

IDENTIFICATION OF REGULATORY ELEMENTS AND FUNCTIONAL VARIANTS AT GWAS LOCI FOR
HUMAN METABOLIC TRAITS

Tamara S. Roman

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment
of the requirements for the degree of Doctor of Philosophy in the Curriculum of Genetics and Molecular
Biology in the School of Medicine.

Chapel Hill
2016

Approved by:

Karen Mohlke

Praveen Sethupathy

Nobuyo Maeda

John Rawls

Ian Davis

©2016
Tamara S. Roman
ALL RIGHTS RESERVED

ABSTRACT

Tamara S. Roman: Identification of regulatory elements and functional variants at GWAS loci for human metabolic traits
(Under the direction of Karen Mohlke)

Cardiovascular disease (CVD) and type 2 diabetes (T2D) are major public health burdens in the United States. Metabolic risk factors for these diseases include increased triglyceride levels, decreased high-density lipoprotein cholesterol levels (HDL-C), and increased waist-hip ratio (WHR, a measure of body fat distribution). Genome-wide association studies (GWAS) have identified hundreds of loci for human metabolic traits and diseases. For many of these loci, the molecular and biological mechanisms are unknown. Many association signals are located in non-coding regions of the genome, and some of the associated variants are located within regulatory elements. Due to linkage disequilibrium, the GWAS identified variants may serve as proxies for the true functional variants. Functional follow-up studies are necessary to identify the functional variant(s) at a locus, determine the target gene(s) and mechanisms for how the variants and genes influence metabolic traits.

I identified functional regulatory variants and elements at three GWAS loci for metabolic traits. At the *GALNT2* locus for HDL-C, at least two regulatory variants demonstrated allelic differences in transcriptional enhancer activity and transcription factor binding. The alleles associated with increased HDL-C demonstrated increased *GALNT2* expression in human hepatocytes and subcutaneous adipose tissue samples. At the *ADCY5* locus for T2D and glucose-related traits, I identified a regulatory variant located within an enhancer element in human pancreatic islets that exhibited allelic differences in enhancer activity, differential transcription factor binding, and allelic differences in regulatory H3K27ac ChIP-seq reads from human islets. The T2D risk allele showed lower *ADCY5* expression in human islets. At the *PLXND1* locus for WHR, I identified at least 4 enhancer elements in human umbilical vein endothelial cells, one of which also showed enhancer activity in zebrafish endothelial cells.

Functional studies helped to identify regulatory variants, elements and target genes underlying these three association signals, and provided evidence that some GWAS loci have multiple regulatory

variants that act to influence gene expression. Identification of functional variants will enable the characterization of the molecular mechanisms and direction of effect in humans, leading to a greater understanding of the relationship between these variants and target genes with metabolic traits and disease.

ACKNOWLEDGEMENTS

Thank you to my advisor, Karen Mohlke, for giving me the opportunity to be a part of your lab and for all of your advice, guidance and teaching throughout my graduate career. I have learned so much from your strong science writing skills and strong ability to collaborate with others. I am grateful for my experiences while in your lab, including presenting my research at conferences and learning how to think more analytically about scientific questions and research. You have always been supportive of my career goals, and I thank you for providing me with great preparation for the next stage of my scientific career.

Thank you to my committee members, Praveen Sethupathy, Nobuyo Maeda, John Rawls and Ian Davis, for your helpful advice and feedback regarding my projects, taking the time to attend my committee meetings and for teaching me to think critically about my research. I am very appreciative to have all of you as members of my committee.

Thank you to all members of the Mohlke lab, past and present, for being such a great group of individuals to work with! You have all helped me to think about my research in different ways, aided my growth as a scientist and have always supported me with all aspects of graduate school. I am grateful for having such a positive and interactive work environment for the past several years.

Thank you to my collaborators for all of your guidance, advice, manuscript and abstract editing and contributions to my research. Many of the data supporting my experimental evidence highlighted in this dissertation would not have been possible without these smart and hard-working individuals. I have learned a lot from my interactions with collaborators and appreciate their fostering of a positive collaborative environment.

For Chapter 2, I would like to thank Swarooparani Vadlamudi, Amanda Marvelle, and Arlene Gonzalez for help with luciferase reporter assays; Marie Fogarty, Karen Mohlke, Kyle Gaulton and Praveen Sethupathy for help with data interpretation and experimental design; Kyle Gaulton and Praveen Sethupathy for contributing transcription factor binding predictions for *GALNT2* HDL-C-associated variants and Martin Buchkovich for performing allelic-imbalance analysis of sequence reads from CEBPB

ChIP-seq and DNase-seq reads. Thank you Jeroen Huyghe for fine-mapping and conditional analyses of METSIM data; Christian Fuchsberger for imputation of genotypes from METSIM data; Anne Jackson and Jeroen Huyghe for association analyses of lipid and lipoprotein particle-serum concentrations from the METSIM study; Ying Wu and Mete Civelek for performing eQTL analyses from subcutaneous adipose tissue samples from the METSIM study; Antti Kangas, Pasi Soininen and Mika Ala-Korpela for measuring and providing lipid and lipoprotein particle-serum concentrations from METSIM study samples and Aldons Lusis, Johanna Kuusisto, Markku Laakso, Francis Collins, Michael Boehnke and Karen Mohlke for contributing data and analyses from the METSIM study. Thank you Jennifer Kulzer, Marie Fogarty, Kyle Gaulton, Praveen Sethupathy, Mika Ala-Korpela, Francis Collins, Michael Boehnke and Karen Mohlke for editing the *GALNT2* manuscript. I would also like to acknowledge the GoT2D Consortium for access to the reference panel used for imputation; human organ donors whose liver samples were used in the study; the ENCODE and NIH Roadmap Epigenomics Consortia, Data Analysis and Coordination Centers, and production laboratories that generated the chromatin, histone-modification, and chromatin-immunoprecipitation (ChIP) sequencing data used for variant annotation. Thank you Terry Furey and Greg Crawford for interpreting ENCODE data, Yun Li and Qing Duan for providing 1000 Genomes linkage-disequilibrium data, Scott Bultman and Dallas Donohoe for providing advice on ChIP experiments and Sumeet Khetarpal and Daniel Rader for sharing unpublished results.

For Chapter 3, thank you Swarooparani Vadlamudi for performing transcriptional reporter assays in MIN6 cells; Michael Stitzel and Mario Morken for H3K27ac ChIP-seq experiment data from primary human islets; Michael Stitzel and Stephen Parker for human islet regulatory data, including RNA-seq and chromatin state data; Ryan Welch for performing eQTL analyses in human islet samples; Michael Erdos for human islet collection and RNA-seq, DNase-seq and genotyping experiments and data; Brooke Wolford and Martin Buchkovich for allelic imbalance analyses of H3K27ac ChIP-seq reads in a human islet sample and Peter Chines for transcription factor binding prediction analyses. Thank you Anne Jackson for performing association analyses for T2D and glucose-related traits in METSIM and Ying Wu for performing lookups for rs11708067 associations with glucose and insulin-related traits in METSIM. Thank you Karen Mohlke, Francis Collins, Kyle Gaulton, Swarooparani Vadlamudi, Michael Stitzel, Stephen Parker, Ryan Welch, Michael Erdos, Brooke Wolford, Martin Buchkovich, Anne Jackson, Peter

Chines, Mario Morken and Ying Wu for their contributions to data interpretation and/or editing of abstracts for this project.

Chapter 4 was a collaboration with Dr. John Rawls' laboratory at Duke University. Thank you James Minchin for contributing to the experimental design of CREs and cloning of CREs into the EGFP vector for zebrafish transgenesis assays. Thank you James Minchin and Patrick Williams for performing zebrafish injections, visualizing GFP and mCherry expression on the fluorescence microscope and interpreting data for the zebrafish transgenesis assays. Thank you Kristin Young for assisting with site-directed mutagenesis of rs4488824 in CRE 13 in the gateway-compatible pGL4.23 luciferase reporter vector and purification of plasmids. Thank you Karen Mohlke, John Rawls and James Minchin for guidance and editing of *PLXND1* project summaries, presentations and abstracts.

Thank you to The Integrative Vascular Biology Training Program (IVB) and the opportunity to be a trainee of this program. I have learned many skills throughout my time in IVB, including how to give presentations of my research to a broader audience and how to establish collaborations with other departments. I appreciate getting to know all of the professors and students that I shared the experiences of the program with.

Finally, thank you to my family, especially Mom, Dad, Kevin, Tyler and Papa. You have always supported me through both successes and struggles! Thank you for always being there for me and listening, cheering me on and always providing a constant source of love and encouragement. I am grateful to have such a loving family of which I hold very dear. I love you.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS.....	xiv
CHAPTER 1: INTRODUCTION	1
1.1 Cardiovascular disease	1
1.2 Lipoproteins and cholesterol	1
1.3 Type 2 diabetes.....	2
1.4 Obesity and body fat distribution.....	2
1.5 Genetics of cardiometabolic traits	3
1.6 Identifying functional regulatory DNA variants from GWAS.....	3
1.7 Aims and overview	4
CHAPTER 2: MULTIPLE HEPATIC REGULATORY VARIANTS AT THE <i>GALNT2</i> GWAS LOCUS ASSOCIATED WITH HIGH-DENSITY LIPOPROTEIN CHOLESTEROL	7
2.1 Introduction.....	7
2.2 Materials and Methods	8
2.2.1 Defining the candidate set of variants	8
2.2.2 Genotyping and imputation	9
2.2.3 Fine-mapping and conditional analyses.....	9
2.2.4 Cell culture	10
2.2.5 Transcriptional reporter assays.....	10
2.2.6 Electrophoretic mobility shift assays	11
2.2.7 ChIP assays	12
2.2.8 Allelic-imbalance analysis of sequence reads from CEBPB ChIP-seq and DNase-seq.....	13
2.2.9 Preparation of cDNA for <i>GALNT2</i> mRNA-expression and AEI assays	14

2.2.10	AEI assays	14
2.2.11	<i>GALNT2</i> hepatocyte mRNA expression	14
2.2.12	Lookup of expression quantitative trait loci in subcutaneous adipose tissue	15
2.2.13	siRNA-mediated knockdown of <i>CEBPB</i>	15
2.3	Results	16
2.3.1	Fine mapping shows evidence of a single association signal strongest for total cholesterol in medium HDL	16
2.3.2	Open-chromatin, histone-modification, and transcription-factor marks indicate potential regulatory elements overlapping the <i>GALNT2</i> association signal	16
2.3.3	All 25 candidate regulatory variants were evaluated for allelic differences in luciferase activity	17
2.3.4	rs2281721 exhibits allelic differences in transcriptional activity.....	18
2.3.5	A segment containing rs4846913, rs2144300, and rs6143660 shows haplotype differences in transcriptional activity	18
2.3.6	The haplotype variants act together to regulate enhancer activity	19
2.3.7	USF1 binds to rs2281721	20
2.3.8	CEBPB binds differentially to the alleles of rs4846913.....	21
2.3.9	rs2144300, rs1555290, and rs6143660 also show suggestive evidence of protein binding	22
2.3.10	Variants associated with <i>GALNT2</i> expression.....	22
2.4	Discussion.....	23
2.5	Conflicts of interest.....	27
	ENDNOTES.....	69
CHAPTER 3: A FUNCTIONAL REGULATORY VARIANT ASSOCIATED WITH TYPE 2 DIABETES, FASTING GLUCOSE AND DECREASED <i>ADCY5</i> EXPRESSION		70
3.1	Introduction.....	70
3.2	Materials and Methods.....	71
3.2.1	Lookup of expression quantitative trait loci in primary human islets.....	71
3.2.2	H3K27ac CHIP-seq in a primary human islet sample	72
3.2.3	Cell culture	72
3.2.4	Transcriptional reporter assays.....	72

3.2.5	Electrophoretic mobility shift assays	73
3.3	Results	74
3.3.1	rs11708067 is associated with <i>ADCY5</i> expression in primary human pancreatic islets	74
3.3.2	rs11708067 overlaps evidence of open chromatin and a predicted strong enhancer region in human pancreatic islets	74
3.3.3	Allelic imbalance of H3K27ac reads overlapping rs11708067 in human pancreatic islets	75
3.3.4	rs11708067 shows allelic differences in enhancer activity	75
3.3.5	The alleles of rs11708067 exhibit differential protein binding	75
3.4	Discussion	76
CHAPTER 4: REGULATORY ELEMENTS IN ENDOTHELIAL CELLS AT THE <i>PLXND1</i> GWAS LOCUS ASSOCIATED WITH WHR.....		89
4.1	Introduction.....	89
4.2	Materials and Methods.....	90
4.2.1	Luciferase reporter vectors containing <i>PLXND1</i> candidate cis-regulatory elements.....	90
4.2.2	Generation of a gateway-compatible pGL4.23 luciferase reporter vector	91
4.2.3	Cell culture	92
4.2.4	Cloning of <i>PLXND1</i> candidate <i>cis</i> -regulatory elements into EGFP expression vector	92
4.2.5	Injection and imaging of human <i>PLXND1</i> CREs in zebrafish	92
4.2.6	Electrophoretic mobility shift assays	93
4.3	Results	94
4.3.1	Four candidate CREs show enhancer activity in HUVEC.....	94
4.3.2	The alleles of rs11718169 in CRE 17 show similar enhancer activities in HUVEC	94
4.3.3	CRE 17 shows enhancer activity in endothelial cells in zebrafish	95
4.3.4	rs4488824 in CRE 13 exhibits differential protein binding	95
4.4	Discussion	96
CHAPTER 5: CONCLUSIONS		107
REFERENCES.....		116

LIST OF TABLES

Table 2.1	Definitions of lipoprotein subclass measurements for association analyses in the METSIM study.....	55
Table 2.2	Primer sequences for functional assays	56
Table 2.3	Evidence of association of <i>GALNT2</i> SNP rs2144300 with 72 METSIM lipid and cholesterol traits	59
Table 2.4	Association of <i>GALNT2</i> locus variants with total cholesterol in medium HDL in the METSIM study.....	62
Table 2.5	Variants in strong linkage disequilibrium ($r^2 > 0.7$) with the GWAS HDL-C-associated variant rs4846914 overlap ENCODE open chromatin, histone ChIP-seq and transcription factor ChIP-seq peaks in liver cell types	65
Table 2.6	SNP associations with gene expression levels in primary subcutaneous adipose tissue samples from 1381 individuals from the METSIM study	67
Table 2.7	Initial and residual associations with <i>GALNT2</i> expression in primary subcutaneous adipose tissue samples from 1381 individuals from the METSIM study.....	68
Table 3.1	Summary of published <i>ADCY5</i> SNP associations with T2D and glucose-related traits.....	86
Table 3.2	Primer sequences for luciferase reporter and electrophoretic mobility shift assays	88
Table 4.1	Summary of candidate <i>cis</i> -regulatory elements designed to test in luciferase reporter assays and zebrafish transgenesis assays	105
Table 4.2	Primer sequences for functional assays	106

LIST OF FIGURES

Figure 2.1	Non-coding variants at <i>GALNT2</i> are associated with total cholesterol in medium HDL in the METSIM study.....	28
Figure 2.2	HDL-C-associated variants overlap open chromatin and histone modifications indicating potential regulatory regions in <i>GALNT2</i> intron 1	29
Figure 2.3	HDL-C associated variants overlap open chromatin, histone modifications, and predicted enhancer and transcribed chromatin states in liver cell types identifying potential regulatory regions in <i>GALNT2</i> intron 1	30
Figure 2.4	Two of 14 elements containing variants in strong linkage disequilibrium ($r^2 > 0.7$) with rs4846914 show allelic or haplotype differences in transcriptional enhancer activity in both orientations.....	32
Figure 2.5	Twenty-one <i>GALNT2</i> intronic variants tested in reporter assays do not exhibit allelic or haplotype differences in transcriptional enhancer activity in both orientations.....	33
Figure 2.6	Haplotype and allelic differences in transcriptional activity at the <i>GALNT2</i> locus	34
Figure 2.7	Haplotype and allelic differences in transcriptional activity at the <i>GALNT2</i> locus are consistent in Huh-7 cells.....	35
Figure 2.8	Haplotype variants act together to increase transcriptional activity.....	36
Figure 2.9	Haplotype variants act together to increase transcriptional activity in Huh-7 cells.....	37
Figure 2.10	The 780-bp segment shows haplotype differences in enhancer activity in a promoterless vector.....	38
Figure 2.11	SNPs rs4846913, rs2144300, rs1555290 and 21-bp indel rs6143660 do not show allelic differences in enhancer activity when cloned and tested as separate segments in luciferase reporter assays	39
Figure 2.12	The rs2281721 C allele shows binding to USF1.....	40
Figure 2.13	Transcription factors not bound to rs2281721 or rs4846913.....	41
Figure 2.14	CEBPB binds differentially to the alleles of rs4846913.....	42
Figure 2.15	The alleles of rs4846913 show consistent allelic differences in binding.....	43
Figure 2.16	The alleles at rs1555290 exhibit allele-specific protein binding.....	44
Figure 2.17	DNA sequence containing the 21-bp rs6143660 deletion allele shows evidence of protein binding	45
Figure 2.18	The alleles at rs2144300 show suggestive differential protein binding	46
Figure 2.19	Specificity of FOXO3 binding to the deletion allele of rs6143660.....	47
Figure 2.20	<i>GALNT2</i> expression in human hepatocytes from 50 individuals	48

Figure 2.21	The rs4846914 A Allele associated with increased HDL-C is associated with higher <i>GALNT2</i> RNA expression in primary human hepatocytes and subcutaneous adipose tissue.....	49
Figure 2.22	Non-coding variants at <i>GALNT2</i> are associated with <i>GALNT2</i> expression in adipose tissue samples from 1381 individuals from the METSIM study.....	50
Figure 2.23	An additional SNP, rs10864726, shows modest allelic differences in enhancer activity in the forward but not reverse orientation	52
Figure 2.24	siRNA-mediated knockdown of <i>CEBPB</i> in HepG2 cells results in a modest decrease in <i>GALNT2</i> expression	53
Figure 2.25	Multiple HDL-C-associated variants bind transcription factors at <i>GALNT2</i>	54
Figure 3.1	The A allele of rs11708067 is associated with decreased <i>ADCY5</i> expression in 112 human primary pancreatic islet samples.....	79
Figure 3.2	rs11708067 overlaps pancreatic islet open chromatin and a strong enhancer chromatin state in <i>ADCY5</i> intron 3.....	80
Figure 3.3	Evidence of allelic imbalance in H3K27ac CHIP-seq reads at rs11708067	82
Figure 3.4	rs11708067 exhibits allelic differences in transcriptional activity in MIN6 mouse insulinoma and 832/13 rat insulinoma cells	83
Figure 3.5	The alleles of rs11708067 show allelic differences in nuclear protein binding.....	84
Figure 3.6	rs11708067 may alter transcription factor binding at <i>ADCY5</i>	85
Figure 4.1	Open chromatin and histone modifications marking potential regulatory regions at the <i>PLXND1</i> WHR-associated locus.....	99
Figure 4.2	Four <i>PLXND1</i> candidate <i>cis</i> -regulatory elements show enhancer activity in HUVEC transcriptional reporter assays	100
Figure 4.3	Both alleles of rs11708067 in CRE 17 exhibit enhancer activity in HUVEC transcriptional reporter assays.....	102
Figure 4.4	CRE 17 shows enhancer activity in endothelial cells in zebrafish 120 hours post-fertilization	103
Figure 4.5	rs4488824 (in CRE 13) shows allelic differences in protein binding.....	104
Figure 5.1	Allele and haplotype differences in transcriptional enhancer activity are dependent on size of segment cloned into the luciferase reporter vector	115

LIST OF ABBREVIATIONS

AEI	allelic expression imbalance
<i>ADCY5</i>	Adenylate Cyclase 5; Adenylyl Cyclase 5
ATAC-seq	Assay for transposase-accessible chromatin sequencing
BMI	Body mass index
ChIP-seq	Chromatin immunoprecipitation sequencing
CRE	<i>Cis</i> -regulatory element
CVD	Cardiovascular disease
DNA	Deoxyribonucleic acid
DNase-seq	DNase I hypersensitivity sequencing
EMSA	Electrophoretic mobility shift assay
ENCODE	Encyclopedia of DNA Elements
eQTL	Expression quantitative trait locus
FAIRE	Formaldehyde-assisted isolation of regulatory elements
FUSION	Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus
<i>GALNT2</i>	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2
GIANT	Genetic Investigation of Anthropometric Traits
GWAS	Genome-wide association study
HDL-C	High-density lipoprotein cholesterol
HUVEC	Human umbilical vein endothelial cells
IPSC	Induced pluripotent stem cell
LD	Linkage disequilibrium
LDL-C	Low-density lipoprotein cholesterol
METSIM	Metabolic Syndrome in Men
NHGRI	National Human Genome Research Institute
<i>PLXND1</i>	Plexin D1
T2D	Type 2 diabetes
WHR	Waist-hip ratio

CHAPTER 1: INTRODUCTION

1.1 Cardiovascular disease

Cardiovascular disease (CVD) is a leading cause of death worldwide and a major public health burden[1, 2]. An estimated 86 million adults in the United States suffer from CVD, a term that encompasses numerous medical conditions, including high blood pressure, myocardial infarction, stroke, and heart failure[3]. An estimated 43.9% of the United States population is predicted to develop some form of CVD by 2030, with medical costs projected to be about \$918 billion[3, 4]. There are many major risk factors for CVD, including increased triglyceride and low-density lipoprotein cholesterol (LDL-C) levels, and decreased high-density lipoprotein cholesterol (HDL-C) levels[5-8]. Obesity and fat deposition also contribute to CVD risk[3, 9]. Increased visceral abdominal and extra-pericardial mediastinal fat deposition were found to be associated with an enhanced cardiovascular disease risk profile[10].

1.2 Lipoproteins and cholesterol

High total cholesterol levels increase an individual's risk of cardiovascular disease and other adverse metabolic traits[11]. HDL-C (high-density lipoprotein cholesterol) and LDL-C (low-density lipoprotein cholesterol) are two major components of total cholesterol levels. LDL-C is a major focus with regards to risk of cardiovascular disease[12]. LDL particles consist of a core of triglycerides and cholesteryl esters surrounded by phospholipids, free cholesterol and the protein apolipoprotein B-100[13] and can become oxidized and accumulate in arteries, where inflammation and atherosclerosis can result[14]. In contrast to LDL, HDL particles are smaller in size[15]. HDL-C levels are inversely associated with atherosclerotic cardiovascular disease[16, 17]. Like LDL, HDL consists of phospholipids, free cholesterol, apolipoproteins, and a core of triglycerides and cholesteryl esters, however, its major protein component is Apolipoprotein A-1[18]. Apolipoprotein A-1 facilitates reverse cholesterol transport, which may be a critical function in preventing atherosclerosis[19-23]. Reverse cholesterol transport

involves the efflux of cholesterol from macrophages in the arteries and cholesterol transport back to the liver[24-26].

1.3 Type 2 diabetes

Diabetes has also become a health epidemic within the United States. An estimated 21.1 million adults have diagnosed diabetes, with estimated medical costs at \$245 billion in 2012[2, 27]. Type 2 diabetes (T2D) is the primary type of diabetes in adults, consisting of 90-95% of diagnosed cases[2]. T2D is characterized by impaired fasting glucose and/or impaired glucose tolerance[28]. These characteristics can be the result of insulin resistance or impaired insulin secretion through pancreatic beta cell dysfunction[29]. Individuals with type 2 diabetes can also develop vascular and circulatory abnormalities that can lead to the development of cardiovascular disease[30]. Circulatory disorders associated with diabetes include coronary heart disease, congestive heart failure and stroke[31]. Fasting insulin levels were found to be associated with coronary heart disease incidence[32] and individuals with type 2 diabetes are at increased risk of coronary atherosclerosis[33] and death from coronary heart disease[34].

1.4 Obesity and body fat distribution

Obesity and excess body fat contribute type 2 diabetes and cardiovascular disease risk. An estimated 35% of adults in the United States are obese[2]. Obesity can occur with other comorbid conditions such as hypertension, hypercholesterolemia, and hypertriglyceridemia[35], all of which can increase risk of type 2 diabetes and cardiovascular disease[36]. Waist-hip ratio (WHR) a measurement of body fat distribution, also influences cardiometabolic traits. Increased WHR is a common anthropometric index for visceral fat accumulation[37]; WHR values of 0.9 to 1.0 in men and 0.8 to 0.9 in women have been thought to confer an increased risk of adverse metabolic risk factors[37-40]. Increased abdominal visceral adipose tissue (VAT) has been shown to be associated with higher triglyceride and lower HDL-C in men and women[41]. Additionally, increased expression of inflammatory genes[42] and macrophage infiltration[43] have been observed in omental (visceral) fat compared with subcutaneous fat. Elevated serum free fatty acids from visceral fat lipolysis can increase pancreatic insulin secretion and endothelial dysfunction, and decrease insulin sensitivity in muscle and liver[44]. Increased free fatty acids from

visceral fat can also contribute to increased hepatic triglycerides, nonalcoholic fatty liver disease and insulin resistance[45]. Studies have shown that insulin resistance, increased triglycerides, and central obesity tend to cluster together[46, 47] and contribute to a metabolic disease risk profile.

1.5 Genetics of cardiometabolic traits

While environment, lifestyle and diet can influence cardiovascular disease, blood lipid and cholesterol levels, type 2 diabetes and body fat distribution[2, 48, 49], there is also a heritable component to these traits and diseases[50-52]. Heritability estimates for blood lipid and cholesterol levels range from 40-90%[53-58]. Heritability estimates for type 2 diabetes range from 30-77%[59, 60] and heritability estimates for body fat distribution (as measured by WHR) range from 31-70%[40, 55, 61, 62].

Genome-wide association studies (GWAS) have identified over 150 loci associated with blood lipid and cholesterol levels[63] over 45 loci associated with WHR[64, 65] and over 70 loci associated with type 2 diabetes[66, 67]. GWAS provide an unbiased approach to identify candidate genes that may influence a particular trait or disease. Although GWAS have identified DNA variants and candidate genes associated with a trait or disease, they do not indicate which variants are functional. DNA variants inherited together and in strong linkage disequilibrium (LD) with the trait-associated variant may be responsible for driving the association signal, with the identified associated variant serving as a proxy[68]. Therefore, functional follow-up studies of association signals are necessary to determine the true functional variants and genes and to identify novel drug targets.

1.6 Identifying functional regulatory DNA variants from GWAS

Identification of the functional variants underlying an association signal from GWAS can help to characterize a molecular mechanism for how the variants may be acting on a gene. Further, it can lead to the better understanding of how genes play a role in a biological process or disease. In one example, a study at the *SORT1* locus identified a functional variant and molecular mechanism underlying association with LDL-C and myocardial infarction risk, and also elucidated novel biology by connecting *SORT1* gene function to LDL-C metabolism[69]. In another example, GWAS identified SNPs associated with HDL-C, LDL-C, triglyceride levels and myocardial infarction near the *TRIB1* gene[70], and subsequent studies of

Trib1 overexpression in mouse liver resulted in decreased lipid and cholesterol levels[71], providing a direct connection between gene function and CVD risk factors. Many DNA variants identified through GWAS are located within noncoding or intergenic regions of the genome[68], therefore, the variants may be located within enhancer or repressor elements that may ultimately alter gene expression, in contrast to altering the structure or function of a protein.

The ENCODE Project has numerous datasets to annotate functional genomic elements, including regulatory elements[72]. The NIH Roadmap Epigenomics Consortium has also created a resource that includes regulatory datasets for over 100 reference human epigenomes from a wide variety of cell types and tissues[73]. Both the ENCODE Project and NIH Roadmap Epigenomics Consortium resources include datasets that distinguish open chromatin regions that are devoid of histone proteins and are therefore more accessible to transcription factors. Two well-established methods for identifying open chromatin include DNase I hypersensitivity sequencing (DNase-seq) and formaldehyde-assisted isolation of regulatory elements (FAIRE) sequencing. DNase-seq involves DNase I enzyme digestion of DNA in regions devoid of histones (“open chromatin”) and subsequent sequencing[74]. FAIRE involves crosslinking histones to DNA and then isolating the open chromatin regions by phenol-chloroform extraction[75]. Both methods detect more active regulatory regions of DNA, including promoters and enhancers, which are more accessible to transcription factors. Chromatin immunoprecipitation sequencing (ChIP-seq) of histone modifications can also distinguish regulatory regions of DNA. Enhancer regions are often marked by H3K4me1 and/or H3K4me2 (mono- or di-methylation of lysine 4 on histone H3), while promoter regions typically are marked by H3K4me3[76]. Furthermore, active enhancers often can be distinguished by the histone modification H3K27ac[77]. Transcription factor ChIP-seq data can also be used to detect regulatory regions in which transcription factors are bound[78]. Open chromatin and histone modifications have been used to guide the identification of functional regulatory variants[79-85].

1.7 Aims and overview

The main aims of this dissertation are to identify the functional variants and regulatory elements at the HDL-C-associated *GALNT2* locus, the type 2 diabetes and glycemic trait-associated *ADCY5* locus,

and the WHR-associated *PLXND1* locus, and to further characterize the molecular mechanisms by which the functional variants act. The association signals at the *GALNT2*, *ADCY5* and *PLXND1* loci are all located within noncoding regions, therefore, I aimed to investigate whether any of the variants are located within regulatory regions and whether the variants show allelic differences in regulatory activity. At these three loci, open chromatin, histone modification ChIP-seq and transcription factor ChIP-seq data were considered in order to identify potential regulatory elements and variants.

In Chapter 2, the *GALNT2* HDL-C-association signal was further characterized by using fine-mapping and conditional analyses in 10,079 individuals from the METSIM study. I comprehensively evaluated the transcriptional regulatory activity of all candidate variants in strong linkage disequilibrium ($r^2 > 0.7$) with the lead HDL-C GWAS variant, rs4846914. For the variants that showed allelic or haplotype differences in transcriptional regulatory activity, I tested for evidence of differential protein binding to the alleles using electrophoretic mobility shift assays (EMSAs), and confirmed the identity of the factors bound to 2 of the regulatory variants using EMSA supershift and ChIP assays. I also evaluated whether using regulatory datasets (i.e. open chromatin, histone modification ChIP-seq and transcription factor ChIP-seq data) can help to distinguish variants that show allelic differences in regulatory activity from a set of candidate variants. Finally, I describe the direction of effect using allelic expression imbalance assays in primary hepatocytes and expression quantitative trait locus (eQTL) lookups in primary subcutaneous adipose samples to show that the rs4846914-A allele (associated with increased HDL-C) is associated with increased *GALNT2* expression.

In Chapter 3, I used open chromatin and chromatin state regulatory datasets in human pancreatic islets to identify a candidate functional regulatory variant, rs11708067, which was associated with type 2 diabetes and glycemic traits at the *ADCY5* locus. I tested and confirmed that this variant showed allelic differences in transcriptional activity in mouse and rat insulinoma beta cell lines, as well as allelic differences in protein binding in EMSAs. Additionally, H3K27ac ChIP-seq reads spanning this variant in a human pancreatic islet sample were evaluated for allelic imbalance. Finally, I describe the direction of effect using both new and previously published eQTL data and provide a proposed model in which the T2D risk- and glucose-raising-A allele of rs11708067 results in decreased *ADCY5* expression.

In Chapter 4, I describe the identification of regulatory elements at the *PLXND1* locus. GWAS identified a human WHR association signal that spans upstream of the *PLXND1* promoter and *PLXND1* intron 1. Considering its established role in vascular biology, visceral adipose tissue morphology and lipid storage, *PLXND1* is a strong candidate gene that may influence human body fat distribution. Through the use of complementary assays in human cell lines and zebrafish, I examine the regulatory activity of candidate regulatory elements at the *PLXND1* locus and show that one of these candidate elements shows regulatory enhancer activity in both HUVEC and in zebrafish endothelial cells.

In Chapter 5, I summarize the findings highlighted in Chapters 2, 3 and 4 and discuss the lessons learned from my functional follow-up studies on three GWAS signals for metabolic complex traits.

CHAPTER 2: MULTIPLE HEPATIC REGULATORY VARIANTS AT THE GALNT2 GWAS LOCUS ASSOCIATED WITH HIGH-DENSITY LIPOPROTEIN CHOLESTEROL¹

2.1 Introduction

Genome-wide association studies (GWASs) have identified more than 150 loci associated with blood lipid and cholesterol levels[63, 86-89]. One of the first novel GWAS signals for high-density lipoprotein cholesterol (HDL-C) levels in Europeans was reported for variants rs2144300 and rs4846914, located within intron 1 of *GALNT2*[70, 90] (MIM: 602274). These two lead GWAS variants are in perfect linkage disequilibrium (LD; $r^2 = 1$). The association signal for rs4846914 has been replicated ($n = 187,000$ individuals, $p = 4 \times 10^{-41}$) in subsequent studies with larger sample sizes[63, 87] and in Japanese and Mexican populations[91, 92]. HDL-C-associated variant rs4846914 is also associated ($n = 178,000$, $p = 7 \times 10^{-31}$) with triglycerides[63] and nominally associated ($n = 2,744-3,481$, $p < 0.05$) with large HDL concentration, low-density lipoprotein (LDL) size, HDL size, HDL-2 subfraction, and the ratio of total cholesterol to HDL-C[93]. The alleles associated with increased HDL-C are also nominally associated ($n = 84,068$, $p = 0.04$) with decreased risk of coronary artery disease[63].

According to 1000 Genomes phase 1 version 3 European (EUR) data[94] 24 variants exhibit strong LD ($r^2 > 0.7$) with the lead *GALNT2* HDL-C-associated SNP, rs4846914, and all 25 variants are noncoding. According to available chromatin data from the ENCODE Project[72] and Human Epigenome Atlas[73] these variants overlap many candidate regulatory regions. Therefore, we hypothesized that one or more of these variants regulate gene expression. Many variants identified through GWASs are located within noncoding or intergenic regions[68, 95] and variants at the *GALNT2* locus might also alter regulatory elements.

GALNT2, encoding UDP-N-acetylgalactosamine:polypeptide N-acetylgalactosaminyltransferase (GALNT2), is a reasonable positional candidate within the HDL-C association signal. GALNT2 is an

enzyme that transfers an N-acetylgalactosamine to serine or threonine residues in target proteins in the initial step of O-linked glycosylation[96]. *GALNT2* is expressed in many tissues, including liver, heart, lung, muscle, pancreas, ovary, and colon[97-99]. *GALNT2* might influence HDL-C levels by catalyzing O-glycosylation on target proteins that play a role in lipid metabolism. In vitro, *GALNT2* has been shown to O-glycosylate lecithin-cholesterol acyltransferase (LCAT), phospholipid transfer protein (PLTP), and angiopoietin-like protein 3 (ANGPTL3), and O-glycosylation of ANGPTL3 was shown to inhibit activation of this protein[100]. Individuals heterozygous for a *GALNT2* missense variant (c.941A>C [p.Asp314Ala]) shown to decrease *GALNT2* function in vitro exhibit decreased glycosylation of apoC-III and high (>95th percentile for age and gender) plasma HDL levels[101]. In mice, liver-specific *Galnt2* overexpression and knockdown have been shown to decrease and increase HDL-C levels, respectively[87].

We aimed to identify the functional regulatory variant(s) responsible for the *GALNT2* HDL-C GWAS signal by fine mapping the association with lipoprotein traits in the METSIM (Metabolic Syndrome in Men) study[102] and by examining a comprehensive set of candidate variants for evidence of allelic differences in enhancer function. We identified a single signal driven by at least two regulatory variants, rs4846913 and rs2281721, that exhibited binding of transcription factors and allelic differences in enhancer activity, as well as additional variants that might contribute to enhancer function. In human hepatocyte and subcutaneous adipose tissue samples, we observed an association between this GWAS signal and *GALNT2* expression. Together, these data show a consistent direction of regulatory effect in which increased expression of *GALNT2* is implicated in increased HDL-C.

2.2 Materials and Methods

2.2.1 Defining the candidate set of variants

We used the 1000 Genomes Project Phase 1 version 3 EUR dataset[94] including 24 variants in strong LD ($r^2 > 0.7$) with the lead HDL-C GWAS SNP rs4846914, to calculate LD. We used ENCODE data[72] available through the UCSC Genome Browser to determine which of the 25 total variants overlapped open-chromatin peaks and chromatin-immunoprecipitation-sequencing (ChIP-seq) peaks of histone modifications H3K4me1, H3K4me2, H3K4me3, H3K9ac, and H3K27ac and transcription factors in liver cell types (HepG2 cells, Huh-7 cells, and primary human hepatocytes) and ChromHMM chromatin

states[103] in multiple cell types. We used the Human Epigenome Atlas[73] to determine overlap with peaks of H3K4me1 and H3K9ac in primary adult liver and with ChromHMM chromatin states in multiple cell types and tissues.

2.2.2 Genotyping and imputation

We used the Illumina HumanOmniExpress and HumanCoreExome[104] Beadchips to genotype 10,134 Finnish men from the METSIM study[102]. Sample-level and SNP-level quality control included detecting sample contamination[105] confirming sex and relationships, and using principal-component analysis to detect population outliers. After we filtered SNPs to retain those with a call rate of >95% and Hardy-Weinberg Equilibrium $p > 10^{-6}$, we successfully analyzed 10,082 individuals and 681,803 SNPs. The METSIM study was approved by the ethics committee of the University of Kuopio and Kuopio University Hospital, and informed consent was obtained from all study participants. To impute ungenotyped SNPs, we used a panel of 5,474 reference haplotypes derived from genome sequences of 2,737 central-northern European individuals sequenced as part of the Genetics of Type 2 Diabetes study (C.F., J. Flannick, K.J.G., H. Kang, and the GoT2D Consortium, unpublished data). The minimum MaCH imputation quality score for the imputed variants was $R^2 = 0.971$. We used a two-step imputation strategy wherein individuals were pre-phased with ShapIT version 2 before imputation using Minimac[106]. We also used these data to verify LD proxies for allelic-expression-imbalance (AEI) assays.

2.2.3 Fine-mapping and conditional analyses

We analyzed 72 measures of lipid and lipoprotein particle-serum concentration obtained via nuclear-magnetic-resonance (NMR) metabolomics (65 traits) or enzymatic assays (7 traits) in up to 10,079 Finnish men. The NMR platform has been described previously[107, 108]. The methodology for measuring lipoprotein subclasses has been described previously[15] and subclasses are defined in Table 2.1. Trait values with skewed distributions were log transformed, and all traits were Winsorized at 5 SDs from the mean. After adjustment for age, squared age, smoking status, and lipid-lowering-medication status, we transformed residuals to a standard normal. We tested for association between normalized residuals and SNPs with a minor allele frequency (MAF) > 0.0005 (minor allele count > 10) by assuming

an additive genetic model and using a linear mixed model with an empirical kinship matrix to account for relatedness, as implemented in EMMAX[109]. We repeated the analysis while excluding individuals on lipid-lowering medication (2,844 participants), individuals with type 1 or 2 diabetes (1,420 participants), or both (3,560 participants) to assess sensitivity to these exclusions, and in each instance we obtained qualitatively similar results. To identify any additional independent signals in the region, we performed a conditional analysis by using the rs17315646 allele count as an additional covariate in the model. LocusZoom[110] plots were generated to include 2,079 variants in a 350-kb region surrounding the lead variant rs17315646 and spanning *GALNT2*.

2.2.4 Cell culture

Human HepG2 hepatocellular carcinoma cells (ATCC, HB-8065) were grown in Eagle's minimum essential medium (MEM) alpha supplemented with 10% fetal bovine serum (FBS) and 1 mM sodium pyruvate at 37°C and 5% CO₂. Human Huh-7 hepatocellular carcinoma cell lines (JCRB0403, Japanese Collection of Research Bioresources Cell Bank, National Institute of Biomedical Innovation) were grown in DMEM supplemented with 10% FBS, 1 mM sodium pyruvate, 1× MEM non-essential amino acids, and 2 mM L-glutamine. HepG2 and Huh-7 cells were seeded into 24-well plates (100,000 cells per well) 1 day prior to transfection experiments.

2.2.5 Transcriptional reporter assays

To test candidate variants and haplotypes for allele-specific effects on transcriptional activity, we amplified segments of 109–262 bp (for single variants) and 349–780 bp (for multiple variants) from DNA of individuals homozygous for each allele or haplotype. Segment size was chosen to include the ~147 nucleotides of DNA spanning one nucleosome. However, because of the proximity of other candidate variants, larger segments of 349–780 bp were designed to include multiple variants. Primer sequences are listed in Table 2.2. Amplicons were cloned into the KpnI and XhoI restriction sites of the firefly luciferase transcriptional reporter vector pGL4.23 in both forward and reverse orientations with respect to the minimal promoter (Promega). Three to seven independent plasmids for each allele or haplotype were isolated and confirmed by sequencing. When additional variants were identified, we selected clones for

which the alleles matched at these variants. For simplicity of presentation, Figures 2.4, 2.6 and 2.7 do not show the differing alleles of rs1555290; these data are provided in Figures 2.8 and 2.9. We then transfected each purified clone into HepG2 or Huh-7 cells in duplicate (720 ng in each well) by using FuGENE 6 (Promega) and Opti-MEM (Life Technologies). To control for transfection efficiency, we co-transfected a phRL-TK *Renilla* luciferase reporter vector (80 ng in each well) into cells. For empty-vector controls, two independent preparations of empty vector were each transfected into HepG2 or Huh-7 cells in duplicate. After 48 hr, cell-lysate luciferase activity was measured with the Dual-Luciferase Reporter Assay System (Promega), normalized, and compared to readings for empty-vector controls. These control readings were very similar for the two independent preparations. We performed two-tailed t tests to compare the luciferase activities between variant alleles. For comparisons of multiple haplotypes, we performed ANOVA and Tukey's post hoc tests by using JMP 10.0.1 software (SAS Institute). We also assessed the transcriptional activity of the 780-bp segment by cloning the haplotype into the promoterless pGL4.10 firefly luciferase reporter vector (Promega). To examine individual variant effects within the 780-bp segment, we altered alleles by using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) and confirmed them by sequencing.

2.2.6 Electrophoretic mobility shift assays

Complementary DNA oligonucleotides (17–19 bp) centered on variant alleles (Table 2.2) were synthesized by Integrated DNA Technologies. The rs6143660 insertion allele consisted of 39-bp complementary oligonucleotides containing the 21-bp insertion (9 bp flanking each side of the insertion), and the rs6143660 deletion allele consisted of 18-bp complementary oligonucleotides (9 bp flanking each side of the deleted sequence). Labeled oligonucleotides included biotin on the 5' end. We performed assays as previously described[83] by using 3.5–6 µg of HepG2 nuclear lysate and 30- to 300-fold excess unlabeled probe. For supershift reactions, 4–8 µg CEBPB antibody (sc-150X), USF1 antibody (sc-229X), and FOXO3 antibody (sc-34895X), all from Santa Cruz Biotechnology, were incubated with binding buffer, poly(dI-dC), and HepG2 nuclear lysate for 20 min at room temperature before the addition of labeled DNA probes and incubation. Additional control antibodies (4–6 µg, all from Santa Cruz Biotechnology) were chosen on the basis of transcription-factor binding motifs, ENCODE ChIP-seq peaks, expression of

transcription factors in liver, or plausible roles of the factors in cholesterol metabolism. These included antibodies to ARNT (sc-271801X), SF1 (sc-10976X), HNF4A (sc-6556X), RXRA (sc-553X), CEBPA (sc-61X), CEBPB (sc-150X), NR1H3 (sc-1202X), and MAX (sc-765X). Reactions were loaded into a 6% DNA retardation gel (Life Technologies), subjected to electrophoresis, transferred to Biodyne B nylon membranes (Life Technologies), and UV crosslinked. Wash and detection steps were performed according to instructions in the Chemiluminescent Nucleic Acid Detection Module (Life Technologies). Experiments involving electrophoretic mobility shift assays (EMSAs) were repeated on 2–7 separate days, and all had consistent results.

To predict transcription-factor binding sites, we searched databases for transcription-factor binding-site motifs in 17- to 21-bp genomic sequences containing each allele of candidate variants. For JASPAR[111] we searched all available matrix models with a relative profile-score threshold of 80%. We also searched positional-weight matrices (PWMs) from vertebrates in TRANSFAC by using default parameters in the TRANSFAC Professional's Match tool and by using PWM-SCAN[112].

2.2.7 ChIP assays

We used a TaqMan SNP Genotyping Assay (Life Technologies) to genotype HepG2 and Huh-7 cells at rs4846913. Cells were crosslinked with 1% formaldehyde, and glycine was added to stop fixation. Fixed cells were resuspended in SDS lysis buffer (1% SDS, 10 mM EDTA, and 50 mM Tris [pH 8.1]), diluted with immunoprecipitation (IP) buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris [pH 8.1], and 167 mM NaCl), and sonicated on ice with a Branson sonifier for the generation of 100- to 500-bp DNA fragments. Each CEBPB IP or immunoglobulin G (IgG) reaction used two to three million cells. After preclearing with Protein A Agarose beads (sc-2001, Santa Cruz Biotechnology), 10 µg of CEBPB antibody (sc-150X, Santa Cruz Biotechnology) or 10 µg of normal rabbit IgG (sc-2027, Santa Cruz Biotechnology) was added to HepG2 and Huh-7 cell lysates and incubated overnight. Protein A Agarose beads were then added for 3 hr and washed separately with low-salt buffer, high-salt buffer, LiCl buffer, 10 mM Tris-EDTA, and elution buffer (1% SDS and 0.1 M NaHCO₃). The crosslinks were then reversed by the addition of 5 M NaCl and overnight incubation. Samples were incubated with 20 µl 1M Tris, 10 µl 0.5 M EDTA, and 0.03 mg proteinase K. DNA was purified by phenol-chloroform extraction and

ethanol precipitation. For qPCR, a 120-bp region spanning rs4846913 was amplified with FAST SYBR Green Master Mix (Life Technologies). qPCR was performed in triplicate and quantified with respect to a standard curve generated from sonicated HepG2 DNA standards. The mean quantities of CEBPB IP and IgG control sample were normalized to input HepG2 or Huh-7 DNA. CEBPB and IgG ChIP experiments with HepG2 and Huh-7 cells were each performed twice and had similar results. We used two-tailed t tests to compare CEBPB enrichment of the rs4846913 region in HepG2 and Huh-7 cells.

We performed USF1 ChIP assays similarly, except that we used a Diagenode Bioruptor Standard sonicator. We used 10 µg of USF1 antibody (sc-229X, Santa Cruz Biotechnology) or 10 µg IgG (sc-2027, Santa Cruz Biotechnology) and purified DNA with the QIAquick PCR Purification Kit (QIAGEN) to amplify and quantify a 164-bp region spanning rs2281721. We performed USF1 and IgG ChIP experiments with HepG2 and Huh-7 cells twice each, and they had similar results.

2.2.8 Allelic-imbalance analysis of sequence reads from CEBPB ChIP-seq and DNase-seq

Sequence reads from HepG2 for DNase-I-hypersensitivity-site sequencing (DNase-seq) and CEBPβ ChIP-seq experiments generated by the ENCODE Consortium[72] were aligned to UCSC Genome Browser build hg19 with AA-ALIGNER[113]. In brief, HepG2 genotypes were obtained from the Illumina Human-1M-Duo BeadChip array genotyped at HudsonAlpha Institute of Biotechnology, and imputation was performed with the 1000 Genomes Phase 1 EUR reference panel[94]. Using this data, we verified that HepG2 is diploid in the chromosome 1 region containing *GALNT2*, and we created a personalized HepG2 reference genome containing non-reference alleles for sites at which HepG2 is homozygous for the non-reference allele. We aligned sequence reads to the personalized genome by using HepG2 heterozygous sites identified by imputation and GSNAP[114] with the following parameters: -m 1, -k 11, --basesize = 11, -sampling = 1, -terminal-threshold = 10, -n 1, -query-unk-mismatch = 1, -genome-unk-mismatch = 1, -trim-mismatch-score = 0, -t 7, and --A sam. We filtered the alignments to remove sequences aligned to more than one genomic location, sequences aligned to regions underrepresented in the reference sequence (ENCODE blacklisted[72] regions), and duplicate reads that might represent PCR artifacts. We determined the significance of allelic imbalance at rs4846913 by using an exact binomial test, based on the number of reads containing the reference allele and the total number

of reads at the heterozygous site. For the HepG2 DNase-seq data and USF1 HepG2 ChIP-seq data, insufficient reads were available for allelic imbalance analysis at rs2281721.

2.2.9 Preparation of cDNA for *GALNT2* mRNA-expression and AEI assays

RNA and DNA were isolated from hepatocyte samples of 50 individuals as described previously[115] RNA for each sample was treated with DNase I with the DNA-free Kit (Life Technologies) and added to a final concentration of 24 ng/μl in an RT-PCR reaction in the SuperScript III First-Strand Synthesis System (Life Technologies), which includes both oligo(dT)₂₀ and random hexamer primers. The synthesized cDNA for each sample was then diluted with diethylpyrocarbonate (DEPC)-treated water for use in mRNA-expression and AEI assays.

2.2.10 AEI assays

We used a TaqMan SNP genotyping assay (Life Technologies) to genotype human hepatocyte genomic DNA (gDNA) from ADMET Technologies for the HDL-C index SNP rs4846914. To quantify allele-specific expression, we diluted hepatocyte cDNA and gDNA from 36 individuals heterozygous for rs4846914 and performed subsequent qPCR reactions in triplicate. To generate a standard curve, we mixed gDNAs from samples homozygous for each rs4846914 allele in the following ratios: 95:5, 72.5:27.5, 61.25:38.75, 50:50, 38.75:61.25, 27.5:72.5, and 5:95. We generated a standard curve by plotting the quantity of one allele against the difference between the cycle-threshold (Ct) values of the two alleles. For each heterozygous sample, we estimated the expression percentage of one allele by using the difference between the mean Ct values of the alleles and the standard curve[115]. We used two-tailed t tests to compare gDNA and cDNA values and used F-tests to determine equal or unequal variance between gDNA and cDNA samples[115, 116].

2.2.11 *GALNT2* hepatocyte mRNA expression

We measured expression of *GALNT2* in 50 human hepatocyte samples by qPCR with a standard curve and FAST SYBR Green Master Mix (Life Technologies). We performed triplicate qPCR reactions including 9 ng of total cDNA in each well, Taqman Gene Expression Master Mix (Life Technologies), and

primer-set sequences within *GALNT2* exons (Table 2.2). *GALNT2* expression values were normalized to the expression of beta-2-microglobulin (*B2M* [MIM: 109700]), natural-log transformed, and plotted according to the rs4846914 genotype (AA, AG, and GG). Using linear regression and an additive model including sex, ancestry, and age as covariates, we tested for association between the level of *GALNT2* mRNA and the rs4846914 genotype.

2.2.12 Lookup of expression quantitative trait loci in subcutaneous adipose tissue

We looked for evidence of association between the HDL-C GWAS variant region and gene expression by using preliminary microarray expression data from subcutaneous adipose tissue from the METSIM study. In brief, we used EPACTS-multi (with EMMAX implemented to account for family relatedness) Affymetrix Human Genome U219 Array data for 1381 individuals was adjusted for 40 confounding factors[117] inverse-normal transformed, and tested for association with variants. We analyzed variants within 1 Mb of rs4846914 and identified one probe set for *GALNT2* and other genes located in this region. Subsequent reciprocal conditional analyses included two variants in each model.

2.2.13 siRNA-mediated knockdown of *CEBPB*

HepG2 cells (80,000 per well) were plated into 24-well collagen-coated plates and then treated with 50 nM *CEBPB* Silencer Select small interfering RNA (siRNA; s2893, Thermo Fisher Scientific) or Silencer Select Negative Control No. 1 siRNA (4390843, Thermo Fisher Scientific) with Opti-MEM (Life Technologies) and DharmaFECT1 transfection reagent (GE Healthcare Life Sciences) on the following day. Cells were incubated at 37°C with 5% CO₂ for 48 hr, and HepG2 medium was changed the day following siRNA transfection. RNA was harvested with the RNeasy Plus Mini Kit (QIAGEN), and cDNA was prepared with the SuperScript III First-Strand Synthesis System (Life Technologies). *CEBPB* and *GALNT2* expression was measured by qPCR with a standard curve. Raw expression values were normalized to the expression of *B2M*, and expression percentages were calculated by comparison to expression values in HepG2 cells transfected with the negative control siRNA. Primers for gene expression are listed in Table 2.2. We used two-tailed t tests to compare normalized *GALNT2* expression between HepG2 cells treated with *CEBPB* siRNA and HepG2 cells treated with negative control siRNA.

2.3 Results

2.3.1 Fine mapping shows evidence of a single association signal strongest for total cholesterol in medium HDL

We tested one of the HDL-C-associated lead GWAS variants, rs2144300 ($r^2 = 1$ with rs4846914), for association with 72 lipid and cholesterol traits in up to 10,079 Finnish men from the METSIM study. The strongest evidence of association across all 72 traits was with total cholesterol in medium HDL ($n = 9,810$, $\beta = 0.10$, $p = 5.3 \times 10^{-12}$; Table 2.3). Using this trait and variants that were directly genotyped or imputed from a reference panel of 2,737 genomes of central-northern European individuals, we performed fine-mapping association and conditional analyses. Among 2,079 total variants in the region spanning 350 kb surrounding the lead variant, the strongest evidence of association was observed for rs17315646 ($n = 9,810$, $p = 3.5 \times 10^{-12}$), which is in perfect LD ($r^2 = 1$) with HDL-C GWAS lead SNPs rs4846914 ($n = 9,810$, $p = 5.3 \times 10^{-12}$) and rs2144300 ($n = 9,810$, $p = 5.3 \times 10^{-12}$). Conditioning on rs17315646 attenuated the 13.7-kb association signal (all $p > 0.01$; Figure 2.1 and Table 2.4). The new lead variants after conditional analysis ($p_{\text{cond}} \sim 10^{-4} - 10^{-6}$) showed weak initial evidence of association ($p \sim 10^{-2} - 10^{-4}$) and have a MAF < 0.04 (Table 2.4). These data provide evidence that common variants in strong LD with rs4846914 and rs17315646 are most likely responsible for the HDL-C association signal. All 25 candidate variants ($r^2 > 0.7$ with HDL-C GWAS lead variant rs4846914) are located within intron 1 of *GALNT2*. These variants all exhibited strong evidence of association with total cholesterol in medium HDL ($p \leq 1.45 \times 10^{-9}$; Table 2.4).

2.3.2 Open-chromatin, histone-modification, and transcription-factor marks indicate potential regulatory elements overlapping the *GALNT2* association signal

We hypothesized that one or more of the 25 HDL-C-associated non-coding variants affect transcriptional activity. We asked whether regulatory datasets from the ENCODE Project and Human Epigenome Atlas could help us identify variants that exhibit regulatory activity. We compared the location of the 25 candidate variants to regions of open chromatin depicted by DNase I hypersensitivity and formaldehyde-assisted isolation of regulatory elements (FAIRE); histone-modification ChIP-seq peaks H3K4me1, H3K4me2, H3K4me3, H3K9ac, and H3K27ac, which often mark enhancer or promoter

elements; and transcription-factor ChIP-seq peaks. On the basis of the liver's key role in HDL synthesis and transport, and the enrichment of lipid GWAS signals in liver[63] we focused on datasets from human liver cell lines, primary human hepatocytes, and human adult liver cells. All 25 variants overlap broad patterns of H3K4me1-enriched domains in HepG2 cells (Table 2.5), and 13 overlap narrower H3K4me1-enriched peaks[118]. Sixteen variants overlap two or more histone-modification peaks, and six variants (rs4631704, rs4846913, rs2144300, rs6143660, rs2281721, and rs11122453) overlap at least one transcription-factor ChIP-seq peak (Figure 2.2 and Table 2.5). Three variants (rs4846913, rs2144300, and rs6143660) overlap the most marks of potential regulatory function (at least 22 peaks, including open-chromatin, histone-modification, and transcription-factor peaks). According to ChromHMM data[103] from ENCODE[72] and the Roadmap Epigenomics Project[73] all 25 variants are located within predicted enhancer or transcribed-region chromatin states in HepG2 cells and primary liver cells (Figure 2.3). The regulatory peaks are not specific to liver cells, given that open-chromatin, histone-modification, and transcription-factor peaks were also observed in GM12878 lymphoblastoid cells, human umbilical-vein endothelial cells (HUVECs), K562 leukemia cells, CD20⁺ B cells, CD14⁺ monocytes, human skeletal-muscle myoblasts (HSMs), normal human astrocytes (NHAs), normal human lung fibroblasts (NHLFs), and osteoblasts (Figure 2.3).

2.3.3 All 25 candidate regulatory variants were evaluated for allelic differences in luciferase activity

Given the limited resolution of open-chromatin peaks and histone-modification and transcription-factor ChIP-seq peaks, and the knowledge that these marks do not predict allelic differences in regulatory activity, we tested all 25 variants in transcriptional reporter assays in luciferase vectors containing a minimal promoter in HepG2 cells (Figures 2.4 and 2.5). Tested elements contained one to four variants, and three to seven independent clones for each allele or haplotype in the element were tested. We considered elements exhibiting >1.5-fold more activity than empty-vector controls as enhancers. We focused on elements whose enhancer activity in both forward and reverse orientations was >1.5-fold higher than that in the luciferase reporter gene alone, as well as differences ($p < 0.05$) between the alleles or haplotypes in both orientations.

2.3.4 rs2281721 exhibits allelic differences in transcriptional activity

Among all the segments tested, a 154-bp DNA segment containing rs2281721 showed the strongest enhancement of luciferase activity, and this activity also differed between the alleles (Figure 2.6A). In the forward orientation, the DNA segment containing the rs2281721 T allele, associated with increased HDL-C, showed 75-fold more luciferase activity than the empty-vector control, whereas the segment containing the rs2281721 C allele showed 27-fold more, and significant differences were observed between the alleles ($p = 2.4 \times 10^{-6}$). In the reverse orientation, the T and C alleles exhibited 37-fold and 14-fold, respectively, more luciferase activity than did the control, and there were significant differences between the alleles ($p = 5.8 \times 10^{-5}$). The rs2281721 T allele also showed stronger enhancer activity than did the rs2281721 C allele in both forward (205-fold versus 59-fold more than in the control) and reverse (49-fold versus 19-fold more than in the control) orientations in a second hepatocellular carcinoma cell line, Huh-7 (both $p < 3 \times 10^{-4}$; Figure 2.7A). The segment and the position of rs2281721 overlaps H3K4me1 peaks in HepG2 cells and adult liver; H3K4me2, H3K9ac, and H3K27ac peaks in HepG2 cells; and a USF1 transcription-factor ChIP-seq peak in HepG2 cells (Figure 2.2 and Table 2.5).

2.3.5 A segment containing rs4846913, rs2144300, and rs6143660 shows haplotype differences in transcriptional activity

A 780-bp DNA segment containing three variants exhibited significant haplotype differences in luciferase activity (Figure 2.6B). We analyzed two haplotypes of rs4846913, rs2144300, and rs6143660 (a 21-bp indel): AT⁻ (containing the alleles associated with increased HDL-C) and CC⁺ (containing the alleles associated with decreased HDL-C). In the forward orientation, the AT⁻ and CC⁺ haplotypes showed 49-fold and 16-fold, respectively, more luciferase activity than did the empty-vector control, and significant differences were observed between the haplotypes ($p = 1.8 \times 10^{-5}$). In the reverse orientation, the AT⁻ and CC⁺ haplotypes exhibited 11-fold and 4-fold, respectively, more luciferase activity than did the empty-vector control, and significant differences were observed between the haplotypes ($p = 2.0 \times 10^{-4}$). The three HDL-C-associated variants rs4846913, rs2144300, and rs6143660 (a 21-bp indel) each overlapped ≥ 22 open-chromatin, histone-modification, and transcription-factor peaks (Figure 2.2 and Table 2.5).

Haplotype effects on transcriptional activity were similar in Huh-7 cells. In the forward orientation, haplotypes AT⁻ and CC⁺ showed 36-fold and 13-fold, respectively, more luciferase activity than did the empty-vector control; the differences observed between the haplotypes were significant ($p = 1.4 \times 10^{-5}$ Figure 2.7B). In the reverse orientation, haplotypes AT⁻ and CC⁺ showed 12-fold and 6-fold, respectively, more luciferase activity than did the empty-vector control ($p = 0.08$, Figure 2.7B). The direction of effect was the same as that of rs2281721; alleles associated with increased HDL-C showed increased luciferase activity.

Taken together, our data show consistent haplotype differences in luciferase activity for a 780-bp segment consisting of rs4846913, rs2144300, and rs6143600. The results suggest that these variants are located within an enhancer element that can affect transcription and that one or more of them might have an allelic effect on transcriptional activity.

2.3.6 The haplotype variants act together to regulate enhancer activity

We then created additional haplotypes of the 780-bp segment by performing site-directed mutagenesis to investigate the variant responsible for allelic differences in transcriptional activity. In addition to including the candidate variants rs4846913, rs2144300, and rs6143660, the segment included an additional common variant, rs1555290, in moderate LD with rs4846914 ($D' = 1$, $r^2 = 0.26$). We analyzed and tested natural haplotypes of rs4846913, rs2144300, rs1555290, and rs6143660 (ATA⁻, CCC⁺, and CCA⁺), as well as constructed haplotypes CTA⁻, ACA⁻, CCA⁻, and CCC⁻. These seven haplotypes were then tested separately in luciferase assays (Figure 2.8). ATA⁻, CTA⁻, ACA⁻, and CCA⁻ haplotypes showed 44-fold, 38-fold, 31-fold, and 23-fold, respectively, more luciferase activity than did the empty-vector control, suggesting that both of the first two variants, rs4846913 and rs2144300, contribute to haplotype differences in transcriptional activity. Specifically, significant differences were observed between the ATA⁻ and ACA⁻ haplotypes ($p = 0.04$), between the CTA⁻ and CCA⁻ haplotypes ($p = 0.04$), and between the ATA⁻ versus CCA⁻ haplotypes ($p = 0.001$), supporting a contribution from both rs4846913 and rs2144300 to haplotype differences in luciferase activity. Compared to the control, the CCA⁺, CCC⁻, and CCC⁺ haplotypes showed similar 12- to 14-fold increases in luciferase activity ($p > 0.05$; an intermediate between the CCA⁻ haplotype and the empty vector), suggesting that rs1555290

and/or rs6143660 might also contribute to increased transcriptional activity. We observed similar results in Huh-7 cells: the ATA⁻, ACA⁻, and CCA⁻ haplotypes showed 36-fold, 30-fold, and 27-fold, respectively, more luciferase activity than did the empty-vector control, and significant differences were observed between the ATA⁻ and CCC⁻ haplotypes ($p = 0.0001$). Compared to the empty-vector control, the CCA⁺, CCC⁻, and CCC⁺ haplotypes showed similar ($p > 0.05$) 13- to 15-fold increases in luciferase activity (Figure 2.9). Taken together, these data suggest a role for at least two variants, rs4846913 and rs2144300, in haplotype differences in enhancer activity.

Because the SNPs in the 780-bp segment overlap H3K1me1 and H3K4me2 peaks, which are frequently present in enhancer regions, as well as a H3K4me3 peak, which is often found at promoters, we also evaluated the haplotypes in a promoterless vector (Figure 2.10). In the forward orientation, haplotypes ATA⁻, CCA⁺, and CCC⁺ showed 44-fold, 27-fold, 16-fold, respectively, more luciferase activity than did the empty-vector control, and significant differences were observed between all haplotype comparisons ($p < 0.001$). In the reverse orientation, haplotypes ATA⁻, CCA⁺, and CCC⁺ exhibited 12-fold, 9-fold, and 5-fold, respectively, more luciferase activity than did the empty-vector control, and significant differences were observed between all the haplotype comparisons ($p < 0.04$).

We subsequently analyzed the four variants individually in 100- to 200-bp DNA segments. Of the four variants tested individually, only the element containing rs2144300 exhibited enhanced luciferase activity (average 6.5-fold more than in the empty-vector control). Almost none of these segments showed allelic differences ($p > 0.1$; Figure 2.11), but rs4846913 did show allelic differences in only the forward orientation ($p = 0.03$). These data suggest that the larger segment is necessary for observing allelic differences in enhancer activity in this assay.

2.3.7 USF1 binds to rs2281721

To investigate whether transcription factors bind differentially to rs2281721, rs4846913, rs2144300, rs1555290, and rs6143660, we performed EMSAs with HepG2 nuclear lysate. The rs2281721 C probe showed stronger protein binding than the rs2281721 T probe (lane 2 versus 7, arrow, Figure 2.12A). The addition of 40-fold excess unlabeled rs2281721 C probe competed away the signal more effectively than did unlabeled rs2281721 T probe (lane 3 versus 4). rs2281721 overlaps a

USF1 ChIP-seq peak from ENCODE data in HepG2 cells, as well as a predicted USF1 motif. In EMSAs, the addition of a USF1 antibody resulted in a disruption of the band observed with the C allele (lane 5). As negative controls, we tested ARNT and SF1 antibodies; we did not observe disruption of the rs2281721 C allele band (Figure 2.13). These data provide evidence supporting USF1 binding to the C allele of rs2281721.

To validate the USF1 binding in a native chromatin context, we performed ChIP assays in both HepG2 (rs2281721 genotype T/C) and Huh-7 (rs2281721 genotype C/C) cells. The ChIP assays provided evidence supporting USF1 binding to a 164-bp DNA region spanning rs2281721; however, this binding was not allele specific (Figure 2.12B).

2.3.8 CEBPB binds differentially to the alleles of rs4846913

We also observed more protein binding to the A allele than to the C allele of rs4846913 (lane 2 versus 7, arrow, Figures 2.14A and 2.15). We observed a greater decrease in band intensity upon addition of 63-fold excess unlabeled rs4846913 A probe than upon addition of 63-fold excess unlabeled rs4846913 C probe, suggesting that competition of the lane 2 band with unlabeled rs4846913 A probe is more effective than competition with the unlabeled rs4846913 C probe (lane 3 versus 4). rs4846913 overlaps CEBPB ChIP-seq peaks from ENCODE data in HepG2 cells and a predicted CEBPB binding motif. Incubation of the EMSA reactions with a CEBPB antibody generated a strong supershift band for the A allele and a detectable supershift band for the C allele (lane 5 versus 10, Figure 6A). As negative controls, we tested HNF4A or RXRA antibodies; we did not observe evidence of supershifts (Figure 2.13).

To validate the differential CEBPB binding in a native chromatin context, we performed ChIP assays in both HepG2 (rs4846913 genotype A/C) and Huh-7 (rs4846913 genotype C/C) cells. A 120-bp region of DNA containing rs4846913 showed 31-fold more CEBPB binding in HepG2 cells than in Huh-7 cells ($p = 0.006$; Figure 2.14B). These results are consistent with the EMSA result of increased CEBPB binding to rs4846913 A.

The CEBPB ChIP-seq signal at rs4846913 was sufficiently strong to permit an analysis of allelic imbalance reflecting in vivo CEBPB binding. We re-aligned HepG2 CEBPB ChIP-seq reads and HepG2 DNase-seq reads by using an allele-aware approach to avoid reference-allele bias (see Material and

Methods). Among CEBPB reads spanning rs4846913, 57 of 78 (73%) contained the rs4846913 A allele (binomial $p = 5.6 \times 10^{-5}$; Figure 2.14C). Among DNase-seq reads spanning rs4846913, 31 of 83 (37%) contained the rs4846913 A allele (binomial $p = 0.03$; Figure 2.14C), suggesting that CEBPB might help protect the DNA sequence containing the A allele from being accessible to the DNase I enzyme. Taken together, the EMSA, CHIP, and allelic imbalance in CHIP-seq and DNase-seq reads all show consistent evidence suggesting stronger CEBPB binding to the A allele of rs4846913.

2.3.9 rs2144300, rs1555290, and rs6143660 also show suggestive evidence of protein binding

EMSAs using HepG2 nuclear lysate for the other variants in the four-variant haplotype (rs2144300, rs1555290, and rs6143660) showed suggestive evidence of differential allelic protein binding (Figures 2.16, 2.17, 2.18). The rs1555290 C probe showed a reproducible band (arrow, Figure 2.16), and the addition of 30-fold excess unlabeled rs1555290 C probe competed away the signal (lane 7) more effectively than did the addition of 30-fold excess unlabeled rs1555290 A probe (lane 8). The rs6143660 deletion allele showed multiple protein-binding bands (lane 2, Figure 2.17) that were altered by incubation with FOXO3 antibody (Figure 2.19), although it is challenging to interpret results when using EMSA probes of different sequence lengths to compare the 21-bp-insertion and 21-bp-deletion alleles. However, we did not observe any disruption of the band corresponding to the rs6143660 deletion allele when we tested additional antibodies CEBPA, CEBPB, HR1H3, MAX, and RXRA (Figure 2.19). A weak, reproducible, allele-specific band for rs2144300 C was observed, but it was not fully competed away with excess unlabeled rs2144300 C probe (Figure 2.18). Overall, these EMSA data suggest that these other three variants in the haplotype might exhibit differential transcription-factor binding.

2.3.10 Variants associated with *GALNT2* expression

On the basis of the location of the HDL-C-associated variants within *GALNT2* intron 1, we hypothesized that these variants might be acting in *cis* to regulate *GALNT2* expression. We assessed *GALNT2* expression stratified by genotype of rs4846914 ($r^2 = 1$ with rs4846913) in 50 primary human hepatocyte samples, and we observed a trend toward association between genotype and *GALNT2* expression ($p = 0.08$; Figure 2.20). To increase the sensitivity by removing the sample-specific

contribution of environmental factors, we performed AEI assays in 36 human hepatocyte samples heterozygous for rs4846914. Of 142 variants in $r^2 > 0.2$ with rs4846914 in METSIM, none are transcribed, so we used qPCR to measure relative allelic cDNA levels from *GALNT2* pre-mRNA at rs4846914. Previous studies have similarly used intronic SNPs in unspliced RNA to measure allelic expression [119, 120]. The rs4846914 A allele, which is associated with increased HDL-C, showed higher *GALNT2* cDNA expression ($p = 5.4 \times 10^{-7}$; Figure 2.21A). When normalized to the allelic difference detected in heterozygous genomic DNA, the A allele of rs4846914 showed a 7.4% increase in expression. Finally, we examined rs4846914 in preliminary expression-quantitative-trait-locus (eQTL) data from subcutaneous adipose tissue from 1,381 individuals in the METSIM study (M.C. and Y.W., unpublished data) and observed consistent association between the rs4846914 A allele and increased expression of *GALNT2* ($p = 2.2 \times 10^{-14}$; Figure 2.21B), but not expression of 16 other potential target genes ($p > 0.05/17 = 0.003$ on the basis of 17 tests; Figure 2.22 and Table 2.6). Conditioning on the lead eQTL variant, rs4846922, attenuated the association signal with *GALNT2* expression ($p > 0.2$; Figure 2.22 and Table 2.7).

2.4 Discussion

We identified multiple functional regulatory variants that contribute to the single strong HDL-C GWAS signal at *GALNT2*. Variants rs2281721 and rs4846913, located 2 kb apart, showed strong and consistent allelic and haplotype differences in enhancer activity and exhibited transcription-factor binding of USF1 and CEBPB, respectively. Variants rs2144300 and rs6143660 showed moderate effects on transcriptional activity and suggestive allelic differences in binding of nuclear proteins, although we did not confirm the identity of these potential regulatory proteins. In addition, rs1555290, in moderate LD ($r^2 = 0.26$) with the GWAS variants, also showed suggestive haplotype differences in enhancer activity and allelic differences in binding of a nuclear protein. For all of these variants, the alleles associated with increased HDL-C showed higher transcriptional activity in reporter assays and were associated with higher expression of *GALNT2* in human hepatocyte and human subcutaneous adipose tissue samples. By testing all variants in strong LD ($r^2 > 0.7$) with lead HDL-C GWAS variant rs4846914 in luciferase reporter assays in both forward and reverse orientations, we identified the most likely functional regulatory candidates. Although we used a threshold of $r^2 > 0.7$ to identify the most likely functional

candidates, we cannot rule out the possibility that we could have missed additional functional variants in weaker LD with the lead HDL-C-associated variant. Of 14 DNA segments tested, we observed that a segment containing four variants and a separate segment containing rs2281721 showed up to 49-fold and 75-fold, respectively, more enhancer activity than did the empty-vector control. One additional element containing one SNP, rs10864726, showed more-moderate allelic differences in enhancer activity (2.5-fold [T allele] versus 1.8-fold [C allele] more than the empty vector, $p = 0.003$) in only the forward orientation (Figure 2.23). Compared to a vector control, the remaining 11 segments did not exhibit enhancer activity in both orientations, nor did they exhibit significant differences between alleles. These 11 segments included one of the lead GWAS SNPs, rs4846914, and the lead SNP from our fine-mapping analysis, rs17315646. These two SNPs are located in regions containing H3K4me1, H3K4me2, H3K4me3, H3K9ac, and H3K27ac peaks in HepG2 cells and H3K4me1 peaks in adult liver cells, demonstrating that not all GWAS variants located in regulatory elements contribute to regulatory function.

Open-chromatin peaks and histone-modification and transcription-factor ChIP-seq data in liver cell types partially predicted the variants that showed regulatory activity. Variants rs4846913, rs2144300, rs1555290, and rs6143660 overlapped the most peaks (≥ 22 each) of any candidate variant, but rs2281721 did not overlap any DNase or FAIRE peaks; instead, it overlapped only four histone-modification peaks (H3K4me1, H3K4me2, H3K9ac, and H3K27ac) and one transcription-factor ChIP-seq peak (USF1). We compared our results to the Probabilistic Identification of Causal SNPs (PICS) algorithm's [121] prediction of candidate causal variants in this region. The variants that overlapped the most marks of open chromatin and histone modification in liver cell types (rs4846914, rs10127775, rs4846913, rs2144300, rs17315646, and rs10864726) showed the highest prediction scores, whereas the remaining 19 variants (including rs2281721) showed a prediction score of 0. Thus, testing many candidate variants in experimental assays was necessary for identifying all variants that showed strong allelic differences in enhancer activity. We also conclude that testing all candidate variants in strong LD in combination with regulatory datasets provides the best chance of identifying regulatory variants that exhibit allelic differences in enhancer activity.

One or more of these regulatory variants might also influence gene expression in other cell types. In addition to being present in HepG2 cells, rs4846913, rs2144300, rs1555290, rs6143660, and

rs2281721 are also located within predicted ChromHMM[103] strong enhancer chromatin states in K562 leukemia cells from ENCODE data[72] and ChromHMM active enhancer chromatin states in multiple cell types and tissues from Roadmap Epigenomics data[73] including adipose nuclei, CD34 cells, brain, pancreas, and skeletal muscle (Figure 2.3).

One or more of the 21 candidate variants that did not show haplotype or allelic differences in enhancer activity in both orientations, especially those variants tested individually in smaller segments, might also contribute to the *GALNT2* regulatory mechanism. Testing smaller segments (~100–200 bp) in reporter assays can help with focusing on individual variants; however, these segments might fail to capture full regulatory elements. Testing larger segments can be beneficial because they could encompass a regulatory element, although segment size must be balanced with the fact that larger elements might contain both enhancer and repressor regions and mask modest allelic effects.

The observation that the 780-bp segment showed similar haplotype differences in enhancer activity in both a luciferase vector with a minimal promoter and a promoterless luciferase vector further confirms that rs4846913, rs2144300, rs1555290, and rs6143660 are located within a regulatory element that drives expression. We did not find annotated evidence of an alternative promoter or alternative splicing in this region; however, we cannot rule out alternate mechanisms. Using qRT-PCR and strand-specific primers, we detected hepatocyte and HepG2 RNA transcribed from both strands within *GALNT2* intron 1 (T.R., unpublished data). This result is not unexpected, given that transcription initiation has been found to occur in both directions at both promoters and enhancers[122].

The functional regulatory variants at this locus might bind multiple transcription factors and act together in a complex to influence transcriptional enhancer activity. Increased binding of CEBPB to the rs4846913 A allele, associated with increased HDL-C, is consistent with the fact that CEBPB acts as a transcriptional activator[123]. CEBPB is a CCAAT/enhancer binding transcription factor that plays diverse roles in cell proliferation, development, adipocyte differentiation, immune response, and liver gene expression[124-128]. Knockdown of *CEBPB* (71% of negative control) by siRNA in HepG2 cells resulted in a modest 22% decrease in *GALNT2* expression ($p = 0.058$; Figure 2.24), suggesting that other transcription factors within a complex might bind to rs4846913 and be sufficient to preserve substantial

enhancer activity and *GALNT2* expression in HepG2 cells. Nonetheless, our data suggest that differential binding of CEBPB to rs4846913 is one potential molecular mechanism underlying the HDL-C association.

The other candidate regulatory variants, rs2281721, rs2144300, rs1555290, and rs6143660, might also bind one or more transcription factors. Binding of USF1 to rs2281721 was allele specific in EMSAs; however, ChIP assays in HepG2 (C/T at rs2281721) and Huh-7 (C/C at rs2281721) cells suggest that USF1 binds to both alleles of rs2281721 in a native chromatin context. Therefore, USF1 might not drive the allelic differences observed in EMSAs and in HepG2 and Huh-7 cell transcriptional reporter assays. USF1 is an upstream stimulatory factor and has been shown to bind to lipid- and glucose-metabolism-related genes[129-131] and affect cholesterol homeostasis, insulin sensitivity, and body-fat mass[132]. The other three potential regulatory variants, rs2144300, rs1555290, and rs6143660, each overlap ≥ 13 transcription-factor ChIP-seq peaks in HepG2 cells. Future experiments will be valuable for confirming the role of CEBPB and USF1 in *GALNT2* expression, identifying other transcription factors contributing to the haplotype differences in transcriptional enhancer activity, detecting physical interactions of this enhancer region, and fully characterizing this complex molecular mechanism at *GALNT2*.

The transcriptional reporter assays all exhibit the same direction of effect; alleles associated with increased HDL-C are also associated with increased enhancer activity. In the assay of liver AEI (Figure 2.21A) and the study on adipose eQTLs (Figure 2.21B and Figure 2.22), the rs4846914 A allele associated with increased HDL-C was associated with higher *GALNT2* expression. Although the HDL-C GWAS variants were not associated with *GALNT2* expression in an eQTL dataset of more than 400 liver samples ($p < 3.95 \times 10^{-8}$)[133] or in our eQTL analysis of primary hepatocyte samples from 50 individuals (Figure 2.20), significant associations were previously observed[134] in 146 human liver biopsy samples ($p = 0.002$) and 105 carotid-atherosclerotic-plaque biopsy samples ($p = 0.001$), consistent with our observed direction of effect. The evidence of variant association with the level of *GALNT2* expression in both liver and adipose tissue is consistent with the frequent observation that eQTLs are shared across multiple tissues[135]. This direction of effect is opposite to the effect on HDL-C of a previous study that used adeno-associated viral vectors to overexpress and knock down mouse *Galnt2*[87] and a study that observed a rare *GALNT2* missense variant in humans[101]. Notably, our observed direction of effect of

the human GWAS variants is consistent with recent unpublished studies demonstrating that total loss of function of *GALNT2* in humans, mice, rats, and cynomolgus monkeys consistently results in lower HDL-C (S. Khetarpal, D. Rader, personal communication). *GALNT2* remains a likely candidate gene, considering the association between the GWAS variants and *GALNT2* expression, but not expression of 16 other potential target genes within 1 Mb in samples of subcutaneous adipose tissue. However, the regulatory variants we identified might also act to increase expression of other nearby genes that might contribute to the HDL-C association signal. Identification of the functional variants responsible for the human GWAS signal can provide further understanding of the direction of effect in humans and lead to a greater insight into the molecular mechanisms of how these variants might influence gene expression and function.

In the METSIM cohort, the studied variants were most strongly associated with traits correlated with total cholesterol in medium HDL, including cholesterol esters in medium HDL ($r^2 = 1.0$), concentrations of medium HDL particles (>0.9), phospholipids in medium HDL (>0.9), and free cholesterol in medium HDL (>0.9). Sub-phenotype associations can provide additional insight into the function of GWAS loci[136]. *GALNT2* might act during specific steps of HDL-particle formation or remodeling by directly glycosylating lipid modifiers or enzymes, such as ANGPTL3[100] or LIPG, an enzyme that has been shown to hydrolyze phospholipids in HDL[137, 138] and remodel HDL particles[138]. More work is necessary for determining the mechanism(s) by which *GALNT2* might influence the size of HDL particles.

Overall, this study demonstrates multiple lines of evidence suggesting that at least two regulatory variants might act to regulate expression of *GALNT2*, a gene involved in HDL-C metabolism (Figure 2.25). The eQTL and AEI data clarify the direction of effect by which the GWAS variants act. Our study joins a growing set of studies that implicate multiple functional regulatory variants at a GWAS locus[139, 140], highlighting the complexity of molecular mechanisms underlying GWAS loci, and emphasizes that multiple common functional regulatory variants might work in concert.

2.5 Conflicts of interest

A.J.K., P.S., and M.A.-K. are shareholders of Brainshake Ltd., which offers NMR-based metabolite profiling.

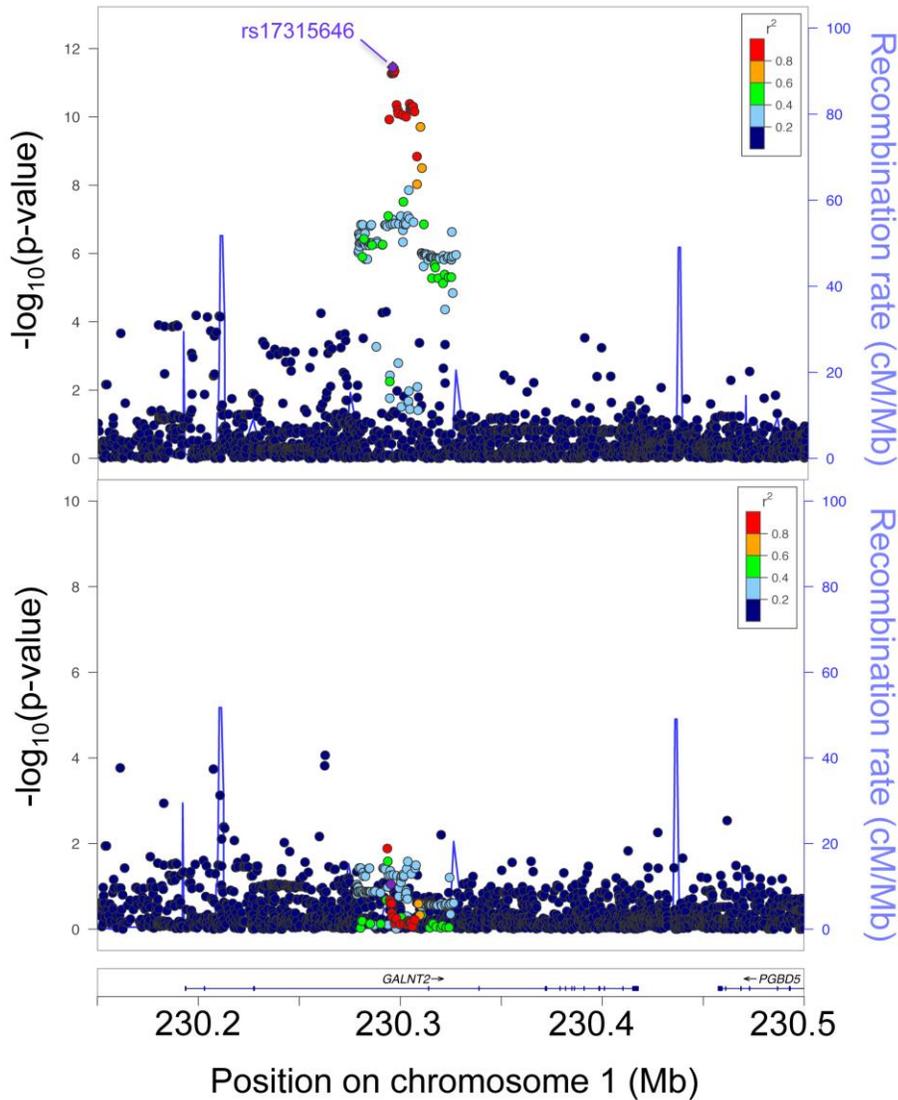


Figure 2.1 Non-coding variants at *GALNT2* are associated with total cholesterol in medium HDL in the METSIM study

The entire initial association signal (upper panel) was reduced after conditioning on lead SNP rs17315646 (lower panel). Circles represent genotyped and imputed DNA variants and their LD r^2 values with rs17315646 in the METSIM study (2,079 variants are shown). Chromosome coordinates correspond to UCSC Genome Browser build hg19. The left y axis indicates the $-\log_{10}(\text{p value})$, the right y axis indicates the recombination rate (cM/Mb), and the x axis indicates position on chromosome 1 (Mb).

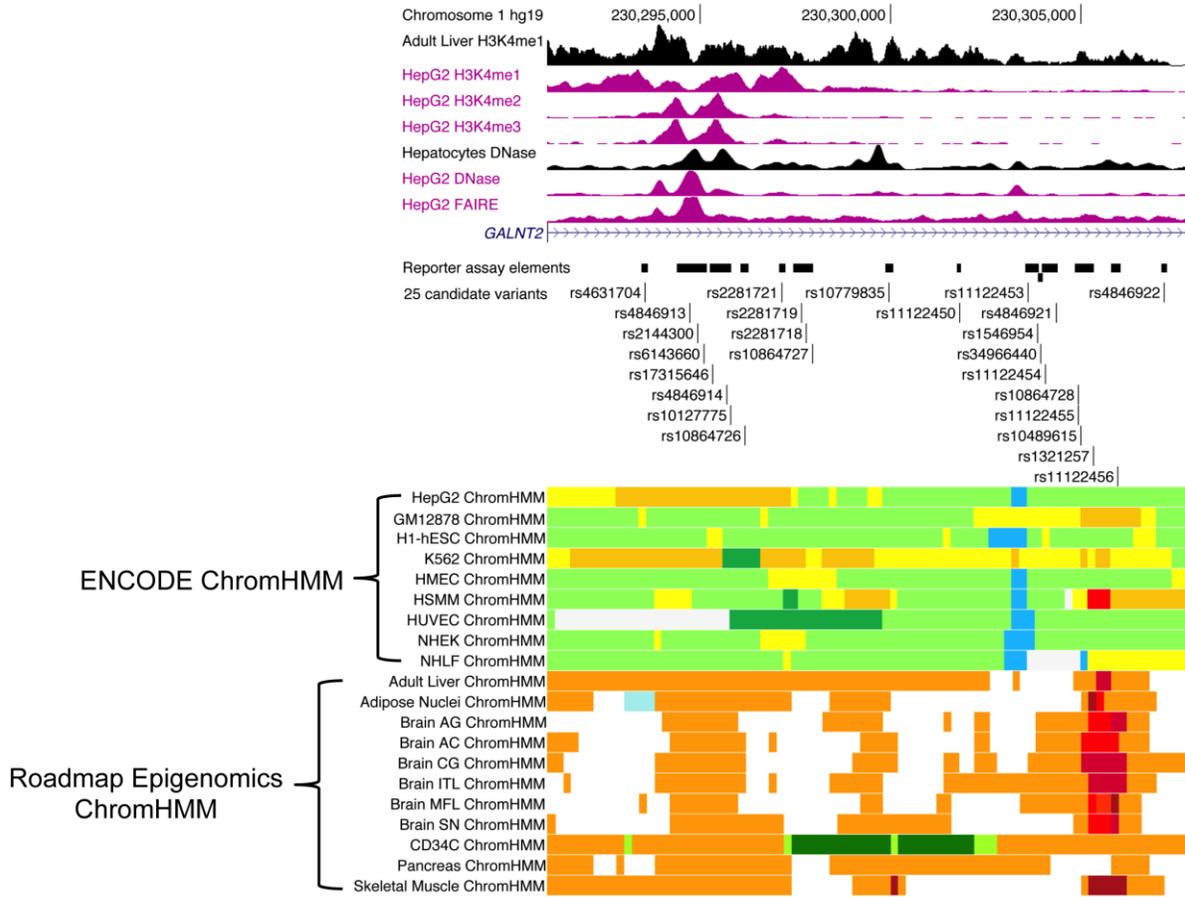


Figure 2.3 HDL-C associated variants overlap open chromatin, histone modifications, and predicted enhancer and transcribed chromatin states in liver cell types identifying potential regulatory regions in *GALNT2* intron 1

Figure 2.3 HDL-C associated variants overlap open chromatin, histone modifications, and predicted enhancer and transcribed chromatin states in liver cell types identifying potential regulatory regions in *GALNT2* intron 1

A 13.7-kb region includes all 24 variants in strong linkage disequilibrium ($r^2 > .7$) with the HDL-C-associated index SNP rs4846914 (25 total candidate variants). Selected Roadmap Epigenomics Human Epigenome Atlas histone modification and chromatin state segmentation tracks and ENCODE open chromatin, histone modification and ChromHMM chromatin state tracks are shown. Rectangular bars represent elements containing the variant(s) that were tested in luciferase reporter assays. GM12878, B-lymphocyte, lymphoblastoid cells; H1-hESC, embryonic stem cells; K562, leukemia cell line; HMEC, human mammary epithelial cells; HSMM, human skeletal muscle myoblasts; HUVEC, human umbilical vein endothelial cells; NHEK, epidermal keratinocytes; NHLF, lung fibroblasts, Brain AG, angular gyrus; Brain AC, anterior caudate; Brain CG, cingulate gyrus; Brain ITL, inferior temporal lobe; Brain MFL, mid-frontal lobe; Brain SN, substantia nigra; CD34C, CD34 cultured cells. Yellow and orange ChromHMM regions indicate weak and active enhancers, respectively, white indicates quiescent/lowly expressed regions, red indicates promoter regions, green indicates transcribed regions, and blue indicates insulator regions.

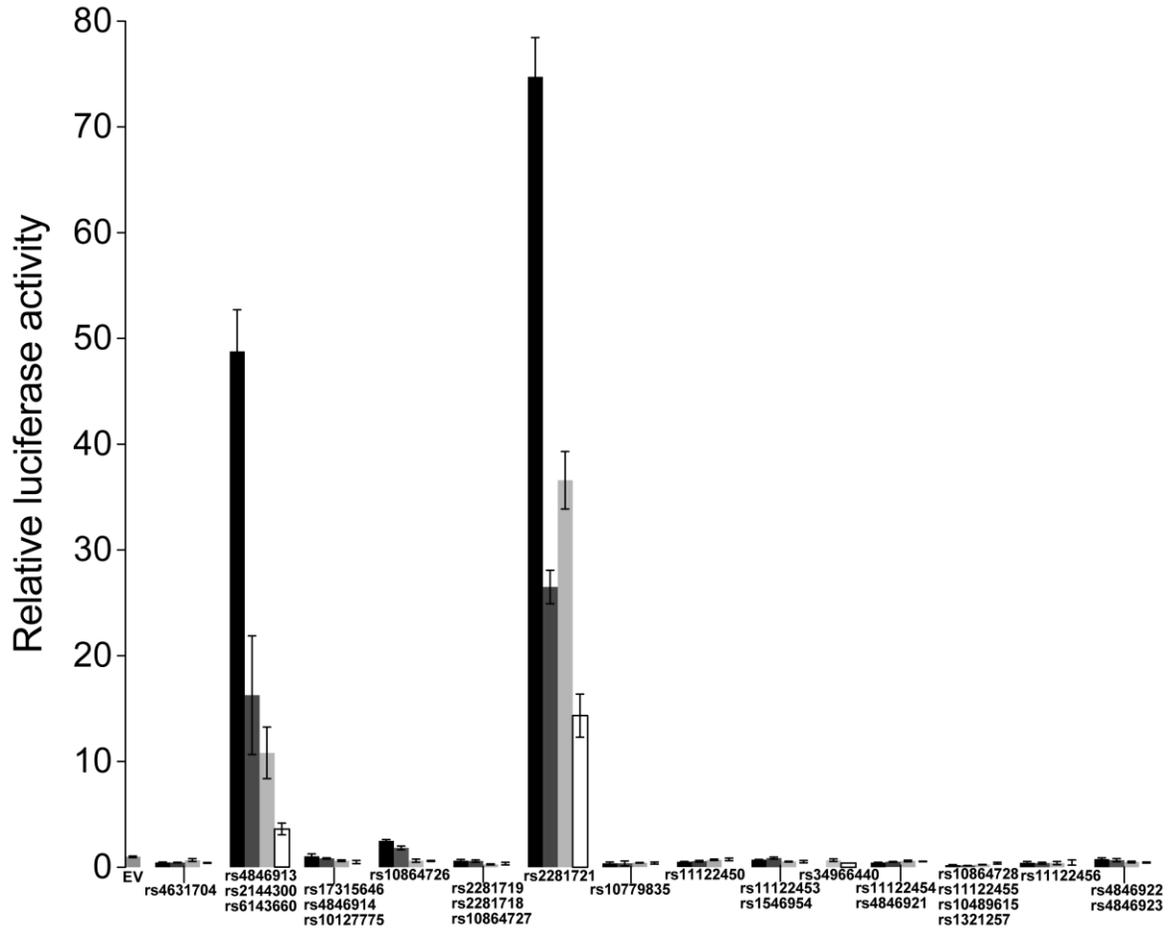


Figure 2.4 Two of 14 elements containing variants in strong linkage disequilibrium ($r^2 > 0.7$) with rs4846914 show allelic or haplotype differences in transcriptional enhancer activity in both orientations

Luciferase activity of DNA segments containing one to four variants cloned upstream of a minimal promoter luciferase vector and transfected into HepG2 cells. Black bars indicate luciferase activity of segments containing the allele/haplotype associated with increased HDL-C, and dark gray bars indicate segments with the other allele/haplotype, both cloned in the forward orientation. Light gray bars represent luciferase activity of segments containing the allele/haplotype associated with HDL-C, and white bars indicate fragments containing the other allele/haplotype, both cloned in the reverse orientation. Luciferase activity is shown normalized to an empty vector control (EV). Error bars represent standard deviation of 2-6 independent clones per allele/haplotype, except for 2-bp indel rs34966440, which was tested with 1-3 independent clones in the reverse orientation. The element containing rs4846922 ($LD\ r^2 > 0.7$ with rs4846914) contained an additional SNP, rs4846923 ($r^2 > 0.6$ with rs4846914).

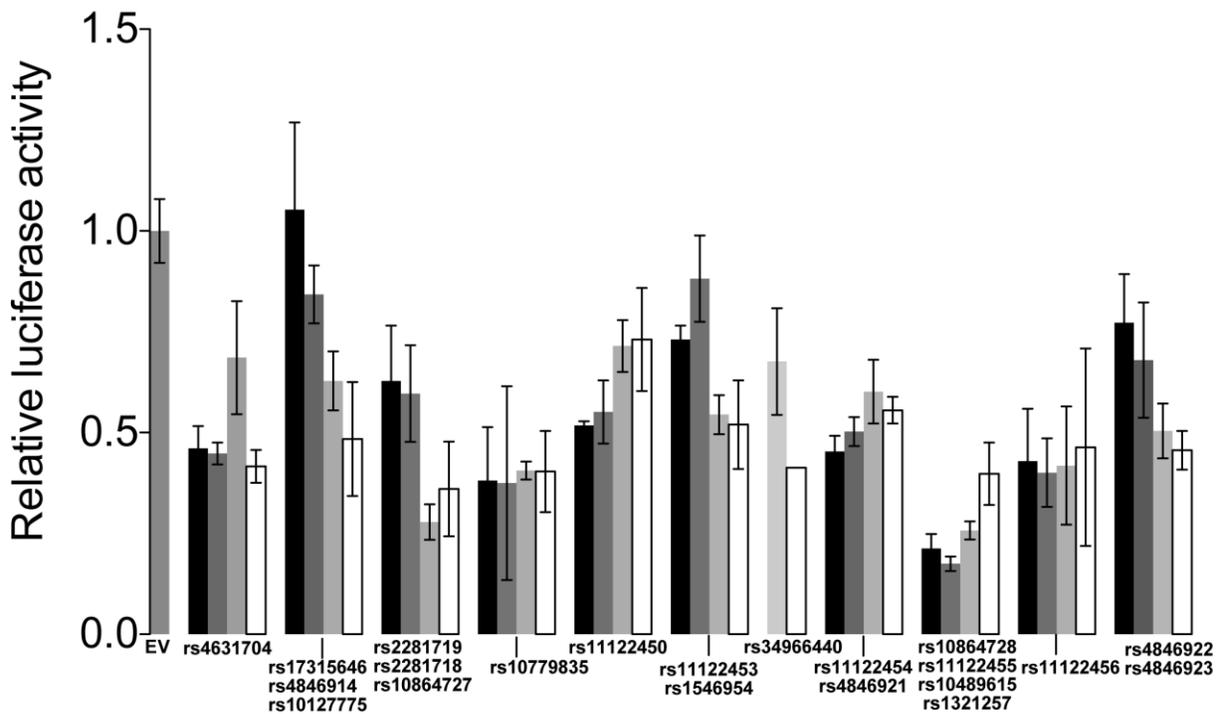


Figure 2.5 Twenty-one *GALNT2* intronic variants tested in reporter assays do not exhibit allelic or haplotype differences in transcriptional enhancer activity in both orientations

Luciferase activity of DNA segments containing one to four variants cloned upstream of a minimal promoter luciferase vector and transfected into HepG2 cells. These 11 segments did not show allelic differences in enhancer activity (<1.5-fold) in both orientations compared to a vector control. Black bars indicate luciferase activity of segments containing the allele/haplotype associated with increased HDL-C, and dark gray bars indicate segments with the other allele/haplotype, both cloned in the forward orientation. Light gray bars represent luciferase activity of segments containing the allele/haplotype associated with HDL-C, and white bars indicate fragments containing the other allele/haplotype, both cloned in the reverse orientation. Luciferase activity is shown normalized to an empty vector control (EV). Error bars represent standard deviation of 2-6 independent clones per allele/haplotype, except for 2-bp indel rs34966440, which was tested with 1-3 independent clones in the reverse orientation. The element containing rs4846922 (LD $r^2 > 0.7$ with rs4846914) contained an additional SNP, rs4846923 ($r^2 > 0.6$ with rs4846914).

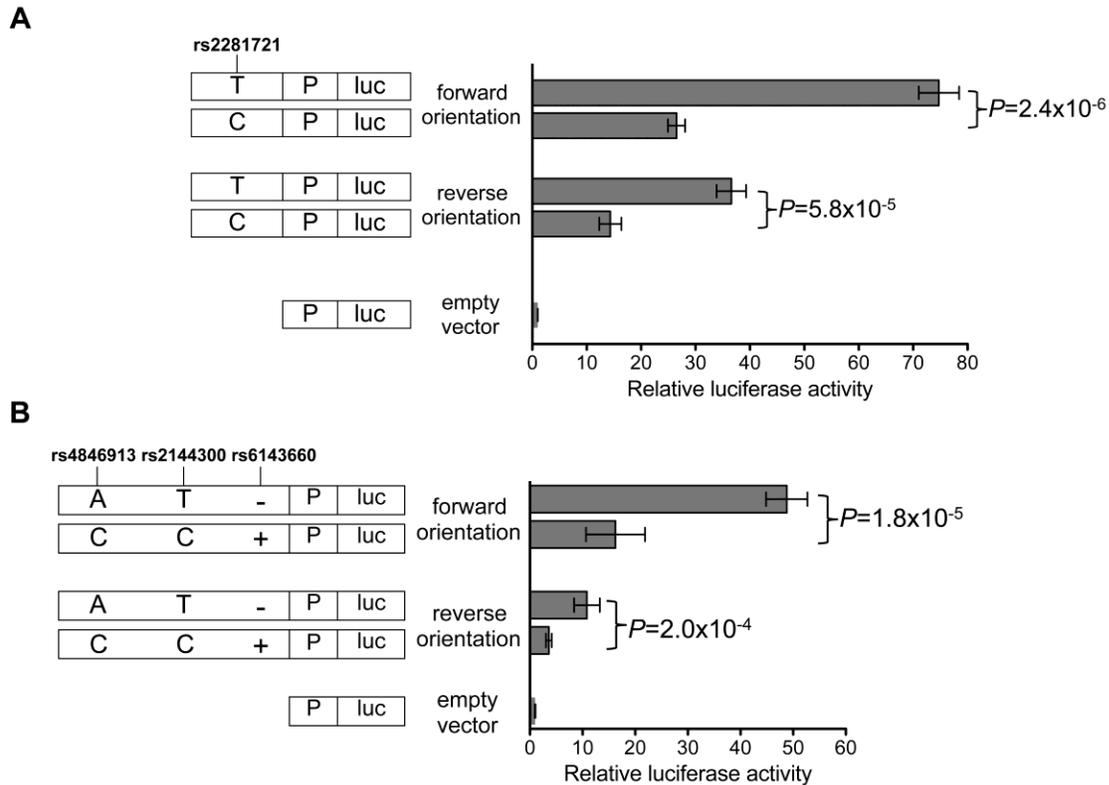


Figure 2.6 Haplotype and allelic differences in transcriptional activity at the *GALNT2* locus

Segments containing each haplotype or allele were cloned into a pGL4.23 luciferase reporter vector upstream of the minimal promoter in both orientations. The vectors were transfected into HepG2 cells, and luciferase expression normalized to that of an empty vector control is shown. Error bars represent the pairwise SD of three to six independent clones per allele or haplotype (t tests). Abbreviations are as follows: P, promoter; and luc, luciferase.

A Luciferase activity of 154-bp DNA segments containing rs2281721 alleles.

B Luciferase activity of 780-bp DNA segments of two different haplotypes. The haplotypes contained three candidate variants in strong LD: rs4846913, rs2144300, and 21-bp indel, rs6143660. An additional variant, rs1555290, was detected in the segment, as shown in Figures 2.8 and 2.9.

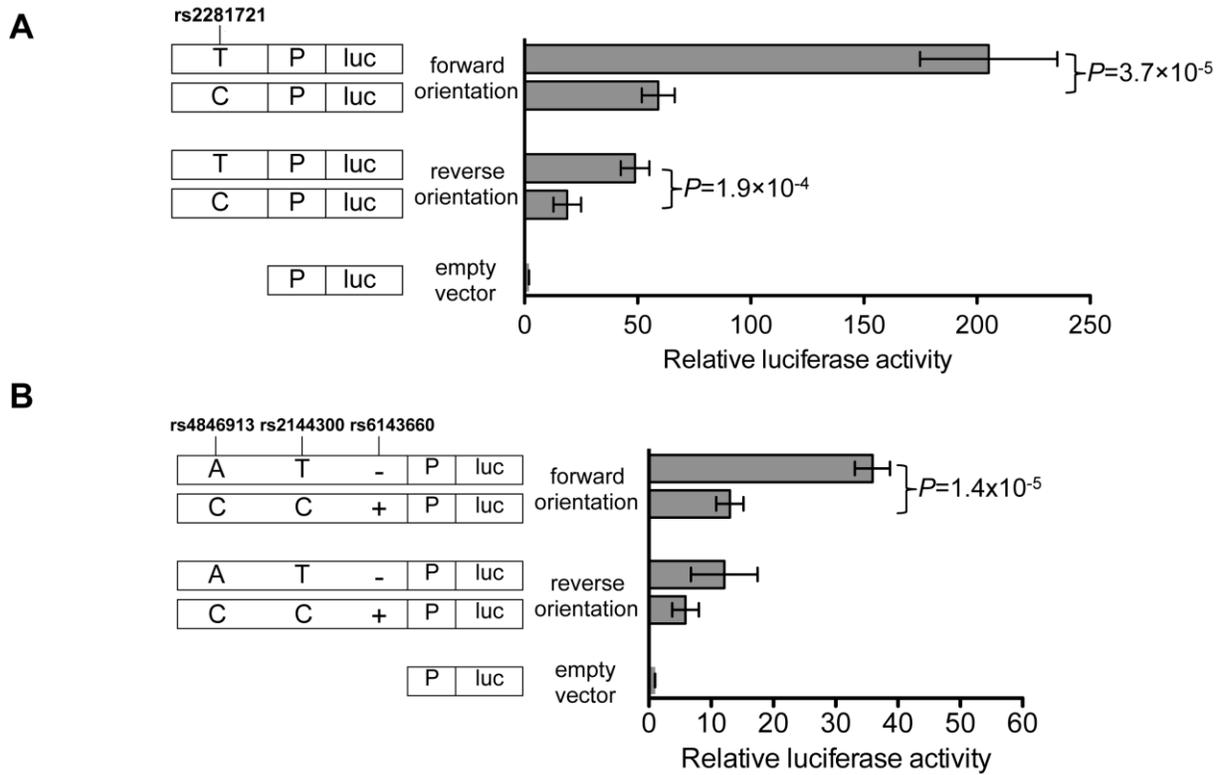


Figure 2.7 Haplotype and allelic differences in transcriptional activity at the *GALNT2* locus are consistent in Huh-7 cells

Segments containing each haplotype/allele were cloned into a pGL4.23 luciferase reporter vector upstream of the minimal promoter in both orientations. The vectors were transfected into Huh-7 cells, and luciferase expression is shown normalized to an empty vector control. Error bars represent pairwise standard deviation of 3-5 independent clones per allele/haplotype (t-tests). P, promoter; luc, luciferase gene.

A Luciferase activity of 154-bp DNA segments containing rs2281721 alleles.

B Luciferase activity of 780-bp DNA segments of 2 different haplotypes. The haplotypes contained three candidate variants in strong LD, rs4846913, rs2144300, and 21-bp indel, rs6143660. Luciferase activity measurements were performed on the same day as the site-directed haplotypes and are also shown in Figure 2.9.

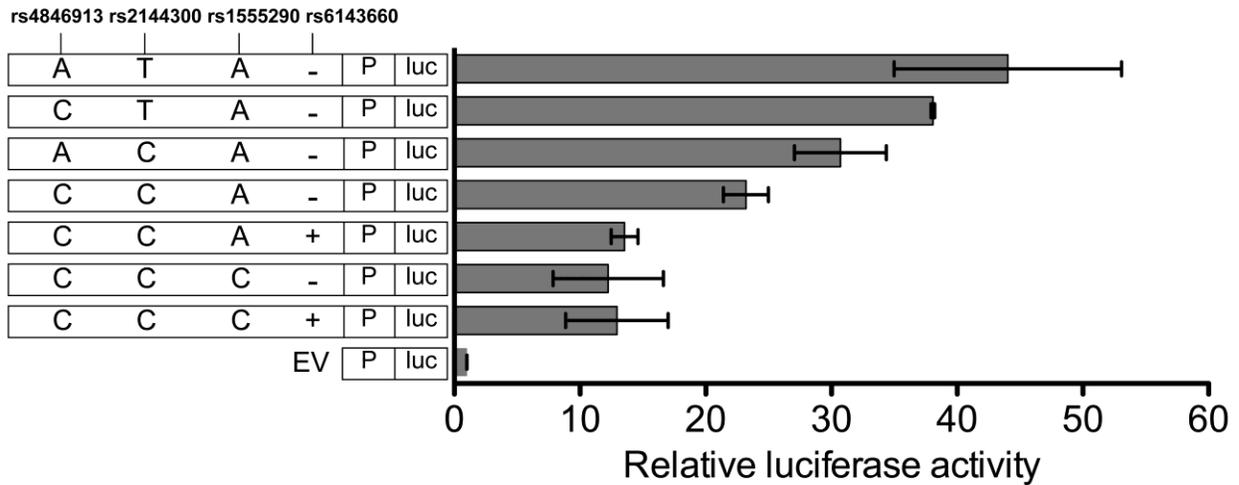


Figure 2.8 Haplotype variants act together to increase transcriptional activity

Additional haplotypes were created by site-directed mutagenesis of haplotypes cloned into a pGL4.23 luciferase vector in the forward orientation. All constructs were transfected separately into HepG2 cells, and this experiment was performed separately from the experiment presented in Figure 2.6. Luciferase activity was measured and normalized to that of an empty vector control. Error bars represent the pairwise SD of three to four independent clones per haplotype (ANOVA and Tukey's post hoc tests). Abbreviations are as follows: P, promoter; and luc, luciferase.

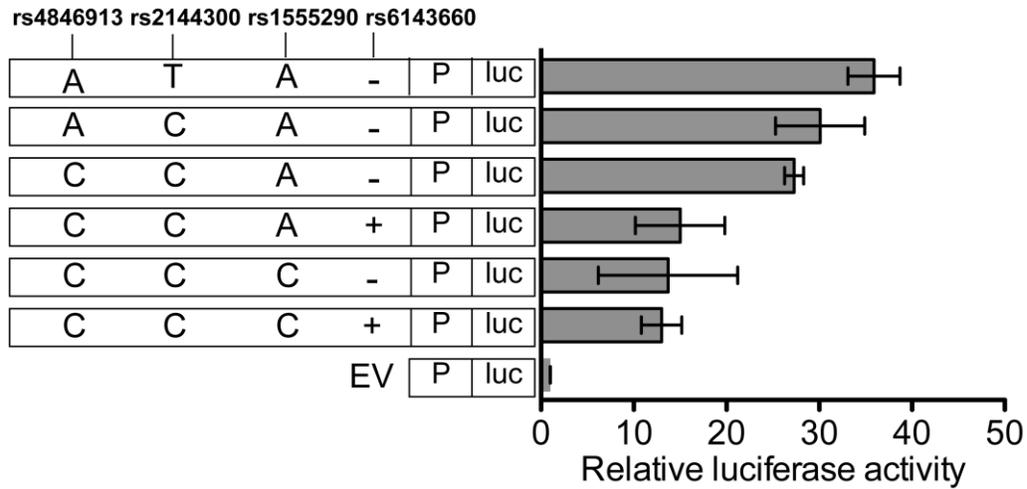


Figure 2.9 Haplotype variants act together to increase transcriptional activity in Huh-7 cells

Additional haplotypes were created by site-directed mutagenesis of haplotypes cloned into pGL4.23 luciferase vector in the forward orientation. All constructs were transfected separately into Huh-7 cells. Luciferase activity was measured and normalized to empty vector control. Error bars represent standard deviation of 3-4 independent clones per haplotype (ANOVA and Tukey post-hoc tests). P, promoter; luc, luciferase. All haplotypes are present as in Figure 2.8 except for CTA-.

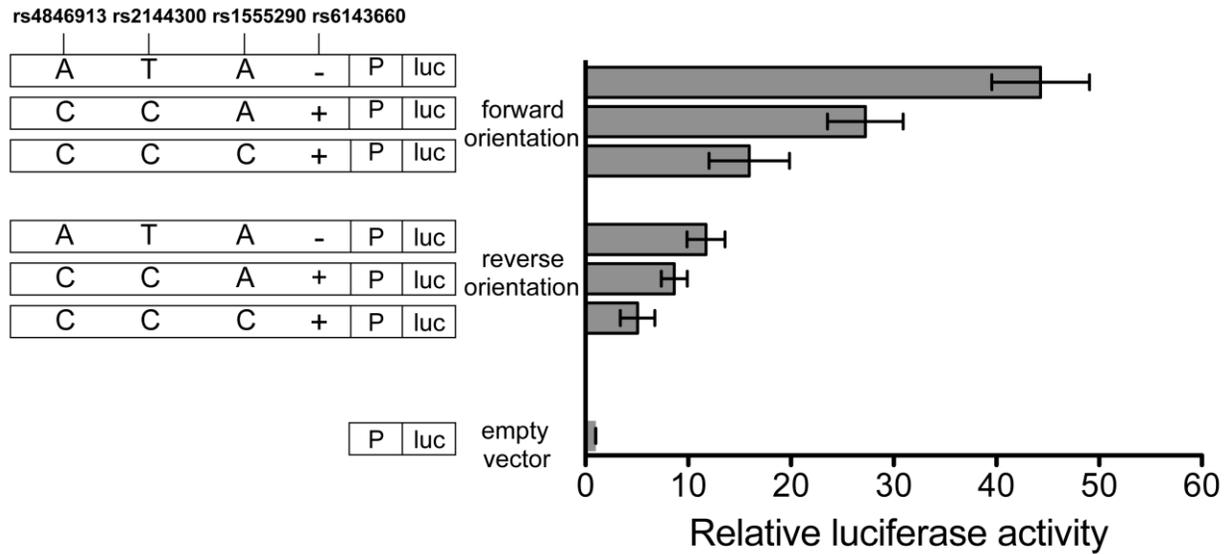


Figure 2.10 The 780-bp segment shows haplotype differences in enhancer activity in a promoterless vector

Luciferase activity of 780-bp DNA segments in 3 different allelic combinations cloned upstream of a promoterless pGL4.10 luciferase reporter after transfection into HepG2 cells. The segments contained two SNPs, rs4846913, rs2144300, and a 21-bp indel, rs6143660 in strong pairwise LD, and an additional SNP, rs1555290, in moderate LD ($r^2 = 0.2$) with the HDL-C-associated variants. Luciferase activity is shown normalized to an empty vector control. Error bars represent standard deviation of 4-7 independent clones per haplotype. P, promoter; luc, luciferase.

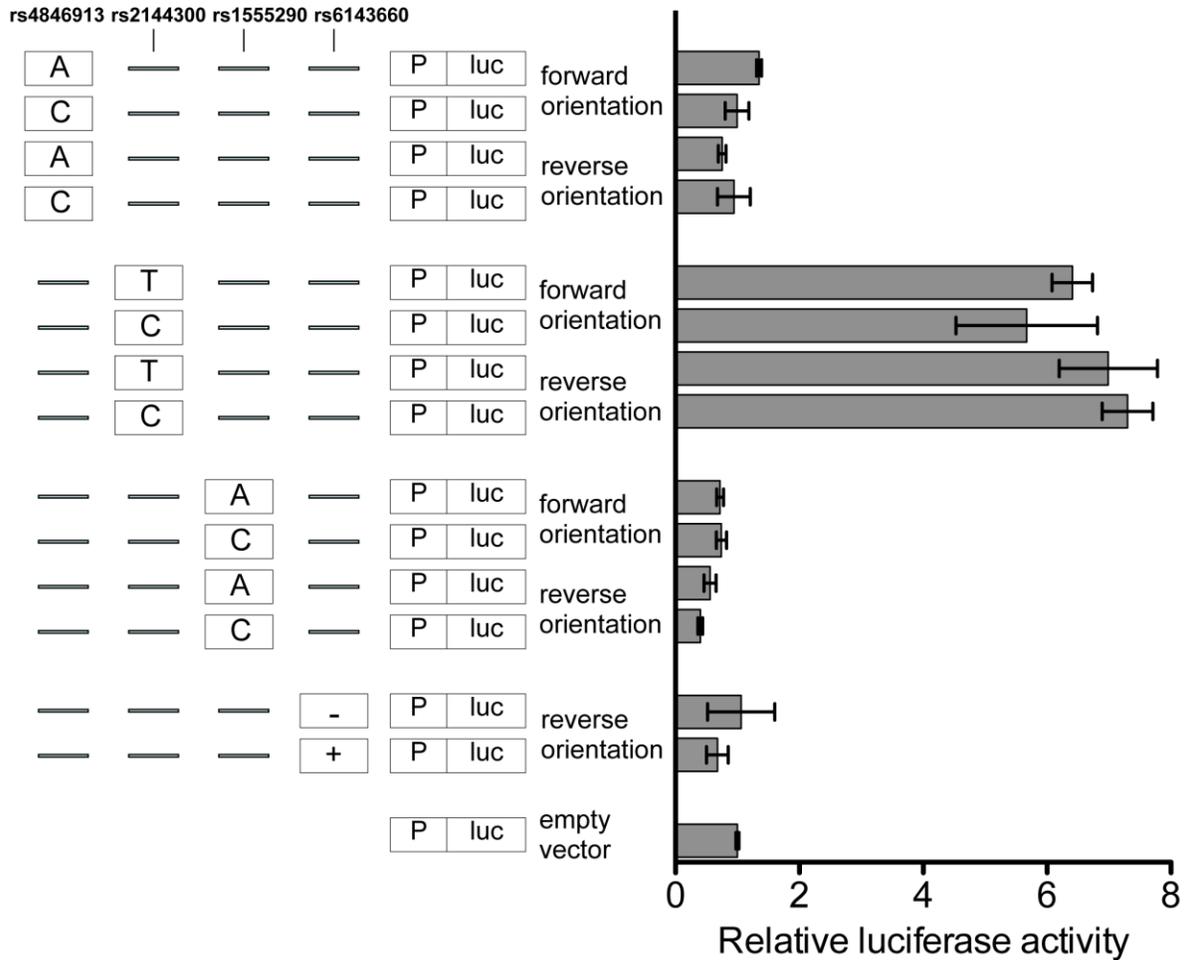


Figure 2.11 SNPs rs4846913, rs2144300, rs1555290 and 21-bp indel rs6143660 do not show allelic differences in enhancer activity when cloned and tested as separate segments in luciferase reporter assays

Luciferase activity of 120-bp DNA segments containing rs4846913 A or C alleles, 201-bp DNA segments containing rs2144300 T or C alleles, 121-bp segments containing rs1555290 A or C alleles, and 173-bp DNA segments containing a 21-bp deletion or 21-bp insertion (rs6143660). Error bars represent standard deviation of 2-5 independent clones per haplotype. Luciferase activity is shown normalized to an empty vector control. P, promoter; luc, luciferase.

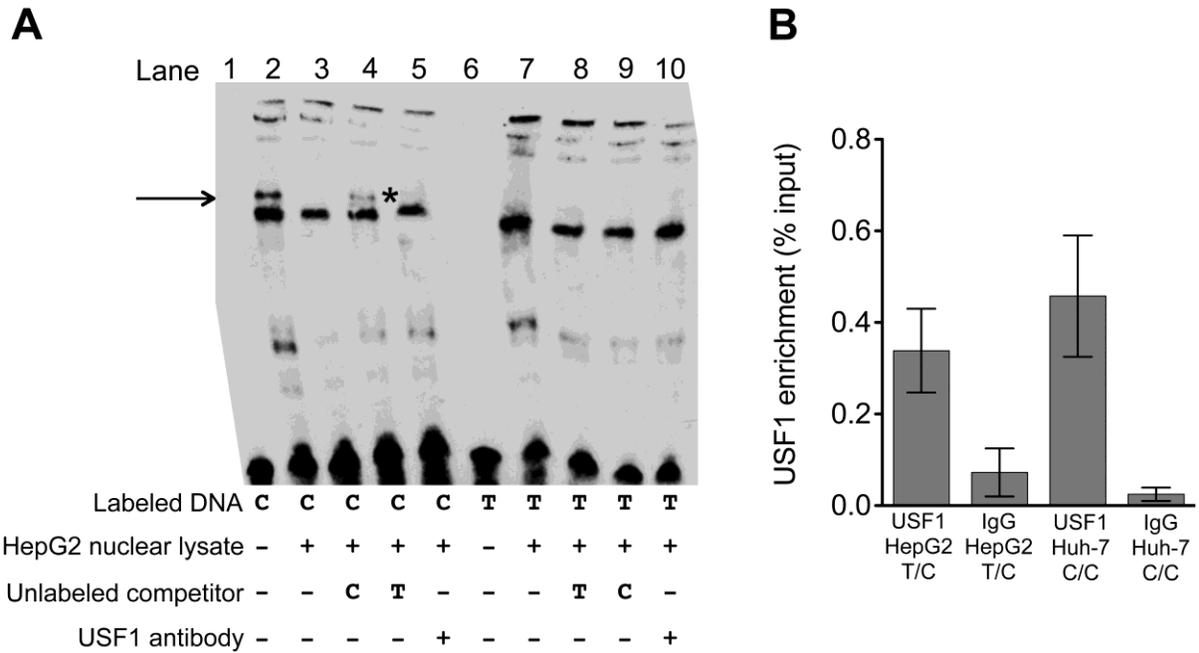


Figure 2.12 The rs2281721 C Allele shows binding to USF1

A EMSAs with biotin-labeled probes containing either the T or C allele of rs2281721 and incubated with 5 μ g HepG2 nuclear lysate. The arrow indicates an allele-specific band (lanes 2 and 7), and the asterisk indicates the C-allele-specific band that was disrupted upon incubation with USF1 antibody (lane 5). For competition reactions, 40-fold excess unlabeled probe was added.

B ChIP experiments were performed in HepG2 cells (T/C at rs2281721) and Huh-7 cells (C/C at rs2281721) with USF1 antibody or rabbit IgG control, and a 164-bp DNA region containing rs2281721 was amplified via qPCR and quantified with a standard curve. Results are shown as percentages of input DNA. Error bars represent the SEM of two independent USF1 and IgG ChIP experiments each in HepG2 and Huh-7 cells.

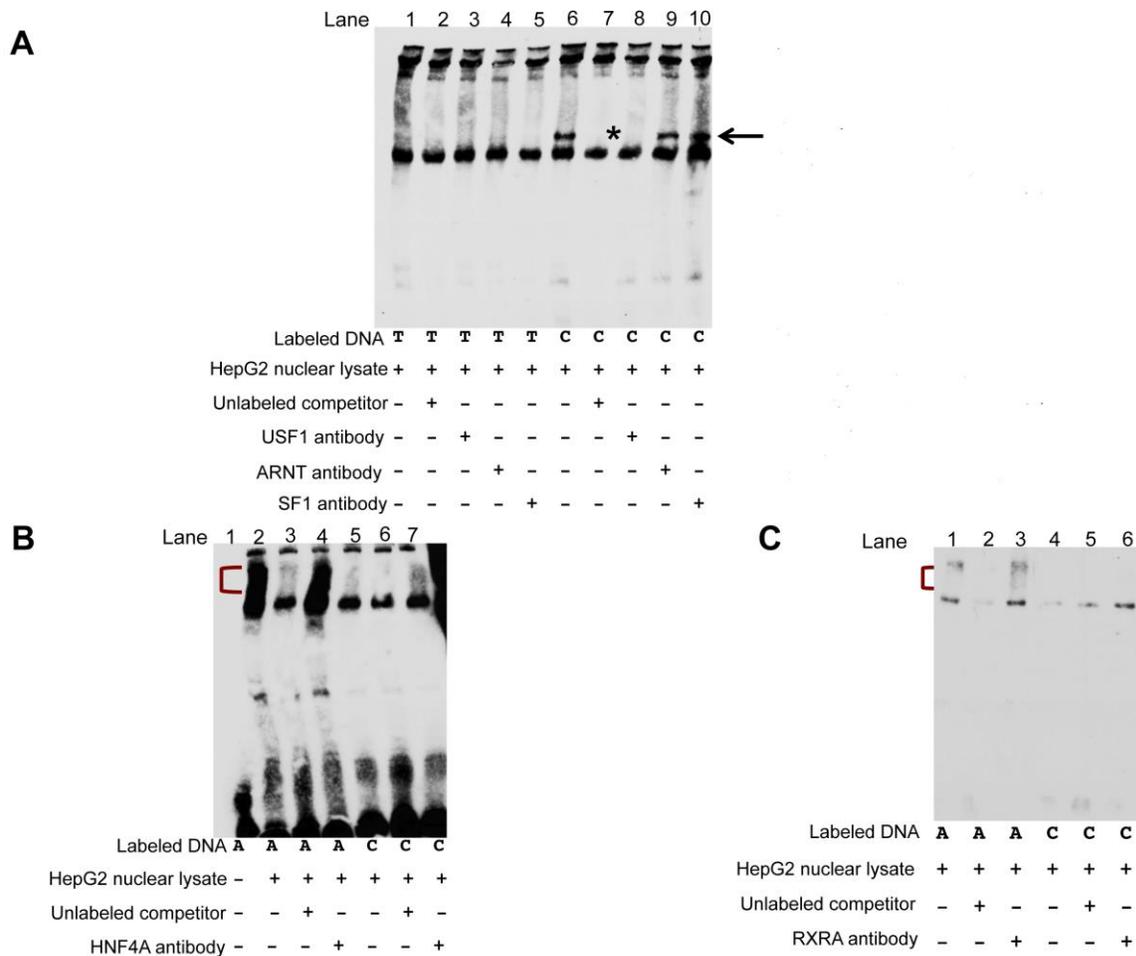


Figure 2.13 Transcription factors not bound to rs2281721 or rs4846913

To confirm specificity of USF1 binding to rs2281721 and CEBPB binding to rs4846913, additional antibodies were tested in EMSAs as negative controls. Neither ARNT nor SF1 bind to the alleles of rs2281721, and neither HNF4A nor RXRA bind to the alleles of rs4846913.

A Probes containing alleles of rs2281721 were incubated with 6 μ g HepG2 nuclear lysate and 4 μ g USF1 antibody (lanes 3 and 8), ARNT antibody (lanes 4 and 9) or SF1 antibody (lanes 5 and 10). For competition reactions, 119-fold excess unlabeled probe was added.

B Probes containing alleles of rs4846913 were incubated with 6 μ g HepG2 nuclear lysate and 4 μ g HNF4A antibody (lanes 4 and 7). For competition reactions, 63-fold excess unlabeled probe was added.

C Probes containing alleles of rs4846913 were incubated with 3.5 μ g HepG2 nuclear lysate and 4 μ g RXRA antibody (lanes 3 and 6). For competition reactions, 63-fold excess unlabeled probe was added.

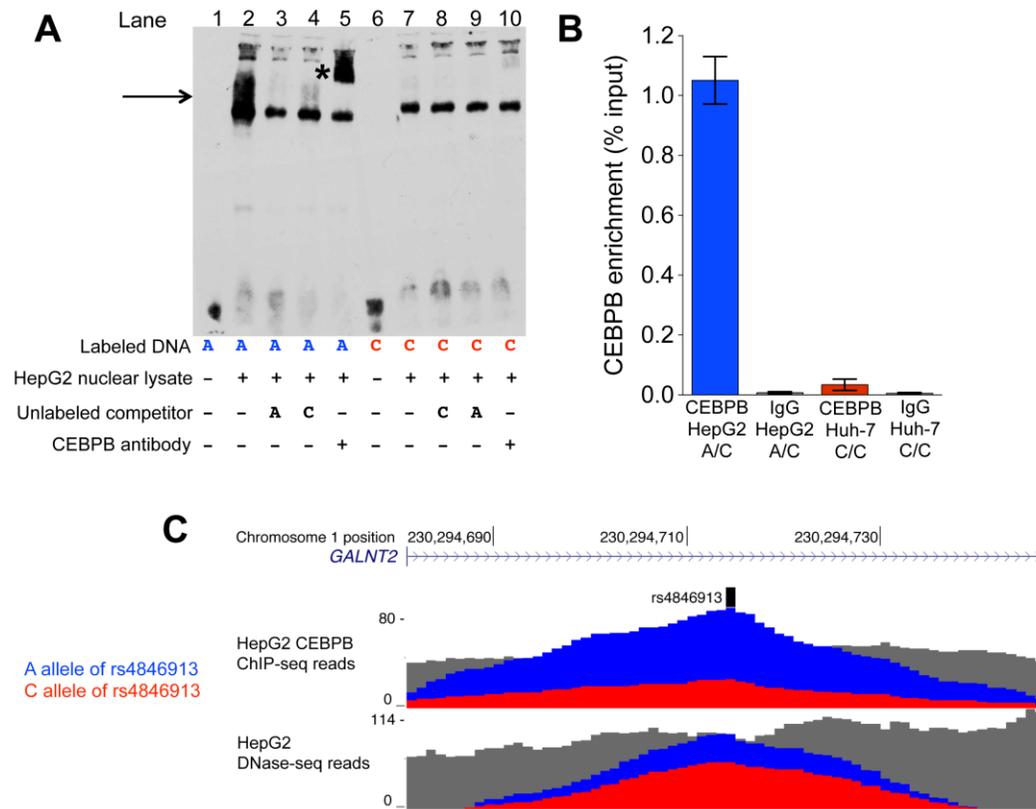


Figure 2.14 CEBPB binds differentially to the alleles of rs4846913

A EMSAs with biotin-labeled probes containing either the A or C allele of rs4846913 and incubated with 6 μ g HepG2 nuclear lysate. The arrow indicates increased protein binding to the A allele (lanes 2 versus 7). The asterisk indicates evidence of a CEBPB supershift. For competition reactions, 63-fold excess unlabeled probe was added.

B ChIP experiments were performed in HepG2 cells (A/C at rs4846913) and Huh-7 cells (C/C at rs4846913) with CEBPB antibody or rabbit IgG control, and a 120-bp DNA region containing rs4846913 was amplified via qPCR and quantified with a standard curve. Results are shown as percentages of input DNA. Error bars represent the SEM of two independent CEBPB and IgG ChIP experiments each in HepG2 and Huh-7 cells.

C CEBPB ChIP-seq reads and DNase-seq reads from ENCODE data in a region containing rs4846913 (UCSC Genome Browser hg19 chromosome coordinates). Blue indicates reads that contain the A allele of rs4846913, red indicates reads that contain the C allele of rs4846913, and gray indicates reads that do not overlap rs4846913 in the peak.

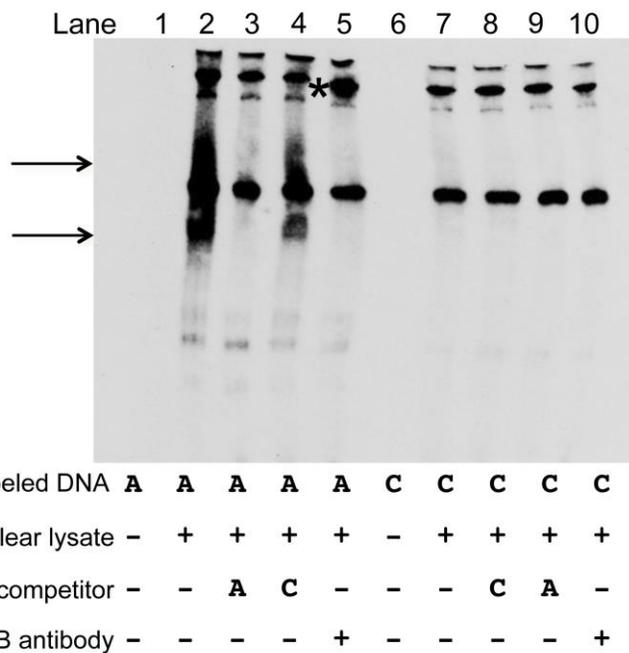


Figure 2.15 The alleles of rs4846913 show consistent allelic differences in binding

Electrophoretic mobility shift assays with biotin-labeled probes containing either the A or C allele of rs4846913. Supershift reactions were incubated with a different lot number of CEBPB antibody, resulting in a slightly different supershift band compared to Figure 2.14. Probes were incubated with 5 μ g HepG2 nuclear lysate. The arrows indicate allele-specific protein binding to the A allele, which is competed away effectively by 40-fold excess competitor DNA containing the A allele (lane 3) compared to excess competitor containing the C allele (lane 4). An additional allele-specific band was present compared to Figure 2.14 (indicated by the lower arrow). The asterisk indicates a supershift upon incubation with a CEBPB antibody.

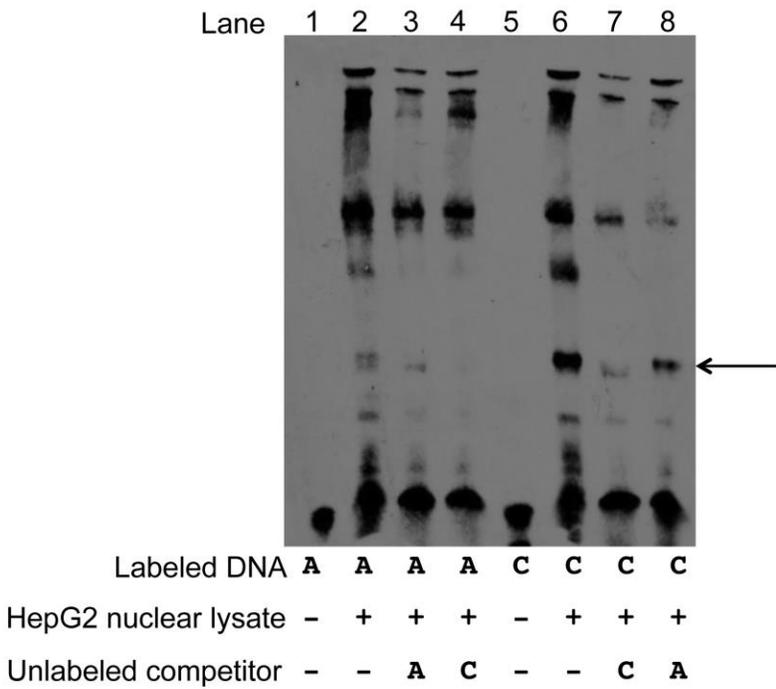


Figure 2.16 The alleles at rs1555290 exhibit allele-specific protein binding

Electrophoretic mobility shift assays with biotin-labeled probes containing either the A or C allele of rs1555290. Probes were incubated with HepG2 nuclear lysate. The arrow indicates allele-specific protein binding to the C allele, which was reproducible and observed on at least 3 separate days. For competition reactions, 30-fold excess unlabeled probe was added.

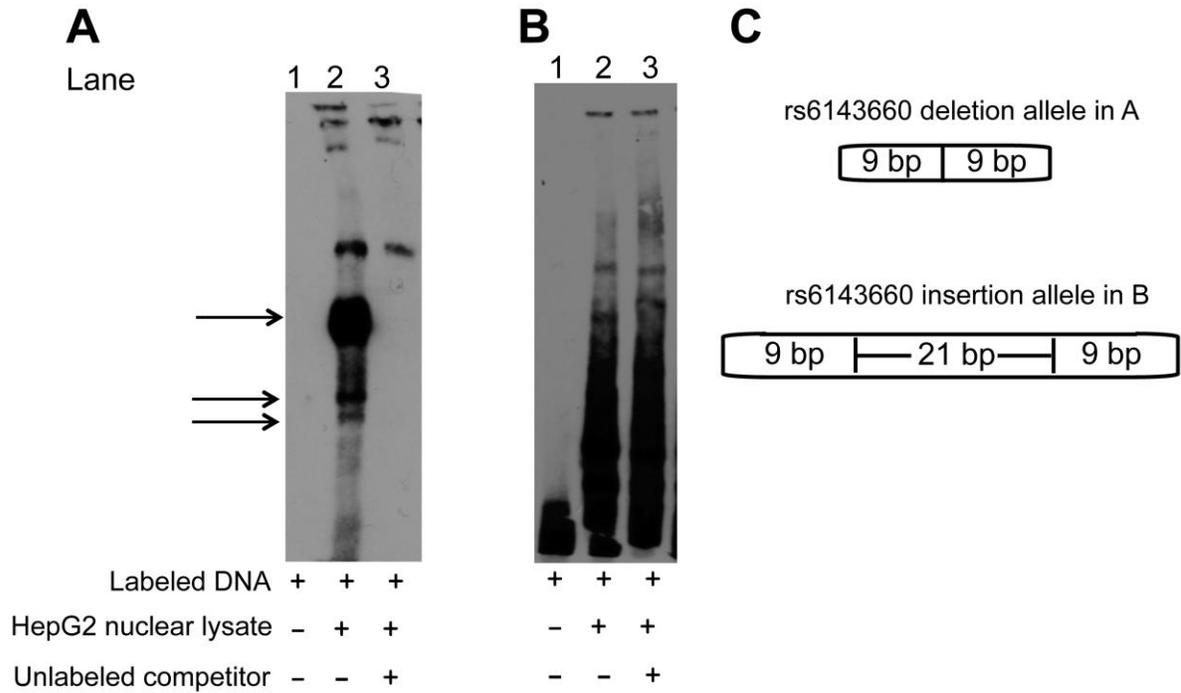


Figure 2.17 DNA sequence containing the 21-bp rs6143660 deletion allele shows evidence of protein binding

A Electrophoretic mobility shift assays with biotin-labeled probes containing a 18-bp DNA sequence flanking the 21-bp deletion. Probes were incubated with HepG2 nuclear lysate. Arrows indicate specific bands that are competed away by excess unlabeled competitor. For competition reactions, 300-fold excess unlabeled probe was added.

B EMSAs with biotin-labeled probes containing the alternate insertion allele of rs6143660, a 39-bp DNA sequence containing the 21-bp insertion and 9 bp of DNA sequence flanking either side. For competition reactions, 150-fold excess unlabeled probe was added.

C Schematic of labeled DNA EMSA probes containing the 21-bp rs6143660 deletion and insertion alleles. The deletion allele probes contain 9 bp of flanking DNA sequence (total 18 bp) and the insertion allele probes contain the 21-bp insertion sequence plus 9 bp of flanking DNA sequence on either side (total 39 bp).

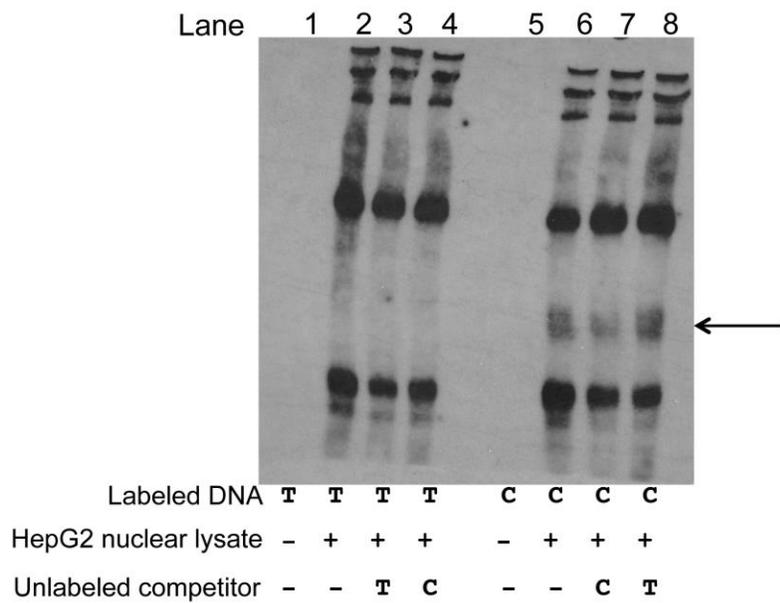


Figure 2.18 The alleles at rs2144300 show suggestive differential protein binding

Electrophoretic mobility shift assays with biotin-labeled probes containing either the T or C allele of rs2144300. Probes were incubated with HepG2 nuclear lysate. The arrow indicates allele-specific protein binding to the C allele, however, the band is not competed away upon addition of excess unlabeled rs2144300-C probe (lane 7). For competition reactions, 240-fold excess unlabeled probe was added. To resolve allele-specific bands, the unbound probes were run off the gel and were not transferred to the membrane for visualization.

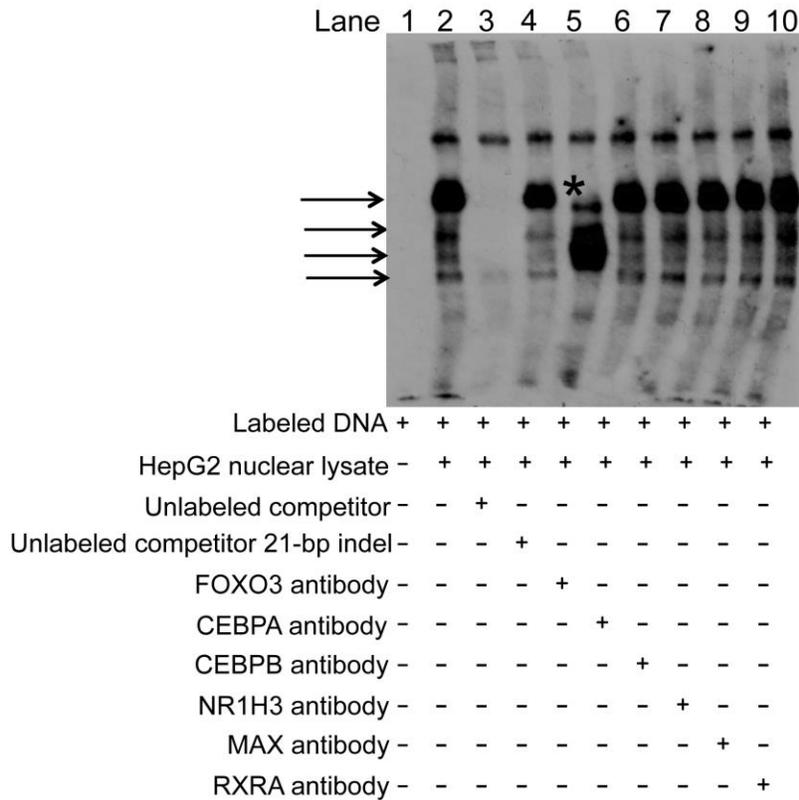


Figure 2.19 Specificity of FOXO3 binding to the deletion allele of rs6143660

To confirm specificity of FOXO3 binding to the deletion allele of rs6143660, additional antibodies were tested in EMSAs as negative controls. Biotin-labeled probes containing a 18-bp DNA sequence flanking the 21-bp deletion were incubated with HepG2 nuclear lysate. Arrows indicate specific bands that are competed away by excess unlabeled competitor and the asterisk indicates the band that is disrupted upon incubation with FOXO3 antibody (lane 5). For competition reactions, 200-fold excess unlabeled probe containing the 18-bp sequence (lane 3) or the 21-bp insertion sequence (lane 4) was added.

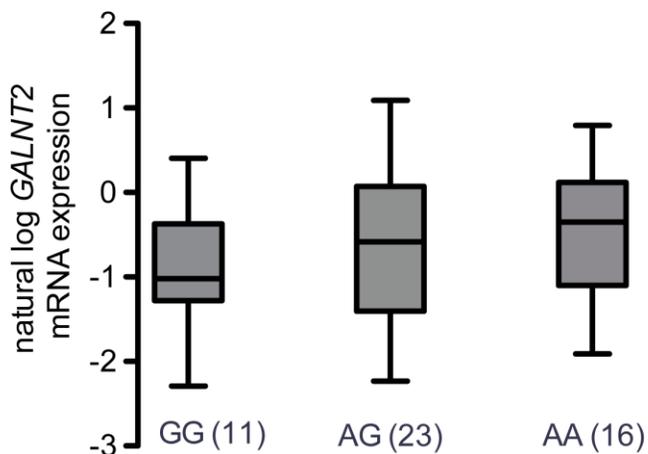


Figure 2.20 GALNT2 expression in human hepatocytes from 50 individuals

GALNT2 mRNA was measured by quantitative PCR and normalized to a housekeeping gene (Beta-2 microglobulin). Normalized *GALNT2* expression was then grouped based on genotype at the lead HDL-C-associated SNP rs4846914. *GALNT2* expression values were natural log transformed to obtain an approximate normal distribution. Horizontal lines within the boxes represent the median expression values, and whiskers represent the minimum and maximum expression values observed for each genotype.

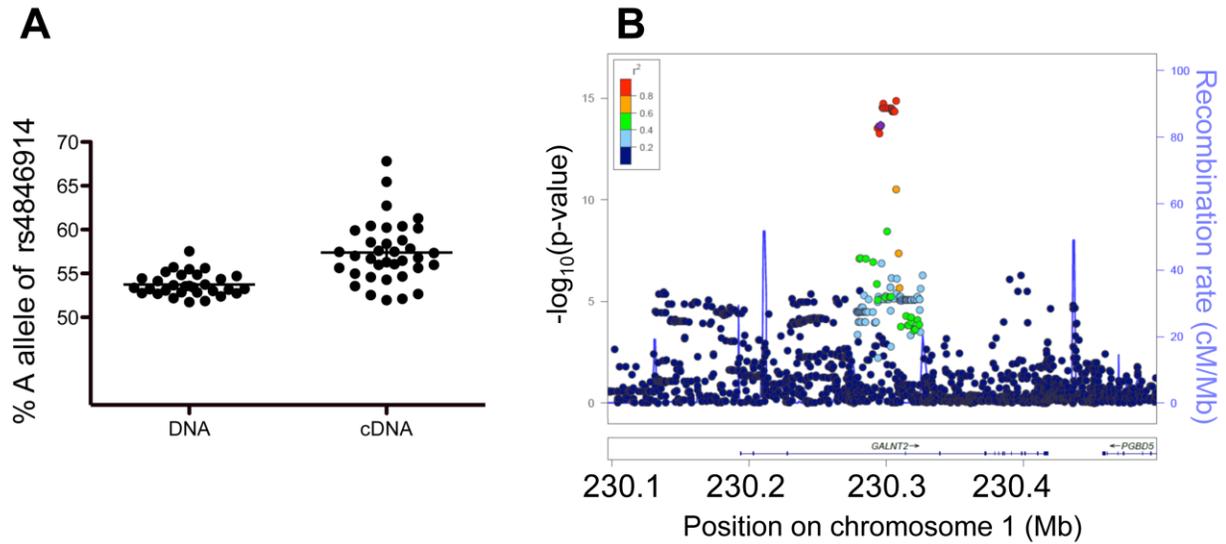


Figure 2.21 The rs4846914 A Allele associated with increased HDL-C is associated with higher *GALNT2* RNA expression in primary human hepatocytes and subcutaneous adipose tissue

A AEI assays were performed in primary human hepatocytes from 36 individuals heterozygous for the intronic HDL-C-associated lead SNP rs4846914 ($p = 5.4 \times 10^{-7}$). rs4846914 was used as a proxy for SNP rs4846913. DNA results from genomic DNA are shown as a control, and cDNA results in this intronic region represent pre-mRNA.

B eQTL queries were performed for the noncoding *GALNT2* variants and *GALNT2* expression in subcutaneous adipose tissue samples from 1,381 individuals from the METSIM study. Circles represent genotyped and imputed DNA variants and the LD r^2 values with lead eQTL variant rs4846922. LD was calculated from METSIM genotypes imputed from 1000 Genomes Phase 1 EUR dataset. Chromosome coordinates correspond to UCSC Genome Browser build hg19. The left y axis indicates the $-\log_{10}(\text{p value})$, the right y axis indicates the recombination rate (cM/Mb), and the x axis indicates position on chromosome 1 (Mb).

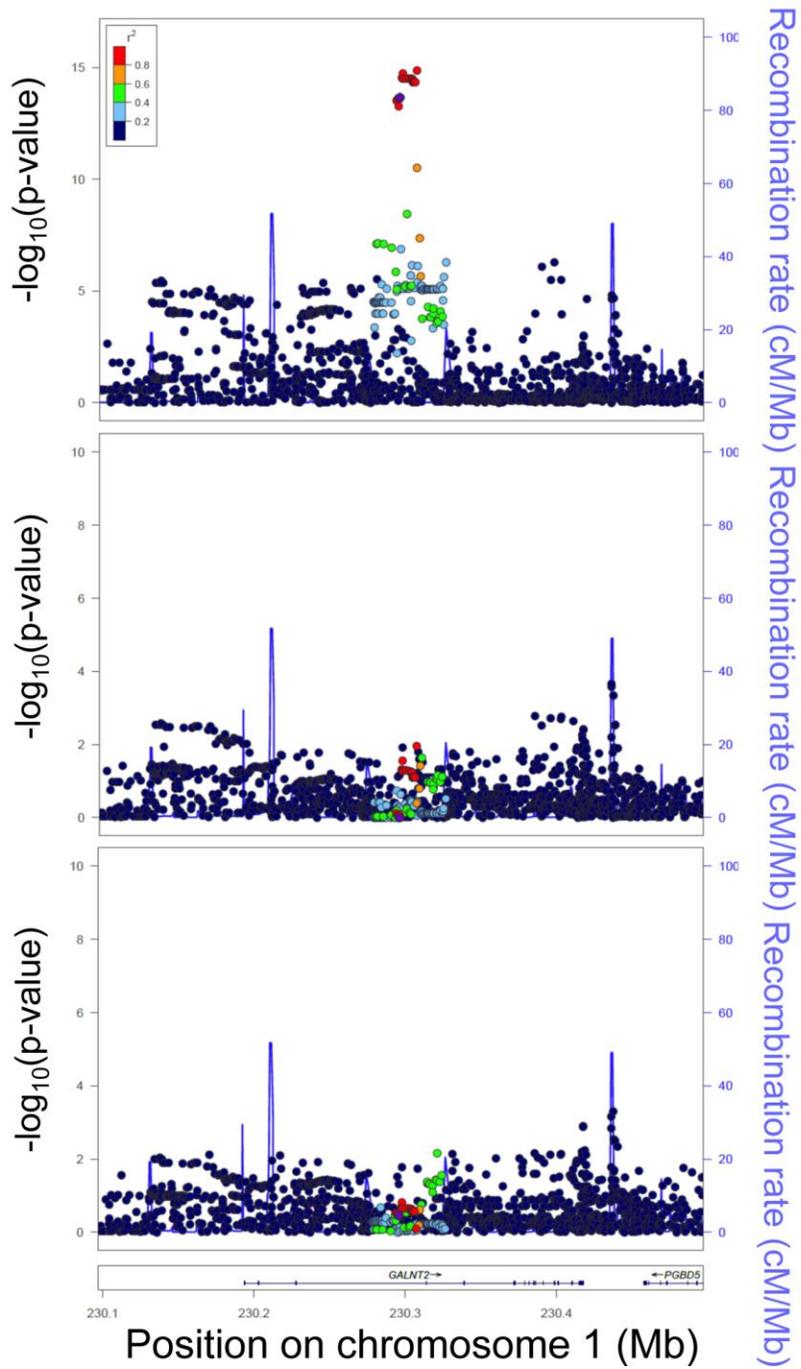


Figure 2.22 Non-coding variants at *GALNT2* are associated with *GALNT2* expression in adipose tissue samples from 1381 individuals from the METSIM study

Figure 2.22 Non-coding variants at *GALNT2* are associated with *GALNT2* expression in adipose tissue samples from 1381 individuals from the METSIM study

The entire association signal (upper panel) is reduced after conditioning on the lead GWAS HDL-C-associated variant rs4846914 (middle panel). The association signal is also diminished after conditioning on the lead eQTL SNP, rs4846922 (lower panel). rs4846922 is in strong LD ($r^2 = 0.82$) with rs4846914. LD was calculated from METSIM genotypes imputed from 1000 Genomes Phase I EUR. Association plot in the upper panel is also shown in Figure 2.21B. The left Y-axis indicates the $-\log_{10}(\text{p-value})$, the right Y-axis indicates the recombination rate (cM/Mb) and the X-axis indicates position on chromosome 1 (Mb).

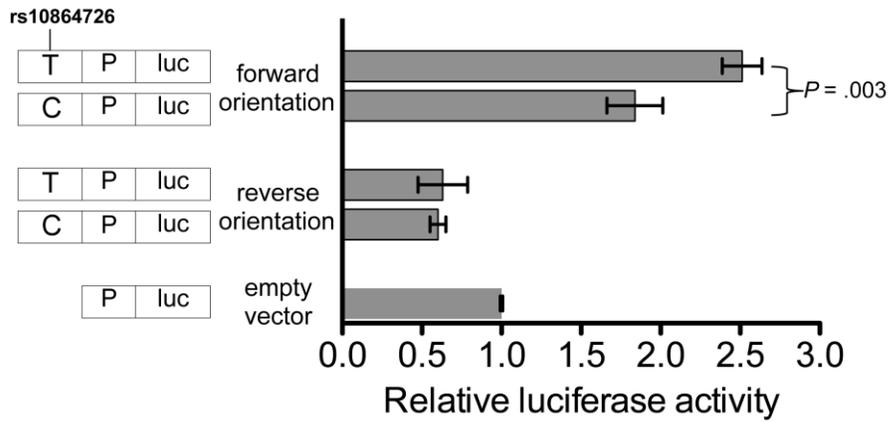


Figure 2.23 An additional SNP, rs10864726, shows modest allelic differences in enhancer activity in the forward but not reverse orientation

Luciferase activity of 204-bp DNA segments containing rs10864726 in 2 different allelic combinations cloned into a pGL4.23 luciferase reporter upstream of the minimal promoter in both orientations. Error bars represent standard deviation of 2-5 independent clones per haplotype. P, promoter; luc, luciferase.

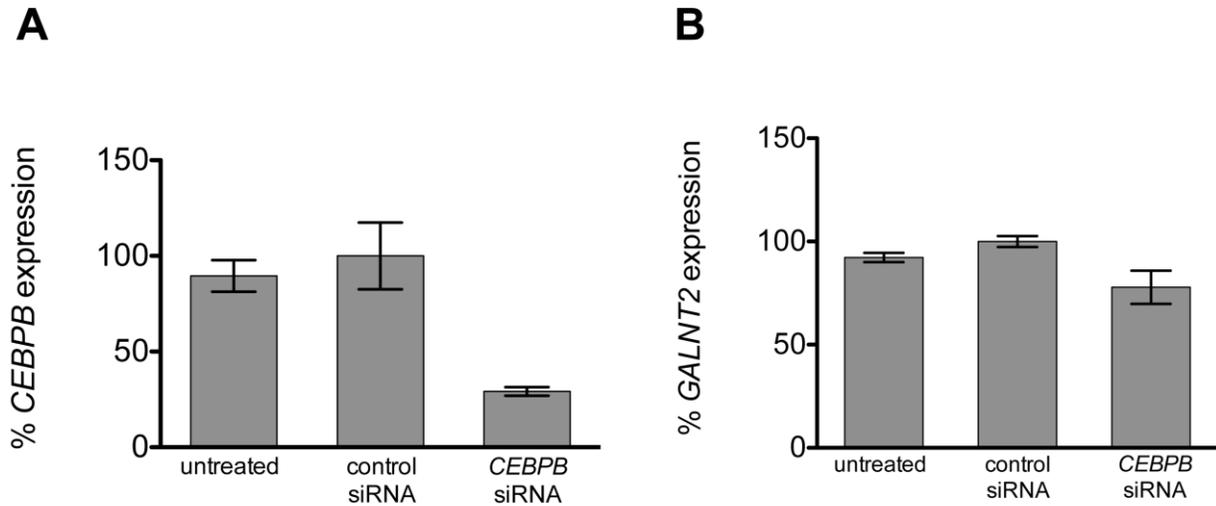


Figure 2.24 siRNA-mediated knockdown of *CEBPB* in HepG2 cells results in a modest decrease in *GALNT2* expression

HepG2 cells were untreated, treated with 50 nM negative control siRNA or treated with 50 nM *CEBPB* siRNA in triplicate, and RNA was harvested 48 hours after transfection of siRNAs. Expression of *CEBPB* and *GALNT2* were measured by qPCR and normalized to *B2M* expression. The y-axes show normalized *CEBPB* or *GALNT2* expression in untreated or *CEBPB* siRNA-treated cells as a percent of normalized *CEBPB* or *GALNT2* expression in negative control siRNA-treated cells. Error bars indicate standard error of the mean of triplicate samples of one experiment.

A % gene expression of *CEBPB*.

B % gene expression of *GALNT2*.

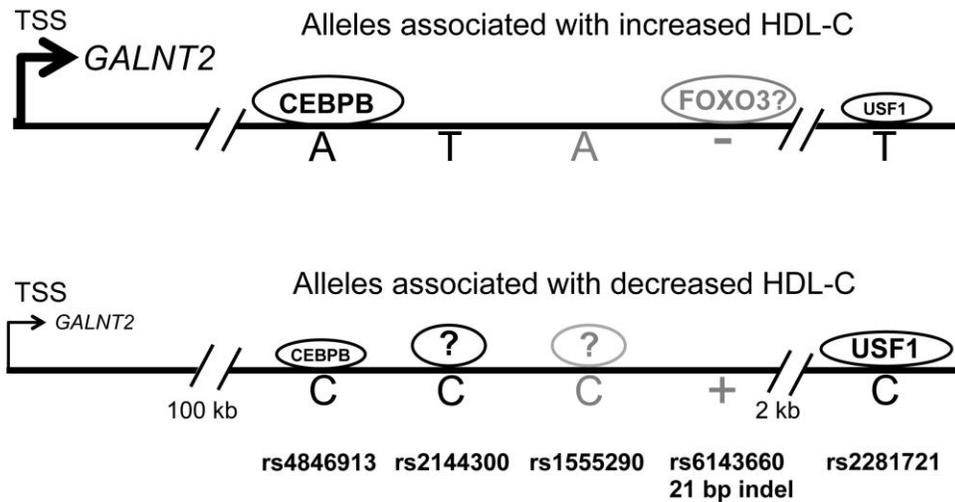


Figure 2.25 Multiple HDL-C-associated variants bind transcription factors at *GALNT2*

Arrows indicate the transcription start site (TSS) of the *GALNT2* gene. Ovals represent proteins that are bound to DNA variant alleles. rs4846913 and rs2281721, shown in black, demonstrated allelic differences in enhancer activity. rs4846913 showed greater CEBPB binding to the A allele in EMSAs and ChIP assays. rs2281721 showed greater USF1 binding to the C allele in EMSAs and USF1 binding to both alleles in ChIP assays. rs2144300, shown in black, showed allelic differences in enhancer reporter activity as part of a haplotype and suggestive differences in protein binding. rs1555290 and rs6143660, shown in gray, showed allelic differences in protein binding in EMSAs. The alleles that showed stronger enhancer reporter activity are the alleles that showed higher *GALNT2* expression in liver and adipose tissues. These alleles are also associated with increased HDL-C.

Lipoprotein Subclass Name	Particle diameter (nm)
Chylomicrons and extremely large VLDL particles	75 nm and above
very large VLDL	64 nm
large VLDL	53.6 nm
medium VLDL	44.5 nm
small VLDL	36.8 nm
very small VLDL	31.3 nm
IDL	28.6 nm
large LDL	25.5 nm
medium LDL	23.0 nm
small LDL	18.7 nm
very large HDL	14.3 nm
large HDL	12.1 nm
medium HDL	10.9 nm
small HDL	8.7 nm

Table 2.1 Definitions of lipoprotein subclass measurements for association analyses in the METSIM study

VLDL, very-low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein;

HDL, high-density lipoprotein

Primer sequences used to amplify DNA for luciferase reporter assays	Sequence 5'-3'	Chromosome 1 position (hg19)
rs4631704_F rs4631704_R	GCATTAAGAAGCCAGGTCCA AGAACAGTGCCACCATCCTC	230293458 230293625
rs4846913- rs2144300- rs1555290- rs6143660_F rs4846913- rs2144300- rs1555290- rs6143660_R	GGCTCTGGCAAAGTGTCTTG TGAATTTCTCCGGTTGACCT	230294386 230295165
rs17315646- rs4846914- rs10127775_F rs17315646- rs4846914- rs10127775_R	TTAGTTGAGGATCAGATGTGTCA TCAGTGAGCAGAACTAAGGACA	230295248 230295812
rs10864726_F rs10864726_R	GCGAGAGTGAGCCTTAGGAA GGGTGGTCTCAAACCTCCTGA	230296062 230296265
rs2281719- rs2281718- rs10864727_F rs2281719- rs2281718- rs10864727_R	TTCATCTGTGAAATGGATATGATG GCTGGGAGACTTAAAGGAAGG	230297453 230297964
rs2281721_F rs2281721_R	CAGGGCAGCATTCTAAGAGG TATTTCTTTGCCACGTCTGC	230297074 230297227
rs10779835_F rs10779835_R	ACGTAACCTGCAATCGCACAC GTACGGGCAATCACACACAC	230299854 230300062
rs11122450_F rs11122450_R	ATGGAGTCTGGGCTTTGATG CAAAGGATTTGTGCAGAGGA	230301742 230301850
rs11122453- rs1546954_F rs11122453- rs1546954_R	GCTCCACTGTGGCCCTTG AACACTTGCTCCGAGTCCAC	230303544 230303893
rs34966440_F rs34966440_R	GTGGACTCGGAGCAAGTGTT CCATTCCTGAAGGCAAGAAC	230303874 230304003
rs11122454- rs4846921_F rs11122454- rs4846921_R	AAAGGACCTGCTGTTCTTGC AAGGGAAAGCAACTACACTGAAA	230303972 230304392
rs10864728- rs11122455- rs10489615- rs1321257_F rs10864728- rs11122455- rs10489615- rs1321257_R	TTGGGAGAGGAAGACCACAG TCATCCTGCACATTGATTCC	230304838 230305341

rs11122456_F rs11122456_R	GCCTTGGGAAATCCAAA GTGCCCATGTAGCAGAAGT	230305793 230306034
rs4846922- rs4846923_F rs4846922- rs4846923_R	AAATTTCTGGCAACCTGTTTT CCGAAAATACCATCTGAATGA	230307110 230307260
rs4846913_F rs4846913_R	ATGCCGCTTTGGATTTCA CATCCAAAAGCAGTGTG	230294687 230294806
rs2144300_F rs2144300_R	GCTGACACTGCTTTTGGAT ACCAAACCTGTCAATTCC	230294787 230294987
rs1555290_F rs1555290_R	CACCTTTGCTGTTGATGT CCTTCGTGTGCAAAGATG	230294932 230295053
rs6143660_F rs6143660_R	ATTGCTCCTTTCTCCGCA TGAATTTCTCCGGTTGAC	230294993 230295165
Primer sequences for site-directed mutagenesis of 780-bp element		
CTA-_F CTA-_R	CTTTGGATTTCAAGTGGCCTCTGCAGC TGATGTAAGAATAATAAATTGCTGCA	
ACA-_F ACA-_R	AAAGGACGCTGTACCCCTGCCCTATT CAGCCAATAGGGCAGGGGTACAGCG	
CCA-_F CCA-_R	CTTTGGATTTCAAGTGGCCTCTGCAGC TGATGTAAGAATAATAAATTGCTGCA	
CCC-_F CCC-_R	ATTGACAGGGTTTTGGTCCATGATTG GGAGAAAGGAGCAATCATGGACCAAA	
Probe sequences for electrophoretic mobility shift assays		
rs4846913_Aforward rs4846913_Areverse	TGGCCTCTGAAGCAATTTA TAAATTGCTTCAGAGGCCA	
rs4846913_Cforward rs4846913_Creverse	TGGCCTCTGCAGCAATTTA TAAATTGCTGCAGAGGCCA	
rs2144300_Tforward rs2144300_Treverse	CGCTGTACCTCTGCCCTAT ATAGGGCAGAGGTACAGCG	
rs2144300_Cforward rs2144300_Creverse	CGCTGTACCCCTGCCCTAT ATAGGGCAGGGGTACAGCG	
rs1555290_Aforward rs1555290_Areverse	GTTTTGGTCAATGATTGCT AGCAATCATTGACCAAAAC	
rs1555290_Cforward rs1555290_Creverse	GTTTTGGTCCATGATTGCT AGCAATCATGGACCAAAAC	
rs2281721_Tforward rs2281721_Treverse	TCCTTGCCTTGTGGTTGCT AGCAACCACAAGGCAAGGA	
rs2281721_Cforward rs2281721_Creverse	TCCTTGCCTCGTGGTTGCT AGCAACCACGAGGCAAGGA	
rs6143660_deletionforward rs6143660_deletionreverse	TTTTTCTCTGAGTTTATT AATAAACTCAGAGAAAAA	
rs6143660_insertionforward rs6143660_insertionreverse	TTTTTCTCTTAAAGTGTTCAGCACTCCCCTGAGTTTATT AATAAACTCAGGGGAGTGCTGAACACTTTAAGAGAAAAA	
Primer sequences for CEBPB ChIP assay amplifying rs4846913 region		
rs4846913_forward rs4846913_reverse	ATGCCGCTTTGGATTTCA CATCCAAAAGCAGTGTG	

Primer sequences for USF1 ChIP assay amplifying rs2281721 region	Sequence 5'-3'
rs2281721_forward rs2281721_reverse	GGTCTCAGGGCAGCATTCTA GGCCATATTTCTTTGCCACG
Primer sequences for GALNT2 mRNA expression assays	Sequence 5'-3'
GALNT2 forward GALNT2 reverse	AAGGAGAAGTCCGGTGAAGCA TTGAGCGTGAACCTTCCACTG
B2M forward B2M reverse	TGTCTGGGTTTCATCCATCCGACA TCACACGGCAGGCATACTCATCTT
Primer sequences for siRNA-mediated CEBPB knockdown experiments	Sequence 5'-3'
CEBPB forward CEBPB reverse	TACTACGAGGCGGACTGCTT GAGAGGAAGTCGTGGTGCTG
GALNT2 forward GALNT2 reverse	AGCTTATGTTGGAGGGACGA CCGACCTGGCTTCATTGT
B2M forward and reverse	Same as primer sequences for GALNT2 mRNA expression assays

Table 2.2 Primer sequences for functional assays

F indicates forward primer, R indicates reverse primer (5'-3').

Trait	N	Genotypes (CC/CT/TT)	MAF	Effect	P Value
Total cholesterol in medium HDL	9810	2101/4883/2826	0.463	0.1015	5.28E-12
Cholesterol esters in medium HDL	9810	2101/4883/2826	0.463	0.1012	6.03E-12
Concentration of medium HDL particles	9810	2101/4883/2826	0.463	0.09769	2.17E-11
Phospholipids in medium HDL	9810	2101/4883/2826	0.463	0.09535	6.43E-11
Ratio of ApoB to ApoA1	10077	2179/5013/2885	0.465	-0.09166	2.93E-10
Free cholesterol in medium HDL	9810	2101/4883/2826	0.463	0.09245	3.34E-10
Phospholipids in large HDL	9810	2101/4883/2826	0.463	0.07902	1.11E-07
Total triglycerides	10079	2179/5015/2885	0.465	-0.07402	3.20E-07
Concentration of large HDL particles	9810	2101/4883/2826	0.463	0.07484	4.92E-07
Mean diameter for HDL particles	9810	2101/4883/2826	0.463	0.06731	5.42E-06
Triglycerides in small VLDL	9810	2101/4883/2826	0.463	-0.06447	1.13E-05
Cholesterol esters in large HDL	9810	2101/4883/2826	0.463	0.06356	1.75E-05
ApoB	10077	2179/5013/2885	0.465	-0.06265	1.88E-05
HDL cholesterol	10077	2178/5014/2885	0.465	0.06243	2.37E-05
ApoA1	10077	2179/5013/2885	0.465	0.06103	2.58E-05
Concentration of small VLDL particles	9810	2101/4883/2826	0.463	-0.06168	2.95E-05
Total cholesterol in large HDL	9810	2101/4883/2826	0.463	0.06177	2.99E-05
Triglycerides in medium VLDL	9810	2101/4883/2826	0.463	-0.06067	3.59E-05
Concentration of medium VLDL particles	9810	2101/4883/2826	0.463	-0.05927	5.88E-05
Free cholesterol in large HDL	9810	2101/4883/2826	0.463	0.05878	7.80E-05
Mean diameter for VLDL particles	9810	2101/4883/2826	0.463	-0.05809	8.27E-05
Free cholesterol in small VLDL	9810	2101/4883/2826	0.463	-0.05712	9.59E-05
Phospholipids in medium VLDL	9810	2101/4883/2826	0.463	-0.05729	0.000103
Phospholipids in small VLDL	9810	2101/4883/2826	0.463	-0.05699	0.0001146
Triglycerides in large VLDL	9810	2101/4883/2826	0.463	-0.05556	0.0001449
Concentration of large VLDL particles	9810	2101/4883/2826	0.463	-0.05540	0.0001521
Cholesterol esters in large VLDL	9810	2101/4883/2826	0.463	-0.05492	0.0001748
Free cholesterol in medium VLDL	9810	2101/4883/2826	0.463	-0.05465	0.0002142
Total cholesterol in medium VLDL	9810	2101/4883/2826	0.463	-0.05383	0.0002357
Triglycerides in small HDL	9810	2101/4883/2826	0.463	-0.05384	0.0002677
Total cholesterol in small VLDL	9810	2101/4883/2826	0.463	-0.05282	0.0003104
Phospholipids in large VLDL	9810	2101/4883/2826	0.463	-0.05264	0.0003617
Cholesterol esters in medium LDL	9810	2101/4883/2826	0.463	-0.05228	0.0004001
Cholesterol esters in medium VLDL	9810	2101/4883/2826	0.463	-0.05169	0.0004156
Total cholesterol in medium LDL	9810	2101/4883/2826	0.463	-0.05181	0.0004519
Total cholesterol in large VLDL	9810	2101/4883/2826	0.463	-0.05049	0.0005621
Total cholesterol in small LDL	9810	2101/4883/2826	0.463	-0.05016	0.0006098

Mean diameter for LDL particles	9810	2101/4883/2826	0.463	0.04941	0.0007732
Triglycerides in very small VLDL	9810	2101/4883/2826	0.463	-0.04859	0.0009471
Concentration of medium LDL particles	9810	2101/4883/2826	0.463	-0.04837	0.001062
Concentration of small LDL particles	9810	2101/4883/2826	0.463	-0.04631	0.00156
Free cholesterol in large VLDL	9810	2101/4883/2826	0.463	-0.04615	0.001693
Concentration of very large VLDL particles	9810	2101/4883/2826	0.463	-0.04601	0.001755
Phospholipids in medium LDL	9810	2101/4883/2826	0.463	-0.04332	0.003093
Triglycerides in very large VLDL	9810	2101/4883/2826	0.463	-0.04344	0.003127
Cholesterol esters in large LDL	9810	2101/4883/2826	0.463	-0.04144	0.005026
LDL cholesterol	10076	2177/5014/2885	0.465	-0.03815	0.008521
Total cholesterol in large LDL	9810	2101/4883/2826	0.463	-0.03813	0.009843
Concentration of large LDL particles	9810	2101/4883/2826	0.463	-0.03517	0.01736
Triglycerides in chylomicrons and largest VLDL	9810	2101/4883/2826	0.463	-0.03199	0.02879
Phospholipids in very large VLDL	9810	2101/4883/2826	0.463	-0.03104	0.0334
Free cholesterol	9985	2164/4960/2861	0.465	-0.03048	0.03626
Phospholipids in very large HDL	9810	2101/4883/2826	0.463	0.03016	0.04097
Triglycerides in IDL	9810	2101/4883/2826	0.463	-0.02929	0.04635
Phospholipids in large LDL	9810	2101/4883/2826	0.463	-0.02903	0.04858
Free cholesterol in large LDL	9810	2101/4883/2826	0.463	-0.02721	0.06557
Concentration of very large HDL particles	9810	2101/4883/2826	0.463	0.02434	0.0992
Total cholesterol	10079	2179/5015/2885	0.465	-0.02111	0.1453
Free cholesterol in very large HDL	9810	2101/4883/2826	0.463	0.01893	0.2023
Esterified cholesterol	9985	2164/4960/2861	0.465	-0.01687	0.2487
Concentration of small HDL particles	9810	2101/4883/2826	0.463	0.01519	0.3002
Total cholesterol in IDL	9810	2101/4883/2826	0.463	-0.01403	0.3423
Concentration of very small VLDL particles	9810	2101/4883/2826	0.463	-0.01391	0.3466
Concentration of IDL particles	9810	2101/4883/2826	0.463	-0.01194	0.4195
Total cholesterol in very large HDL	9810	2101/4883/2826	0.463	0.01114	0.4489
Concentration of chylomicrons and largest VLDL particles	9810	2101/4883/2826	0.463	-0.009914	0.4966
Cholesterol esters in very large HDL	9810	2101/4883/2826	0.463	0.008723	0.5532
Phospholipids in chylomicrons and largest VLDL	9810	2101/4883/2826	0.463	0.008239	0.5732
Phospholipids in IDL	9810	2101/4883/2826	0.463	-0.008056	0.5824
Triglycerides in very large HDL	9810	2101/4883/2826	0.463	0.003751	0.8012
Phospholipids in very small VLDL	9810	2101/4883/2826	0.463	-0.003168	0.8303
Free cholesterol in IDL	9810	2101/4883/2826	0.463	-0.0004156	0.9774

Table 2.3 Evidence of association of GALNT2 SNP rs2144300 with 72 METSIM lipid and cholesterol

traits

Table 2.3 Evidence of association of *GALNT2* SNP rs2144300 with 72 METSIM lipid and cholesterol traits

Traits were adjusted for age, age², smoking status and lipid-lowering medication status and resulting residuals were transformed to a standard normal. N=number of samples, MAF=minor allele frequency.

For rs2144300 (C/T), C is the minor allele, and T is the major allele. Effect corresponds to the major allele (T) of rs2144300.

Unconditioned **Conditioned**
on rs17315646

Variant	Chr1 Position (hg19)	LD (r2)	EA	MAF	N	Effect	P Value	Effect	P Value
rs17315646*	230295307	0.98789	G	0.464	9810	0.102	3.51E-12	-	-
rs10864726*	230296153	0.99074	T	0.463	9810	0.102	4.55E-12	-0.205	0.5216
rs10127775*	230295789	0.9879	T	0.463	9810	0.102	4.72E-12	-0.272	0.4186
rs4846914*	230295691	1	A	0.463	9810	0.102	5.27E-12	-0.383	0.2513
rs2144300*	230294916	0.99003	T	0.463	9810	0.102	5.28E-12	-0.385	0.2486
rs4846913*	230294715	0.9886	A	0.463	9810	0.102	5.39E-12	-0.427	0.2103
rs11122453*	230303611	0.91864	A	0.445	9810	0.0973	4.18E-11	-0.0140	0.7889
rs2281721*	230297136	0.9284	T	0.445	9810	0.0971	4.50E-11	-0.0184	0.7295
rs10489615*	230304988	0.92146	G	0.445	9810	0.0968	5.02E-11	-0.0197	0.7071
rs1321257*	230305312	0.92075	A	0.445	9810	0.0968	5.03E-11	-0.0198	0.7062
rs10864728*	230304914	0.92146	G	0.445	9810	0.0968	5.06E-11	-0.0199	0.7037
rs4846921*	230304352	0.92146	A	0.445	9810	0.0968	5.12E-11	-0.0203	0.6988
rs11122455*	230304930	0.92005	G	0.445	9810	0.0966	5.52E-11	-0.0216	0.6781
rs11122454*	230304051	0.92213	T	0.445	9810	0.0965	5.92E-11	-0.0245	0.6405
rs1546954*	230303848	0.92072	G	0.445	9810	0.0964	6.00E-11	-0.0249	0.6353
rs2281719*	230297659	0.92351	T	0.445	9810	0.0963	6.37E-11	-0.0266	0.612
rs2281718*	230297778	0.92281	T	0.445	9810	0.0963	6.37E-11	-0.0266	0.612
rs11122456*	230305966	0.91145	G	0.437	9810	0.0960	7.03E-11	-0.00772	0.8637
rs10864727*	230297939	0.9221	G	0.446	9810	0.0958	8.07E-11	-0.0299	0.5613
rs10779835*	230299949	0.91479	C	0.438	9810	0.0955	8.83E-11	-0.0126	0.7796
rs11122450*	230301811	0.90916	G	0.438	9810	0.0953	9.93E-11	-0.0145	0.7469
rs4631704*	230293530	0.93949	T	0.464	9810	0.0949	1.20E-10	-0.217	0.01306
rs609526	230308906	0.69241	A	0.450	9810	-0.0943	1.97E-10	-0.0295	0.2546
rs4846922*	230307182	0.7818	C	0.415	9810	0.0899	1.45E-09	-0.0176	0.6066
rs590820	230309619	0.66157	G	0.470	9810	-0.0875	3.16E-09	-0.0172	0.4719
rs4846923	230307222	0.58652	G	0.353	9810	0.0879	9.39E-09	0.0112	0.652
rs10779836	230303150	0.33687	T	0.227	9810	0.0988	1.41E-08	0.0438	0.03919
rs11122449	230300481	0.54112	T	0.324	9810	0.0866	3.07E-08	0.0145	0.5311
rs10779833	230293037	0.34238	G	0.345	9810	0.0836	8.01E-08	0.0253	0.209
rs4846917	230299222	0.38327	C	0.254	9810	0.0901	8.06E-08	0.0292	0.1715
rs4846841	230302661	0.38327	G	0.254	9810	0.0901	8.08E-08	0.0292	0.1716
rs4846840	230302521	0.38327	G	0.254	9810	0.0901	8.10E-08	0.0292	0.1718
rs6672758	230303512	0.36541	T	0.260	9810	0.0890	9.56E-08	0.0276	0.197
rs1555290	230294989	0.25101	A	0.196	9810	0.0977	1.00E-07	0.0433	0.0435
rs4846919	230301451	0.25918	G	0.197	9810	0.0972	1.19E-07	0.0418	0.05218

rs966333	230305477	0.25101	C	0.195	9810	0.0974	1.21E-07	0.0420	0.05129
rs4846916	230296492	0.2536	A	0.195	9810	0.0970	1.34E-07	0.0415	0.05416
rs4846915	230296470	0.2536	A	0.195	9810	0.0970	1.34E-07	0.0415	0.0542
rs2296065	230301776	0.26045	A	0.196	9810	0.0967	1.37E-07	0.0411	0.05672
rs2103827	230298285	0.26143	A	0.196	9810	0.0967	1.37E-07	0.0411	0.05676
rs12065546	230295245	0.2536	T	0.195	9810	0.0969	1.37E-07	0.0414	0.05498
rs4846918	230300586	0.26143	T	0.196	9810	0.0967	1.39E-07	0.0410	0.05709
rs627108	230310568	0.51376	T	0.416	9810	-0.0787	1.40E-07	-0.0148	0.4676
rs910502	230294185	0.25266	T	0.195	9810	0.0968	1.42E-07	0.0413	0.05584
rs4846920	230301574	0.26143	G	0.197	9810	0.0963	1.44E-07	0.0405	0.06021
rs4846905	230279582	0.23473	G	0.239	9810	0.0898	1.44E-07	0.0409	0.03751
rs2144299	230280366	0.23473	G	0.239	9810	0.0898	1.44E-07	0.0409	0.03754
rs4846907	230284703	0.23473	C	0.239	9810	0.0898	1.45E-07	0.0409	0.03763
rs2352723	230291868	0.23422	C	0.238	9810	0.0898	1.46E-07	0.0409	0.03757
rs4846910	230291441	0.23473	A	0.238	9810	0.0898	1.46E-07	0.0409	0.03764
Top SNPs (p < 0.001) associated with total cholesterol in medium HDL after conditional analysis on rs17315646									
rs188527357	230262751	<0.2	A	0.034	9810	0.102	0.0116	0.161	0.00009
rs185682763	230262536	<0.2	A	0.035	9810	0.0957	0.0166	0.154	0.00015
rs76230683	230161280	<0.2	T	0.001 1	9810	-1.09	0.000218	-1.11	0.00017
rs186540557	230207407	<0.2	T	0.000 90	9810	-1.29	0.000262	-1.32	0.00018
rs148414409	230210718	<0.2	G	0.029	9810	0.0947	0.0327	0.151	0.00075

Table 2.4 Association of *GALNT2* locus variants with total cholesterol in medium HDL in the METSIM study

Table 2.4 Association of *GALNT2* locus variants with total cholesterol in medium HDL in the METSIM study

Evidence of *GALNT2* variant association with total cholesterol in medium HDL-C in 9,810 individuals from the METSIM study. Effect represents the change in standard-normalized residuals of total cholesterol in medium HDL level after adjustment for age, age², smoking status and lipid-lowering medication status. Positive signs for effect indicate positive direction of effect. Variants were either directly genotyped (9 variants) or imputed (91 variants) using a central-northern European reference panel (the top 50 variants are shown in this table). Positions on chromosome 1 (Chr1 position) are hg19 coordinates, variants names are from dbSNP138. Asterisks indicate variants tested in luciferase reporter assays. LD (r^2), linkage disequilibrium r^2 values with lead HDL-C-associated variant rs4846914 calculated from the METSIM cohort; EA, effect allele; MAF, minor allele frequency; N, number of samples.

Variant	Chr1 position (hg19)	r ²	ENCODE open chromatin and histone modifications in liver cell types	Epigenome Atlas histone modifications in adult liver	TF ChIP-seq peaks in HepG2
rs4631704	230293530	0.96	HepG2 H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac	Adult liver H3K4me1	POLR2A
rs4846913	230294715	1	HepG2 FAIRE; HepG2, hepatocytes and Huh-7 DNase; HepG2 H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac	Adult liver H3K4me1	ARID3A, BHLHE40, CEBPB, CEBPD, CHD2, FOSL2, FOXA1, FOXA2, HDAC2, HNF4A, HNF4G, JUND, MAFF, MAFK, MAX, MAZ, MBD4, MXI1, MYBL2, MYC, NFIC, NR2F2, EP300, RAD21, RCOR1, REST, RXRA, SIN3A, SMC3, SP1, TBP, TCF4, TCF12, TEAD4, USF1
rs2144300	230294916	1	HepG2 FAIRE; HepG2, hepatocytes and Huh-7 DNase; HepG2 H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac	Adult liver H3K4me1	CEBPB, FOSL2, FOXA1, FOXA2, HDAC2, HNF4A, HNF4G, JUND, MAFF, MAFK, MAX, MBD4, MYBL2, MYC, NFIC, NR2F2, EP300, RAD21, REST, RXRA, SIN3A, SP1, TCF12, TEAD4
rs6143660	230295079	0.96	HepG2 FAIRE; HepG2 and Huh-7 DNase; HepG2 H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac	Adult liver H3K4me1	FOXA1, FOXA2, HDAC2, HNF4G, MAX, MBD4, MYBL2, NFIC, NR2F2, EP300, RXRA, SP1, TEAD4
rs17315646	230295307	1	HepG2 H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac	Adult liver H3K4me1	
rs4846914	230295691	1	Hepatocytes DNase; HepG2 H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac	Adult liver H3K4me1	
rs10127775	230295789	1	Hepatocytes DNase; HepG2 H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac	Adult liver H3K4me1	
rs10864726	230296153	0.99	HepG2 H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac	Adult liver H3K4me1	
rs2281721	230297136	0.96	HepG2 H3K4me1, H3K4me2, H3K9ac, H3K27ac	Adult liver H3K4me1	USF1
rs2281719	230297659	0.96	HepG2 H3K4me1, H3K4me2, H3K27ac	Adult liver H3K4me1	
rs2281718	230297778	0.91	HepG2 H3K4me1, H3K4me2, H3K27ac	Adult liver H3K4me1	
rs10864727	230297939	0.96	HepG2 H3K4me1, H3K4me2, H3K27ac	Adult liver H3K4me1	
rs10779835	230299949	0.96	HepG2 H3K4me1, H3K27ac	Adult Liver H3K4me1, H3K9ac	
rs11122450	230301811	0.96	HepG2 H3K4me1, H3K27ac	Adult Liver H3K4me1	
rs11122453	230303611	0.95	HepG2 H3K4me1		RAD21, CTCF
rs1546954	230303848	0.95	HepG2 H3K4me1		

rs34966440	230303931	0.94	HepG2 H3K4me1		
rs11122454	230304051	0.95	HepG2 H3K4me1		
rs4846921	230304352	0.95	HepG2 H3K4me1		
rs10864728	230304914	0.94	HepG2 H3K4me1		
rs11122455	230304930	0.94	HepG2 H3K4me1		
rs10489615	230304988	0.94	HepG2 H3K4me1		
rs1321257	230305312	0.94	HepG2 H3K4me1	Adult Liver H3K9ac	
rs11122456	230305966	0.95	HepG2 H3K4me1	Adult Liver H3K9ac	
rs4846922	230307182	0.77	HepG2 H3K4me1		

Table 2.5 Variants in strong linkage disequilibrium ($r^2 > 0.7$) with the GWAS HDL-C-associated variant rs4846914 overlap ENCODE open chromatin, histone ChIP-seq and transcription factor ChIP-seq peaks in liver cell types

ENCODE open chromatin and histone modification peaks in liver cell types were based on ENCODE Open Chromatin by DNase I HS and FAIRE tracks and Broad Histone and UW Histone tracks.

Transcription factor ChIP-seq peaks were based on Integrated Regulation from ENCODE tracks and ENCODE Transcription Factor Binding tracks. Linkage disequilibrium values from 1000 Genomes phase 1 version 3 CEU, Northern Europeans from Utah, are shown in the r^2 column.

Gene	eQTL P value					Probeset
	rs4846914	rs4846913	rs2281721	rs2144300	rs4846922 *	
<i>GALNT2</i>	2.2E-14	2.4E-14	2.9E-15	2.4E-14	1.4E-15	11737879_s_at
<i>PGBD5</i>	0.0084	0.010	0.018	0.010	0.0069	11722329_at
<i>CAPN9</i>	0.023	0.028	0.016	0.028	0.011	11737581_a_at
<i>FAM89A</i>	0.071	0.084	0.046	0.084	0.095	11731560_at
<i>ARV1</i>	0.12	0.099	0.13	0.099	0.75	11738018_a_at
<i>CCSAP</i>	0.16	0.18	0.080	0.18	0.12	11720144_s_at
<i>TAF5L</i>	0.18	0.22	0.23	0.22	0.082	11736606_a_at
<i>C1orf198</i>	0.18	0.22	0.16	0.22	0.34	11741655_a_at
<i>TRIM67</i>	0.20	0.25	0.104	0.25	0.067	11752381_a_at
<i>NUP133</i>	0.29	0.29	0.48	0.29	0.78	11723924_a_at
<i>ABCB10</i>	0.30	0.34	0.25	0.34	0.096	11724634_s_at
<i>URB2</i>	0.41	0.45	0.34	0.45	0.79	11746690_a_at
<i>COG2</i>	0.44	0.39	0.61	0.39	0.75	11743607_a_at
<i>ACTA1</i>	0.54	0.57	0.45	0.57	0.45	11743108_a_at
<i>RAB4A</i>	0.60	0.53	0.56	0.53	0.39	11721301_a_at
<i>TTC13</i>	0.77	0.70	0.91	0.70	0.86	11723590_a_at
<i>AGT</i>	0.85	0.91	0.92	0.90	0.39	11730726_s_at

Table 2.6 SNP associations with gene expression levels in primary subcutaneous adipose tissue samples from 1381 individuals from the METSIM study

Primary subcutaneous adipose eQTL p-values for genes located within 1 Mb from the lead HDL-C-associated variant rs4846914 and functional candidate variants rs4846913, rs2144300 and rs2281721 (all variants $r^2 > 0.7$ with rs4846914). *The lead eQTL-associated variant, rs4846922, exhibited the strongest evidence of association with *GALNT2* expression. Probesets, gene probesets from the Affymetrix Human Genome U219 Array.

SNP	conditional variant					
	none	rs4846913	rs2144300	rs4846914	rs2281721	rs4846922 *
rs4846913	2.4E-14	n.a.	0.8756	0.87	0.98	0.29
rs2144300	2.4E-14	0.88	n.a.	0.87	0.98	0.29
rs4846914	2.2E-14	0.70	0.71	n.a.	0.79	0.35
rs2281721	2.9E-15	0.042	0.044	0.047	n.a.	0.21
rs4846922 *	1.4E-15	0.0074	0.0077	0.011	0.062	n.a.

Table 2.7 Initial and residual associations with *GALNT2* expression in primary subcutaneous adipose tissue samples from 1381 individuals from the METSIM study

P-values of the lead HDL-C-associated variant rs4846914 and other functional candidate variants, rs4846913, rs2144300, and rs2281721 (all variants $r^2 > 0.7$ with rs4846914) after conditioning on each respective variant (conditional variant). The asterisk indicates the lead eQTL variant, rs4846922.

ENDNOTES

¹Chapter 2 is featured as an article, “Multiple hepatic regulatory variants at the *GALNT2* GWAS locus associated with high-density lipoprotein cholesterol” in *The American Journal of Human Genetics*, Volume 97 Issue 6, Pages 801-815.

CHAPTER 3: A FUNCTIONAL REGULATORY VARIANT ASSOCIATED WITH TYPE 2 DIABETES, FASTING GLUCOSE AND DECREASED *ADCY5* EXPRESSION

3.1 Introduction

Genome-wide association studies (GWAS) have identified more than 80 loci associated with type 2 diabetes[66]; however the underlying biological and molecular mechanisms responsible for many of these associations are still unknown. One association signal for type 2 diabetes is located within introns 1-3 of the *ADCY5* gene. The A allele of rs11708067 has been associated with risk of type 2 diabetes in a European population GWAS meta-analysis (n = 34,840 T2D cases, 114,981 controls, $p = 7.2 \times 10^{-14}$)[141], in a candidate variant study in a South Asian population (n = 1,678 T2D cases, 1,584 controls, $p = 9.1 \times 10^{-4}$)[142] and in a candidate gene study in an African American population (n = 2,806 T2D cases, 4,265 controls, $p = 4.7 \times 10^{-3}$)[143]. The T allele of rs11717195, (LD $r^2 = 0.91$ with rs11708067, 1000 Genomes Phase 1 EUR[144], was also associated with risk of type 2 diabetes in a trans-ethnic GWAS meta-analysis (n = 26,488 T2D cases, 83,964 controls, $p = 2.2 \times 10^{-8}$)[67] and in a European population GWAS meta-analysis (n = 34,840 T2D cases, 114,981 controls, $p = 6.5 \times 10^{-14}$)[141]. In an African population, rs143882978 (LD $r^2 = 0.96$ with rs11708067, 1000 Genomes YRI) was also found to be associated with T2D (n = 1,035 T2D cases, 740 controls, $p = 0.02$)[145].

Variants at the *ADCY5* locus are also associated with other glycemic traits. The A allele of rs11708067 was shown to be associated with increased fasting glucose levels (n = 118,475, $p = 7.1 \times 10^{-22}$) and decreased HOMA-B (n = 94,212, $p = 2.5 \times 10^{-12}$) in Europeans[146]. An additional variant, rs2877716, in strong linkage disequilibrium (LD $r^2 = 0.80$, 1000 Genomes Phase 1, EUR [144]) with rs11708067, was found to be associated with 2-hour glucose level (n = 44,225, $p = 4.2 \times 10^{-16}$)[147]. Associations with fasting glucose, 2-hour glucose, insulinogenic index, and 2-hour insulin have also been reported in Asian Indians (rs9883204, n=2,151, $p \leq 0.05$)[148]. A summary of published *ADCY5* variant associations with T2D and glucose-related traits is shown in Table 3.1. *ADCY5* variant alleles associated

with increased glucose have also been reported to be associated with decreased birthweight (rs9883204, LD $r^2 = 0.79$, 1000 Genomes Phase 1 EUR, $n = 61,509$ $p = 5.5 \times 10^{-20}$)[149].

ADCY5 encodes adenylate cyclase 5, which catalyzes generation of cyclic AMP from ATP[150]. Cyclic AMP, ATP and calcium play crucial roles in insulin secretion in the pancreatic beta cell[151]. Previous studies have established *ADCY5* as a compelling biological candidate influencing glucose metabolism, showing association of the T2D risk alleles with decreased *ADCY5* mRNA expression[152, 153]. Furthermore, *ADCY5* knockdown in human islets reduced glucose-stimulated cAMP levels, inhibiting glucose metabolism toward ATP[152].

In this study, we aimed to identify the variant(s) responsible for the T2D and glycemic trait associations at this locus. Based on 1000 Genomes phase 1 version 3 EUR data[94], 13 variants exhibit strong linkage disequilibrium ($r^2 > 0.8$) with rs11708067, and all 14 candidates including rs11708067 are located within *ADCY5* introns 1-3. We hypothesized that one or more of these variants may regulate gene expression. To identify the most plausible functional regulatory variants, we used multiple regulatory datasets in human islets, including open chromatin, chromatin states, and H3K27ac histone modification ChIP-seq data[72, 79, 80, 154]. Of the 14 candidate variants, only one variant, rs11708067, overlaps both DNase and FAIRE peaks and a predicted strong enhancer chromatin state in primary human islets. We demonstrate that rs11708067 exhibits allelic differences in reporter assays of transcriptional activity consistent with the association between the T2D risk alleles and decreased expression of *ADCY5*, and show evidence of differential transcription factor binding to the rs11708067-A T2D risk allele.

3.2 Materials and Methods

3.2.1 Lookup of expression quantitative trait loci in primary human islets

We looked for evidence of association between the T2D and fasting glucose GWAS variant region and gene expression using primary human islet RNA-seq eQTL data from the Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus (FUSION) study and from a set of previously published islet data[155]. RNA-seq data from 82 individuals were aligned using STAR (version 2.3.1y)[156]. Analysis of cis-eQTL data for SNPs within 1 Mb of the closest upstream TSS of each gene was performed using Matrix eQTL[157]. Gene expression values were calculated based on inverse

normalized fragments per kilobase of transcript per million mapped reads (FPKM), factor analysis via probabilistic estimation of expression residuals (PEER)[158] on the inverse normalized FPKM and inverse normalization of the resulting residuals. A linear regression model considering an additive genetic effect was used.

3.2.2 H3K27ac ChIP-seq in a primary human islet sample

Human pancreatic islets were obtained through the Islet Cell Resource (ICR) Basic Science Islet Distribution Program and National Disease Research Interchange in accordance with regulations of Human Subjects research. H3K27ac ChIP-seq experiments and read mapping were performed as described[80, 154]. For H3K27 ChIP experiments, rabbit anti-H3K27ac antibodies from Abcam were used (ab4729)[154]. We aligned sequence reads to the UCSC Genome Browser build hg19 considering variants from the 1000 Genomes Project using GSNAP[114] with AA-ALIGNER[113]. We determined significance of allelic imbalance at rs11708067 by using an exact binomial test, based on the number of reads containing the reference allele and the total number of reads at the heterozygous site.

3.2.3 Cell culture

MIN6 mouse insulinoma beta cells[159] were grown in Dulbecco's modified Eagle's Medium (Sigma-Aldrich) with 10% fetal bovine serum, 1 mM sodium pyruvate, and 100 μ M beta-mercaptoethanol. INS-1-derived 832/13 rat insulinoma beta cells (a gift from C. Newgard, Duke University) were grown in RPMI-1640 medium (Corning) supplemented with 10% fetal bovine serum, 10 mM HEPES, 2 mM L-glutamine, 1 mM sodium pyruvate, and .05 mM beta-mercaptoethanol. Cell lines were maintained at 37° C and 5% CO₂. MIN6 or 832/13 cells (200,000 cells per well) were seeded in a 24-well tissue culture treated plate one day before transfections.

3.2.4 Transcriptional reporter assays

To test candidate variants for allelic differences in regulatory activity, 231-base pair segments were amplified from DNA of individuals homozygous for each allele by PCR. Primers are listed in Table 3.2. Amplicons were digested using XhoI and KpnI restriction enzymes (New England BioLabs) and

cloned into the multiple cloning region of firefly luciferase reporter vector pGL4.23 in both orientations with respect to the minimal promoter (Promega). Three to five independent plasmids for each allele of rs11708067 were purified and confirmed by Sanger sequencing. Plasmids containing each allele were transfected into MIN6 mouse or 832/13 rat insulinoma cells in duplicate. For transfection of MIN6 cells, 250 ng of plasmid DNA was added to each well with Lipofectamine LTX (ThermoFisher Scientific) and Opti-MEM® (Gibco®). For transfection of 832/13 cells, 770 ng of plasmid was added to each well using Lipofectamine 2000 (ThermoFisher Scientific) and Opti-MEM® (Gibco®). To control for transfection efficiency, a pHRL-TK *Renilla* luciferase reporter vector was also transfected into MIN6 and 832/13 cells (80 ng each well). Cells were incubated at 37° C with 5% CO₂ for 48 hours, and luciferase assays were performed with 832/13 cell lysates using the Dual-Luciferase® Reporter Assay System (Promega). Firefly luciferase readings were normalized to *Renilla* luciferase readings and compared to pGL4.23 minimal promoter empty vector control activity using two-tailed t-tests.

3.2.5 Electrophoretic mobility shift assays

Complementary oligonucleotide probes (21 base-pairs) centered on rs11708067 variant alleles (Table 3.2) were synthesized by Integrated DNA Technologies, Inc. Probes were labeled with biotin on the 5' end. To create double stranded DNA probes, single-stranded oligonucleotides were incubated in 10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 8.0 at 95°C for 5 minutes followed by 70 one-minute cycles with temperature decreasing by 1°. Nuclear lysates from 832/13 rat insulinoma beta cells were prepared using Pierce NE-PER® Nuclear and Cytoplasmic Extraction Reagents according to manufacturer's instructions. EMSA reactions were performed using the LightShift® Chemiluminescent Kit (Thermo Fisher Scientific). Briefly, binding buffer and Poly(dI·dC) were added to 4 µg of 832/13 nuclear lysate and incubated at room temperature for ten minutes. Biotin-labeled DNA probes were added and the reaction was incubated for 30 minutes. For unlabeled competition reactions 125-fold excess unlabeled double-stranded probe was added to binding buffer, poly(dI·dC) and 832/13 nuclear lysate and incubated for 10 minutes before addition of biotin-labeled DNA. Reactions were loaded into a 6% DNA retardation gel (Life Technologies), subjected to electrophoresis for 2 hours at 100V, transferred to Biodyne B Nylon Membranes (Thermo Fisher Scientific) and crosslinked by UV. Subsequent wash and detection steps were performed

according to Chemiluminescent Nucleic Acid Detection Module (Thermo Fisher Scientific) instructions. For EMSA supershift assays, 6-8 µg of CEBPA (sc-61X), CEBPB (sc-150X) or NKX2.2 (sc-25404X) antibodies (all from Santa Cruz Biotechnology, Inc.) were added to 832/13 nuclear lysate, binding buffer, and poly(dI-dC) for 20 minutes before addition of labeled DNA probes.

3.3 Results

3.3.1 rs11708067 is associated with *ADCY5* expression in primary human pancreatic islets

We examined evidence of association between rs11708067 and expression of nearby genes in a preliminary analysis of islet eQTL data from 112 human pancreatic islet samples, including eQTL data previously reported from 81 islet donors of European ancestry[155]. The T2D risk rs11708067-A allele showed the strongest association with decreased *ADCY5* expression ($p = 4.2 \times 10^{-9}$, Figure 3.1), consistent with previous reports[152, 153].

3.3.2 rs11708067 overlaps evidence of open chromatin and a predicted strong enhancer region in human pancreatic islets

We aimed to determine whether the GWAS lead variants or variants in strong linkage disequilibrium showed evidence of regulatory activity. Open chromatin data, including DNase I hypersensitivity-sequencing (DNase-seq) data and formaldehyde-assisted isolation of regulatory elements (FAIRE)-sequencing data, often mark regulatory elements[74, 75]. We used DNase and FAIRE data in primary human pancreatic islets to investigate potential regulatory regions overlapping the association signal at the *ADCY5* locus[79, 80]. We asked if any of the variants in strong linkage disequilibrium ($r^2 > 0.8$, 1000 Genomes phase 1 EUR) with any of the lead T2D or glycemic trait lead GWAS variants overlapped regions of open chromatin. Among 14 candidate variants, rs11708067 overlaps a predicted strong enhancer region in human pancreatic islets[154], a transposase-accessible peak from an Assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-seq) in pancreatic islet alpha cells[160] and is the only variant to overlap both a DNase and FAIRE peak in human pancreatic islets (Figure 3.2).

3.3.3 Allelic imbalance of H3K27ac reads overlapping rs11708067 in human pancreatic islets

In addition to open chromatin data, histone modifications, such as H3K4me1, H3K4me2, and H3K27ac, also often mark regulatory enhancer regions. Using H3K27ac ChIP-seq data[154] in a primary human islet sample, we investigated the region spanning rs11708067. H3K27ac ChIP-seq reads containing rs11708067 showed evidence of allelic imbalance, with 69 reads containing the rs11708067-G allele compared to 9 reads containing rs11708067-A reference allele (Figure 3.3, binomial $p = 1.2 \times 10^{-4}$). These data are consistent with the eQTL data; rs11708067-G allele shows higher expression and greater H3K27ac enrichment.

3.3.4 rs11708067 shows allelic differences in enhancer activity

To investigate whether rs11708067 shows allelic differences in transcriptional activity, we tested a 231-bp DNA segment containing rs11708067 for allelic differences in enhancer activity in luciferase reporter assays in MIN6 mouse insulinoma and 832/13 rat insulinoma cell lines. In MIN6 cells, we observed 2.6- to 8.3-fold enhancer activity compared to an empty vector control in the forward orientation, with significant differences between the alleles (Figure 3.4A, $p = 1.9 \times 10^{-3}$). In the reverse orientation, the segment exhibited 2.3- to 13.9-fold enhancer activity compared to empty vector with significant differences between the alleles (Figure 3.4A, $p = 4.9 \times 10^{-5}$). In 832/13 cells, we observed 1.6- to 3.3-fold enhancer activity compared to an empty vector control in the forward orientation, with significant differences between the alleles (Figure 3.4B, $p = 1.0 \times 10^{-3}$). In the reverse orientation, we observed 0.84- to 1.3-fold luciferase activity, with significant differences between the alleles (Figure 3.4B, $p = 3.0 \times 10^{-3}$). In both cell types and both the forward and reverse orientations, the rs11708067-A allele showed lower luciferase activity, suggesting that the rs11708067-G allele may be located within an enhancer element that increases transcriptional activity of *ADCY5* or another nearby gene.

3.3.5 The alleles of rs11708067 exhibit differential protein binding

To determine whether there was differential transcription factor binding at rs11708067, we performed EMSAs with 21-bp biotinylated probes containing either the rs11708067-A or rs11708067-G alleles. We observed three bands specific to the rs11708067-A allele compared to the rs11708067-G

allele (Figure 3.5, lane 2 vs. lane 6). Addition of 125-fold unlabeled probe containing the A allele of rs11708067 competed away the A-allele-specific bands more effectively than unlabeled probe containing the G allele (Figure 3.5, lanes 3 and 4). Using bioinformatic prediction tools (JASPAR[111], TRANSFAC, Regulome Browser[161]) to search for transcription factor motif matches to the sequence spanning rs11708067, we found that DNA region spanning rs11708067 overlaps a NKX2.2 transcription factor ChIP-seq peak[161]. However, incubation of the EMSA reaction with an antibody to NKX2.2 did not result in a supershift (data not shown). Additional antibodies tested in EMSAs, including CEBPA and CEBPB, also did not demonstrate evidence of supershifts. Additional experiments will be necessary to identify the transcription factor(s) binding to rs11708067 and to determine whether the transcription factors influence allelic differences in transcriptional activity and modulate *ADCY5* expression levels.

3.4 Discussion

In this study, we show evidence of a functional regulatory variant at the *ADCY5* association signal for fasting glucose levels, 2-hour glucose levels and type 2 diabetes[141, 146, 147]. The proposed regulatory variant, rs11708067, overlaps multiple regulatory datasets, including DNase and FAIRE peaks in human pancreatic islets[79, 80], a transposase-accessible ATAC-seq peak in pancreatic islet alpha cells[160] and a predicted strong enhancer region in human pancreatic islets[154]. Additionally, the rs11708067-G allele exhibited increased enrichment of H3K27ac regulatory marks compared to the rs11708067-A allele, which is consistent with the observation that the G allele showed higher enhancer activity in transcriptional reporter assays. The rs11708067-A allele showed lower transcriptional activity and allele-specific binding of nuclear proteins, suggesting that the A allele binds a transcriptional repressor protein. The rs11708067-A allele also showed lower expression of *ADCY5* in our lookup of eQTL data in human pancreatic islets and in previously published data from human pancreatic islets[152, 153].

Determining the functional regulatory variants at GWAS loci and the genes they may regulate can help elucidate the mechanisms by which variants increase or decrease gene expression. Previously, the A allele of rs11708067 was shown to be associated with lower *ADCY5* expression in 7 and 118 human islet samples[152, 153]. We confirmed that the rs11708067-A T2D risk and fasting glucose-raising allele

showed lower *ADCY5* expression in 112 human islets (81 published[155] samples and 31 new islet samples). These data suggest that *ADCY5* is a likely target gene, and that lower *ADCY5* expression may increase fasting glucose and lead to increased risk of T2D.

ADCY5 encodes an enzyme that influences the generation of cyclic AMP (cAMP), a crucial signaling molecule in the pancreatic beta cell[152]. cAMP plays roles in Ca²⁺ signaling and influx[162], as well as in insulin granule exocytosis[163]. In *ADCY5* knockdown studies in primary human pancreatic islets, *ADCY5* was shown to be required to couple glucose signals to intracellular Ca²⁺ increases to drive insulin granule exocytosis, supporting a role for *ADCY5* in glucose and insulin metabolism[152].

ADCY5 is expressed in a number of different tissues, including adipose tissue, pancreatic islets, and pituitary, with highest levels of expression in the heart, tibial artery, brain and esophagus[73, 154]. RNA-seq data from many other cell types, including human umbilical vein endothelial cells and human skeletal muscle myoblasts, show low or no expression of *ADCY5*[154]. We did not observe evidence of eQTLs for the GWAS variant rs11708067 and *ADCY5* expression in peripheral blood cells[164]. *ADCY5* may play a functional role in islets based on *ADCY5* biology studies[152], although *ADCY5* expression in other cell types might also contribute to biological mechanisms of T2D, as alterations in *TCF7L2* T2D-associated gene expression were shown to alter glucose and insulin metabolism in both liver and pancreatic beta cells[165-169]. The T2D-associated lead GWAS variant, rs11708067, was associated with impaired proinsulin-to-insulin conversion, as individuals homozygous for the T2D-risk A allele showed higher proinsulin levels and a higher proinsulin/insulin ratio after an oral glucose tolerance test ($p = 2.0 \times 10^{-3}$)[170]. Considering that proinsulin is synthesized in the endoplasmic reticulum of pancreatic islet beta cells[171], these data provide additional evidence that *ADCY5* influences glucose and insulin metabolism in islets and may be a likely target gene underlying the association with multiple glucose-related traits.

Our study has some limitations. We did not examine other candidate variants that were in strong linkage disequilibrium ($r^2 > 0.8$) with rs11708067 for regulatory activity in luciferase reporter assays. rs11708067 was the strongest candidate for predicted regulatory function, however, 2 additional variants overlap predicted enhancer states in islets[154]. Considering that GWAS loci may have multiple functional regulatory variants responsible for the association[85], future studies should include testing at

least these 2 variants (if not the entire set of candidate variants) for evidence of regulatory activity. The identities of the nuclear protein(s) binding to rs11708067 are still unknown; it will be necessary to identify these proteins to determine if they alter the regulatory activity of rs11708067. These putative proteins could then be tested for binding in ChIP assays in a more *in vivo* context in primary human islets. Future studies could also include allelic expression imbalance analyses in human islets heterozygous for rs11708067, or genomic editing of rs11708067 in embryonic stem cells or human pancreatic beta cell lines and measuring the effects on *ADCY5* gene expression. These experiments would ultimately contribute to further characterization of the molecular mechanism(s) responsible for the T2D and glucose-related associations at the *ADCY5* locus. Despite these limitations, we provide multiple lines of evidence for rs11708067 as a functional regulatory variant at the *ADCY5* locus. This is in line with rs11708067 showing the highest posterior probability of driving the association signal from a credible set analysis that included enrichment estimates using chromatin annotations from 12 cell types ([172] and Kyle Gaulton, personal communication). In our studies, we also found consistent evidence that the rs11708067-A allele is associated with decreased *ADCY5* expression, highlighting the relationship between the GWAS variant and gene expression.

Taken together, our data identify a functional variant at the *ADCY5* locus, rs11708067, and provide evidence of a mechanism by which this variant affects expression of *ADCY5*, leading to altered fasting glucose levels and risk of T2D (Figure 3.6). The T2D-risk allele rs11708067-A appears to bind one or more repressor proteins that reduce enhancer activity and reduce *ADCY5* expression. The lower transcriptional activity of *ADCY5* may reduce the generation of cAMP and lead to decreases in Ca^{2+} influx and glucose-stimulated insulin secretion, which could result in increased plasma glucose levels and T2D. Additional experiments will be necessary to fully elucidate the molecular and biological mechanism(s) at this locus.

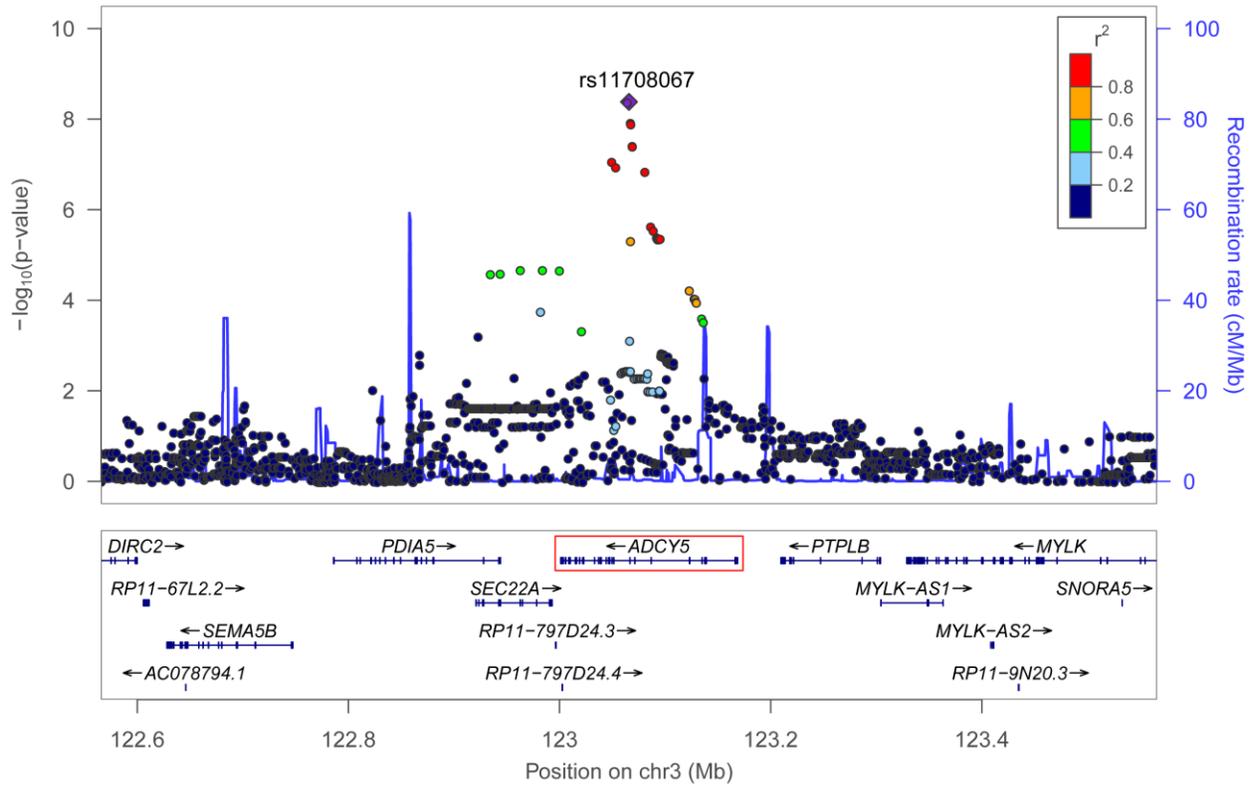


Figure 3.1 The A allele of rs11708067 is associated with decreased *ADCY5* expression in 112 human primary pancreatic islet samples

eQTL data are based on RNA-sequencing of 81 published[155] and 31 new islet samples. The diamond represents the lead variant, rs11708067, of the eQTL analysis. Circles represent genotyped and imputed DNA variants and LD (r^2) is colored based on 1000 Genomes Phase 3 EUR. Chromosome coordinates correspond to the UCSC Genome Browser build hg19. The left Y-axis indicates the $-\log_{10}(p\text{-value})$, the right Y-axis indicates the recombination rate (cM/Mb) and the X-axis indicates the position on chromosome 3 (Mb).

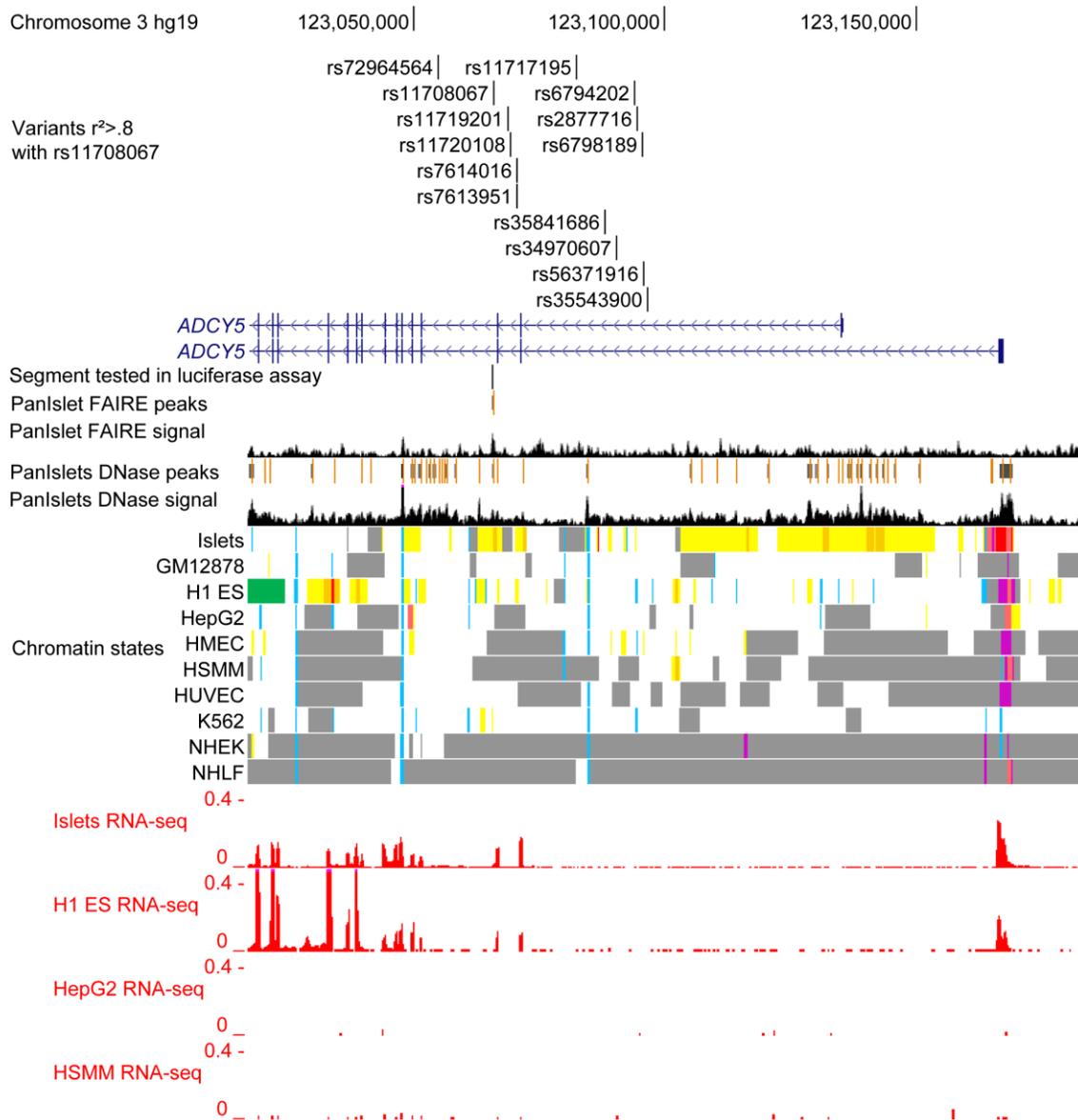


Figure 3.2 rs11708067 overlaps pancreatic islet open chromatin and a strong enhancer chromatin state in *ADCY5* intron 3

Figure 3.2 rs11708067 overlaps pancreatic islet open chromatin and a strong enhancer chromatin state in *ADCY5* intron 3

The 42-kb region of association shown includes all 13 variants in strong linkage disequilibrium ($r^2 > 0.8$) with the T2D-associated variant rs11708067 (14 total noncoding candidate variants located within *ADCY5* introns 1-3). Selected ENCODE open chromatin (DNase, FAIRE), and RNA-seq and chromatin state tracks are shown. The black vertical line above the pancreatic islet open chromatin tracks represents the 231-bp segment containing rs11708067 that was tested in luciferase reporter assays. rs11708067 is located 102 kb from the *ADCY5* transcription start site (TSS) (longer isoform) and 69 kb from the shorter isoform *ADCY5* TSS. PanIslets, pancreatic islets; GM12878, B-lymphocyte, lymphoblastoid cells; H1 ES, embryonic stem cells; HepG2, human hepatocellular carcinoma cells; HMEC, human mammary epithelial cells; HSMM, human skeletal muscle myoblasts; HUVEC, human umbilical vein endothelial cells; K562, leukemia cell line; NHEK, epidermal keratinocytes; NHLF, lung fibroblasts.

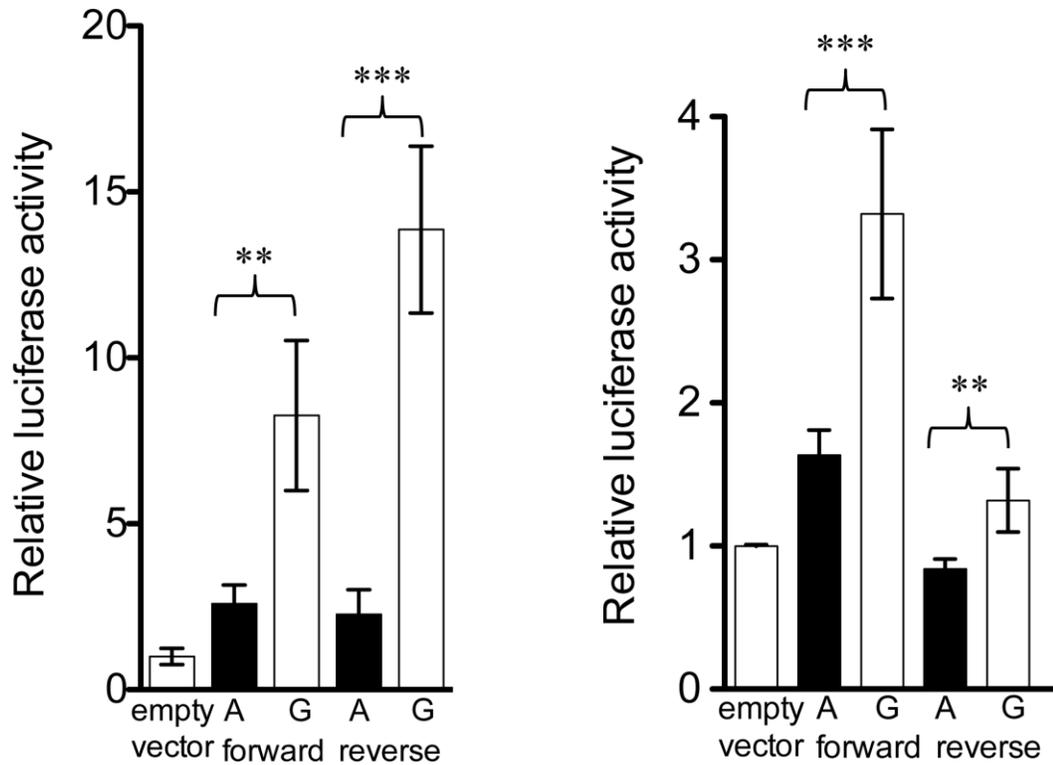


Figure 3.4 rs11708067 exhibits allelic differences in transcriptional activity in MIN6 mouse insulinoma and 832/13 rat insulinoma cells

231-bp segments containing allele A or G of rs11708067 were cloned into a pGL4.23 luciferase reporter vector upstream of the minimal promoter in both forward and reverse orientations.

A Relative luciferase activity of plasmids transfected into MIN6 cells. Luciferase activity is shown normalized to an empty vector control. Error bars represent pairwise standard deviation of 3-5 independent clones per allele (t-tests). * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p \leq 0.001$.

B Relative luciferase activity of plasmids transfected into 832/13 cells.

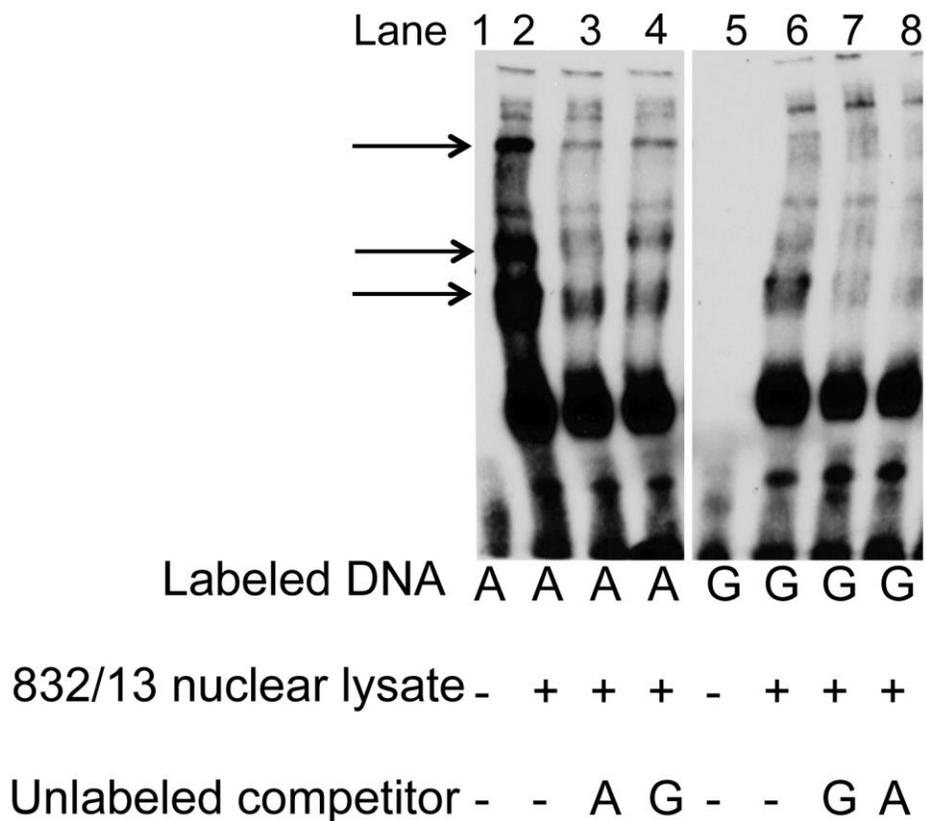


Figure 3.5 The alleles of rs11708067 show allelic differences in nuclear protein binding

Electrophoretic mobility shift assays with biotin-labeled probes containing either the A or G allele of rs11708067. Probes were incubated with 832/13 rat insulinoma nuclear lysate. The arrows indicate differential protein binding to the A allele, which is competed away by 125-fold excess competitor DNA containing the A allele (lane 3). EMSAs were repeated 8 independent times with consistent results.

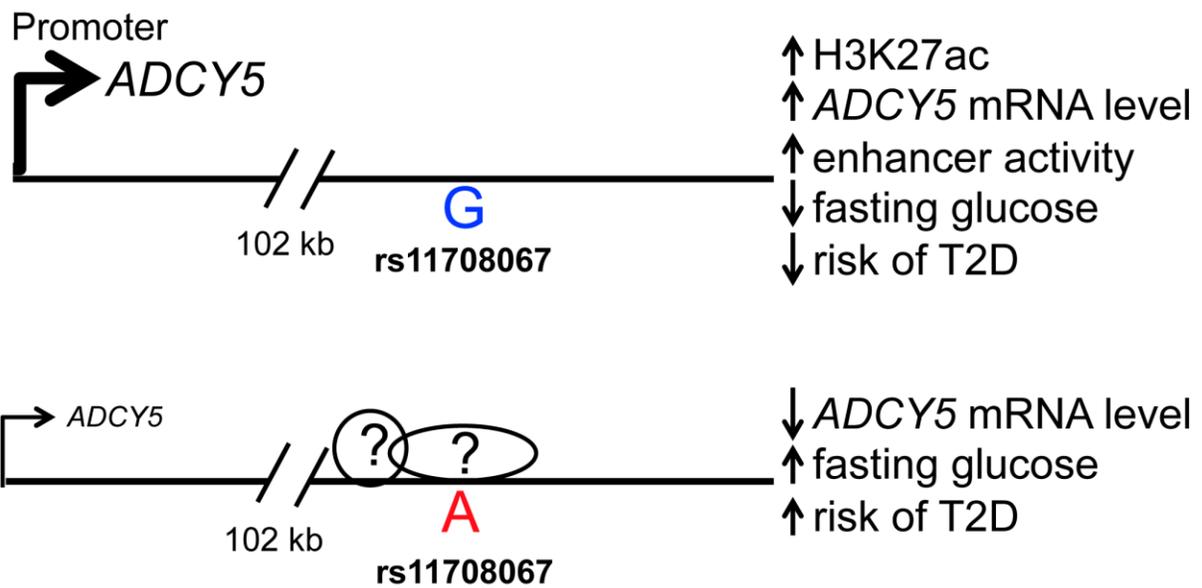


Figure 3.6 rs11708067 may alter transcription factor binding at *ADCY5*

Arrows indicate the transcription start site (TSS) of the *ADCY5* gene. Circles/ovals represent proteins that appear to be bound specifically and differentially to the A allele of rs11708067. rs11708067 showed allelic differences in transcriptional enhancer activity, differential protein binding in EMSAs, and evidence of allelic imbalance in H3K27ac ChIP-seq reads. The rs11708067-G allele, which showed greater enhancer reporter activity and greater H3K27ac enrichment, is the allele that showed higher *ADCY5* expression in human pancreatic islets. The G allele is also associated with decreased risk of T2D.

Lead SNP	T2D risk, glucose raising allele	Alt. allele	Effect size or Odds Ratio	N	P value	Trait	Study	Pop
rs11717195	T	C	1.09-1.18	26,488 T2D cases and 83,964 controls	2.2E-08	T2D	GWAS, DIAGRAM 2014	Trans-ethnic meta-analysis
rs11717195	T	C	1.11	34,840 T2D cases, 114,981 controls	6.5E-14	T2D	GWAS, Morris 2012	European
rs2877716	C	T	+0.09 mmol/l	44,225	4.2E-16	2-hr glucose	GWAS, Saxena 2010	European
rs2877716	C	T	1.12	35,869 T2D cases, 89,798 controls	4.8E-18	T2D	GWAS, Saxena 2010	European
rs11708067	A	G	+0.03 mmol/l	118,475	7.1E-22	fasting glucose	GWAS, Dupuis 2010	European
rs11708067	A	G	-0.02	94,212	2.5E-12	HOMA-B	GWAS, Dupuis 2010	European
rs11708067	A	G	1.18	2,806 T2D cases, 4,265 controls	4.7E-03	T2D	Candidate gene, Ng 2013	African American
rs11708067	A	G	1.23	1678 T2D cases, 1584 controls	9.1E-04	T2D	Candidate variant, Rees 2011	South Asian
rs143882978	C	T	-0.462	1035 cases, 740 controls	0.02	T2D	Candidate gene, Adeyemo 2015	African

rs9883204	C	T	0.041, 0.127, -0.106, -0.058	2,151	0.03, 0.02, 0.05, 0.01	Fasting glucose ; 2-hr glucose ; reduced I-I; 2-hr insulin	Candidate variant, Vasan 2011	Asian Indians
-----------	---	---	---------------------------------------	-------	---------------------------------	---	-------------------------------------	------------------

Table 3.1 Summary of published *ADCY5* SNP associations with T2D and glucose-related traits

Alt. allele, alternate allele (T2D non-risk allele, or glucose decreasing allele); Pop, population; I-I, insulinogenic index. LD (r^2 , 1000 Genomes Phase 1 EUR) between rs11717195 and rs11708067: 0.91; LD between rs2877716 and rs11708067: 0.8; LD between rs9883204 and rs11708067: 0.79. LD (r^2 , 1000 Genomes YRI) between rs143882978 and rs11708067: 0.96.

Primer sequences used to amplify DNA for luciferase reporter assays	
rs11708067_Forward Chromosome 3 position (hg19)	CCTGGGGAGAAGGAACTCTC 123065631
rs11708067_Reverse Chromosome 3 position (hg19)	GTCCTTTCACTGCGTGTTT 123065861
Oligonucleotide sequences for EMSAs	
rs11708067_A_Forward	CAGATTTTGCACTCTATTAAT
rs11708067_A_Reverse	ATTAATAGAGTGCAAAATCTG
rs11708067_G_Forward	CAGATTTTGCGCTCTATTAAT
rs11708067_G_Reverse	ATTAATAGAGCGCAAAATCTG

Table 3.2 Primer sequences for luciferase reporter and electrophoretic mobility shift assays

Forward indicates forward primer, reverse indicates reverse primer (5'-3').

CHAPTER 4: REGULATORY ELEMENTS IN ENDOTHELIAL CELLS AT THE *PLXND1* GWAS LOCUS ASSOCIATED WITH WHR

4.1 Introduction

Increased waist-hip ratio (WHR) is associated with risk of metabolic diseases, including type 2 diabetes[173] and coronary heart disease[174]. Increased WHR is often characterized by increased adiposity in the central part of the body[175]. Insulin resistance, increased triglycerides, decreased HDL cholesterol level and central obesity tend to cluster together in individuals and contribute to metabolic disease risk[46, 47]. To date, GWAS have identified 49 loci that are associated with human WHR, a measure of body fat distribution[64, 65]. Of these loci, at least 5 contain genes that have roles in angiogenesis and/or vascular biology. Angiogenesis is required for adipogenesis and adipose tissue expansion[176, 177]; therefore, studying these genes may provide further insight into the biology of body fat distribution in humans.

The WHR association signals include one represented by lead GWAS variant rs10804591 on chromosome 3 ($p = 6.6 \times 10^{-9}$, $n = 209,921$ total individuals)[65]. Like many WHR association loci, this signal showed a stronger association in women ($p = 6.1 \times 10^{-13}$, $n = 116,667$ women)[65]. The alleles associated with increased WHR are also associated with decreased hip circumference ($p = 7.4 \times 10^{-18}$, $n = 210,953$ total individuals; $p = 5.5 \times 10^{-15}$, $n = 117,265$ women)[65].

The WHR-associated lead GWAS variant rs10804591 is located 8.6 kb upstream of the promoter of *PLXND1*. *PLXND1* (MIM: 604282) encodes Plexin D1 protein, which acts as a receptor for the semaphorins 3E and 4A[178-183]. Semaphorins are secreted and membrane-bound proteins that play a role in the function and formation of the cardiovascular, endocrine, gastrointestinal, hepatic, immune, musculoskeletal, renal, reproductive, and respiratory systems[184]. Semaphorin-Plexin D1 signaling was also found to be important for synaptic signaling and development of the nervous system[185] as well as

regulation of angiogenesis[186]. In zebrafish, loss of *plxnd1* function by morpholino injection or in mutants resulted in blood vessel patterning defects[179].

In addition to roles in angiogenesis, Plexin D1 has also been shown to influence adipocyte morphology and fat storage. Homozygous null *plxnd1* mutant zebrafish exhibit decreased lipid storage and a hyperplastic morphology specifically in visceral adipose tissue (VAT)[187]. After being fed a high-fat diet, VAT in *plxnd1* mutants did not expand, nor did these zebrafish display hyperglycemia, leading to the conclusion that loss of *plxnd1* offers protection from high-fat diet-induced insulin resistance[187]. In humans, *PLXND1* mRNA levels in VAT were positively associated with a hypertrophic morphology ($p = 0.03$, $n = 79$ individuals) and *PLXND1* mRNA was increased in VAT of obese patients with type 2 diabetes ($p = 0.03$, $n = 30$ individuals)[187]. Plexin D1's established role in vascular development and patterning and adipose tissue morphology suggests *PLXND1* may have biological relevance and may be a novel gene influencing WHR.

According to 1000 Genomes phase 1 European (EUR) data[144], 35 variants exhibit strong linkage disequilibrium (LD, $r^2 > 0.8$) with rs10804591. All 36 variants are noncoding and overlap at least one mark indicative of regulatory regions, including open chromatin, histone modification ChIP-seq and/or transcription factor binding ChIP-seq peaks from the ENCODE Project[72] or Human Epigenome Atlas[73]. We therefore hypothesized that one or more of these variants may be located in enhancer regions and act to regulate gene expression. Based on evidence of *plxnd1*'s role in visceral adipose endothelial cells in zebrafish[187], we sought to identify the functional variants and molecular mechanisms underlying the *PLXND1* WHR association by identifying regulatory enhancer elements in both a human umbilical vein endothelial cell line (HUVEC) and zebrafish.

4.2 Materials and Methods

4.2.1 Luciferase reporter vectors containing *PLXND1* candidate cis-regulatory elements

Candidate cis-regulatory elements (CREs) to test in reporter assays (Figure 4.1 and Table 4.1) were designed based on overlap with variants in strong LD ($r^2 > 0.8$, 1000 Genomes phase 1 EUR) with the lead WHR-associated GWAS variant rs10804591[144] as well as open chromatin DNase and FAIRE peaks and histone modification H3K4me1, H3K9ac and H3K27ac ChIP-seq peaks from ENCODE

data[72]. Elements were amplified using Phusion® High-Fidelity DNA polymerase (New England BioLabs). Primer sequences for amplification of each CRE are in Table 4.2. Site-specific recombination sequences (attB sites) were added to the forward and reverse primers to facilitate recombination of the elements into vectors using the Gateway® recombination cloning technology (Thermo Fisher Scientific). PCR amplicon bands were isolated from agarose gels and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). The CRE elements were then cloned into a Gateway pDONR™221 entry clone using BP Clonase™ (Thermo Fisher Scientific) and transformed into Library Efficiency® DH5α™ Chemically Competent *E. coli* strains (Thermo Fisher Scientific) according to manufacturer's instructions. Plasmids were purified using QIAprep Spin Miniprep Kit (QIAGEN) and the elements and recombination junctions were verified by sequencing. Candidate CRE elements in pDONR™221 vectors were then recombined into the destination gateway-compatible pGL4.23 luciferase vectors by LR recombination using LR Clonase™ according to manufacturer's instructions (Thermo Fisher Scientific) and incubated overnight. The recombination reactions were transformed into Library Efficiency® DH5α™ Chemically Competent *E. coli* strains, purified using PureLink® HiPure Plasmid Miniprep Kit (Thermo Fisher Scientific) and vectors were verified by sequencing. To create the alternate alleles of candidate variants, site-directed mutagenesis was performed using the Q5® Site-Directed Mutagenesis Kit (New England BioLabs) and alleles were confirmed by sequencing.

4.2.2 Generation of a gateway-compatible pGL4.23 luciferase reporter vector

For recombination of CRE elements into pGL4.23 luciferase vector (Promega), a gateway-compatible pGL4.23 luciferase vector was created. Briefly, the Gateway® Vector Conversion System (Thermo Fisher Scientific) was used to recombine a reading frame cassette into the pGL4.23 luciferase vector. The pGL4.23 luciferase vector was first linearized using *EcoRV* (New England BioLabs), 5' phosphates were removed with calf intestinal alkaline phosphatase and reading frame cassette B (Thermo Fisher Scientific) was ligated into the pGL4.23 luciferase vector according to manufacturer's instructions. The ligation reaction was then transformed into competent DB3.1™ *E. coli* (Thermo Fisher Scientific) and resulting vectors were purified using the QIAprep Spin Miniprep Kit (QIAGEN) and sequence verified.

4.2.3 Cell culture

On day 1, HUVEC cells (40,000 per well) were plated into 24-well tissue culture plates. On day 2, 250 ng of luciferase vector and 27.8 ng of renilla vector per well were co-transfected into HUVEC. Luciferase and renilla vectors were added to 250 μ l of Opti-MEM® Reduced Serum Medium (Life Technologies), and 0.5 μ l of *TransIT*®-2020 transfection reagent (Mirus Bio) per well was added to the vectors and Opti-MEM. The transfection mixture was incubated for 30 minutes at room temperature and then transfected into HUVEC, and cells were incubated for 2 days at 37°C and 5% CO₂. Luciferase assays were then performed using a GloMax® luminometer (Promega) after 48 hours.

4.2.4 Cloning of *PLXND1* candidate *cis*-regulatory elements into EGFP expression vector

CRE elements that were cloned into a Gateway pDONR™221 entry clone (Thermo Fisher Scientific) as described for the luciferase vectors above were recombined into a pGW_*cfos*EGFP expression vector[188] containing Tol2 sites necessary for transposition into the zebrafish genome. CRE elements were recombined by LR recombination using LR Clonase™ according to manufacturer's instructions (Thermo Fisher Scientific) and incubated overnight. The recombination reactions were then transformed into Library Efficiency® DH5 α ™ Chemically Competent *E. coli* strains, purified using PureLink® HiPure Plasmid Miniprep Kit (Thermo Fisher Scientific) and vectors were verified by sequencing.

4.2.5 Injection and imaging of human *PLXND1* CREs in zebrafish

Zebrafish experiments were approved by The University of North Carolina at Chapel Hill and Duke University Institutional Animal Care and Use Committees. All injections were performed in *Tg(kdrl:mCherry)^{is5Tg}* zebrafish embryos (hereafter referred to as *kdrl:mCherry*)[189, 190]. On the day of injection, a mix was prepared containing 1 μ l pTol2DEST:PLXND1-CRE (125 ng/ μ l), 1 μ l Tol2 transposase mRNA (175 ng/ μ l), 0.5 μ l Phenol Red (2% in H₂O) and 2.5 μ l RNase-free water. This mix was kept on ice until injection. A 4 nl volume of mix was injected into the cell of 1-cell stage embryos using a Picospritzer III Microinjection system (Parker Hannifin Corporation). Injected embryos were kept in Embryo Medium in petri dishes at 28.5°C. Embryo Medium consisted of 400 ml of 20X salt (17.5 g

NaCl, 0.75 g KCl, 2.9 g CaCl₂, 2.39 g MgSO₄, 0.41g KH₂PO₄, 0.13 g Na₂HPO₄, 16 ml of bicarbonate stock (1.5 g NaHCO₃ in 50 ml of dH₂O) and dH₂O. After 8 hours post fertilization (hpf), embryo medium was changed and dead embryos were removed. At 24 hpf, embryos were screened for eGFP fluorescence on a Leica M205 FA fluorescence stereomicroscope (Leica Microsystems). Embryos that exhibited expression of eGFP were then raised to adulthood. These adults were then crossed to *kdr1:mCherry* fish and their offspring were screened for eGFP fluorescence. Three independent transgenic founders were identified with germline transmission of *Tg(HSa.CRE17.PLXND1-Mmu.FOS:GFP)* with the following allele designation of *rdu1*, *rdu2* and *rdu3*. The resulting F1 offspring were raised to establish stable transgenic lines. The expression dynamics of eGFP was similar between all alleles. F1 *Tg(HSa.CRE17.PLXND1-Mmu.FOS:GFP)^{rdu3}* adults were again outcrossed to *kdr1:mCherry* fish and the F2 offspring were analyzed by fluorescence stereomicroscopy (as above) at 22hpf and 120hpf.

4.2.6 Electrophoretic mobility shift assays

Complementary DNA oligonucleotides (19-bp) centered on variant alleles for rs4488824 were synthesized by Integrated DNA Technologies and sequences are noted in Table 4.2. Labeled oligonucleotides included biotin on the 5' end. We performed assays as previously described [83, 85] by using 4 µg of HUVEC nuclear lysate and 50-fold excess unlabeled probe. Reactions were loaded into a 6% DNA retardation gel (Thermo Fisher Scientific), subjected to electrophoresis, transferred to Biodyne B nylon membranes (Thermo Fisher Scientific), and UV crosslinked. Wash and detection steps were performed according to instructions in the Chemiluminescent Nucleic Acid Detection Module (Thermo Fisher Scientific). EMSAs were repeated on a separate day with consistent results. TRANSFAC and JASPAR[111] databases were used to search for transcription factor position weight matrices that were similar to the human genome sequence containing rs4488824. For supershift assays, 6 µg of MAFB (sc-22830X) or 6 µg MYOG (Myogenin, sc-31943X) (both from Santa Cruz Biotechnology, Inc.) antibodies were incubated with HUVEC nuclear lysate, binding buffer, and poly(dI-dC) for 20 minutes before addition of labeled DNA probes.

4.3 Results

4.3.1 Four candidate CREs show enhancer activity in HUVEC

Considering the relatively wide 31-kb WHR association signal (defined by all variants $r^2 > 0.8$ with the GWAS associated variant rs10804591) and our observation that all of the WHR-associated variants at *PLXND1* overlap open chromatin DNase and FAIRE peaks and/or histone modification H3K4me1, H3K4me2, H3K9ac, and H3K27ac ChIP-seq peaks in HUVEC (Table 4.1, Figure 4.1), we aimed to identify regulatory enhancer elements that influence gene expression. We designed candidate *cis*-regulatory elements (CREs) to test in complementary luciferase reporter assays in HUVEC and transgenesis reporter assays in zebrafish. Fifteen candidate CREs were originally designed to span DNase, FAIRE, H3K4me1, H3K4me2, H3K9ac, and H3K27ac peaks in HUVEC from ENCODE data[72] and variants in strong LD ($r^2 > 0.8$, 1000 Genomes phase 1 EUR[144]) with the WHR-associated GWAS lead variant rs10804591. We were not successful in amplifying 5 CREs that were larger than 3 kb, therefore, we revised the candidate CRE list and designed smaller candidate CREs. A total of 10 candidate CREs span 20 candidate variants (Table 4.1). We successfully PCR-amplified and cloned 7 candidate CREs to date (CRE 2, 6, 8, 9, 13, 14, 17). Of these, 4 CREs (CRE 8, 9, 13, 17) exhibited evidence of transcriptional enhancer activity (1.4-fold to 3.4-fold compared to empty vector control, Figure 4.2) for the one allele or haplotype that was cloned. Of these, CRE 13 exhibited the strongest enhancer activity, suggesting that it could be the strongest enhancer among elements tested.

4.3.2 The alleles of rs11718169 in CRE 17 show similar enhancer activities in HUVEC

We next investigated whether there were haplotype or allelic differences for candidate CREs that demonstrated transcriptional enhancer activity in HUVEC. Specifically, we performed preliminary experiments for two of the 4 CREs that contain only a single variant in strong LD ($r^2 > 0.8$, 1000 Genomes phase 1 EUR[144]) with the lead GWAS WHR-associated variant rs10804591 (CRE 13 and 17). CRE 13 contains rs4488824 ($r^2 = 0.99$ with rs10804591) and CRE 17 contains rs11718169 ($r^2 = 0.91$ with rs10804591), therefore, it was relatively straightforward to change the alleles of these variants using site-directed mutagenesis and test for allelic differences in transcriptional enhancer activity. In preliminary experiments, the alleles of rs11718169 in CRE 17 showed 7.6- to 10.3-fold enhancer activity compared to

empty vector control, however, there were no significant differences observed between the alleles (Figure 4.3). Plasmids containing both alleles of rs4488824 in CRE 13 have been generated and verified by sequencing, however, luciferase reporter assays in HUVEC still need to be performed.

4.3.3 CRE 17 shows enhancer activity in endothelial cells in zebrafish

To determine whether the CREs could drive tissue-specific expression, we tested whether candidate CRE 17 (the CRE tested for allelic differences in transcriptional enhancer activity in HUVEC) could drive expression of the reporter gene GFP in zebrafish. We cloned the human CRE 17 element into the pGW_*cfos*EGFP expression vector[188] and injected the CRE 17-containing vector into 1-cell stage zebrafish embryos. At 22 hpf, the CRE 17 element (*hsPLXND1-CRE17:eGFP*) drove GFP expression throughout the zebrafish embryo specifically in endothelial cells. GFP expression was observed in the cranial vasculature (lateral dorsal aorta), endothelial cells migrating over the yolk (common cardinal vein) and truncal vasculature (segmental vessels, dorsal aorta and posterior cardinal vein). These results are shown in Figure 4.4 (left panels). To better compare the expression pattern to sites of endothelial cells, we crossed *hsPLXND1-CRE17:eGFP* to an established endothelial cell-specific transgenic line, *kdr1:mCherry*[189, 190]. At 120 hpf, the progeny showed GFP expression co-localized with mCherry expression in endothelial cells in the dorsal aorta and posterior cardinal vein (Figure 4.4, right panels). The green autofluorescence in pigment cells and intestine (Figure 4.4, lower right panel) does not represent specific CRE 17 enhancer activity. The transgenesis assays in zebrafish suggest that CRE 17 is an endothelial cell-specific regulatory enhancer element in zebrafish.

4.3.4 rs4488824 in CRE 13 exhibits differential protein binding

Based on CRE 13's strong enhancer activity in HUVEC transcriptional reporter assays, we evaluated whether the WHR-associated variant in CRE 13, rs4488824, exhibits differential protein binding of nuclear proteins such as transcription factors. In EMSAs, rs4488824 exhibited a reproducible band for the rs4488824-A allele that was not observed for the rs4488824-T allele. This band likely represents better binding of a transcription factor to the A allele (Figure 4.5, lane 2 vs. lane 6). The differential band was confirmed to be specific through better competition by excess unlabeled DNA containing the

rs4488824-A allele than the rs4488824-T allele. The rs4488824 variant is located within predicted MYOG and MAFB protein sites. We tested whether antibodies against these proteins could demonstrate specific recognition of the bound protein in an EMSA, however, we did not observe evidence of a supershift or disruption of the A allele-specific band. Taken together, our EMSA data suggest that rs4488824 within CRE 13 binds an unknown transcriptional regulatory protein.

4.4 Discussion

We identified four candidate *cis*-regulatory elements that affect regulatory activity at the *PLXND1* locus. Candidate CREs 8, 9, 13 and 17 showed enhancer activity in HUVEC, and the human *PLXND1* CRE 17 element was expressed in endothelial cells in stable transgenic *kdr1:mCherry* zebrafish lines. Together, these data suggest that the transcriptional regulation of CRE17 is conserved between zebrafish and humans and may be a functional regulatory element that affects enhancer activity in an endothelial cell-specific manner. CRE 13 showed the strongest enhancer activity of the 7 candidate CREs tested in HUVEC transcriptional reporter assays, and contains a variant, rs4488824, which showed allelic differences in binding of a nuclear protein. These data suggest that rs4488824 may be a functional regulatory variant located within an enhancer element in human endothelial cells. While neither the full set of candidate WHR variants nor all CREs have been examined, these data support the previous findings that *plxnd1*'s effect on fat deposition and VAT morphology was observed in vascular endothelial cells[187].

This study highlights the value of using complementary human cell-based and zebrafish transgenesis assays to identify regulatory elements. Reporter assays in zebrafish and the generation of stable zebrafish lines can provide evidence that a regulatory element acts during a specific developmental time point or in a specific cell or tissue. Luciferase reporter assays have been useful in identifying regulatory elements[82-85, 191, 192], however, the identification of these elements may be hindered by the lack of prior knowledge of the correct cell type in which to conduct the assays. There are also limitations to using animal models, as absence of regulatory activity of human candidate CREs could be attributed to lack of conservation between humans and zebrafish.

The EMSA data for rs4488824 in CRE 13 suggest that rs4488824 could be a functional regulatory variant responsible for the WHR association that is located within a regulatory enhancer

element at the *PLXND1* locus. Additional experiments will be necessary to determine the identity and validate the protein bound differentially to rs4488824. Additional experiments are also needed to determine whether rs4488824 within CRE 13 shows allelic differences in transcriptional regulatory activity, as the CRE 13 element containing the rs4488824-T allele was tested in the HUVEC transcriptional reporter assays, but not the CRE 13 element containing the rs4488824-A allele. Furthermore, it will also be necessary to test CRE 13 in zebrafish transgenesis assays to investigate whether this CRE drives temporal or tissue-specific GFP expression.

CRE 17 showed enhancer activity in HUVEC as well as in endothelial cells in zebrafish. Preliminary transcriptional reporter assay experiments in HUVEC did not show allelic differences in enhancer activity for rs11718169 (the WHR-associated variant located within CRE 17). These results are not unexpected, considering that we observed no noticeable differences in GFP expression between the two alleles in the stable transgenic zebrafish lines containing human CRE 17. These data suggest that rs11718169 may not drive allelic differences in enhancer activity or be a functional regulatory variant responsible for the WHR association. Additional experiments, such as EMSAs, could be performed to test whether there is evidence of differential nuclear protein binding at rs11718169.

More work is necessary to identify all the potential regulatory elements in this region, including testing the remaining 3 candidate CREs in both transcriptional reporter assays in HUVEC and transgenesis assays in zebrafish. The transcriptional reporter assays in HUVEC took months of optimizing due to low transfection efficiency and the fact that HUVEC are sensitive to endotoxins, therefore, it will be important to repeat these assays with the optimized conditions to confirm that the transcriptional enhancer activities observed for CREs 8, 9, 13 and 17 are reproducible. Low transfection efficiency, toxicity and cell death were observed when transfecting greater than 250 ng of plasmids into HUVEC; these reasons are likely attributed to the fact we only observed 1.4-fold enhancer activity for CRE 17 in our preliminary experiments (Figure 4.2) but 7.6- to 10-fold enhancer activity for CRE 17 in an independent experiment after the conditions were optimized (Figure 4.3). It will also be important to test both alleles/haplotypes of the WHR-associated variants contained within the candidate CREs to identify the potential functional regulatory variants and gain a greater understanding of the direction of effect on WHR in humans.

Functional investigation of GWAS signals can be challenging, especially if the association signal is wide (i.e. there are many candidate variants in strong LD with the GWAS lead variant). Therefore, it can be beneficial to identify larger regulatory elements to determine the functional regulatory variant(s) at a GWAS locus. Open chromatin and histone modification ChIP-seq peaks have been useful for identifying regulatory variants that are responsible for the association signal[85]; however, if all of the variants overlap regulatory marks, these marks may not be useful for distinguishing regulatory variants. This study shows the benefits of using regulatory datasets combined with experiments in human cell-based assays and animal models to identify the functional regulatory elements and variants, as well as the cell types in which they act. Another challenge to functional follow-up studies of GWAS loci is determining the target gene(s). *PLXND1* is a likely target gene influencing WHR, considering *plxnd1*'s role in adipose morphology and lipid storage and the observation that *PLXND1* mRNA levels were associated with a hypertrophic VAT morphology in humans[187]. However, we cannot rule out the possibility of other target genes, considering the published eQTL highlighting an association between the WHR-associated variant rs10804591 and *TMCC1* expression in 5,311 peripheral blood samples ($p= 3.5 \times 10^{-27}$)[164]. The *TMCC1* gene is located 41 kb from *PLXND1* and encodes transmembrane and coiled-coil domain family 1 protein, which has a suggested role in endoplasmic reticulum organization[193]. Overall, the human GWAS study that identified a WHR association signal at the *PLXND1* locus[65] and studies on the biological role of *plxnd1* in adipose morphology and fat storage in zebrafish[187] were undoubtedly valuable for gaining greater understanding of *PLXND1*'s potential role in WHR and body fat distribution.

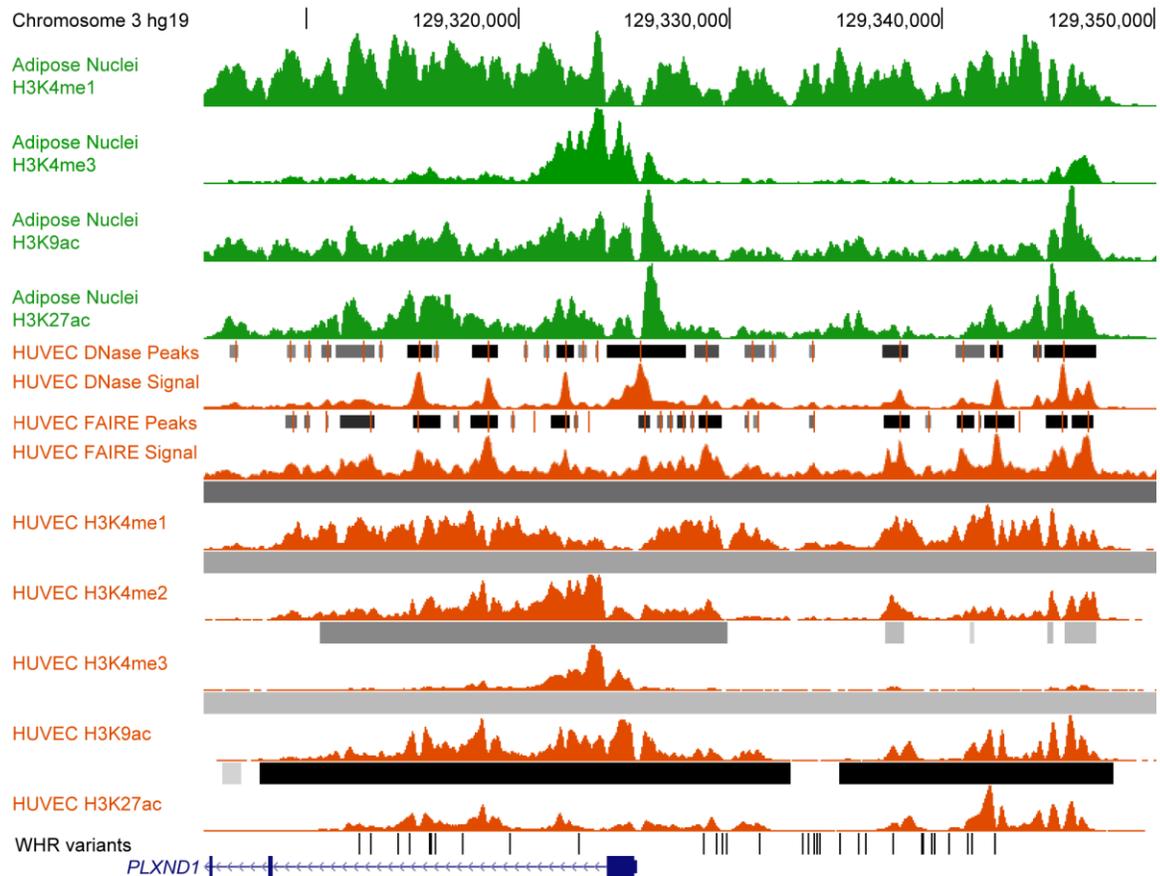


Figure 4.1 Open chromatin and histone modifications marking potential regulatory regions at the *PLXND1* WHR-associated locus

A 31-kb region includes all 35 variants in strong linkage disequilibrium ($r^2 > 0.8$, 1000 Genomes phase 1 EUR) with the WHR-associated lead GWAS SNP rs10804591 (36 total candidate variants). Selected Roadmap Epigenomics Human Epigenome Atlas histone modification ChIP-seq and ENCODE open chromatin, and histone modification ChIP-seq signal and peak tracks are shown for adipose nuclei (green) and human umbilical vein endothelial cells (HUVEC, designated by orange). ENCODE histone modification ChIP-seq peaks in HUVEC show broad patterns of H3K4me1-, H3K4me2-, H3K4me3-, H3K9ac- and H3K27ac-enriched domains[118]. Darker-shaded enrichment domains for these histone marks (indicated by the gray and black-shaded rectangles over the histone modification signals) signify higher browser extensible data (BED) score values. The black vertical lines above the *PLXND1* gene represent all 36 total candidate variants (labeled as WHR variants).

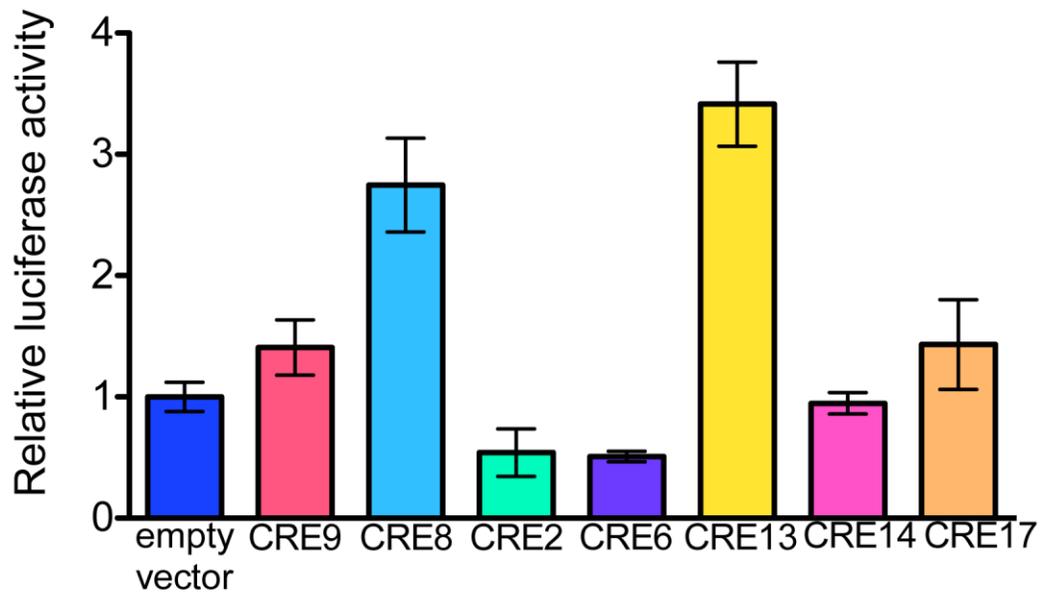
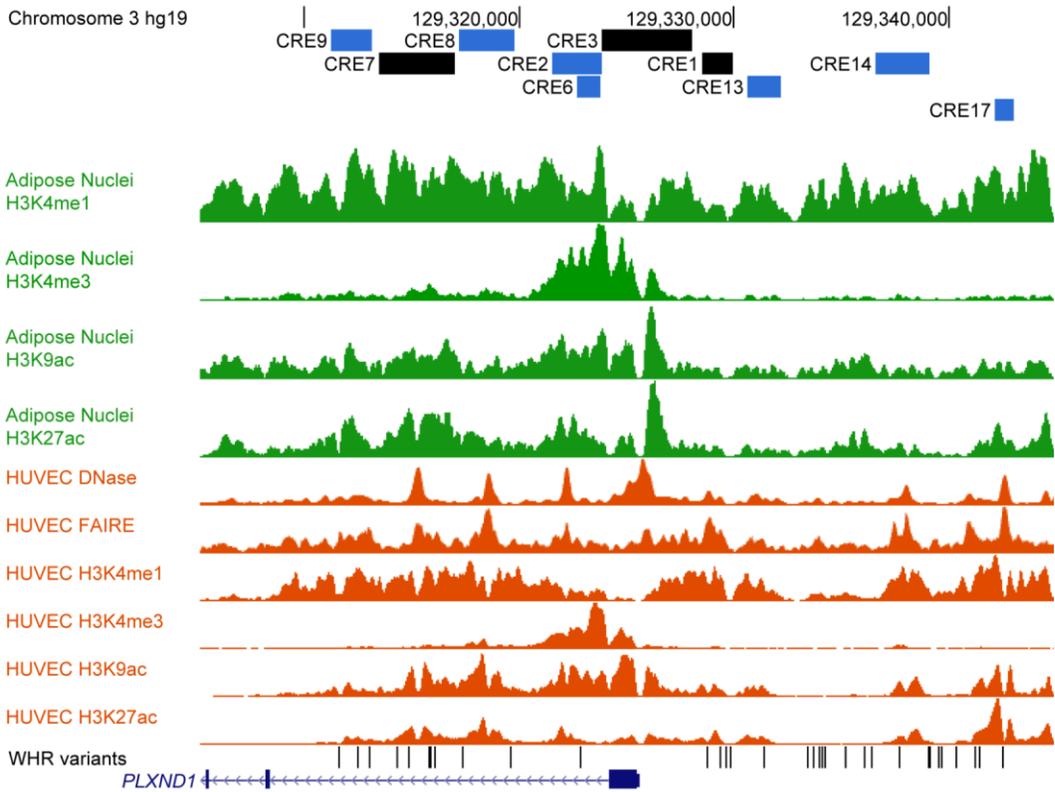


Figure 4.2 Four *PLXND1* candidate *cis*-regulatory elements show enhancer activity in HUVEC transcriptional reporter assays

Figure 4.2 Four *PLXND1* candidate *cis*-regulatory elements show enhancer activity in HUVEC transcriptional reporter assays

The rectangles designated by “CRE” indicate candidate *cis*-regulatory elements that overlap with open chromatin or histone modification ChIP-seq marks and WHR-associated variants in *PLXND1* intron 1 or upstream of the *PLXND1* promoter. The blue rectangles indicate CRE elements that have been cloned into gateway compatible luciferase plasmids and tested in luciferase assays in HUVEC, and black rectangles indicate CREs that have been designed but not cloned. The bottom of the figure indicates the relative luciferase expression of the candidate CREs tested (one haplotype for each CRE was tested). WHR variants indicate variants in strong LD ($r^2 > 0.8$, 1000 Genomes phase 1 EUR) with the GWAS-associated lead variant rs10804591 (36 total candidate variants). CRE, *cis*-regulatory element.

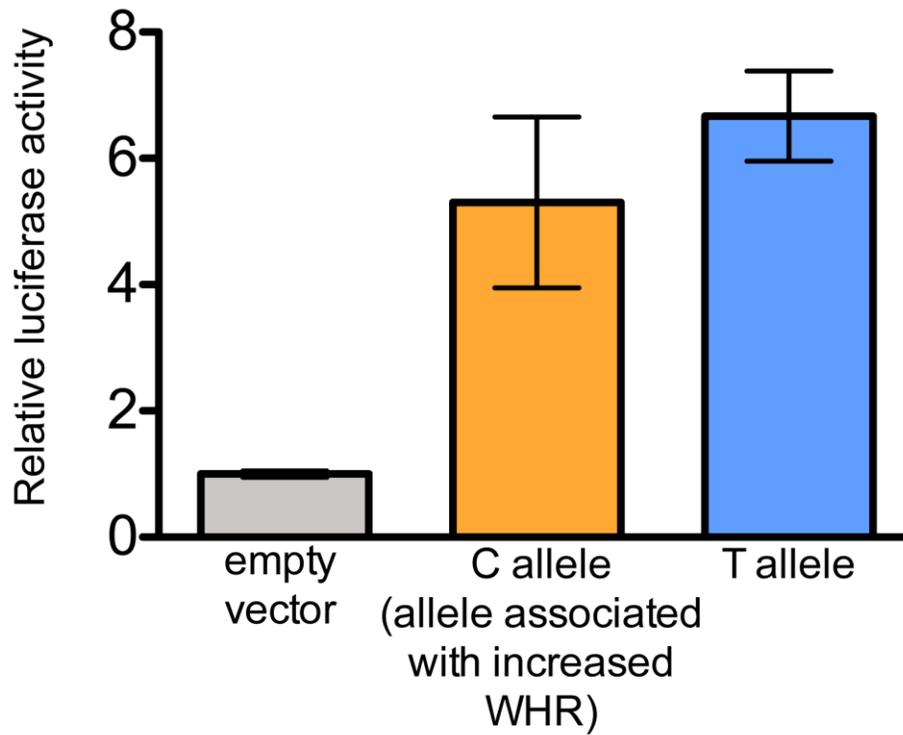


Figure 4.3 Both alleles of rs11708067 in CRE 17 exhibit enhancer activity in HUVEC transcriptional reporter assays

888-bp segments containing each allele of rs11708067 were cloned into a gateway-compatible pGL4.23 luciferase reporter vector upstream of the minimal promoter. The vectors were transfected into HUVEC cells, and luciferase expression normalized to that of an empty vector control is shown. Error bars represent the pairwise standard deviation of four independent clones per allele.

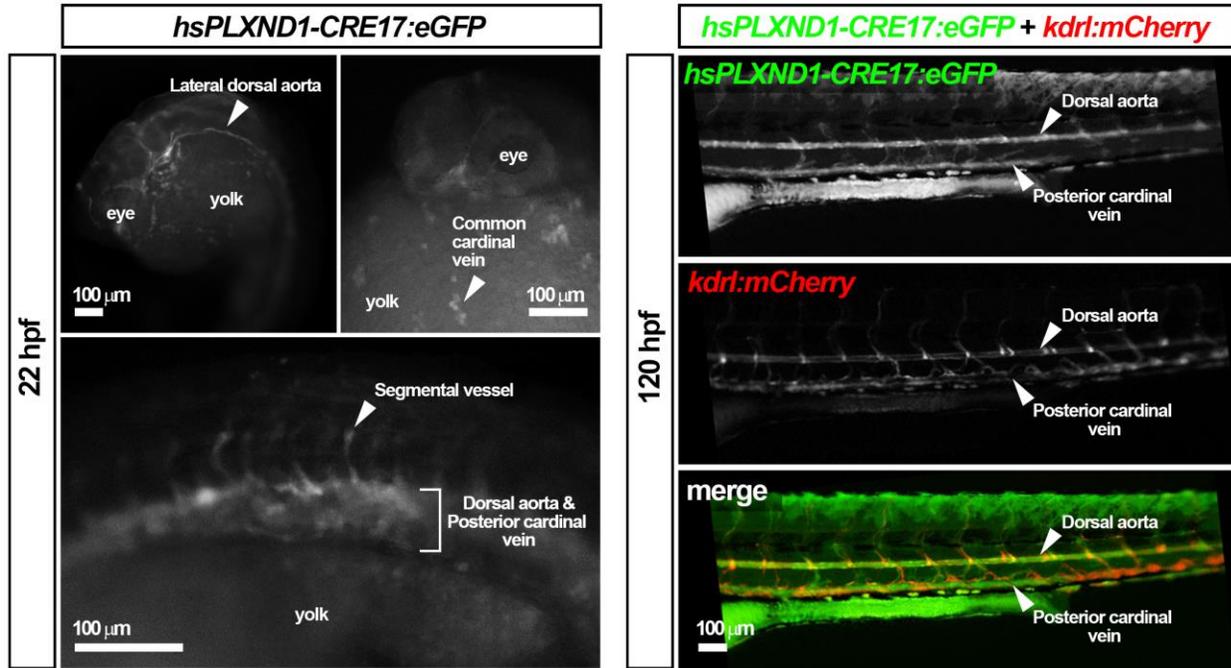


Figure 4.4 CRE 17 shows enhancer activity in endothelial cells in zebrafish 120 hours post-fertilization

1-cell stage zebrafish embryos were injected with 125 ng of plasmid containing *hsPLXND1-CRE17:eGFP* with transposase mRNA. Embryos that exhibited eGFP expression were raised to adulthood and crossed to *kdr1:mCherry* zebrafish. F1 offspring were raised to establish stable transgenic lines and F1 *hsPLXND1-CRE17:eGFP* adults were outcrossed to *kdr1:mCherry* zebrafish. F2 offspring were analyzed by fluorescence stereomicroscopy at 22 hpf (left panels) and 120 hpf (right panels).

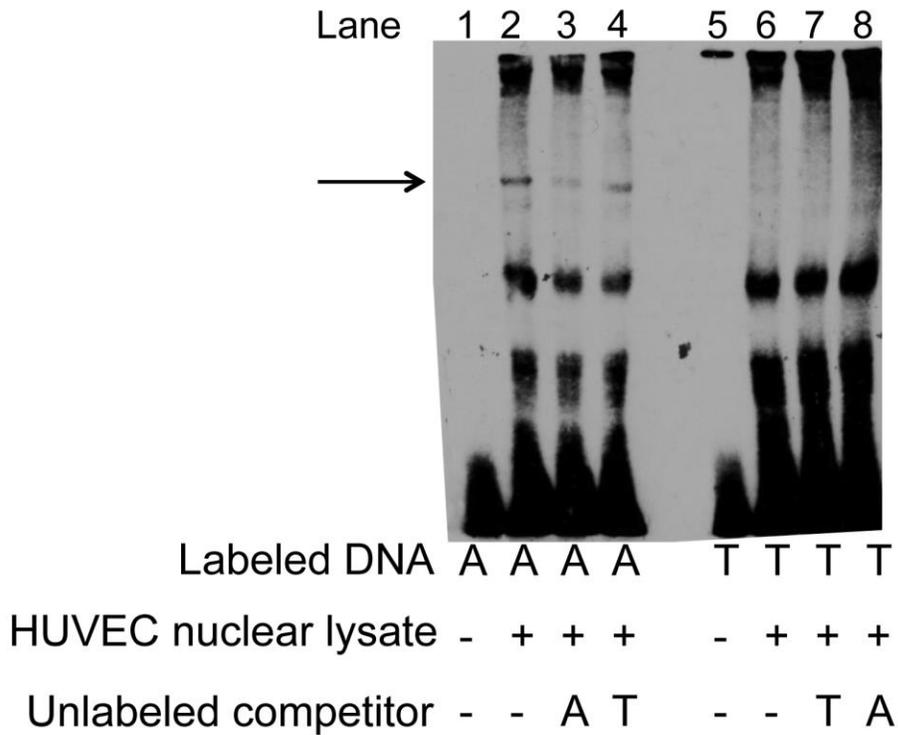


Figure 4.5 rs4488824 (in CRE 13) shows allelic differences in protein binding

EMSA with biotin-labeled probes containing either the A or T allele of rs4488824 and incubated with 4 μ g HUVEC nuclear lysate. The arrow indicates specific protein binding to the A allele (lane 2 versus 6).

For competition reactions, 50-fold excess unlabeled probe containing either allele were added.

Candidate CRE	Location with respect to <i>PLXND1</i>	Size of element in bp	Overlapping WHR-associated variants	Regulatory marks in HUVEC
1	3 kb upstream <i>PLXND1</i>	1479	rs9840468, rs11929498, rs9845755, rs9808900	DNase, FAIRE, H3K4me1, 2, 3; H3K9ac, H3K27ac
2	<i>PLXND1</i> intron 1	2262	rs1872113	DNase, FAIRE, H3K4me1, 2, 3; H3K9ac, H3K27ac
3	<i>PLXND1</i> exon 1, 5'UTR, TSS and 3 kb upstream	4180, contains <i>PLXND1</i> TSS	-	DNase, FAIRE, H3K4me1, 2, 3; H3K9ac, H3K27ac
6	<i>PLXND1</i> intron 1	1072	rs1872113	DNase, H3K4me1, 2, 3; H3K9ac, H3K27ac
7	<i>PLXND1</i> intron 1	3472	rs281448, rs57098603, rs11711654, rs9837325, rs4579013	DNase, FAIRE, H3K4me1, 2, 3; H3K9ac, H3K27ac
8	<i>PLXND1</i> intron 1	2614	rs6792344, rs9682596	DNase, FAIRE, H3K4me1, 2, 3; H3K9ac, H3K27ac
9	<i>PLXND1</i> intron 1	1878	rs6786987, rs6774878, rs2811450	DNase, FAIRE, H3K4me1, 2, 3; H3K9ac, H3K27ac
13	5 kb upstream of <i>PLXND1</i>	1589	rs4488824	DNase, FAIRE, H3K4me1, 2; H3K9ac, H3K27ac
14	11 kb upstream of <i>PLXND1</i>	2506	rs6778532, rs4688808, rs2811437	DNase, FAIRE, H3K4me1, 2, 3; H3K9ac, H3K27ac
17	26 kb upstream of <i>PLXND1</i>	888	rs11718169	DNase, FAIRE, H3K4me1, 2; H3K9ac, H3K27ac

Table 4.1 Summary of candidate *cis*-regulatory elements designed to test in luciferase reporter assays and zebrafish transgenesis assays

CRE, *cis*-regulatory element; TSS, transcriptional start site; 5' UTR, 5' untranslated region; HUVEC, human umbilical vein endothelial cells. Regulatory marks in HUVEC indicate that the CRE overlaps at least 200 base pairs of an ENCODE-defined regulatory peak (DNase, FAIRE, H3K4me1, H3K4me2, H3K4me3, H3K9ac or H3K27ac peaks).

Primer sequences for amplifying <i>PLXND1</i> CREs	
CRE1_Forward CRE1_Reverse	ACACCTTCCCTGGCCTTACT GTGTCCCCATCTGCAAAAGT
CRE2_Forward CRE2_Reverse	CTCCAGGACCAAGCAGAGAC CATGCCTGACCAAAGTTGTG
CRE3_Forward CRE3_Reverse	CGTGCATACCTGGGCTTTAT GAGCCTTGTTGTGCTTGACA
CRE6_Forward CRE6_Reverse	CTGCATTTTGCTGATGGAGA GTTACCATTGCACGTTGGTG
CRE7_Forward CRE7_Reverse	CCTGGAAAAGAAAGCAGACG CCTGCAGGCTAAAGACTTGG
CRE8_Forward CRE8_Reverse	ACCTCCTCCCTCACCATTCT TCAGTGCAGCATTTTCAAGG
CRE9_Forward CRE9_Reverse	CTCCCACCAAGGGAACATAA CAGCTCTGTGGCAAATTTCA
CRE13_Forward CRE13_Reverse	TTGACCGTGTTTCCAGACAG GGCCTGTGTGCATCTGATAA
CRE14_Forward CRE14_Reverse	CCAGAAGGATCTGGGTTCAA CTCCTGTGGGTCTGTCCTGT
CRE17_Forward CRE17_Reverse	GGGGACTGTGAAGACAGGAG CCTGGCCCTAATCTCCTCAT
Probe sequences for EMSAs	
rs4488824_A_Forward (with respect to the genome) rs4488824_T_Reverse	CAGCTGTGCAGCTGACCCC GGGGTCAGCTGCACAGCTG
rs4488824_T_Forward (with respect to the genome) rs4488824_A_Reverse	CAGCTGTGCTGCTGACCCC GGGGTCAGCAGCACAGCTG

Table 4.2 Primer sequences for functional assays

All sequences are shown 5'-3' with respect to the genome.

CHAPTER 5: CONCLUSIONS

GWAS have identified thousands of variants associated with complex traits and diseases[194, 195]. However, many of the molecular and biological mechanisms responsible for these associations are complex and identifying the causal variants and genes can be challenging. Contributing to this complexity is the fact that many of the variants discovered through GWAS are located within noncoding or intergenic regions[68, 95] which can make it difficult to identify the target genes, especially if the variant is located within a gene-dense region of the genome. The noncoding variants also make it challenging to determine the molecular mechanism(s) responsible for the association, because the variants often act to regulate expression of a gene, in contrast to altering the structure or function of a protein. Variants can also act to regulate genes across long distances[196, 197], which can make the identification of target genes challenging. Identifying the functional causal variant(s) at a locus and the gene(s) they may regulate is a crucial step in understanding the molecular mechanism(s) and biology underlying an association. This can ultimately provide insight into the direction of effect in humans and lead to a greater understanding of how genetic variation can contribute to complex metabolic traits, such as lipid and cholesterol levels, type 2 diabetes and WHR.

In the previous chapters, I described the functional follow-up of three GWAS loci associated with metabolic traits and disease. I identified multiple regulatory variants at the *GALNT2* HDL-C-associated locus[85]; at least two of these variants, rs4846913 and rs2281721, exhibited strong haplotype and allelic differences in transcriptional enhancer activity and differential protein binding in EMSAs (described in Chapter 2). The alleles associated with increased HDL-C also showed an association with increased *GALNT2* expression in primary human hepatocyte samples and subcutaneous adipose tissue samples, providing new clues into the direction of effect in humans that contrasts with previous findings based on *GALNT2* and HDL studies in mice[87]. At the *ADCY5* locus associated with T2D and glucose-related traits, I identified a functional regulatory variant, rs11708067, which is located within enhancer elements in human pancreatic islets (described in Chapter 3). Histone modification H3K27ac ChIP-seq reads,

which often mark transcriptional enhancer regions, showed significant allelic imbalance spanning rs11708067 in human pancreatic islets, and this variant also exhibited allelic differences in enhancer activity and differential protein binding. The T2D-risk allele of rs11708067 also showed an association with decreased *ADCY5* expression in human pancreatic islets. Finally, at the *PLXND1* locus associated with human WHR, I identified four regulatory elements that demonstrated enhancer activity in human umbilical vein endothelial cells, one of which also exhibited enhancer activity in endothelial cells in zebrafish (described in Chapter 4). These studies highlight the often complex molecular and biological mechanisms that are responsible for genome-wide associations, and underscore the value of using functional follow-up studies to further understand the genetic contribution of complex traits.

Many functional studies have reported single regulatory variants that likely drive the association signal[69, 81-84]. At other GWAS loci, multiple variants may influence the association signal. Functional studies on the *GALNT2* GWAS HDL-C locus showed that multiple noncoding variants in strong (or even moderate) linkage disequilibrium likely contribute to gene regulatory function and bind various transcription factors in a complex in an allele-specific manner. At least 2 regulatory variants, rs4846913 and rs2281721, showed strong allelic differences in transcriptional enhancer activity and differential protein binding in EMSAs. rs2144300 also showed evidence of allelic effects on enhancer activity when tested together with rs4846913 in a 780-bp element, and showed suggestive differences in transcription factor binding in EMSAs. Finally, variants rs1555290 and rs6143660 also exhibited differential protein binding in EMSAs. The observation that at least 4 *GALNT2* variants showed allelic differences in protein binding is not unexpected; a previous study showed that the majority of candidate regulatory variants at platelet quantitative trait loci overlapped transcription factor binding sites and affected binding[198]. Considering the numerous transcription factor ChIP-seq peaks from ENCODE data[72] that overlap *GALNT2* variants rs4846913, rs2144300, rs1555290 and rs6143660, it is possible that additional factors beyond those we identified bind to one or more of these variant(s) and may regulate expression of *GALNT2* or some other gene. More experiments are necessary to determine the identities of these proteins and to further investigate whether these factors contribute to allelic differences in regulatory activity. At the *ADCY5* locus, I identified a single variant, rs11708067, which influences regulatory activity and is associated with *ADCY5* expression. However, it is important to note that additional experiments will

be necessary to determine whether other variants also show evidence of allelic differences in regulatory activity or protein binding. The two additional variants in strong LD with rs11708067 are also located within predicted pancreatic islet enhancer regions, supporting their possible role in *ADCY5* regulation. At the *PLXND1* locus, we have identified multiple regulatory elements, which overlap numerous variants in strong linkage disequilibrium ($r^2 > 0.8$) with the WHR-associated lead GWAS variant. At least one of these variants affects transcription factor binding. Identification of multiple regulatory elements at the *PLXND1* locus suggests that there are also multiple functional regulatory variants, however, more experiments will be necessary to determine the full set of variants that exhibit regulatory activity.

Open chromatin, chromatin state, histone modification ChIP-seq and transcription factor ChIP-seq data have been successful at guiding the identification of regulatory variants at GWAS loci[79-84]. At the *ADCY5* locus, rs11708067 was a plausible candidate regulatory variant based on its location within a strong enhancer chromatin region and overlap with DNase and FAIRE peaks in human pancreatic islets. As predicted, it showed allelic differences in transcriptional enhancer activity and differential transcription factor binding. However, at other loci, regulatory data may not be informative in pinpointing functional variants, as is the case with the *PLXND1* WHR-associated locus. Of 35 variants in strong linkage disequilibrium ($r^2 > 0.8$, 1000 Genomes phase 1 EUR) with the lead GWAS variant, all of the variants overlap at least one mark of open chromatin and/or histone modifications marking active regulatory regions. Therefore, a more comprehensive approach (i.e. testing all candidate variants in strong linkage disequilibrium or testing larger regions that encompass many variants) are necessary to identify the causal variants. At the *GALNT2* HDL-C-associated locus, regulatory datasets were useful in predicting rs4846913, rs2144300, rs1555290 and rs6143660 as likely functional regulatory variants, as these variants overlapped the most regulatory marks of all candidate variants. However, testing all candidate variants in strong linkage disequilibrium ($r^2 > 0.7$, 1000 Genomes phase 1 EUR) with the lead GWAS variant enabled us to identify rs2281721 as an additional regulatory variant that showed the strongest evidence of enhancer activity in reporter assays at the *GALNT2* locus. Overall, regulatory datasets such as those generated by the ENCODE project[72] and Roadmap Epigenomics Consortium[73] can be used as a guide for identification of functional regulatory variants, but solely using these datasets to prioritize variants for functional follow-up studies may result in overlooking others that exhibit regulatory function.

My studies on the *GALNT2* locus have highlighted some of the limitations to using *in vitro* assays, such as luciferase reporter assays and EMSAs, for functional follow-up of GWAS loci. One of these limitations is the luciferase reporter assay relies on a segment of DNA cloned into a reporter vector, which does not reflect the actual chromatin conformation of the element in the genomic context of cells. While we were able to identify multiple regulatory variants at the *GALNT2* locus through luciferase assays, it is possible that additional variants also exhibit regulatory function, but were not detected because the DNA element was not in its natural chromatin conformation in the cell. Another important issue to consider when deciding which elements to test in these assays is the size of the element cloned within the vector. rs4846913 only showed strong allelic differences in enhancer activity within a larger 780-bp segment with additional variants rs2144300, rs1555290, and rs6143660; it did not show allelic effects on enhancer activity when tested individually in a 120-bp segment. This suggests that some variants at GWAS signals may show regulatory activity that is genomic context-specific; that is, larger segments of DNA that encompass regulatory marks such as open chromatin, histone modifications or transcription factor binding sites may be necessary to exhibit allelic effects on transcription. Another study[199] has confirmed that the context of a larger 489-bp segment encompassing rs4846913, as well as a 608-bp segment containing both rs4846913 and rs2144300 are necessary to observe allelic differences in transcriptional enhancer activity (Figure 5.1) which supports the conclusion that for some variants, it is necessary to test larger segment sizes in order to incorporate the whole regulatory element.

There are also limitations to *in vitro* EMSAs, which examine protein binding by taking the DNA out of the chromatin context in a cell. rs2281721 showed C-allele specific binding of USF-1 in EMSAs, however, I did not observe statistically significant allelic differences in USF-1 binding in two independent ChIP experiments (a more *in vivo* context) in HepG2 and Huh-7 cells. These inconsistent results could be the result of sub-optimal conditions for protein binding in the *in vitro* context of an EMSA. The strong allelic differences in transcriptional enhancer activity and USF-1 ChIP results for the rs2281721 region suggest that another transcription factor besides USF-1 may also bind to rs2281721 and cause the allelic differences in enhancer activity. USF-1 has been shown to bind with other transcription factors[200-202], therefore, other transcription factors bound to USF-1 may be necessary to affect allele-specific regulatory activity. Experiments to alter USF-1 expression (via overexpression or knockdown methods) will be

necessary to fully interrogate whether USF-1 influences expression of *GALNT2*, although the effects on expression may be modest and similar to what we observed with siRNA-mediated knockdown of CEBPB.

As we progress through the GWAS era into follow-up studies, fine-mapping analyses, as well as functional studies, are useful tools to refine GWAS signals and determine the functional candidate variants responsible for these signals. Fine-mapping and imputation strategies involve using linkage disequilibrium patterns to infer haplotypes in many individuals; therefore, these methods take into consideration numerous additional variants besides those that were directly genotyped[203-205]. Fine-mapping and conditional analyses indicate whether an association is a single signal consisting of common variants, or if there are underlying secondary (or additional) signals that are the result of other functional variants in moderate to weak LD with the lead GWAS variant[206, 207]. These fine-mapping and conditional analyses were very valuable for the further characterization of the HDL-C association signal at *GALNT2* and helped to distinguish that the HDL-C association consists of a single signal of the common GWAS-identified variants in strong linkage disequilibrium in intron 1. Fine-mapping has also aided in the characterization of association signals for fasting proinsulin and/or T2D[83, 172]. Additionally, trans-ancestry fine-mapping approaches can also be useful for refining association signals for complex traits because they take into consideration the differing LD patterns between populations. Trans-ancestry fine-mapping has helped to refine association signals, including associations for T2D[67, 208] and lipids[207]. These approaches also refined the LDL-C association signal at the *SORT1* locus and guided the identification of a functional regulatory variant[69]. Incorporating fine-mapping and/or trans-ancestry fine-mapping methods in future studies on the T2D and WHR association signals at the *ADCY5* and *PLXND1* loci, respectively, may be useful for further characterization of these association signals and for identification of the likely functional variants.

Functional studies in model organisms are also useful tools to understanding the underlying biology of GWAS loci. Using zebrafish transgenesis assays and creating stable lines were valuable in confirming that a regulatory element showing enhancer activity in HUVEC was conserved and also showed enhancer activity in endothelial cells in zebrafish. Using zebrafish transgenesis assays to identify *cis*-regulatory elements at the *PLXND1* locus were beneficial because they provided information about the spatial and temporal regulation by the enhancer element, CRE 17, which is not possible with

luciferase assays in cell lines. Using model systems such as zebrafish and mice may also provide a greater understanding of the relationship between GWAS target genes and metabolic traits and the direction of effect in humans. At the *GALNT2* HDL-C-associated locus, we observed a consistent direction of effect with the human GWAS variants and recent studies showing that total loss of function of *GALNT2* in human, mouse, rat and cynomolgus monkey all resulted in lower HDL-C (Khetarpal et al. 2013 abstract for The American Society of Human Genetics annual meeting). However, previously published *Galnt2* overexpression and knockdown studies in mice using adeno-associated viral vectors showed the opposite direction of effect[87]. These observations may be attributed to differences between partial loss of function knockdown and total knockout studies, or due to the fact that adenoviral vectors can have toxic effects[209], which may be confounding the effects on HDL-C. It is important to note that in some cases, using model organisms to study the molecular mechanisms that are responsible for GWAS signals can also be challenging, because of the innate species differences (i.e. lack of conservation of gene structure or gene regulatory elements) between animal models and humans. However, utilizing both human cell-based and animal models is ideal to gain further understanding of the relationship between the GWAS variants and their target genes, as well as how the target genes influence the trait or disease of study.

Additional experiments will be important for fully characterizing the molecular mechanisms responsible for the GWAS signals at the *GALNT2*, *ADCY5* and *PLXND1* loci. While we observed associations between the alleles of GWAS-associated variant rs4846914 and *GALNT2* expression in both human hepatocyte samples and subcutaneous adipose tissue samples[85] and an association between the alleles of rs11708067 and *ADCY5* expression in primary human islets, these associations do not prove that there is a direct interaction between the GWAS variants and the *GALNT2* or *ADCY5* gene promoters. It would be ideal to perform chromosome conformation capture experiments[210] to interrogate the potential gene targets of the identified enhancer elements at the *GALNT2*, *ADCY5* and *PLXND1* loci and to further characterize the molecular mechanisms for the *GALNT2* and *ADCY5* expression quantitative trait loci. Future studies utilizing human induced pluripotent stem cells (iPSCs) may also be helpful to gain a better understanding of the molecular and biological mechanisms at the *GALNT2*, *ADCY5* and *PLXND1* loci. The iPSCs could be differentiated into various cell types of interest

(i.e. hepatocytes for *GALNT2*, pancreatic islets for *ADCY5* or adipose endothelial cells for *PLXND1*) and transcriptional reporter assays could be performed to determine in which cell types the regulatory enhancers act. Additionally genome editing by CRISPR-Cas9 technology[211-213] could be used to generate cells of different genotypes in order to test which variants at the *ADCY5* and *PLXND1* loci affect expression of *ADCY5* and *PLXND1*, respectively, or if the variant affect expression of other nearby genes. Additional experiments will be necessary to further characterize the molecular mechanisms at the *ADCY5* and *PLXND1* loci. These experiments include testing whether there are additional functional regulatory variants at the *ADCY5* locus, as there are other likely candidates located within predicted regulatory elements, and to identify the protein differentially bound to rs11708067. At the *PLXND1* locus, amplifying, cloning and testing the remainder of the candidate CREs in HUVEC transcriptional reporter assays and zebrafish transgenesis assays will be necessary to determine if there are additional regulatory elements. Further investigation of which candidate CREs at the *PLXND1* locus contain the functional variants that drive the WHR association is also essential to characterizing how these variants act to influence body fat distribution.

My dream experiments for future follow-up studies include using genome editing technologies (i.e. CRISPR-Cas9) to alter regulatory enhancer regions and to investigate the effects on expression of candidate target genes. I would perform these genome editing experiments in primary human cells and tissues. Additionally, I would conduct these experiments using induced pluripotent stem cells before and after differentiation into multiple cell types to investigate whether functional regulatory variants or enhancer elements function at a specific developmental time point or in a specific cell type. I would also perform chromosome conformation capture experiments to validate a direct connection between identified regulatory enhancer elements at GWAS loci and target gene promoters. I would use a more high-throughput approach to investigate the functional regulatory variants and elements at GWAS loci; this approach would incorporate high-throughput transcriptional reporter assays and chromosome conformation capture experiments to more comprehensively identify regulatory variants and target genes and characterize the molecular mechanisms at GWAS loci for metabolic traits. Finally, I would also study the biology of novel genes identified through GWAS using model organisms such as mice and zebrafish.

Studies such as gene knockout experiments in animal models can ultimately help to characterize the biological mechanisms and pathways that influence the metabolic trait or disease of interest.

Overall, the studies described in Chapters 2, 3 and 4 provide examples of what GWAS functional follow-up studies entail and highlight the complexity of determining the underlying molecular and biological mechanisms that are responsible for complex metabolic traits. Multiple regulatory variants can contribute to association signal(s) and be located within regulatory elements that influence expression of target genes. Using regulatory datasets can be useful to pinpoint the most likely candidate regulatory elements or variants, but performing functional assays and experiments to determine all of the potential regulatory elements and variants are essential. Ultimately, these GWAS follow-up studies can help to identify the direction of effect in humans and how the associated variants influence metabolic traits and diseases, such as HDL-C, T2D and WHR.

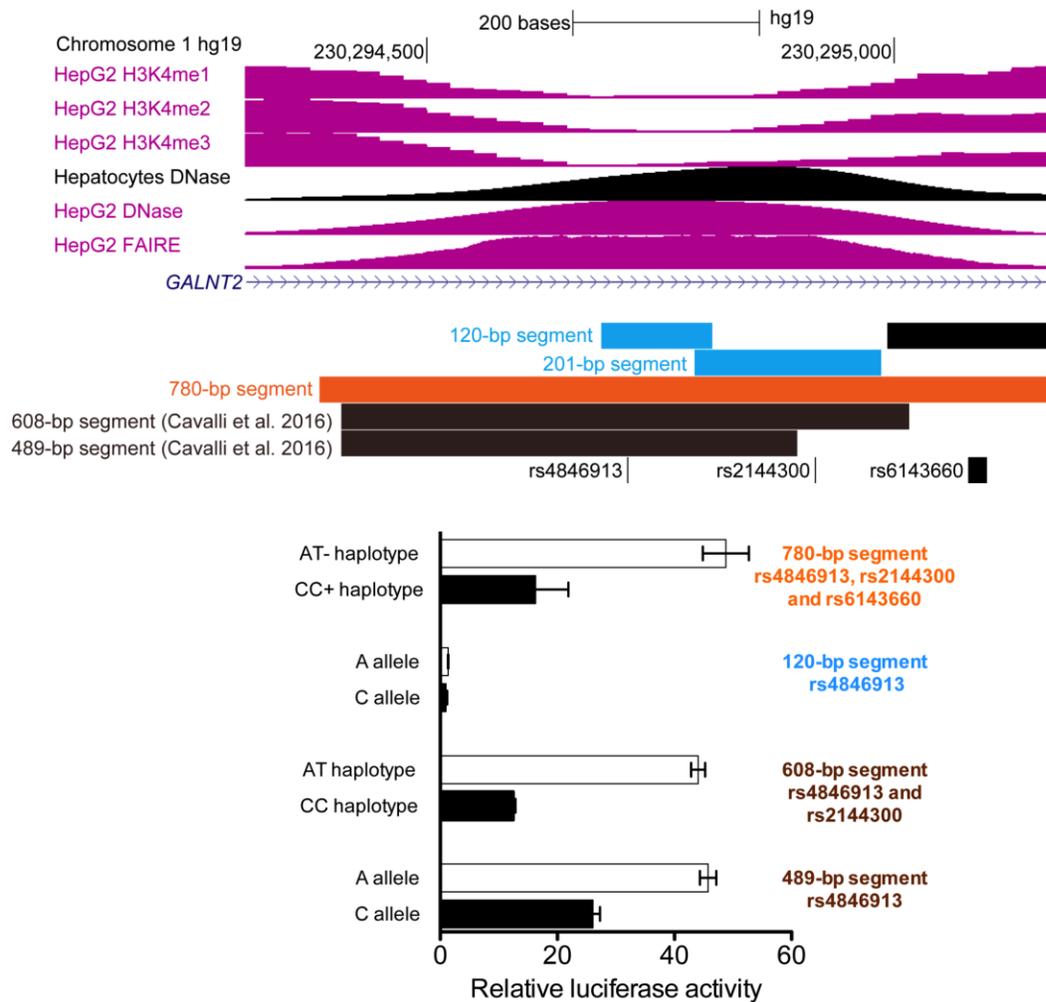


Figure 5.1 Allele and haplotype differences in transcriptional enhancer activity are dependent on size of segment cloned into the luciferase reporter vector

Top: Genome browser view of elements tested in luciferase reporter assays (blue, orange and black rectangles) and their overlap with HDL-C-associated variants rs4846913, rs2144300, rs6143660 and open chromatin and histone modifications.

Bottom: Relative luciferase activities of a 780-bp segment containing the AT- haplotype or CC+ haplotype of rs4846913, rs2144300 and rs6143660; 120-bp segment containing the A or C allele of rs4846913; 608-bp segment containing the AT haplotype or CC haplotype of rs4846913 and rs2144300 and 489-bp segment containing the A or C allele of rs4846913. Relative luciferase activity is shown on the x-axis.

REFERENCES

- [1] R. Arena, M. Guazzi, L. Lianov, L. Whitsel, K. Berra, C.J. Lavie, L. Kaminsky, M. Williams, M.F. Hivert, N.C. Franklin, J. Myers, D. Dengel, D.M. Lloyd-Jones, F.J. Pinto, F. Cosentino, M. Halle, S. Gielen, P. Dendale, J. Niebauer, A. Pelliccia, P. Giannuzzi, U. Corra, M.F. Piepoli, G. Guthrie, D. Shurney, Healthy Lifestyle Interventions to Combat Noncommunicable Disease—A Novel Nonhierarchical Connectivity Model for Key Stakeholders: A Policy Statement From the American Heart Association, European Society of Cardiology, European Association for Cardiovascular Prevention and Rehabilitation, and American College of Preventive Medicine, *Mayo Clin Proc*, 90 (2015) 1082-1103.
- [2] D. Mozaffarian, E.J. Benjamin, A.S. Go, D.K. Arnett, M.J. Blaha, M. Cushman, S.R. Das, S. de Ferranti, J.P. Després, H.J. Fullerton, V.J. Howard, M.D. Huffman, C.R. Isasi, M.C. Jiménez, S.E. Judd, B.M. Kissela, J.H. Lichtman, L.D. Lisabeth, S. Liu, R.H. Mackey, D.J. Magid, D.K. McGuire, E.R. Mohler, C.S. Moy, P. Muntner, M.E. Mussolino, K. Nasir, R.W. Neumar, G. Nichol, L. Palaniappan, D.K. Pandey, M.J. Reeves, C.J. Rodriguez, W. Rosamond, P.D. Sorlie, J. Stein, A. Towfighi, T.N. Turan, S.S. Virani, D. Woo, R.W. Yeh, M.B. Turner, A.H.A.S.C.a.S.S. Subcommittee, Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association, *Circulation*, 133 (2016) e38-e360.
- [3] D. Mozaffarian, E.J. Benjamin, A.S. Go, D.K. Arnett, M.J. Blaha, M. Cushman, S. de Ferranti, J.P. Després, H.J. Fullerton, V.J. Howard, M.D. Huffman, S.E. Judd, B.M. Kissela, D.T. Lackland, J.H. Lichtman, L.D. Lisabeth, S. Liu, R.H. Mackey, D.B. Matchar, D.K. McGuire, E.R. Mohler, C.S. Moy, P. Muntner, M.E. Mussolino, K. Nasir, R.W. Neumar, G. Nichol, L. Palaniappan, D.K. Pandey, M.J. Reeves, C.J. Rodriguez, P.D. Sorlie, J. Stein, A. Towfighi, T.N. Turan, S.S. Virani, J.Z. Willey, D. Woo, R.W. Yeh, M.B. Turner, A.H.A.S.C.a.S.S. Subcommittee, Heart disease and stroke statistics--2015 update: a report from the American Heart Association, *Circulation*, 131 (2015) e29-322.
- [4] P.A. Heidenreich, J.G. Trogdon, O.A. Khavjou, J. Butler, K. Dracup, M.D. Ezekowitz, E.A. Finkelstein, Y. Hong, S.C. Johnston, A. Khera, D.M. Lloyd-Jones, S.A. Nelson, G. Nichol, D. Orenstein, P.W. Wilson, Y.J. Woo, A.H.A.A.C. Committee, S. Council, C.o.C.R.a. Intervention, C.o.C. Cardiology, C.o.E.a. Prevention, C.o. Arteriosclerosis, T.a.V. Biology, C.o. Cardiopulmonary, C. Care, P.a. Resuscitation, C.o.C. Nursing, C.o.t.K.i.C. Disease, a.I.C.o.Q.o.C.a.O.R. Council on Cardiovascular Surgery and Anesthesia, Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association, *Circulation*, 123 (2011) 933-944.
- [5] A.R. Sharrett, C.M. Ballantyne, S.A. Coady, G. Heiss, P.D. Sorlie, D. Catellier, W. Patsch, A.R.i.C.S. Group, 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 (2001) 1108-1113.
- [6] M.T. Cooney, A. Dudina, D. De Bacquer, L. Wilhelmsen, S. Sans, A. Menotti, G. De Backer, P. Jousilahti, U. Keil, T. Thomsen, P. Whincup, I.M. Graham, S. investigators, HDL cholesterol protects against cardiovascular disease in both genders, at all ages and at all levels of risk, *Atherosclerosis*, 206 (2009) 611-616.
- [7] J.A. Farmer, J. Liao, Evolving concepts of the role of high-density lipoprotein in protection from atherosclerosis, *Curr Atheroscler Rep*, 13 (2011) 107-114.
- [8] B.G. Talayero, F.M. Sacks, The role of triglycerides in atherosclerosis, *Curr Cardiol Rep*, 13 (2011) 544-552.
- [9] M.A. Cornier, J.P. Després, N. Davis, D.A. Grossniklaus, S. Klein, B. Lamarche, F. Lopez-Jimenez, G. Rao, M.P. St-Onge, A. Towfighi, P. Poirier, A.H.A.O.C.o.t.C.o. Nutrition, P.A.a. Metabolism, C.o. Arteriosclerosis, T.a.V. Biology, C.o.C.D.i.t. Young, C.o.C.R.a. Intervention, C.u.o.E.a.P. Council on Cardiovascular Nursing, a.S.C. Council on the Kidney in Cardiovascular Disease, Assessing adiposity: a scientific statement from the American Heart Association, *Circulation*, 124 (2011) 1996-2019.

- [10] A.M. Sironi, R. Petz, D. De Marchi, E. Buzzigoli, D. Ciociaro, V. Positano, M. Lombardi, E. Ferrannini, A. Gastaldelli, Impact of increased visceral and cardiac fat on cardiometabolic risk and disease, *Diabet Med*, 29 (2012) 622-627.
- [11] E.a. Expert Panel on Detection, and Treatment of High Blood Cholesterol in Adults, Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III), *JAMA*, 285 (2001) 2486-2497.
- [12] R.J. Chilton, Pathophysiology of coronary heart disease: a brief review, *J Am Osteopath Assoc*, 104 (2004) S5-8.
- [13] L. Badimon, G. Vilahur, LDL-cholesterol versus HDL-cholesterol in the atherosclerotic plaque: inflammatory resolution versus thrombotic chaos, *Ann N Y Acad Sci*, 1254 (2012) 18-32.
- [14] E.M. deGoma, J.W. Knowles, F. Angeli, M.J. Budoff, D.J. Rader, The evolution and refinement of traditional risk factors for cardiovascular disease, *Cardiol Rev*, 20 (2012) 118-129.
- [15] M. Inouye, J. Kettunen, P. Soininen, K. Silander, S. Ripatti, L.S. Kumpula, E. Hämmäläinen, P. Jousilahti, A.J. Kangas, S. Männistö, M.J. Savolainen, A. Jula, J. Leiviskä, A. Palotie, V. Salomaa, M. Perola, M. Ala-Korpela, L. Peltonen, Metabonomic, transcriptomic, and genomic variation of a population cohort, *Mol Syst Biol*, 6 (2010) 441.
- [16] D.J. Gordon, B.M. Rifkind, High-density lipoprotein--the clinical implications of recent studies, *N Engl J Med*, 321 (1989) 1311-1316.
- [17] P.H. Joshi, P.P. Toth, S.T. Lirette, M.E. Griswold, J.M. Massaro, S.S. Martin, M.J. Blaha, K.R. Kulkarni, A.A. Khokhar, A. Correa, R.B. D'Agustino, S.R. Jones, L.I.C.L.S. Group, Association of high-density lipoprotein subclasses and incident coronary heart disease: The Jackson Heart and Framingham Offspring Cohort Studies, *Eur J Prev Cardiol*, 23 (2016) 41-49.
- [18] D.J. Rader, Molecular regulation of HDL metabolism and function: implications for novel therapies, *J Clin Invest*, 116 (2006) 3090-3100.
- [19] D.J. Rader, E.T. Alexander, G.L. Weibel, J. Billheimer, G.H. Rothblat, The role of reverse cholesterol transport in animals and humans and relationship to atherosclerosis, *J Lipid Res*, 50 Suppl (2009) S189-194.
- [20] D.J. Rader, A.R. Tall, The not-so-simple HDL story: Is it time to revise the HDL cholesterol hypothesis?, *Nat Med*, 18 (2012) 1344-1346.
- [21] R.S. Rosenson, H.B. Brewer, B. Ansell, P. Barter, M.J. Chapman, J.W. Heinecke, A. Kontush, A.R. Tall, N.R. Webb, Translation of high-density lipoprotein function into clinical practice: current prospects and future challenges, *Circulation*, 128 (2013) 1256-1267.
- [22] D.J. Rader, G.K. Hovingh, HDL and cardiovascular disease, *Lancet*, 384 (2014) 618-625.
- [23] A. Bhatt, A. Rohatgi, HDL Cholesterol Efflux Capacity: Cardiovascular Risk Factor and Potential Therapeutic Target, *Curr Atheroscler Rep*, 18 (2016) 2.
- [24] D.J. Rader, Pathophysiology and management of low high-density lipoprotein cholesterol, *Am J Cardiol*, 83 (1999) 22F-24F.
- [25] M.A. Kawashiri, C. Maugeais, D.J. Rader, High-density lipoprotein metabolism: molecular targets for new therapies for atherosclerosis, *Curr Atheroscler Rep*, 2 (2000) 363-372.

- [26] D.J. Rader, Regulation of reverse cholesterol transport and clinical implications, *Am J Cardiol*, 92 (2003) 42J-49J.
- [27] A.D. Association, Economic costs of diabetes in the U.S. in 2012, *Diabetes Care*, 36 (2013) 1033-1046.
- [28] M. Stumvoll, B.J. Goldstein, T.W. van Haeften, Type 2 diabetes: principles of pathogenesis and therapy, *Lancet*, 365 (2005) 1333-1346.
- [29] M. Laakso, Is Insulin Resistance a Feature of or a Primary Risk Factor for Cardiovascular Disease?, *Curr Diab Rep*, 15 (2015) 105.
- [30] S. Casella, A. Bielli, A. Mauriello, A. Orlandi, Molecular Pathways Regulating Macrovascular Pathology and Vascular Smooth Muscle Cells Phenotype in Type 2 Diabetes, *Int J Mol Sci*, 16 (2015) 24353-24368.
- [31] J.B. Buse, H.N. Ginsberg, G.L. Bakris, N.G. Clark, F. Costa, R. Eckel, V. Fonseca, H.C. Gerstein, S. Grundy, R.W. Nesto, M.P. Pignone, J. Plutzky, D. Porte, R. Redberg, K.F. Stitzel, N.J. Stone, A.H. Association, A.D. Association, Primary prevention of cardiovascular diseases in people with diabetes mellitus: a scientific statement from the American Heart Association and the American Diabetes Association, *Diabetes Care*, 30 (2007) 162-172.
- [32] S. Faghihi-Kashani, F. Bonnet, N. Hafezi-Nejad, B. Heidari, A. Aghajani Nargesi, S. Sheikhabaei, M. Ebadi, A. Esteghamati, Fasting hyperinsulinaemia and 2-h glycaemia predict coronary heart disease in patients with type 2 diabetes, *Diabetes Metab*, 42 (2016) 55-61.
- [33] J.B. Meigs, M.G. Larson, R.B. D'Agostino, D. Levy, M.E. Clouse, D.M. Nathan, P.W. Wilson, C.J. O'Donnell, Coronary artery calcification in type 2 diabetes and insulin resistance: the framingham offspring study, *Diabetes Care*, 25 (2002) 1313-1319.
- [34] C.S. Fox, L. Sullivan, R.B. D'Agostino, P.W. Wilson, F.H. Study, The significant effect of diabetes duration on coronary heart disease mortality: the Framingham Heart Study, *Diabetes Care*, 27 (2004) 704-708.
- [35] R.T. Hurt, C. Kulisek, L.A. Buchanan, S.A. McClave, The obesity epidemic: challenges, health initiatives, and implications for gastroenterologists, *Gastroenterol Hepatol (N Y)*, 6 (2010) 780-792.
- [36] L.H. Opie, D.M. Yellon, B.J. Gersh, Controversies in the cardiovascular management of type 2 diabetes, *Heart*, 97 (2011) 6-14.
- [37] S. Lemieux, D. Prud'homme, C. Bouchard, A. Tremblay, J.P. Després, A single threshold value of waist girth identifies normal-weight and overweight subjects with excess visceral adipose tissue, *Am J Clin Nutr*, 64 (1996) 685-693.
- [38] P. Björntorp, Regional patterns of fat distribution, *Ann Intern Med*, 103 (1985) 994-995.
- [39] C.S. Fox, J.M. Massaro, U. Hoffmann, K.M. Pou, P. Maurovich-Horvat, C.Y. Liu, R.S. Vasan, J.M. Murabito, J.B. Meigs, L.A. Cupples, R.B. D'Agostino, C.J. O'Donnell, Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study, *Circulation*, 116 (2007) 39-48.
- [40] D. Schleinitz, Y. Böttcher, M. Blüher, P. Kovacs, The genetics of fat distribution, *Diabetologia*, 57 (2014) 1276-1286.

- [41] S.H. Scheuer, K. Færch, A. Philipsen, M.E. Jørgensen, N.B. Johansen, B. Carstensen, D.R. Witte, I. Andersen, T. Lauritzen, G.S. Andersen, Abdominal Fat Distribution and Cardiovascular Risk in Men and Women With Different Levels of Glucose Tolerance, *J Clin Endocrinol Metab*, 100 (2015) 3340-3347.
- [42] O. Poulain-Godefroy, C. Lecoeur, F. Pattou, G. Frühbeck, P. Froguel, Inflammation is associated with a decrease of lipogenic factors in omental fat in women, *Am J Physiol Regul Integr Comp Physiol*, 295 (2008) R1-7.
- [43] I. Harman-Boehm, M. Blüher, H. Redel, N. Sion-Vardy, S. Ovadia, E. Avinoach, I. Shai, N. Klötting, M. Stumvoll, N. Bashan, A. Rudich, Macrophage infiltration into omental versus subcutaneous fat across different populations: effect of regional adiposity and the comorbidities of obesity, *J Clin Endocrinol Metab*, 92 (2007) 2240-2247.
- [44] M.J. Lee, Y. Wu, S.K. Fried, Adipose tissue heterogeneity: Implication of depot differences in adipose tissue for obesity complications, *Mol Aspects Med*, 34 (2013) 1-11.
- [45] E. Vanni, E. Bugianesi, A. Kotronen, S. De Minicis, H. Yki-Järvinen, G. Svegliati-Baroni, From the metabolic syndrome to NAFLD or vice versa?, *Dig Liver Dis*, 42 (2010) 320-330.
- [46] M.I. Schmidt, R.L. Watson, B.B. Duncan, P. Metcalf, F.L. Brancati, A.R. Sharrett, C.E. Davis, G. Heiss, Clustering of dyslipidemia, hyperuricemia, diabetes, and hypertension and its association with fasting insulin and central and overall obesity in a general population. Atherosclerosis Risk in Communities Study Investigators, *Metabolism*, 45 (1996) 699-706.
- [47] P. Bunnag, S. Chanprasertyothin, A. Kongsuksai, B. Ongphiphadhanakul, R. Rajatanavin, G. Puavilai, Correlation between serum insulin and features of metabolic syndrome in Thais, *J Med Assoc Thai*, 83 (2000) 783-789.
- [48] E. Mansfield, R. McPherson, K.G. Koski, Diet and waist-to-hip ratio: important predictors of lipoprotein levels in sedentary and active young men with no evidence of cardiovascular disease, *J Am Diet Assoc*, 99 (1999) 1373-1379.
- [49] L.O. Schulz, P.H. Bennett, E. Ravussin, J.R. Kidd, K.K. Kidd, J. Esparza, M.E. Valencia, Effects of traditional and western environments on prevalence of type 2 diabetes in Pima Indians in Mexico and the U.S., *Diabetes Care*, 29 (2006) 1866-1871.
- [50] R.P. Middelberg, N.G. Martin, J.B. Whitfield, Longitudinal genetic analysis of plasma lipids, *Twin Res Hum Genet*, 9 (2006) 550-557.
- [51] I.J. Kullo, K. Ding, Mechanisms of disease: The genetic basis of coronary heart disease, *Nat Clin Pract Cardiovasc Med*, 4 (2007) 558-569.
- [52] G. Jermendy, T. Horváth, L. Littvay, R. Steinbach, A.L. Jermendy, A.D. Tárnoki, D.L. Tárnoki, J. Métneki, J. Osztoivits, Effect of genetic and environmental influences on cardiometabolic risk factors: a twin study, *Cardiovasc Diabetol*, 10 (2011) 96.
- [53] A. Iliadou, H. Snieder, X. Wang, F.A. Treiber, C.L. Davis, Heritabilities of lipids in young European American and African American twins, *Twin Res Hum Genet*, 8 (2005) 492-498.
- [54] R.P. Middelberg, N.G. Martin, J.B. Whitfield, A longitudinal genetic study of plasma lipids in adolescent twins, *Twin Res Hum Genet*, 10 (2007) 127-135.
- [55] N.Y. Souren, A.D. Paulussen, R.J. Loos, M. Gielen, G. Beunen, R. Fagard, C. Derom, R. Vlietinck, M.P. Zeegers, Anthropometry, carbohydrate and lipid metabolism in the East Flanders Prospective Twin Survey: heritabilities, *Diabetologia*, 50 (2007) 2107-2116.

- [56] D. Weissglas-Volkov, P. Pajukanta, Genetic causes of high and low serum HDL-cholesterol, *J Lipid Res*, 51 (2010) 2032-2057.
- [57] S.R. Browning, B.L. Browning, Identity-by-descent-based heritability analysis in the Northern Finland Birth Cohort, *Hum Genet*, 132 (2013) 129-138.
- [58] J. van Dongen, G. Willemsen, W.M. Chen, E.J. de Geus, D.I. Boomsma, Heritability of metabolic syndrome traits in a large population-based sample, *J Lipid Res*, 54 (2013) 2914-2923.
- [59] P. Almgren, M. Lehtovirta, B. Isomaa, L. Sarelin, M.R. Taskinen, V. Lyssenko, T. Tuomi, L. Groop, B.S. Group, Heritability and familiarity of type 2 diabetes and related quantitative traits in the Botnia Study, *Diabetologia*, 54 (2011) 2811-2819.
- [60] G. Willemsen, K.J. Ward, C.G. Bell, K. Christensen, J. Bowden, C. Dalgård, J.R. Harris, J. Kaprio, R. Lyle, P.K. Magnusson, K.A. Mather, J.R. Ordoñana, F. Perez-Riquelme, N.L. Pedersen, K.H. Pietiläinen, P.S. Sachdev, D.I. Boomsma, T. Spector, The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium, *Twin Res Hum Genet*, 18 (2015) 762-771.
- [61] J.V. Selby, B. Newman, C.P. Quesenberry, R.R. Fabsitz, D. Carmelli, F.J. Meaney, C. Slemenda, Genetic and behavioral influences on body fat distribution, *Int J Obes*, 14 (1990) 593-602.
- [62] K.M. Rose, B. Newman, E.J. Mayer-Davis, J.V. Selby, Genetic and behavioral determinants of waist-hip ratio and waist circumference in women twins, *Obes Res*, 6 (1998) 383-392.
- [63] C.J. Willer, E.M. Schmidt, S. Sengupta, G.M. Peloso, S. Gustafsson, S. Kanoni, A. Ganna, J. Chen, M.L. Buchkovich, S. Mora, J.S. Beckmann, J.L. Bragg-Gresham, H.Y. Chang, A. Demirkan, H.M. Den Hertog, R. Do, L.A. Donnelly, G.B. Ehret, T. Esko, M.F. Feitosa, T. Ferreira, K. Fischer, P. Fontanillas, R.M. Fraser, D.F. Freitag, D. Gurdasani, K. Heikkilä, E. Hyppönen, A. Isaacs, A.U. Jackson, A. Johansson, T. Johnson, M. Kaakinen, J. Kettunen, M.E. Kleber, X. Li, J. Luan, L.P. Lytykäinen, P.K. Magnusson, M. Mangino, E. Mihailov, M.E. Montasser, M. Müller-Nurasyid, I.M. Nolte, J.R. O'Connell, C.D. Palmer, M. Perola, A.K. Petersen, S. Sanna, R. Saxena, S.K. Service, S. Shah, D. Shungin, C. Sidore, C. Song, R.J. Strawbridge, I. Surakka, T. Tanaka, T.M. Teslovich, G. Thorleifsson, E.G. Van den Herik, B.F. Voight, K.A. Volcik, L.L. Waite, A. Wong, Y. Wu, W. Zhang, D. Absher, G. Asiki, I. Barroso, L.F. Been, J.L. Bolton, L.L. Bonnycastle, P. Brambilla, M.S. Burnett, G. Cesana, M. Dimitriou, A.S. Doney, A. Döring, P. Elliott, S.E. Epstein, G.I. Eyjolfsson, B. Gigante, M.O. Goodarzi, H. Grallert, M.L. Gravito, C.J. Groves, G. Hallmans, A.L. Hartikainen, C. Hayward, D. Hernandez, A.A. Hicks, H. Holm, Y.J. Hung, T. Illig, M.R. Jones, P. Kaleebu, J.J. Kastelein, K.T. Khaw, E. Kim, N. Klopp, P. Komulainen, M. Kumari, C. Langenberg, T. Lehtimäki, S.Y. Lin, J. Lindström, R.J. Loos, F. Mach, W.L. McArdle, C. Meisinger, B.D. Mitchell, G. Müller, R. Nagaraja, N. Narisu, T.V. Nieminen, R.N. Nsubuga, I. Olafsson, K.K. Ong, A. Palotie, T. Papamarkou, C. Pomilla, A. Pouta, D.J. Rader, M.P. Reilly, P.M. Ridker, F. Rivadeneira, I. Rudan, A. Ruukonen, N. Samani, H. Scharnagl, J. Seeley, K. Silander, A. Stancáková, K. Stirrups, A.J. Swift, L. Tiret, A.G. Uitterlinden, L.J. van Pelt, S. Vedantam, N. Wainwright, C. Wijmenga, S.H. Wild, G. Willemsen, T. Wilsgaard, J.F. Wilson, E.H. Young, J.H. Zhao, L.S. Adair, D. Arveiler, T.L. Assimes, S. Bandinelli, F. Bennett, M. Bochud, B.O. Boehm, D.I. Boomsma, I.B. Borecki, S.R. Bornstein, P. Bovet, M. Burnier, H. Campbell, A. Chakravarti, J.C. Chambers, Y.D. Chen, F.S. Collins, R.S. Cooper, J. Danesh, G. Dedoussis, U. de Faire, A.B. Feranil, J. Ferrières, L. Ferrucci, N.B. Freimer, C. Gieger, L.C. Groop, V. Gudnason, U. Gyllenstein, A. Hamsten, T.B. Harris, A. Hingorani, J.N. Hirschhorn, A. Hofman, G.K. Hovingh, C.A. Hsiung, S.E. Humphries, S.C. Hunt, K. Hveem, C. Iribarren, M.R. Jarvelin, A. Julia, M. Kähönen, J. Kaprio, A. Kesäniemi, M. Kivimäki, J.S. Kooner, P.J. Koudstaal, R.M. Krauss, D. Kuh, J. Kuusisto, K.O. Kyvik, M. Laakso, T.A. Lakka, L. Lind, C.M. Lindgren, N.G. Martin, W. März, M.I. McCarthy, C.A. McKenzie, P. Meneton, A. Metspalu, L. Moilanen, A.D. Morris, P.B. Munroe, I. Njølstad, N.L. Pedersen, C. Power, P.P. Pramstaller, J.F. Price, B.M. Psaty, T. Quertermous, R. Rauramaa, D. Saleheen, V. Salomaa, D.K. Sanghera, J. Saramies, P.E. Schwarz, W.H. Sheu, A.R. Shuldiner, A. Siegbahn, T.D. Spector, K. Stefansson, D.P. Strachan, B.O. Tayo, E. Tremoli, J. Tuomilehto, M. Uusitupa, C.M. van Duijn, P. Vollenweider, L. Wallentin, N.J. Wareham, J.B. Whitfield, B.H. Wolfenbuttel, J.M.

Ordovas, E. Boerwinkle, C.N. Palmer, U. Thorsteinsdottir, D.I. Chasman, J.I. Rotter, P.W. Franks, S. Ripatti, L.A. Cupples, M.S. Sandhu, S.S. Rich, M. Boehnke, P. Deloukas, S. Kathiresan, K.L. Mohlke, E. Ingelsson, G.R. Abecasis, G.L.G. Consortium, Discovery and refinement of loci associated with lipid levels, *Nat Genet*, 45 (2013) 1274-1283.

[64] I.M. Heid, A.U. Jackson, J.C. Randall, T.W. Winkler, L. Qi, V. Steinthorsdottir, G. Thorleifsson, M.C. Zillikens, E.K. Speliotes, R. Mägi, T. Workalemahu, C.C. White, N. Bouatia-Naji, T.B. Harris, S.I. Berndt, E. Ingelsson, C.J. Willer, M.N. Weedon, J. Luan, S. Vedantam, T. Esko, T.O. Kilpeläinen, Z. Kutalik, S. Li, K.L. Monda, A.L. Dixon, C.C. Holmes, L.M. Kaplan, L. Liang, J.L. Min, M.F. Moffatt, C. Molony, G. Nicholson, E.E. Schadt, K.T. Zondervan, M.F. Feitosa, T. Ferreira, H. Lango Allen, R.J. Weyant, E. Wheeler, A.R. Wood, K. Estrada, M.E. Goddard, G. Lettre, M. Mangino, D.R. Nyholt, S. Purcell, A.V. Smith, P.M. Visscher, J. Yang, S.A. McCarroll, J. Nemesl, B.F. Voight, D. Absher, N. Amin, T. Aspelund, L. Coin, N.L. Glazer, C. Hayward, N.L. Heard-Costa, J.J. Hottenga, A. Johansson, T. Johnson, M. Kaakinen, K. Kapur, S. Ketkar, J.W. Knowles, P. Kraft, A.T. Kraja, C. Lamina, M.F. Leitzmann, B. McKnight, A.P. Morris, K.K. Ong, J.R. Perry, M.J. Peters, O. Polasek, I. Prokopenko, N.W. Rayner, S. Ripatti, F. Rivadeneira, N.R. Robertson, S. Sanna, U. Sovio, I. Surakka, A. Teumer, S. van Wingerden, V. Vitart, J.H. Zhao, C. Cavalcanti-Proença, P.S. Chines, E. Fisher, J.R. Kulzer, C. Lecoeur, N. Narisu, C. Sandholt, L.J. Scott, K. Silander, K. Stark, M.L. Tammesoo, T.M. Teslovich, N.J. Timpson, R.M. Watanabe, R. Welch, D.I. Chasman, M.N. Cooper, J.O. Jansson, J. Kettunen, R.W. Lawrence, N. Pellikka, M. Perola, L. Vandenput, H. Alavere, P. Almgren, L.D. Atwood, A.J. Bennett, R. Biffar, L.L. Bonnycastle, S.R. Bornstein, T.A. Buchanan, H. Campbell, I.N. Day, M. Dei, M. Dörr, P. Elliott, M.R. Erdos, J.G. Eriksson, N.B. Freimer, M. Fu, S. Gaget, E.J. Geus, A.P. Gjesing, H. Grallert, J. Grässler, C.J. Groves, C. Guiducci, A.L. Hartikainen, N. Hassanali, A.S. Havulinna, K.H. Herzig, A.A. Hicks, J. Hui, W. Igl, P. Jousilahti, A. Jula, E. Kajantie, L. Kinnunen, I. Kolcic, S. Koskinen, P. Kovacs, H.K. Kroemer, V. Krzelj, J. Kuusisto, K. Kvaloy, J. Laitinen, O. Lantieri, G.M. Lathrop, M.L. Lokki, R.N. Luben, B. Ludwig, W.L. McArdle, A. McCarthy, M.A. Morken, M. Nelis, M.J. Neville, G. Paré, A.N. Parker, J.F. Peden, I. Pichler, K.H. Pietiläinen, C.G. Platou, A. Pouta, M. Ridderstråle, N.J. Samani, J. Saramies, J. Sinisalo, J.H. Smit, R.J. Strawbridge, H.M. Stringham, A.J. Swift, M. Teder-Laving, B. Thomson, G. Usala, J.B. van Meurs, G.J. van Ommen, V. Vatin, C.B. Volpato, H. Wallaschofski, G.B. Walters, E. Widen, S.H. Wild, G. Willemssen, D.R. Witte, L. Zgaga, P. Zitting, J.P. Beilby, A.L. James, M. Kähönen, T. Lehtimäki, M.S. Nieminen, C. Ohlsson, L.J. Palmer, O. Raitakari, P.M. Ridker, M. Stumvoll, A. Tönjes, J. Viikari, B. Balkau, Y. Ben-Shlomo, R.N. Bergman, H. Boeing, G.D. Smith, S. Ebrahim, P. Froguel, T. Hansen, C. Hengstenberg, K. Hveem, B. Isomaa, T. Jørgensen, F. Karpe, K.T. Khaw, M. Laakso, D.A. Lawlor, M. Marre, T. Meitinger, A. Metspalu, K. Midthjell, O. Pedersen, V. Salomaa, P.E. Schwarz, T. Tuomi, J. Tuomilehto, T.T. Valle, N.J. Wareham, A.M. Arnold, J.S. Beckmann, S. Bergmann, E. Boerwinkle, D.I. Boomsma, M.J. Caulfield, F.S. Collins, G. Eiriksdottir, V. Gudnason, U. Gyllensten, A. Hamsten, A.T. Hattersley, A. Hofman, F.B. Hu, T. Illig, C. Iribarren, M.R. Jarvelin, W.H. Kao, J. Kaprio, L.J. Launer, P.B. Munroe, B. Oostra, B.W. Penninx, P.P. Pramstaller, B.M. Psaty, T. Quertermous, A. Rissanen, I. Rudan, A.R. Shuldiner, N. Soranzo, T.D. Spector, A.C. Syvanen, M. Uda, A. Uitterlinden, H. Völzke, P. Vollenweider, J.F. Wilson, J.C. Witterman, A.F. Wright, G.R. Abecasis, M. Boehnke, I.B. Borecki, P. Deloukas, T.M. Frayling, L.C. Groop, T. Haritunians, D.J. Hunter, R.C. Kaplan, K.E. North, J.R. O'Connell, L. Peltonen, D. Schlessinger, D.P. Strachan, J.N. Hirschhorn, T.L. Assimes, H.E. Wichmann, U. Thorsteinsdottir, C.M. van Duijn, K. Stefansson, L.A. Cupples, R.J. Loos, I. Barroso, M.I. McCarthy, C.S. Fox, K.L. Mohlke, C.M. Lindgren, MAGIC, Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution, *Nat Genet*, 42 (2010) 949-960.

[65] D. Shungin, T.W. Winkler, D.C. Croteau-Chonka, T. Ferreira, A.E. Locke, R. Mägi, R.J. Strawbridge, T.H. Pers, K. Fischer, A.E. Justice, T. Workalemahu, J.M. Wu, M.L. Buchkovich, N.L. Heard-Costa, T.S. Roman, A.W. Drong, C. Song, S. Gustafsson, F.R. Day, T. Esko, T. Fall, Z. Kutalik, J. Luan, J.C. Randall, A. Scherag, S. Vedantam, A.R. Wood, J. Chen, R. Fehrmann, J. Karjalainen, B. Kahali, C.T. Liu, E.M. Schmidt, D. Absher, N. Amin, D. Anderson, M. Beekman, J.L. Bragg-Gresham, S. Buyske, A. Demirkan, G.B. Ehret, M.F. Feitosa, A. Goel, A.U. Jackson, T. Johnson, M.E. Kleber, K. Kristiansson, M. Mangino, I. Mateo Leach, C. Medina-Gomez, C.D. Palmer, D. Pasko, S. Pechlivanis, M.J. Peters, I. Prokopenko, A. Stančáková, Y. Ju Sung, T. Tanaka, A. Teumer, J.V. Van Vliet-Ostaptchouk, L. Yengo, W. Zhang, E. Albrecht, J. Ärnlöv, G.M. Arscott, S. Bandinelli, A. Barrett, C. Bellis, A.J. Bennett, C. Berne, M. Blüher, S. Böhringer, F. Bonnet, Y. Böttcher, M. Bruinenberg, D.B. Carba, I.H. Caspersen, R. Clarke, E.W. Daw, J.

Deelen, E. Deelman, G. Delgado, A.S. Doney, N. Eklund, M.R. Erdos, K. Estrada, E. Eury, N. Friedrich, M.E. Garcia, V. Giedraitis, B. Gigante, A.S. Go, A. Golay, H. Grallert, T.B. Grammer, J. Gräßler, J. Grewal, C.J. Groves, T. Haller, G. Hallmans, C.A. Hartman, M. Hassinen, C. Hayward, K. Heikkilä, K.H. Herzig, Q. Helmer, H.L. Hillege, O. Holmen, S.C. Hunt, A. Isaacs, T. Ittermann, A.L. James, I. Johansson, T. Juliusdottir, I.P. Kalafati, L. Kinnunen, W. Koenig, I.K. Kooner, W. Kratzer, C. Lamina, K. Leander, N.R. Lee, P. Lichtner, L. Lind, J. Lindström, S. Lobbens, M. Lorentzon, F. Mach, P.K. Magnusson, A. Mahajan, W.L. McArdle, C. Menni, S. Merger, E. Mihailov, L. Milani, R. Mills, A. Moayyeri, K.L. Monda, S.P. Mooijaart, T.W. Mühleisen, A. Mulas, G. Müller, M. Müller-Nurasyid, R. Nagaraja, M.A. Nalls, N. Narisu, N. Glorioso, I.M. Nolte, M. Olden, N.W. Rayner, F. Renstrom, J.S. Ried, N.R. Robertson, L.M. Rose, S. Sanna, H. Scharnagl, S. Scholtens, B. Sennblad, T. Seufferlein, C.M. Sittani, A. Vernon Smith, K. Stirrups, H.M. Stringham, J. Sundström, M.A. Swertz, A.J. Swift, A.C. Syvänen, B.O. Tayo, B. Thorand, G. Thorleifsson, A. Tomaschitz, C. Troffa, F.V. van Oort, N. Verweij, J.M. Vonk, L.L. Waite, R. Wennauer, T. Wilsgaard, M.K. Wojczynski, A. Wong, Q. Zhang, J. Hua Zhao, E.P. Brennan, M. Choi, P. Eriksson, L. Folkersen, A. Franco-Cereceda, A.G. Gharavi, Å. Hedman, M.F. Hivert, J. Huang, S. Kanoni, F. Karpe, S. Keildson, K. Kiryluk, L. Liang, R.P. Lifton, B. Ma, A.J. McKnight, R. McPherson, A. Metspalu, J.L. Min, M.F. Moffatt, G.W. Montgomery, J.M. Murabito, G. Nicholson, D.R. Nyholt, C. Olsson, J.R. Perry, E. Reinmaa, R.M. Salem, N. Sandholm, E.E. Schadt, R.A. Scott, L. Stolk, E.E. Vallejo, H.J. Westra, K.T. Zondervan, P. Amouyel, D. Arveiler, S.J. Bakker, J. Beilby, R.N. Bergman, J. Blangero, M.J. Brown, M. Burnier, H. Campbell, A. Chakravarti, P.S. Chines, S. Claudi-Boehm, F.S. Collins, D.C. Crawford, J. Danesh, U. de Faire, E.J. de Geus, M. Dörr, R. Erbel, J.G. Eriksson, M. Farrall, E. Ferrannini, J. Ferrières, N.G. Forouhi, T. Forrester, O.H. Franco, R.T. Gansevoort, C. Gieger, V. Gudnason, C.A. Haiman, T.B. Harris, A.T. Hattersley, M. Heliövaara, A.A. Hicks, A.D. Hingorani, W. Hoffmann, A. Hofman, G. Homuth, S.E. Humphries, E. Hyppönen, T. Illig, M.R. Jarvelin, B. Johansen, P. Jousilahti, A.M. Jula, J. Kaprio, F. Kee, S.M. Keinanen-Kiukaanniemi, J.S. Kooner, C. Kooperberg, P. Kovacs, A.T. Kraja, M. Kumari, K. Kuulasmaa, J. Kuusisto, T.A. Lakka, C. Langenberg, L. Le Marchand, T. Lehtimäki, V. Lyssenko, S. Männistö, A. Marette, T.C. Matise, C.A. McKenzie, B. McKnight, A.W. Musk, S. Möhlenkamp, A.D. Morris, M. Nelis, C. Ohlsson, A.J. Oldehinkel, K.K. Ong, L.J. Palmer, B.W. Penninx, A. Peters, P.P. Pramstaller, O.T. Raitakari, T. Rankinen, D.C. Rao, T.K. Rice, P.M. Ridker, M.D. Ritchie, I. Rudan, V. Salomaa, N.J. Samani, J. Saramies, M.A. Sarzynski, P.E. Schwarz, A.R. Shuldiner, J.A. Staessen, V. Steinthorsdottir, R.P. Stolk, K. Strauch, A. Tönjes, A. Tremblay, E. Tremoli, M.C. Vohl, U. Völker, P. Vollenweider, J.F. Wilson, J.C. Witterman, L.S. Adair, M. Bochud, B.O. Boehm, S.R. Bornstein, C. Bouchard, S. Cauchi, M.J. Caulfield, J.C. Chambers, D.I. Chasman, R.S. Cooper, G. Dedoussis, L. Ferrucci, P. Froguel, H.J. Grabe, A. Hamsten, J. Hui, K. Hveem, K.H. Jöckel, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, W. März, P.B. Munroe, I. Njølstad, B.A. Oostra, C.N. Palmer, N.L. Pedersen, M. Perola, L. Pérusse, U. Peters, C. Power, T. Quertermous, R. Rauramaa, F. Rivadeneira, T.E. Saaristo, D. Saleheen, J. Sinisalo, P.E. Slagboom, H. Snieder, T.D. Spector, U. Thorsteinsdottir, M. Stumvoll, J. Tuomilehto, A.G. Uitterlinden, M. Uusitupa, P. van der Harst, G. Veronesi, M. Walker, N.J. Wareham, H. Watkins, H.E. Wichmann, G.R. Abecasis, T.L. Assimes, S.I. Berndt, M. Boehnke, I.B. Borecki, P. Deloukas, L. Franke, T.M. Frayling, L.C. Groop, D.J. Hunter, R.C. Kaplan, J.R. O'Connell, L. Qi, D. Schlessinger, D.P. Strachan, K. Stefansson, C.M. van Duijn, C.J. Willer, P.M. Visscher, J. Yang, J.N. Hirschhorn, M.C. Zillikens, M.I. McCarthy, E.K. Speliotes, K.E. North, C.S. Fox, I. Barroso, P.W. Franks, E. Ingelsson, I.M. Heid, R.J. Loos, L.A. Cupples, A.P. Morris, C.M. Lindgren, K.L. Mohlke, A. Consortium, C.D. Consortium, C. Consortium, G. Consortium, G. Consortium, GLGC, ICBP, I.E. Consortium, L.C. Study, M. Investigators, M. Consortium, P. Consortium, R. Consortium, New genetic loci link adipose and insulin biology to body fat distribution, *Nature*, 518 (2015) 187-196.

[66] K.L. Mohlke, M. Boehnke, Recent advances in understanding the genetic architecture of type 2 diabetes, *Hum Mol Genet*, 24 (2015) R85-92.

[67] A. Mahajan, M.J. Go, W. Zhang, J.E. Below, K.J. Gaulton, T. Ferreira, M. Horikoshi, A.D. Johnson, M.C. Ng, I. Prokopenko, D. Saleheen, X. Wang, E. Zeggini, G.R. Abecasis, L.S. Adair, P. Almgren, M. Atalay, T. Aung, D. Baldassarre, B. Balkau, Y. Bao, A.H. Barnett, I. Barroso, A. Basit, L.F. Been, J. Beilby, G.I. Bell, R. Benediktsson, R.N. Bergman, B.O. Boehm, E. Boerwinkle, L.L. Bonnycastle, N. Burt, Q. Cai, H. Campbell, J. Carey, S. Cauchi, M. Caulfield, J.C. Chan, L.C. Chang, T.J. Chang, Y.C. Chang, G. Charpentier, C.H. Chen, H. Chen, Y.T. Chen, K.S. Chia, M. Chidambaram, P.S. Chines, N.H. Cho, Y.M. Cho, L.M. Chuang, F.S. Collins, M.C. Cornelis, D.J. Couper, A.T. Crenshaw, R.M. van Dam, J. Danesh, D.

Das, U. de Faire, G. Dedoussis, P. Deloukas, A.S. Dimas, C. Dina, A.S. Doney, P.J. Donnelly, M. Dorkhan, C. van Duijn, J. Dupuis, S. Edkins, P. Elliott, V. Emilsson, R. Erbel, J.G. Eriksson, J. Escobedo, T. Esko, E. Eury, J.C. Florez, P. Fontanillas, N.G. Forouhi, T. Forsen, C. Fox, R.M. Fraser, T.M. Frayling, P. Froguel, P. Frossard, Y. Gao, K. Gertow, C. Gieger, B. Gigante, H. Grallert, G.B. Grant, L.C. Grrop, C.J. Groves, E. Grundberg, C. Guiducci, A. Hamsten, B.G. Han, K. Hara, N. Hassanali, A.T. Hattersley, C. Hayward, A.K. Hedman, C. Herder, A. Hofman, O.L. Holmen, K. Hovingh, A.B. Hreidarsson, C. Hu, F.B. Hu, J. Hui, S.E. Humphries, S.E. Hunt, D.J. Hunter, K. Hveem, Z.I. Hydrie, H. Ikegami, T. Illig, E. Ingelsson, M. Islam, B. Isomaa, A.U. Jackson, T. Jafar, A. James, W. Jia, K.H. Jöckel, A. Jonsson, J.B. Jowett, T. Kadowaki, H.M. Kang, S. Kanoni, W.H. Kao, S. Kathiresan, N. Kato, P. Katulanda, K.M. Keinanen-Kiukaanniemi, A.M. Kelly, H. Khan, K.T. Khaw, C.C. Khor, H.L. Kim, S. Kim, Y.J. Kim, L. Kinnunen, N. Klopp, A. Kong, E. Korpi-Hyövälti, S. Kowlessur, P. Kraft, J. Kravic, M.M. Kristensen, S. Krithika, A. Kumar, J. Kumate, J. Kuusisto, S.H. Kwak, M. Laakso, V. Lagou, T.A. Lakka, C. Langenberg, C. Langford, R. Lawrence, K. Leander, J.M. Lee, N.R. Lee, M. Li, X. Li, Y. Li, J. Liang, S. Liju, W.Y. Lim, L. Lind, C.M. Lindgren, E. Lindholm, C.T. Liu, J.J. Liu, S. Lobbens, J. Long, R.J. Loos, W. Lu, J. Luan, V. Lyssenko, R.C. Ma, S. Maeda, R. Mägi, S. Männistö, D.R. Matthews, J.B. Meigs, O. Melander, A. Metspalu, J. Meyer, G. Mirza, E. Mihailov, S. Moebus, V. Mohan, K.L. Mohlke, A.D. Morris, T.W. Mühleisen, M. Müller-Nurasyid, B. Musk, J. Nakamura, E. Nakashima, P. Navarro, P.K. Ng, A.C. Nica, P.M. Nilsson, I. Njølstad, M.M. Nöthen, K. Ohnaka, T.H. Ong, K.R. Owen, C.N. Palmer, J.S. Pankow, K.S. Park, M. Parkin, S. Pechlivanis, N.L. Pedersen, L. Peltonen, J.R. Perry, A. Peters, J.M. Pinidiyathirige, C.G. Platou, S. Potter, J.F. Price, L. Qi, V. Radha, L. Rallidis, A. Rasheed, W. Rathman, R. Rauramaa, S. Raychaudhuri, N.W. Rayner, S.D. Rees, E. Rehnberg, S. Ripatti, N. Robertson, M. Roden, E.J. Rossin, I. Rudan, D. Rybin, T.E. Saaristo, V. Salomaa, J. Saltevo, M. Samuel, D.K. Sanghera, J. Saramies, J. Scott, L.J. Scott, R.A. Scott, A.V. Segrè, J. Sehmi, B. Sennblad, N. Shah, S. Shah, A.S. Shera, X.O. Shu, A.R. Shuldiner, G. Sigurdsson, E. Sijbrands, A. Silveira, X. Sim, S. Sivapalaratnam, K.S. Small, W.Y. So, A. Stančáková, K. Stefansson, G. Steinbach, V. Steinthorsdóttir, K. Stirrups, R.J. Strawbridge, H.M. Stringham, Q. Sun, C. Suo, A.C. Syvänen, R. Takayanagi, F. Takeuchi, W.T. Tay, T.M. Teslovich, B. Thorand, G. Thorleifsson, U. Thorsteinsdóttir, E. Tikkanen, J. Trakalo, E. Tremoli, M.D. Trip, F.J. Tsai, T. Tuomi, J. Tuomilehto, A.G. Uitterlinden, A. Valladares-Salgado, S. Vedantam, F. Veglia, B.F. Voight, C. Wang, N.J. Wareham, R. Wennauer, A.R. Wickremasinghe, T. Wilsgaard, J.F. Wilson, S. Wiltshire, W. Winckler, T.Y. Wong, A.R. Wood, J.Y. Wu, Y. Wu, K. Yamamoto, T. Yamauchi, M. Yang, L. Yengo, M. Yokota, R. Young, D. Zabaneh, F. Zhang, R. Zhang, W. Zheng, P.Z. Zimmet, D. Altshuler, D.W. Bowden, Y.S. Cho, N.J. Cox, M. Cruz, C.L. Hanis, J. Kooner, J.Y. Lee, M. Seielstad, Y.Y. Teo, M. Boehnke, E.J. Parra, J.C. Chambers, E.S. Tai, M.I. McCarthy, A.P. Morris, D.G.R.A.M.-a.D. Consortium, A.G.E.N.T.D.A.-T.D. Consortium, S.A.T.D.S.D. Consortium, M.A.T.D.M.D. Consortium, T.D.G.E.b.N.-g.s.i.m.-E.S.T.D.-G. Consortium, Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility, *Nat Genet*, 46 (2014) 234-244.

[68] P.M. Visscher, M.A. Brown, M.I. McCarthy, J. Yang, Five years of GWAS discovery, *Am J Hum Genet*, 90 (2012) 7-24.

[69] K. Musunuru, A. Strong, M. Frank-Kamenetsky, N.E. Lee, T. Ahfeldt, K.V. Sachs, X. Li, H. Li, N. Kuperwasser, V.M. Ruda, J.P. Pirruccello, B. Muchmore, L. Prokunina-Olsson, J.L. Hall, E.E. Schadt, C.R. Morales, S. Lund-Katz, M.C. Phillips, J. Wong, W. Cantley, T. Racie, K.G. Ejebe, M. Orho-Melander, O. Melander, V. Koteliansky, K. Fitzgerald, R.M. Krauss, C.A. Cowan, S. Kathiresan, D.J. Rader, From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus, *Nature*, 466 (2010) 714-719.

[70] C.J. Willer, S. Sanna, A.U. Jackson, A. Scuteri, L.L. Bonnycastle, R. Clarke, S.C. Heath, N.J. Timpson, S.S. Najjar, H.M. Stringham, J. Strait, W.L. Duren, A. Maschio, F. Busonero, A. Mulas, G. Albai, A.J. Swift, M.A. Morken, N. Narisu, D. Bennett, S. Parish, H. Shen, P. Galan, P. Meneton, S. Hercberg, D. Zelenika, W.M. Chen, Y. Li, L.J. Scott, P.A. Scheet, J. Sundvall, R.M. Watanabe, R. Nagaraja, S. Ebrahim, D.A. Lawlor, Y. Ben-Shlomo, G. Davey-Smith, A.R. Shuldiner, R. Collins, R.N. Bergman, M. Uda, J. Tuomilehto, A. Cao, F.S. Collins, E. Lakatta, G.M. Lathrop, M. Boehnke, D. Schlessinger, K.L. Mohlke, G.R. Abecasis, Newly identified loci that influence lipid concentrations and risk of coronary artery disease, *Nat Genet*, 40 (2008) 161-169.

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

[72] E.P. Consortium, An integrated encyclopedia of DNA elements in the human genome, *Nature*, 489 (2012) 57-74.

[73] A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M.J. Ziller, V. Amin, J.W. Whitaker, M.D. Schultz, L.D. Ward, A. Sarkar, G. Quon, R.S. Sandstrom, M.L. Eaton, Y.C. Wu, A.R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R.A. Harris, N. Shores, C.B. Epstein, E. Gjoneska, D. Leung, W. Xie, R.D. Hawkins, R. Lister, C. Hong, P. Gascard, A.J. Mungall, R. Moore, E. Chuah, A. Tam, T.K. Canfield, R.S. Hansen, R. Kaul, P.J. Sabo, M.S. Bansal, A. Carles, J.R. Dixon, K.H. Farh, S. Feizi, R. Karlic, A.R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T.R. Mercer, S.J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R.C. Sallari, K.T. Siebenthal, N.A. Sinnott-Armstrong, M. Stevens, R.E. Thurman, J. Wu, B. Zhang, X. Zhou, A.E. Beaudet, L.A. Boyer, P.L. De Jager, P.J. Farnham, S.J. Fisher, D. Haussler, S.J. Jones, W. Li, M.A. Marra, M.T. McManus, S. Sunyaev, J.A. Thomson, T.D. Tlsty, L.H. Tsai, W. Wang, R.A. Waterland, M.Q. Zhang, L.H. Chadwick, B.E. Bernstein, J.F. Costello, J.R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J.A. Stamatoyannopoulos, T. Wang, M. Kellis, R.E. Consortium, Integrative analysis of 111 reference human epigenomes, *Nature*, 518 (2015) 317-330.

[74] L. Song, G.E. Crawford, DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells, *Cold Spring Harb Protoc*, 2010 (2010) pdb.prot5384.

[75] P.G. Giresi, J. Kim, R.M. McDaniell, V.R. Iyer, J.D. Lieb, FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin, *Genome Res*, 17 (2007) 877-885.

[76] T.K. Kim, R. Shiekhattar, Architectural and Functional Commonalities between Enhancers and Promoters, *Cell*, 162 (2015) 948-959.

[77] M.P. Creighton, A.W. Cheng, G.G. Welstead, T. Kooistra, B.W. Carey, E.J. Steine, J. Hanna, M.A. Lodato, G.M. Frampton, P.A. Sharp, L.A. Boyer, R.A. Young, R. Jaenisch, Histone H3K27ac separates active from poised enhancers and predicts developmental state, *Proc Natl Acad Sci U S A*, 107 (2010) 21931-21936.

[78] G. Robertson, M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O.L. Griffith, A. He, M. Marra, M. Snyder, S. Jones, Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing, *Nat Methods*, 4 (2007) 651-657.

[79] K.J. Gaulton, T. Nammo, L. Pasquali, J.M. Simon, P.G. Giresi, M.P. Fogarty, T.M. Panhuis, P. Mieczkowski, A. Secchi, D. Bosco, T. Berney, E. Montanya, K.L. Mohlke, J.D. Lieb, J. Ferrer, A map of open chromatin in human pancreatic islets, *Nat Genet*, 42 (2010) 255-259.

[80] M.L. Stitzel, P. Sethupathy, D.S. Pearson, P.S. Chines, L. Song, M.R. Erdos, R. Welch, S.C. Parker, A.P. Boyle, L.J. Scott, E.H. Margulies, M. Boehnke, T.S. Furey, G.E. Crawford, F.S. Collins, N.C.S. Program, Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci, *Cell Metab*, 12 (2010) 443-455.

[81] D.S. Paul, J.P. Nisbet, T.P. Yang, S. Meacham, A. Rendon, K. Hautaviita, J. Tallila, J. White, M.R. Tijssen, S. Sivapalaratnam, H. Basart, M.D. Trip, B. Göttgens, N. Soranzo, W.H. Ouwehand, P. Deloukas, C. Consortium, M. Consortium, Maps of open chromatin guide the functional follow-up of genome-wide association signals: application to hematological traits, *PLoS Genet*, 7 (2011) e1002139.

- [82] M.P. Fogarty, T.M. Panhuis, S. Vadlamudi, M.L. Buchkovich, K.L. Mohlke, Allele-Specific Transcriptional Activity at Type 2 Diabetes-Associated Single Nucleotide Polymorphisms in Regions of Pancreatic Islet Open Chromatin at the JAZF1 Locus, *Diabetes*, (2013).
- [83] J.R. Kulzer, M.L. Stitzel, M.A. Morken, J.R. Huyghe, C. Fuchsberger, J. Kuusisto, M. Laakso, M. Boehnke, F.S. Collins, K.L. Mohlke, A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell, *Am J Hum Genet*, 94 (2014) 186-197.
- [84] M.P. Fogarty, M.E. Cannon, S. Vadlamudi, K.J. Gaulton, K.L. Mohlke, Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus, *PLoS Genet*, 10 (2014) e1004633.
- [85] T.S. Roman, A.F. Marvelle, M.P. Fogarty, S. Vadlamudi, A.J. Gonzalez, M.L. Buchkovich, J.R. Huyghe, C. Fuchsberger, A.U. Jackson, Y. Wu, M. Civelek, A.J. Lusic, K.J. Gaulton, P. Sethupathy, A.J. Kangas, P. Soininen, M. Ala-Korpela, J. Kuusisto, F.S. Collins, M. Laakso, M. Boehnke, K.L. Mohlke, Multiple Hepatic Regulatory Variants at the GALNT2 GWAS Locus Associated with High-Density Lipoprotein Cholesterol, *Am J Hum Genet*, 97 (2015) 801-815.
- [86] J.A. Kuivenhoven, R.A. Hegele, Mining the genome for lipid genes, *Biochim Biophys Acta*, 1842 (2014) 1993-2009.
- [87] T.M. Teslovich, K. Musunuru, A.V. Smith, A.C. Edmondson, I.M. Stylianou, M. Koseki, J.P. Pirruccello, S. Ripatti, D.I. Chasman, C.J. Willer, C.T. Johansen, S.W. Fouchier, A. Isaacs, G.M. Peloso, M. Barbalic, S.L. Ricketts, J.C. Bis, Y.S. Aulchenko, G. Thorleifsson, M.F. Feitosa, J. Chambers, M. Orholm, O. Melander, T. Johnson, X. Li, X. Guo, M. Li, Y. Shin Cho, M. Jin Go, Y. Jin Kim, J.Y. Lee, T. Park, K. Kim, X. Sim, R. Twee-Hee Ong, D.C. Croteau-Chonka, L.A. Lange, J.D. Smith, K. Song, J. Hua Zhao, X. Yuan, J. Luan, C. Lamina, A. Ziegler, W. Zhang, R.Y. Zee, A.F. Wright, J.C. Witteman, J.F. Wilson, G. Willemssen, H.E. Wichmann, J.B. Whitfield, D.M. Waterworth, N.J. Wareham, G. Waeber, P. Vollenweider, B.F. Voight, V. Vitart, A.G. Uitterlinden, M. Uda, J. Tuomilehto, J.R. Thompson, T. Tanaka, I. Surakka, H.M. Stringham, T.D. Spector, N. Soranzo, J.H. Smit, J. Sinisalo, K. Silander, E.J. Sijbrands, A. Scuteri, J. Scott, D. Schlessinger, S. Sanna, V. Salomaa, J. Saharinen, C. Sabatti, A. Ruukonen, I. Rudan, L.M. Rose, R. Roberts, M. Rieder, B.M. Psaty, P.P. Pramstaller, I. Pichler, M. Perola, B.W. Penninx, N.L. Pedersen, C. Pattaro, A.N. Parker, G. Pare, B.A. Oostra, C.J. O'Donnell, M.S. Nieminen, D.A. Nickerson, G.W. Montgomery, T. Meitinger, R. McPherson, M.I. McCarthy, W. McArdle, D. Masson, N.G. Martin, F. Marroni, M. Mangino, P.K. Magnusson, G. Lucas, R. Luben, R.J. Loos, M.L. Lokki, G. Lettre, C. Langenberg, L.J. Launer, E.G. Lakatta, R. Laaksonen, K.O. Kyvik, F. Kronenberg, I.R. König, K.T. Khaw, J. Kaprio, L.M. Kaplan, A. Johansson, M.R. Jarvelin, A.C. Janssens, E. Ingelsson, W. Igl, G. Kees Hovingh, J.J. Hottenga, A. Hofman, A.A. Hicks, C. Hengstenberg, I.M. Heid, C. Hayward, A.S. Havulinna, N.D. Hastie, T.B. Harris, T. Haritunians, A.S. Hall, U. Gyllensten, C. Guiducci, L.C. Groop, E. Gonzalez, C. Gieger, N.B. Freimer, L. Ferrucci, J. Erdmann, P. Elliott, K.G. Ejebe, A. Döring, A.F. Dominiczak, S. Demissie, P. Deloukas, E.J. de Geus, U. de Faire, G. Crawford, F.S. Collins, Y.D. Chen, M.J. Caulfield, H. Campbell, N.P. Burt, L.L. Bonnycastle, D.I. Boomsma, S.M. Boekholdt, R.N. Bergman, I. Barroso, S. Bandinelli, C.M. Ballantyne, T.L. Assimes, T. Quertermous, D. Altshuler, M. Seielstad, T.Y. Wong, E.S. Tai, A.B. Feranil, C.W. Kuzawa, L.S. Adair, H.A. Taylor, I.B. Borecki, S.B. Gabriel, J.G. Wilson, H. Holm, U. Thorsteinsdottir, V. Gudnason, R.M. Krauss, K.L. Mohlke, J.M. Ordovas, P.B. Munroe, J.S. Kooner, A.R. Tall, R.A. Hegele, J.J. Kastelein, E.E. Schadt, J.I. Rotter, E. Boerwinkle, D.P. Strachan, V. Mooser, K. Stefansson, M.P. Reilly, N.J. Samani, H. Schunkert, L.A. Cupples, M.S. Sandhu, P.M. Ridker, D.J. Rader, C.M. van Duijn, L. Peltonen, G.R. Abecasis, M. Boehnke, S. Kathiresan, Biological, clinical and population relevance of 95 loci for blood lipids, *Nature*, 466 (2010) 707-713.
- [88] L. Dumitrescu, C.L. Carty, K. Taylor, F.R. Schumacher, L.A. Hindorf, J.L. Ambite, G. Anderson, L.G. Best, K. Brown-Gentry, P. Bůžková, C.S. Carlson, B. Cochran, S.A. Cole, R.B. Devereux, D. Duggan, C.B. Eaton, M. Fornage, N. Franceschini, J. Haessler, B.V. Howard, K.C. Johnson, S. Laston, L.N. Kolonel, E.T. Lee, J.W. MacCluer, T.A. Manolio, S.A. Pendergrass, M. Quibrera, R.V. Shohet, L.R. Wilkens, C.A. Haiman, L. Le Marchand, S. Buyske, C. Kooperberg, K.E. North, D.C. Crawford, Genetic determinants of

lipid traits in diverse populations from the population architecture using genomics and epidemiology (PAGE) study, *PLoS Genet*, 7 (2011) e1002138.

[89] D. Weissglas-Volkov, C.A. Aguilar-Salinas, E. Nikkola, K.A. Deere, I. Cruz-Bautista, O. Arellano-Campos, L.L. Muñoz-Hernandez, L. Gomez-Munguia, M.L. Ordoñez-Sánchez, P.M. Reddy, A.J. Lusi, N. Matikainen, M.R. Taskinen, L. Riba, R.M. Cantor, J.S. Sinsheimer, T. Tusie-Luna, P. Pajukanta, Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci, *J Med Genet*, 50 (2013) 298-308.

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

[91] K. Nakayama, T. Bayasgalan, K. Yamanaka, M. Kumada, T. Gotoh, N. Utsumi, Y. Yanagisawa, M. Okayama, E. Kajii, S. Ishibashi, S. Iwamoto, J.C.G.T. (JCOG), Large scale replication analysis of loci associated with lipid concentrations in a Japanese population, *J Med Genet*, 46 (2009) 370-374.

[92] D. Weissglas-Volkov, C.A. Aguilar-Salinas, J.S. Sinsheimer, L. Riba, A. Huertas-Vazquez, M.L. Ordoñez-Sánchez, R. Rodriguez-Guillen, R.M. Cantor, T. Tusie-Luna, P. Pajukanta, Investigation of variants identified in caucasian genome-wide association studies for plasma high-density lipoprotein cholesterol and triglycerides levels in Mexican dyslipidemic study samples, *Circ Cardiovasc Genet*, 3 (2010) 31-38.

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

[94] G.R. Abecasis, A. Auton, L.D. Brooks, M.A. DePristo, R.M. Durbin, R.E. Handsaker, H.M. Kang, G.T. Marth, G.A. McVean, G.P. Consortium, An integrated map of genetic variation from 1,092 human genomes, *Nature*, 491 (2012) 56-65.

[95] T.A. Manolio, Bringing genome-wide association findings into clinical use, *Nat Rev Genet*, 14 (2013) 549-558.

[96] K.G. Ten Hagen, T.A. Fritz, L.A. Tabak, All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases, *Glycobiology*, 13 (2003) 1R-16R.

[97] T. White, E.P. Bennett, K. Takio, T. Sørensen, N. Bonding, H. Clausen, Purification and cDNA cloning of a human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase, *J Biol Chem*, 270 (1995) 24156-24165.

[98] E.P. Bennett, H. Hassan, H. Clausen, cDNA cloning and expression of a novel human UDP-N-acetyl-alpha-D-galactosamine. Polypeptide N-acetylgalactosaminyltransferase, GalNAc-t3, *J Biol Chem*, 271 (1996) 17006-17012.

- [99] E.P. Bennett, U. Mandel, H. Clausen, T.A. Gerken, T.A. Fritz, L.A. Tabak, Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family, *Glycobiology*, 22 (2012) 736-756.
- [100] K.T. Schjoldager, M.B. Vester-Christensen, E.P. Bennett, S.B. Levery, T. Schwientek, W. Yin, O. Blixt, H. Clausen, O-glycosylation modulates proprotein convertase activation of angiotensin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids, *J Biol Chem*, 285 (2010) 36293-36303.
- [101] A.G. Holleboom, H. Karlsson, R.S. Lin, T.M. Beres, J.A. Sierts, D.S. Herman, E.S. Stroes, J.M. Aerts, J.J. Kastelein, M.M. Motazacker, G.M. Dallinga-Thie, J.H. Levels, A.H. Zwinderman, J.G. Seidman, C.E. Seidman, S. Ljunggren, D.J. Lefeber, E. Morava, R.A. Wevers, T.A. Fritz, L.A. Tabak, M. Lindahl, G.K. Hovingh, J.A. Kuivenhoven, Heterozygosity for a loss-of-function mutation in GALNT2 improves plasma triglyceride clearance in man, *Cell Metab*, 14 (2011) 811-818.
- [102] A. Stancáková, M. Javorský, T. Kuulasmaa, S.M. Haffner, J. Kuusisto, M. Laakso, Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men, *Diabetes*, 58 (2009) 1212-1221.
- [103] J. Ernst, P. Kheradpour, T.S. Mikkelsen, N. Shores, L.D. Ward, C.B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis, B.E. Bernstein, Mapping and analysis of chromatin state dynamics in nine human cell types, *Nature*, 473 (2011) 43-49.
- [104] J.R. Huyghe, A.U. Jackson, M.P. Fogarty, M.L. Buchkovich, A. Stančáková, H.M. Stringham, X. Sim, L. Yang, C. Fuchsberger, H. Cederberg, P.S. Chines, T.M. Teslovich, J.M. Romm, H. Ling, I. McMullen, R. Ingersoll, E.W. Pugh, K.F. Doheny, B.M. Neale, M.J. Daly, J. Kuusisto, L.J. Scott, H.M. Kang, F.S. Collins, G.R. Abecasis, R.M. Watanabe, M. Boehnke, M. Laakso, K.L. Mohlke, Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion, *Nat Genet*, 45 (2013) 197-201.
- [105] G. Jun, M. Flickinger, K.N. Hetrick, J.M. Romm, K.F. Doheny, G.R. Abecasis, M. Boehnke, H.M. Kang, Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data, *Am J Hum Genet*, 91 (2012) 839-848.
- [106] B. Howie, C. Fuchsberger, M. Stephens, J. Marchini, G.R. Abecasis, Fast and accurate genotype imputation in genome-wide association studies through pre-phasing, *Nat Genet*, 44 (2012) 955-959.
- [107] P. Soininen, A.J. Kangas, P. Würtz, T. Tukiainen, T. Tynkkynen, R. Laatikainen, M.R. Järvelin, M. Kähönen, T. Lehtimäki, J. Viikari, O.T. Raitakari, M.J. Savolainen, M. Ala-Korpela, High-throughput serum NMR metabolomics for cost-effective holistic studies on systemic metabolism, *Analyst*, 134 (2009) 1781-1785.
- [108] P. Soininen, A.J. Kangas, P. Würtz, T. Suna, M. Ala-Korpela, Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics, *Circ Cardiovasc Genet*, 8 (2015) 192-206.
- [109] H.M. Kang, J.H. Sul, S.K. Service, N.A. Zaitlen, S.Y. Kong, N.B. Freimer, C. Sabatti, E. Eskin, Variance component model to account for sample structure in genome-wide association studies, *Nat Genet*, 42 (2010) 348-354.
- [110] R.J. Pruim, R.P. Welch, S. Sanna, T.M. Teslovich, P.S. Chines, T.P. Gliedt, M. Boehnke, G.R. Abecasis, C.J. Willer, LocusZoom: regional visualization of genome-wide association scan results, *Bioinformatics*, 26 (2010) 2336-2337.

- [111] A. Sandelin, W. Alkema, P. Engström, W.W. Wasserman, B. Lenhard, JASPAR: an open-access database for eukaryotic transcription factor binding profiles, *Nucleic Acids Res*, 32 (2004) D91-94.
- [112] S. Levy, S. Hannenhalli, Identification of transcription factor binding sites in the human genome sequence, *Mamm Genome*, 13 (2002) 510-514.
- [113] M.L. Buchkovich, K. Eklund, Q. Duan, Y. Li, K.L. Mohlke, T.S. Furey, Removing reference mapping biases using limited or no genotype data identifies allelic differences in protein binding at disease-associated loci, *BMC Med Genomics*, 8 (2015) 43.
- [114] T.D. Wu, S. Nacu, Fast and SNP-tolerant detection of complex variants and splicing in short reads, *Bioinformatics*, 26 (2010) 873-881.
- [115] M.P. Fogarty, R. Xiao, L. Prokunina-Olsson, L.J. Scott, K.L. Mohlke, Allelic expression imbalance at high-density lipoprotein cholesterol locus MMAB-MVK, *Hum Mol Genet*, 19 (2010) 1921-1929.
- [116] R. Xiao, L.J. Scott, Detection of cis-acting regulatory SNPs using allelic expression data, *Genet Epidemiol*, 35 (2011) 515-525.
- [117] O. Stegle, L. Parts, R. Durbin, J. Winn, A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies, *PLoS Comput Biol*, 6 (2010) e1000770.
- [118] T. Bailey, P. Krajewski, I. Ladunga, C. Lefebvre, Q. Li, T. Liu, P. Madrigal, C. Taslim, J. Zhang, Practical guidelines for the comprehensive analysis of ChIP-seq data, *PLoS Comput Biol*, 9 (2013) e1003326.
- [119] D.J. Verlaan, B. Ge, E. Grundberg, R. Hoberman, K.C. Lam, V. Koka, J. Dias, S. Gurd, N.W. Martin, H. Mallmin, O. Nilsson, E. Harmsen, K. Dewar, T. Kwan, T. Pastinen, Targeted screening of cis-regulatory variation in human haplotypes, *Genome Res*, 19 (2009) 118-127.
- [120] H.Q. Qu, D.J. Verlaan, B. Ge, Y. Lu, K.C. Lam, R. Grabs, E. Harmsen, T.J. Hudson, H. Hakonarson, T. Pastinen, C. Polychronakos, A cis-acting regulatory variant in the IL2RA locus, *J Immunol*, 183 (2009) 5158-5162.
- [121] K.K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W.J. Housley, S. Beik, N. Shores, H. Whitton, R.J. Ryan, A.A. Shishkin, M. Hatan, M.J. Carrasco-Alfonso, D. Mayer, C.J. Luckey, N.A. Patsopoulos, P.L. De Jager, V.K. Kuchroo, C.B. Epstein, M.J. Daly, D.A. Hafler, B.E. Bernstein, Genetic and epigenetic fine mapping of causal autoimmune disease variants, *Nature*, 518 (2015) 337-343.
- [122] L.J. Core, A.L. Martins, C.G. Danko, C.T. Waters, A. Siepel, J.T. Lis, Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers, *Nat Genet*, 46 (2014) 1311-1320.
- [123] T.X. Cui, G. Lin, C.R. LaPensee, A.A. Calinescu, M. Rathore, C. Streeter, G. Piwien-Pilipuk, N. Lanning, H. Jin, C. Carter-Su, Z.S. Qin, J. Schwartz, C/EBP β mediates growth hormone-regulated expression of multiple target genes, *Mol Endocrinol*, 25 (2011) 681-693.
- [124] T. Tanaka, N. Yoshida, T. Kishimoto, S. Akira, Defective adipocyte differentiation in mice lacking the C/EBP β and/or C/EBP δ gene, *EMBO J*, 16 (1997) 7432-7443.
- [125] E. Sterneck, L. Tessarollo, P.F. Johnson, An essential role for C/EBP β in female reproduction, *Genes Dev*, 11 (1997) 2153-2162.

- [126] G.W. Robinson, P.F. Johnson, L. Hennighausen, E. Sterneck, The C/EBP β transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland, *Genes Dev*, 12 (1998) 1907-1916.
- [127] R. Huber, D. Pietsch, T. Panterodt, K. Brand, Regulation of C/EBP β and resulting functions in cells of the monocytic lineage, *Cell Signal*, 24 (2012) 1287-1296.
- [128] S.M. Rahman, M. Choudhury, R.C. Janssen, K.C. Baquero, M. Miyazaki, J.E. Friedman, CCAAT/enhancer binding protein β deletion increases mitochondrial function and protects mice from LXR-induced hepatic steatosis, *Biochem Biophys Res Commun*, 430 (2013) 336-339.
- [129] G.J. Botma, A.J. Verhoeven, H. Jansen, Hepatic lipase promoter activity is reduced by the C-480T and G-216A substitutions present in the common LIPC gene variant, and is increased by Upstream Stimulatory Factor, *Atherosclerosis*, 154 (2001) 625-632.
- [130] C. Weigert, K. Brodbeck, M. Sawadogo, H.U. Häring, E.D. Schleicher, Upstream stimulatory factor (USF) proteins induce human TGF- β 1 gene activation via the glucose-response element-1013/-1002 in mesangial cells: up-regulation of USF activity by the hexosamine biosynthetic pathway, *J Biol Chem*, 279 (2004) 15908-15915.
- [131] J.W. Kim, H. Monila, A. Pandey, M.D. Lane, Upstream stimulatory factors regulate the C/EBP alpha gene during differentiation of 3T3-L1 preadipocytes, *Biochem Biophys Res Commun*, 354 (2007) 517-521.
- [132] S. Wu, R. Mar-Heyming, E.Z. Dugum, N.A. Kolaitis, H. Qi, P. Pajukanta, L.W. Castellani, A.J. Lusis, T.A. Drake, Upstream transcription factor 1 influences plasma lipid and metabolic traits in mice, *Hum Mol Genet*, 19 (2010) 597-608.
- [133] E.E. Schadt, C. Molony, E. Chudin, K. Hao, X. Yang, P.Y. Lum, A. Kasarskis, B. Zhang, S. Wang, C. Suver, J. Zhu, J. Millstein, S. Sieberts, J. Lamb, D. GuhaThakurta, J. Derry, J.D. Storey, I. Avila-Campillo, M.J. Kruger, J.M. Johnson, C.A. Rohl, A. van Nas, M. Mehrabian, T.A. Drake, A.J. Lusis, R.C. Smith, F.P. Guengerich, S.C. Strom, E. Schuetz, T.H. Rushmore, R. Ulrich, Mapping the genetic architecture of gene expression in human liver, *PLoS Biol*, 6 (2008) e107.
- [134] L. Folkersen, F. van't Hooft, E. Chernogubova, H.E. Agardh, G.K. Hansson, U. Hedin, J. Liska, A.C. Syvänen, G. Paulsson-Berne, G. Paulsson-Berne, A. Franco-Cereceda, A. Hamsten, A. Gabrielsen, P. Eriksson, B.a.A.s. groups, Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease, *Circ Cardiovasc Genet*, 3 (2010) 365-373.
- [135] G. Consortium, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans, *Science*, 348 (2015) 648-660.
- [136] J. Kettunen, T. Tukiainen, A.P. Sarin, A. Ortega-Alonso, E. Tikkanen, L.P. Lyytikäinen, A.J. Kangas, P. Soininen, P. Würtz, K. Silander, D.M. Dick, R.J. Rose, M.J. Savolainen, J. Viikari, M. Kähönen, T. Lehtimäki, K.H. Pietiläinen, M. Inouye, M.I. McCarthy, A. Jula, J. Eriksson, O.T. Raitakari, V. Salomaa, J. Kaprio, M.R. Jarvelin, L. Peltonen, M. Perola, N.B. Freimer, M. Ala-Korpela, A. Palotie, S. Ripatti, Genome-wide association study identifies multiple loci influencing human serum metabolite levels, *Nat Genet*, 44 (2012) 269-276.
- [137] M. Jaye, K.J. Lynch, J. Krawiec, D. Marchadier, C. Maugeais, K. Doan, V. South, D. Amin, M. Perrone, D.J. Rader, A novel endothelial-derived lipase that modulates HDL metabolism, *Nat Genet*, 21 (1999) 424-428.
- [138] M.G. McCoy, G.S. Sun, D. Marchadier, C. Maugeais, J.M. Glick, D.J. Rader, Characterization of the lipolytic activity of endothelial lipase, *J Lipid Res*, 43 (2002) 921-929.

[139] O. Corradin, A. Saiakhova, B. Akhtar-Zaidi, L. Myeroff, J. Willis, R. Cowper-Salari, M. Lupien, S. Markowitz, P.C. Scacheri, Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits, *Genome Res*, 24 (2014) 1-13.

[140] H. He, W. Li, S. Liyanarachchi, M. Srinivas, Y. Wang, K. Akagi, D. Wu, Q. Wang, V. Jin, D.E. Symer, R. Shen, J. Phay, R. Nagy, A. de la Chapelle, Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer, *Proc Natl Acad Sci U S A*, 112 (2015) 6128-6133.

[141] A.P. Morris, B.F. Voight, T.M. Teslovich, T. Ferreira, A.V. Segrè, V. Steinthorsdottir, R.J. Strawbridge, H. Khan, H. Grallert, A. Mahajan, I. Prokopenko, H.M. Kang, C. Dina, T. Esko, R.M. Fraser, S. Kanoni, A. Kumar, V. Lagou, C. Langenberg, J. Luan, C.M. Lindgren, M. Müller-Nurasyid, S. Pechlivanis, N.W. Rayner, L.J. Scott, S. Wiltshire, L. Yengo, L. Kinnunen, E.J. Rossin, S. Raychaudhuri, A.D. Johnson, A.S. Dimas, R.J. Loos, S. Vedantam, H. Chen, J.C. Florez, C. Fox, C.T. Liu, D. Rybin, D.J. Couper, W.H. Kao, M. Li, M.C. Cornelis, P. Kraft, Q. Sun, R.M. van Dam, H.M. Stringham, P.S. Chines, K. Fischer, P. Fontanillas, O.L. Holmen, S.E. Hunt, A.U. Jackson, A. Kong, R. Lawrence, J. Meyer, J.R. Perry, C.G. Platou, S. Potter, E. Rehnberg, N. Robertson, S. Sivapalaratnam, A. Stančáková, K. Stirrups, G. Thorleifsson, E. Tikkanen, A.R. Wood, P. Almgren, M. Atalay, R. Benediktsson, L.L. Bonnycastle, N. Burt, J. Carey, G. Charpentier, A.T. Crenshaw, A.S. Doney, M. Dorkhan, S. Edkins, V. Emilsson, E. Eury, T. Forsen, K. Gertow, B. Gigante, G.B. Grant, C.J. Groves, C. Guiducci, C. Herder, A.B. Hreidarsson, J. Hui, A. James, A. Jonsson, W. Rathmann, N. Klopp, J. Kravic, K. Krjutškov, C. Langford, K. Leander, E. Lindholm, S. Lobbens, S. Männistö, G. Mirza, T.W. Mühleisen, B. Musk, M. Parkin, L. Rallidis, J. Saramies, B. Sennblad, S. Shah, G. Sigurdsson, A. Silveira, G. Steinbach, B. Thorand, J. Trakalo, F. Veglia, R. Wennauer, W. Winckler, D. Zabaneh, H. Campbell, C. van Duijn, A.G. Uitterlinden, A. Hofman, E. Sijbrands, G.R. Abecasis, K.R. Owen, E. Zeggini, M.D. Trip, N.G. Forouhi, A.C. Syvänen, J.G. Eriksson, L. Peltonen, M.M. Nöthen, B. Balkau, C.N. Palmer, V. Lyssenko, T. Tuomi, B. Isomaa, D.J. Hunter, L. Qi, A.R. Shuldiner, M. Roden, I. Barroso, T. Wilsgaard, J. Beilby, K. Hovingh, J.F. Price, J.F. Wilson, R. Rauramaa, T.A. Lakka, L. Lind, G. Dedoussis, I. Njølstad, N.L. Pedersen, K.T. Khaw, N.J. Wareham, S.M. Keinanen-Kiukkaanniemi, T.E. Saaristo, E. Korpi-Hyövälti, J. Saltevo, M. Laakso, J. Kuusisto, A. Metspalu, F.S. Collins, K.L. Mohlke, R.N. Bergman, J. Tuomilehto, B.O. Boehm, C. Gieger, K. Hveem, S. Cauchi, P. Froguel, D. Baldassarre, E. Tremoli, S.E. Humphries, D. Saleheen, J. Danesh, E. Ingelsson, S. Ripatti, V. Salomaa, R. Erbel, K.H. Jöckel, S. Moebus, A. Peters, T. Illig, U. de Faire, A. Hamsten, A.D. Morris, P.J. Donnelly, T.M. Frayling, A.T. Hattersley, E. Boerwinkle, O. Melander, S. Kathiresan, P.M. Nilsson, P. Deloukas, U. Thorsteinsdottir, L.C. Groop, K. Stefansson, F. Hu, J.S. Pankow, J. Dupuis, J.B. Meigs, D. Altshuler, M. Boehnke, M.I. McCarthy, W.T.C.C. Consortium, M.-A.o.G.a.l.-r.t.C.M. Investigators, G.I.o.A.T.G. Consortium, A.G.E.N.T.D.A.-T.D. Consortium, S.A.T.D.S.D. Consortium, D.G.R.A.M.-a.D. Consortium, Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes, *Nat Genet*, 44 (2012) 981-990.

[142] S.D. Rees, M.Z. Hydrie, J.P. O'Hare, S. Kumar, A.S. Shera, A. Basit, A.H. Barnett, M.A. Kelly, Effects of 16 genetic variants on fasting glucose and type 2 diabetes in South Asians: ADCY5 and GLIS3 variants may predispose to type 2 diabetes, *PLoS One*, 6 (2011) e24710.

[143] M.C. Ng, R. Saxena, J. Li, N.D. Palmer, L. Dimitrov, J. Xu, L.J. Rasmussen-Torvik, J.M. Zmuda, D.S. Siscovick, S.R. Patel, E.D. Crook, M. Sims, Y.D. Chen, A.G. Bertoni, M. Li, S.F. Grant, J. Dupuis, J.B. Meigs, B.M. Psaty, J.S. Pankow, C.D. Langefeld, B.I. Freedman, J.I. Rotter, J.G. Wilson, D.W. Bowden, Transferability and fine mapping of type 2 diabetes loci in African Americans: the Candidate Gene Association Resource Plus Study, *Diabetes*, 62 (2013) 965-976.

[144] L.D. Ward, M. Kellis, HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants, *Nucleic Acids Res*, 40 (2012) D930-934.

[145] A.A. Adeyemo, F. Tekola-Ayele, A.P. Doumatey, A.R. Bentley, G. Chen, H. Huang, J. Zhou, D. Shriner, O. Fasanmade, G. Okafor, B. Eghan, K. Agyenim-Boateng, J. Adeleye, W. Balogun, A. Elkahloun, S. Chandrasekharappa, S. Owusu, A. Amoah, J. Acheampong, T. Johnson, J. Oli, C.

Adebamowo, F. Collins, G. Dunston, C.N. Rotimi, Evaluation of Genome Wide Association Study Associated Type 2 Diabetes Susceptibility Loci in Sub Saharan Africans, *Front Genet*, 6 (2015) 335.

[146] J. Dupuis, C. Langenberg, I. Prokopenko, R. Saxena, N. Soranzo, A.U. Jackson, E. Wheeler, N.L. Glazer, N. Bouatia-Naji, A.L. Gloyn, C.M. Lindgren, R. Mägi, A.P. Morris, J. Randall, T. Johnson, P. Elliott, D. Rybin, G. Thorleifsson, V. Steinthorsdottir, P. Henneman, H. Grallert, A. Dehghan, J.J. Hottenga, C.S. Franklin, P. Navarro, K. Song, A. Goel, J.R. Perry, J.M. Egan, T. Lajunen, N. Grarup, T. Sparsø, A. Doney, B.F. Voight, H.M. Stringham, M. Li, S. Kanoni, P. Shrader, C. Cavalcanti-Proença, M. Kumari, L. Qi, N.J. Timpson, C. Gieger, C. Zabena, G. Rocheleau, E. Ingelsson, P. An, J. O'Connell, J. Luan, A. Elliott, S.A. McCarroll, F. Payne, R.M. Ruccasecca, F. Pattou, P. Sethupathy, K. Ardlie, Y. Ariyurek, B. Balkau, P. Barter, J.P. Beilby, Y. Ben-Shlomo, R. Benediktsson, A.J. Bennett, S. Bergmann, M. Bochud, E. Boerwinkle, A. Bonnefond, L.L. Bonnycastle, K. Borch-Johnsen, Y. Böttcher, E. Brunner, S.J. Bumpstead, G. Charpentier, Y.D. Chen, P. Chines, R. Clarke, L.J. Coin, M.N. Cooper, M. Cornelis, G. Crawford, L. Crisponi, I.N. Day, E.J. de Geus, J. Delplanque, C. Dina, M.R. Erdos, A.C. Fedson, A. Fischer-Rosinsky, N.G. Forouhi, C.S. Fox, R. Frants, M.G. Franzosi, P. Galan, M.O. Goodarzi, J. Graessler, C.J. Groves, S. Grundy, R. Gwilliam, U. Gyllensten, S. Hadjadj, G. Hallmans, N. Hammond, X. Han, A.L. Hartikainen, N. Hassanali, C. Hayward, S.C. Heath, S. Hercberg, C. Herder, A.A. Hicks, D.R. Hillman, A.D. Hingorani, A. Hofman, J. Hui, J. Hung, B. Isomaa, P.R. Johnson, T. Jørgensen, A. Jula, M. Kaakinen, J. Kaprio, Y.A. Kesaniemi, M. Kivimaki, B. Knight, S. Koskinen, P. Kovacs, K.O. Kyvik, G.M. Lathrop, D.A. Lawlor, O. Le Bacquer, C. Lecoeur, Y. Li, V. Lyssenko, R. Mahley, M. Mangino, A.K. Manning, M.T. Martinez-Larrad, J.B. McAteer, L.J. McCulloch, R. McPherson, C. Meisinger, D. Melzer, D. Meyre, B.D. Mitchell, M.A. Morcken, S. Mukherjee, S. Naitza, N. Narisu, M.J. Neville, B.A. Oostra, M. Orrù, R. Pakyz, C.N. Palmer, G. Paolisso, C. Pattaro, D. Pearson, J.F. Peden, N.L. Pedersen, M. Perola, A.F. Pfeiffer, I. Pichler, O. Polasek, D. Posthuma, S.C. Potter, A. Pouta, M.A. Province, B.M. Psaty, W. Rathmann, N.W. Rayner, K. Rice, S. Ripatti, F. Rivadeneira, M. Roden, O. Rolandsson, A. Sandbaek, M. Sandhu, S. Sanna, A.A. Sayer, P. Scheet, L.J. Scott, U. Seedorf, S.J. Sharp, B. Shields, G. Sigurdsson, E.J. Sijbrands, A. Silveira, L. Simpson, A. Singleton, N.L. Smith, U. Sovio, A. Swift, H. Syddall, A.C. Syvänen, T. Tanaka, B. Thorand, J. Tichet, A. Tönjes, T. Tuomi, A.G. Uitterlinden, K.W. van Dijk, M. van Hoek, D. Varma, S. Visvikis-Siest, V. Vitart, N. Vogelzang, G. Waeber, P.J. Wagner, A. Walley, G.B. Walters, K.L. Ward, H. Watkins, M.N. Weedon, S.H. Wild, G. Willemssen, J.C. Witteman, J.W. Yarnell, E. Zeggini, D. Zelenika, B. Zethelius, G. Zhai, J.H. Zhao, M.C. Zillikens, I.B. Borecki, R.J. Loos, P. Meneton, P.K. Magnusson, D.M. Nathan, G.H. Williams, A.T. Hattersley, K. Silander, V. Salomaa, G.D. Smith, S.R. Bornstein, P. Schwarz, J. Spranger, F. Karpe, A.R. Shuldiner, C. Cooper, G.V. Dedoussis, M. Serrano-Rios, A.D. Morris, L. Lind, L.J. Palmer, F.B. Hu, P.W. Franks, S. Ebrahim, M. Marmot, W.H. Kao, J.S. Pankow, M.J. Sampson, J. Kuusisto, M. Laakso, T. Hansen, O. Pedersen, P.P. Pramstaller, H.E. Wichmann, T. Illig, I. Rudan, A.F. Wright, M. Stumvoll, H. Campbell, J.F. Wilson, R.N. Bergman, T.A. Buchanan, F.S. Collins, K.L. Mohlke, J. Tuomilehto, T.T. Valle, D. Altshuler, J.I. Rotter, D.S. Siscovick, B.W. Penninx, D.I. Boomsma, P. Deloukas, T.D. Spector, T.M. Frayling, L. Ferrucci, A. Kong, U. Thorsteinsdottir, K. Stefansson, C.M. van Duijn, Y.S. Aulchenko, A. Cao, A. Scuteri, D. Schlessinger, M. Uda, A. Ruokonen, M.R. Jarvelin, D.M. Waterworth, P. Vollenweider, L. Peltonen, V. Mooser, G.R. Abecasis, N.J. Wareham, R. Sladek, P. Froguel, R.M. Watanabe, J.B. Meigs, L. Groop, M. Boehnke, M.I. McCarthy, J.C. Florez, I. Barroso, D. Consortium, G. Consortium, G.B. Consortium, A.H.o.b.o.P. Consortium, M. investigators, New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk, *Nat Genet*, 42 (2010) 105-116.

[147] R. Saxena, M.F. Hivert, C. Langenberg, T. Tanaka, J.S. Pankow, P. Vollenweider, V. Lyssenko, N. Bouatia-Naji, J. Dupuis, A.U. Jackson, W.H. Kao, M. Li, N.L. Glazer, A.K. Manning, J. Luan, H.M. Stringham, I. Prokopenko, T. Johnson, N. Grarup, T.W. Boesgaard, C. Lecoeur, P. Shrader, J. O'Connell, E. Ingelsson, D.J. Couper, K. Rice, K. Song, C.H. Andreassen, C. Dina, A. Köttgen, O. Le Bacquer, F. Pattou, J. Taneera, V. Steinthorsdottir, D. Rybin, K. Ardlie, M. Sampson, L. Qi, M. van Hoek, M.N. Weedon, Y.S. Aulchenko, B.F. Voight, H. Grallert, B. Balkau, R.N. Bergman, S.J. Bielinski, A. Bonnefond, L.L. Bonnycastle, K. Borch-Johnsen, Y. Böttcher, E. Brunner, T.A. Buchanan, S.J. Bumpstead, C. Cavalcanti-Proença, G. Charpentier, Y.D. Chen, P.S. Chines, F.S. Collins, M. Cornelis, G. J Crawford, J. Delplanque, A. Doney, J.M. Egan, M.R. Erdos, M. Firmann, N.G. Forouhi, C.S. Fox, M.O. Goodarzi, J. Graessler, A. Hingorani, B. Isomaa, T. Jørgensen, M. Kivimaki, P. Kovacs, K. Krohn, M. Kumari, T. Lauritzen, C. Lévy-Marchal, V. Mayor, J.B. McAteer, D. Meyre, B.D. Mitchell, K.L. Mohlke, M.A. Morcken,

N. Narisu, C.N. Palmer, R. Pakyz, L. Pascoe, F. Payne, D. Pearson, W. Rathmann, A. Sandbaek, A.A. Sayer, L.J. Scott, S.J. Sharp, E. Sijbrands, A. Singleton, D.S. Siscovick, N.L. Smith, T. Sparsø, A.J. Swift, H. Syddall, G. Thorleifsson, A. Tönjes, T. Tuomi, J. Tuomilehto, T.T. Valle, G. Waeber, A. Walley, D.M. Waterworth, E. Zeggini, J.H. Zhao, T. Illig, H.E. Wichmann, J.F. Wilson, C. van Duijn, F.B. Hu, A.D. Morris, T.M. Frayling, A.T. Hattersley, U. Thorsteinsdottir, K. Stefansson, P. Nilsson, A.C. Syvänen, A.R. Shuldiner, M. Walker, S.R. Bornstein, P. Schwarz, G.H. Williams, D.M. Nathan, J. Kuusisto, M. Laakso, C. Cooper, M. Marmot, L. Ferrucci, V. Mooser, M. Stumvoll, R.J. Loos, D. Altshuler, B.M. Psaty, J.I. Rotter, E. Boerwinkle, T. Hansen, O. Pedersen, J.C. Florez, M.I. McCarthy, M. Boehnke, I. Barroso, R. Sladek, P. Froguel, J.B. Meigs, L. Groop, N.J. Wareham, R.M. Watanabe, G. consortium, M. investigators, Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge, *Nat Genet*, 42 (2010) 142-148.

[148] S.K. Vasan, M.J. Neville, B. Antonisamy, P. Samuel, C.H. Fall, F.S. Geethanjali, N. Thomas, P. Raghupathy, K. Brismar, F. Karpe, Absence of birth-weight lowering effect of ADCY5 and near CCNL, but association of impaired glucose-insulin homeostasis with ADCY5 in Asian Indians, *PLoS One*, 6 (2011) e21331.

[149] M. Horikoshi, H. Yaghoobkar, D.O. Mook-Kanamori, U. Sovio, H.R. Taal, B.J. Hennig, J.P. Bradfield, B. St Pourcain, D.M. Evans, P. Charoen, M. Kaakinen, D.L. Cousminer, T. Lehtimäki, E. Kreiner-Møller, N.M. Warrington, M. Bustamante, B. Feenstra, D.J. Berry, E. Thiering, T. Pfab, S.J. Barton, B.M. Shields, M. Kerkhof, E.M. van Leeuwen, A.J. Fulford, Z. Kutalik, J.H. Zhao, M. den Hoed, A. Mahajan, V. Lindi, L.K. Goh, J.J. Hottenga, Y. Wu, O.T. Raitakari, M.N. Harder, A. Meirhaeghe, I. Ntalla, R.M. Salem, K.A. Jameson, K. Zhou, D.M. Monies, V. Lagou, M. Kirin, J. Heikkinen, L.S. Adair, F.S. Alkuraya, A. Al-Odaib, P. Amouyel, E.A. Andersson, A.J. Bennett, A.I. Blakemore, J.L. Buxton, J. Dallongeville, S. Das, E.J. de Geus, X. Estivill, C. Flexeder, P. Froguel, F. Geller, K.M. Godfrey, F. Gottrand, C.J. Groves, T. Hansen, J.N. Hirschhorn, A. Hofman, M.V. Hollegaard, D.M. Hougaard, E. Hyppönen, H.M. Inskip, A. Isaacs, T. Jørgensen, C. Kanaka-Gantenbein, J.P. Kemp, W. Kiess, T.O. Kilpeläinen, N. Klopp, B.A. Knight, C.W. Kuzawa, G. McMahon, J.P. Newnham, H. Niinikoski, B.A. Oostra, L. Pedersen, D.S. Postma, S.M. Ring, F. Rivadeneira, N.R. Robertson, S. Sebert, O. Simell, T. Slowinski, C.M. Tiesler, A. Tönjes, A. Vaag, J.S. Viikari, J.M. Vink, N.H. Vissing, N.J. Wareham, G. Willemsen, D.R. Witte, H. Zhang, J. Zhao, J.F. Wilson, M. Stumvoll, A.M. Prentice, B.F. Meyer, E.R. Pearson, C.A. Boreham, C. Cooper, M.W. Gillman, G.V. Dedoussis, L.A. Moreno, O. Pedersen, M. Saarinen, K.L. Mohlke, D.I. Boomsma, S.M. Saw, T.A. Lakka, A. Körner, R.J. Loos, K.K. Ong, P. Vollenweider, C.M. van Duijn, G.H. Koppelman, A.T. Hattersley, J.W. Holloway, B. Hocher, J. Heinrich, C. Power, M. Melbye, M. Guxens, C.E. Pennell, K. Bønnelykke, H. Bisgaard, J.G. Eriksson, E. Widén, H. Hakonarson, A.G. Uitterlinden, A. Pouta, D.A. Lawlor, G.D. Smith, T.M. Frayling, M.I. McCarthy, S.F. Grant, V.W. Jaddoe, M.R. Jarvelin, N.J. Timpson, I. Prokopenko, R.M. Freathy, M.-A.o.G.-a.l.-r.t.C. (MAGIC), E.G.G.E. Consortium, New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism, *Nat Genet*, 45 (2013) 76-82.

[150] S.F. Vatner, M. Park, L. Yan, G.J. Lee, L. Lai, K. Iwatsubo, Y. Ishikawa, J. Pessin, D.E. Vatner, Adenylyl cyclase type 5 in cardiac disease, metabolism, and aging, *Am J Physiol Heart Circ Physiol*, 305 (2013) H1-8.

[151] T. Shibasaki, Y. Sunaga, S. Seino, Integration of ATP, cAMP, and Ca²⁺ signals in insulin granule exocytosis, *Diabetes*, 53 Suppl 3 (2004) S59-62.

[152] D.J. Hodson, R.K. Mitchell, L. Marselli, T.J. Pullen, S. Gimeno Brias, F. Semplici, K.L. Everett, D.M. Cooper, M. Bugliani, P. Marchetti, V. Lavallard, D. Bosco, L. Piemonti, P.R. Johnson, S.J. Hughes, D. Li, W.H. Li, A.M. Shapiro, G.A. Rutter, ADCY5 couples glucose to insulin secretion in human islets, *Diabetes*, 63 (2014) 3009-3021.

[153] M. van de Bunt, J.E. Manning Fox, X. Dai, A. Barrett, C. Grey, L. Li, A.J. Bennett, P.R. Johnson, R.V. Rajotte, K.J. Gaulton, E.T. Dermitzakis, P.E. MacDonald, M.I. McCarthy, A.L. Gloyn, Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors, *PLoS Genet*, 11 (2015) e1005694.

- [154] S.C. Parker, M.L. Stitzel, D.L. Taylor, J.M. Orozco, M.R. Erdos, J.A. Akiyama, K.L. van Bueren, P.S. Chines, N. Narisu, B.L. Black, A. Visel, L.A. Pennacchio, F.S. Collins, N.C.S. Program, N.I.o.H.I.S.C.C.S.P. Authors, N.C.S.P. Authors, Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants, *Proc Natl Acad Sci U S A*, 110 (2013) 17921-17926.
- [155] J. Fadista, P. Vikman, E.O. Laakso, I.G. Mollet, J.L. Esguerra, J. Taneera, P. Storm, P. Osmark, C. Ladvall, R.B. Prasad, K.B. Hansson, F. Finotello, K. Uvebrant, J.K. Ofori, B. Di Camillo, U. Krus, C.M. Cilio, O. Hansson, L. Eliasson, A.H. Rosengren, E. Renström, C.B. Wollheim, L. Groop, Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism, *Proc Natl Acad Sci U S A*, 111 (2014) 13924-13929.
- [156] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, *Bioinformatics*, 29 (2013) 15-21.
- [157] A.A. Shabalín, Matrix eQTL: ultra fast eQTL analysis via large matrix operations, *Bioinformatics*, 28 (2012) 1353-1358.
- [158] O. Stegle, L. Parts, M. Piipari, J. Winn, R. Durbin, Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses, *Nat Protoc*, 7 (2012) 500-507.
- [159] J. Miyazaki, K. Araki, E. Yamato, H. Ikegami, T. Asano, Y. Shibasaki, Y. Oka, K. Yamamura, Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms, *Endocrinology*, 127 (1990) 126-132.
- [160] A.M. Ackermann, Z. Wang, J. Schug, A. Najj, K.H. Kaestner, Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes, *Mol Metab*, 5 (2016) 233-244.
- [161] L. Pasquali, K.J. Gaulton, S.A. Rodríguez-Seguí, L. Mularoni, I. Miguel-Escalada, I. Akerman, J.J. Tena, I. Morán, C. Gómez-Marín, M. van de Bunt, J. Ponsa-Cobas, N. Castro, T. Nammo, I. Cebola, J. García-Hurtado, M.A. Maestro, F. Pattou, L. Piemonti, T. Berney, A.L. Gloyn, P. Ravassard, J.L. Gómez-Skarmeta, F. Müller, M.I. McCarthy, J. Ferrer, Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants, *Nat Genet*, 46 (2014) 136-143.
- [162] L.S. Ramos, J.H. Zippin, M. Kamenetsky, J. Buck, L.R. Levin, Glucose and GLP-1 stimulate cAMP production via distinct adenylyl cyclases in INS-1E insulinoma cells, *J Gen Physiol*, 132 (2008) 329-338.
- [163] T. Shibasaki, Y. Sunaga, K. Fujimoto, Y. Kashima, S. Seino, Interaction of ATP sensor, cAMP sensor, Ca²⁺ sensor, and voltage-dependent Ca²⁺ channel in insulin granule exocytosis, *J Biol Chem*, 279 (2004) 7956-7961.
- [164] H.J. Westra, M.J. Peters, T. Esko, H. Yaghoobkar, C. Schurmann, J. Kettunen, M.W. Christiansen, B.P. Fairfax, K. Schramm, J.E. Powell, A. Zhernakova, D.V. Zhernakova, J.H. Veldink, L.H. Van den Berg, J. Karjalainen, S. Withoff, A.G. Uitterlinden, A. Hofman, F. Rivadeneira, P.A. 't Hoen, E. Reinmaa, K. Fischer, M. Nelis, L. Milani, D. Melzer, L. Ferrucci, A.B. Singleton, D.G. Hernandez, M.A. Nalls, G. Homuth, M. Nauck, D. Radke, U. Völker, M. Perola, V. Salomaa, J. Brody, A. Suchy-Dacey, S.A. Gharib, D.A. Enquobahrie, T. Lumley, G.W. Montgomery, S. Makino, H. Prokisch, C. Herder, M. Roden, H. Grallert, T. Meitinger, K. Strauch, Y. Li, R.C. Jansen, P.M. Visscher, J.C. Knight, B.M. Psaty, S. Ripatti, A. Teumer, T.M. Frayling, A. Metspalu, J.B. van Meurs, L. Franke, Systematic identification of trans eQTLs as putative drivers of known disease associations, *Nat Genet*, 45 (2013) 1238-1243.
- [165] V. Lyssenko, R. Lupi, P. Marchetti, S. Del Guerra, M. Orho-Melander, P. Almgren, M. Sjögren, C. Ling, K.F. Eriksson, A.L. Lethagen, R. Mancarella, G. Berglund, T. Tuomi, P. Nilsson, S. Del Prato, L.

Groop, Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes, *J Clin Invest*, 117 (2007) 2155-2163.

[166] A.H. Rosengren, M. Braun, T. Mahdi, S.A. Andersson, M.E. Travers, M. Shigeto, E. Zhang, P. Almgren, C. Ladenvall, A.S. Axelsson, A. Edlund, M.G. Pedersen, A. Jonsson, R. Ramracheya, Y. Tang, J.N. Walker, A. Barrett, P.R. Johnson, V. Lyssenko, M.I. McCarthy, L. Groop, A. Salehi, A.L. Gloyn, E. Renström, P. Rorsman, L. Eliasson, Reduced insulin exocytosis in human pancreatic β -cells with gene variants linked to type 2 diabetes, *Diabetes*, 61 (2012) 1726-1733.

[167] O. Le Bacquer, J. Kerr-Conte, S. Gargani, N. Delalleau, M. Huyvaert, V. Gmyr, P. Froguel, B. Neve, F. Pattou, TCF7L2 rs7903146 impairs islet function and morphology in non-diabetic individuals, *Diabetologia*, 55 (2012) 2677-2681.

[168] S.F. Boj, J.H. van Es, M. Huch, V.S. Li, A. José, P. Hatzis, M. Mokry, A. Haegebarth, M. van den Born, P. Chambon, P. Voshol, Y. Dor, E. Cuppen, C. Fillat, H. Clevers, Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand, *Cell*, 151 (2012) 1595-1607.

[169] M.I. McCarthy, P. Rorsman, A.L. Gloyn, TCF7L2 and diabetes: a tale of two tissues, and of two species, *Cell Metab*, 17 (2013) 157-159.

[170] R. Wagner, K. Dudziak, S.A. Herzberg-Schäfer, F. Machicao, N. Stefan, H. Staiger, H.U. Häring, A. Fritsche, Glucose-raising genetic variants in MADD and ADCY5 impair conversion of proinsulin to insulin, *PLoS One*, 6 (2011) e23639.

[171] D.L. Eizirik, M. Cnop, ER stress in pancreatic beta cells: the thin red line between adaptation and failure, *Sci Signal*, 3 (2010) pe7.

[172] K.J. Gaulton, T. Ferreira, Y. Lee, A. Raimondo, R. Mägi, M.E. Reschen, A. Mahajan, A. Locke, N.W. Rayner, N. Robertson, R.A. Scott, I. Prokopenko, L.J. Scott, T. Green, T. Sparso, D. Thuillier, L. Yengo, H. Grallert, S. Wahl, M. Frånberg, R.J. Strawbridge, H. Kestler, H. Chheda, L. Eisele, S. Gustafsson, V. Steinthorsdottir, G. Thorleifsson, L. Qi, L.C. Karssen, E.M. van Leeuwen, S.M. Willems, M. Li, H. Chen, C. Fuchsberger, P. Kwan, C. Ma, M. Linderman, Y. Lu, S.K. Thomsen, J.K. Rundle, N.L. Beer, M. van de Bunt, A. Chalisey, H.M. Kang, B.F. Voight, G.R. Abecasis, P. Almgren, D. Baldassarre, B. Balkau, R. Benediktsson, M. Blüher, H. Boeing, L.L. Bonnycastle, E.P. Bottinger, N.P. Burt, J. Carey, G. Charpentier, P.S. Chines, M.C. Cornelis, D.J. Couper, A.T. Crenshaw, R.M. van Dam, A.S. Doney, M. Dorkhan, S. Edkins, J.G. Eriksson, T. Esko, E. Eury, J. Fadista, J. Flannick, P. Fontanillas, C. Fox, P.W. Franks, K. Gertow, C. Gieger, B. Gigante, O. Gottesman, G.B. Grant, N. Grarup, C.J. Groves, M. Hassinen, C.T. Have, C. Herder, O.L. Holmen, A.B. Hreidarsson, S.E. Humphries, D.J. Hunter, A.U. Jackson, A. Jonsson, M.E. Jørgensen, T. Jørgensen, W.H. Kao, N.D. Kerrison, L. Kinnunen, N. Klopp, A. Kong, P. Kovacs, P. Kraft, J. Kravic, C. Langford, K. Leander, L. Liang, P. Lichtner, C.M. Lindgren, E. Lindholm, A. Linneberg, C.T. Liu, S. Lobbens, J. Luan, V. Lyssenko, S. Männistö, O. McLeod, J. Meyer, E. Mihailov, G. Mirza, T.W. Mühleisen, M. Müller-Nurasyid, C. Navarro, M.M. Nöthen, N.N. Oskolkov, K.R. Owen, D. Palli, S. Pechlivanis, L. Peltonen, J.R. Perry, C.G. Platou, M. Roden, D. Ruderfer, D. Rybin, Y.T. van der Schouw, B. Sennblad, G. Sigurðsson, A. Stančáková, G. Steinbach, P. Storm, K. Strauch, H.M. Stringham, Q. Sun, B. Thorand, E. Tikkanen, A. Tonjes, J. Trakalo, E. Tremoli, T. Tuomi, R. Wennauer, S. Wiltshire, A.R. Wood, E. Zeggini, I. Dunham, E. Birney, L. Pasquali, J. Ferrer, R.J. Loos, J. Dupuis, J.C. Florez, E. Boerwinkle, J.S. Pankow, C. van Duijn, E. Sijbrands, J.B. Meigs, F.B. Hu, U. Thorsteinsdottir, K. Stefansson, T.A. Lakka, R. Rauramaa, M. Stumvoll, N.L. Pedersen, L. Lind, S.M. Keinänen-Kiukaanniemi, E. Korpi-Hyövälti, T.E. Saaristo, J. Saltevo, J. Kuusisto, M. Laakso, A. Metspalu, R. Erbel, K.H. Jöcke, S. Moebus, S. Ripatti, V. Salomaa, E. Ingelsson, B.O. Boehm, R.N. Bergman, F.S. Collins, K.L. Mohlke, H. Koistinen, J. Tuomilehto, K. Hveem, I. Njølstad, P. Deloukas, P.J. Donnelly, T.M. Frayling, A.T. Hattersley, U. de Faire, A. Hamsten, T. Illig, A. Peters, S. Cauchi, R. Sladek, P. Froguel, T. Hansen, O. Pedersen, A.D. Morris, C.N. Palmer, S. Kathiresan, O. Melander, P.M. Nilsson, L.C. Groop, I. Barroso, C. Langenberg, N.J. Wareham, C.A. O'Callaghan, A.L. Gloyn, D. Altshuler, M. Boehnke, T.M. Teslovich, M.I.

- McCarthy, A.P. Morris, D.G.R.A.M.-a.D. Consortium, Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci, *Nat Genet*, 47 (2015) 1415-1425.
- [173] Y. Wang, E.B. Rimm, M.J. Stampfer, W.C. Willett, F.B. Hu, Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men, *Am J Clin Nutr*, 81 (2005) 555-563.
- [174] D. Canoy, Distribution of body fat and risk of coronary heart disease in men and women, *Curr Opin Cardiol*, 23 (2008) 591-598.
- [175] R. Huxley, S. Mendis, E. Zheleznyakov, S. Reddy, J. Chan, Body mass index, waist circumference and waist:hip ratio as predictors of cardiovascular risk--a review of the literature, *Eur J Clin Nutr*, 64 (2010) 16-22.
- [176] Z. Tahergorabi, M. Khazaei, The relationship between inflammatory markers, angiogenesis, and obesity, *ARYA Atheroscler*, 9 (2013) 247-253.
- [177] Y. Cao, Angiogenesis as a therapeutic target for obesity and metabolic diseases, *Chem Immunol Allergy*, 99 (2014) 170-179.
- [178] A.D. Gitler, M.M. Lu, J.A. Epstein, PlexinD1 and semaphorin signaling are required in endothelial cells for cardiovascular development, *Dev Cell*, 7 (2004) 107-116.
- [179] J. Torres-Vázquez, A.D. Gitler, S.D. Fraser, J.D. Berk, Van N Pham, M.C. Fishman, S. Childs, J.A. Epstein, B.M. Weinstein, Semaphorin-plexin signaling guides patterning of the developing vasculature, *Dev Cell*, 7 (2004) 117-123.
- [180] C. Gu, Y. Yoshida, J. Livet, D.V. Reimert, F. Mann, J. Merte, C.E. Henderson, T.M. Jessell, A.L. Kolodkin, D.D. Ginty, Semaphorin 3E and plexin-D1 control vascular pattern independently of neuropilins, *Science*, 307 (2005) 265-268.
- [181] A. Watakabe, S. Ohsawa, T. Hashikawa, T. Yamamori, Binding and complementary expression patterns of semaphorin 3E and plexin D1 in the mature neocortices of mice and monkeys, *J Comp Neurol*, 499 (2006) 258-273.
- [182] T. Toyofuku, M. Yabuki, J. Kamei, M. Kamei, N. Makino, A. Kumanogoh, M. Hori, Semaphorin-4A, an activator for T-cell-mediated immunity, suppresses angiogenesis via Plexin-D1, *EMBO J*, 26 (2007) 1373-1384.
- [183] S. Chauvet, S. Cohen, Y. Yoshida, L. Fekrane, J. Livet, O. Gayet, L. Segu, M.C. Buhot, T.M. Jessell, C.E. Henderson, F. Mann, Gating of Sema3E/PlexinD1 signaling by neuropilin-1 switches axonal repulsion to attraction during brain development, *Neuron*, 56 (2007) 807-822.
- [184] U. Yazdani, J.R. Terman, The semaphorins, *Genome Biol*, 7 (2006) 211.
- [185] J.B. Ding, W.J. Oh, B.L. Sabatini, C. Gu, Semaphorin 3E-Plexin-D1 signaling controls pathway-specific synapse formation in the striatum, *Nat Neurosci*, 15 (2012) 215-223.
- [186] C. Meda, F. Molla, M. De Pizzol, D. Regano, F. Maione, S. Capano, M. Locati, A. Mantovani, R. Latini, F. Bussolino, E. Giraud, Semaphorin 4A exerts a proangiogenic effect by enhancing vascular endothelial growth factor-A expression in macrophages, *J Immunol*, 188 (2012) 4081-4092.
- [187] J.E. Minchin, I. Dahlman, C.J. Harvey, N. Mejhert, M.K. Singh, J.A. Epstein, P. Arner, J. Torres-Vázquez, J.F. Rawls, Plexin D1 determines body fat distribution by regulating the type V collagen microenvironment in visceral adipose tissue, *Proc Natl Acad Sci U S A*, 112 (2015) 4363-4368.

- [188] S. Fisher, E.A. Grice, R.M. Vinton, S.L. Bessling, A. Urasaki, K. Kawakami, A.S. McCallion, Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish, *Nat Protoc*, 1 (2006) 1297-1305.
- [189] Y. Wang, M.S. Kaiser, J.D. Larson, A. Nasevicius, K.J. Clark, S.A. Wadman, S.E. Roberg-Perez, S.C. Ekker, P.B. Hackett, M. McGrail, J.J. Essner, *Moesin1* and *Ve-cadherin* are required in endothelial cells during *in vivo* tubulogenesis, *Development*, 137 (2010) 3119-3128.
- [190] S.H. Oehlers, M.R. Cronan, N.R. Scott, M.I. Thomas, K.S. Okuda, E.M. Walton, R.W. Beerman, P.S. Crosier, D.M. Tobin, Interception of host angiogenic signalling limits mycobacterial growth, *Nature*, 517 (2015) 612-615.
- [191] L. Yao, B.P. Berman, P.J. Farnham, Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes, *Crit Rev Biochem Mol Biol*, 50 (2015) 550-573.
- [192] K. Eun, S.U. Hwang, H.M. Jeon, S.H. Hyun, H. Kim, Comparative Analysis of Human, Mouse, and Pig Glial Fibrillary Acidic Protein Gene Structures, *Anim Biotechnol*, 27 (2016) 126-132.
- [193] C. Zhang, Y.S. Kho, Z. Wang, Y.T. Chiang, G.K. Ng, P.C. Shaw, Y. Wang, R.Z. Qi, Transmembrane and coiled-coil domain family 1 is a novel protein of the endoplasmic reticulum, *PLoS One*, 9 (2014) e85206.
- [194] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, H. Parkinson, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations, *Nucleic Acids Res*, 42 (2014) D1001-1006.
- [195] Y.G. Tak, P.J. Farnham, Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome, *Epigenetics Chromatin*, 8 (2015) 57.
- [196] N.F. Wasserman, I. Aneas, M.A. Nobrega, An 8q24 gene desert variant associated with prostate cancer risk confers differential *in vivo* activity to a MYC enhancer, *Genome Res*, 20 (2010) 1191-1197.
- [197] S. Wang, F. Wen, K.L. Tessneer, P.M. Gaffney, TALEN-mediated enhancer knockout influences TNFAIP3 gene expression and mimics a molecular phenotype associated with systemic lupus erythematosus, *Genes Immun*, (2016).
- [198] D.S. Paul, C.A. Albers, A. Rendon, K. Voss, J. Stephens, P. van der Harst, J.C. Chambers, N. Soranzo, W.H. Ouwehand, P. Deloukas, H. Consortium, Maps of open chromatin highlight cell type-restricted patterns of regulatory sequence variation at hematological trait loci, *Genome Res*, 23 (2013) 1130-1141.
- [199] M. Cavalli, G. Pan, H. Nord, C. Wadelius, Looking beyond GWAS: allele-specific transcription factor binding drives the association of GALNT2 to HDL-C plasma levels, *Lipids Health Dis*, 15 (2016) 18.
- [200] J.M. Coulson, C.E. Fiskerstrand, P.J. Woll, J.P. Quinn, E-box motifs within the human vasopressin gene promoter contribute to a major enhancer in small-cell lung cancer, *Biochem J*, 344 Pt 3 (1999) 961-970.
- [201] C.I. Rodríguez, N. Gironès, M. Fresno, Cha, a basic helix-loop-helix transcription factor involved in the regulation of upstream stimulatory factor activity, *J Biol Chem*, 278 (2003) 43135-43145.
- [202] S. Yan, D.T. Jane, M.J. Dufresne, B.F. Sloane, Transcription of cathepsin B in glioma cells: regulation by an E-box adjacent to the transcription initiation site, *Biol Chem*, 384 (2003) 1421-1427.

[203] L.M. Pardo, G. Piras, R. Asproni, K.J. van der Gaag, A. Gabbas, A. Ruiz-Linares, P. de Knijff, M. Monne, P. Rizzu, P. Heutink, Dissecting the genetic make-up of North-East Sardinia using a large set of haploid and autosomal markers, *Eur J Hum Genet*, 20 (2012) 956-964.

[204] C.J. Willer, K.L. Mohlke, Finding genes and variants for lipid levels after genome-wide association analysis, *Curr Opin Lipidol*, 23 (2012) 98-103.

[205] M. Horikoshi, R. Mägi, M. van de Bunt, I. Surakka, A.P. Sarin, A. Mahajan, L. Marullo, G. Thorleifsson, S. Hägg, J.J. Hottenga, C. Ladenvall, J.S. Ried, T.W. Winkler, S.M. Willems, N. Pervjakova, T. Esko, M. Beekman, C.P. Nelson, C. Willenborg, S. Wiltshire, T. Ferreira, J. Fernandez, K.J. Gaulton, V. Steinthorsdottir, A. Hamsten, P.K. Magnusson, G. Willemsen, Y. Milaneschi, N.R. Robertson, C.J. Groves, A.J. Bennett, T. Lehtimäki, J.S. Viikari, J. Rung, V. Lyssenko, M. Perola, I.M. Heid, C. Herder, H. Grallert, M. Müller-Nurasyid, M. Roden, E. Hyppönen, A. Isaacs, E.M. van Leeuwen, L.C. Karssen, E. Mihailov, J.J. Houwing-Duistermaat, A.J. de Craen, J. Deelen, A.S. Havulinna, M. Blades, C. Hengstenberg, J. Erdmann, H. Schunkert, J. Kaprio, M.D. Tobin, N.J. Samani, L. Lind, V. Salomaa, C.M. Lindgren, P.E. Slagboom, A. Metspalu, C.M. van Duijn, J.G. Eriksson, A. Peters, C. Gieger, A. Jula, L. Groop, O.T. Raitakari, C. Power, B.W. Penninx, E. de Geus, J.H. Smit, D.I. Boomsma, N.L. Pedersen, E. Ingelsson, U. Thorsteinsdottir, K. Stefansson, S. Ripatti, I. Prokopenko, M.I. McCarthy, A.P. Morris, E. Consortium, Discovery and Fine-Mapping of Glycaemic and Obesity-Related Trait Loci Using High-Density Imputation, *PLoS Genet*, 11 (2015) e1005230.

[206] A.M. Delgado-Vega, M.G. Dozmorov, M.B. Quirós, Y.Y. Wu, B. Martínez-García, S.V. Kozyrev, J. Frostegård, L. Truedsson, E. de Ramón, M.F. González-Escribano, N. Ortego-Centeno, B.A. Pons-Estel, S. D'Alfonso, G.D. Sebastiani, T. Witte, B.R. Lauwerys, E. Endreffy, L. Kovács, C. Vasconcelos, B.M. da Silva, J.D. Wren, J. Martin, C. Castillejo-López, M.E. Alarcón-Riquelme, Fine mapping and conditional analysis identify a new mutation in the autoimmunity susceptibility gene BLK that leads to reduced half-life of the BLK protein, *Ann Rheum Dis*, 71 (2012) 1219-1226.

[207] Y. Wu, L.L. Waite, A.U. Jackson, W.H. Sheu, S. Buyske, D. Absher, D.K. Arnett, E. Boerwinkle, L.L. Bonnycastle, C.L. Carty, I. Cheng, B. Cochran, D.C. Croteau-Chonka, L. Dumitrescu, C.B. Eaton, N. Franceschini, X. Guo, B.E. Henderson, L.A. Hindorf, E. Kim, L. Kinnunen, P. Komulainen, W.J. Lee, L. Le Marchand, Y. Lin, J. Lindström, O. Lingaas-Holmen, S.L. Mitchell, N. Narisu, J.G. Robinson, F. Schumacher, A. Stančáková, J. Sundvall, Y.J. Sung, A.J. Swift, W.C. Wang, L. Wilkens, T. Wilsgaard, A.M. Young, L.S. Adair, C.M. Ballantyne, P. Bůžková, A. Chakravarti, F.S. Collins, D. Duggan, A.B. Feranil, L.T. Ho, Y.J. Hung, S.C. Hunt, K. Hveem, J.M. Juang, A.Y. Kesäniemi, J. Kuusisto, M. Laakso, T.A. Lakka, I.T. Lee, M.F. Leppert, T.C. Matise, L. Moilanen, I. Njølstad, U. Peters, T. Quertermous, R. Rauramaa, J.I. Rotter, J. Saramies, J. Tuomilehto, M. Uusitupa, T.D. Wang, M. Boehnke, C.A. Haiman, Y.D. Chen, C. Kooperberg, T.L. Assimes, D.C. Crawford, C.A. Hsiung, K.E. North, K.L. Mohlke, Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained, *PLoS Genet*, 9 (2013) e1003379.

[208] M. Horikoshi, L. Pasquali, S. Wiltshire, J.R. Huyghe, A. Mahajan, J.L. Asimit, T. Ferreira, A.E. Locke, N.R. Robertson, X. Wang, X. Sim, H. Fujita, K. Hara, R. Young, W. Zhang, S. Choi, H. Chen, I. Kaur, F. Takeuchi, P. Fontanillas, D. Thuillier, L. Yengo, J.E. Below, C.H. Tam, Y. Wu, G. Abecasis, D. Altshuler, G.I. Bell, J. Blangero, N.P. Burt, R. Duggirala, J.C. Florez, C.L. Hanis, M. Seielstad, G. Atzmon, J.C. Chan, R.C. Ma, P. Froguel, J.G. Wilson, D. Bharadwaj, J. Dupuis, J.B. Meigs, Y.S. Cho, T. Park, J.S. Kooner, J.C. Chambers, D. Saleheen, T. Kadowaki, E.S. Tai, K.L. Mohlke, N.J. Cox, J. Ferrer, E. Zeggini, N. Kato, Y.Y. Teo, M. Boehnke, M.I. McCarthy, A.P. Morris, T.D.-G. Consortium, Transancestral fine-mapping of four type 2 diabetes susceptibility loci highlights potential causal regulatory mechanisms, *Hum Mol Genet*, (2016).

[209] M.K. Chuah, D. Collen, T. VandenDriessche, Biosafety of adenoviral vectors, *Curr Gene Ther*, 3 (2003) 527-543.

[210] J. Dekker, K. Rippe, M. Dekker, N. Kleckner, Capturing chromosome conformation, *Science*, 295 (2002) 1306-1311.

[211] F.A. Ran, P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, F. Zhang, Genome engineering using the CRISPR-Cas9 system, *Nat Protoc*, 8 (2013) 2281-2308.

[212] J.D. Sander, J.K. Joung, CRISPR-Cas systems for editing, regulating and targeting genomes, *Nat Biotechnol*, 32 (2014) 347-355.

[213] J.A. Doudna, E. Charpentier, Genome editing. The new frontier of genome engineering with CRISPR-Cas9, *Science*, 346 (2014) 1258096.