Title
Searching for Variants, Genes, and Pathways Involved in Hyperlipidemia

Author
Haas, Blake Edwin

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Searching for Variants, Genes, and Pathways Involved in Hyperlipidemia

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Human Genetics

by

Blake Edwin Haas

2012

ABSTRACT OF THE DISSERTATION

Searching for Variants, Genes, and Pathways Involved in Hyperlipidemia

by

Blake Edwin Haas

Doctor of Philosophy in Human Genetics

University of California, Los Angeles, 2012

Professor Päivi Pajukanta, Chair

Genome-wide association studies (GWAS) have been successful in identifying variants with low to moderate effects on serum lipid levels. However, determining the causal variant and underlying mechanism in an associated region has been difficult since linkage equilibrium (LD) of the associated variants often extends to large regions covering multiple genes, and GWAS provide no functional information. Using adipose gene expression data, we provide evidence of mechanism underlying a low-density lipoprotein cholesterol (LDL-C) GWAS signal in the apolipoprotein B (*APOB*) region. First, we determined that an LDL-C GWAS signal uncovered in a population of European ancestry (rs7575840) replicates in a Mexican study sample (1). Mexicans are an understudied population with a high prevalence of dyslipidemia; 44% of the population has high total cholesterol levels (above 200 mg/dl) (2). To further elucidate the mechanism underlying the association between rs7575840 and LDL-C, we measured lipid particle subclasses using Nuclear Magnetic Resonance (NMR), and we determined that rs7575840 is also associated with apoB-containing lipid particles, including very small very-low-density lipoprotein, intermediate lipoprotein, and LDL particles (1). We also discovered a

possible mechanism underlying the increase in apoB-containing particles; rs7575840 disrupts a transcription factor binding site of the transcription factors CCAAT/enhancer binding protein alpha and CCAAT/enhancer binding protein beta, impacting gene expression of *APOB* and a noncoding RNA BU630349 in adipose tissue (1).

In our second study, we searched for whether novel genes and biological gene expression networks associated with triglyceride (TG) levels replicated and shared across populations (Haas et al. submitted (3)). Adipose tissue has previously been shown to be involved in regulation of serum TG levels in humans (4). High serum TG levels predispose to coronary heart disease (CHD). Searching for novel TG-associated biological networks in adipose tissue may provide novel insights into TG regulation. We measured gene expression in adipose samples from two Finnish and one Mexican study sample. In each study sample, we observed a module (biological gene expression network) that was significantly associated with serum TG levels (3). The most significant TG modules observed in each of the three study samples significantly overlapped ($p<10^{-10}$) and shared 34 genes, utilizing two unique methods of measuring gene expression, microarrays and RNA sequencing (RNAseq). Thus, the results should be robust to the measurement method of gene expression. In the 34 genes shared between the three TG modules, more nonsynonymous variants ($p=0.034$) and overall variants ($p=0.018$) were observed in individuals with high TGs when compared with the individuals with low TGs (3). Seven of the 34 genes (*ARHGAP30, CCR1, CXCL16, FERMT3, HCST, RNASET2, SELPG*) were identified as the key hub genes of all three TG modules ($p<10^{-7}$) (3). As only 11 of the 34 genes have prior evidence of involvement in CHD, type 2 diabetes, or obesity, our study provides 23 new candidates for TG regulation. Furthermore, two of the 34 genes (*ARHGAP9, LST1*) reside in

previous TG GWAS regions, suggesting them as the regional candidates underlying the GWAS signals. This study presents a novel TG biological network shared across populations.

Our next study focused on determining whether Procadherin 15 (*PCDH15*) variants are involved with Familial Combined Hyperlipidemia (FCHL). Previous studies have revealed that *PCDH15* resides in a region linked to FCHL (5-7), so we hypothesized that nonsynonymous variants in *PCDH15* may be associated with FCHL in 92 Finnish and Dutch families. We discovered that one variant (rs10825269) associates with serum TG, apoB, and TC (Total Cholesterol) levels in these families (8). Next, we made a *PCDH15* knockout model in collaboration with investigators at Case Western Reserve University and discovered that serum TG and TC levels decreased in mice with both copies of *PCDH15* knocked out (8).

The dissertation of Blake Edwin Haas is approved.

Eleazar Eskin

Aldons J. Lusis

Janet Sinsheimer

Yibin Wang

Päivi Pajukanta, Committee Chair

University of California, Los Angeles

2012

DEDICATION

I dedicate this thesis to my grandparents, parents, and brothers for always being there for me.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

ADR – Adverse Drug Reaction

APOB – Apolipoprotein B

CEBPA – CCAAT/enhancer binding protein alpha

CEBPB - CCAAT/enhancer binding protein beta

CHD – Coronary Heart Disease

cis-eQTL – *cis*-acting Expression Quantitative Trait Locus

FCHL – Familial Combined Hyperlipidemia

FH – Familial Hypercholesterolemia

FPKM - Fragments per Kilobase of Exon per Million Mapped Reads

GWAS – Genome-wide Association Study

HDL-C – High Density Lipoprotein Cholesterol

IDL – Intermediate Density Lipoprotein

LD – Linkage Disequilibrium

LDL-C – Low Density Lipoprotein Cholesterol

ME – Module Eigengene

NEO – Network Edge Orienting

NMR – Nuclear Magnetic Resonance

PCDH15 – Procadherin 15

RNAseq – RNA Sequencing

SNP – Single Nucleotide Polymorphism

SNV – Single Nucleotide Variant

TFBS – Transcription Factor Binding Site

TC – Total Cholesterol

TG – Triglycerides

VLDL – Very Low Density Lipoprotein

WGCNA – Weighted Gene Co-expression Network Analysis

VITA

2007	B.S., Physiological Science Cum Laude College of Letters and Science Honors University of California, Los Angeles Los Angeles, California
2009, 2011	Teaching Assistant Life Sciences Core Curriculum Los Angeles, California

PUBLICATIONS AND PRESENTATIONS

Haas BE, Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, Vergnes L, Cruz-Bautista I, Riba L, Stancakova A, Kuusisto J, Soininen P, Kangas A, Ala-Korpela M, Tusie-Luna T, Laakso M, Pajukanta P. **Evidence of How rs7575840 Influences Apolipoprotein B-Containing Lipid Particles.** *Arterioscler Thromb Vasc Biol*. 2011 May;31:1201-1207. PMID: 21393584

Haas BE, Horvath S, Pietiläinen KH, Cantor RM, Nikkola E, Weissglas-Volkov D, Rissanen A, Civelek M, Cruz-Bautista I, Riba L, Kuusisto J, Kaprio J, Tusie-Luna T, Laakso M, Aguilar-Salinas CA, Pajukanta P. **Adipose Transcript Networks Across Finns and Mexicans Identify Novel Triglyceride Genes.** Manuscript submitted for publication.

Ochoa MT, Teles R, Haas BE, Zaghi D, Li H, Sarno EN, Rea TH, Modlin RL, Lee DJ. **A role for IL-5 in promoting increased IgM at the site of disease in leprosy.** *Immunology*. 2010 Nov;131(3):405-414. PMID: 20561085

Huertas-Vazquez A, Plaisier CL, Geng R, Haas BE, Lee J, Greevenbroek MM, van der

Kallen C, de Bruin TW, Taskinen MR, Alagramam KN, Pajukanta P. **A nonsynonymous SNP within PCDH15 is associated with lipid traits in familial combined hyperlipidemia.** *Hum Genet.* 2010 Jan;127(1):83-89. PMID: 19816713

Haas BE, Plaisier CL, Weisglass-Volkov W, Laakso M, Pajukanta P. **Comparison of genome-wide transcript and sequence variation data identifies a DNA sequence variant for LDL cholesterol levels.** Genomic Analysis Training Program Retreat, May 4, 2010. (**Oral Presentation**)

Haas BE, Plaisier CL, Weissglas-Volkov D, Laakso M, Pajukanta P. Searching for genetic variants affecting serum cholesterol levels. Genomic Analysis Training Program Underrepresented Minority Recruitment to Scientific Research, October 23, 2009. (**Oral Presentation**)

Plaisier CL, Haas BE, Laakso M, Pajukanta P. Cross-species Comparisons of GWAS Data as a Screening Tool to Identify Novel Loci for LDL-C. International Symposium on Atherosclerosis XV Annual Meeting, June 15, 2009. (**Oral Presentation**)

Haas BE, Tusié-Luna T, Aguilar-Salinas CA, Riba L, Plaisier CL, Weissglas-Volkov D, Laakso M, Pajukanta P. **A cluster of SNPs in the apolipoprotein B region is associated with high serum apolipoprotein B levels.** American Society of Human Genetics 59th Annual Meeting, October 22, 2009. (**Poster Presentation**).

CHAPTER 1:

INTRODUCTION

1.1 - Overview of Hypercholesterolemia, Hypertriglyceridemia, and Familial Combined Hyperlipidemia

Hypercholesterolemia is defined as the presence of high cholesterol levels in the blood. Previous studies have discovered that hypercholesterolemia increases the risk of coronary heart disease (the development of fatty plaque deposits in the walls of coronary arteries) and death (9). The most successful and well-proven drug to slow the development of coronary heart disease and prevent mortality has been statins (10), which lower serum low-density lipoprotein cholesterol (LDL-C) and total cholesterol (TC) levels by inhibiting cholesterol synthesis in the liver. Currently, common variation in the genome explains 12.2% of the variation in LDL-C levels (11). Previous studies have estimated that the heritability of LDL-C is between 40%-50%, leaving a large portion of the genetic variation impacting cholesterol levels undiscovered (12). Monogenic forms of familial hypercholesterolemia have been discovered in four genes: *LDLR*, *APOB*, *PCSK9*, and *LDLRAP1* (13). Searching for additional variants and genes that explain variation in cholesterol levels will lead to a better understanding of the underlying biology of cholesterol regulation and may reveal potential novel drug targets.

Patients with hypertriglyceridemia have high serum triglyceride levels and a higher risk of coronary heart disease (14). Hypertriglyceridemia is a common condition in the United States, with an estimated 31% of the U.S. adults being affected (15). Some controversy still exists on whether triglyceride levels are directly causing increased CHD risk (16), but multiple studies have recently concluded that serum TG levels are indeed an independent risk factor for CHD (17-20). Fibrates, which lower TG levels and raise high density lipoprotein cholesterol (HDL-C) levels, have been successful at lowering the risk of cardiovascular events in individuals with

hypertriglyceridemia (21). Currently, common variation in the genome explains 9.6% of the variation in triglyceride levels (11), which is a small portion of the expected 35%-48% heritability of serum triglyceride levels (12). Monogenic forms of hypertriglyceridemia have previously been discovered in the genes *LPL, APOC2, APOA5, LMF1,* and *GPIHBP1*, which are typically passed through families in an autosomal recessive fashion (22). A recent review article suggests additional research is necessary to develop novel TG drugs and better elucidate the role triglycerides play in CHD progression (14). Thus, our goal of determining genes and genetic variation that influence serum triglyceride levels may further advance the goal of preventing coronary heart disease by improving our understanding of triglyceride regulation.

Familial combined hyperlipidemia (FCHL) is a complex disorder characterized by the presence of high serum TC levels, high serum TG levels, or both (23,24). Originally, FCHL was thought to be a monogenic, dominant disorder (25), but later studies provided evidence that FCHL is actually a polygenic, complex disorder (26-28). FCHL greatly increases an affected individual's risk of developing early onset CHD (29-30). FCHL is a common complex disease, with approximately 1-6% of Western populations affected (29,31). Although both common and rare genetic variants from multiple genes are expected to play a role in FCHL, the genetic factors influencing FCHL remain elusive, with only three genes (*APOA1/C3/A4/A5*, *USF1*, and *LPL*) displaying associations with FCHL in multiple studies (32). Physicians currently treat FCHL patients with drugs that lower TC and/or TG levels, yet FCHL patients are still at a high risk of CHD (33), indicating that there is a public need to better biologically understand FCHL and develop additional treatment options. Searching for variants and genes that influence FCHL risk

may lead to a better understanding of the disorder, a step necessary for more tailored treatment and management of the disorder in the future.

1.2 Discovering Genetic Factors Predisposing to Hypercholesterolemia, Hypertriglyceridemia, and Familial Combined Hyperlipidemia

Linkage studies were previously a popular unbiased method for searching for genomic regions that may influence lipid levels in large extended families. Linkage uses genetic markers genotyped throughout the genome to search for large genomic regions that tend to be passed along with the trait of interest from one generation to another. For FCHL, a linkage study in Finnish families identified a large region on chromosome 1q21–q23 that segregated with FCHL status (34). Ultimately, *USF1* was identified as the gene influencing FCHL risk in the linkage peak (35). Linkage was also the initial method used that implicated the gene cluster *APOA1/C3/A4/A5* in FCHL development (36,37).

The candidate gene approach of investigating genes for trait association has yielded success in discovering genes involved in lipid regulation. The candidate gene approach uses information from previous research to determine whether logical candidates impact the trait of interest. This method has successfully identified *LPL* as a gene influencing FCHL risk (38). An association between *LPL* and FCHL status was observed because *LPL* was already known to be involved in lipid metabolism, making *LPL* a good candidate to search for single nucleotide polymorphisms (SNPs) impacting FCHL status.

Genome-wide association studies (GWAS) are a popular method of discovering common SNPs that are associated with serum lipid levels in a population. GWAS is able to identify variants without the need for *a priori* knowledge. GWAS genotype SNPs that tag most of the common variation throughout the genome to discover associations between common variants (minor allele

frequency > 0.05) and a quantitative or qualitative trait of interest. The linkage disequilibrium (LD) from the European HapMap samples are usually used to determine which SNPs should be genotyped to tag common genomic variation. GWAS have been successful at identifying common variants associated with LDL-C and TG, but fail to explain a large proportion of the genetic heritability of these traits. Figure 1.2.1 displays the primary success of the GWAS identifying novel genomic regions associated with a lipid trait. These regions can then be explored using regional LD analysis, resequencing, and functional studies to identify the full spectrum of susceptibility variants, common and rare, as well as underlying mechanisms responsible for the association. As shown in Figure 1.2.1, GWAS also detect loci previously identified for extreme lipid levels, further validating the GWAS approach. For example, all 4 known genes causing Familial Hypercholesterolemia (FH) were observed in a recent lipid metaGWAS (11). On the other hand, figure 1.2.2 also illustrates that the TG GWAS identified several novel genomic regions not previously known to cause hypertriglyceridemia.

Figure 1.2.1 TC GWAS Signals (11) and Monogenic FH Genes (13).

TC GWAS Signals

gene	ANGPTL3	EVI5	IRF2BP2	LDLRAP1	MOSC1	PCSK9	SORT1	ABCG5/8
chr	1	1	1	1	1	1	1	2
gene	APOB	GCKR	RAB3GAP1	RAF1	HMGCR	TIMD4	C6orf106	FRK
chr	2	2	2	3	5	5	6	6
gene	HFE	HLA	LPA	MYLIP	DNAH11	NPC1L1	CYP7A1	NAT2
chr	6	6	6	6	7	7	8	8
gene	PLEC1	PPP1R3B	TRIB1	TRPS1	TTC39B	ABCA1	ABO	GPAM
chr	8	8	8	8	8	9	9	10
gene	APOA1	FADS1-2-3	SPTY2D1	ST3GAL4	UBASH3B	BRAP	HNF1A	LIPC
chr	11	11	11	11	11	12	12	15
gene	CETP	HPR	OSBPL7	LIPG	APOE	CILP2	FLJ36070	LDLR
chr	16	16	17	18	19	19	19	19
gene	ERGIC3	HNF4A	MAFB	TOP1				
chr	20	20	20	20				

TC GWAS Heritability – 12.4%
LDL-C Heritability – 40%-50%
HDL-C Heritability – 40%-60%

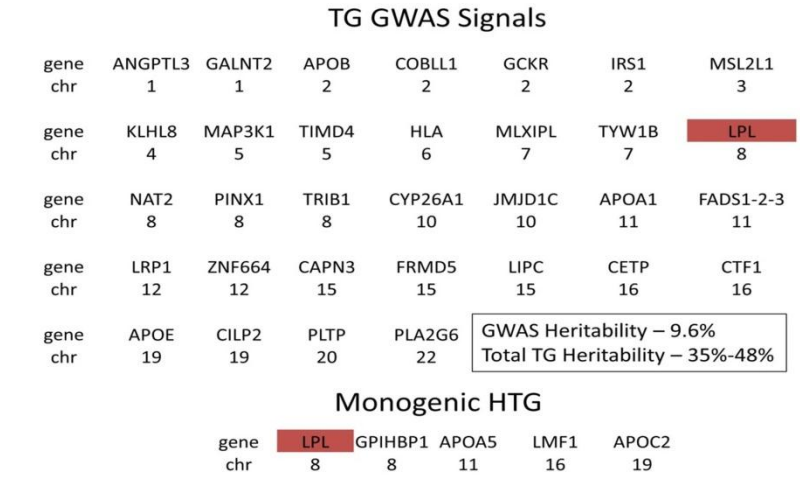
Monogenic FH Genes

gene	LDLRAP1	PCSK9	APOB	LDLR
chr	1	1	2	19

Genes found in both lists are highlighted in brown. Lipid heritability estimates were performed by Weiss et al. (12).

Resequencing GWAS candidate genes in individuals with extreme lipid values has been successful in identifying rare variants that impact lipid levels (39). Because lipid GWAS identified genes known to harbor mutations for monogenic forms of the lipid disorders, it is possible that genes identified by GWAS may also harbor rare variants that impact a trait. Resequencing individuals with extreme lipid values may thus reveal rare, novel variants that may have a large impact on serum lipid levels. For instance, resequencing four genes identified in a TG GWAS (*APOA5*, *GCKR*, *LPL* and *APOB)* in high TG cases and controls lead to the discovery of 154 rare nonsense and missense variants found only in individuals with high TG (39). Additionally, resequencing individuals in the phenotypic extremes require 5 to 10 times fewer individuals to discover an association than a GWAS (40), saving the time and money associated with fewer study participants.

Figure 1.2.2 TG GWAS Genes (11) and Monogenic HTG Genes (22).



Genes found in both lists are highlighted in brown. Lipid heritability estimates were performed by Weiss et al. (12).

One important way to prioritize variants for follow-up in functional studies is to utilize bioinformatics tools. Several software programs, such as SIFT (41-45) and PolyPhen2 (46),

have been developed to determine whether a nonsynonymous variant is likely to impact gene function. These software packages utilize protein structure and conservation to categorize variants as damaging or benign. Annovar (47) is a commonly used software program that determines the location within a gene a variant falls in, as well as whether or not the SNP is nonsynonymous. Additionally, Annovar returns protein prediction scores from various programs, including SIFT and PolyPhen2, making it easier to prioritize variants in a region for follow-up in functional experiments.

Comparison of GWAS data and regional LD structure in multiple populations may help discover causal variants, because different populations have different allele frequencies and different LD patterns. Utilizing LD differences between populations may thus break up large LD blocks associated with a trait and elucidate a causal variant in a genomic region. To date, most GWAS have been performed in European study samples using European LD patterns to tag common variation. Mexicans can serve as an excellent population for studying lipid regulation because they represent an admixed population with a high prevalence of hypercholesterolemia (44% of the population has TC > 200mg/dl) and hypertriglyceridemia (32% of the population has TG > 150 mg/dl) (2). Utilizing a Mexican population can reveal novel loci in this understudied population; help validate European lipid loci; and narrow the typically large LD blocks in GWAS regions.

1.3 Gene Expression as a Method to Investigate Hyperlipidemia

Microarrays have been an effective way to measure gene expression over the past few decades. Microarrays are glass slides which contain probes complementary to genes throughout the genome. Fluorescently labeled RNA binds the probe with complementary sequence, and the amount of light emitted by a portion of the microarray slide is proportional to the amount of bound RNA, enabling computer software to estimate gene expression.

RNAseq is a newer method of measuring gene expression. RNAseq converts mRNA into cDNA, which is then sequenced in a high throughput sequencer. Gene expression is measured based on the number of sequenced reads mapping to the gene. To account for the large variation in number of reads mapped between lanes and that longer genes are more likely to be sequenced, RNAseq gene expression units are represented as fragments per kilobase of exon per million mapped reads (FPKM).

RNAseq has several advantages over microarrays for measuring gene expression. RNAseq does not require *a priori* knowledge of which regions of the genome are expressed. In RNAseq, all mRNA is converted to cDNA and sequenced and mapped to the genome, while microarrays require probes to be made ahead of time. Thus, microarrays can only measure expression of known genes, while RNAseq can discover novel genes. Another RNAseq advantage is that it can discover novel isoforms of genes, unlike microarrays which are restricted due to probe selection. Finally, RNAseq is more accurate at measuring genes expressed at very high and very low levels. Microarrays are less accurate at measuring gene expression in the extremes, because at low levels of expression, a gene may not be called due to background noise, while at high

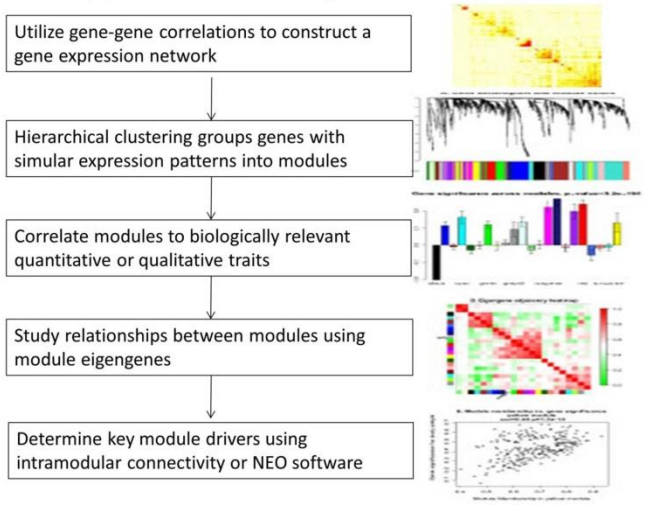
levels of expression, the probes corresponding to the gene may become too saturated to accurately measure expression.

One of the main disadvantages of RNAseq in gene expression analysis is that it is currently more expensive than the microarray-based technologies. As sequencing technologies develop and improve overtime, it is, however, likely that the costs of performing RNAseq will decrease. Other concerns related to RNAseq include the GC content sequencing bias (i.e. GC-rich and GC-poor sequencing reads may be expressed at higher levels than reported by RNAseq) and large variation in the number of mapped reads observed between sequencing lanes. Besides price, the primary advantage microarrays have over RNAseq is that microarrays require fewer hours of man power. This cost savings allows several additional samples to be measured for the same price and time commitment as RNAseq.

Utilizing gene expression data has been an important tool to discover genes involved in regulating lipid levels. One method of discovering the possible mechanism underlying a SNP association with a lipid trait has been determining whether the SNP acts as a *cis*-acting expression quantitative trait locus (*cis*-eQTL). Because GWAS do not provide any functional information, it is important to determine whether a GWAS SNP may influence the trait of interest by acting as a *cis*-eQTL in a specific tissue. For instance, a recent LDL-C, HDL-C, and TG GWAS checked whether all SNPs significantly associated with a lipid trait act as a *cis*-eQTL in tissues known to be relevant in lipid regulation, including liver and adipose tissue (11).

Several studies have searched for adipose tissue gene expression differences between obese and lean subjects (48,49). Because adipose tissue is known to be a TG regulator, searching for differences between lean and obese subjects may reveal the genes responsible for the phenotypic difference between the two groups. Searching for gene expression differences in people with extreme lipid values has also lead to successful identification of lipid genes and gene variants. Using weighted gene co-expression network analysis (WGCNA), Plaisier et al. discovered several novel FCHL genes by analyzing adipose tissue gene expression network differences between FCHL cases and controls (50).

Figure 1.3 Schematic Overview of Weighted Gene Coexpression Network Analysis (WGCNA) (modified from Langfelder et al. (51)).



WGCNA is a popular method of analyzing gene expression differences between two groups (51,52). WGCNA places genes into groups called modules based on similar expression patterns between samples, and then WGCNA tests for association between a module and a qualitative or quantitative trait. Chapter 3 of this dissertation utilized WGCNA to discover novel TG networks in Finnish and Mexican samples (Haas et al. submitted (3)).

As figure 1.3 summarizes, WGCNA has several benefits over searching for gene expression and TG correlations. One useful aspect of WGCNA is that it places genes into modules based on similar expression patterns across samples. Genes that follow the same expression pattern across samples are more likely to be involved in similar biological processes. Thus, the grouping WGCNA performs may reveal novel direct or indirect gene interactions, whereas searching only for gene expression correlation with a quantitative or qualitative trait gives no information regarding interactions between genes.

Second, WGCNA can determine correlations between an entire module and a trait. In each sample, WGCNA represents the expression of an entire module with a single value called the module eigengene (ME), which can then be tested for correlation with a trait of interest. This module trait association can determine whether an entire biological pathway is correlated with a trait. This test is unique to WGCNA and cannot be done with gene expression and trait correlations alone.

Another advantage WGCNA has over the conventional differential gene expression analysis is that WGCNA can test whether a genetic marker is impacting the expression of an entire module. Network edge orienting (NEO) software can determine whether a SNP may be causing a change in the module expression as a whole (53). With gene-trait correlations, there is no way to determine whether a SNP influences the expression of a group of genes.

1.4 Variant rs7575840 Influences Apolipoprotein B-Containing Lipid Particles

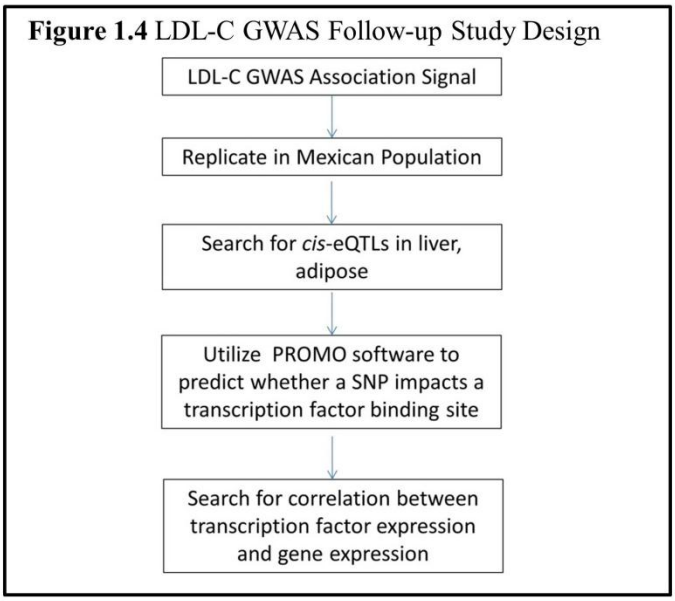
High serum LDL-C is a well-established risk factor for coronary heart disease (54). LDL-C GWAS have successfully identified genomic regions of interest (11), but determining which variant in a region may be causal is difficult to ascertain from a GWAS alone, as GWAS do not provide functional information to break-down large LD regions. These large LD blocks often span several genes, making it difficult to determine which gene is involved in LDL-C regulation.

Determining the specific lipid sub-particles an LDL-C GWAS association signal impacts may lead to a better understanding of the mechanism underlying an LDL-C association signal. A recent mega-GWAS study tested only LDL-C, HDL-C, and TG, leaving out several lipid sub-particles, including apolipoprotein B (apoB), very low density lipoprotein (VLDL) and intermediate density lipoprotein (IDL). ApoB is strongly correlated with CHD progression (55) and is a major component of LDL-C particles. Additionally, the sizes of lipoproteins are not routinely tested in lipid GWAS, yet current research suggests that small LDL particles are strongly correlated with CHD progression (54). Thus, discovering an association between a genetic variant and lipid sub-particles would provide a possible causal mechanism underlying the GWAS association signal.

LDL-C GWAS signals can be influencing serum LDL-C levels by acting as a *cis*-eQTL. Thus, determining whether LDL-C GWAS variants are *cis*-eQTLs in tissues relevant to lipid regulation may reveal the causal mechanism underlying the GWAS association. Adipose tissue is known to be a regulator of lipid levels in humans, so the possibility of an LDL-C GWAS variant impacting serum LDL-C levels through influencing gene expression in adipose tissue is worth exploring.

One way a genetic variant may affect gene expression is through disrupting a transcription factor binding site (TFBS). A SNP may impact the transcription factor's binding affinity to a specific DNA sequence. The software program PROMO is commonly used to determine whether a SNP may influence a TFBS. Figure 1.4 depicts the overall study design employed in Haas et al. to follow-up the LDL-C GWAS variant near *APOB* (1).

Figure 1.4 LDL-C GWAS Follow-up Study Design



1.5 Adipose Transcript Networks Identify Novel Triglyceride Genes

Serum TG levels are known to be an independent risk factor for CHD (17-19). Adipose tissue is a key regulator of serum TG levels by controlling TG storage in adipose tissue. Searching for novel genes involved in regulating TG levels may lead to a better biological understanding of hypertriglyceridemia. We used three different adipose RNA study samples derived from two distinct populations, Mexicans and Finns, to identify novel TG genes and adipose transcript networks utilizing WGCNA (3). The gene expression was measured by microarrays in two of the study samples (Mexican TG Cases/Controls and Finnish Twin Cohort) and by RNAseq in one of the study sample (METSIM TG Cases/Controls) (Table 1.5).

Mexicans have a high prevalence of hyperlipidemia, with 44% of the population having hypercholesterolemia (2), making Mexicans a particularly important population to study. Mexicans are also an understudied population, with the vast majority of lipid GWAS focusing on European origin populations.

Table 1.5 Clinical Characteristics of Mexican and Finnish Study Samples

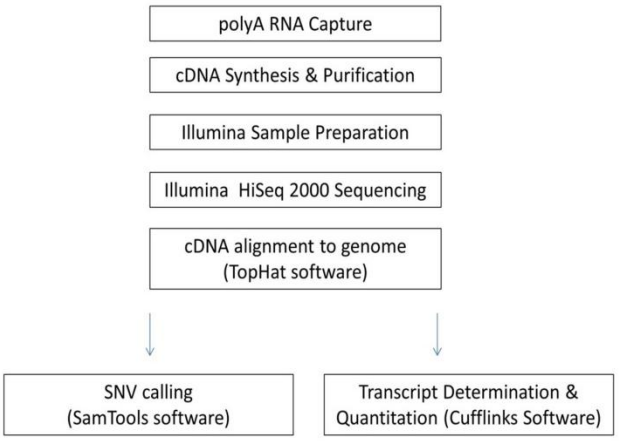
Trait	Mexican TG Cases/Controls	Finnish Twin Cohort	METSIM TG Cases/Controls
n (Male)	70 (36)	53 (25)	20 (20)
Age	37.8 ± 9.2	28.9 ± 4.2	54.1 ± 4.1
TG (mmol/L)	3.29 ± 2.54	1.08 ± 0.37	2.07 ± 1.94
BMI	25.9 ± 2.7	27.7 ± 4.2	27.2 ± 3.7
TC (mmol/L)	5.78 ± 1.47	4.39 ± 0.74	5.74 ± 0.95
HDL (mmol/L)	1.15 ± 0.28	1.40 ± 0.37	1.6 ± 0.74
Fasting Glucose (mmol/L)	5.32 ± 1.09	5.25 ± 0.50	5.84 ± 0.71
Fasting Insulin (mmol/L)	12.24 ± 8.49	6.53 ± 3.80	7.89 ± 4.36

Searching for genes associated with TGs across multiple populations may reveal novel TG regulators. Different populations often have unique LD patterns and minor allele frequencies at

various SNPs, however, genes involved in regulating lipid levels are expected to be consistent across populations. Therefore, searching for genes associated with TG levels in multiple populations will provide strong evidence that the gene is truly involved in TG regulation. Table 1.5 displays the clinical characteristics of the three population samples used in our study (3).

WGCNA has been a successful method at finding groups of genes and pathways that are biologically relevant to a qualitative or quantitative trait. Utilizing WGCNA has previously lead to novel gene discoveries involved in regulating lipid levels (50), making this method an excellent tool to discover novel TG genes with adipose gene expression data.

Figure 1.5 RNA Sequencing Experiment Pipeline



The RNAseq data analysis pipeline used in Haas et al. (3). HiSeq 2000 is a sequencing platform by Illumina.

Figure 1.5 illustrates the study design used to analyze RNAseq data in our study (3). Basically, mRNA is captured and prepped for sequencing using the Illumina mRNA Sequencing Kit, and the resulting cDNA is sequenced on the HiSeq 2000 sequencing platform. HiSeq2000 is a recently developed sequencing platform by Illumina that sequences more reads per lane than the

previous sequencer Genome Analyzer IIx by Illumina. HiSeq 2000 can sequence up to 2 billion paired-end 100 basepair reads per flow cell (8 sequencing lanes), while the Genome Analyzer IIx can obtain 640 million paired-end 150 basepair reads in a flow cell. Tophat (56,57), a software tool developed specifically for RNAseq data, is used to align the cDNA to the reference genome, and Cufflinks (58-60), the sister program of Tophat, is then utilized to determine and quantitate the transcripts. Finally, Samtools (61), a popular single nucleotide variant (SNV) calling program, is used to call SNVs.

1.6 A PCDH15 SNP is Associated with Familial Combined Hyperlipidemia

FCHL is a common familial lipid disorder found in approximately 20% of individuals with premature CHD (62). Determining novel susceptibility genes and variants is necessary to better understand the largely unknown pathophysiology of this prevalent lipid disorder. The genetic characterization of FCHL may also ultimately help prevention and develop better targeted treatments in order to combat premature death in affected FCHL individuals. Determining causal variants in chromosomal regions previously linked to FCHL consistently in multiple populations can lead to a better biological understanding of this common disorder.

PCDH15 is a member of the cadherin superfamily of proteins known to be involved in calcium-dependent cell adhesion. A linkage peak near *PCDH15* was previously reported for FCHL (5-7). In fact the key microsatellite marker, D10S546, resides within the *PCDH15* gene, making it our prime candidate under the linkage peak. No prior genome-wide significant GWAS signals for TG or TC have been identified for *PCDH15* in prior GWAS studies (11). However, lipid GWAS have been performed on unrelated mostly normolipidemic individuals, which leaves the possibility that rare or common *PCDH15* variants may be responsible for FCHL (or its metabolic component traits) in large, extended dyslipidemic families. Therefore, to test whether nonsynonymous variants in this gene may be impacting TG or TC levels through disrupting PCDH15 function, we tested whether 4 common nonsynonymous *PCDH15* variants are associated with lipid levels in 92 Finnish and Dutch FCHL families.

The mouse has served as an important modeling tool for understanding lipid regulation. Knocking out a gene of interest in mice and then measuring lipid levels can provide direct evidence that a specific gene is involved in lipid regulation *in vivo*. Thus, we created a *PCDH15* loss of function knockout mouse line in collaboration with investigators at Case Western Reserve University and measured the resulting lipid levels to determine whether knocking out *PCDH15* may impact lipid levels.

CHAPTER 2:

EVIDENCE OF HOW RS7575840 INFLUENCES APOLIPOPROTEIN B-CONTAINING LIPID PARTICLES

This work has been published and is freely accessible beginning May 1, 2012 in PubMed Central. PMCID # 3081410

Permission has been granted to reprint the publication (1) in this dissertation.

Evidence of mechanism how rs7575840 influences apolipoprotein B containing lipid particles

Blake E. Haas[1], **Daphna Weissglas-Volkov**[1], **Carlos A. Aguilar-Salinas**[2], **Elina Nikkola**[1], **Laurent Vergnes**[1], **Ivette Cruz-Bautista**[2], **Laura Riba**[3], **Alena Stancakova**[4], **Johanna Kuusisto**[4], **Pasi Soininen**[4], **Antti J. Kangas**[5], **Mika Ala-Korpela**[4,5], **Teresa Tusie-Luna**[3], **Markku Laakso**[4], and **Päivi Pajukanta**[1,*]

[1]Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, California, USA [2]Department of Endocrinology and Metabolism, Instituto Nacional de Ciencias Médicas y Nutrición, Salvador Zubiran, Mexico City, Mexico [3]Molecular Biology and Genomic Medicine Unit, Instituto de Investigaciones Biomédicas de la UNAM, Instituto Nacional de Ciencias Médicas y Nutrición, Salvador Zubiran, Mexico City, Mexico [4]University of Eastern Finland, Kuopio, Finland [5]Computational Medicine Research Group, University of Oulu and Biocenter Oulu, Oulu, Finland

Abstract

Objective—Recent genome-wide association studies (GWAS) identified a variant rs7575840 in the apolipoprotein B (APOB) gene region to be associated with LDL-C. However, the underlying functional mechanism of this variant that resides 6.5 kb upstream of APOB has remained unknown. Our objective was to investigate rs7575840 for association with refined apoB containing lipid particles; for replication in a non-Caucasian Mexican population; and for underlying functional mechanism.

Methods and Results—Our data show that rs7575840 is associated with serum apoB levels ($P=4.85\times10^{-10}$) and apoB containing lipid particles, very small VLDL, IDL and LDL particles ($P=2\times10^{-5}$ - 9×10^{-7}) in the Finnish METSIM study sample (n=7,710). Fine mapping of the APOB region using 43 SNPs replicated the association of rs7575840 with apoB in a Mexican study sample (n=2,666, $P=3.33\times10^{-05}$). Furthermore, our transcript analyses of adipose RNA samples from 175 Finnish METSIM subjects indicate that rs7575840 alters expression of APOB ($P=1.13\times10^{-10}$) and a regional non-coding RNA (BU630349) ($P=7.86\times10^{-6}$) in adipose tissue.

Conclusions—It has been difficult to convert GWAS associations into mechanistic insights. Our data show that rs7575840 is associated with serum apoB levels and apoB containing lipid particles as well as influences expression of APOB and a regional transcript BU630349 in adipose tissue. We thus provide evidence how a common genome-wide significant SNP rs7575840 may affect serum apoB, LDL-C, and TC levels.

Keywords

Apolipoprotein B; association analysis; gene expression; adipose tissue; Mexicans

*Address Correspondence to: Paivi Pajukanta, MD, PhD, Professor, Department of Human Genetics, David Geffen School of Medicine at UCLA, Gonda Center, Room 6335B, 695 Charles E. Young Drive South, Los Angeles, California 90095-7088, USA, ppajukanta@mednet.ucla.edu .

Disclosures

None.

INTRODUCTION

Genome-wide association studies (GWAS) have been successful in discovering single nucleotide polymorphisms (SNPs) that affect LDL-C levels with low to moderate effect sizes[1]. However, many of the significantly associated SNPs are intergenic and belong to large linkage disequilibrium (LD) blocks, making it difficult to evaluate their functional relevance.

The SNP rs7575840 has been implicated in a recent GWAS for LDL-C[1]. As rs7575840 resides in the APOB gene region, 6.5 kb upstream of the gene, we tested it for association with refined lipid phenotypes and proton NMR spectroscopy measurements of lipid particles in a Caucasian population sample, the METabolic Syndrome in Men (METSIM) cohort (n=7,710). Furthermore, to also explore the role of this GWAS variant in a non-Caucasian population, we fine mapped the APOB region in a Mexican dyslipidemic case/control study sample and compared the regional LD between Caucasians and Mexicans. The Mexican population has a 44% prevalence of hypercholesterolemia defined by total cholesterol (TC) > 200mg/dL[2]. Other forms of dyslipidemias are highly common in this population as well. Yet to the best of our knowledge, no GWAS has been performed to date exploring lipid levels in Mexicans. Nor have the Caucasian GWAS signals for LDL-C been explored thoroughly in the Mexican population as of yet. Thus, Mexicans represent a population with a high susceptibility to dyslipidemia underinvestigated for the underlying genetic factors2.

ApoB is the sole protein component of the LDL-C particle and an essential core of triglyceride-rich lipoproteins. A mutation in APOB is known to cause a monogenic autosomal dominant disease called familial hypercholesterolemia[3], further establishing the role of APOB in LDL metabolism. We hypothesized that rs7575840 and/or SNPs in LD with it may function as a regulatory *cis*-eQTL, and tested 175 adipose RNA samples from the Finnish METSIM study for allele-specific expression of transcripts within the APOB gene region.

METHODS

The study design was approved by the ethics committees of the participating centers and all subjects gave a written informed consent.

Study samples

METSIM cohort—The Finnish population-based cohort, METSIM (METabolic Syndrome In Men), was collected at the University of Kuopio, Kuopio, Finland as described previously4. The METSIM cohort consisted of 7,710 male subjects, age 50-70 years, randomly selected from the population of Kuopio in Eastern Finland. Each patient participated in an interview session to measure factors that may affect cardiovascular disease risk, including prescription drug use, weekly exercise, metabolites, cardiovascular family history, and other health questions. Phenotypic determinations were performed as described previously4. Plasma lipoproteins from the METSIM subjects were fractionated using proton NMR into HDL, LDL, intermediate lipoproteins (IDL), very low density lipoproteins, and chylomicron subclasses, as described previously5·6.

Mexican hypertriglyceridemia cases and controls—A total of 2,666 Mexican hypertriglyceridemic cases and controls were recruited at the Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, as described previously7. Briefly, the inclusion criteria were fasting serum triglycerides > 2.3 mmol/L (200 mg/dL) for the cases and < 1.7 mmol/L (150 mg/dL) for the controls. Exclusion criteria were type 2 diabetes mellitus or morbid obesity (body mass index (BMI) > 40 kg/m2), and the use of lipid lowering drugs for

the controls. Fasting lipid levels were measured using commercially available standardized methods as described previously7. Serum LDL-C levels were calculated using the Friedewald formula for subjects with TG < 400 mg/dL.

Genotyping and Real-time PCR

The SNP rs7575840 was genotyped in the METSIM study sample using the Sequenom genotyping platform. The 43 SNPs in the APOB region (+/−100 kb) were genotyped in the Mexican study sample using both Pyrosequencing and Illumina BeadArray technology platforms. The SNPs were in Hardy-Weinberg Equilibrium (p-value>0.05) in both study samples and had a genotyping call rate >90%. The real-time PCR of 175 METSIM RNAs for BU630349 was performed using Quantitect Reverse Transcription kit (Qiagen), as described in detail in the Supplementary methods.

Tagging of the APOB region in the Mexican study sample

We genotyped 43 SNPs for the APOB region (+/−100 kb) in 2,310 subjects of the Mexican dyslipidemic study sample. Using Haploview v4.2[8] and CEU HapMap data, we calculated that these 43 SNPs capture 90% of the variation with minor allele frequency (MAF) ≥5% and $r^2 \geq 0.80$. We utilized the CEU HapMap data for these calculations as the coverage of the Mexican American HapMap data is still incomplete.

Microarray data and gene expression analysis in METSIM fat biopsy samples

A total of 175 METSIM subjects underwent subcutaneous fat biopsies for adipose RNA isolation. The RNA isolation was performed according to the manufacturer's instructions (Qiagen). The adipose RNA samples were hybridized to the Illumina HT-12 v3.0 expression chips. Genome Studio was used to perform quantile normalization and background subtraction. An absent call was made with the detection p-values>0.01. Probes with >50% absent calls were excluded from the analysis. All METSIM subjects were unrelated males. Microarray data for the transcripts in the APOB gene region (+/−500 kb) will be submitted to the NCBI's Gene Expression Omnibus repository in MIAME compliant format (GSE27666).

To test whether the APOB or BU630349 expression in 175 METSIM fat biopsies are influenced by the rs7575840 genotypes, we log transformed the APOB and BU630349 expression values. Expression values more than 4 standard deviations from the mean were removed form the analysis. The two-sided Student's T-test was performed to compare the means between carriers of the T rare allele and the homozygous common G group (i.e. dominant model). The ECR Vista browser was used to look at the conservation of BU630349 across 13 species.

Association analysis in METSIM

To test for association between the SNPs and lipid traits, and lipoprotein subclasses, we performed multivariate linear regression analysis for the additive genetic model using SPSS software. Trait values were adjusted for age and log transformed BMI. Subjects with BMI>40, lipid lowering medication, diabetes, or a trait value greater than 4 standard deviations from the mean were excluded from the analysis, leaving 5,054 METSIM individuals in the association analyses. Pearson correlation analyses were performed between the raw lipid particle measurements and gene expression (APOB and BU630349). The gene expression values were log transformed, and values greater than 4 standard deviations from the mean were removed.

Association analysis in Mexican hypertriglyceridemic cases and controls

To unify the association analysis with the METSIM analysis, we performed the multivariate linear regression analysis using SPSS software by adjusting the trait values for age, sex, log transformed BMI, and hypertriglyceridemia case-control status. Subjects with BMI>40, lipid lowering medication, diabetes, or a trait value greater than 4 standard deviations from the mean were excluded from the statistical analysis. Bonferroni correction (43 tested SNPs; $P<0.001$) was employed to evaluate significance. We did not correct for three tested traits, as apoB, LDL-C and TC are known to be highly correlated.

Individual ancestry (IA) analyses

We had genotype data for 82 evenly distributed ancestry informative markers (AIMs) for 2,310 of the Mexican hypertriglyceridemia case-control samples7. These AIMs were selected based on a published list of European/Amerindian AIMs9 and were used to calculate individual ancestry (IA) estimates using the STRUCTURE 2.2 software10, as described in the Supplementary methods.

RESULTS

The SNP rs7575840 resides in the APOB region, 6.5 kb upstream of APOB. To further investigate the association signal of rs7575840, we first tested the SNP for association with refined lipid levels and proton NMR spectroscopy measurements of lipid particles in the Finnish population sample METSIM. Table 1 shows that the rare allele T (MAF=28%) of rs7575840 is significantly associated with elevated apoB, LDL-C and TC levels in METSIM (n=7,710), apoB providing the most significant evidence of association ($P=4.85\times10^{-10}$). The proportion of 1 SD change in standardized apoB residual values for each copy of the risk allele was 0.14, and the contribution of this genetic effect can explain 0.8% of the total variance of apoB levels in METSIM (Table 1). The most associated lipid particles were the very small VLDL, IDL as well as all of the LDL subclasses ($P=2\times10^{-5}$-9×10^{-7}) (Supplementary Table 1). Among the tested lipoprotein subclasses, rs7575840 explains the most variance for these same subclasses (Supplementary Figure 1). Furthermore, the strongest signals classified by positive beta were also observed for the VLDL, IDL and LDL subclasses (Beta=8×10^{-3}-3×10^{-2}) (Supplementary Table 1).

Different populations may have different signals due to the underlying differences in LD. To investigate whether the effect of rs7575840 on serum apoB levels also extends to a non-Caucasian population, we first fine mapped the APOB gene region (±100 kb) by genotyping 43 SNPs in the Mexican dyslipidemic study sample (n=2,310). By utilizing the Caucasian HapMap CEU samples and the Mexican apoB controls (apoB levels <50th age/sex specific Mexican population percentile), we first observed that the LD pattern in Caucasians closely resembles the LD in Mexicans (Supplementary figure 2). When testing the 43 SNPs for association with apoB, we observed 4 non-redundant independent signals ($r^2<0.6$) passing the Bonferroni correction ($P<0.001$) (Supplementary Table 2, Supplementary figure 2). Among the 43 tested SNPs, rs693, rs1367117, and rs7575840, that have been implicated in previous GWAS for LDL-C[12,13], provided strong evidence of association with apoB, though rs1367117 and rs7575840 represent redundant signals, as described below. The p-values for these three SNPs were ranked in the top 5 for all of the three tested traits, apoB, LDL-C, and TC (Supplementary Tables 2-4). To further validate these associations, we extended the genotyping of rs693, rs1367117, and rs7575840 to 356 additional Mexican dyslipidemic samples available for study (total n=2,666, Table 2), which further strengthened the association signals for apoB (p-values of 2.90×10^{-06}, 6.83×10^{-06}, and 3.33×10^{-05}; explaining 0.74%, 0.90%, and 0.76% of the apoB levels, respectively). The LDL-C and TC traits provided somewhat less significant p-values (Table 2). Interestingly, rs1367117, is a

nonsynonymous SNP (T71I) in APOB. However, according to VISTA browser the amino acid T71 is only conserved in mammals, and furthermore, the amino acid change T71I was previously predicted to be benign using the PolyPhen, SIFT, and PANTHER algorithms[11]. The SNP rs693 is a synonymous variant. It is worth noting that the association signals were all for the same risk allele as in Caucasians. Hence, our results suggest that these Caucasian GWAS signals extend to the Mexican population.

We next investigated the pairwise LD of rs7575840 with all other SNPs in the APOB gene region (±100 kb) using the Caucasian HapMap CEU sample and Mexican apoB controls (Supplementary figure 3). We observed that the regional differences in pairwise LD between the two populations are not large (Supplementary figure 3). The SNP rs693 was not in strong LD with rs7575840 or rs1367117 (r^2 <0.55 both in HapMap CEU and Mexican apoB controls). The pairwise LD (using r^2) between rs7575840 and rs1367117 in the Mexican apoB controls was 0.80 and in the CEU HapMap sample 0.91, indicating that in both populations rs7575840 and rs1367117 are proxies, likely reflecting the same functional signal.

Mexicans are an admixed population that descends from a recent mix of Amerindian and European ancestry with a small proportion of African ancestry14·[15]. Population admixture may confound allelic association if both the trait distribution and the allele frequency differ between ancestries. We previously demonstrated that the apoB distribution did not differ with individual ancestry (IA) estimates in 2,310 subjects of the Mexican hypertriglyceridemia cases and controls[5]. However, to further eliminate the possibility of spurious associations due to admixture, we also performed the association analyses of rs7575840, rs1367117, and rs693 while including IA estimates as a covariate in the regression model. In these adjusted analyses, we obtained P-values of 3.76×10^{-05}, 6.55×10^{-06} and 1.30×10^{-06} for residual apoB levels, respectively, suggesting that the associations are not confounded by population admixture.

As rs7575840 resides 6.5 kb upstream of APOB, it may modify regulatory elements. To investigate whether rs7575840 affects the expression of APOB and/or transcripts in the APOB gene region, we searched for gene expression probes within 500 kb of APOB. The APOB probe ILMN_1664024, which corresponds with the APOB refseq ID NM_000384, was expressed in the 175 METSIM samples. No other expressed probes were observed within 500 kb of APOB on the microarray. We also tested a local transcript BU630349 for differential expression between the rs7575840 genotypes by RT-PCR, because it was previously found to be significantly differentially expressed based on the rs7575840 genotypes (i.e. cis-eQTL) in liver ($P=1.62\times10^{-54}$), passing a multiple testing correction of FDR <10%[16]. No effect was observed between the rs7575840 genotypes and APOB expression in liver[16]. BU630349 is a 662-bp long predicted non-coding RNA located 872 bp upstream of APOB and conserved in primates. To ensure we did not obtain spurious associations due to a small number of homozygous individuals, we utilized a dominant model to test for cis-eQTLs. The METSIM subjects carrying at least one minor allele of rs7575840 had significantly more APOB expression ($P=1.13\times10^{-10}$) and BU630349 expression ($P=7.86\times10^{-06}$) in adipose tissue relative to the homozygous common group (figures 1-2). The fold change between the two groups was 1.17X for APOB and 1.78X for BU630349, rs7575840 explaining 21.5% and 11.2% of the APOB and BU630349 variance, respectively (figures 1-2). The associations were also significant using an additive model ($P=5.37\times10^{-09}$ for APOB, $P=2.18\times10^{-06}$ for BU630349). The detected APOB differential expression is in line with the observation that individuals with the rare allele of rs7575840 have higher serum apoB levels (Table 1).

To determine whether expression of APOB or BU630349 may be related to lipoprotein subclasses, we performed Pearson correlations in the 175 METSIM fat biopsy samples. Both expression of APOB and BU630349 were positively correlated with HDL-C concentration/particles and negatively correlated with triglycerides and VLDL concentration/particles (Supplementary figure 4). We also observed that raw serum apoB levels were negatively correlated with APOB expression in adipose tissue ($p=0.005$, $r=-0.21$).

To search for a possible mechanism of how rs7575840 influences APOB and BU630349 expression, we performed in silico analysis using the PROMO program to search for transcription factor binding sites (TFBS) that differ based on rs7575840 genotype. We found that the major allele uniquely codes for a TFBS for CEBPA and CEBPB transcription factors, and the minor allele codes for a YY1 TFBS. CEBPA expression was positively correlated with APOB ($r=0.50$, $P=1.57\times10^{-12}$) and BU630349 expression ($r=0.21$, $P=6.91\times10^{-03}$) in the 175 METSIM adipose RNA samples. This is in line with the fact that APOB and BU630349 expression were also correlated in adipose ($r=0.49$, $p=1.46\times10^{-11}$). After using the rs7575840 genotype as a covariate, the correlations between CEBPA and APOB were stronger and about 1000 times more significant ($r=0.56$ and $P=8.38\times10^{-16}$ for additive model; $r=0.58$ and $P=6.92\times10^{-17}$ for dominant model). The correlation between CEBPA and BU630349 were slightly stronger when correcting for rs7575840 genotype ($r=0.24$ and $P=1.97\times10^{-03}$ for additive genotype; $r=0.23$ and $P=2.42\times10^{-03}$ for dominant genotype). SNP rs7575840 was not significantly associated with CEBPB or YY1 expression. CEPBA and CEBPB were positively correlated with each other ($r=0.37$, $p=3.30\times10^{-07}$). Taken together, our in silico and gene expression data suggest that rs7575840 may affect APOB and BU630349 expression in adipose by influencing a CEBPA binding site.

DISCUSSION

Our data in two different populations, Finns and Mexicans, implicate the rs7575840 variant near the APOB gene for apoB, LDL-C and TC levels, the most affected traits in our analyses being apoB and apoB containing lipid particles. Furthermore, our novel data also revealed a possible functional mechanism how rs7575840 influences apoB, LDL-C and TC levels by altering the expression level of the APOB gene itself and a regional non-coding RNA (BU630349) in adipose tissue. The latter has also been shown to be differentially expressed by rs7575840 in 427 liver samples[16]. A nonsynonymous SNP, rs1367117, that has also been identified in a recent GWAS meta-analysis for LDL-C1 is in high LD with rs7575840 both in Caucasians and Mexicans. Our functional findings suggest a mechanism how these redundant, genome-wide significant GWAS signals, rs7575840 and rs1367117, may alter apoB levels.

Variants in APOB have been investigated for associations with lipids in multiple previous candidate gene studies mostly based on Caucasian populations[11,17,18]. In a recent Caucasian mega meta-analysis GWAS of mainly population-based samples, the amino acid change rs1367117 in APOB resulted in a p-value of 4.48×10^{-114} for LDL-C in 95,402 subjects, whereas rs7575840 (LD=0.9 by r^2) resulted in a p-value of 1.67×10^{-98} for LDL-C in 41,912 subjects[1]. It is important to confirm whether the known loci have a consistent effect across ethnic groups to determine which variants are to be used in cardiovascular risk assessment. Furthermore, despite the high prevalence of high LDL-C (46%) in Mexicans[2], little is known about the genetic factors predisposing to high cholesterol levels in this non-Caucasian population. Our trans-ethnic fine mapping shows the first replication of these two GWAS variants and another previous GWAS variant, rs69312, in the Mexican population with a high susceptibility to dyslipidemia.

The recent Caucasian mega GWAS study[1] focused on TC and LDL-C, whereas apoB levels were not available for the study. Although we also obtained significant association evidence for TC and LDL-C, we observed the strongest association signal with apoB. ApoB is the sole protein component of LDL particles and present in most atherogenic lipoprotein particles. Previous prospective epidemiological studies have clearly shown that apoB levels predict the CAD risk better than LDL-C and TC levels[19]. Our study demonstrates that rs7575840 is associated with VLDL, IDL, and LDL subclasses. These data further implicate the role of rs7575840 and/or SNPs in LD with it in apoB metabolism.

Discovering the underlying mechanisms behind GWAS signals has been very difficult, largely due to large linkage disequilibrium blocks, spanning many genes, and the absence of functional evidence provided by the GWAS. To search for a possible function of rs7575840, we utilized one of the largest subcutaneous fat biopsy gene expression cohorts published to date (n=175). Because adipose tissue is an important regulator of lipid metabolism in humans, we decided to determine whether rs7575840 genotype alters gene expression in this tissue. We discovered that rs7575840 is associated with APOB and BU630349 expression in adipose tissue, providing possible regulatory mechanisms how this SNP may alter apoB, LDL-C and TC levels. Importantly, the same cis-eQTL between rs7575840 and BU630349 has previously been observed in liver, although the APOB expression did not differ between the rs7575840 genotypes[16]. Differences in the used microarray platforms may have contributed to these discrepancies. However, even though the hepatic APOB expression did not differ by the rs7575840 genotypes, the regional non-coding RNA BU630349 may still serve as a post-transcriptional regulator of APOB, because besides expression long non-coding RNAs may affect splicing, transport, translation, and epigenetic regulation of genes20. Interestingly, we also found that the binding site of a transcription factor, CEBPA, is predicted to differ based on the rs7575840 genotypes, and its expression was positively correlated with APOB and BU630349 expression in the 175 METSIM adipose RNA samples. CEBPA is highly expressed both in adipose tissue and liver; known to modulate leptin and adiponectin expression; and induces expression of genes involved in differentiation of granulocytes, monocytes, adipocytes and hepatocytes21·22. Future studies are warranted to determine the detailed molecular mechanisms how rs7575840 or SNPs in LD with it ultimately impact apoB levels. Nevertheless, taken together our data provide the first piece of evidence how rs7575840 influences apoB and apoB containing lipid particles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Finnish and Mexican individuals who participated in this study. We also thank Cindy Montes, Maribel Rodríguez-Torres, and Salvador Ramírez for laboratory technical assistance.

Sources of funding

This research was supported by the grants HL095056 and HL-28481 from the National Institutes of Health. B.E.H. is supported by NHGRI grant T32 HG02536. This work has also been supported by the Academy of Finland's Responding to Public Health Challenges Research Programme (grant 129429 to MAK), and the Finnish Cardiovascular Research Foundation (MAK), and the Jenny and Antti Wihuri Foundation (AJK).

REFERENCES

1. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M,

Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, König IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Cecile JW, Janssens A, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Döring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA Jr, Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010;466:707–713.

2. Aguilar-Salinas CA, Gómez-Pérez FJ, Rull J, Villalpando S, Barquera S, Rojas R. Prevalence of dyslipidemias in the Mexican National Health and Nutrition Survey 2006. Salud Publica Mex. 2010;52:S44–53.
3. Innerarity TL, Weisgraber KH, Arnold KS, Mahley RW, Krauss RM, Vega GL, Grundy SM. Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding. Proc Natl Acad Sci U S A. 1987;84:6919–6923. [PubMed: 3477815]
4. Lee JC, Weissglas-Volkov D, Kyttälä M, Dastani Z, Cantor RM, Sobel EM, Plaisier CL, Engert JC, van Greevenbroek MM, Kane JP, Malloy MJ, Pullinger CR, Huertas-Vazquez A, Aguilar-Salinas CA, Tusie-Luna T, de Bruin TW, Aouizerat BE, van der Kallen CC, Croce CM, Aqeilan RI, Marcil M, Viikari JS, Lehtimäki T, Raitakari OT, Kuusisto J, Laakso M, Taskinen MR, Genest J, Pajukanta P. WW-domain-containing oxidoreductase is associated with low plasma HDL-C levels. Am J Hum Genet. 2008;83:180–192. [PubMed: 18674750]
5. Soininen P, Kangas AJ, Würtz P, Tukiainen T, Tynkkynen T, Laatikainen R, Järvelin MR, Kähönen M, Lehtimäki T, Viikari J, Raitakari OT, Savolainen MJ, Ala-Korpela M. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. Analyst. 2009;134:1781–1785. [PubMed: 19684899]
6. Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, Hämäläinen E, Jousilahti P, Kangas AJ, Männistö S, Savolainen MJ, Palotie A, Salomaa V, Perola M, Ala-Korpela M, Peltonen L. Metabonomic, transcriptomic, and genomic variation of a population cohort. Mol Syst Biol. 2010;6:441. [PubMed: 21179014]
7. Weissglas-Volkov D, Plaisier CL, Huertas-Vazquez A, Cruz-Bautista I, Riaño-Barros D, Herrera-Hernandez M, Riba L, Cantor RM, Sinsheimer JS, Aguilar-Salinas CA, Tusie-Luna T, Pajukanta P. Identification of Two Common Variants Contributing to Serum Apolipoprotein B Levels in Mexicans. Arterioscler Thromb Vasc Biol. 2009;30:353–359. [PubMed: 19965785]
8. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263–265. [PubMed: 15297300]
9. Tian C, Hinds DA, Shigeta R, Adler SG, Lee A, Pahl MV, Silva G, Belmont JW, Hanson RL, Knowler WC, Gregersen PK, Ballinger DG, Seldin MF. A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. Am J Hum Genet. 2007;80:1014–1023. [PubMed: 17557415]

10. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–959. [PubMed: 10835412]
11. Benn M, Nordestgaard BG, Jensen JS, Grande P, Sillesen H, Tybjaerg-Hansen A. Polymorphism in APOB associated with increased low-density lipoprotein levels in both genders in the general population. J Clin Endocrinol Metab. 2005;90:5797–5803. [PubMed: 16030169]
12. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet. 2008;40:189–197. [PubMed: 18193044]
13. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, Strait J, Duren WL, Maschio A, Busonero F, Mulas A, Albai G, Swift AJ, Morken MA, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen WM, Li Y, Scott LJ, Scheet PA, Sundvall J, Watanabe RM, Nagaraja R, Ebrahim S, Lawlor DA, Ben-Shlomo Y, Davey-Smith G, Shuldiner AR, Collins R, Bergman RN, Uda M, Tuomilehto J, Cao A, Collins FS, Lakatta E, Lathrop GM, Boehnke M, Schlessinger D, Mohlke KL, Abecasis GR. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet. 2008;40:161–169. [PubMed: 18193043]
14. Salari K, Burchard E. Latino populations: a unique opportunity for epidemiological research of asthma. Paediatr Perinat Epidemiol. 2007;21(Suppl 3):15–22. [PubMed: 17935571]
15. Price A, Patterson N, Yu F, et al. A genomewide admixture map for Latino populations. Am J Hum Genet. 2007;80:1024–1036. [PubMed: 17503322]
16. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008;6:e107. [PubMed: 18462017]
17. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, Hirschhorn JN, Berglund G, Hedblad B, Groop L, Altshuler DM, Newton-Cheh C, Orho-Melander M. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med. 2008;358:1240–1249. [PubMed: 18354102]
18. Ronald J, Rajagopalan R, Ranchalis JE, Marshall JK, Hatsukami TS, Heagerty PJ, Jarvik GP. Analysis of recently identified dyslipidemia alleles reveals two loci that contribute to risk for carotid artery disease. Lipids Health Dis. 2009;8:52. [PubMed: 19951432]
19. Barter PJ, Ballantyne CM, Carmena R, Castro Cabezas M, Chapman MJ, Couture P, de Graaf J, Durrington PN, Faergeman O, Frohlich J, Furberg CD, Gagne C, Haffner SM, Humphries SE, Jungner I, Krauss RM, Kwiterovich P, Marcovina S, Packard CJ, Pearson TA, Reddy KS, Rosenson R, Sarrafzadegan N, Sniderman AD, Stalenhoef AF, Stein E, Talmud PJ, Tonkin AM, Walldius G, Williams KM. Apo B versus cholesterol in estimating cardiovascular risk and in guiding therapy: report of the thirty-person/ten-country panel. J Intern Med. 2006;259:247–258. [PubMed: 16476102]
20. Chen L-C, Carmichael GG. Decoding the function of nuclear long non-coding RNAs. Curr Opin Cell Biol. 2010;22:357–364. [PubMed: 20356723]
21. Miller SG, De Vos P, Guerre-Millo M, Wong K, Hermann T, Staels B, Briggs MR, Auwerx J. The adipocyte specific transcription factor C/EBPalpha modulates human ob gene expression. Proc Natl Acad Sci U S A. 1996;93:5507–5511. [PubMed: 8643605]
22. Segawa K, Matsuda M, Fukuhara A, Morita K, Okuno Y, Komuro R, Shimomura I. Identification of a novel distal enhancer in human adiponectin gene. J Endocrinol. 2009;200:107–116. [PubMed: 18931025]

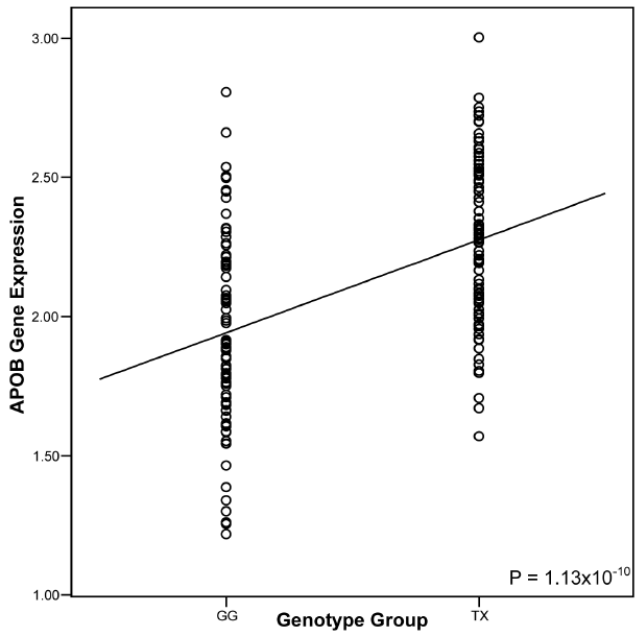


Figure 1.
APOB expression in 175 METSIM fat biopsies based on the rs7575840 genotypes. The TX genotype group (n=85) contains carriers of the minor allele (T), and the GG genotype group (n=90) are homozygous for the common (G) allele. We added a regression line to show the trend of the difference in means.

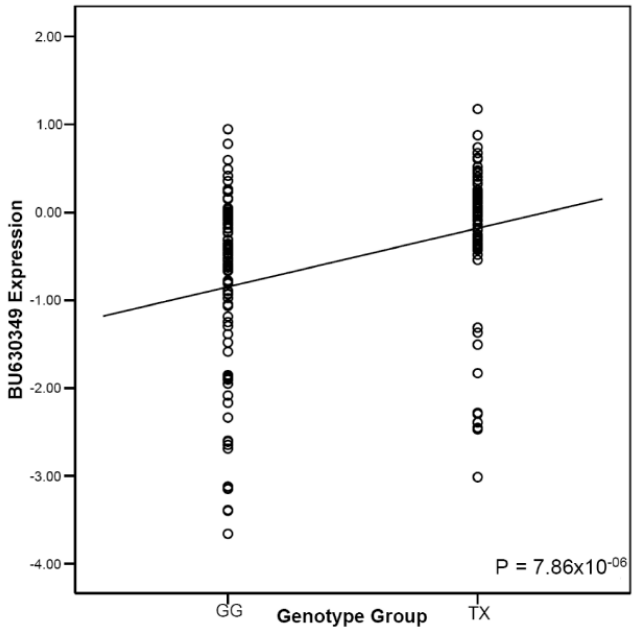


Figure 2.
BU630349 expression in 175 METSIM fat biopsies using the same rs7575840 genotype groups as described in legend to figure 1. We added a regression line to show the trend of the difference in means.

Table 1

Association results of rs7575840 in the METSIM cohort

SNP	Trait	p-value	n	Standardized Effect size		Variance explained	Effect Allele	Effect Allele Freq‡
				B*	SE†			
rs7575840	ApoB	4.85×10^{-10}	5026	0.138	0.022	0.77%	T	0.28
	LDL-C	3.88×10^{-09}	5054	0.131	0.022	0.68%		
	TC	8.05×10^{-09}	5052	0.128	0.022	0.66%		

Linear regression was performed on the METSIM cohort using an additive model. The effect allele is the allele associated with an increased trait value. Variance explained represents the percent of the overall variation in the standardized residuals that can be attributed to the additive model.

*B (beta) represents the proportion of 1 SD change in standardized trait values for each copy of the rare allele

†standard error of beta

‡the frequency of the effect allele.

Table 2

Association results of rs1367117, rs7575840, and rs693 in the Mexican study sample

SNP	Trait	p-value	n	Standardized Effect Size		Variance explained	Effect Allele	Effect Allele Freq
				B	SE			
rs1367117	ApoB	6.83×10^{-06}	2250	0.146	0.032	0.90%	A	0.27
	LDL-C*	6.06×10^{-03}	1551	0.107	0.039	0.49%		
	TC	2.07×10^{-05}	2384	0.135	0.032	0.76%		
rs7575840	ApoB	3.33×10^{-05}	2272	0.134	0.032	0.76%	T	0.28
	LDL-C	2.2×10^{-02}	1563	0.089	0.039	0.34%		
	TC	9.65×10^{-05}	2406	0.123	0.031	0.63%		
rs693	ApoB	2.90×10^{-06}	2312	0.140	0.030	0.94%	A	0.37
	LDL-C	3.86×10^{-03}	1598	0.103	0.036	0.52%		
	TC	6.41×10^{-05}	2446	0.116	0.029	0.65%		

Linear regression was performed in the Mexican study sample using an additive model, as described in the footnote to Table 1.

*Serum LDL-C levels were calculated using the Friedewald formula only for subjects with TGs<400 mg/dL.

CHAPTER 3:

INVESTIGATION OF ADIPOSE TRANSCRIPT NETWORKS ACROSS POPULATIONS IDENTIFIES NOVEL TRIGLYCERIDE GENES

The data in this chapter have been submitted for publication (Haas et al. submitted[1])

Summary of the submitted manuscript

Searching for novel TG-associated biological networks in adipose tissue may provide novel insights into TG regulation. We measured gene expression in adipose samples from two Finnish and one Mexican study sample. In each study sample, we observed a module (biological gene expression network) that was significantly associated with serum TG levels using the weighted gene co-expression network analysis (WGCNA).[2,3] The most significant TG modules observed in each of the three study samples significantly overlapped ($p<10^{-10}$) and shared 34 genes, utilizing two unique methods of measuring gene expression, microarrays and RNAseq. In the 34 genes shared between the three TG modules, more nonsynonymous variants ($p = 0.034$) and overall variants ($p = 0.018$) were observed in individuals with high TGs when compared with the individuals with low TGs. Seven of the 34 genes were identified as the key hub genes of all three TG modules (p<10-7). As only 11 of the 34 genes have prior evidence of involvement in CHD, type 2 diabetes, or obesity, our study provides 23 new candidates for TG regulation. Furthermore, two of the 34 genes (*ARHGAP9, LST1*) reside in previous TG GWAS regions, suggesting them as the regional candidates underlying the GWAS signals. This study presents a novel TG biological network shared across populations.

Chapter 3 References

1. Haas BE, Horvath S, Pietiläinen KH, Cantor RM, Nikkola E, Weissglas-Volkov D, Rissanen A, Civelek M, Cruz-Bautista I, Riba L, Kuusisto J, Kaprio J, Tusie-Luna T, Laakso M, Aguilar-Salinas CA, Pajukanta P. Adipose Transcript Networks Across Finns and Mexicans Identify Novel Triglyceride Genes. Manuscript submitted for publication.

2. Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.

3. Langfelder P, Zhang B, Horvath S. 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 24:719–720.

CHAPTER 4:

A NONSYNONYMOUS SNP WITHIN PCDH15 IS ASSOCIATED WITH LIPID TRAITS IN FAMILIAL COMBINED HYPERLIPIDEMIA

This work has been published and is freely accessible in PubMed Central

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2793376/

Hum Genet (2010) 127:83–89
DOI 10.1007/s00439-009-0749-z

ORIGINAL INVESTIGATION

A nonsynonymous SNP within *PCDH15* is associated with lipid traits in familial combined hyperlipidemia

Adriana Huertas-Vazquez · Christopher L. Plaisier · Ruishuang Geng · Blake E. Haas · Jenny Lee · Marleen M. Greevenbroek · Carla van der Kallen · Tjerk W. A. de Bruin · Marja-Riitta Taskinen · Kumar N. Alagramam · Päivi Pajukanta

Received: 28 May 2009 / Accepted: 21 September 2009 / Published online: 9 October 2009

Abstract Familial combined hyperlipidemia (FCHL) is a common lipid disorder characterized by the presence of multiple lipoprotein phenotypes that increase the risk of premature coronary heart disease. In a previous study, we identified an intragenic microsatellite marker within the protocadherin 15 (*PCDH15*) gene to be associated with high triglycerides (TGs) in Finnish dyslipidemic families. In this study we analyzed all four known nonsynonymous SNPs within *PCDH15* in 1,268 individuals from Finnish and Dutch multigenerational families with FCHL. Association analyses of quantitative traits for SNPs were performed using the QTDT test. The nonsynonymous SNP rs10825269 resulted in a $P = 0.0006$ for the quantitative TG trait. Additional evidence for association was observed with the same SNP for apolipoprotein B levels (apo-B) ($P = 0.0001$) and total cholesterol (TC) levels ($P = 0.001$). None of the other three SNPs tested showed a significant association with any lipid-related trait. We investigated the expression of *PCDH15* in different human tissues and observed that *PCDH15* is expressed in several tissues including liver and pancreas. In addition, we measured the plasma lipid levels in mice with loss-of-function mutations in Pcdh15 (Pcdh15$^{av\text{-}Tg}$ and Pcdh15$^{av\text{-}3J}$) to investigate possible abnormalities in their lipid profile. We observed a significant difference in plasma TG and TC concentrations for the Pcdh15$^{av\text{-}3J}$ carriers when compared with the wild type ($P = 0.013$ and $P = 0.044$, respectively). Our study suggests that *PCDH15* is associated with lipid abnormalities.

A. Huertas-Vazquez
Heart Institute, Cedars-Sinai Medical Center,
Los Angeles, CA, USA
e-mail: huertasa@cshs.org

C. L. Plaisier · B. E. Haas · J. Lee · P. Pajukanta (✉)
Department of Human Genetics, University of California,
David Geffen School of Medicine at UCLA, Gonda 6335B,
695 Charles Young Drive South, Los Angeles, CA 90095, USA
e-mail: ppajukanta@mednet.ucla.edu

R. Geng · K. N. Alagramam
Department of Otolaryngology Head and Neck Surgery,
University Hospitals Case Medical Center,
Case Western Reserve University, Cleveland, OH, USA

M. M. Greevenbroek · C. van der Kallen · T. W. A. de Bruin
Department of Internal Medicine and the Cardiovascular Research
Institute Maastricht, Maastricht University, Maastricht,
The Netherlands

M.-R. Taskinen
Helsinki University Central Hospital,
University of Helsinki, Helsinki, Finland

Introduction

Familial combined hyperlipidemia (FCHL) is a complex disease characterized by hypertriglyceridemia, hypercholesterolemia or both (Goldstein et al. 1973). In addition, high serum levels of apolipoprotein-B (apo-B) are often observed in FCHL affected individuals (Brunzell et al. 1983; Ayyobi et al. 2003). Several genome-wide scans have been performed to detect susceptibility loci for FCHL (Pajukanta et al. 1999; Aouizerat et al. 1999; Allayee et al. 2002). In a previous study, we identified an intragenic microsatellite marker (D10S546) within the protocadherin 15 (*PCDH15*) to be associated with high serum triglycerides (TGs) in Finnish dyslipidemic families (Lilja et al. 2004). Furthermore, *PCDH15* resides in a region on chromosome 10q11 that has been linked to lipid abnormalities in several studies (Pajukanta et al. 1999; Lilja et al. 2004; Huertas-Vazquez et al. 2005). *PCDH15* is a member of the

cadherin superfamily and encodes an integral membrane protein that mediates calcium-dependent cell–cell adhesion. Mutations in *PCDH15* have been associated with hearing-loss and visual-loss due to retinitis pigmentosa (Ahmed et al. 2001; Alagramam et al. 2001a). Several previous epidemiological studies have demonstrated a relationship between hearing loss and hyperlipidemia (Rosen et al. 1964; Rosen and Olin 1965; Evans et al. 2006; Chang et al. 2007). In this study, we investigated all known nonsynonymous SNPs within *PCDH15*, rs11004439, rs10825269, rs4935502 rs2135720, for association with the FCHL component traits, TGs, total cholesterol (TC) and apo-B in multigenerational Finnish and Dutch families with FCHL as well as the *PCDH15* expression pattern in different human tissues. In addition, we investigated the lipid profile in mice with two different loss-of-function mutations in Pcdh15.

Subjects and methods

Finnish FCHL families

A total of 60 Finnish FCHL families comprising 719 individuals were included in this study. The families were recruited in the Helsinki and Turku University Central Hospitals. The inclusion and exclusion criteria for FCHL have been described in detail previously (Pajukanta et al. 1999; Soro et al. 2002). All subjects gave their informed consent. The study design was approved by the ethics committees of the participating centers.

Dutch FCHL families

A total of 32 Dutch FCHL families comprising 549 individuals were included in this study. The families were recruited at the Lipid Clinic of the Utrecht Academic University Hospital, the Netherlands. The inclusion and exclusion criteria for FCHL have been described in detail previously (Allayee et al. 2002). All subjects provided written informed consent. The study design was approved by the ethics committee of the participating center.

Biochemical analysis and SNP genotyping

Serum lipid parameters were measured as described earlier (Pajukanta et al. 1999; Soro et al. 2002; Allayee et al. 2002). We selected all nonsynonymous SNPs within the *PCDH15* gene, rs11004439, rs10825269, rs4935502 and rs2135720, for genotyping. The SNP primers were designed for PCR using the Primer3 program, and for detection, using the SNP Primer Design software (Pyrosequencing). Genotyping of the 1,268 Finnish and Dutch FCHL family members was performed with the Pyrosequencing technique on the automated PSQ HS96A platform. All SNPs had at least 92% genotype call rate. For quality control, we replicated 3.5% of the genotyped samples. The percentage agreement between samples was >99%. All SNPs were tested for a possible violation of Hardy–Weinberg equilibrium (HWE).

Statistical analysis

All of the association analyses were performed using quantitative lipid traits. The quantitative transmission disequilibrium test (QTDT) (Abecasis et al. 2000) implemented in the genetic analysis package SOLAR. QTDT was performed for each analyzed trait in the Finnish and Dutch families, both separately and in the combined dataset. We analyzed the quantitative TG, TC, and apo-B, traits, as they are the key component traits of FCHL. The residuals for these traits were adjusted by age and sex in the total sample, using the SPSS 12.0 program. The PedCheck program was used to assess the genotype data for pedigree inconsistencies (O'Connell and Weeks 1998). P values of less than <0.05 after Bonferroni correction for multiple testing were considered statistically significant. However, it is worth noting that the Bonferroni correction for the probability values obtained in these analyses can be considered conservative, because we investigated highly correlated lipid traits. Apo-A1 and HDL-C traits were analyzed as secondary traits for rs10825269 after establishing the significant associations with TGs and apo-B.

To analyze whether rs10825269 affects a combined trait of HDL-C and TGs, we utilized option 19 of Mendel software (Lange et al. 1976, 2001; Lange and Boehnke 1983). Mendel option 19 performs QTL association using a variance components model. We used a bivariate model consisting of an additive polygenic effect, a random environmental effect, and an additive SNP regression coefficient. Standardized residuals for HDL and Log(TG) were age and sex corrected, and proband ascertainment was corrected for within Mendel. A likelihood ratio test was performed using the formula: $\mathrm{LRT} = 2[\ln(L_{H1}) - \ln(L_{H2})]$, where L_{H1} = maximum likelihood of the bivariate model, and L_{H2} = maximum likelihood of the bivariate model without the additive SNP regression coefficient.

Simulation for functional change in coding nsSNPs

The PolyPhen software was used to investigate a possible impact of all nonsynonymous changes on the structure and function of the *PCDH15 in silico*.

Cross species comparisons

The cross species conservation of the nonsynonymous SNPs was evaluated using the UCSC genome browser.

RT-PCR analysis of *PCDH15* mRNA expression

The expression of *PCDH15* mRNA was analyzed using the human multiple tissue cDNA panel 1 (Clontech). Specific primers for the PCR amplification of *PCDH15* were 5′CCAGGACAAGCTATG TACTTCGAGTCCAAG-3′ (forward) and 5′-GACGAGTACATCGGCTTTGCCG CTCAGTC-3′ (reverse), amplifying a 396 bp fragment (Rouget-Quermalet et al. 2006). Amplification of specific DNA fragments was performed by adding 3 µl of cDNA from the Human Tissue panel I to a PCR mixture containing 0.2 mM dNTPs, 0.4 µM of each primer, 2 µl of 10× reaction buffer, 1.5 µM $MgCl_2$ and 0.2 µl of Taq DNA polymerase. PCR conditions were as follows: After initial denaturation for 10 min at 94°C, the reaction was subjected to 35 cycles of denaturation (30 s, 94°C), annealing (30 s, 61°C) and extension (1 min, 72°C). The amplified products were separated on a 1% agarose gel electrophoresis. G3PDH was used as a reference gene.

Animals

All the mice serum samples were collected at the Department of Otolaryngology-Head and Neck Surgery, Case Western Reserve University, University Hospitals-Case Medical Center. All mice used in this study were maintained on regular mouse diet (6% fat IsoPro 3000 from Purina that contains 6% fat).The mice were fasted for 12 h, beginning one hour after the start of their light cycle. At the conclusion of the fast, the blood was collected from each mouse using a retro-orbital bleed. A total of 41 mice serums were collected for the FVB/N genetic background $Pcdh15^{av\text{-}Tg}$ ($n = 13$ mutants: 3 males and 10 females; and $n = 8$ controls: 4 males and 4 females), and for the C57BL/6J genetic background $Pcdh15^{av\text{-}3J}$ ($n = 9$ mutants: 5 males, 4 females; and $n = 11$ controls: 4 males, 7 females). Mutant mice were generated as described previously (Alagramam et al. 2001b). All animal experimental protocols were approved by Institutional Animal Care and Use Committee, Case Western Reserve University.

Mice serum lipid measurement

All mice were fed a normal diet for 100 days and lipid concentrations were determined. TC and TGs were determined as described previously (Castellani et al. 2004). Each lipid determination was measured in triplicate. The statistical analysis to evaluate differences in the mice lipid measurements was determined by using the unpaired, two tailed Student's *t* test. Sex was included as a covariate in these analyses. Values of $P \leq 0.05$ were considered to be significant.

Results

The mean lipid values of the 92 Dutch and Finnish FCHL families included in this study are shown in Table 1. All nonsynonymous SNPs within *PCDH15* were genotyped in these 92 Finnish and Dutch FCHL families. Genotype distributions for the four investigated SNPs in both populations were consistent with the Hardy–Weinberg equilibrium in nonrelated groups of family members ($P > 0.05$). Of the four nonsynonymous SNPs investigated, SNP rs10825269 showed significant evidence for association for the different quantitative lipid traits, TGs ($P = 0.001$), apo-B ($P = 0.002$) and TC ($P = 0.04$) in the Finns, and for the quantitative apo-B trait in the Dutch ($P = 0.04$) for the same allele (C). No significant association signals were observed with the other three SNPs rs11004439, rs4935502 and rs2135720 ($P > 0.05$). None of the investigated SNPs were in linkage disequilibrium with each other. Next we performed a combined data analysis of both the Finnish and Dutch families with FCHL, and observed a significant increase of statistical significance for all investigated quantitative lipid traits (uncorrected $P = 0.001$–0.0001, Bonferroni corrected $P = 0.02$–0.002). Association results for the combined study sample for SNP rs10825269 are presented in Table 2. We also investigated the nonsynonymous SNP rs10825269 within PCDH15 for associations with quantitative Apo-A1 and HDL-C levels in the Finnish and Dutch FCHL families. No evidence for association was observed for these traits ($P > 0.05$). The frequency of the minor allele of the SNP rs10825269 in both populations was 10%, which is in a good agreement with the allele frequency reported by the International HapMap project in the CEPH samples (http://www.hapmap.org).

The chromosomal region on 10q11, where PCDH15 is located, was also implicated for a combined trait of HDL-C and TGs in our previous study (Lilja et al. 2004). Therefore, we investigated whether rs10825269 affects the combined trait of HDL-C and TGs in the Dutch and Finnish families with FCHL. For this analysis, we utilized option

Table 1 Mean lipid values of the 92 FCHL families included in the study

	Finnish FCHL	Dutch FCHL
No. of families	60	32
No. of subjects (M/F)	719 (356/363)	549 (273/276)
TG, mg/dl (mean ± SD)	316.4 ± 151.0	315.0 ± 201.7
TC, mg/dl (mean ± SD)	298.8 ± 41.3	301.7 ± 61.2
Apo-B, mg/dl (mean ± SD)	146.7 ± 31.1	141.8 ± 24.6
HDL-C, mg/dl (mean ± SD)	40.9 ± 13.1	39.4 ± 12.8

M/F male/female, *TG* triglycerides, *TC* total cholesterol, *Apo-B* apolipoprotein-B, *HDL-C* HDL cholesterol

Table 2 Quantitative family-based association analysis of lipid phenotypes with SNP rs10825269 in the Finnish and Dutch FCHL families using the QTDT program

Trait	Major/minor allele	Minor allele frequency	QTDT[a]	QTDT[b]
TG	**C/T**	0.10	0.0006	0.01
Apo-B	**C/T**	0.10	0.0001	0.002
TC	**C/T**	0.10	0.001	0.02

The risk allele is indicated in bold

TG triglycerides, *TC* total cholesterol, *apo-B* apolipoprotein-B

[a] Uncorrected *P* values (The Bonferroni correction for the probability values obtained in these analyses can be considered conservative, because we investigate highly correlated lipid traits)

[b] *P* values obtained after Bonferroni correction for 24 test (4 SNPs, 3 traits, 2 different populations)

19 of Mendel software (Lange et al. 1976, 2001; Lange and Boehnke 1983) (see Subjects and methods section). We observed that rs10825269 does not significantly alter this combined trait ($P = 0.08$).

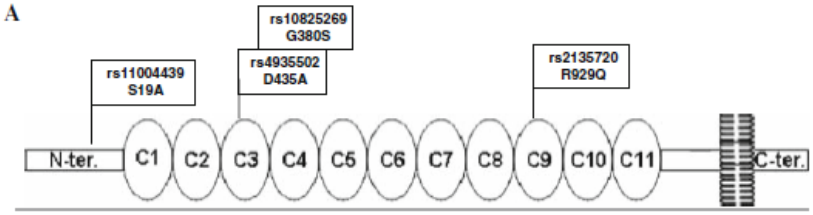
The nonsynonymous changes of rs10825269, rs11004439 and rs2135720 were predicted to be benign by the PolyPhen software (PSIC score difference: 0.057, 1.023 and 1.034, respectively). The SNP rs4935720, 166 bp away from SNP rs10825269, was predicted to be possibly damaging (PSIC score 1.7). We also examined the sequence conservation across species of the nonsynonymous variants within *PCDH15*. The cross-species conservations of these nonsynonymous SNPs are shown in Fig. 1.

Next, we investigated the tissue distribution of *PCDH15* in different human tissues using a commercial human multiple tissue cDNA panel of eight different tissues. We observed that *PCDH15* was expressed in brain, heart, kidney, liver, lung and pancreas. Figure 2 shows the expression patterns of *PCDH15* in eight human adult tissues.

To investigate possible alterations in the lipid profiles of the Pcdh15 mouse mutants, we measured the lipid levels of two mouse mutants homozygous for different loss of function mutation in Pcdh15 ($Pcdh15^{av\text{-}Tg}$ and $Pcdh15^{av\text{-}3J}$) (Fig. 3a). We observed a significant decrease in plasma TG and TC concentrations between the $Pcdh15^{av\text{-}3J}$ homozygotes and age-match wild type siblings ($P = 0.013$ and $P = 0.044$, respectively) (Fig. 3b). No statistically significant differences were observed between the $Pcdh15^{av\text{-}Tg}$ homozygotes and controls for any lipid trait (data not shown).

Discussion

Results from our study suggest that the common allele of SNP rs10825269 within *PCDH15* is associated with TG, apo-B and TC levels in FCHL. This SNP resides in the same exon as the microsatellite D10S546 that was previously associated with high TGs in the Finnish families with FCHL (Lilja et al. 2004). The functional role of the amino acid substitution G380S in the lipid metabolism is unknown. This amino acid substitution is located in the extracellular domain and resides in a highly conserved residue. Although the amino acid change was predicted to be benign using the Polyphen software (Ramensky and Bork 2002), nonsynonymous SNPs in the coding region of a gene could affect the structure and function of the protein. It



B

Variant	hs	pt	rm	mm	rt	cf	dn	la	ec
rs11004439	S	S	S	S	S	A	S	N	S
rs10825269	G	S	S	S	S	S	S	S	S
rs4935502	D	D	D	D	D	D	D	D	D
rs2135720	R	R	R	R	R	R	R	R	R

Fig. 1 Nonsynonymous sequence variants in *PCDH15*. **a** All nonsynonymous sequence variants within *PCDH15*. *N-ter* N-terminus of the amino acid sequence, *C1* cadherin domain 1, *C2* cadherin domain 2, etc.; *C-ter* C-terminus of the amino acid sequence. **b** Sequence conservation across species of nonsynonymous variants in *PCDH15*. The alignment includes Human (*hs*), chimpanzee (*pt*), rhesus monkey (*rm*), mouse (*mm*), rat (*rt*), dog (*cf*), armadillo (*dn*), elephant (*la*), horse (*ec*). *S* serine, *A* alanine, *G* glycine, *D* aspartic acid, *R* Arginine, *Q* Glutamine, *N* aligning has one or more unalignable bases in the gap region

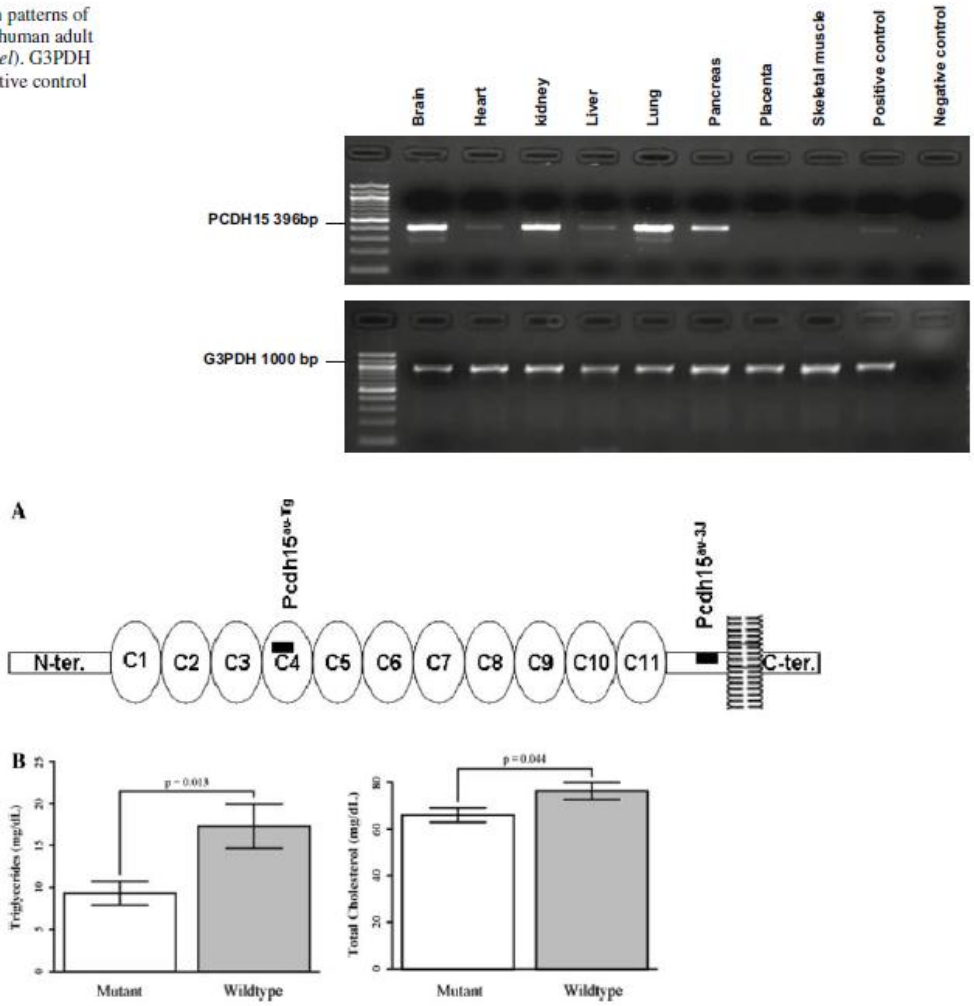


Fig. 2 Expression patterns of *PCDH15* in eight human adult tissues (*upper panel*). G3PDH was used as a positive control (*lower panel*)

Fig. 3 **a** Mouse loss-of function mutations in Pcdh15 investigated in this study. The *solid rectangle* indicates protein truncation due to premature stop mutations in the mutant mouse. *N-ter* N-terminus of the amino acid sequence, *C1* cadherin domain 1, *C2* cadherin domain 2, etc.; *C-ter* C-terminus of the amino acid sequence. **b** Levels of TGs and total cholesterol in the Pcdh15 mouse mutant in the loss-of-function allele Pcdh15$^{av\text{-}3J}$, when compared to sibling controls. Groups of mice were as follows: 9 mutant and 11 control mice (C57BL/6J genetic background). TG and total cholesterol levels are expressed in mg/dl

also remains possible that SNP rs10825269 is in linkage disequilibrium with another functional DNA variant at this locus.

PCDH15 is a member of the cadherin superfamily of calcium-dependent cell–cell adhesion molecules. *PCDH15* plays an essential role in the maintenance of normal retinal and cochlear function, and mutations in *PCDH15* have been associated with nonsyndromic (DFNB23) (Ahmed et al. 2003) and syndromic hearing loss (the Usher syndrome type 1F, USH1F) (Ahmed et al. 2001; Alagramam et al. 2001a). Although *PCDH15* has not been directly related with lipid abnormalities, previous biochemical analysis suggested that USH1F patients have decreased levels of long-chain polyunsaturated fatty acids in plasma (Maude et al. 1998). In addition, previous epidemiological studies have linked hearing loss to lipid abnormalities, showing that hyperlipidemia and atherosclerosis can induce alteration in cochlear function (Rosen et al. 1964; Rosen and Olin 1965; Evans et al. 2006; Chang et al. 2007). The biological role of protocadherins in lipid abnormalities is unclear. The large size and diversity of members of the protocadherin family suggest the participation of these

proteins in a wide variety of biological processes. Previous studies of the Usher syndrome and visual abnormalities have shown that *PCDH15* is expressed in several tissues including retina, brain, cerebellum, kidney, cochlea and liver (Alagramam et al. 2001a; Rouget-Quermalet et al. 2006). In this study, the expression pattern of *PCDH15* in human was consistent with the pattern previously observed in mice (Alagramam et al. 2001b; Rouget-Quermalet et al. 2006). Importantly, we also demonstrate that *PCDH15* is expressed in human pancreas. Further investigation is necessary to confirm the role of PCDH15 in lipid abnormalities.

To the best of our knowledge, this is the first time that lipid traits have been investigated in the Pcdh15 mouse mutant. Although additional studies are necessary to confirm our findings, these observations suggest a possible alteration in the lipid profile of the Pcdh15 mouse mutant due to the $Pcdh15^{av\text{-}3J}$ loss-of-function mutation. No statistically significant differences were observed in the $Pcdh15^{av\text{-}Tg}$ loss-of-function mutation. The observed results suggest differences in the genetic background between the FVB/N and C57BL/6J strains. This suggestion is indirectly supported by a previous study demonstrating that the FVB/N strain is susceptible to diet induce-atherosclerosis whereas the C57BL/6J strain is resistant (Hoover-Plow et al. 2006). A given phenotype could be obvious in one inbred genetic background but it could be suppressed in another genetic background (potential genetic modifier effect; Nadeau 2001).

Genome-wide association analyses in unrelated individuals have identified several loci associated with lipid abnormalities. However, the variants identified so far explain a small fraction of the disease risk, suggesting that many genes implicated in the lipid metabolism still remain undiscovered. We have previously identified several genes associated with FCHL using family-based studies and replicated our results in different cohorts (Pajukanta et al. 2004; Huertas-Vazquez et al. 2005, 2008; Weissglas-Volkov et al. 2006; Lee et al. 2008; Plaisier et al. 2009). Family-based studies are more robust to population stratification and families ascertained for the disease of interest provide a powerful tool for association studies.

In conclusion, we have identified a nonsynonymous variant in *PCDH15* associated with TG, apo-B and TC levels in multigenerational Caucasian FCHL families. Replication in additional FCHL study samples and sequencing of *PCDH15* are warranted to further explore the effects of *PCDH15* in FCHL.

Acknowledgments We thank the patients and family members for their participation in this study. We thank L. Peltonen, I. Nuotio and J. S. A. Viikari for sample collection. We thank Janet Sinsheimer and Daphna Weissglas-Volkov for their advice in the statistical analysis; as well as E. Nikkola and M. Lupsakko for laboratory technical assistance. C. L. Plaisier is supported by the National Human Genome Research Institute grant T32 HG02536. M.-R. Taskinen is supported by a grant from Finnish Heart Foundation. This research was supported by the National Institutes of Health grant HL-28481 and the American Heart Association grant 0725232Y. T. W. A. de Bruin has been employed by GlaxoSmithKline. This work was not supported in full or part by GlaxoSmithKline. We would like to thank Janis Paulsey in the laboratory of K. Alagramam for technical assistance with animal work.

References

Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279–292

Ahmed ZM, Riazuddin S, Bernstein SL, Ahmed Z, Khan S, Griffith AJ, Morell RJ, Friedman TB, Riazuddin S, Wilcox ER (2001) Mutations of the protocadherin gene *PCDH15* cause Usher syndrome type 1F. Am J Hum Genet 69(1):25–34

Ahmed ZM, Riazuddin S, Ahmad J, Bernstein SL, Guo Y, Sabar MF, Sieving P, Riazuddin S, Griffith AJ, Friedman TB, Belyantseva IA, Wilcox ER (2003) *PCDH15* is expressed in the neurosensory epithelium of the eye and ear and mutant alleles are responsible for both USH1F and DFNB23. Hum Mol Genet 12(24):3215–3223

Alagramam KN, Yuan H, Kuehn MH, Murcia CL, Wayne S, Srisailpathy CR, Lowry RB, Knaus R, Van Laer L, Bernier FP, Schwartz S, Lee C, Morton CC, Mullins RF, Ramesh A, Van Camp G, Hageman GS, Woychik RP, Smith RJ (2001a) Mutations in the novel protocadherin *PCDH15* cause Usher syndrome type 1F. Hum Mol Genet 10(16):1709–1718

Alagramam KN, Murcia CL, Kwon HY, Pawlowski KS, Wright CG, Richard P, Woychik RP (2001b) The mouse Ames waltzer hearing-loss mutant is caused by mutation of Pcdh15, a novel protocadherin gene. Nat Genet 27:99–102

Allayee H, Krass KL, Pajukanta P, Cantor RM, van der Kallen CJ, Mar R, Rotter JI, de Bruin TW, Peltonen L, Lusis AJ (2002) Locus for elevated apolipoprotein B levels on chromosome 1p31 in families with familial combined hyperlipidemia. Circ Res 90:926–931

Aouizerat BE, Allayee H, Cantor RM, Dallinga-Thie GM, Lanning CD, de Bruin TW, Rotter JI, Lusis AJ (1999) A genome scan for familial combined hyperlipidemia reveals evidence of linkage with a locus on chromosome 11. Am J Hum Genet 65:397–412

Ayyobi AF, McGladdery SH, McNeely MJ, Austin MA, Motulsky AG, Brunzell JD (2003) Small, dense LDL and elevated apolipoprotein B are the common characteristics for the three major lipid phenotypes of familial combined hyperlipidemia. Arterioscler Thromb Vasc Biol 23:1289–1294

Brunzell JD, Albers JJ, Chait A, Grundy SM, Groszek E, McDonald GB (1983) Plasma lipoproteins in familial combined hyperlipidemia and monogenic familial hypertriglyceridemia. J Lipid Res 24:147–155

Castellani LW, Gargalovic P, Febbraio M, Charugundla S, Jien ML, Lusis AJ (2004) Mechanisms mediating insulin resistance in transgenic mice overexpressing mouse apolipoprotein A-II. J Lipid Res 12:2377–2387

Chang N, Yu M, Ho K, Ho C (2007) Hyperlipidemia in noise-induced hearing loss. Otolaryngol Head Neck Surg 137(4):603–606

Evans M, Tonini R, Shope C, Oghalai J, Jerger J, Insull W Jr, Brownell W (2006) Dyslipidemia and auditory function. Otol Neurotol 27(5):609–614

Goldstein JL, Schrott HG, Hazzard WR, Bierman EL, Motulsky AG (1973) Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. J Clin Invest 52:1544–1568

Hoover-Plow J, Shchurin A, Hart E, Sha J, Hill AE, Singer JB, Nadeau JH (2006) Genetic background determines response to hemostasis and thrombosis. BMC Blood Disord 6:6

Huertas-Vazquez A, Aguilar-Salinas C, Lusis AJ, Cantor RM, Canizales-Quinteros S, Lee JC, Mariana-Nunez L, Riba-Ramirez RM, Jokiaho A, Tusie-Luna T, Pajukanta P (2005) Familial combined hyperlipidemia in Mexicans: association with upstream transcription factor 1 and linkage on chromosome 16q24.1. Arterioscler Thromb Vasc Biol 25:1985–1991

Huertas-Vazquez A, Plaisier C, Weissglas-Volkov D, Sinsheimer J, Canizales-Quinteros S, Cruz-Bautista I, Nikkola E, Herrera-Hernandez M, Davila-Cervantes A, Tusie-Luna T, Taskinen MR, Aguilar-Salinas C, Pajukanta P (2008) TCF7L2 is associated with high serum triacylglycerol and differentially expressed in adipose tissue in families with familial combined hyperlipidaemia. Diabetologia 51:62–69

Lange K, Boehnke M (1983) Extensions to pedigree analysis. IV. Covariance components models for multivariate traits. Am J Med Genet 14:513–524

Lange K, Westlake J, Spence MA (1976) Extensions to pedigree analysis. III. Variance components by the scoring method. Ann Hum Genet 39:485–491

Lange K, Cantor R, Horvath S, Perola M, Sabatti C, Sinsheimer J, Sobel E (2001) MENDEL version 4.0: a complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. Am J Hum Genet 69(Suppl):504

Lee JC, Weissglas-Volkov D, Kyttala M, Dastani Z, Cantor RM, Sobel EM, Plaisier CL, Engert JC, van Greevenbroek MM, Kane JP, Malloy MJ, Pullinger CR, Huertas-Vazquez A, Aguilar-Salinas CA, Tusie-Luna T, de Bruin TW, Aouizerat BE, van der Kallen CC, Croce CM, Aqeilan RI, Marcil M, Viikari JS, Lehtimaki T, Raitakari OT, Kuusisto J, Laakso M, Taskinen MR, Genest J, Pajukanta P (2008) WW-domain-containing oxidoreductase is associated with low plasma HDL-C levels. Am J Hum Genet 83:180–192

Lilja HE, Suviolahti E, Soro-Paavonen A, Hiekkalinna T, Day A, Lange K, Sobel E, Taskinen MR, Peltonen L, Perola M, Pajukanta P (2004) Locus for quantitative HDL-cholesterol on chromosome 10q in Finnish families with dyslipidemia. J Lip Res 45:1876–1884

Maude MB, Anderson EO, Anderson RE (1998) Polyunsaturated fatty acids are lower in blood lipids of Usher's type I but not Usher's type II. Invest Ophthalmol Vis Sci 39:2164–2166

Nadeau JH (2001) Modifier genes in mice and humans. Nat Rev Genet 2(3):165–174

O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259–266

Pajukanta P, Terwilliger JD, Perola M, Hiekkalinna T, Nuotio I, Ellonen P, Parkkonen M, Hartiala J, Porkka K, Laakso M, Viikari JSA, Ehnholm C, Taskinen M-R, Peltonen L (1999) Genomewide scan for familial combined hyperlipidemia genes in Finnish families, suggesting multiple susceptibility loci influencing triglyceride, cholesterol and apolipoprotein B levels. Am J Hum Genet 64:1453–1463

Pajukanta P, Lilja HE, Sinsheimer JS, Cantor RM, Lusis AJ, Gentile M, Duan XJ, Soro-Paavonen A, Naukkarinen J, Saarela J, Laakso M, Ehnholm C, Taskinen MR, Peltonen L (2004) Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). Nat Genet 36:371–376

Plaisier CL, Kyttala M, Weissglas-Volkov D, Sinsheimer JS, Huertas-Vazquez A, Riba L, Ramirez-Jimenez S, de Bruin TW, Tusie-Luna T, Aouizerat BE, Pullinger CR, Malloy MJ, Kane JP, Cruz-Bautista I, Herrera MF, Aguilar-Salinas C, Kuusisto J, Laakso M, Taskinen MR, van der Kallen CJ, Pajukanta P (2009) Galanin preproprotein is associated with elevated plasma triglycerides. Arterioscler Thromb Vasc Biol 29:147–152

Ramensky V, Bork P (2002) Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30:3894–3900

Rosen S, Olin P (1965) Hearing loss and coronary heart disease. Arch Otolaryngol 82:236–243

Rosen S, Plester D, EL-Mofty A, Rosen H (1964) Relation of hearing loss to cardiovascular disease. Trans Am Acad Ophthalmol Otolaryngol 68:433–444

Rouget-Quermalet V, Giustiniani J, Marie-Cardine A, Beaud G, Besnard F, Loyaux D, Ferrara P, Leroy K, Shimizu N, Gaulard P, Bensussan A, Schmitt C (2006) Protocadherin 15 (*PCDH15*): a new secreted isoform and a potential marker for NK/T cell lymphomas. Oncogene 25(19):2807–2811

Soro A, Pajukanta P, Lilja HE, Ylitalo K, Hiekkalinna T, Perola M, Cantor RM, Viikari JS, Taskinen MR, Peltonen L (2002) Genome scans provide evidence for low-HDL-C loci on chromosomes 8q23, 16q24.1-24.2, and 20q13.11 in Finnish families. Am J Hum Genet 70(5):1333–1340

Weissglas-Volkov D, Huertas-Vazquez A, Suviolahti E, Lee J, Plaisier C, Canizales-Quinteros S, Tusie-Luna T, Aguilar-Salinas C, Taskinen MR, Pajukanta P (2006) Common hepatic nuclear factor-4alpha variants are associated with high serum lipid levels and the metabolic syndrome. Diabetes 55:1970–1977

CHAPTER 5:

Conclusions

5.1 Summary of Approaches

The first part of my thesis focused on discovering genetic variation and the underlying mechanisms regulating serum lipid levels. A common method used to reveal the underlying mechanism of a GWAS signal is to determine whether the SNP acts as a *cis*-eQTL in a tissue known to be relevant to lipid regulation. Knowing whether a SNP impacts gene expression can offer a potential biological mechanism responsible for a GWAS signal. This biological mechanism can then be further tested and explored using functional studies. In this thesis we studied an LDL GWAS signal (1).

Determining the specific lipid sub-particles an LDL-C GWAS signal is associated with can also provide novel information regarding the underlying mechanism influencing serum LDL-C levels. The refined subphenotypes and endophenotypes can help further dissect the critical phenotype – genotype relationship of a variant. As demonstrated in chapter 2, we showed that LDL-C GWAS SNP rs7575840 is associated with apoB-containing lipid sub-particles, including IDL and VLDL (1). Thus, future research can search for ways rs7575840 may act on IDL and VLDL regulatory pathways.

The second part of my thesis utilized WGCNA across multiple populations to discover biologically relevant pathways and networks impacting serum lipid levels in humans (3). As demonstrated in chapter 3, we used WGCNA to determine novel TG regulatory genes in adipose tissue of Finnish and Mexican subjects. Using two distinct populations and replication in multiple study samples helps overcome some of known error sources of the gene expression

analysis, and thus helps reveal lipid correlated genes and pathways with true biological relevance.

The third part of my thesis investigated nonsynonymous variants for association in gene residing in a genomic region previously linked to FCHL and its component traits to discover novel FCHL variants (8). As nonsynonymous SNPs change the primary structure of a protein, the amino acid change caused by some SNPs may be deleterious to protein function which can ultimately impact lipid levels. As demonstrated in chapter 4, this method was successful at identifying variants in *PCDH15* that are associated with TGs, apoB, and TC variants (8). Creating a knockout mouse to validate the effect the non-functional gene has on lipid levels provides strong additional evidence that the gene is involved in FCHL.

5.2 Summary and Future Insights

In chapter 2 of the dissertation, we introduce a novel mechanism underlying LDL-C GWAS signal rs7575840. We also show that the variant is associated with LDL-C in a Mexican study sample. Furthermore, we demonstrate that rs7575840 is associated with serum apoB levels and several apoB-containing lipid sub-particles of various sizes. In adipose tissue, we discovered that rs7575840 acts as a *cis*-eQTL for both *APOB* and the non-coding RNA BU630349. The likely mechanism responsible for this *cis*-eQTL is that rs7575840 disrupts a binding site for the transcription factor CCAAT/enhancer binding protein alpha (CEBPA) and CCAAT/enhancer binding protein beta (CEBPB). Thus, we provide a testable hypothesis of how this variant may be influencing serum lipid levels, which should be explored further in future studies.

In chapter 3, we discovered novel TG genes consistently across two different populations and three study samples using WGCNA on adipose gene expression data. When we performed WGCNA on three study samples (two Finnish study samples, and one Mexican study sample), we discovered that 34 genes were part of modules associated with serum TG levels in all three study samples. Only 11 of the 34 genes had been implicated in obesity, diabetes, or CHD in prior publications, leaving 23 genes with a novel role in TG regulation. This set of 34 genes is related to TG levels in two distinct populations, and furthermore, different methods of gene expression analysis were used; one Finnish study sample underwent RNA sequencing to measure gene expression, while microarrays measured gene expression in the other two study samples. Thus, the relation these 34 genes have to serum TG levels are robust to both gene expression measurement and population, providing strong evidence that this set of genes is important in TG regulation.

Chapter 4 of this dissertation focused on *PCDH15*, a gene residing in a region implicated in previous FCHL linkage studies (5-7). We first tested whether 4 nonsynonymous *PCDH15* variants were associated with lipid levels in 92 Finnish and Dutch FCHL families. One nonsynonymous SNP (rs10825269) significantly associated with TG, apoB, and TC. To determine whether this gene is in fact the causal gene, we created a loss-of-function *PCDH15* knockout mouse, and observed that there was a significant decrease in TG and TC levels in a mouse carrying two copies loss-of-function construct. Future studies should explore the underlying mechanism and biological pathways how the nonsynonymous *PCDH15* variants impact lipid levels.

5.3 Concluding Remarks

Lipid GWAS studies have been successful in identifying genomic regions associated with lipid levels (63). However, the genes and biological pathways responsible for the lipid association remain elusive for most signals (63). As demonstrated in chapter 2 of this thesis, we uncovered a possible mechanism underlying an LDL-C GWAS signal rs7575840 (1). Revealing the causal variants, underlying genes, and biological pathways that regulate lipid levels is essential for identifying potential drug targets. Very few lipid drugs have been developed based on the function of lipid-associated GWAS genes as of yet (63), leaving many viable drug development opportunities to be explored in the future once more is known about the underlying mechanisms behind these associations.

In addition to GWAS regions, uncovering the causal variant(s) and gene(s) in linkage regions has been very challenging (32,62). In chapter 4 of this thesis, we genotyped common nonsynonymous variants in *PCDH15*, the gene containing a microsatellite marker (D10S546) linked to FCHL in several studies (5-7). To search for causal variants in this region, we genotyped 4 nonsynonymous variants in *PCDH15* in FCHL families and discovered a nonsynonymous variant (rs10825269) associated with FCHL component traits (8). Importantly, PCDH15 knockout mice displayed a phenotype similar to FCHL (8).

Another way to better understand the biological pathways regulating lipid levels is through expression network analysis (51,52). We successfully identified novel TG regulatory genes and pathways by overlaying TG-associated modules in two unique populations, Finnish and Mexicans utilizing WGCNA (3). GWAS signals may alter lipid levels through perturbing

networks, therefore, determining whether GWAS signals alter specific lipid regulatory pathways can reveal the causal variant and gene in lipid GWAS regions spanning several genes. In chapter 3 of this thesis, we discovered one GWAS signal associated with expression of a TG-associated gene expression network in human adipose tissue (3). Thus, we identified a potential TG regulatory mechanism that can be further explored by performing functional studies in tissue cultures and model organisms.

The vast majority of GWAS have focused on SNPs, yet there are various other kinds of genomic differences that need to be further explored in the future. Copy number variations, epigenetic modifications, chromosomal rearrangements, insertions and deletions, and RNA editing can influence gene expression and protein structure. Thus, SNPs are only one part in a multi-faceted approach to understanding genetics of lipid regulation in humans.

Determining the genes and genetic variation that affect lipid levels is important to better understand of the biological pathways involved in lipid regulation. Additional functional studies are necessary to gain a better biological understanding of the genetic variation associated with lipids. Proteomic studies are likely to become essential to understand the mechanism underlying genetic variation. Often, GWAS signals associated with a lipid trait span many genes, making it difficult to identify the causal gene without functional experimentation. Furthermore, GWAS have solely focused on common variation (minor allele frequencies > 5%), and a study of 100,000 individuals revealed that the vast majority of the common GWAS lipid-associated variants have small effect sizes and do not explain all of the expected genetic heritability of serum lipid levels (11). Scientists have hypothesized that rare variants in the genes identified by

GWAS may have larger effect sizes and explain some of the unexplained genetic heritability of lipid traits. A recent publication successfully identified novel rare variants in TG GWAS regions (39). However, sequencing the exomes of up to 10,000 people may be necessary to discover rare exome variants throughout the genome that influence lipid levels (63).

Besides DNA, studying other aspects of the central dogma of molecular biology is necessary to better understand lipid biology. Proteomic studies may determine whether genetic variation impacts protein structure, localization, binding, or amount in tissues or model organisms. Thus, systematic approaches of proteomics may reveal novel genes in lipid regulation and identify the regional causal genes from metaGWAS. For instance, pulldown assays of proteins that lie within a GWAS signal would reveal direct and indirect binding partners. Discovering that a GWAS-region protein binds to proteins already known to be involved in lipid regulation would reveal a likely causal candidate in the GWAS region. The candidate causal proteins can then be screened for nonsynonymous variants that impact protein binding. A copy of the gene that contains the variant altering protein binding can be knocked into a model organism (such as a mouse). Finally, the resulting mouse can be used to determine whether the nonsynonymous variant impacts lipid levels, which would provide strong evidence for a causal gene within the GWAS signal.

After discovering the genes regulating lipid levels, model organisms can be utilized for experiments that cannot be done in humans. For instance, in a model organism, a researcher can knockout the gene of interest in a relevant tissue such as liver that is seldom available in human to determine whether lipid levels are impacted, or a researcher can determine whether a receptor

localizes to the proper area in the relevant tissue in this knockout line. Model organisms serve as essential tools necessary for elucidating the biological impact of genetic variation. Furthermore, most human genes have mouse orthologs, so mice and humans may have variation in the same genes or genetic regions that affect lipid levels. Discovering lipid-associated genetic variation that map near orthologous genes in mice may reveal causal variants and genes in a large DNA region associated with a lipid trait. Thus, utilizing cross-species mapping can serve as an excellent tool to better elucidate regulation of serum lipid levels.

Some of the currently undiscovered genetic variation influencing lipid levels may be hidden due to gene-environment interactions. People have different combinations of genetic variants and environmental exposures, and the same genetic variant may impact a phenotype in some individuals but not in others. Teasing out genetic variants that influence lipid levels in certain scenarios of environmental exposures will likely lead to a better understanding of lipid regulation and lipid disorders.

Identifying genes involved with lipid regulation can ultimately lead to potential drug targets. The discovery of genes which regulate lipid levels is the initial first step necessary for drug development. Although we have lipid lowering medication, heart disease is still the most common cause of death in developed nations, which emphasizes the need for additional drugs that lower lipids and extend lifespan. The most common lipid lowering drug (statins) have several relatively common side effects, including muscle pain, rhapdomyolysis, and liver damage (64). In conclusion, lipid research is a crucial tool for identifying genes and genetic variants that influence lipid levels, which can yield novel drug targets and benefit society as a whole.

Pharmacogenomics is one of the fields that can drastically improve people's lives in the future. It has been estimated that adverse drug reactions (ADR) are between the fourth and sixth most common cause of death in the United States (65). Determining whether a person is genetically predisposed to having an ADR may greatly reduce fatal ADRs in the future. One of the few instances in medicine where genetic information is already used is in calculating Warfarin dosage (66). Future advancements in pharmacogenomics may help doctors prescribe the optimal medicine, which can save lives by getting patients on the optimal drug sooner and reducing the amount of ADRs. As sequencing costs are substantially decreasing overtime, upcoming studies may reveal novel variants that influence ADR risk or drug effectiveness. Future medical care may thus be personalized based on a person's genetic make-up.

References

1. Haas BE, Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, Vergnes L, Cruz-Bautista I, Riba L, Stancakova A, Kuusisto J, Soininen P, Kangas AJ, Ala-Korpela M, Tusie-Luna T, Laakso M, Pajukanta P. 2011. Evidence of how rs7575840 influences apolipoprotein B-containing lipid particles. *Arterioscler Thromb Vasc Biol* 31:1201-1207.

2. Aguilar-Salinas CA, Gómez-Pérez FJ, Rull J, Villalpando S, Barquera S, Rojas R. 2006. Prevalence of dyslipidemias in the Mexican National Health and Nutrition Survey 2006. *Salud Publica Mex* 52:S44–S53.

3. Haas BE, Horvath S, Pietiläinen KH, Cantor RM, Nikkola E, Weissglas-Volkov D, Rissanen A, Civelek M, Cruz-Bautista I, Riba L, Kuusisto J, Kaprio J, Tusie-Luna T, Laakso M, Aguilar-Salinas CA, Pajukanta P. Adipose Transcript Networks Across Finns and Mexicans Identify Novel Triglyceride Genes. Manuscript submitted for publication.

4. Björntorp P, Karlsson M. 1970. Triglyceride Synthesis in Human Subcutaneous Adipose Tissue Cells of Different Size. *Eur J Clin Invest* 1:112-117.

5. Lilja HE, Suviolahti E, Soro-Paavonen A, Hiekkalinna T, Day A, Lange K, Sobel E, Taskinen MR, Peltonen L, Perola M, Pajukanta P. 2004. Locus for quantitative HDL-cholesterol on chromosome 10q in Finnish families with dyslipidemia. *J Lip Res* 45:1876–1884.

6. Huertas-Vazquez A, Aguilar-Salinas C, Lusis AJ, Cantor RM, Canizales-Quinteros S, Lee JC, Mariana-Nunez L, Riba-Ramirez RM, Jokiaho A, Tusie-Luna T, Pajukanta P. 2005. Familial combined hyperlipidemia in Mexicans: association with upstream transcription factor 1 and linkage on chromosome 16q24.1. *Arterioscler Thromb Vasc Biol* 25:1985–1991.

7. Pajukanta P, Terwilliger JD, Perola M, Hiekkalinna T, Nuotio I, Ellonen P, Parkkonen M, Hartiala J, Ylitalo K, Pihlajamäki J, Porkka K, Laakso M, Viikari J, Ehnholm C, Taskinen MR, Peltonen L. 1999. Genomewide scan for familial combined hyperlipidemia genes in finnish families, suggesting multiple susceptibility loci influencing triglyceride, cholesterol, and apolipoprotein B levels. *Am J Hum Gen* 64:1453-1463.

8. Huertas-Vazquez A, Plaisier CL, Geng R, Haas BE, Lee J, Greevenbroek MM, van der Kallen C, de Bruin TW, Taskinen MR, Alagramam KN, Pajukanta P. 2010. A nonsynonymous SNP within PCDH15 is associated with lipid traits in familial combined hyperlipidemia. *Hum Genet* 127:83-89.

9. Durrington P. 2003. Dyslipidaemia. *Lancet* 362:717-731.

10. Pahan K. 2006. Lipid-lowering drugs. *Cell Mol Life Sci.* 63:1165-1178.

11. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso

GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, König IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Döring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA Jr, Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466:707-713.

12. Weiss LA, Pan L, Abney M, Ober C. 2006. The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet.* 38: 218-222.

13. Day, INM. 2011. Monogenic Hypercholesterolemia: Genetics. eLS.

14. Goldberg IJ, Eckel RH, McPherson R. 2011. Triglycerides and heart disease: still a hypothesis? *Arterioscler Thromb Vasc Biol.* 31:1716-1725.

15. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). 2002. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *Circulation* 106:3143–421.

16. Miller M, Stone NJ, Ballantyne C, Bittner V, Criqui MH, Ginsberg HN, Goldberg AC, Howard WJ, Jacobson MS, Kris-Etherton PM, Lennie TA, Levi M, Mazzone T, Pennathur S; American Heart Association Clinical Lipidology, Thrombosis, and Prevention Committee of the Council on Nutrition, Physical Activity, and Metabolism; Council on Arteriosclerosis,

Thrombosis and Vascular Biology; Council on Cardiovascular Nursing; Council on the Kidney in Cardiovascular Disease. 2011. Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association. *Circulation* 123:2292-2333 .

17. Assmann G, Schulte H, Funke H, von Eckardstein A. 1998. The emergence of triglycerides as a significant independent risk factor in coronary artery disease. *Eur Heart J* 19(suppl M):M8–M14.

18. Castelli WP. 1992. Epidemiology of triglycerides: a view from Framingham. *Am J Cardiol* 70:3H–9H.

19. Fontbonne A, Eschwège E, Cambien F, Richard JL, Ducimetière P, Thibult N, Warnet JM, Claude JR, Rosselin GE. 1989. Hypertriglyceridaemia as a risk factor of coronary heart disease mortality in subjects with impaired glucose tolerance or diabetes. Results from the 11-year follow-up of the Paris Prospective Study. *Diabetologia* 32:300–304.

20. Triglyceride Coronary Disease Genetics Consortium and Emerging Risk Factors Collaboration, Sarwar N, Sandhu MS, Ricketts SL, Butterworth AS, Di Angelantonio E, Boekholdt SM, Ouwehand W, Watkins H, Samani NJ, Saleheen D, Lawlor D, Reilly MP, Hingorani AD, Talmud PJ, Danesh J. 2010. Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. *Lancet.* 375:1634–1639.

21. Jun M, Foote C, Lv J, Neal B, Patel A, Nicholls SJ, Grobbee DE, Cass A, Chalmers J, Perkovic V.2010. Effects of fibrates on cardiovascular outcomes: a systematic review and meta-analysis. *Lancet* 375:1875–1884.

22. Johansen CT, Hegele RA. 2011. Genetic bases of hypertriglyceridemic phenotypes. *Curr Opin Lipidol* 22:247-253.

23. Goldstein JL, Schrott HG, Hazzard WR, Bierman EL, Motulsky AG. 1973. Hyperlipidemia in coronary heart disease II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J. Clin. Invest* 52:1544–1568.

24. Nikkilä EA, Aro A. 1973. Family study of serum lipids and lipoproteins in coronary heart disease. *Lancet* 1:954–959.

25. Rose HG, Kranz P, Weinstock M, Juliano J, Haft JI. 1973. Inheritance of combined hyperlipoproteinemia: evidence for a new lipoprotein phenotype. *Am J Med* 54:148–160.

26. de Graaf J, Stalenhoef AF. 1998. Defects of lipoprotein metabolism in familial combined hyperlipidaemia. *Curr Opin Lipidol* 9:189–196.

27. Cullen P, Farren B, Scott J, Farrall M. 1994. Complex segregation analysis provides evidence for a major gene acting on serum triglyceride levels in 55 British families with familial combined hyperlipidemia. *Arterioscler Thromb* 14:1233–1249.

28. Jarvik GP, Brunzell JD, Austin MA, Krauss RM, Motulsky AG, Wijsman E. 1994. Genetic predictors of FCHL in four large pedigrees. Influence of ApoB level major locus predicted genotype and LDL subclass phenotype. *Arterioscler Thromb* 14:1687–1694.

29. Hopkins PN, Heiss G, Ellison RC, Province MA, Pankow JS, Eckfeldt JH, Hunt SC. 2003. Coronary artery disease risk in familial combined hyperlipidemia and familial hypertriglyceridemia: a case-control comparison from the National Heart, Lung, and Blood Institute Family Heart Study. *Circulation* 108:519–523.

30. Pitsavos C, Skoumas I, Masoura C, Aznaouridis K, Papadimitriou L, Chrysohoou C, Giotsas N, Toutouza M, Stefanadis C. 2008. Prevalence and determinants of coronary artery disease in males and females with familial combined hyperlipidaemia. *Atherosclerosis* 199:402–407.

31. Goldstein JL, Schrott HG, Hazzard WR, Bierman EL, Motulsky AG. 1973. Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J Clin Invest* 52:1544–1568.

32. Suviolahti E, Lilja HE, Pajukanta P. 2006. Unraveling the complex genetics of familial combined hyperlipidemia. *Ann Med* 38:337-351.

33. Thomas EL, Potter E, Tosi I, Fitzpatrick J, Hamilton G, Amber V, Hughes R, North C, Holvoet P, Seed M, Betteridge DJ, Bell JD, Naoumova RP. 2007. Pioglitazone added to conventional lipid-lowering treatment in familial combined hyperlipidaemia improves parameters of metabolic control: relation to liver, muscle and regional body fat content. *Atherosclerosis* 195:e181-190.

34. Pajukanta P, Nuotio I, Terwilliger JD, Porkka KV, Ylitalo K, Pihlajamäki J, Suomalainen AJ, Syvänen AC, Lehtimäki T, Viikari JS, Laakso M, Taskinen MR, Ehnholm C, Peltonen L. 1998. Linkage of familial combined hyperlipidaemia to chromosome 1q21–q23. *Nat Genet* 18:369–373.

35. Pajukanta P, Lilja HE, Sinsheimer JS, Cantor RM, Lusis AJ, Gentile M, Duan XJ, Soro-Paavonen A, Naukkarinen J, Saarela J, Laakso M, Ehnholm C, Taskinen MR, Peltonen L. 2004. Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). *Nat Genet* 36:371-376.

36. Eichenbaum-Voline S, Olivier M, Jones EL, Naoumova RP, Jones B, Gau B, Patel HN, Seed M, Betteridge DJ, Galton DJ, Rubin EM, Scott J, Shoulders CC, Pennacchio LA. 2004. Linkage and association between distinct variants of the APOA1/C3/A4/A5 gene cluster and familial combined hyperlipidemia. *Arterioscler Thromb Vasc Biol* 24:167-174.

37. Wojciechowski AP, Farrall M, Cullen P, Wilson TME, Bayliss JD, Farren B, Griffin BA, Caslake MJ, Packard CJ, Shepherd J, Thakker R, Scott J. 1990. Familial combined hyperlipidaemia linked to the apolipoprotein AI-CIII-AIV gene cluster on chromosome 11q23-q24. *Nature* 349:161-164.

38. Campagna F, Montali A, Baroni MG, Maria AT, Ricci G, Antonini R, Verna R, Arca M. 2002. Common variants in the lipoprotein lipase gene, but not those in the insulin receptor substrate-1, the beta3-adrenergic receptor, and the intestinal fatty acid binding protein-2 genes, influence the lipid phenotypic expression in familial combined hyperlipidemia. *Metabolism* 51:1298-305.

39. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Hegele RA. 2010. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 42:684-687.

40. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M. 2008. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 2008; 40:189–197.

41. Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4:1073-1081.

42. Ng PC, Henikoff S. 2006. Predicting the Effects of Amino Acid Substitutions on Protein Function *Annu Rev Genomics Hum Genet* 7:61-80.

43. Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31:3812-3814.

44. Ng PC, Henikoff S. 2002. Accounting for Human Polymorphisms Predicted to Affect Protein Function. *Genome Res* 12:436-446.

45. Ng PC, Henikoff S. 2001. Predicting Deleterious Amino Acid Substitutions. *Genome Res* 11:863-874.

46. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* 7:248-249.

47. Wang K, Li M, Hakonarson H. 2010. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Res* 38:e164.

48. Jansson PA, Larsson A, Smith U, Lönnroth P 1992. Glycerol production in subcutaneous adipose tissue in lean and obese humans. *J Clin Invest* 89:1610-1617.

49. Hotamisligil GS, Arner P, Caro JF, Atkinson RL, Spiegelman BM. 1995. Increased adipose tissue expression of tumor necrosis factor-alpha in human obesity and insulin resistance. *J Clin Invest* 95:2409-2415.

50. Plaisier CL, Horvath S, Huertas-Vazquez A, Cruz-Bautista I, Herrera MF, Tusie-Luna T, Aguilar-Salinas C, Pajukanta P. 2009. A systems geneics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipdemia. *PLoS Genet* 5:e1000642.

51. Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.

52. Langfelder P, Zhang B, Horvath S. 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 24:719–720.

53. Aten JE, Fuller TF, Lusis AJ, Horvath S. 2008. Using genetic markers to orient the edges in quantitative trait networks: the NEO software. *BMC Syst Biol* 2:34.

54. Vakkilainen J, Steiner G, Ansquer JC, Aubin F, Rattier S, Foucher C, Hamsten A, Taskinen MR; DAIS Group. 2003. Relationships between low-density lipoprotein particle size, plasma lipoproteins, and progression of coronary artery disease: the Diabetes Atherosclerosis Intervention Study (DAIS). *Circulation* 107:1733-1737.

55. Avogaro P, Bon GB, Cazzolato G, Quinci GB. 1979. Are apolipoproteins better discriminators than lipids for atherosclerosis? *Lancet* 1:901-903.

56. Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105-1111.

57. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.

58. Trapnell C, Williams BA, Pertea G, Mortazavi AM, Kwan G, van Baren MJ, Salzberg SL, Wold B, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511-515.

59. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. 2011. Improving RNA-Seq expression estimates by correcting for fragment bias *Genome Biol* 12:R22.

60. Roberts A, Pimentel H, Trapnell C, Pachter L. 2011. Identification of novel transcripts in annotated genomes using RNA-Seq *Bioinformatics* 27:2325-2329.

61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.

62. Lee JC, Lusis AJ, Pajukanta P. 2006. Familial combined hyperlipidemia: upstream transcription factor 1 and beyond. *Curr Opin Lipidol* 17:101–109.

63. Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. 2012. *Cell* 148:1242-1257.

64. Law M, Rudnicka AR. 2006. Statin safety: A systemic review. *Am J Cardiol* 97:52C-60C.

65. Lazarou J, Pomeranz BH, Corey PN. 1998. Incidence of Adverse Drug Reactions in Hospitalized Patients. *JAMA* 279:1200-1205.

66. Schwarz UI. 2003. Clinical relevance of genetic polymorphisms in the human CYP2C9 gene. *Eur J Clin Invest* 33:23–30.