1943973	0.98499	0.10835	3.4066	0.0006578

Table 5.6: Single variant associations with high HDL-C. Variants (SNPs and INDELs) across targets were compared for frequency of the minor allele in high vs. low HDL participants by Score test statistic. $P < 0.01$ was considered nominally statistically significant. MAF refers to minor allele frequency within the sequencing cohort. Call rate refers to the fraction of 1,532 samples for which variant position was sequenced and passed variant quality filtering. *Contributions: All data in this table generated from bench experiments performed by SAK and analyses of sequencing data performed by Paul Babb and then by SAK. Final data analyses and QC performed by SAK.*

Table 5.7

Table 5.7: Replication of GWAS-significant HDL-C associations with MIP sequencing													
Region	Chr.	Position	Ref. allele	Alt. allele	rsID	GLGC+MetaboChip GWAS	1000 Genomes	MIP sequencing					
						P-value, association with HDL-C	MAF	Call rate	Overall MAF	MAF - High HDL Cohort	MAF - Low HDL Cohort	Score statistic	P-value
GALNT2	1	230295245	C	T	rs12065546	1.50E-15	0.8443	0.96932	0.13737	0.87857	0.84545	2.9757	0.0029235
	1	230295307	C	G	rs17315646	1.35E-36	0.5844	0.99413	0.48424	0.57785	0.44884	2.7679	0.0056412
	1	230295691	G	A	rs4846914	3.51E-41	0.5844	0.98368	0.47545	0.59073	0.45411	2.7894	0.0052813
	1	230296153	C	T	rs10864726	8.65E-24	0.5858	0.98368	0.48142	0.58869	0.44376	3.3998	0.00067442
APOC5-APOC3	11	116662331	G	T	rs12287066	1.08E-20	0.94195	0.99413	0.082075	0.061083	0.10494	-3.1879	0.0014332
	11	116662407	G	C	rs3135506	7.74E-16	0.94327	0.87663	0.078183	0.055866	0.10367	-3.4676	0.00052513
	11	116662579	C	T	rs651821	7.72E-26	0.91953	0.98042	0.096538	0.93266	0.87133	3.4607	0.00053868
	11	116663707	G	A	rs662799	4.16E-37	0.91953	0.99217	0.096053	0.92522	0.88098	2.8851	0.0039128
CCDC92-ZNF664	12	124427306	T	A	rs11057401	4.53E-08	0.2504	0.99543	0.33967	0.37894	0.29713	4.5426	5.56E-06
	12	124428331	T	C	rs4930726	1.53E-09	0.3641	0.9752	0.37383	0.3947	0.35139	3.9729	7.10E-05
CETP	16	56995236	C	A	rs1800775	3.33E-644	0.4802	0.87337	0.32175	0.75943	0.59852	7.307	2.73E-13
	16	56996211	G	A	rs711752	1.28E-641	0.4222	0.97911	0.42967	0.52635	0.32548	7.9663	1.63E-15
	16	56999258	A	C	rs7203984	3.59E-517	0.777	0.96345	0.23103	0.12092	0.34951	-8.5873	8.90E-18
	16	56999328	C	T	rs11508026	2.63E-318	0.4142	0.97846	0.3956	0.51677	0.26588	9.7472	1.90E-22
	16	57004889	G	A	rs7205804	5.27E-675	0.4235	0.96345	0.35637	0.46414	0.24412	7.3489	2.00E-13
	16	57005301	C	T	rs1532625	2.25E-397	0.4235	0.96997	0.35834	0.46084	0.24931	8.4272	3.54E-17
	16	57007353	C	T	rs5883	1.76E-31	0.06069	0.97454	0.073007	0.094716	0.049512	4.9778	6.43E-07
	16	57015091	G	C	rs5880	1.37E-233	0.94063	0.99021	0.034608	0.019133	0.05116	-4.7289	2.26E-06
16	57016092	G	A	rs5882	2.21E-58	0.3325	0.99151	0.37261	0.61111	0.64512	-4.38	1.19E-05	

	16	57017474	G	A	rs289741	7.64E-160	0.3219	0.91971	0.3577	0.61172	0.67471	-4.7505	2.03E-06
	16	57017662	G	A	rs1801706	1.09E-15	0.1939	0.98172	0.17121	0.20968	0.13032	5.0081	5.50E-07
LIPG	18	47093864	C	T	rs2000813	1.08E-23	0.2902	0.99347	0.26905	0.31266	0.22199	3.2313	0.0012322
	18	47157400	T	C	rs2000825	2.93E-24	0.8245	0.9876	0.14243	0.87739	0.8363	3.3521	0.00080214
	18	47157439	C	T	rs2000826	1.36E-07	0.6649	0.98172	0.34973	0.6982	0.59904	3.6989	0.00021654
	18	47158186	T	C	rs10438978	1.56E-27	0.8179	0.99608	0.19397	0.86398	0.74317	5.0649	4.09E-07
	18	47158234	C	T	rs9304381	3.06E-24	0.8193	0.99608	0.17857	0.86839	0.77049	4.7841	1.72E-06
	18	47164926	T	C	rs6507937	6.52E-24	0.8259	0.98042	0.13848	0.87773	0.84414	3.3464	0.00081856
	18	47167214	T	C	rs4939883	1.80E-66	0.8193	0.99739	0.20844	0.85076	0.72752	5.0238	5.07E-07
	18	47167407	T	C	rs4939884	5.13E-24	0.8206	0.98107	0.13373	0.87916	0.85229	3.3699	0.00075194
	18	47179516	G	A	rs1943973	2.75E-54	0.843	0.98499	0.10835	0.90857	0.87345	3.4066	0.0006578

Table 5.7: Replication of GWAS-significant HDL-C associations with MIP sequencing. Variants identified by MIP sequencing were compared for associations with HDL-C with the Global Lipids Genetics Consortium+MetaboChip GWAS results [117]. 30 variants identified by MIP sequencing in the targeted regions were also associated with HDL-C in GWAS. All 30 variants displayed consistent directionality of association with HDL-C through MIP sequencing ($P < 0.01$, Score test), as observed by GWAS. MAFs were obtained from 1000 Genomes for the given variants and compared with the MAFs observed in the MIP sequencing discovery cohort. *Contributions: All data in this table generated from bench experiments performed by SAK and analyses of sequencing data performed by Paul Babb and then by SAK. Final data analyses and QC performed by SAK.*

Table 5.8

Table 5.8: Gene-burden level association of variants from MIP sequencing with HDL-C										
Aggregation of coding variants in each gene region		Disruptive variants ¹			Disruptive + Missense (1/5) ²			Disruptive + Missense (5/5) ³		
Gene	Region	Number of variants	β (SE)	P-value (Collapsing test)	Number of variants	β (SE)	P-value (Collapsing test)	Number of variants	β (SE)	P-value (Collapsing test)
<i>GALNT2</i>	1:230338882-230415203	50	0.17737(0.1656)	0.28414	75	0.37613(0.13308)	0.0047096	50	0.17737(0.1656)	0.28414
<i>APOA5</i>	11:116660886-116663095	47	-0.1072(0.14206)	0.45047	69	-0.28607(0.12272)	0.01975	47	-0.1072(0.14206)	0.45047
<i>APOC3</i>	11:116701515-116703575	17	-0.30959(0.26308)	0.23928	20	-0.35348(0.25918)	0.17262	17	-0.30959(0.26308)	0.23928
<i>CCDC92</i>	12:124421620-124428847	53	0.18717(0.18344)	0.30759	73	0.53371(0.12312)	1.46E-05	53	0.18717(0.18344)	0.30759
<i>ZNF664</i>	12:124488089-124497235	8	0.84513(0.46749)	0.07064	22	0.13866(0.17381)	0.425	8	0.84513(0.46749)	0.07064
<i>SCARB1</i>	12:125267297-125348261	68	0.17916(0.13531)	0.18547	92	0.042176(0.11697)	0.71842	68	0.17916(0.13531)	0.18547
<i>CETP</i>	16:56995891-57017577	46	-0.34005(0.18686)	0.068783	59	-0.48798(0.13732)	0.00037992	46	-0.34005(0.18686)	0.068783
<i>LIPG</i>	18:47088681-47110121	48	-0.003131(0.1718)	0.98546	67	0.21127(0.14962)	0.15793	48	-0.003131(0.1718)	0.98546
Aggregation of all noncoding variants across targeted regions		Number of variants		Number of single variants		p statistic		P-value (SKAT-O test)		
		3436		467		0		0.05		

Table 5.8: Gene-burden associations of identified variants with high HDL-C

¹Disruptive coding variants included nonsense (stop-gained), frameshift, splice-donor and splice-acceptor variants, as annotated from dbSNP.

²Missense variants were defined as nonsynonymous amino acid-altering variants. Variants were included in this grouping if they were identified as ‘deleterious’ or ‘damaging’ by one of the five in silico prediction tools: SIFT (deleterious), PolyPhen2 HDIV (‘possibly

damaging' or 'probably damaging'), PolyPhen2 HVAR ('possibly damaging' or 'probably damaging'), MutationTaster, and LRT (disruptive).

³Missense variants were included in this grouping if they were predicted to be 'deleterious' or 'damaging' by all five prediction tools as described in the Materials and Methods.

Contributions: All data in this table generated from bench experiments performed by SAK and analyses of sequencing data performed by Paul Babb and then by SAK. Final data analyses and QC performed by SAK.

CHAPTER 6 – SUMMARY AND FUTURE DIRECTIONS

Summary

At the time this dissertation work started, there had already been ample progress in both the understanding of the heritable basis of plasma lipid traits and the development of life-saving therapies to reduce the burden of CHD and MI through lowering atherogenic lipoprotein concentrations in the blood. Statins were already recognized as remarkably effective drugs for LDL-C lowering and disease risk reduction [409-411], and the translation of human genetics discoveries in the PCSK9 realm, fueled in part by *in vivo* physiological and biochemical investigation, was already well underway with phase I studies of PCSK9-inhibitory monoclonal antibodies showing promise in LDL-C lowering [412]. The major players in the physiological pathways regulating LDL-C, HDL-C and TG metabolism were likewise discovered and their biology explored through study of Mendelian disorders of lipoprotein metabolism and animal models of gain- and loss-of-function. GWAS for lipids had expanded to more than 100,000 participants, yielding 95 genomic loci, most of which were novel, associated with at least one blood lipid trait [86]. This era was aptly termed by one investigator as a “treasure trove for lipoprotein biology” [413] because it marked the beginning of an entirely new approach for mechanistic and physiological discovery of the regulation of lipid metabolism and offered the potential for new therapeutic targets. Finally, the application of exome sequencing to the study of rare human phenotypes had just begun, and one of the first uses of this methodology had successfully identified LOF of the lipase inhibitory gene *ANGPTL3* as a genetic cause of a novel Mendelian disorder of lipoprotein metabolism, Familial Combined Hypolipidemia [118].

However, for all of these successes and potentials for further discovery, this period of time also bore much uncertainty about the biological and clinical relevance of some lipid traits to disease and the ways to translate the novel genomics findings to discrete biological pathways and targets. The role of HDL-C as a protective factor against CHD became hotly questioned after low-frequency genetic variant studies using Mendelian randomization did not support a causal role for HDL-C elevation itself in CHD risk reduction [204,414], and several therapeutic efforts to raise HDL-C failed to demonstrate any clinical benefit [19]. The causal relationship of TGs with CHD risk was also debated because of the pleiotropic effects of TG-altering variants on other blood lipids [124], and no existing drug trials completed to appropriately address the question of whether TRL lowering itself in hyperTG patients could orthogonally reduce CHD risk in addition to conventional LDL-C reduction. The need for additional therapies was especially pressing as increasing evidence mounted that LDL-C lowering alone still may leave significant residual vascular risk, and many patients were intolerant to the doses of statins needed to achieve optimal LDL-C levels [19]. And while the era of GWAS and next-generation sequencing allowed the discovery of additional pathways for targeting, the progress from locus discovery to causal gene, causal variant, tissue of action, and mechanism had proven difficult [415]. The goal of this dissertation work was thus to use the existing and newly developed methodologies to address key questions and controversies that have arisen regarding the disease causality, novel biology, and potential clinical promise of HDL and TGs from the vantage point of human genetics.

I first sought to address the controversial relationship of HDL to CHD risk by discovering and investigating a LOF variant in *SCARB1*, a gene encoding the HDL cholesterol receptor that in experimental models is critical for the hepatic clearance of

HDL cholesterol and the RCT pathway postulated to confer protection from CHD [207,277]. This study identified the first homozygote for a LOF variant in *SCARB1* and showed that human SR-BI deficiency results in extremely high HDL-C levels due to ineffective CE clearance by the liver from HDLs. Consistent with prior animal model data but challenging the human epidemiology, *SCARB1* LOF was associated with increased CHD risk despite the marked HDL-C elevation. This work supports and provides the foundation for further study of the specific contribution of the RCT pathway of cholesterol removal via HDL that may be a more clinically relevant measure of disease risk than HDL-C alone.

I next asked a question central to the utility of GWAS for complex traits – what is the causal gene and mechanism of action of a newly identified GWAS-associated locus? I addressed the role of *GALNT2*, one of the top novel genes implicated in HDL-C and TG metabolism through GWAS [86,117], through studies of LOF in humans, mice, rats and nonhuman primates. We compared the lipid phenotypes and consequences of absent *GALNT2* across models and identified a biochemical target of *GALNT2*'s gene product whose function in mediating HDL-C levels was impaired by *GALNT2* deficiency. Through this work, we provide an example for the systematic evaluation of GWAS-implicated genes to uncover underlying conserved and species-specific biological functions.

Thirdly, I sought to understand how recently identified genetic variation underlying TGs and CHD could be leveraged for therapeutic gain. I performed a structure-function analysis of a the A43T LOF variant in *APOC3* identified from exome sequencing [91,106] that reduces TGs and CHD risk independently of LDL-C. Through studying human A43T carriers and humanized mice expressing A43T, I have shown that

this variant confers LOF through promoting ApoC-III clearance and have outlined a mechanism that may be informative for developing ApoC-III-focused therapies.

Lastly, I tested methods to better uncover the noncoding genome for variants underlying phenotypic extremes of lipid traits. I applied a recently reported multiplexed sequencing capture method [368] for targeted resequencing of both the coding and noncoding regions of HDL-C candidate genes in high HDL-C vs. low HDL-C participants. Through a proof-of-principle study, I confirmed reported associations between common and low frequency variants and HDL-C and identified several novel variants in these genes for further study. This work supports the potential for large-scale targeted sequencing to better explain the heritable basis of the extremes of lipid traits. As is the case for any scientific endeavor, my work on these projects raised many questions, perhaps more than they answered, but thus afford many opportunities for exciting new and continued exploration, which I will discuss below.

Future Directions

Role of SCARB1 in HDL metabolism and CHD risk in humans

The discovery that genetic LOF of *SCARB1* in humans is tolerated and largely recapitulates the phenotype of the murine models allows an exciting opportunity to directly assess the contribution of this gene and the RCT pathway to CHD risk in humans. We found a moderate association of the variant with increased CHD in our study of heterozygotes from a multi-ethnic population ($P=0.018$). Though this variant is extremely rare in the general population (MAF in Europeans of 0.0003 in our study), the variant appears to be enriched in the Ashkenazi Jewish community. Indeed, our high HDL-C cohort demonstrated that almost all of the identified carriers were of Ashkenazi

descent. Independently, genotyping in 256 participants of the Ashkenazi Genome Consortium (<https://ashkenazigenome.org/>) showed that the P376L variant was present with a MAF of 2.344%, a nearly ~100-fold higher frequency than in our cohort. Thus, search for this variant specifically in the Ashkenazi population may yield more heterozygous and homozygous carriers that would allow further study of the relationship of the variant to CHD, along with insight about the penetrance of the HDL-C and other metabolic phenotypes in carriers.

Additional P376L carriers provide the opportunity to directly assess RCT in humans via novel methods. Our laboratory developed a protocol to study macrophage-to-feces RCT *in vivo* in animal models that has become the mainstay of RCT studies in the HDL field [223,416-418]. We recently modified this protocol for the study of RCT in humans (under review for publication). Briefly, an isotope-labeled tracer of cholesterol is administered intravenously to human participants and the appearance of this tracer in the plasma, relative distribution among lipoproteins, proportion of free vs. esterified cholesterol, and clearance is measured, along with the appearance of the tracer in the feces (a measure of biliary excretion). This assay has been applied in a limited number of human participants to date, including a patient with ABCA1 deficiency (Tangier disease) and LCAT deficiency, two Mendelian disorders of HDL metabolism leading to extremely low HDL-C levels. With the identification of homozygous *SCARB1* deficiency, there may now be potential to directly interrogate the RCT pathway via this assay in human subjects homozygous for the *SCARB1* P376L variant or other confirmed LOF variants yet to be identified.

Systematic gene-by-gene dissection of the RCT pathway through human homozygotes for LOF variants could provide unequivocal evidence of the existence and

relevance of this pathway in humans. If this is coupled to measures of atherosclerotic burden (e.g. coronary calcium scores, carotid intimal medial thickness), this method could help relate the RCT hypothesis and HDL functionality over HDL-C levels with CHD risk in humans. This effort could also resolve a controversy raised from a prior study of cholesterol flux using tracer-labeled isotopes in humans, which purported that most of the hepatic cholesterol taken up from lipoproteins occurs through the LDLR pathway [269]. This study posited that CETP drives CE flux from HDLs to ApoB-containing lipoproteins which then deliver the CE to the liver. This contrasts with the conclusions from my data showing massive HDL-C and specifically CE elevation with human *SCARB1* deficiency. Our work supports the SR-BI pathway as a mediator of at least some of the CE flux to the liver via HDLs. An alternative conclusion from our findings is that ‘constipation’ of this pathway through SR-BI deficiency may have indirect effects on the ability of the CE to be delivered via the other receptors by ApoB-lipoprotein mediated transport. The comparison of RCT in humans with SR-BI, LDLR, and CETP deficiency, along with those with variants in upstream mediators of the pathway (LOF of APOA1, ABCA1, and LCAT) would provide thus invaluable insight. Moreover, such an experiment would aid in pinpointing specific nodes of the pathway most relevant for therapeutic intervention.

The identification of *SCARB1* deficiency also allows the exploration of the other, non-RCT related roles of SR-BI. SR-BI is critical for cholesterol uptake for steroid hormone biosynthesis in the adrenal glands and gonads in mice, and indeed SR-BI deficiency in these animals results in profound metabolic consequences, including infertility, adrenal glucocorticoid insufficiency, thrombocytopenia and platelet activity dysfunction [208,277]. Many of these findings were absent in the *SCARB1* homozygote

we identified, who presented with normal fertility and no observable endocrine anomalies. In this regard, it is noteworthy that the phenotype of the P376L homozygote resembles that of the PDZK1 KO mouse, a mouse deficient in an adapter protein for SR-BI localization to the cell membrane in hepatocytes [275]. Thus, it is possible that our homozygote displays the liver findings of SR-BI deficiency but not the findings seen from the global *SCARB1* KO mouse. Additional variant discovery to identify more LOF variants of *SCARB1* in humans may help further dissect the set of phenotypes of SR-BI deficiency that affect HDL metabolism and other metabolic phenotypes in humans.

Likewise, additional studies of the P376L variant, both in animal models recapitulating the variant such as CRISPR/Cas9 knock-in mice and also in human derived cell types from iPSCs, may further elucidate the myriad functions of SR-BI to vascular homeostasis that are altered with this variant. Paramount to these studies is a deeper inspection of how this variant's impact on SR-BI post-translational processing ultimately modulates its varied functions (**Figure 6.1**). Specifically, previous investigators have shown that endothelial SR-BI is a critical mediator of vasodilation through an HDL-mediated signaling program that drives nitric oxide production [276], and thus SR-BI may confer atheroprotection through this mechanism as well. Future ongoing work will undoubtedly assess the role of SR-BI in CHD through all of these mechanisms using the vantage point of the natural variant and potential for identifying more carriers for phenotyping and translational investigation as a starting point.

Biology of GALNT2 and lipoprotein metabolism

Our interrogation of the *GALNT2* locus started with the GWAS association of this locus with HDL-C and TGs in 2008 [131,132] and resulted in a cross-species comparison of

GALNT2 deficiency across mammalian models. Our eight-year effort has resulted in several new areas for potential study. These include the molecular functions of *GALNT2* in lipid metabolism and its broader biological roles, the reasons for the cross-species differences in phenotypes across the spectrum of *GALNT2* LOF, and the implications of our approach for the elucidation of other GWAS loci for complex traits.

The impact of *GALNT2* on lipoprotein metabolism remains an important area for continued focus, as it will shed light on the broader biological roles of GalNAc transferases and O-glycosylation. We identified a consistent reduction in HDL-C with *GALNT2* deficiency across the models tested, however there were notable differences in the O-glycoproteomes between WT vs. KO organisms for each species. This suggests that multiple glycosylation modifications of different protein targets may act to elicit the same HDL-C phenotype that we saw across the different species. In addition to PLTP, we identified hundreds of target proteins that were differentially glycosylated between WT vs. KO animals or control vs. LOF homozygous humans. Each of these targets is worthy of further inspection to understand the role of the glycosylation modification on the target protein's function. For example, the mouse plasma differential glycoproteomes also identified glycopeptides in hepatic triglyceride lipase (HL), one of the extracellular lipases involved in remodeling of TRL remnants and HDLs through its TG lipase and phospholipase activities [419]. Additionally, mouse liver differential glycoproteomics identified glycopeptides in LRP1, one of the LDL receptor family members that mediates TRL remnant clearance and mediates signaling in extrahepatic tissues [420]. Systematic surveying of the results from our O-glycoproteomics across the humans, mice and rats will facilitate the understanding of the varied functions of O-glycosylation in regulating protein functions.

With regard to the observation that *GALNT2* modulates PLTP to regulate HDL metabolism, additional experiments will be required to determine the physiological impact of O-glycosylation on the molecular and biochemical properties of PLTP. Studies such as site-directed mutagenesis of the glycosylation sites at the C-terminus of PLTP will help determine the exact residue and fold the glycosylation modulates, and will be a key first step in identifying the molecular mechanism. Further studies of isolated or recombinant glycosylated vs. nonglycosylated PLTP such as Michaelis-Menten kinetics will also define the biochemical consequences of PLTP's molecular activity. Likewise, measurement of the association of glycosylated vs. nonglycosylated PLTP with HDLs, TRLs, and phospholipid vesicles will help determine if O-glycosylation stabilizes PLTP's molecular interactions.

Further work in animal models will also be critical in determining the molecular and physiological impact of PLTP O-glycosylation on both PLTP metabolism itself and atherosclerosis (**Figure 6.2**). To study glycosylation and PLTP metabolism, PLTP could be isolated from *GALNT2* WT vs. KO animals or synthesized with or without the glycosylation and radiolabeled to measure PLTP turnover and production, in ways analogous to the ApoC-III turnover studies described to study the A43T missense variant in Chapter 4. This experiment would help determine if PLTP clearance is affected by the O-glycosylation and could be related to the binding of the glycosylated vs. nonglycosylated protein to lipoproteins as mentioned above. To better understand the epistatic relationship between *GALNT2* and PLTP, the *Pltp/Galnt2* double knockout (DKO) mouse could be generated and compared to the single KO models. Similarly, the overexpression of PLTP in the setting of *GALNT2* deficiency could be used to understand if glycosylation is a necessity for maximally functional PLTP activity and HDL

maintenance. Finally, the impact of *GALNT2* function on the physiological consequences of altered PLTP activity could be assessed using these animal models to study RCT, HDL and TRL kinetics, and atherosclerosis *in vivo*. Given the small modulation in HDL-C levels with the absence of *GALNT2* and moderately reduced PLTP activity, the likelihood of a profound effect on atherosclerosis is low, but these studies would be informative to better understand the regulation of PLTP function nonetheless.

The additional targets of *GALNT2* that were identified in humans, ANGPTL3 and ApoC-III, also open the door to many questions about the role of *GALNT2* in the regulation of TRL metabolism pathway proteins. At this point, our data and prior evidence suggest that while ApoC-III is a bona fide target of GalNAc-T2 in humans, the absence of O-glycosylation does not appear to impact steady-state TG levels [95-97]. Yet it is still unclear what the function of the O-glycosylation modification of this target is. For ANGPTL3, evidence establishing a physiological role for glycosylation of ANGPTL3 *in vivo* is also lacking. For both cases, the study of the human homozygotes for *GALNT2* LOF variants will undoubtedly be informative. Mass spectrometry methods may help resolve cleaved vs. intact ANGPTL3 in the plasma of variant carriers vs. controls, an experiment which may confirm results from the prior biochemical studies suggesting that O-glycosylation inhibits ANGPTL3 cleavage and activation [186,279]. For the study of both proteins, glycosylated vs. nonglycosylated forms could be generated and studied for their metabolism in animal model using isotope-labeling methods to track the clearance of the different proteins and their association with the various lipoprotein fractions.

As in the case for the *SCARB1* variant carrier, much is yet to be learned from further study of the humans homozygous for LOF variants in *GALNT2*. We are especially

interested in the possible role of GalNAc-T2 mediated O-glycosylation in the regulation of neuronal development and behavior, because limited observation of the homozygous carriers has demonstrated neurodevelopmental delay and seizure disorders in addition to the lipid phenotypes central for our work. Through additional unpublished work using our *GALNT2* KO mouse, I have found that global *GALNT2* deficiency reduces inhibition, increases fearlessness, decreases social interaction and reduces cerebellar learning capacity *in vivo*. Glycomic measurements of the sugar structures in the plasma of the human *GALNT2* LOF carriers as well as plasma and CNS tissues of *GALNT2* KO mice have shown a consistent and dramatic reduction in O-GalNAc glycosylation with compensatory increases in other forms of glycosylation. While this work is certainly preliminary, the potential to connect *GALNT2* to a syndrome of neurobehavioral function through studies in humans and mice would be exciting. Indeed, it would be one of the few studies establishing deficiency of a GalNAc transferase gene as the cause for a genetic disorder, reminiscent of the seminal discovery of *GALNT3* deficiency as the cause of the Mendelian bone disorder Familial Tumoral Calcinosis [421]. Further studies in the *GALNT2* KO mice and *GALNT2* deficient humans to identify glycoprotein targets differentially regulated in the CNS tissues and their relationship to the observed phenotypes is certainly needed to further this effort.

Therapeutically relevant mechanisms of APOC3 loss-of-function

Our work on *APOC3* LOF variants associated with reduced TGs and protection from CHD set out to identify the mechanisms of putative ApoC-III lowering or inactivation conferred by the variants. We found that three of the four variants expectedly reduced ApoC-III levels likely through reducing expression of the protein. Meanwhile, the fourth

variant, A43T, increased ApoC-III catabolism due to perturbation of binding of ApoC-III to lipoproteins in the circulation, a mechanism distinct from that of the other three variants. Nevertheless, all four variants in aggregate have been associated with a favorable lipoprotein profile and protection from CHD, suggesting that these various mechanisms of ApoC-III LOF are potentially avenues for therapeutic targeting.

One can view the currently developing anti-*APOC3* antisense oligonucleotide (ASO) as a therapy that attempts to mimic the effect of the three *APOC3* LOF variants that reduce *APOC3* expression. This therapy has shown some promise so far in preclinical models and phase I and II studies in severely hyperTG patients [102-104]; however, recent reports of profound thrombocytopenia with ASO treatment have temporarily halted ASO trials, including the *APOC3* ASO program. By elucidating the mechanism of the A43T variant (**Figure 6.3**), I hope that our findings will serve as a foundation for pharmacological endeavors to reduce circulating ApoC-III levels through promoting its clearance. One strategy for this would be to design monoclonal antibodies that target ApoC-III by binding the dynamic region of the protein harboring the Ala43 residue and promoting dissociation of ApoC-III from lipoproteins to augment renal clearance. Alternatively, antibodies that promote the dissociation of ApoC-III without allowing the renal catabolism of the ApoC-III:antibody complex could still reduce TRLs if the complexes are incapable of inhibiting LPL or remnant particle clearance. Other strategies that could be envisioned involving monoclonal antibodies would include those that promote the ApoC-III:LDLR interaction and thus promote hepatic TRL remnant uptake. However, experimental data supporting the plausibility of such a mechanism does not yet exist. Indeed, the current scope for novel *APOC3*-focused therapies is broad and many are looking to the examples from other genetics discoveries-turned-

therapies, namely the PCSK9 chronicle, for guidance in designing pharmacological approaches here.

From a physiological vantage point, the study of *APOC3* LOF variants has also allowed the potential to address the relative contributions of the different functions of ApoC-III on TRL metabolism. Our study of the A43T variant identified clear consequences of the variant on LPL function *in vivo* and *in vitro*, but it did not rule out the possibility that the variant confers an atheroprotective lipid profile through accelerating remnant particle clearance. The issue of a 'primary-lipolysis' vs. 'primary-clearance' effect of ApoC-III is an age-old debate that started when the *APOC3* transgenic mouse was shown to have 'normal' tissue LPL activity but impaired remnant particle clearance *in vivo* [99]. Since then, several mouse studies have attempted to address the issue of whether ApoC-III does actually compete with ApoE for mediating remnant clearance or if the LPL inhibitory functions are predominant [317,422]. Most recently, use of the mouse anti-*Apoc3* ASO in a series of mouse models lacking different hepatic lipoprotein receptors suggested that promotion of remnant clearance was the primary means by which *APOC3* silencing reduced TGs [333]. This study utilized an inducible LPL KO mouse to further suggest that LPL activity was not required to elicit a TG reduction in the setting of *APOC3* silencing, thus supporting the notion that the remnant clearance effect is more relevant. Clear from all these studies is that the methodologies to assess *in vivo* LPL activity and *in vivo* remnant particle clearance in mouse models are not straightforward and subject to significant inter-user differences in performance. Simultaneous studies of gain- and loss-of-function of ApoC-III, such as through the identification and *in vivo* study of gain-of-function variants, could help in

assessing the relative contributions of the differing functions of ApoC-III in addition to methods standardization.

Finally, the study here of *APOC3* LOF variant carriers illustrates the utility of heterozygotes for LOF variants to study gene function. We were able to assess the impact of variants that putatively reduce *APOC3* expression by 50% on downstream effects on ApoC-III protein levels. Doing so allowed us to remark that half-normal *APOC3* expression may reduce the total circulating pool of ApoC-III by >50%. For the three LOF variants that reduce *APOC3* expression, we observed robust reductions of ApoC-III levels beyond 50% of the expected values from noncarrier controls. I postulate that this may be due to the enhanced clearance of TRLs and remnants that are partially devoid of ApoC-III due to improved lipolysis, more efficient particle remodeling, and possibly more accessibility for key ligands such as ApoE to be retained on the remnants. These properties increase the clearance of the remnants by the liver and thus may account for the very low (often zero) levels of ApoC-III we observed in the heterozygous carriers. This has implications for developing therapies as drugs that even partially reduce ApoC-III levels may still yield benefits by promoting further ApoC-III clearance indirectly by this proposed mechanism. Through collaboration with the Pakistan Risk of Myocardial Infarction Study (PROMIS) led by Dr. Danish Saleheen at Penn, we have been able to identify dozens of *APOC3* R19X homozygotes born in consanguineous unions and have initiated studies of noncarriers, R19X heterozygotes and homozygotes to address questions related to the impact of gene-dosage of *APOC3* on lipoprotein phenotypes, kinetics, and other metabolic processes in addition to evaluating any adverse effects of complete *APOC3* deficiency. These studies will hopefully also shape future pharmacological efforts of ApoC-III lowering.

Resequencing the noncoding genome for rare causal variation

Targeted sequencing of a handful of HDL-C candidate loci through the use of molecular inversion probes was an economical means to interrogate the coding and noncoding regions of selected targets and identified almost 5,000 variants, most of which was novel and in noncoding regions. Appreciation of the value of rare noncoding variants will only increase with the advent of whole-genome sequencing studies, which is currently in its earliest phase. Targeted sequencing of candidate regions will remain a reasonable way to improve coverage of targets of known relevance to a disease or trait process.

In the case of our targeted sequencing, the identified rare, noncoding novel variants we identified will have to be tested for their impact on gene expression through a number of ways. These include assessment of their abilities to impact reporter gene expression, endogenous gene expression in patient-derived cells, and elucidation of the transcriptional machinery that regulates the regions harboring these variants. For promising variants, human carriers could also be recruited for deeper phenotyping related to their extreme HDL-C phenotype. Broadly speaking, the approach we have undertaken here can be applied with the workflow and analysis tools we have generated to large population-based resequencing efforts for noncoding regions or larger phenotypic extreme case-control designs. Such studies are likely to benefit from the combination of other variant interrogation tools that have been used throughout this dissertation work, such as exome chip genotyping, GWAS, and exome sequencing. Through such parallel investigations of different types of variants in the same participants, we may also better answer age-old questions regarding the missing heritability underlying these complex traits from the coding vs. noncoding genome.

Figure 6.1

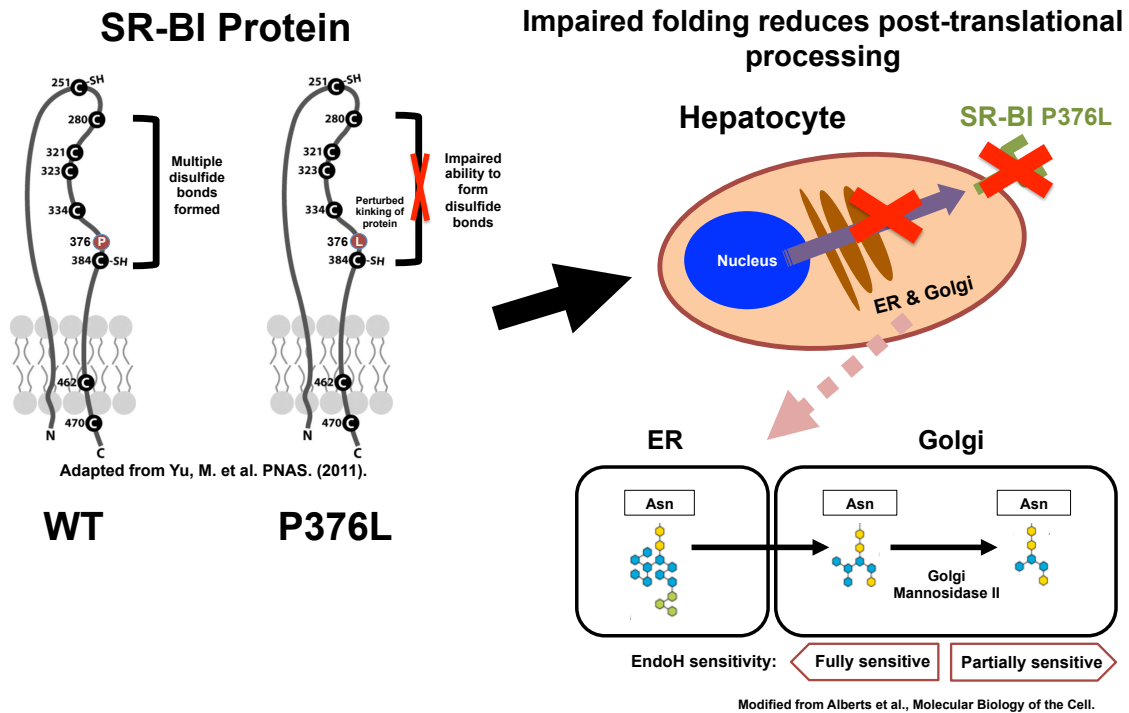


Figure 6.1: Proposed mechanism of disruption of folding of SR-BI due to *SCARB1* P376L variant. On the left is shown a model of the SR-BI protein, which forms a loop with two transmembrane components at the cell surface. Proline 376 spans a region of multiple cysteines and disulfide bonds at the C-terminus of the protein at the extracellular loop. Disruption of the natural fold through the P376L variant may perturb disulfide bond formation, proper folding, and thus post-translational N-glycosylation and mature, endoglycosidase H-resistant forms of the SR-BI normally destined for the cell surface as functional receptors (right). This may result in the observed reduction in cell-surface SR-BI and selective cholesterol uptake from HDLs due to the variant, as discussed in Chapter 2.

Figure 6.2

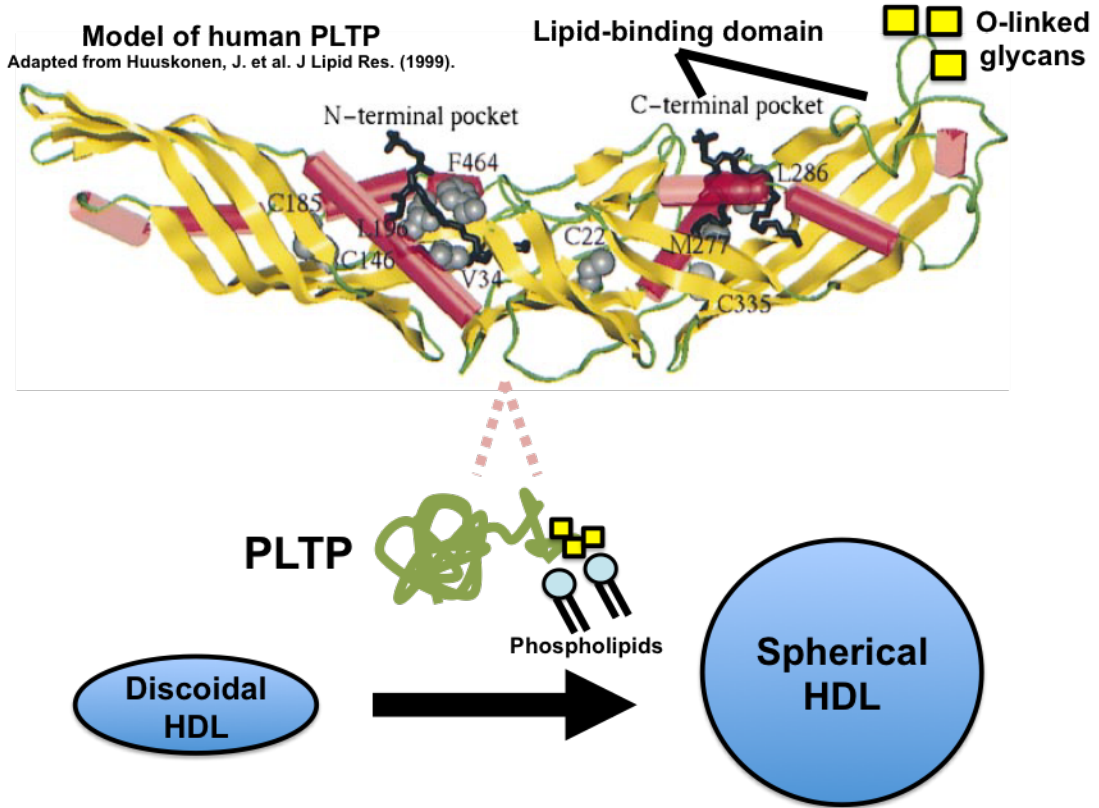


Figure 6.2: Possible role of C-terminal O-glycosylation of PLTP in mediating phospholipid binding capacity and transfer to HDLs. On the top is shown a molecular model for human PLTP, with the C-terminal pocket serving as the lipid binding domain. Multiple residues at the extreme C-terminus in this pocket were determined to be O-glycosylated and isoform-specific targets of GalNAc-T2 in vitro and ex vivo. These residues may impact the phospholipid binding affinity of PLTP and ability to mediate transfers from ApoB-containing lipoproteins to discoidal (immature) HDLs to yield spherical (mature) HDLs. Alternatively, O-glycosylation of PLTP may modulate its ability to associate with HDL proteins and already present on immature particles.

Figure 6.3

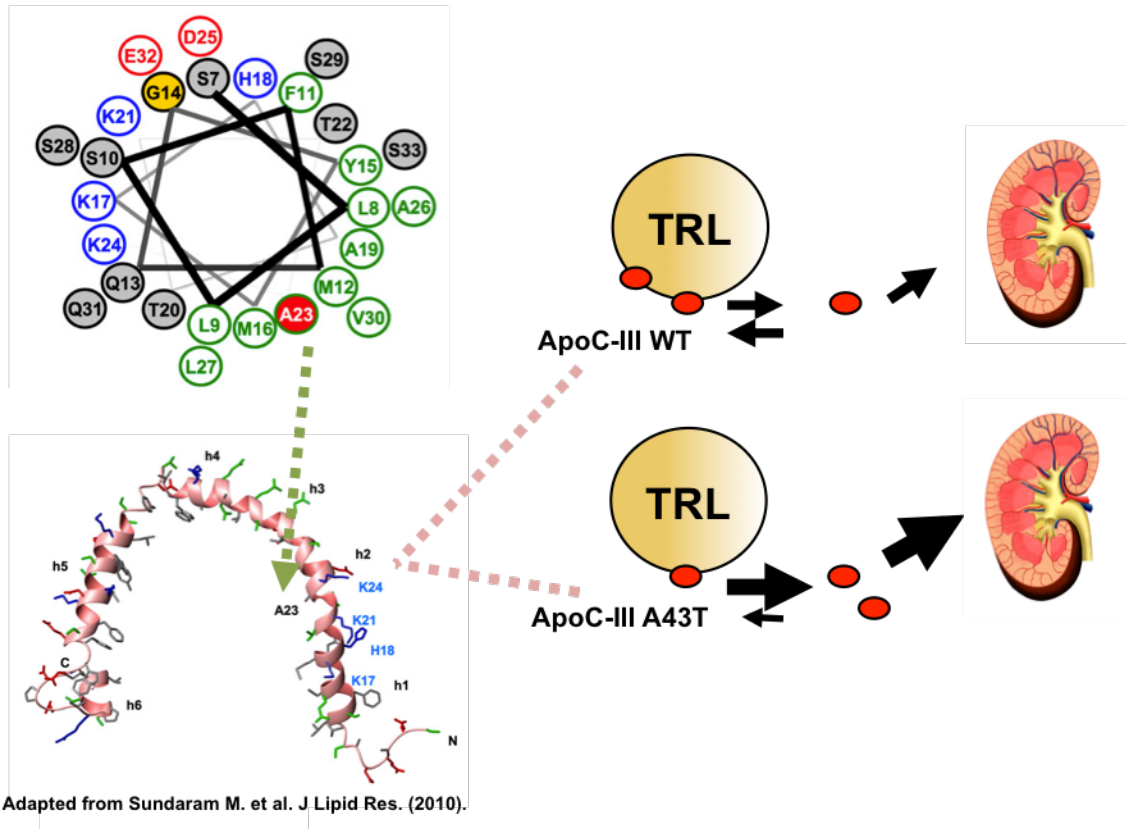


Figure 6.3: Impact of *APOC3* A43T variant on lipid binding affinity of ApoC-III. On the top left is shown a helical wheel diagram for ApoC-III, demonstrating that the alanine 43 residue (residue 23 of the mature protein after signal peptide cleavage) lies in a hydrophobic (lipid-opposing) face of an alpha helix. Below is a model of ApoC-III indicating that this alanine is part of a lipid-facing helix. Our data discussed in Chapter 4 demonstrates a reduced lipid binding affinity for this variant likely due to perturbed interaction with lipid surfaces due to the more polar threonine residue imparted by the variant. This results in reduced binding to lipoprotein surfaces such as that of TRLs, causing dissociation to an unbound state and increased renal clearance (right).

BIBLIOGRAPHY

1. Libby P, Theroux P (2005) Pathophysiology of coronary artery disease. *Circulation* 111: 3481-3488.
2. Lilly LS, Harvard Medical School. (2011) Pathophysiology of heart disease : a collaborative project of medical students and faculty. Baltimore, MD: Wolters Kluwer/Lippincott Williams & Wilkins. xiv, 461 p. p.
3. Herrington W, Lacey B, Sherliker P, Armitage J, Lewington S (2016) Epidemiology of Atherosclerosis and the Potential to Reduce the Global Burden of Atherothrombotic Disease. *Circ Res* 118: 535-546.
4. Mortality GBD, Causes of Death C (2015) Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 385: 117-171.
5. Writing Group M, Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, et al. (2016) Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. *Circulation* 133: e38-360.
6. Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, et al. (2007) Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. *N Engl J Med* 356: 2388-2398.
7. Libby P, Ridker PM, Hansson GK (2011) Progress and challenges in translating the biology of atherosclerosis. *Nature* 473: 317-325.
8. Rader DJ, Daugherty A (2008) Translating molecular discoveries into new therapies for atherosclerosis. *Nature* 451: 904-913.
9. Greaves DR, Gordon S (2009) The macrophage scavenger receptor at 30 years of age: current knowledge and future challenges. *J Lipid Res* 50 Suppl: S282-286.
10. Brown MS, Goldstein JL (1983) Lipoprotein metabolism in the macrophage: implications for cholesterol deposition in atherosclerosis. *Annu Rev Biochem* 52: 223-261.
11. Steinberg D, Witztum JL (2010) Oxidized low-density lipoprotein and atherosclerosis. *Arterioscler Thromb Vasc Biol* 30: 2311-2316.
12. Libby P, Tabas I, Fredman G, Fisher EA (2014) Inflammation and its resolution as determinants of acute coronary syndromes. *Circ Res* 114: 1867-1879.
13. Boudoulas KD, Triposciadis F, Geleris P, Boudoulas H (2016) Coronary Atherosclerosis: Pathophysiologic Basis for Diagnosis and Management. *Prog Cardiovasc Dis* 58: 676-692.
14. Yang Q, Cogswell ME, Flanders WD, Hong Y, Zhang Z, et al. (2012) Trends in cardiovascular health metrics and associations with all-cause and CVD mortality among US adults. *JAMA* 307: 1273-1283.
15. Berry JD, Dyer A, Cai X, Garside DB, Ning H, et al. (2012) Lifetime risks of cardiovascular disease. *N Engl J Med* 366: 321-329.
16. Rosenson RS, Brewer HB, Rader DJ (2014) Lipoproteins as biomarkers and therapeutic targets in the setting of acute coronary syndrome. *Circ Res* 114: 1880-1889.
17. Emerging Risk Factors C, Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, et al. (2009) Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* 302: 1993-2000.

18. Suhre K, Gieger C (2012) Genetic variation in metabolic phenotypes: study designs and applications. *Nat Rev Genet* 13: 759-769.
19. Rader DJ (2016) New Therapeutic Approaches to the Treatment of Dyslipidemia. *Cell Metab* 23: 405-412.
20. Goldstein JL, Brown MS (2015) A century of cholesterol and coronaries: from plaques to genes to statins. *Cell* 161: 161-172.
21. Finking G, Hanke H (1997) Nikolaj Nikolajewitsch Anitschkow (1885-1964) established the cholesterol-fed rabbit as a model for atherosclerosis research. *Atherosclerosis* 135: 1-7.
22. Steinberg D (2005) Thematic review series: the pathogenesis of atherosclerosis. An interpretive history of the cholesterol controversy: part II: the early evidence linking hypercholesterolemia to coronary disease in humans. *J Lipid Res* 46: 179-190.
23. Steinberg D (2005) Thematic review series: the pathogenesis of atherosclerosis: an interpretive history of the cholesterol controversy, part III: mechanistically defining the role of hyperlipidemia. *J Lipid Res* 46: 2037-2051.
24. Brown MS, Goldstein JL (1974) Familial hypercholesterolemia: defective binding of lipoproteins to cultured fibroblasts associated with impaired regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity. *Proc Natl Acad Sci U S A* 71: 788-792.
25. Brown MS, Goldstein JL (1986) A receptor-mediated pathway for cholesterol homeostasis. *Science* 232: 34-47.
26. Gofman JW, Rubin L, Mc GJ, Jones HB (1954) Hyperlipoproteinemia. *Am J Med* 17: 514-520.
27. Gofman JW, Glazier F, Tamplin A, Strisower B, De Lalla O (1954) Lipoproteins, coronary heart disease, and atherosclerosis. *Physiol Rev* 34: 589-607.
28. Endo A, Kuroda M, Tanzawa K (1976) Competitive inhibition of 3-hydroxy-3-methylglutaryl coenzyme A reductase by ML-236A and ML-236B fungal metabolites, having hypocholesterolemic activity. *FEBS Lett* 72: 323-326.
29. Brautbar A, Leary E, Rasmussen K, Wilson DP, Steiner RD, et al. (2015) Genetics of familial hypercholesterolemia. *Curr Atheroscler Rep* 17: 491.
30. Cohen JC, Kimmel M, Polanski A, Hobbs HH (2003) Molecular mechanisms of autosomal recessive hypercholesterolemia. *Curr Opin Lipidol* 14: 121-127.
31. Horton JD, Cohen JC, Hobbs HH (2009) PCSK9: a convertase that coordinates LDL catabolism. *J Lipid Res* 50 Suppl: S172-177.
32. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, et al. (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* 34: 154-156.
33. Cunningham D, Danley DE, Geoghegan KF, Griffor MC, Hawkins JL, et al. (2007) Structural and biophysical studies of PCSK9 and its mutants linked to familial hypercholesterolemia. *Nat Struct Mol Biol* 14: 413-419.
34. Pandit S, Wisniewski D, Santoro JC, Ha S, Ramakrishnan V, et al. (2008) Functional analysis of sites within PCSK9 responsible for hypercholesterolemia. *J Lipid Res* 49: 1333-1343.
35. Naoumova RP, Tosi I, Patel D, Neuwirth C, Horswell SD, et al. (2005) Severe hypercholesterolemia in four British families with the D374Y mutation in the

- PCSK9 gene: long-term follow-up and treatment response. *Arterioscler Thromb Vasc Biol* 25: 2654-2660.
36. Maxwell KN, Breslow JL (2004) Adenoviral-mediated expression of Pcsk9 in mice results in a low-density lipoprotein receptor knockout phenotype. *Proc Natl Acad Sci U S A* 101: 7100-7105.
 37. Park SW, Moon YA, Horton JD (2004) Post-transcriptional regulation of low density lipoprotein receptor protein by proprotein convertase subtilisin/kexin type 9a in mouse liver. *J Biol Chem* 279: 50630-50638.
 38. Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* 354: 1264-1272.
 39. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, et al. (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* 37: 161-165.
 40. Kathiresan S, Myocardial Infarction Genetics C (2008) A PCSK9 missense variant associated with a reduced risk of early-onset myocardial infarction. *N Engl J Med* 358: 2299-2300.
 41. McPherson R, Kavaslar N (2007) Statins for primary prevention of coronary artery disease. *Lancet* 369: 1078; author reply 1079.
 42. Roth EM, McKenney JM, Hanotin C, Asset G, Stein EA (2012) Atorvastatin with or without an antibody to PCSK9 in primary hypercholesterolemia. *N Engl J Med* 367: 1891-1900.
 43. Stein EA, Mellis S, Yancopoulos GD, Stahl N, Logan D, et al. (2012) Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. *N Engl J Med* 366: 1108-1118.
 44. Kastelein JJ, Ginsberg HN, Langslet G, Hovingh GK, Ceska R, et al. (2015) ODYSSEY FH I and FH II: 78 week results with alirocumab treatment in 735 patients with heterozygous familial hypercholesterolaemia. *Eur Heart J* 36: 2996-3003.
 45. Sjouke B, Kusters DM, Kindt I, Besseling J, Defesche JC, et al. (2015) Homozygous autosomal dominant hypercholesterolaemia in the Netherlands: prevalence, genotype-phenotype relationship, and clinical outcome. *Eur Heart J* 36: 560-565.
 46. Raal FJ, Honarpour N, Blom DJ, Hovingh GK, Xu F, et al. (2015) Inhibition of PCSK9 with evolocumab in homozygous familial hypercholesterolaemia (TESLA Part B): a randomised, double-blind, placebo-controlled trial. *Lancet* 385: 341-350.
 47. Raal FJ, Stein EA, Dufour R, Turner T, Civeira F, et al. (2015) PCSK9 inhibition with evolocumab (AMG 145) in heterozygous familial hypercholesterolaemia (RUTHERFORD-2): a randomised, double-blind, placebo-controlled trial. *Lancet* 385: 331-340.
 48. Stroes E, Colquhoun D, Sullivan D, Civeira F, Rosenson RS, et al. (2014) Anti-PCSK9 antibody effectively lowers cholesterol in patients with statin intolerance: the GAUSS-2 randomized, placebo-controlled phase 3 clinical trial of evolocumab. *J Am Coll Cardiol* 63: 2541-2548.
 49. Sabatine MS, Wasserman SM, Stein EA (2015) PCSK9 Inhibitors and Cardiovascular Events. *N Engl J Med* 373: 774-775.

50. Sabatine MS, Giugliano RP, Wiviott SD, Raal FJ, Blom DJ, et al. (2015) Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. *N Engl J Med* 372: 1500-1509.
51. Everett BM, Smith RJ, Hiatt WR (2015) Reducing LDL with PCSK9 Inhibitors--The Clinical Benefit of Lipid Drugs. *N Engl J Med* 373: 1588-1591.
52. Rader DJ (2006) Molecular regulation of HDL metabolism and function: implications for novel therapies. *J Clin Invest* 116: 3090-3100.
53. Rader DJ, Hovingh GK (2014) HDL and cardiovascular disease. *Lancet* 384: 618-625.
54. Kontush A (2014) HDL-mediated mechanisms of protection in cardiovascular disease. *Cardiovasc Res* 103: 341-349.
55. Kannel WB, Dawber TR, Friedman GD, Glennon WE, McNamara PM (1964) Risk Factors in Coronary Heart Disease. An Evaluation of Several Serum Lipids as Predictors of Coronary Heart Disease; the Framingham Study. *Ann Intern Med* 61: 888-899.
56. Assmann G, Cullen P, 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.
57. Rosenson RS, Brewer HB, Jr., Davidson WS, Fayad ZA, Fuster V, et al. (2012) Cholesterol efflux and atheroprotection: advancing the concept of reverse cholesterol transport. *Circulation* 125: 1905-1919.
58. Khera AV, Rader DJ (2010) Future therapeutic directions in reverse cholesterol transport. *Curr Atheroscler Rep* 12: 73-81.
59. Rye KA, Barter PJ (2014) Cardioprotective functions of HDLs. *J Lipid Res* 55: 168-179.
60. Rader DJ, deGoma EM (2012) Approach to the patient with extremely low HDL-cholesterol. *J Clin Endocrinol Metab* 97: 3399-3407.
61. Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJ, et al. (2007) Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med* 357: 2109-2122.
62. Kastelein JJ, van Leuven SI, Burgess L, Evans GW, Kuivenhoven JA, et al. (2007) Effect of torcetrapib on carotid atherosclerosis in familial hypercholesterolemia. *N Engl J Med* 356: 1620-1630.
63. Connelly MA, Parry TJ, Giardino EC, Huang Z, Cheung WM, et al. (2010) Torcetrapib produces endothelial dysfunction independent of cholesteryl ester transfer protein inhibition. *J Cardiovasc Pharmacol* 55: 459-468.
64. Schwartz GG, Olsson AG, Abt M, Ballantyne CM, Barter PJ, et al. (2012) Effects of dalcetrapib in patients with a recent acute coronary syndrome. *N Engl J Med* 367: 2089-2099.
65. Investigators A-H, Boden WE, Probstfield JL, Anderson T, Chaitman BR, et al. (2011) Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy. *N Engl J Med* 365: 2255-2267.
66. Group HTC, Landray MJ, Haynes R, Hopewell JC, Parish S, et al. (2014) Effects of extended-release niacin with laropiprant in high-risk patients. *N Engl J Med* 371: 203-212.
67. Kingwell BA, Chapman MJ, Kontush A, Miller NE (2014) HDL-targeted therapies: progress, failures and future. *Nat Rev Drug Discov* 13: 445-464.

68. Nissen SE, Tsunoda T, Tuzcu EM, Schoenhagen P, Cooper CJ, et al. (2003) Effect of recombinant ApoA-I Milano on coronary atherosclerosis in patients with acute coronary syndromes: a randomized controlled trial. *JAMA* 290: 2292-2300.
69. Tardif JC, Gregoire J, L'Allier PL, Ibrahim R, Lesperance J, et al. (2007) Effects of reconstituted high-density lipoprotein infusions on coronary atherosclerosis: a randomized controlled trial. *JAMA* 297: 1675-1682.
70. Tardif JC, Ballantyne CM, Barter P, Dasseux JL, Fayad ZA, et al. (2014) Effects of the high-density lipoprotein mimetic agent CER-001 on coronary atherosclerosis in patients with acute coronary syndromes: a randomized trial. *Eur Heart J* 35: 3277-3286.
71. Kohli P, Cannon CP (2012) Triglycerides: how much credit do they deserve? *Med Clin North Am* 96: 39-55.
72. Khetarpal SA, Qamar A, Millar JS, Rader DJ (2016) Targeting ApoC-III to Reduce Coronary Disease Risk. *Curr Atheroscler Rep* 18: 54.
73. Johansen CT, Kathiresan S, Hegele RA (2011) Genetic determinants of plasma triglycerides. *J Lipid Res* 52: 189-206.
74. Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR (1977) High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am J Med* 62: 707-714.
75. Group AS, Ginsberg HN, Elam MB, Lovato LC, Crouse JR, 3rd, et al. (2010) Effects of combination lipid therapy in type 2 diabetes mellitus. *N Engl J Med* 362: 1563-1574.
76. Boekholdt SM, Arsenault BJ, Mora S, Pedersen TR, LaRosa JC, et al. (2012) Association of LDL cholesterol, non-HDL cholesterol, and apolipoprotein B levels with risk of cardiovascular events among patients treated with statins: a meta-analysis. *JAMA* 307: 1302-1309.
77. Gagne SE, Larson MG, Pimstone SN, Schaefer EJ, Kastelein JJ, et al. (1999) A common truncation variant of lipoprotein lipase (Ser447X) confers protection against coronary heart disease: the Framingham Offspring Study. *Clin Genet* 55: 450-454.
78. Groenemeijer BE, Hallman MD, Reymer PW, Gagne E, Kuivenhoven JA, et al. (1997) Genetic variant showing a positive interaction with beta-blocking agents with a beneficial influence on lipoprotein lipase activity, HDL cholesterol, and triglyceride levels in coronary artery disease patients. The Ser447-stop substitution in the lipoprotein lipase gene. REGRESS Study Group. *Circulation* 95: 2628-2635.
79. Henderson HE, Kastelein JJ, Zwinderman AH, Gagne E, Jukema JW, et al. (1999) Lipoprotein lipase activity is decreased in a large cohort of patients with coronary artery disease and is associated with changes in lipids and lipoproteins. *J Lipid Res* 40: 735-743.
80. Humphries SE, Nicaud V, Margalef J, Tiret L, Talmud PJ (1998) Lipoprotein lipase gene variation is associated with a paternal history of premature coronary artery disease and fasting and postprandial plasma triglycerides: the European Atherosclerosis Research Study (EARS). *Arterioscler Thromb Vasc Biol* 18: 526-534.

81. Rip J, Nierman MC, Ross CJ, Jukema JW, Hayden MR, et al. (2006) Lipoprotein lipase S447X: a naturally occurring gain-of-function mutation. *Arterioscler Thromb Vasc Biol* 26: 1236-1245.
82. Varbo A, Benn M, Tybjaerg-Hansen A, Jorgensen AB, Frikke-Schmidt R, et al. (2013) Response: Lipoprotein subclass profiling reveals pleiotropy in the genetic variants of lipid risk factors for coronary heart disease: a note on Mendelian randomization studies. *J Am Coll Cardiol* 62: 1908-1909.
83. Wittrup HH, Tybjaerg-Hansen A, Nordestgaard BG (1999) Lipoprotein lipase mutations, plasma lipids and lipoproteins, and risk of ischemic heart disease. A meta-analysis. *Circulation* 99: 2901-2907.
84. Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, et al. (2011) Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARE Project. *PLoS Genet* 7: e1001300.
85. Pearce LR, Atanassova N, Banton MC, Bottomley B, van der Klaauw AA, et al. (2013) KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell* 155: 765-777.
86. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466: 707-713.
87. Triglyceride Coronary Disease Genetics C, Emerging Risk Factors C, Sarwar N, Sandhu MS, Ricketts SL, et al. (2010) Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. *Lancet* 375: 1634-1639.
88. Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, et al. (2010) Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler Thromb Vasc Biol* 30: 2264-2276.
89. Dewey FE, Gusarova V, O'Dushlaine C, Gottesman O, Trejos J, et al. (2016) Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. *N Engl J Med* 374: 1123-1133.
90. Myocardial Infarction G, Investigators CAEC, Stitzel NO, Stirrups KE, Masca NG, et al. (2016) Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N Engl J Med* 374: 1134-1144.
91. Jorgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjaerg-Hansen A (2014) Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med* 371: 32-41.
92. Do R, Stitzel NO, Won HH, Jorgensen AB, Duga S, et al. (2015) Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* 518: 102-106.
93. Ooi EM, Barrett PH, Chan DC, Watts GF (2008) Apolipoprotein C-III: understanding an emerging cardiovascular risk factor. *Clin Sci (Lond)* 114: 611-624.
94. Sundaram M, Yao Z (2012) Intrahepatic role of exchangeable apolipoproteins in lipoprotein assembly and secretion. *Arterioscler Thromb Vasc Biol* 32: 1073-1078.
95. Maeda H, Hashimoto RK, Ogura T, Hiraga S, Uzawa H (1987) Molecular cloning of a human apoC-III variant: Thr 74----Ala 74 mutation prevents O-glycosylation. *J Lipid Res* 28: 1405-1409.
96. Maeda H, Uzawa H, Kamei R (1981) Unusual familial lipoprotein C-III associated with apolipoprotein C-III-O preponderance. *Biochim Biophys Acta* 665: 578-585.

97. Roghani A, Zannis VI (1988) Mutagenesis of the glycosylation site of human ApoCIII. O-linked glycosylation is not required for ApoCIII secretion and lipid binding. *J Biol Chem* 263: 17925-17932.
98. Eisenberg S, Patsch JR, Sparrow JT, Gotto AM, Olivecrona T (1979) Very low density lipoprotein. Removal of Apolipoproteins C-II and C-III-1 during lipolysis in vitro. *J Biol Chem* 254: 12603-12608.
99. Aalto-Setälä K, Fisher EA, Chen X, Chajek-Shaul T, Hayek T, et al. (1992) Mechanism of hypertriglyceridemia in human apolipoprotein (apo) CIII transgenic mice. Diminished very low density lipoprotein fractional catabolic rate associated with increased apo CIII and reduced apo E on the particles. *J Clin Invest* 90: 1889-1900.
100. Aalto-Setälä K, Weinstock PH, Bisgaier CL, Wu L, Smith JD, et al. (1996) Further characterization of the metabolic properties of triglyceride-rich lipoproteins from human and mouse apoC-III transgenic mice. *J Lipid Res* 37: 1802-1811.
101. Qin W, Sundaram M, Wang Y, Zhou H, Zhong S, et al. (2011) Missense mutation in APOC3 within the C-terminal lipid binding domain of human ApoC-III results in impaired assembly and secretion of triacylglycerol-rich very low density lipoproteins: evidence that ApoC-III plays a major role in the formation of lipid precursors within the microsomal lumen. *J Biol Chem* 286: 27769-27780.
102. Gaudet D, Brisson D, Tremblay K, Alexander VJ, Singleton W, et al. (2014) Targeting APOC3 in the familial chylomicronemia syndrome. *N Engl J Med* 371: 2200-2206.
103. Graham MJ, Lee RG, Bell TA, 3rd, Fu W, Mullick AE, et al. (2013) Antisense oligonucleotide inhibition of apolipoprotein C-III reduces plasma triglycerides in rodents, nonhuman primates, and humans. *Circ Res* 112: 1479-1490.
104. Gaudet D, Alexander VJ, Baker BF, Brisson D, Tremblay K, et al. (2015) Antisense Inhibition of Apolipoprotein C-III in Patients with Hypertriglyceridemia. *N Engl J Med* 373: 438-447.
105. Pollin TI, Damcott CM, Shen H, Ott SH, Shelton J, et al. (2008) A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science* 322: 1702-1705.
106. Tg, Hdl Working Group of the Exome Sequencing Project NHL, Blood I, Crosby J, Peloso GM, et al. (2014) Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med* 371: 22-31.
107. Sharma V, Forte TM, Ryan RO (2013) Influence of apolipoprotein A-V on the metabolic fate of triacylglycerol. *Curr Opin Lipidol* 24: 153-159.
108. Pennacchio LA, Olivier M, Hubacek JA, Cohen JC, Cox DR, et al. (2001) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* 294: 169-173.
109. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, et al. (2010) Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 42: 684-687.
110. Kersten S (2014) Physiological regulation of lipoprotein lipase. *Biochim Biophys Acta* 1841: 919-933.
111. Lee EC, Desai U, Gololobov G, Hong S, Feng X, et al. (2009) Identification of a new functional domain in angiopoietin-like 3 (ANGPTL3) and angiopoietin-like 4

- (ANGPTL4) involved in binding and inhibition of lipoprotein lipase (LPL). *J Biol Chem* 284: 13735-13745.
112. Yau MH, Wang Y, Lam KS, Zhang J, Wu D, et al. (2009) A highly conserved motif within the NH₂-terminal coiled-coil domain of angiopoietin-like protein 4 confers its inhibitory effects on lipoprotein lipase by disrupting the enzyme dimerization. *J Biol Chem* 284: 11942-11952.
 113. Larsson M, Vorrso E, Talmud P, Lookene A, Olivecrona G (2013) Apolipoproteins C-I and C-III inhibit lipoprotein lipase activity by displacement of the enzyme from lipid droplets. *J Biol Chem* 288: 33997-34008.
 114. Yin W, Romeo S, Chang S, Grishin NV, Hobbs HH, et al. (2009) Genetic variation in ANGPTL4 provides insights into protein processing and function. *J Biol Chem* 284: 13213-13222.
 115. Miida T, Hirayama S (2010) Impacts of angiopoietin-like proteins on lipoprotein metabolism and cardiovascular events. *Curr Opin Lipidol* 21: 70-75.
 116. Romeo S, Yin W, Kozlitina J, Pennacchio LA, Boerwinkle E, et al. (2009) Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. *J Clin Invest* 119: 70-79.
 117. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, et al. (2013) Discovery and refinement of loci associated with lipid levels. *Nat Genet* 45: 1274-1283.
 118. Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, et al. (2010) Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med* 363: 2220-2227.
 119. Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, et al. (2007) Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* 39: 513-516.
 120. Talmud PJ, Smart M, Presswood E, Cooper JA, Nicaud V, et al. (2008) ANGPTL4 E40K and T266M: effects on plasma triglyceride and HDL levels, postprandial responses, and CHD risk. *Arterioscler Thromb Vasc Biol* 28: 2319-2325.
 121. Hatsuda S, Shoji T, Shinohara K, Kimoto E, Mori K, et al. (2007) Association between plasma angiopoietin-like protein 3 and arterial wall thickness in healthy subjects. *J Vasc Res* 44: 61-66.
 122. Korstanje R, Eriksson P, Samnegard A, Olsson PG, Forsman-Semb K, et al. (2004) Locating Ath8, a locus for murine atherosclerosis susceptibility and testing several of its candidate genes in mice and humans. *Atherosclerosis* 177: 443-450.
 123. Folsom AR, Peacock JM, Demerath E, Boerwinkle E (2008) Variation in ANGPTL4 and risk of coronary heart disease: the Atherosclerosis Risk in Communities Study. *Metabolism* 57: 1591-1596.
 124. Nordestgaard BG (2016) Triglyceride-Rich Lipoproteins and Atherosclerotic Cardiovascular Disease: New Insights From Epidemiology, Genetics, and Biology. *Circ Res* 118: 547-563.
 125. Womack JE, Jang HJ, Lee MO (2012) Genomics of complex traits. *Ann N Y Acad Sci* 1271: 33-36.
 126. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9: 356-369.

127. McCarthy MI, Hirschhorn JN (2008) Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 17: R156-165.
128. McCarthy MI, Hirschhorn JN (2008) Genome-wide association studies: past, present and future. *Hum Mol Genet* 17: R100-101.
129. Peters DT, Musunuru K (2012) Functional evaluation of genetic variation in complex human traits. *Hum Mol Genet* 21: R18-23.
130. Willer CJ, Mohlke KL (2012) Finding genes and variants for lipid levels after genome-wide association analysis. *Curr Opin Lipidol* 23: 98-103.
131. Kathiresan S, Melander O, Guiducci C, Surti A, Burt NP, 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.
132. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, et al. (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 40: 161-169.
133. Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, 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.
134. Pollin TI, Jablonski KA, McAteer JB, Saxena R, Kathiresan S, et al. (2011) Triglyceride response to an intensive lifestyle intervention is enhanced in carriers of the GCKR Pro446Leu polymorphism. *J Clin Endocrinol Metab* 96: E1142-1147.
135. Rees MG, Ng D, Ruppert S, Turner C, Beer NL, et al. (2012) Correlation of rare coding variants in the gene encoding human glucokinase regulatory protein with phenotypic, cellular, and kinetic outcomes. *J Clin Invest* 122: 205-217.
136. Kooner JS, Chambers JC, Aguilar-Salinas CA, Hinds DA, Hyde CL, et al. (2008) Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet* 40: 149-151.
137. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, et al. (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41: 56-65.
138. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, et al. (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet* 7: e1002003.
139. Sanna S, Li B, Mulas A, Sidore C, Kang HM, et al. (2011) Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS Genet* 7: e1002198.
140. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, et al. (2010) From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* 466: 714-719.
141. Keebler ME, Deo RC, Surti A, Konieczkowski D, Guiducci C, et al. (2010) Fine-mapping in African Americans of 8 recently discovered genetic loci for plasma lipids: the Jackson Heart Study. *Circ Cardiovasc Genet* 3: 358-364.
142. McKenzie CA, Abecasis GR, Keavney B, Forrester T, Ratcliffe PJ, et al. (2001) Trans-ethnic fine mapping of a quantitative trait locus for circulating angiotensin I-converting enzyme (ACE). *Hum Mol Genet* 10: 1077-1084.

143. Deo RC, Reich D, Tandon A, Akylbekova E, Patterson N, et al. (2009) Genetic differences between the determinants of lipid profile phenotypes in African and European Americans: the Jackson Heart Study. *PLoS Genet* 5: e1000342.
144. Consortium EP (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 9: e1001046.
145. Consortium EP, Bernstein BE, Birney E, Dunham I, Green ED, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57-74.
146. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, et al. (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature* 489: 91-100.
147. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, et al. (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337: 1190-1195.
148. Trynka G, Raychaudhuri S (2013) Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases. *Curr Opin Genet Dev* 23: 635-641.
149. Trynka G, Sandor C, Han B, Xu H, Stranger BE, et al. (2013) Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* 45: 124-130.
150. Consortium GT (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348: 648-660.
151. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, et al. (2015) Human genomics. The human transcriptome across tissues and individuals. *Science* 348: 660-665.
152. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42: 579-589.
153. Small KS, Hedman AK, Grundberg E, Nica AC, Thorleifsson G, et al. (2011) Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet* 43: 561-564.
154. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, et al. (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 45: 1238-1243.
155. Bartz F, Kern L, Erz D, Zhu M, Gilbert D, et al. (2009) Identification of cholesterol-regulating genes by targeted RNAi screening. *Cell Metab* 10: 63-75.
156. Blattmann P, Schuberth C, Pepperkok R, Runz H (2013) RNAi-based functional profiling of loci from blood lipid genome-wide association studies identifies genes with cholesterol-regulatory function. *PLoS Genet* 9: e1003338.
157. Ding Q, Lee YK, Schaefer EA, Peters DT, Veres A, et al. (2013) A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* 12: 238-251.
158. Ding Q, Regan SN, Xia Y, Ostrom LA, Cowan CA, et al. (2013) Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* 12: 393-394.

159. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, et al. (2007) Genomewide association analysis of coronary artery disease. *N Engl J Med* 357: 443-453.
160. Pan M, Maitin V, Parathath S, Andreo U, Lin SX, et al. (2008) Presecretory oxidation, aggregation, and autophagic destruction of apoprotein-B: a pathway for late-stage quality control. *Proc Natl Acad Sci U S A* 105: 5862-5867.
161. Kjolby M, Andersen OM, Breiderhoff T, Fjorback AW, Pedersen KM, et al. (2010) Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. *Cell Metab* 12: 213-223.
162. Hu F, Padukkavidana T, Vaegter CB, Brady OA, Zheng Y, et al. (2010) Sortilin-mediated endocytosis determines levels of the frontotemporal dementia protein, progranulin. *Neuron* 68: 654-667.
163. Jansen P, Giehl K, Nyengaard JR, Teng K, Lioubinski O, et al. (2007) Roles for the pro-neurotrophin receptor sortilin in neuronal development, aging and brain injury. *Nat Neurosci* 10: 1449-1457.
164. Zeng J, Racicott J, Morales CR (2009) The inactivation of the sortilin gene leads to a partial disruption of prosaposin trafficking to the lysosomes. *Exp Cell Res* 315: 3112-3124.
165. Strong A, Ding Q, Edmondson AC, Millar JS, Sachs KV, et al. (2012) Hepatic sortilin regulates both apolipoprotein B secretion and LDL catabolism. *J Clin Invest* 122: 2807-2816.
166. Strong A, Rader DJ (2012) Sortilin as a regulator of lipoprotein metabolism. *Curr Atheroscler Rep* 14: 211-218.
167. Gustafsen C, Kjolby M, Nyegaard M, Mattheisen M, Lundhede J, et al. (2014) The hypercholesterolemia-risk gene SORT1 facilitates PCSK9 secretion. *Cell Metab* 19: 310-318.
168. Patel KM, Strong A, Tohyama J, Jin X, Morales CR, et al. (2015) Macrophage Sortilin Promotes LDL Uptake, Foam Cell Formation, and Atherosclerosis. *Circ Res* 116: 789-796.
169. Goettsch C, Hutcheson JD, Aikawa M, Iwata H, Pham T, et al. (2016) Sortilin mediates vascular calcification via its recruitment into extracellular vesicles. *J Clin Invest* 126: 1323-1336.
170. Mortensen MB, Kjolby M, Gunnarsen S, Larsen JV, Palmfeldt J, et al. (2014) Targeting sortilin in immune cells reduces proinflammatory cytokines and atherosclerosis. *J Clin Invest* 124: 5317-5322.
171. Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, et al. (2013) Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* 45: 25-33.
172. Consortium IKC (2011) Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet* 7: e1002260.
173. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, et al. (2011) Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* 43: 1131-1138.
174. Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, et al. (2012) Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet* 8: e1002607.

175. Angyal A, Kiss-Toth E (2012) The tribbles gene family and lipoprotein metabolism. *Curr Opin Lipidol* 23: 122-126.
176. Bauer RC, Stylianou IM, Rader DJ (2011) Functional validation of new pathways in lipoprotein metabolism identified by human genetics. *Curr Opin Lipidol* 22: 123-128.
177. Burkhardt R, Toh SA, Lagor WR, Birkeland A, Levin M, 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.
178. Bauer RC, Sasaki M, Cohen DM, Cui J, Smith MA, et al. (2015) Tribbles-1 regulates hepatic lipogenesis through posttranscriptional regulation of C/EBPalpha. *J Clin Invest* 125: 3809-3818.
179. Satoh T, Kidoya H, Naito H, Yamamoto M, Takemura N, et al. (2013) Critical role of Trib1 in differentiation of tissue-resident M2-like macrophages. *Nature* 495: 524-528.
180. Hegele RA, Ban MR, Hsueh N, Kennedy BA, Cao H, et al. (2009) A polygenic basis for four classical Fredrickson hyperlipoproteinemia phenotypes that are characterized by hypertriglyceridemia. *Hum Mol Genet* 18: 4189-4194.
181. Pendergrass SA, Brown-Gentry K, Dudek S, Frase A, Torstenson ES, et al. (2013) Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. *PLoS Genet* 9: e1003087.
182. Folkersen L, van't Hooft F, Chernogubova E, Agardh HE, Hansson GK, et al. (2010) Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. *Circ Cardiovasc Genet* 3: 365-373.
183. Roman TS, Marvelle AF, Fogarty MP, Vadlamudi S, Gonzalez AJ, et al. (2015) Multiple Hepatic Regulatory Variants at the GALNT2 GWAS Locus Associated with High-Density Lipoprotein Cholesterol. *Am J Hum Genet* 97: 801-815.
184. Holleboom AG, Karlsson H, Lin RS, Beres TM, Sierts JA, et al. (2011) Heterozygosity for a loss-of-function mutation in GALNT2 improves plasma triglyceride clearance in man. *Cell Metab* 14: 811-818.
185. Schjoldager KT, Vakhrushev SY, Kong Y, Steentoft C, Nudelman AS, et al. (2012) Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proc Natl Acad Sci U S A* 109: 9893-9898.
186. Schjoldager KT, Vester-Christensen MB, Bennett EP, Lavery SB, Schwientek T, et al. (2010) O-glycosylation modulates proprotein convertase activation of angiopoietin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids. *J Biol Chem* 285: 36293-36303.
187. Peloso GM, Timofeev N, Lunetta KL (2009) Principal-component-based population structure adjustment in the North American Rheumatoid Arthritis Consortium data: impact of single-nucleotide polymorphism set and analysis method. *BMC Proc* 3 Suppl 7: S108.
188. Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, et al. (2014) Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet* 94: 223-232.

189. Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, et al. (2014) Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat Genet* 46: 345-351.
190. Smith GD, Timpson N, Ebrahim S (2008) Strengthening causal inference in cardiovascular epidemiology through Mendelian randomization. *Ann Med* 40: 524-541.
191. Thanassoulis G, O'Donnell CJ (2009) Mendelian randomization: nature's randomized trial in the post-genome era. *JAMA* 301: 2386-2388.
192. Schunkert H, Samani NJ (2008) Elevated C-reactive protein in atherosclerosis--chicken or egg? *N Engl J Med* 359: 1953-1955.
193. Frikke-Schmidt R, Nordestgaard BG, Stene MC, Sethi AA, Remaley AT, et al. (2008) Association of loss-of-function mutations in the ABCA1 gene with high-density lipoprotein cholesterol levels and risk of ischemic heart disease. *JAMA* 299: 2524-2532.
194. Haase CL, Tybjaerg-Hansen A, Qayyum AA, Schou J, Nordestgaard BG, et al. (2012) LCAT, HDL cholesterol and ischemic cardiovascular disease: a Mendelian randomization study of HDL cholesterol in 54,500 individuals. *J Clin Endocrinol Metab* 97: E248-256.
195. Johannsen TH, Kamstrup PR, Andersen RV, Jensen GB, Sillesen H, et al. (2009) Hepatic lipase, genetically elevated high-density lipoprotein, and risk of ischemic cardiovascular disease. *J Clin Endocrinol Metab* 94: 1264-1273.
196. Ridker PM, Pare G, Parker AN, Zee RY, Miletich JP, et al. (2009) Polymorphism in the CETP gene region, HDL cholesterol, and risk of future myocardial infarction: Genomewide analysis among 18 245 initially healthy women from the Women's Genome Health Study. *Circ Cardiovasc Genet* 2: 26-33.
197. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, et al. (2013) Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet* 45: 1345-1352.
198. Motazacker MM, Peter J, Treskes M, Shoulders CC, Kuivenhoven JA, et al. (2013) Evidence of a polygenic origin of extreme high-density lipoprotein cholesterol levels. *Arterioscler Thromb Vasc Biol* 33: 1521-1528.
199. Khera AV, Won HH, Peloso GM, Lawson KS, Bartz TM, et al. (2016) Diagnostic Yield and Clinical Utility of Sequencing Familial Hypercholesterolemia Genes in Patients With Severe Hypercholesterolemia. *J Am Coll Cardiol* 67: 2578-2589.
200. Stitzel NO, Fouchier SW, Sjouke B, Peloso GM, Moscoso AM, et al. (2013) Exome sequencing and directed clinical phenotyping diagnose cholesterol ester storage disease presenting as autosomal recessive hypercholesterolemia. *Arterioscler Thromb Vasc Biol* 33: 2909-2914.
201. Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, et al. (2014) Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet* 94: 233-245.
202. Myocardial Infarction Genetics Consortium I, Stitzel NO, Won HH, Morrison AC, Peloso GM, et al. (2014) Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med* 371: 2072-2082.

203. Cannon CP, Blazing MA, Giugliano RP, McCagg A, White JA, et al. (2015) Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes. *N Engl J Med* 372: 2387-2397.
204. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, et al. (2012) Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 380: 572-580.
205. Agerholm-Larsen B, Nordestgaard BG, Steffensen R, Jensen G, Tybjaerg-Hansen A (2000) Elevated HDL cholesterol is a risk factor for ischemic heart disease in white women when caused by a common mutation in the cholesteryl ester transfer protein gene. *Circulation* 101: 1907-1912.
206. Andersen RV, Wittrup HH, Tybjaerg-Hansen A, Steffensen R, Schnohr P, et al. (2003) Hepatic lipase mutations, elevated high-density lipoprotein cholesterol, and increased risk of ischemic heart disease: the Copenhagen City Heart Study. *J Am Coll Cardiol* 41: 1972-1982.
207. Acton S, Rigotti A, Landschulz KT, Xu S, Hobbs HH, et al. (1996) Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. *Science* 271: 518-520.
208. Hoekstra M, Van Berkel TJ, Van Eck M (2010) Scavenger receptor BI: a multi-purpose player in cholesterol and steroid metabolism. *World J Gastroenterol* 16: 5916-5924.
209. Wang N, Arai T, Ji Y, Rinninger F, Tall AR (1998) Liver-specific overexpression of scavenger receptor BI decreases levels of very low density lipoprotein ApoB, low density lipoprotein ApoB, and high density lipoprotein in transgenic mice. *J Biol Chem* 273: 32920-32926.
210. Ueda Y, Royer L, Gong E, Zhang J, Cooper PN, et al. (1999) Lower plasma levels and accelerated clearance of high density lipoprotein (HDL) and non-HDL cholesterol in scavenger receptor class B type I transgenic mice. *J Biol Chem* 274: 7165-7171.
211. Kozarsky KF, Donahee MH, Rigotti A, Iqbal SN, Edelman ER, et al. (1997) Overexpression of the HDL receptor SR-BI alters plasma HDL and bile cholesterol levels. *Nature* 387: 414-417.
212. Ji Y, Wang N, Ramakrishnan R, Sehayek E, Huszar D, et al. (1999) Hepatic scavenger receptor BI promotes rapid clearance of high density lipoprotein free cholesterol and its transport into bile. *J Biol Chem* 274: 33398-33402.
213. Varban ML, Rinninger F, Wang N, Fairchild-Huntress V, Dunmore JH, et al. (1998) Targeted mutation reveals a central role for SR-BI in hepatic selective uptake of high density lipoprotein cholesterol. *Proc Natl Acad Sci U S A* 95: 4619-4624.
214. Rigotti A, Trigatti BL, Penman M, Rayburn H, Herz J, et al. (1997) A targeted mutation in the murine gene encoding the high density lipoprotein (HDL) receptor scavenger receptor class B type I reveals its key role in HDL metabolism. *Proc Natl Acad Sci U S A* 94: 12610-12615.
215. Brundert M, Ewert A, Heeren J, Behrendt B, Ramakrishnan R, et al. (2005) Scavenger receptor class B type I mediates the selective uptake of high-density lipoprotein-associated cholesteryl ester by the liver in mice. *Arterioscler Thromb Vasc Biol* 25: 143-148.

216. Ueda Y, Gong E, Royer L, Cooper PN, Francone OL, et al. (2000) Relationship between expression levels and atherogenesis in scavenger receptor class B, type I transgenics. *J Biol Chem* 275: 20368-20373.
217. Kozarsky KF, Donahee MH, Glick JM, Krieger M, Rader DJ (2000) Gene transfer and hepatic overexpression of the HDL receptor SR-BI reduces atherosclerosis in the cholesterol-fed LDL receptor-deficient mouse. *Arterioscler Thromb Vasc Biol* 20: 721-727.
218. Arai T, Wang N, Bezouevski M, Welch C, Tall AR (1999) Decreased atherosclerosis in heterozygous low density lipoprotein receptor-deficient mice expressing the scavenger receptor BI transgene. *J Biol Chem* 274: 2366-2371.
219. Van Eck M, Twisk J, Hoekstra M, Van Rij BT, Van der Lans CA, et al. (2003) Differential effects of scavenger receptor BI deficiency on lipid metabolism in cells of the arterial wall and in the liver. *J Biol Chem* 278: 23699-23705.
220. Trigatti B, Rayburn H, Vinals M, Braun A, Miettinen H, et al. (1999) Influence of the high density lipoprotein receptor SR-BI on reproductive and cardiovascular pathophysiology. *Proc Natl Acad Sci U S A* 96: 9322-9327.
221. Huszar D, Varban ML, Rinninger F, Feeley R, Arai T, et al. (2000) Increased LDL cholesterol and atherosclerosis in LDL receptor-deficient mice with attenuated expression of scavenger receptor B1. *Arterioscler Thromb Vasc Biol* 20: 1068-1073.
222. Braun A, Trigatti BL, Post MJ, Sato K, Simons M, et al. (2002) Loss of SR-BI expression leads to the early onset of occlusive atherosclerotic coronary artery disease, spontaneous myocardial infarctions, severe cardiac dysfunction, and premature death in apolipoprotein E-deficient mice. *Circ Res* 90: 270-276.
223. Zhang Y, Da Silva JR, Reilly M, Billheimer JT, Rothblat GH, et al. (2005) Hepatic expression of scavenger receptor class B type I (SR-BI) is a positive regulator of macrophage reverse cholesterol transport in vivo. *J Clin Invest* 115: 2870-2874.
224. Vergeer M, Korpelaar SJ, Franssen R, Meurs I, Out R, et al. (2011) Genetic variant of the scavenger receptor BI in humans. *N Engl J Med* 364: 136-145.
225. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, et al. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27: 182-189.
226. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6: 80-92.
227. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-5467.
228. Gonzalez-Perez A, Lopez-Bigas N (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* 88: 440-449.
229. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7: 248-249.
230. Kallberg M, Wang H, Wang S, Peng J, Wang Z, et al. (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7: 1511-1522.

231. Wang Z, Zhao F, Peng J, Xu J (2011) Protein 8-class secondary structure prediction using conditional neural fields. *Proteomics* 11: 3786-3792.
232. Somanathan S, Jacobs F, Wang Q, Hanlon AL, Wilson JM, et al. (2014) AAV vectors expressing LDLR gain-of-function variants demonstrate increased efficacy in mouse models of familial hypercholesterolemia. *Circ Res* 115: 591-599.
233. Chen SJ, Sanmiguel J, Lock M, McMennamin D, Draper C, et al. (2013) Biodistribution of AAV8 vectors expressing human low-density lipoprotein receptor in a mouse model of homozygous familial hypercholesterolemia. *Hum Gene Ther Clin Dev* 24: 154-160.
234. Kassim SH, Li H, Bell P, Somanathan S, Lagor W, et al. (2013) Adeno-associated virus serotype 8 gene therapy leads to significant lowering of plasma cholesterol levels in humanized mouse models of homozygous and heterozygous familial hypercholesterolemia. *Hum Gene Ther* 24: 19-26.
235. Kassim SH, Li H, Vandenberghe LH, Hinderer C, Bell P, et al. (2010) Gene therapy in a humanized mouse model of familial hypercholesterolemia leads to marked regression of atherosclerosis. *PLoS One* 5: e13424.
236. Kitajima K, Marchadier DH, Miller GC, Gao GP, Wilson JM, et al. (2006) Complete prevention of atherosclerosis in apoE-deficient mice by hepatic human apoE gene transfer with adeno-associated virus serotypes 7 and 8. *Arterioscler Thromb Vasc Biol* 26: 1852-1857.
237. Chen SJ, Rader DJ, Tazelaar J, Kawashiri M, Gao G, et al. (2000) Prolonged correction of hyperlipidemia in mice with familial hypercholesterolemia using an adeno-associated viral vector expressing very-low-density lipoprotein receptor. *Mol Ther* 2: 256-261.
238. Hatch FT (1968) Practical methods for plasma lipoprotein analysis. *Adv Lipid Res* 6: 1-68.
239. Tietge UJ, Maugeais C, Cain W, Grass D, Glick JM, et al. (2000) Overexpression of secretory phospholipase A(2) causes rapid catabolism and altered tissue uptake of high density lipoprotein cholesteryl ester and apolipoprotein A-I. *J Biol Chem* 275: 10077-10084.
240. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. *J Biol Chem* 193: 265-275.
241. Azhar S, Tsai L, Reaven E (1990) Uptake and utilization of lipoprotein cholesteryl esters by rat granulosa cells. *Biochim Biophys Acta* 1047: 148-160.
242. Connelly MA, Klein SM, Azhar S, Abumrad NA, Williams DL (1999) Comparison of class B scavenger receptors, CD36 and scavenger receptor BI (SR-BI), shows that both receptors mediate high density lipoprotein-cholesteryl ester selective uptake but SR-BI exhibits a unique enhancement of cholesteryl ester uptake. *J Biol Chem* 274: 41-47.
243. Nieland TJ, Xu S, Penman M, Krieger M (2011) Negatively cooperative binding of high-density lipoprotein to the HDL receptor SR-BI. *Biochemistry* 50: 1818-1830.
244. Yang W, Mills JA, Sullivan S, Liu Y, French DL, et al. (2008) iPSC Reprogramming from Human Peripheral Blood Using Sendai Virus Mediated Gene Transfer. *StemBook*. Cambridge (MA).

245. Cai J, DeLaForest A, Fisher J, Urick A, Wagner T, et al. (2008) Protocol for directed differentiation of human pluripotent stem cells toward a hepatocyte fate. *StemBook*. Cambridge (MA).
246. Tietge UJ, Maugeais C, Cain W, Rader DJ (2003) Acute inflammation increases selective uptake of HDL cholesteryl esters into adrenals of mice overexpressing human sPLA2. *Am J Physiol Endocrinol Metab* 285: E403-411.
247. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9: 671-675.
248. Cuchel M, Bloedon LT, Szapary PO, Kolansky DM, Wolfe ML, et al. (2007) Inhibition of microsomal triglyceride transfer protein in familial hypercholesterolemia. *N Engl J Med* 356: 148-156.
249. Gerdes LU, Gerdes C, Klausen IC, Faergeman O (1992) Generation of analytic plasma lipoprotein profiles using two prepacked superose 6B columns. *Clin Chim Acta* 205: 1-9.
250. Nambi V, Chambless L, Folsom AR, He M, Hu Y, et al. (2010) Carotid intima-media thickness and presence or absence of plaque improves prediction of coronary heart disease risk: the ARIC (Atherosclerosis Risk In Communities) study. *J Am Coll Cardiol* 55: 1600-1607.
251. Stein JH, Korcarz CE, Hurst RT, Lonn E, Kendall CB, et al. (2008) Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force. Endorsed by the Society for Vascular Medicine. *J Am Soc Echocardiogr* 21: 93-111; quiz 189-190.
252. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, et al. (2009) Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 42: 377-381.
253. Chapman MJ, Goldstein S, Lagrange D, Laplaud PM (1981) A density gradient ultracentrifugal procedure for the isolation of the major lipoprotein classes from human serum. *J Lipid Res* 22: 339-358.
254. Guerin M, Lassel TS, Le Goff W, Farnier M, Chapman MJ (2000) Action of atorvastatin in combined hyperlipidemia : preferential reduction of cholesteryl ester transfer from HDL to VLDL1 particles. *Arterioscler Thromb Vasc Biol* 20: 189-197.
255. Camont L, Lhomme M, Rached F, Le Goff W, Negre-Salvayre A, et al. (2013) Small, dense high-density lipoprotein-3 particles are enriched in negatively charged phospholipids: relevance to cellular cholesterol efflux, antioxidative, antithrombotic, anti-inflammatory, and antiapoptotic functionalities. *Arterioscler Thromb Vasc Biol* 33: 2715-2723.
256. Babitt J, Trigatti B, Rigotti A, Smart EJ, Anderson RG, et al. (1997) Murine SR-BI, a high density lipoprotein receptor that mediates selective lipid uptake, is N-glycosylated and fatty acylated and colocalizes with plasma membrane caveolae. *J Biol Chem* 272: 13242-13249.
257. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, et al. (2014) Meta-analysis of gene-level tests for rare variant association. *Nat Genet* 46: 200-204.

258. Neculai D, Schwake M, Ravichandran M, Zunke F, Collins RF, et al. (2013) Structure of LIMP-2 provides functional insights with implications for SR-BI and CD36. *Nature* 504: 172-176.
259. Ghodsizadeh A, Taei A, Totonchi M, Seifinejad A, Gourabi H, et al. (2010) Generation of liver disease-specific induced pluripotent stem cells along with efficient differentiation to functional hepatocyte-like cells. *Stem Cell Rev* 6: 622-632.
260. Cayo MA, Cai J, DeLaForest A, Noto FK, Nagaoka M, et al. (2012) JD induced pluripotent stem cell-derived hepatocytes faithfully recapitulate the pathophysiology of familial hypercholesterolemia. *Hepatology* 56: 2163-2171.
261. Mallanna SK, Duncan SA (2013) Differentiation of hepatocytes from pluripotent stem cells. *Curr Protoc Stem Cell Biol* 26: Unit 1G 4.
262. Si-Tayeb K, Noto FK, Nagaoka M, Li J, Battle MA, et al. (2010) Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* 51: 297-305.
263. Vinals M, Xu S, Vasile E, Krieger M (2003) Identification of the N-linked glycosylation sites on the high density lipoprotein (HDL) receptor SR-BI and assessment of their effects on HDL binding and selective lipid uptake. *J Biol Chem* 278: 5325-5332.
264. Calvo D, Gomez-Coronado D, Lasuncion MA, Vega MA (1997) CLA-1 is an 85-kD plasma membrane glycoprotein that acts as a high-affinity receptor for both native (HDL, LDL, and VLDL) and modified (OxLDL and AcLDL) lipoproteins. *Arterioscler Thromb Vasc Biol* 17: 2341-2349.
265. Nofer JR, van Eck M (2011) HDL scavenger receptor class B type I and platelet function. *Curr Opin Lipidol* 22: 277-282.
266. Dole VS, Matuskova J, Vasile E, Yesilaltay A, Bergmeier W, et al. (2008) Thrombocytopenia and platelet abnormalities in high-density lipoprotein receptor-deficient mice. *Arterioscler Thromb Vasc Biol* 28: 1111-1116.
267. Ma Y, Ashraf MZ, Podrez EA (2010) Scavenger receptor BI modulates platelet reactivity and thrombosis in dyslipidemia. *Blood* 116: 1932-1941.
268. Connelly MA, Williams DL (2003) SR-BI and cholesterol uptake into steroidogenic cells. *Trends Endocrinol Metab* 14: 467-472.
269. Schwartz CC, VandenBroek JM, Cooper PS (2004) Lipoprotein cholesteryl ester production, transfer, and output in vivo in humans. *J Lipid Res* 45: 1594-1607.
270. Hildebrand RB, Lammers B, Meurs I, Korporaal SJ, De Haan W, et al. (2010) Restoration of high-density lipoprotein levels by cholesteryl ester transfer protein expression in scavenger receptor class B type I (SR-BI) knockout mice does not normalize pathologies associated with SR-BI deficiency. *Arterioscler Thromb Vasc Biol* 30: 1439-1445.
271. Lagrost L, Gandjini H, Athias A, Guyard-Dangremont V, Lallemand C, et al. (1993) Influence of plasma cholesteryl ester transfer activity on the LDL and HDL distribution profiles in normolipidemic subjects. *Arterioscler Thromb* 13: 815-825.
272. Miettinen HE, Rayburn H, Krieger M (2001) Abnormal lipoprotein metabolism and reversible female infertility in HDL receptor (SR-BI)-deficient mice. *J Clin Invest* 108: 1717-1722.

273. Hoekstra M, Meurs I, Koenders M, Out R, Hildebrand RB, et al. (2008) Absence of HDL cholesteryl ester uptake in mice via SR-BI impairs an adequate adrenal glucocorticoid-mediated stress response to fasting. *J Lipid Res* 49: 738-745.
274. Hoekstra M, Ye D, Hildebrand RB, Zhao Y, Lammers B, et al. (2009) Scavenger receptor class B type I-mediated uptake of serum cholesterol is essential for optimal adrenal glucocorticoid production. *J Lipid Res* 50: 1039-1046.
275. Kocher O, Yesilaltay A, Cirovic C, Pal R, Rigotti A, et al. (2003) Targeted disruption of the PDZK1 gene in mice causes tissue-specific depletion of the high density lipoprotein receptor scavenger receptor class B type I and altered lipoprotein metabolism. *J Biol Chem* 278: 52820-52825.
276. Mineo C, Shaul PW (2012) Functions of scavenger receptor class B, type I in atherosclerosis. *Curr Opin Lipidol* 23: 487-493.
277. Rigotti A, Miettinen HE, Krieger M (2003) The role of the high-density lipoprotein receptor SR-BI in the lipid metabolism of endocrine and other tissues. *Endocr Rev* 24: 357-387.
278. White T, Bennett EP, Takio K, Sorensen T, Bonding N, et al. (1995) Purification and cDNA cloning of a human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. *J Biol Chem* 270: 24156-24165.
279. Schjoldager KT, Clausen H (2012) Site-specific protein O-glycosylation modulates proprotein processing - deciphering specific functions of the large polypeptide GalNAc-transferase gene family. *Biochim Biophys Acta* 1820: 2079-2094.
280. Kato K, Jeanneau C, Tarp MA, Benet-Pages A, Lorenz-Depiereux B, et al. (2006) Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis. Secretion of fibroblast growth factor 23 requires O-glycosylation. *J Biol Chem* 281: 18370-18377.
281. Leuenberger B, Hahn D, Pischitzis A, Hansen MK, Sterchi EE (2003) Human meprin beta: O-linked glycans in the intervening region of the type I membrane protein protect the C-terminal region from proteolytic cleavage and diminish its secretion. *Biochem J* 369: 659-665.
282. Goth CK, Halim A, Khetarpal SA, Rader DJ, Clausen H, et al. (2015) A systematic study of modulation of ADAM-mediated ectodomain shedding by site-specific O-glycosylation. *Proc Natl Acad Sci U S A* 112: 14623-14628.
283. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297-1303.
284. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079.
285. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38: e164.
286. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, et al. (2010) A map of human genome variation from population-scale sequencing. *Nature* 467: 1061-1073.
287. Schwientek T, Bennett EP, Flores C, Thacker J, Hollmann M, et al. (2002) Functional conservation of subfamilies of putative UDP-N-acetylgalactosamine:polypeptide N-acetylgalactosaminyltransferases in *Drosophila*, *Caenorhabditis elegans*, and mammals. One subfamily composed of I(2)35Aa is essential in *Drosophila*. *J Biol Chem* 277: 22623-22638.

288. Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, et al. (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* 325: 433.
289. Steentoft C, Vakhrushev SY, Vester-Christensen MB, Schjoldager KT, Kong Y, et al. (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* 8: 977-982.
290. Vakhrushev SY, Steentoft C, Vester-Christensen MB, Bennett EP, Clausen H, et al. (2013) - Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells. *Mol Cell Proteomics* 12: 932-944.
291. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, et al. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* 32: 1478-1488.
292. Jauhainen M, Metso J, Pahlman R, Blomqvist S, van Tol A, et al. (1993) Human plasma phospholipid transfer protein causes high density lipoprotein conversion. *J Biol Chem* 268: 4032-4036.
293. Speijer H, Groener JE, van Ramshorst E, van Tol A (1991) Different locations of cholesteryl ester transfer protein and phospholipid transfer protein activities in plasma. *Atherosclerosis* 90: 159-168.
294. Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, et al. (2011) Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet* 7: e1002078.
295. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575.
296. Delaneau O, Marchini J, Zagury JF (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9: 179-181.
297. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5: e1000529.
298. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21.
299. Myocardial Infarction G, Investigators CAEC (2016) Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N Engl J Med* 374: 1134-1144.
300. Abou Jamra R, Wohlfart S, Zweier M, Uebe S, Priebe L, et al. (2011) Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity. *Eur J Hum Genet* 19: 1161-1166.
301. Milac AL, Buchete NV, Fritz TA, Hummer G, Tabak LA (2007) Substrate-induced conformational changes and dynamics of UDP-N-acetylgalactosamine:polypeptide N-acetylgalactosaminyltransferase-2. *J Mol Biol* 373: 439-451.
302. Schjoldager KT, Joshi HJ, Kong Y, Goth CK, King SL, et al. (2015) Deconstruction of O-glycosylation--GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. *EMBO Rep* 16: 1713-1722.
303. Albers JJ, Vuletic S, Cheung MC (2012) Role of plasma phospholipid transfer protein in lipid and lipoprotein metabolism. *Biochim Biophys Acta* 1821: 345-357.

304. Jiang XC, Bruce C, Mar J, Lin M, Ji Y, et al. (1999) Targeted mutation of plasma phospholipid transfer protein gene markedly reduces high-density lipoprotein levels. *J Clin Invest* 103: 907-914.
305. Jiang XC, Jin W, Hussain MM (2012) The impact of phospholipid transfer protein (PLTP) on lipoprotein metabolism. *Nutr Metab (Lond)* 9: 75.
306. Huuskonen J, Jauhiainen M, Ehnholm C, Olkkonen VM (1998) Biosynthesis and secretion of human plasma phospholipid transfer protein. *J Lipid Res* 39: 2021-2030.
307. Jiang XC, Qin S, Qiao C, Kawano K, Lin M, et al. (2001) Apolipoprotein B secretion and atherosclerosis are decreased in mice with phospholipid-transfer protein deficiency. *Nat Med* 7: 847-852.
308. Chatterjee C, Sparks DL (2011) Hepatic lipase, high density lipoproteins, and hypertriglyceridemia. *Am J Pathol* 178: 1429-1433.
309. Fritz TA, Raman J, Tabak LA (2006) Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-2. *J Biol Chem* 281: 8613-8619.
310. Grimm D, Streetz KL, Jopling CL, Storm TA, Pandey K, et al. (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* 441: 537-541.
311. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, et al. (2002) Identification of tissue-specific microRNAs from mouse. *Curr Biol* 12: 735-739.
312. Najafi-Shoushtari SH, Kristo F, Li Y, Shioda T, Cohen DE, et al. (2010) MicroRNA-33 and the SREBP host genes cooperate to control cholesterol homeostasis. *Science* 328: 1566-1569.
313. Rayner KJ, Suarez Y, Davalos A, Parathath S, Fitzgerald ML, et al. (2010) MiR-33 contributes to the regulation of cholesterol homeostasis. *Science* 328: 1570-1573.
314. Goedeke L, Rotllan N, Canfran-Duque A, Aranda JF, Ramirez CM, et al. (2015) MicroRNA-148a regulates LDL receptor and ABCA1 expression to control circulating lipoprotein levels. *Nat Med* 21: 1280-1289.
315. Hansen L, Lind-Thomsen A, Joshi HJ, Pedersen NB, Have CT, et al. (2015) A glycogene mutation map for discovery of diseases of glycosylation. *Glycobiology* 25: 211-224.
316. Gerritsen G, Rensen PC, Kypreos KE, Zannis VI, Havekes LM, et al. (2005) ApoC-III deficiency prevents hyperlipidemia induced by apoE overexpression. *J Lipid Res* 46: 1466-1473.
317. Jong MC, Rensen PC, Dahlmans VE, van der Boom H, van Berkel TJ, et al. (2001) Apolipoprotein C-III deficiency accelerates triglyceride hydrolysis by lipoprotein lipase in wild-type and apoE knockout mice. *J Lipid Res* 42: 1578-1585.
318. Masucci-Magoulas L, Goldberg IJ, Bisgaier CL, Serajuddin H, Francone OL, et al. (1997) A mouse model with features of familial combined hyperlipidemia. *Science* 275: 391-394.
319. Mucaki EJ, Shirley BC, Rogan PK (2013) Prediction of mutant mRNA splice isoforms by information theory-based exon definition. *Hum Mutat* 34: 557-565.
320. Ibrahim S, Somanathan S, Billheimer J, Wilson JM, Rader DJ (2016) Stable liver-specific expression of human IDOL in humanized mice raises plasma cholesterol. *Cardiovasc Res* 110: 23-29.

321. Rader DJ, Castro G, Zech LA, Fruchart JC, Brewer HB, Jr. (1991) In vivo metabolism of apolipoprotein A-I on high density lipoprotein particles LpA-I and LpA-I,A-II. *J Lipid Res* 32: 1849-1859.
322. Cain WJ, Millar JS, Himebauch AS, Tietge UJ, Maugeais C, et al. (2005) Lipoprotein [a] is cleared from the plasma primarily by the liver in a process mediated by apolipoprotein [a]. *J Lipid Res* 46: 2681-2691.
323. Ota T, Gayet C, Ginsberg HN (2008) Inhibition of apolipoprotein B100 secretion by lipid-induced hepatic endoplasmic reticulum stress in rodents. *J Clin Invest* 118: 316-332.
324. McCoy MG, Sun GS, Marchadier D, Maugeais C, Glick JM, et al. (2002) Characterization of the lipolytic activity of endothelial lipase. *J Lipid Res* 43: 921-929.
325. Nalla VK, Rogan PK (2005) Automated splicing mutation analysis by information theory. *Hum Mutat* 25: 334-342.
326. Sundaram M, Zhong S, Bou Khalil M, Links PH, Zhao Y, et al. (2010) Expression of apolipoprotein C-III in McA-RH7777 cells enhances VLDL assembly and secretion under lipid-rich conditions. *J Lipid Res* 51: 150-161.
327. Sundaram M, Zhong S, Bou Khalil M, Zhou H, Jiang ZG, et al. (2010) Functional analysis of the missense APOC3 mutation Ala23Thr associated with human hypotriglyceridemia. *J Lipid Res* 51: 1524-1534.
328. Millar JS, Cromley DA, McCoy MG, Rader DJ, Billheimer JT (2005) Determining hepatic triglyceride production in mice: comparison of poloxamer 407 with Triton WR-1339. *J Lipid Res* 46: 2023-2028.
329. Malmendier CL, Lontie JF, Grutman GA, Delcroix C (1988) Metabolism of apolipoprotein C-III in normolipemic human subjects. *Atherosclerosis* 69: 51-59.
330. Malmendier CL, Delcroix C, Lontie JF (1989) Kinetics of a heterogeneous population of particles in low density lipoprotein apolipoprotein B. *Atherosclerosis* 80: 91-100.
331. Malmendier CL, Lontie JF, Delcroix C, Dubois DY, Magot T, et al. (1989) Apolipoproteins C-II and C-III metabolism in hypertriglyceridemic patients. Effect of a drastic triglyceride reduction by combined diet restriction and fenofibrate administration. *Atherosclerosis* 77: 139-149.
332. Ooi EM, Chan DT, Watts GF, Chan DC, Ng TW, et al. (2011) Plasma apolipoprotein C-III metabolism in patients with chronic kidney disease. *J Lipid Res* 52: 794-800.
333. Gordts PL, Nock R, Son NH, Ramms B, Lew I, et al. (2016) ApoC-III inhibits clearance of triglyceride-rich lipoproteins through LDL family receptors. *J Clin Invest* 126: 2855-2866.
334. de Silva HV, Lauer SJ, Wang J, Simonet WS, Weisgraber KH, et al. (1994) Overexpression of human apolipoprotein C-III in transgenic mice results in an accumulation of apolipoprotein B48 remnants that is corrected by excess apolipoprotein E. *J Biol Chem* 269: 2324-2335.
335. Drenos F, Davey Smith G, Ala-Korpela M, Kettunen J, Wurtz P, et al. (2016) Metabolic Characterization of a Rare Genetic Variation Within APOC3 and Its Lipoprotein Lipase-Independent Effects. *Circ Cardiovasc Genet* 9: 231-239.

336. Dahagam C, Goud A, Abdelqader A, Hendrani A, Feinstein MJ, et al. (2016) PCSK9 inhibitors and their role in high-risk patients in reducing LDL cholesterol levels: alirocumab. *Future Cardiol* 12: 149-157.
337. Dahagam C, Goud A, Abdelqader A, Hendrani A, Feinstein MJ, et al. (2016) PCSK9 inhibitors and their role in high-risk patients in reducing LDL cholesterol levels: evolocumab. *Future Cardiol* 12: 139-148.
338. Schneider JF, Thomas HE, Jr., Kreger BE, McNamara PM, Kannel WB (1979) Newly acquired left bundle-branch block: the Framingham study. *Ann Intern Med* 90: 303-310.
339. Buxton A, Goldberg S, Hirshfeld JW, Wilson J, Mann T, et al. (1980) Refractory ergonovine-induced coronary vasospasm: importance of intracoronary nitroglycerin. *Am J Cardiol* 46: 329-334.
340. Kueppers F, Miller RD, Gordon H, Hepper NG, Offord K (1977) Familial prevalence of chronic obstructive pulmonary disease in a matched pair study. *Am J Med* 63: 336-342.
341. Castelli WP, Doyle JT, Gordon T, Hames CG, Hjortland MC, et al. (1977) HDL cholesterol and other lipids in coronary heart disease. The cooperative lipoprotein phenotyping study. *Circulation* 55: 767-772.
342. Castelli WP, Garrison RJ, Wilson PW, Abbott RD, Kalousdian S, et al. (1986) Incidence of coronary heart disease and lipoprotein cholesterol levels. The Framingham Study. *JAMA* 256: 2835-2838.
343. Sharrett AR, Ballantyne CM, Coady SA, Heiss G, Sorlie PD, et al. (2001) Coronary heart disease prediction from lipoprotein cholesterol levels, triglycerides, lipoprotein(a), apolipoproteins A-I and B, and HDL density subfractions: The Atherosclerosis Risk in Communities (ARIC) Study. *Circulation* 104: 1108-1113.
344. Barter PJ, Rye KA (2015) Targeting High-density Lipoproteins to Reduce Cardiovascular Risk: What Is the Evidence? *Clin Ther* 37: 2716-2731.
345. Barter P, Gotto AM, LaRosa JC, Maroni J, Szarek M, et al. (2007) HDL cholesterol, very low levels of LDL cholesterol, and cardiovascular events. *N Engl J Med* 357: 1301-1310.
346. Opie LH, Schwartz PJ (2012) Dronedronone in high-risk permanent atrial fibrillation. *N Engl J Med* 366: 1159; author reply 1161.
347. Wijns W, Shite J, Jones MR, Lee SW, Price MJ, et al. (2015) Optical coherence tomography imaging during percutaneous coronary intervention impacts physician decision-making: ILUMIEN I study. *Eur Heart J* 36: 3346-3355.
348. Shah T, Zabaneh D, Gaunt T, Swerdlow DI, Shah S, et al. (2013) Gene-centric analysis identifies variants associated with interleukin-6 levels and shared pathways with other inflammation markers. *Circ Cardiovasc Genet* 6: 163-170.
349. Zanon P, Khetarpal SA, Larach DB, Hancock-Cerutti WF, Millar JS, et al. (2016) Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. *Science* 351: 1166-1171.
350. Patel AP, Peloso GM, Pirruccello JP, Johansen CT, Dube JB, et al. (2016) Targeted exonic sequencing of GWAS loci in the high extremes of the plasma lipids distribution. *Atherosclerosis* 250: 63-68.
351. Weissglas-Volkov D, Calkin AC, Tusie-Luna T, Sinsheimer JS, Zelcer N, et al. (2011) The N342S MYLIP polymorphism is associated with high total cholesterol

- and increased LDL receptor degradation in humans. *J Clin Invest* 121: 3062-3071.
352. Service SK, Teslovich TM, Fuchsberger C, Ramensky V, Yajnik P, et al. (2014) Re-sequencing expands our understanding of the phenotypic impact of variants at GWAS loci. *PLoS Genet* 10: e1004147.
 353. Cui G, Li Z, Li R, Huang J, Wang H, et al. (2014) A functional variant in APOA5/A4/C3/A1 gene cluster contributes to elevated triglycerides and severity of CAD by interfering with microRNA 3201 binding efficiency. *J Am Coll Cardiol* 64: 267-277.
 354. Johansen CT, Wang J, McIntyre AD, Martins RA, Ban MR, et al. (2012) Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia. *Circ Cardiovasc Genet* 5: 66-72.
 355. Wang J, Cao H, Ban MR, Kennedy BA, Zhu S, et al. (2007) Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). *Arterioscler Thromb Vasc Biol* 27: 2450-2455.
 356. Minster RL, Hawley NL, Su CT, Sun G, Kershaw EE, et al. (2016) A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat Genet*.
 357. Volckmar AL, Han CT, Putter C, Haas S, Vogel CI, et al. (2016) Analysis of Genes Involved in Body Weight Regulation by Targeted Re-Sequencing. *PLoS One* 11: e0147904.
 358. Sallman Almen M, Rask-Andersen M, Jacobsson JA, Ameer A, Kalnina I, et al. (2013) Determination of the obesity-associated gene variants within the entire FTO gene by ultra-deep targeted sequencing in obese and lean children. *Int J Obes (Lond)* 37: 424-431.
 359. Van Camp JK, Zegers D, Verhulst SL, Van Hoorenbeeck K, Massa G, et al. (2013) Mutation analysis of WNT10B in obese children, adolescents and adults. *Endocrine* 44: 107-113.
 360. Zegers D, Beckers S, Hendrickx R, Van Camp JK, Van Hoorenbeeck K, et al. (2013) Prevalence of rare MC3R variants in obese cases and lean controls. *Endocrine* 44: 386-390.
 361. Harismendy O, Bansal V, Bhatia G, Nakano M, Scott M, et al. (2010) Population sequencing of two endocannabinoid metabolic genes identifies rare and common regulatory variants associated with extreme obesity and metabolite level. *Genome Biol* 11: R118.
 362. Bettens K, Brouwers N, Engelborghs S, Lambert JC, Rogaeva E, et al. (2012) Both common variations and rare non-synonymous substitutions and small insertion/deletions in CLU are associated with increased Alzheimer risk. *Mol Neurodegener* 7: 3.
 363. Jin SC, Pastor P, Cooper B, Cervantes S, Benitez BA, et al. (2012) Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. *Alzheimers Res Ther* 4: 34.
 364. Vardarajan BN, Ghani M, Kahn A, Sheikh S, Sato C, et al. (2015) Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci. *Ann Neurol* 78: 487-498.

365. Cuyvers E, De Roeck A, Van den Bossche T, Van Cauwenberghe C, Bettens K, et al. (2015) Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. *Lancet Neurol* 14: 814-822.
366. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747-753.
367. Lee SH, Wray NR, Goddard ME, Visscher PM (2011) Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 88: 294-305.
368. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, et al. (2012) Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338: 1619-1622.
369. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, et al. (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7: 111-118.
370. Consortium EP (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57-74.
371. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, et al. (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317-330.
372. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120.
373. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754-1760.
374. Liu X, Jian X, Boerwinkle E (2011) dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* 32: 894-899.
375. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43: 491-498.
376. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, et al. (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43: 11 10 11-33.
377. Mills RE, Pittard WS, Mullaney JM, Farooq U, Creasy TH, et al. (2011) Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res* 21: 830-839.
378. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, et al. (2015) A global reference for human genetic variation. *Nature* 526: 68-74.
379. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27: 2156-2158.
380. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, et al. (2016) The Ensembl Variant Effect Predictor. *Genome Biol* 17: 122.
381. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, et al. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26: 2069-2070.
382. Yates A, Akanni W, Amode MR, Barrell D, Billis K, et al. (2016) Ensembl 2016. *Nucleic Acids Res* 44: D710-716.
383. Yates A, Beal K, Keenan S, McLaren W, Pignatelli M, et al. (2015) The Ensembl REST API: Ensembl Data for Any Language. *Bioinformatics* 31: 143-145.

384. Larach DB, Cuchel M, Rader DJ (2013) Monogenic causes of elevated HDL cholesterol and implications for development of new therapeutics. *Clin Lipidol* 8: 635-648.
385. Singaraja RR, Tietjen I, Hovingh GK, Franchini PL, Radomski C, et al. (2014) Identification of four novel genes contributing to familial elevated plasma HDL cholesterol in humans. *J Lipid Res* 55: 1693-1701.
386. Tietjen I, Hovingh GK, Singaraja RR, Radomski C, Barhdadi A, et al. (2012) Segregation of LIPG, CETP, and GALNT2 mutations in Caucasian families with extremely high HDL cholesterol. *PLoS One* 7: e37437.
387. Huyghe JR, Jackson AU, Fogarty MP, Buchkovich ML, Stancakova A, et al. (2013) Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet* 45: 197-201.
388. Di Bartolo B, Takata K, Duong M, Nicholls SJ (2016) CETP Inhibition in CVD Prevention: an Actual Appraisal. *Curr Cardiol Rep* 18: 43.
389. Barter PJ, Brewer HB, Jr., Chapman MJ, Hennekens CH, Rader DJ, et al. (2003) Cholesteryl ester transfer protein: a novel target for raising HDL and inhibiting atherosclerosis. *Arterioscler Thromb Vasc Biol* 23: 160-167.
390. Brousseau ME, Schaefer EJ, Wolfe ML, Bloedon LT, Digenio AG, et al. (2004) Effects of an inhibitor of cholesteryl ester transfer protein on HDL cholesterol. *N Engl J Med* 350: 1505-1515.
391. Cannon CP, Shah S, Dansky HM, Davidson M, Brinton EA, et al. (2010) Safety of anacetrapib in patients with or at high risk for coronary heart disease. *N Engl J Med* 363: 2406-2415.
392. Papp AC, Pinsonneault JK, Wang D, Newman LC, Gong Y, et al. (2012) Cholesteryl Ester Transfer Protein (CETP) polymorphisms affect mRNA splicing, HDL levels, and sex-dependent cardiovascular risk. *PLoS One* 7: e31930.
393. Thompson JF, Wood LS, Pickering EH, Dechairo B, Hyde CL (2007) High-density genotyping and functional SNP localization in the CETP gene. *J Lipid Res* 48: 434-443.
394. Frisdal E, Klerkx AH, Le Goff W, Tanck MW, Lagarde JP, et al. (2005) Functional interaction between -629C/A, -971G/A and -1337C/T polymorphisms in the CETP gene is a major determinant of promoter activity and plasma CETP concentration in the REGRESS Study. *Hum Mol Genet* 14: 2607-2618.
395. Borggreve SE, Hillege HL, Wolffenbuttel BH, de Jong PE, Bakker SJ, et al. (2005) The effect of cholesteryl ester transfer protein -629C->A promoter polymorphism on high-density lipoprotein cholesterol is dependent on serum triglycerides. *J Clin Endocrinol Metab* 90: 4198-4204.
396. Klerkx AH, Tanck MW, Kastelein JJ, Molhuizen HO, Jukema JW, et al. (2003) Haplotype analysis of the CETP gene: not TaqIB, but the closely linked -629C->A polymorphism and a novel promoter variant are independently associated with CETP concentration. *Hum Mol Genet* 12: 111-123.
397. von Kodolitsch Y, Pyeritz RE, Rogan PK (1999) Splice-site mutations in atherosclerosis candidate genes: relating individual information to phenotype. *Circulation* 100: 693-699.
398. Chasman DI, Pare G, Zee RY, Parker AN, Cook NR, et al. (2008) Genetic loci associated with plasma concentration of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, apolipoprotein A1, and

- Apolipoprotein B among 6382 white women in genome-wide analysis with replication. *Circ Cardiovasc Genet* 1: 21-30.
399. Jaye M, Lynch KJ, Krawiec J, Marchadier D, Maugeais C, et al. (1999) A novel endothelial-derived lipase that modulates HDL metabolism. *Nat Genet* 21: 424-428.
 400. Hirata K, Dichek HL, Cioffi JA, Choi SY, Leeper NJ, et al. (1999) Cloning of a unique lipase from endothelial cells extends the lipase gene family. *J Biol Chem* 274: 14170-14175.
 401. Badellino KO, Rader DJ (2004) The role of endothelial lipase in high-density lipoprotein metabolism. *Curr Opin Cardiol* 19: 392-395.
 402. Edmondson AC, Brown RJ, Kathiresan S, Cupples LA, Demissie S, et al. (2009) Loss-of-function variants in endothelial lipase are a cause of elevated HDL cholesterol in humans. *J Clin Invest* 119: 1042-1050.
 403. Khetarpal SA, Edmondson AC, Raghavan A, Neeli H, Jin W, et al. (2011) Mining the LIPG allelic spectrum reveals the contribution of rare and common regulatory variants to HDL cholesterol. *PLoS Genet* 7: e1002393.
 404. deLemos AS, Wolfe ML, Long CJ, Sivapackianathan R, Rader DJ (2002) Identification of genetic variants in endothelial lipase in persons with elevated high-density lipoprotein cholesterol. *Circulation* 106: 1321-1326.
 405. Bachmann-Gagescu R, Dempsey JC, Phelps IG, O'Roak BJ, Knutzen DM, et al. (2015) Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity. *J Med Genet* 52: 514-522.
 406. Schueler M, Halbritter J, Phelps IG, Braun DA, Otto EA, et al. (2016) Large-scale targeted sequencing comparison highlights extreme genetic heterogeneity in nephronophthisis-related ciliopathies. *J Med Genet* 53: 208-214.
 407. Surakka I, Horikoshi M, Magi R, Sarin AP, Mahajan A, et al. (2015) The impact of low-frequency and rare variants on lipid levels. *Nat Genet* 47: 589-597.
 408. Helgadottir A, Gretarsdottir S, Thorleifsson G, Hjartarson E, Sigurdsson A, et al. (2016) Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. *Nat Genet* 48: 634-639.
 409. Gotto AM, Jr., Moon JE (2012) Management of cardiovascular risk: the importance of meeting lipid targets. *Am J Cardiol* 110: 3A-14A.
 410. Gotto AM, Jr. (2012) Update on the evaluation and management of cardiovascular risk. Introduction. *Am J Cardiol* 110: 1A-2A.
 411. Gotto AM, Jr., Moon JE (2012) Recent clinical studies of the effects of lipid-modifying therapies. *Am J Cardiol* 110: 15A-26A.
 412. Lambert G, Sjouke B, Choque B, Kastelein JJ, Hovingh GK (2012) The PCSK9 decade. *J Lipid Res* 53: 2515-2524.
 413. Lusis AJ, Pajukanta P (2008) A treasure trove for lipoprotein biology. *Nat Genet* 40: 129-130.
 414. van Capelleveen JC, Bochem AE, Motazacker MM, Hovingh GK, Kastelein JJ (2013) Genetics of HDL-C: a causal link to atherosclerosis? *Curr Atheroscler Rep* 15: 326.
 415. Kuivenhoven JA, Hegele RA (2014) Mining the genome for lipid genes. *Biochim Biophys Acta* 1842: 1993-2009.

416. Zhang Y, Zanotti I, Reilly MP, Glick JM, Rothblat GH, et al. (2003) Overexpression of apolipoprotein A-I promotes reverse transport of cholesterol from macrophages to feces in vivo. *Circulation* 108: 661-663.
417. Moore RE, Navab M, Millar JS, Zimetti F, Hama S, et al. (2005) Increased atherosclerosis in mice lacking apolipoprotein A-I attributable to both impaired reverse cholesterol transport and increased inflammation. *Circ Res* 97: 763-771.
418. Naik SU, Wang X, Da Silva JS, Jaye M, Macphee CH, et al. (2006) Pharmacological activation of liver X receptors promotes reverse cholesterol transport in vivo. *Circulation* 113: 90-97.
419. Kobayashi J, Miyashita K, Nakajima K, Mabuchi H (2015) Hepatic Lipase: a Comprehensive View of its Role on Plasma Lipid and Lipoprotein Metabolism. *J Atheroscler Thromb* 22: 1001-1011.
420. Pieper-Furst U, Lammert F (2013) Low-density lipoprotein receptors in liver: old acquaintances and a newcomer. *Biochim Biophys Acta* 1831: 1191-1198.
421. Topaz O, Shurman DL, Bergman R, Indelman M, Ratajczak P, et al. (2004) Mutations in GALNT3, encoding a protein involved in O-linked glycosylation, cause familial tumoral calcinosis. *Nat Genet* 36: 579-581.
422. Gerritsen G, van der Hoogt CC, Schaap FG, Voshol PJ, Kypreos KE, et al. (2008) ApoE2-associated hypertriglyceridemia is ameliorated by increased levels of apoA-V but unaffected by apoC-III deficiency. *J Lipid Res* 49: 1048-1055.