Genetic and Functional Studies Provide Insights into the Aetiologies of Familial Combined Hyperlipidemia

Helen Elizabeth Speedy

Thesis submitted to Queen Mary, University of London

For the degree of Doctor of Philosophy

January 2012

Centre for Endocrinology

William Harvey Research Institute

Barts and The London, Queen Mary's School of Medicine and Dentistry

Charterhouse Square

London

EC1M 6BQ

Acknowledgements

I would like to thank my supervisor, Professor Carol Shoulders, for her ongoing support, suggestions and advice through the duration of my project. I am especially grateful for the benefit of her experience, questions and opinions in the preparation of this thesis.

In addition, I would like to thank other members of the Genomic and Molecular Medicine Group, past and present, for all their assistance and support and for providing such a friendly environment in which to work. Firstly, I would like to thank Dr. Rossi Naoumova and her team (Claire Neuwirth, Elizabeth Potter and Yvonne Tan) for assembling the clinical samples and accompanying data that form the basis of this project. In addition, I would like to thank Stephanie Bonney, Emma Duncan, Stuart Horswell, Claire Hutchison, Bethan Jones, Emma Jones, Muddassar Mirza, Ruth Newton, Max Salm, Rasheeta Sivapackianathan and Paul Williams who have all contributed in some way to my studies.

I am especially grateful to Emma Jones for helping my settle into the group from my very first day. I would also like to thank Stuart Horswell for his advice with the computational aspects of this project and for his continued support even though he has now left the group.

I also wish to acknowledge the British Heart Foundation and Queen Mary, University of London for funding my studentship.

Finally, and most of all, I wish to thank my family for being so supportive in my education and career so far. Thank you to my parents for always encouraging me and to my husband, Darren for the countless ways in which he has helped me throughout my Ph.D.

This thesis is dedicated to my mother who, I am sure, would have been very proud to see me complete this work.

Abstract

The integration of biological and genetic data has established that diverse biological processes, involving multiple effectors, influence circulating levels of triglyceride and cholesterol. This diversity may underlie the genetic complexity of human dyslipidemias, including the common and highly atherogenic condition, Familial Combined Hyperlipidemia (FCHL). The aetiologies of FCHL are currently undetermined.

In this thesis, a multi-pronged approach was employed to identify genes/variants contributing to the linkage observed between the chromosome 21q22.2-22.3 interval and lipid traits, in white-British FCHL families. Additionally *GPIHBP1*, which encodes glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1, was studied. *GPIHBP1* represents a strong FCHL candidate gene due to its role in the lipolytic processing of triglyceride-rich lipoproteins.

Combined genetic and gene expression analyses, focussed upon a refined 3.8Mb interval on chromosome 21q22.3 that was linked to lipid abnormalities in subsets of FCHL families, identified two genes (*COL18A1* and *PKNOX1)* that warrant further investigation with regard to their contribution to FCHL. Promising results were also obtained for *C21orf57,* which resides just outside the 3.8Mb interval. Genetic association analyses in 1725 members of 239 FCHL families identified nominal association (P=0.0009) between a *TSPEAR* variant, rs34163868, and plasma triglyceride levels. Furthermore, transcript levels of *CBS* and *TRPM2* were significantly altered by treatment with the PPAR-agonist bezafibrate in a rat hepatoma cell line, thus implicating these genes in triglyceride/fatty acid metabolism.

In combined analysis of five independent cohorts, the minor allele of the *GPIHBP1* variant, rs11538388 was protective against hypertriglyceridemia (P=2.98x10^{-4}). The same allele was associated with decreased risk of coronary heart disease in the prospective Northwick Park Heart Study II (hazard ratio for carriers=0.76, P=0.0480) and delayed age of onset in the Southampton Atherosclerosis Study (odds ratio=0.76, P=0.0146). Collectively, these data demonstrate that the rs11538388 minor allele, or variant in linkage disequilibrium, is associated with more favourable processing of atherogenic lipoproteins.

Table of Contents

List of Figures

List of Tables

Abbreviations

ABC	ATP binding cassette
ACTH	Adrenocorticotropic hormone
ADH	Autosomal dominant hypercholesterolemia
Apo	Apolipoprotein
ARH	Autosomal recessive hypercholesterolemia
BMI	Body mass index
C21orf57	Chromosome 21, open reading frame 57
CBS	Cystathionine beta synthase
CD	Crohn's disease
CEPH	Centre d'Etude du Polymorphisme Humain
CEU	CEPH collection of Utah residents of Northern and Western European ancestry
CHD	Coronary heart disease
CHO	Chinese hamster ovary
CI	Confidence intervals
CNV	Copy number variation
COL18A1	Collagen, type XVIII, alpha 1
COP	Coat protein
COPD	Chronic obstructive pulmonary disease
DAG	Diacylgylcerol acyltransferase
DMSO	Dimethyl Sulfoxide
DS	Down syndrome
EGF	Epidermal growth factor
eQTL	Expression quantitative trait locus
ER	Endoplasmic reticulum
ES	Embryonic stem
EST	Expressed sequence tag
FBAT	Family-based associated test
FCHL	Familial combined hyperlipidemia
FHBL	Familial hypobetalipoproteinemia
GDT	Generalised disequilibrium test
GEE	Generalised estimating equations
GLGC	Global Lipid Genetics Consortium
GPIHBP1	Glycosylphosphatidylinositol-anchored high-density lipoprotein-binding

	protein 1
GWAF	Genome-Wide Association analyses with Family
GWAS	Genome-wide association study
HDL	High-density lipoprotein
HLOD	Heterogeneity LOD
HR	Hazard ratio
HSPG	Heparan sulphate proteoglycan
HTG	Hypertriglyceridemic
HWE	Hardy-Weinberg Equilibrium
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LME	Linear mixed effects
LOD	Logarithm of odds
LXR	Liver-X-receptor
McA-RH7777	McArdle rat hepatoma 7777
MAF	Minor allele frequency
MQLS	More-powerful quasi-likelihood score
NMI	Non-Mendelian inheritance
NPHSII	Northwick Park Heart Study II
NPL	Non-parametric LOD
OSA	Ordered subset analysis
PCR	Polymerase chain reaction
PKNOX1	PBX/knotted 1 homeobox 1
PPAR	Peroxisome proliferator-activated receptor
PUFA	Polyunsaturated fatty acids
RT-qPCR	Real-Time Reverse Transcription PCR
SAS	Southampton Atherosclerosis Study
sd	Standard deviation
SNP	Single nucleotide polymorphism
TBE	Tris-borate EDTA
TDT	Transmission disequilibrium test
TRPM2	Transient receptor potential cation channel, subfamily M, member 2
TSPEAR	Thrombospondin-type laminin G domain and EAR repeats
UTR	Untranslated region
VLDL	Very-low-density lipoprotein

Chapter 1 – General Introduction

According to World Health Organisation statistics, Coronary Heart Disease (CHD) is the leading cause of death across the globe. Based upon 2004 data, it was responsible for over 7.2 million annual deaths (Mackay and Mensah, 2004). More recent British Heart Foundation figures show that in 2007, around 91,000 people in the United Kingdom died as a result of CHD (URL1).

There are numerous established risk factors for the development of CHD. Tobacco exposure (Qiao et al., 2000), a lack of physical activity (Rosengren and Wilhelmsen, 1997) and poor diet (Hu et al., 2000) all increase risk. In a prospective study of 44,875 men, over the 8-year follow-up period, those consuming a 'Western' diet (defined by a greater consumption of red and processed meat, refined grains, sweets and high-fat dairy products) were at a 1.64-times (95% confidence intervals (CI): 1.24-2.17) greater risk of developing CHD than their peers who consumed a more prudent diet, rich in fruit, vegetables, wholegrains, fish and poultry (Hu et al., 2000). The polyunsaturated fats that are prominent in such a 'prudent' diet are believed to be important in reducing CHD risk. Thus in Eskimo populations the low rates of cardiovascular mortality, despite a high overall dietary fat consumption are attributed to their high levels of consumption of oily fish, rich in omega (n)-3 polyunsaturated fatty acids (PUFAs) (Bang et al., 1980). In an analysis of data from 11 cohort studies, including 344,696 participants over a 4-10 year follow-up period (median=6.5 years), it was observed that substituting 5% of daily energy intake from saturated fatty acids with PUFAs was associated with a decreased risk of death from a coronary event (hazard ratio (HR) 0.74, (95% CI: 0.61-0.89)) (Jakobsen et al., 2009).

The beneficial effects of n-3 PUFAs may, in part, be mediated by their impact upon plasma triglyceride levels. The meta-analysis of Balk and colleagues combined data from 17 studies to examine the effect of fish-oil n-3 fatty acids upon this trait (Balk et al., 2006). With daily doses ranging from ~0.1 to 5g per day, a net reduction of 27mg/dl (95% CI: –33 to –20) was observed. Overall, a 1g/day increase in fish oil dose was associated with an ~8mg/dL decrease in plasma triglycerides. This meta-analysis also showed a slight net increase in high-density lipoprotein (HDL)-cholesterol levels with fish-oil consumption of 1.6 mg/dl (95% CI: + 0.8 to +2.3), with no overall effect on total cholesterol levels (Balk et al., 2006).

Mechanistically, the triglyceride-lowering effects of n-3 fatty acids may be mediated by their effect upon lipogenic gene expression. Specifically, PUFAs lead to a reduction in hepatic nuclear protein levels of the transcription factor sterol regulatory element binding-protein (SREBP)-1c (Xu et al., 1999; Yahagi et al., 1999). SREBP-1c regulates the expression lipogenic genes such as fatty acid synthase (Magana and Osborne, 1996) and acetyl coenzyme A carboxylase (Magana et al., 1997).

Elevated circulating cholesterol and triglyceride levels accelerate the underlying pathological processes of heart disease (Arsenault et al., 2009; Bainton et al., 1992; Kannel et al., 1971; Lewington et al., 2007). The impact of high cholesterol on CHD mortality is illustrated by data published by the Prospective Studies Collaboration (Lewington et al., 2007) which involved 892,337 individuals, from 61 studies, of whom, over a mean follow-up period of 13-years, 33,744 died due to CHD. The mean baseline total cholesterol measurement of the study participants was 5.8mmol/l. Regression analysis revealed a log-linear relationship between every 1mmol/l drop in total cholesterol level and the risk of heart disease mortality, across all age groups. This effect was particularly evident in the younger study participants (40-49 years old) where each 1mmol/l drop in cholesterol level halved the risk of mortality (Lewington et al., 2007).

The strength of the association between triglyceride levels and CHD risk is more difficult to estimate since elevated triglyceride levels are concomitant with low HDL-cholesterol levels (Deckelbaum et al., 1984). Recently, Arsenault and colleagues examined data from 21,448 participants of the European Prospective Investigation Into Cancer and Nutrition cohort (Arsenault et al., 2009). Over an 11-year follow-up period, 2086 of the study participants developed heart disease (encompassing the clinical spectrum of the condition, from unstable angina to myocardial infarction). These individuals were more likely to be past or present smokers and undertake lower levels of physical activity than their healthy counterparts. In addition, those with higher (non-fasting) plasma triglyceride levels at the start of the study had increased hazard ratios for developing CHD (triglyceride 150-199mg/dl, HR 1.25 (95% CI: 1.12–1.39); triglyceride 200-249.9mg/dl, HR 1.26 (95% CI: 1.10–1.43) and triglyceride >250mg/dl, HR 1.57 (95% CI: 1.38–1.79). This increased risk was independent of cholesterol level (Arsenault et al., 2009). Hopkins and colleagues similarly reported an independent association between plasma triglyceride levels and CHD risk when examining the lipid levels of 653 premature familial CHD cases and 1029 controls (Hopkins et al., 2005). In their study, plasma triglyceride levels above

200mg/dl, adjusted for HDL-cholesterol level were associated with increased risk of CHD. Specifically, the odds ratios for CHD were 1.7, 2.8 and 11.4 in individuals with triglyceride levels of 200-299, 300-499 and 500-799mg/dl, respectively when compared to those with levels of <100mg/dl ($P \leq 0.01$) (Hopkins et al., 2005).

Plasma cholesterol and triglyceride levels have significant heritable components, each in the order of 50% (Heller et al., 1993; Kathiresan et al., 2007; Weiss et al., 2006). The study of Heller and colleagues examined correlation between lipid levels of 302 pairs of twins raised together (156) or raised apart (146). Using model-fitting approaches, they estimated a heritability of 0.46 for total cholesterol and 0.52 for triglyceride (Heller et al., 1993). Additionally Kathiresan et al., estimated heritability of fasting total cholesterol and triglyceride to be 0.57 and 0.48 (both $P<0.0001$), respectively using data from extended family members participating in the Framingham Heart Study (Kathiresan et al., 2007). Their model employed maximum likelihood estimation to a mixed effects model, incorporating additive genetic and fixed covariate effects to produce heritability estimates broadly in line with those of Heller and colleagues.

Cholesterol and triglyceride are primarily carried through the circulation in the form of lipoprotein particles. The majority of plasma cholesterol exists in low-density lipoproteins (LDL) whilst triglyceride circulates as a component of very low-density lipoprotein (VLDL) and chylomicron particles. These particles consist of a neutral lipid core containing triglyceride and cholesterol esters, with an outer shell of phospholipid and unesterified cholesterol around which apolipoprotein (apo) B, the major structural component of these moieties, is wrapped in a bow-like structure (Chatterton et al., 1995).

A recent lipidomic analysis of pooled fasting plasma samples from 100 individuals, aged 40-50 years, from ethnic backgrounds representative of the United States population, revealed a remarkable diversity in the lipid species present (Quehenberger et al., 2010). In total, 588 different plasma lipid species were recorded, including 36 species of sterol lipids and 73 species of glycerolipid (tri- and diacylglycerol) (Quehenberger et al., 2010). Many of these species may be carried by the apoB-containing lipoproteins and as such could affect their properties.

Cholesterol, the predominant plasma sterol (constituting 99% of free and esterifed sterols (Quehenberger et al., 2010)), has diverse physiological functions. As a

membrane component, cholesterol participates in signalling activities (Simons and Toomre, 2000). It acts as an essential precursor for oxysterols (Schroepfer, 2000) and bile acids (Russell, 2003) and is also required for myelination (Chrast et al., 2011). In addition, 7-dehydrocholesterol, the cholesterol precursor, can be converted into vitamin D (1.2.2). Triglycerides (triacylglycerols) meanwhile provide an essential energy source. Triglyceride is the final product of the multi-step glycerolipid synthesis pathway which may also provide lipid intermediates such as phosphatidic acid, lysophosphatidic acid and diacylglycerol for participation in signalling pathways (Eyster, 2007; Tigyi, 2010).

1.1) Classification of Hyperlipidemias and their Biology

For over four decades, hyperlipidemias have been classified on the basis of the elevated plasma lipoprotein fraction by which they are characterised (Fredrickson and Lees, 1965). However, given the evident diversity in plasma lipid species, this measurement of circulating lipoproteins may, in most people, represent only a simplified readout of the complex biological processes that underpin whole-body lipid metabolism.

It is now apparent that defects in lipid absorption from the diet, *de novo* synthesis, lipoprotein production and secretion, clearance or catabolism disrupt the balance in circulating lipid levels. To illustrate this, Table 1.1 summarises current knowledge regarding the heterogeneous array of genetic variants that predispose to dyslipidemia, while the following sections describe important lipid metabolic processes and how both common and rare genetic variants within the effectors of these processes can cause perturbation in circulating lipid levels. This information serves as a platform for further understanding of where the molecular defects that underlie the complex genetic disorder, Familial Combined Hyperlipidemia may lie.

Table 1.1) Genes Implicated in Monogenic or Polygenic Dyslipidemia and Perturbations of Cellular Lipid Processing

			Common Variants Associated with Plasma TG/TC	
Gene	Main Function	Mendelian Disorder	Variant[a]	Effect of Minor Allele
NPC1L1	Absorption: cholesterol uptake	None[b]	rs2072183	TC +
ABCG5/8	Absorption: cholesterol efflux	Sitosterolemia	rs4299376	TC +
HGMCR	*De novo* cholesterol synthesis	None	rs12916	TC +
DHCR7	*De novo* cholesterol synthesis	Smith-Lemli Opitz syndrome	None	na
AGPAT2	Triglyceride synthesis	Congenital Generalised Lipodystrophy	None	na
APOB	CM & VLDL assembly/secretion/	FHBL/	rs1042034/	TG -
	Lipoprotein clearance/catabolism	ADH-2	rs1367117	TC +
MTTP	CM & VLDL assembly/secretion	Abetalipoproteinemia	None	na
SAR1B	CM & VLDL assembly/secretion	Chylomicron retention disease	None	na
LPL	CM & VLDL catabolism	Lipoprotein lipase deficiency	rs12678919	TG -
LMF1	CM & VLDL catabolism	Combined lipase deficiency	None	na
GPIHBP1	CM & VLDL catabolism	'Monogenic' HTG/Chylomicronemia	None	na
APOC2	CM & VLDL catabolism	'Monogenic' HTG/Chylomicronemia	None	na
APOA5	CM & VLDL catabolism	'Monogenic' HTG/Chylomicronemia	rs964184[c]	TG/TC +
ANGPTL3	CM & VLDL catabolism	'FHBL'/familial combined hypolipidemia	rs2131925	TG/TC -
APOE	ApoE-containing lipoprotein	Dysbetalipoproteinemia	rs4420638	TC +
	clearance		rs439401	TG -

			Common Variants Associated with Plasma TG/TC	
Gene	Main Function	Mendelian Disorder	Variant[a]	Effect of Minor Allele
LDLR	ApoB/ApoE-containing lipoprotein clearance	ADH-1	rs6511720	TC -
LDLRAP	ApoB-containing lipoprotein clearance	ARH	rs12027135	TC -
PCSK9	ApoB-containing Lipoprotein clearance	ADH-3	rs2479409	TC+
NPC1/2	Cholesterol utilisation	Niemann-Pick disease, type C	None	na
MLN64	Cholesterol utilisation	None	rs11869286	[d]
CYP7A1	Cholesterol utilisation	None[e]	rs2081687	TC +

'+'=increased, '-'=decreased, ADH=autosomal dominant hypercholesterolemia, ARH=autosomal recessive hypercholesterolemia, CM=chylomicron, FHBL=familial hypobetalipoproteinemia, HTG=hypertriglyceridemia, TC=total cholesterol, TG=triglyceride, VLDL=very low-density lipoprotein

[a] Lead SNP at locus from Global Lipid Genetics Consortium data (Teslovich et al., 2010)

[b] Low sterol absorbers enriched for rare *NPC1L1* variants (Cohen et al., 2006)

[c] Variant resides at *APOA1/C3/C4/A5* locus

[d] Variant associated with decreased high-density lipoprotein cholesterol

[e] *CYP7A1* frameshift mutation identified in family with statin-resistant hypercholesterolemia (Pullinger et al., 2002)

1.2) Molecular Determinants of Circulating Cholesterol and Triglyceride

1.2.1) Effectors of Dietary Sterol and Triglyceride Absorption

The recommended upper limit for dietary cholesterol consumption in the United States is 300mg per day (Lichtenstein et al., 2006). Cholesterol is the predominant sterol absorbed from the diet; only approximately 5% of plant-derived sterols such as campesterol and sitosterol, and stanols are absorbed (Salen et al., 1992). There is however, a high degree of inter-individual variability in the extent to which dietary sterols are taken up. For example, Bosner and colleagues used dual stable isotopic tracers, given intravenously and orally, to assess cholesterol absorption (Bosner et al., 1999). Absorption was calculated by determining the ratios of plasma tracer to the administered tracer. In 94 adults, absorption values ranged from 29.0% to 80.1% (mean=56.3±12.1%, standard deviation, sd). In a similar study of 18 adults, absorption rates varied from 15.3% to 57.9% (mean=40.7±2.3% sd), though absorption efficiency was decreased when subjects were given high doses (188mg and 421mg) of dietary cholesterol (Ostlund et al., 1999).

A common approach to measuring cholesterol absorption is to use the ratio of serum non-cholesterol sterols to cholesterol. Specifically, the ratios of serum cholesterol to the absorption sterols campesterol and sitosterol are an indicator of cholesterol absorption, whereas ratios of cholesterol to precursor sterols such as lathosterol and desmosterol are an indicator of cholesterol synthesis (Miettinen et al., 1989; Miettinen et al., 1990). Using this methodology in siblings of high and low cholesterol absorbing probands, Gylling and Miettinen showed that the variation in cholesterol absorption efficiency is partly due to inherited factors (Gylling and Miettinen, 2002).

Niemann-Pick C1 like-1 (NPC1L1) protein and adenosine triphosphate binding cassette (ABC) G5/G8 heterodimers are effectors of the sterol absorption process which involves a balance between sterol uptake (NPC1L1) into the enterocyte and efflux back into the intestinal lumen (ABCG5/8) (Figure 1.1).

The properties of NPC1L1, first identified in rodent jejunal cDNA, include those fundamental to a membrane transporter: a signal peptide, predicted transmembrane domains, N-linked glycosylation sites, as well as a sterol-sensing domain (Altmann et al., 2004). In *Npc1l1*-deficient mice cholesterol absorption was reduced ~69% compared to age-matched wild-type mice yet when treated with the drug ezetimibe,

which inhibits the uptake of dietary and biliary cholesterol, no further reduction occurred, thus implying that NPC1L1 is involved in the ezetimibe-sensitive cholesterol absorption pathway (Altmann et al., 2004). *Npc1l1*-deficient mice also have dramatically reduced plasma phytosterol levels and decreased uptake of sitosterol (Davis et al., 2004), illustrating that NPC1L1 is a relatively non-selective transporter of sterols.

ABCG5/ABCG8 are predominantly expressed in the small intestine and liver where they act as obligate heterodimers (Berge et al., 2000; Graf et al., 2002; Graf et al., 2003). In humans, ABCG5 and ABCG8 are normally positioned at the apical membrane of the intestinal enterocytes were they function to efflux sterols back into the intestinal lumen (Figure 1.1). In the liver they are found in cells lining the hepatobiliary tract, both hepatocytes and bile ductular cells and function by promoting sterol efflux into bile (Klett et al., 2004a; Yu et al., 2002).

In *Abcg8*- knockout mice the secretion of cholesterol into bile is impaired (Klett et al., 2004b) whilst in mice overexpressing both *Abcg5/8*, fractional cholesterol absorption is decreased by ~50% and biliary cholesterol secretion is increased three- to six-fold over that of wild-type mice. A compensatory increase in cholesterol synthesis (assessed using plasma lathosterol levels) is observed (Klett et al., 2004a; Yu et al., 2002).

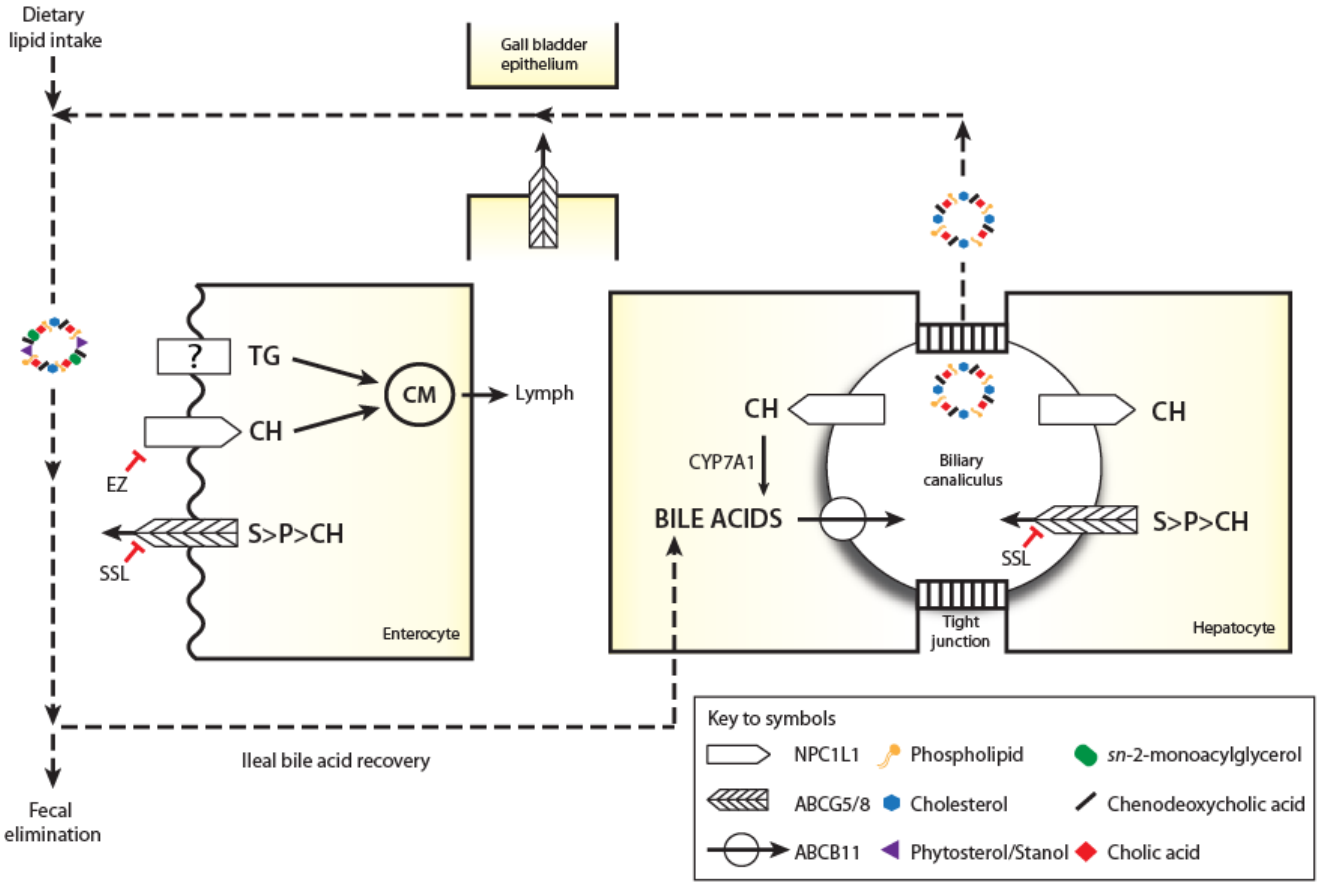


Figure 1.1) Effectors of Intestinal Lipid Uptake

Mixed micelles, synthesised in biliary canaliculi and stored in the gall bladder until required, facilitate the solubilisation and subsequent absorption of dietary lipids. Solubilised sterols are taken up across the brush border of the enterocyte by NPC1L1 in an ezetimibe (EZ)-

sensitive process (Altmann et al., 2004). Three proteins, expressed at the enterocyte brush border, are implicated in mediating the uptake of hydrolysed triglyceride (TG) in the form of s*n-2*-monoacylglycerol/fatty acids: plasma membrane associated fatty acid binding protein (Stremmel et al., 1985), fatty acid transport protein 4 (Stahl et al., 1999) and fatty acid transporter, CD36 (Poirier et al., 1996). However, the involvement of any of these in mediating fatty acid or *sn*-2-monoacylglycerol uptake into the enterocyte has yet to be conclusively determined. Absorbed cholesterol (CH) is packaged, with re-synthesised TG and apolipoprotein-B48 into chylomicron (CM) particles and secreted into the circulation. In contrast, dietary non-cholesterol sterols (P) and stanols (S) are preferentially pumped back into the intestinal lumen via ABCG5/8. This process is blocked in the autosomal recessive condition sitosterolemia, leading to a characteristic accumulation of non-cholesterol sterols in the body. In the liver, sterols, bile acids and phospholipids are secreted into the biliary canaliculus to form mixed micelles which are stored by the gall bladder until release is stimulated by the arrival of chyme in the duodenum. Further sterols may be transported into bile across the gall bladder epithelium by ABCG5/8. In the small intestine, the micelles solubilise and facilitate the uptake of dietary lipids and fat-soluble vitamins. Around 95% of bile acids contained in mixed micelles are recycled to the liver, via the portal circulation, whilst the remainder are eliminated in faeces. (Adapted from (Calandra et al., 2011).

Rare and common sequence variants in *NPC1L1* and *ABCG5/ABCG8* are associated with variation in cholesterol absorption efficiency and plasma sterol levels. However, only rare mutations in *ABCG5* or *ABCG8* cause a truly recessive disorder, namely sitosterolemia (Berge et al., 2000; Lee et al., 2001). Sitosterolemia is characterised by the accumulation of plant sterols in the circulation and in tissues, apart from the brain (Salen et al., 1985). Sitosterol levels in affected individuals tend to be 10-25 times higher than those in unaffected individuals (Lee et al., 2001). The clinical presentation of the disorder includes tendon and tuberous xanthomas as well as accelerated atherosclerosis, leading to premature CHD (Salen et al., 1992).

Genetic variants in *NPC1L1* are associated with both low (Cohen et al., 2006; Fahmi et al., 2008) and high sterol levels (Polisecki et al., 2010). Cohen and colleagues examined the coding region of *NPC1L1* in 128 high cholesterol absorbers and 128 low cholesterol absorbers (as determined by plasma campesterol:lathosterol ratio, with a high ratio indicating a high rate of absorption). They found an enrichment of non-synonymous sequence variants in *NPC1L1* in the low sterol absorbing African American subjects (Cohen et al., 2006). A subsequent study showed that 14 of the 20 non-synonymous variants identified were associated with decreased NPC1L1 expression in a Chinese Hamster Ovary (CHO) cell line. In addition, cholesterol absorption, measured using stable isotope tracers, was decreased in members of two extended families harbouring the I647N or R693C variants (Fahmi et al., 2008).

In the PROSPER cohort, the minor alleles of two common *NPC1L1* variants, rs41279633 and rs207218, were associated with increased plasma LDL-cholesterol levels. Homozygotes for the rs41279633 rare allele were also more likely to develop CHD (fatal or non-fatal) over the 3.2-year follow-up period of the study than carriers of the major allele (HR 1.67 (95%CI: 1.10-2.54), P=0.04) (Polisecki et al., 2010).

Common *ABCG5/ABCG8* variants are also associated with sterol metabolism. For example, rs11887534, a non-synonymous variant which changes the 19th amino acid in ABCG8 from aspartic acid to histidine, is associated with decreased sterol absorption. In a study of 148 unrelated, healthy individuals, carriers of the 19H allele (n=14) had significantly lower plant sterol/stanol:cholesterol ratios than those homozygous for the 19D allele (n=128) (sitosterol:cholesterol 0.94±0.32 versus 1.28±0.45, campesterol:cholesterol 1.21±0.36 versus 1.67±0.62 and cholestanol:cholesterol 1.76± 0.31 versus 2.09±0.40, all P<0.005). These lower

ratios are indicative of decreased cholesterol absorption in 19H homozygotes (Berge et al., 2002).

In a further study involving 263 mildly hypercholesterolemic individuals, the 19H allele was overrepresented in the lowest tertile of cholesterol absorbers whilst the 19D allele was more common in the upper tertile (both n=87, $P<0.001$) (Gylling et al., 2004). These findings were more recently substantiated in a meta-analysis involving over 20,000 individuals where Kathiresan *et al,* demonstrated significant association ($P=1x10^{-10}$) between the D19H variant and LDL-cholesterol levels (Kathiresan et al., 2009b). In addition, the same study identified a second, independent association between another *ABCG8* single nucleotide polymorphism (SNP), rs6544713, and LDL-cholesterol levels (r^2 between the two markers=0.02) (Kathiresan et al., 2009b).

Both D19H and a marker in complete linkage disequilibrium (LD) with rs6544713 (rs4299376) were also associated with risk of gallstone disease (cholelithiasis) in a case-control genome-wide association study (GWAS) (Buch et al., 2007), thus suggesting that *ABCG8* gain-of-function variants increase sterol efflux into bile. This effect could also explain the association of the 19H allele with decreased circulating (LDL) cholesterol levels and decreased sterol absorption.

In summary, both rare variants of moderate to large effect and common variants with a small effect size can perturb the molecular process involved in net sterol absorption and in turn modulate differences in plasma sterol levels.

In contrast to intestinal sterol uptake, relatively little is known about absorption of dietary triglycerides. Since triglycerides cannot be absorbed directly by the enterocytes they must initially undergo hydrolysis to yield *sn*-2-monoacylglycerol (plus fatty acids). This process occurs primarily in the small intestine, under the action of pancreatic lipase, the deficiency of which causes fat-malabsorption, leading to steatorrheoa (Figarella et al., 1980). The resultant fatty acids and *sn*-2-monoacylglycerol are incorporated into bile salt mixed micelles and it is proposed that their uptake into the enterocytes occurs via a protein-mediated process (Murota and Storch, 2005). Using an *in vitro* model of the human post-prandial intestinal milieu, Murota and Storch demonstrated that the uptake of both resultant fatty acids and *sn*-2-monoacylglycerol was saturable at low concentrations of lipid, suggesting that it was facilitated by a protein-carrier. However, at higher concentrations, a linear

relationship existed between uptake and lipid concentration, suggesting that some passive diffusion may also occur (Murota and Storch, 2005).

For transport to peripheral sites, triglyceride must be re-synthesised (1.2.4) and then assembled into chylomicrons (1.2.5) for secretion into the circulation.

1.2.2) *De novo* Cholesterol Biosynthesis

De novo cholesterol synthesis occurs via a multi-step pathway (Figure 1.2) with the liver (Ho and Taylor, 1970), intestine (Sviridov et al., 1986) and brain (Jurevics and Morell, 1995) all being important sites of production. *De novo* cholesterol synthesis is controlled via transcriptional and post-translational mechanisms that regulate the endoplasmic reticulum (ER)-membrane protein, 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG CoA reductase, HMGCR), in response to cellular sterol levels (Kuwabara and Labouesse, 2002). HMGCR consists of N-terminal domain, which has 8 transmembrane helices, including a sterol-sensing domain encompassing helices 2-6, as well as a cytosolic C-terminal domain which possesses the catalytic properties required to generate mevalonate from HMG-CoA (Liscum et al., 1985; Roitelman et al., 1992).

Common *HMGCR* variants have recently been associated with perturbations in LDL-cholesterol levels. An initial GWAS involving the population of the Micronesian island of Kosrae, identified three *HMGCR* variants whose minor alleles were associated with slight increases in LDL-cholesterol level (0.30 mmol/l). Further investigation found that these markers were in LD with a functional variant, rs3846662 which was found to modulate *HMGCR* pre-mRNA alternative splicing (Buckhardt et al., 2008). The results of this study are further substantiated by the findings of the Global Lipid Genetics Consortium (GLGC) meta-analysis which showed significant association between the minor (C) allele of rs12916 (in strong LD with rs3846662) and LDL-cholesterol levels (LDL-cholesterol +0.07mmol/l per copy of C allele, P=9x10^{-47}) (Teslovich et al., 2010). Thus, these studies provide fine examples of genetic determinants that increase circulating cholesterol levels through increased *de novo* cholesterol synthesis.

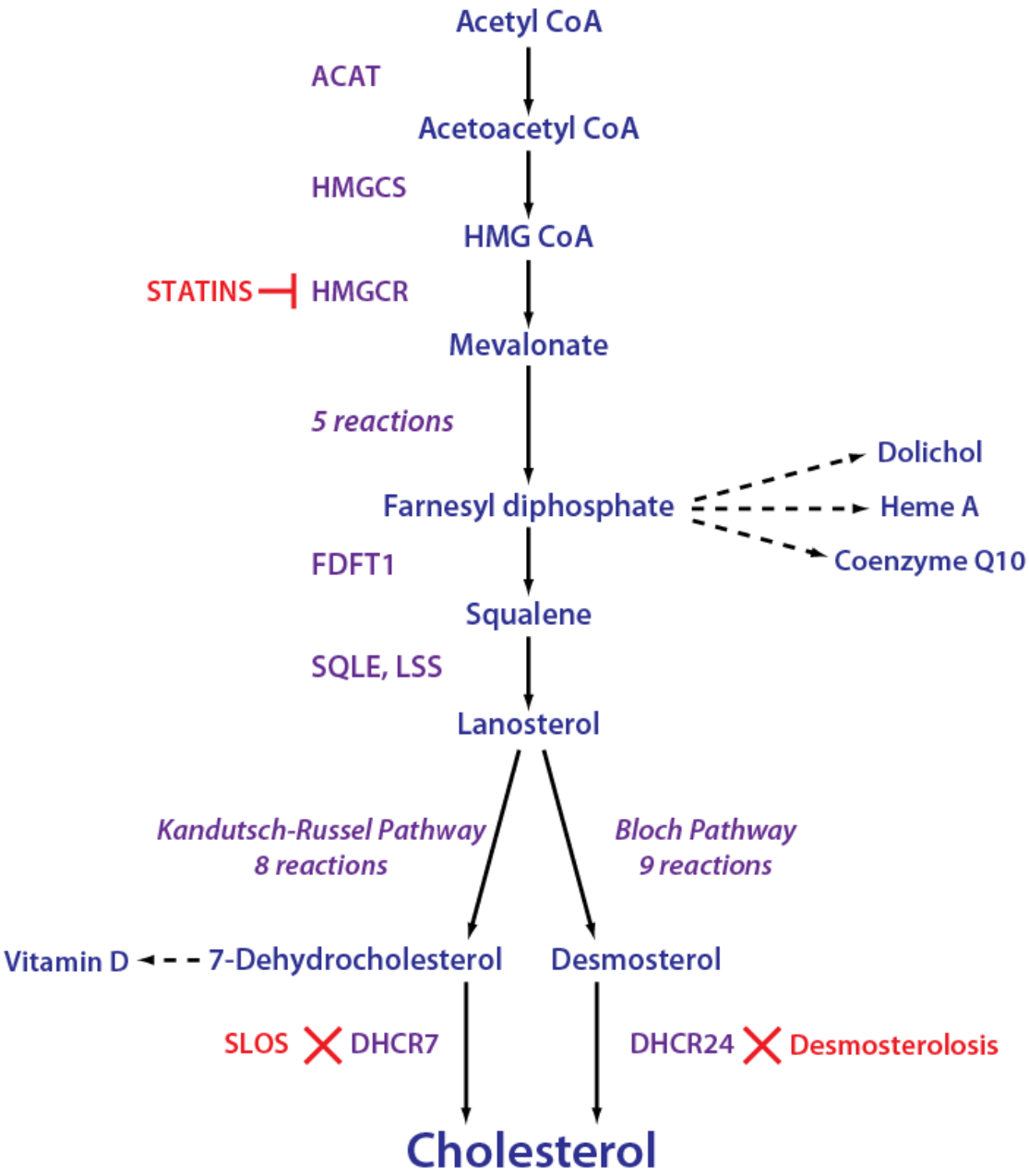


Figure 1.2) Overview of the Cholesterol Biosynthesis Pathway

Cholesterol biosynthesis is multi-step pathway which begins with the precursor molecule acetyl-CoA. Acetyl CoA can be converted into fatty acids for storage or be utilised for the synthesis of cholesterol or non-sterol isoprenoids. The conversion of HMG-CoA to mevalonate by HMG-CoA reductase (HMGCR) is an irreversible reaction and mevalonate is a crucial building block for the formation of isoprenoid groups that are incorporated into a variety of products including cholesterol, heme, dolichol and ubiquinone. HMGCR is the target of the hypolipidemic drugs, statins. Subsequently in the pathway, squalene epoxide undergoes cyclisation to lanosterol, by lanosterol synthase (LSS), and this represents the first committed step to sterol

synthesis. Lanosterol can be converted to cholesterol via two routes, differentiated by the by the timing of the reduction of the Δ24-bond in the aliphatic side chain. The most common pathway is the Kandutsch-Russel pathway (Porter and Herman, 2011). In this pathway, cholesterol is ultimately formed by the reduction of 7-Dehydrocholesterol by 7-Dehydrocholesterol reductase (DHCR7). Mutations of *DHCR7* or sterol Δ24-reductase (*DHCR24)* cause the human malformation syndromes Smith-Lemli Opitz syndrome (SLOS, (Tint et al., 1994)) and desmosterolosis (FitzPatrick et al., 1998), respectively. SLOS can result in plasma cholesterol levels are less than 5% of those measured in controls (Tint et al., 1994). ACAT=acetyl-Coenzyme A acetyltransferase, HMGCS=HMG-CoA synthase, SQLE=squalene epoxidase

1.2.3) Regulation of *De novo* Cholesterol Biosynthesis by Control of HMGCR

When cellular sterol levels are high, accumulation of the cholesterol precursor lanosterol triggers the proteasomal mediated degradation of HMGCR (Song et al., 2005), a process that may involve binding to the sterol-sensing domain of HMGCR. Upon sterol accumulation, HMGCR binds to the ER protein, insulin induced gene 1 (Insig-1) and is subsequently degraded via the proteasome (Sever et al., 2003). The Insig-mediated degradation of HMGCR is thought to involve the interaction of Insig-1 with the E3 ubiquitin ligase, gp78. The prevailing view is that, in concert, gp78, the E2 ubiquitin conjugating enzyme UBE2G2 and other recruited cofactors, facilitate ubiquitination of HMGCR on lysines 89 and 248 and target it to the proteasome for degradation (Cao et al., 2007).

When sterol levels are low, expression of *HMGCR* is facilitated by SREBPs. The active forms of these transcription factors are delivered to the nucleus with the aid of the escort protein, SREBP cleavage activating protein (SCAP), which associates with SREBP by its C-terminus (Hua et al., 1996). However, as sterols accumulate they bind to the sterol-sensing domain of SCAP (Nohturfft et al., 1998). This sterol-SCAP binding causes a conformational change in SCAP (Brown et al., 2002) which permits it to bind to Insig-1 (Yang et al., 2002) or Insig-2 (Yabe et al., 2002). The Insigs cause the SCAP:SREBP complex to be retained in the ER preventing transit to the Golgi and subsequently the nucleus, thus impairing trans-activation of SREBP target genes, including *HMGCR* (Yabe et al., 2002; Yang et al., 2002).

Studies utilising ezetimibe and statin treatment, to block cholesterol absorption (NPC1L1-mediated) and synthesis (via HMG-CoA reductase inhibition), respectively, illustrate the homeostatic interplay that exists between cholesterol absorption and synthesis. For example, in a five-year study of simvastatin treated CHD patients, serum levels of plant sterols (markers of cholesterol absorption) were increased whereas levels of cholesterol precursors (as markers of cholesterol synthesis) were decreased (Miettinen et al., 2000). In another study, two-week treatment of hyperlipidemic men with atorvastatin, led to decreased ratios of the cholesterol precursor lathosterol to the plant sterols campesterol and sitosterol (87.1 and 88.4%, respectively, both $P<0.0001$), when compared to those given a placebo (Tremblay et al., 2011). In addition, intestinal levels of NPC1L1 mRNA were increased by 19% ($P=0.03$) thus collectively demonstrating an increase in cholesterol absorption when cholesterol synthesis is blocked (Tremblay et al., 2011). Conversely, the ratio of

plasma lathosterol to cholesterol in 18 men treated with ezetimibe for two-weeks was significantly higher than when a placebo was given (ezetimibe 1.895μg/mg, placebo 1.103μg/mg, *P*=0.001), indicating an increase in cholesterol biosynthesis (Sudhop et al., 2002). Therefore, when either the endogenous or exogenous source of cholesterol is blocked, mechanisms to upregulate the alternative source appear to be activated.

1.2.4) Triglyceride Biosynthesis

Synthesis of triglyceride is a multi-step process, especially prominent in the small intestine, adipose tissue, liver and mammary gland and is required for diverse biological processes including, dietary lipid absorption, storage, lipoprotein assembly/secretion, lactation and signal transduction (Coleman and Lee, 2004). There are two distinct pathways by which triglyceride can be synthesised; the monoacylglycerol pathway or the glycerol-3-phosphate pathway (Figure 1.3).

The human genome encodes multiple glycerol-3-phosphate acyltransferases (GPATs), acylglycerol-P acyltransferases (AGPATs), lipins and diacylgylcerol acyltransferases (DAGs). This is a likely reflection of the extensive fatty acid (n=31), diacylglycerol (n=55), triacylglycerol (n=18) and phospholipid (n=160) species diversity that exists in human plasma (Quehenberger et al., 2010) and of differences in the sites of expression of these enzymes.

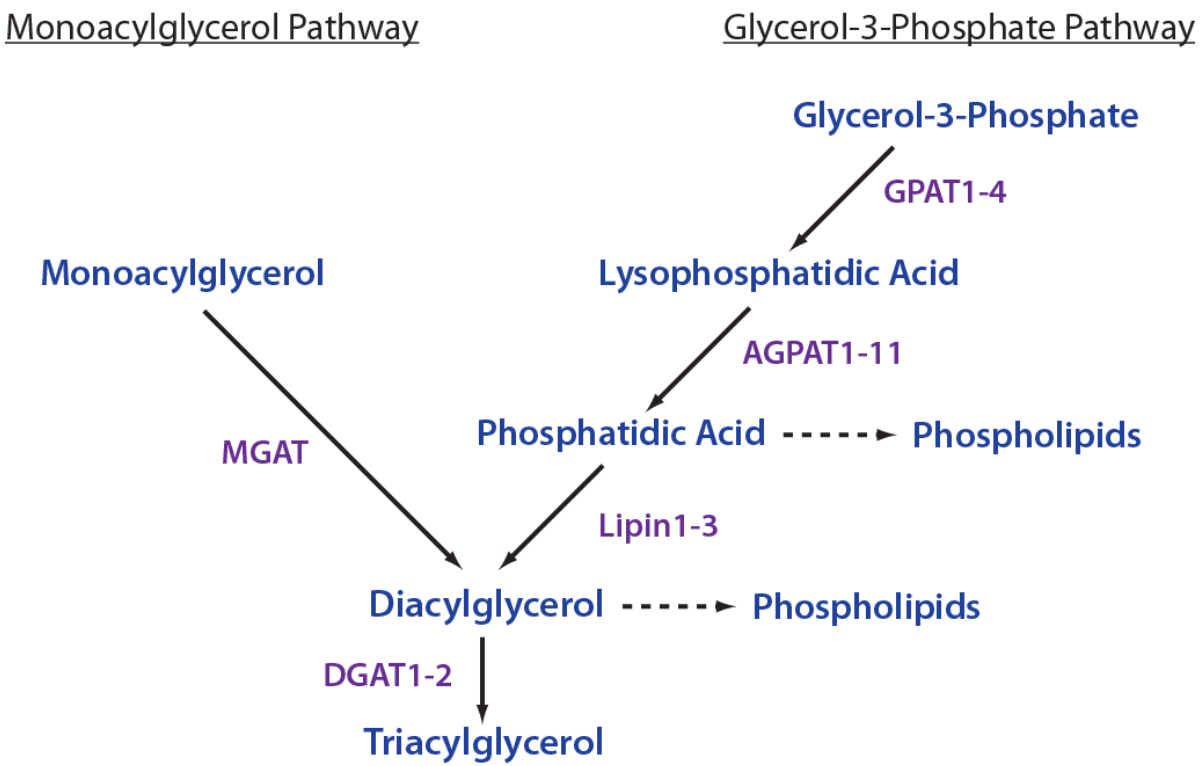


Figure 1.3) Triacylglycerol Synthesis Pathways

In the enterocytes, the monoacylglycerol pathway is the predominant route of diacylglycerol (DAG) synthesis. DAG is synthesised by the acylation of *sn*-2-monoacylglycerol (formed by pancreatic lipase hydrolysis of dietary triacylgycerol), catalysed by monoacylglycerol acyltransferase (MGAT). DAG is then acylated further by the action of (DGAT) to form triacylglycerol.

Most other cell types utilise the glycerol-3-phosphate pathway, whereby successive acylations generate triacylglycerol. The first committed step in triglyceride synthesis is catalysed by glycerol-3-phosphate acyltransferase (GPAT) and involves the acylation of glycerol-3-phosphate by fatty acyl-CoA to form lysophosphatidic acid. The activation of fatty acids by thioester linkage to coenzyme A, mediated by acyl coA synthetase, produces the fatty-acyl-CoA for this initial reaction. Mutations of AGPAT2 cause the autosomal recessive disorder congenital generalised lipodystrophy, which is characterised by lack of visceral and subcutaneous adipose tissue, hypertriglyceridemia and hepatic steatosis (Agarwal et al., 2002).

1.2.5) Assembly and Secretion of ApoB-Containing Lipoproteins

The production and secretion of apoB-containing lipoproteins, for the transport of lipids into the circulation is complex, involving multiple cellular compartments. The first step of lipoprotein assembly involves production of an apoB-containing precursor molecule, through partial lipidation of apoB, which occurs concomitantly with its translation in the endoplasmic reticulum. Subsequently, bulk lipidation of the precursor is thought to take place. This is presumed to occur through fusion of the precursor with a triglyceride-rich (apoB-free) particle. The N-terminal βα1 domain of apoB is vital for the initiation of lipoprotein assembly (Herscovitz et al., 1991; Wang et al., 2010). Lipoprotein assembly involves microsomal triglyceride transfer protein (MTP). In mouse hepatocytes (which can secrete both apoB-48 and apoB-100 containing lipoproteins), treatment with the MTP inhibitor BMS-197636 caused a 90% decrease in the secretion of apoB-100 and led to the secretion of lipid-poor apoB-48 lipoproteins (Kulinski et al., 2002). Inhibition of MTP also led to depletion of non-apoB triglyceride in the microsomal lumen but did not block transfer of these pools to the partially lipidated apoB particle (Kulinski et al., 2002). The significance of this finding is unknown.

Mutations of *APOB* are a cause of familial hypobetalipoproteinemia (FHBL), an autosomal dominant condition characterised by very low plasma levels (<5th percentile of the population distribution) of LDL-cholesterol and apoB (Schonfeld et al., 2005). Heterozygotes for FHBL mutations may be asymptomatic. Rare homozygotes have a more severe phenotype and may exhibit a failure to thrive, intestinal fat malabsorption and develop fatty liver. Most FHBL-*APOB* mutations cause premature truncation of the encoded protein and there is a general inverse correlation between phenotype severity and protein length (Calandra et al., 2011). Truncated proteins shorter than apoB-30 are not detectable in the plasma of patients (Tarugi et al., 2007), resulting in a clinical presentation of severe fat malabsorption (Calandra et al., 2011). Furthermore, whilst lipoproteins containing longer truncated apoB, such as apoB-89, are secreted, these moieties are themselves rapidly cleared from the circulation (Parhofer et al., 1992) thus resulting in hypocholesterolemia.

Mutations of the *MTTP* gene lead to the rare autosomal recessive disorder abetalipoproteinemia (Narcisi et al., 1995). The defect in lipoprotein assembly that occurs in this condition may also lead to a compensatory increase in *de novo* cholesterol synthesis, as illustrated by almost double blood levels of dolichols

(synthesised through the cholesterol biosynthetic pathway) in patients with abetalipoproteinemia compared to control individuals (Kuriyama et al., 1999).

The processes regulating the transport of lipidated-apoB particles from the ER to Golgi are not yet fully understood. Protein transport from the ER to Golgi is typically facilitated by coat protein (COP) complexes. The COPII complex consists of the proteins Sar1 GTPase, Sec23/24p complex, and Sec13/31p (Barlowe, 2002). The transport process initiates with the exchange of guanosine diphosphate (GDP) on Sar1 for guanosine triphosphate (GTP) in a reaction catalysed by Sec12, a guanine nucleotide exchange factor. This reaction, in turn, leads to the recruitment of other COPII proteins to regions of the ER membrane and subsequent vesicle budding, generating ER to Golgi carriers (Barlowe, 2002).

Triglyceride-rich lipoproteins are large; indeed chylomicrons may reach >250nm in diameter (Bhattacharya and Redgrave, 1981) whilst, in contrast, COPII vesicles are generally 60-80nm in diameter. These size differences led Siddiqi and colleagues to propose a specialised transport vesicle for ER to Golgi chylomicron transport (Siddiqi et al., 2003). Although the exact mechanism of ER to Golgi trafficking of chylomicrons is not established, it is known that the COPII protein, Sar1b is required. Mutations of the *SARA2* gene which encodes Sar1b lead to Chylomicron Retention Disorder (Jones et al., 2003). This rare autosomal recessive disorder is characterised by a failure to thrive, severe fat malabsorption and the absence of circulating apoB48-containing lipoprotein particles. Extremely low plasma levels of LDL- and HDL-cholesterol also result (Nemeth et al., 1995; Roy et al., 1987)

1.2.6) Catabolism of Triglyceride-Rich Lipoproteins

Mendelian forms of severe hypertriglyceridemia are genetically heterogeneous, involving mutations in a minimum of five genes. Specifically, lipoprotein lipase (*LPL*), *APOA5*, *APOC2* and the recently characterised genes/proteins glycosylphosphatidylinositol-anchored high-density lipoprotein–binding protein 1 (*GPIHBP1*) and lipase maturation factor 1 (*LMF1*) are all implicated in the catabolism of triglyceride-rich lipoproteins.

The triglyceride-rich cores of chylomicron and VLDL particles are hydrolysed at the capillary endothelium by LPL, thereby releasing fatty acids to peripheral tissues. LPL is abundant in adipose tissue, heart or skeletal muscle and mammary gland

presumably because these tissues require large amounts of fatty acids for storage or energy (Camps et al., 1990).

Active LPL exists as a homodimer (Osborne et al., 1985). Correct folding and assembly of LPL into this form requires LMF1. In mice with combined lipase deficiency (*cld*), severe and progressive hypertriglyceridemia causes death a few hours after birth (Paterniti et al., 1983). This extreme phenotype results from a lack of active LPL (and hepatic lipase) which must occur post-translationally since in *cld/cld* mice LPL mRNA and protein levels are unaffected (Davis et al., 1990). LMF1 was identified as the gene underlying this defect by Péterfy and colleagues: transfection of the full-length *LMF1* cDNA into a *cld* cell line was able to restore LPL activity (Peterfy et al., 2007). The LMF1 mutations Y439X and W464X cause combined lipase deficiency in humans (Cefalu et al., 2009; Peterfy et al., 2007).

GPIHBP1 is thought to transport LPL to the capillary endothelial cell surface from the parenchymal cells within which it is synthesised (Davies et al., 2010). In *Gpihbp1*-deficient mice, LPL is mislocalised to the outside of the parenchymal cells, rather than at the capillary endothelium (Davies et al., 2010). In addition, *Gpihbp1*-null mice are severely chylomicronemic (Beigneux et al., 2007) and rare homozygous mutations of *GPIHBP1* in humans cause severe hypertriglyceridemia since the mutant protein is unable to interact with LPL (Beigneux et al., 2009a; Franssen et al., 2010; Olivecrona et al., 2010).

Rare mutations in *APOA5* and *APOC2* can cause severe hypertriglyceridemia due to a reduction in LPL activity (Calandra et al., 2006; Fojo and Brewer, 1992). Furthermore, in a recent GWAS involving 463 hypertriglyceridemic cases (mean triglyceride=14.3mmol/l) and 1,197 controls (triglyceride<2.3mmol/l), a common variant (rs964184) residing at the *APOA1/C3/C4/A5* locus was significantly associated with risk of hypertriglyceridemia (OR=3.28, P=5.4x10^{-24}) (Johansen et al., 2010).

Of the ~100 human *LPL* mutations described, the majority are missense mutations that affect the essential 'catalytic triad' (serine 132, aspartate 156 and histidine 241) of the enzyme (Emmerich et al., 1992; Wang and Eckel, 2009). Most *LPL* mutations are rare, often being family specific. The G188E substitution which abolishes LPL activity (Peterson et al., 2002) occurs slightly more often, with a frequency of 0.06% in the general population (Nordestgaard et al., 1997). In one meta-analysis, carriers

of the mutation were found to have plasma triglyceride levels increased by approximately 80% as well as an estimated 5-fold higher risk of CHD (Wittrup et al., 1999).

Rare *LPL* sequence variants in the heterozygous state may also increase the risk of developing hypertriglyceridemia. In a recent study, 6.4% of 110 severely hypertriglyceridemic subjects (fasting plasma levels >10 mmol/l) were carriers of rare *LPL* mutations compared to 0/472 controls, all of European descent (Wang et al., 2007a). Furthermore, common *LPL* variants are associated with plasma triglyceride levels (Table 1.1). The common *LPL* variant, S447X is present in 17-22% of Europeans (Fisher et al., 1997). The S447X mutation results in a premature stop codon, two amino acids before the normal end of the protein. This variant is associated with slightly decreased circulating triglyceride levels (Humphries et al., 1998; Jensen et al., 2009; Wittrup et al., 1999). It is also thought to have a cardioprotective effect. In the European Atherosclerosis Research Study, the frequency of the truncated allele was significantly lower in those cases whose fathers had suffered a premature myocardial infarction (before 55 years) than in control subjects (0.09 versus 0.12) (Humphries et al., 1998). One explanation for these effects could be an increase in LPL mass observed both *in vitro* and in plasma from 447X allele carriers leading to enhanced clearance of apoB48-containing lipoproteins from the circulation (Nierman et al., 2005; Zhang et al., 1996).

Angiopoietin-like proteins 3 and 4 (ANGPTL3/ANGPTL4) are additional factors involved in regulating lipolysis of triglyceride-rich lipoproteins, through their inhibitory effect on lipoprotein lipase activity. Human *ANGPTL4* is expressed at high levels in the liver and adipose tissue (Kersten et al., 2000) whereas *ANGPTL3* is found primarily in the liver (Conklin et al., 1999). Overexpression of either protein in mice leads to elevated plasma triglyceride levels (Koishi et al., 2002; Lichtenstein et al., 2007). In contrast, knockout mice are hypotriglyceridemic (Koster et al., 2005). ANGPTL4 is thought increase triglyceride levels by promoting the conversion of active LPL dimers to inactive LPL monomers (Sukonina et al., 2006) whilst ANGPTL3 is thought to inactivate LPL by promoting its cleavage by proprotein convertases such as furin (Liu et al., 2010).

A population-based resequencing study examining the coding regions and intron-exon boundaries of *ANGPTL3* and *ANGPTL4* in 3551 participants in the Dallas Heart Study identified that rare non-synonymous variants in both genes were more

prevalent in those individuals with triglyceride levels found in the lowest quartile of the study distribution, compared to those with levels in the highest quartile (Romeo et al., 2007; Romeo et al., 2009). More recently, whole-exome resequencing of two siblings with FHBL but without a causative mutation of *APOB*, determined that they were both compound heterozygotes for S17X and E129X mutations in *ANGPTL3* (Musunuru et al., 2010).

In addition, common variants residing in or near to *ANGPTL3* and *ANGPTL4* were associated with lipid-traits in the recent GLGC meta-analysis. Specifically, rs2131925, found upstream of *ANGPTL3*, was associated with decreased plasma triglyceride levels (effect size=−0.06mmol/l per copy of G allele, P=9x10^{-43}) whilst the *ANGPTL4* variant rs7255436 was associated with lower plasma HDL levels (effect size= −0.01mmol/l per copy of C allele, P=3x10^{-8}) (Teslovich et al., 2010).

Thus, in summary, both common and rare genetic variants in the effectors of triglyceride-rich lipoprotein catabolism contribute to variance in plasma triglyceride levels. Both genetic and allelic heterogeneity contribute to the complexity of hypertriglyceridemia.

1.2.7) ApoE-Mediated Clearance of Catabolised Lipoproteins

Lipolysis of VLDL and chylomicrons yields lipoprotein particles depleted in triglyceride but enriched in cholesterol and apoE that are highly atherogenic (Karpe et al., 1994). Under normal physiological conditions these are rapidly removed from the circulation in the liver, in a process proposed to involve lipoprotein-bound ligands, including apoE, interacting with receptors such as the LDL-receptor (LDLR) and LDL-receptor-related protein 1 (LRP1), following the initial sequestration of the lipoprotein particles in the space of Disse, mediated by heparan sulphate proteoglycans (HSPGs) (Mahley and Huang, 2007).

ApoE is a vital ligand in the process of remnant clearance. In transgenic mice overexpressing apoE, clearance of injected, labelled chylomicrons occurs five-fold quicker than in controls animals (Shimano et al., 1994). Rare dominant mutations of *APOE* can cause Fredrickson type III hyperlipoproteinemia, defined by the accumulation of remnant lipoproteins in plasma. For example, the R142C mutation prevents normal apoE-HSPG binding (Mahley et al., 1999) and thus leads to increased plasma triglyceride and cholesterol levels by preventing efficient remnant catabolism (Vezeridis et al., 2011). Moreover, homozygosity for the apoE ε2 isoform

is a susceptibility factor for remnant accumulation which requires a 'second hit' from a further genetic or environmental factor for phenotypic presentation (Groenewegen et al., 1994; Schaefer et al., 2004).

In the GLGC meta-analysis, common variants residing downstream of *APOE* were associated with plasma triglyceride and cholesterol levels. Specifically, rs439401 was associated with decreased triglyceride levels (-0.06mmol/l per copy of T allele, $P=1x10^{-30}$), whilst rs4420638 was associated with increased plasma LDL- and total cholesterol levels (LDL: +0.18mmol/l per copy of G allele, $P=9x10^{-147}$, total cholesterol: +0.18mmol/l, $P=5.20x10^{-111}$) (Teslovich et al., 2010). However, since these variants also reside close to *APOC1*, caution needs to be applied in interpreting the functional relevance of these associations.

HSPGs were first implicated in remnant clearance in studies using heparinise treatment to disrupt cellular HSPGs. In three different cell types (LDLR-deficient fibroblasts, normal fibroblasts and HepG2 cells), heparinise treatment reduced apoE-mediated remnant binding by 80, 80 and 95% respectively (Ji et al., 1993). Uptake of labelled apoE-enriched lipoprotein was also reduced in treated HepG2 cells but heparinise did not affect binding of these particles to LRP (Ji et al., 1993). These results led the authors to propose a model in which remnants are initially sequestered through binding to HSPGs via APOE and are subsequently passed to LRP for internalisation (Ji et al., 1993).

Williams and colleagues mooted the HSPG, syndecan 1 as a potential candidate in remnant lipoprotein clearance due to its high expression level at the sinusoidal surface of hepatic parenchymal cells (Hayashi et al., 1987; Williams, 2008). Inhibition of syndecan using an antisense oligonucleotide decreased chylomicron binding to HepG2 cells by around 60% (Zeng et al., 1998). Recent studies using syndecan 1-null mice show that these animals develop plasma hypertriglyceridemia, attributable to delayed clearance of VLDL/chylomicron remnants (Stanford et al., 2009). Currently, no rare mutations or common variants of *SDC1* are reported to contribute to the variance of human plasma cholesterol or triglyceride levels.

1.2.8) LDLR-Mediated Clearance of ApoB-Containing Lipoproteins

A failure in the LDLR-mediated system of LDL-cholesterol catabolism is the basic defect that underlies the known forms of monogenic hypercholesterolemia.

Currently, the disorder is divided into four subtypes (Figure 1.4). Autosomal dominant hypercholesterolemia (ADH)-1, the most common form, resulting from mutations of *LDLR* (Leigh et al., 2008) can be further classified according to the mutational effect upon the LDLR protein ((Calandra et al., 2011; Hobbs et al., 1990; Ranheim et al., 2006), Figure 1.4). Thus monogenic hypercholesterolemia exhibits both genetic and allelic heterogeneity (Bertolini et al., 1999; Medeiros et al., 2010; Vaca et al., 2011). The genetic heterogeneity of monogenic hypercholesterolemia is, at least in part, attributable to the complexity of the pathways that regulate LDLR-mediated clearance of LDL-cholesterol.

LDLR is 843 amino acids in length and has a structure comprising five distinct domains: an N-terminal ligand binding domain, a domain with homology to the epidermal growth factor (EGF) precursor, an O-linked sugar domain, a transmembrane domain and a cytoplasmic domain (Yamamoto et al., 1984). In the ER, a 120kDa LDLR precursor is produced. This undergoes maturation through the ER and Golgi apparatus with removal of the signal sequence and addition of O-linked sugars, forming a mature protein of 160kDa which is then transported to the basolateral cell surface (Tolleshaug et al., 1982).

At the cell surface, LDLR binds to LDL through interaction of the N-terminal LDLR ligand-binding domain with apoB100 on the lipoprotein particle. The LDLR ligand-binding domain consists of seven LDLR type A repeats, each consisting of 40-amino acids and containing six conserved cysteine residues which form disulphide bonds. Repeats 3 to 7 are important for apoB100 binding (Russell et al., 1989). The exact region(s) of apoB-100 important for this interaction are not fully determined. Cross-species conservation suggests that 2 clusters of basic amino acids (amino acids 3147-3157 and 3359-3367) could be important (Law and Scott, 1990). The second of these clusters bears homology to the established ApoE-LDLR binding site (Knott et al., 1985). In addition, by far the most common ADH-2 mutation in Europe, R3500Q, (found in the heterozygous state in 2-5% of hypercholesterolemic patients) (Soutar and Naoumova, 2007) causes defective receptor binding in transgenic mice (Boren et al., 1998)

LDLRAP1 (ARH) interacts with both the C-terminal FDNPVY internalisation motif of LDLR (via its phosphotyrosine binding domain) and the clathrin-coated pit machinery, to mediate endocytosis of receptor-bound LDLR complexes (Chen et al., 1990; Garuti et al., 2005; He et al., 2002). After internalisation, the cargo is delivered to the

endosomes where dissociation of receptor:ligand complexes occurs due to the acidic nature of the endosomal lumen. The EGF-precursor homologous domain of the LDL receptor is important in this process. It consists of 3 EGF repeats (A, B and C) and a β-propeller structure, found in between repeats B and C. At pH5.3, the β-propeller acts as an alternative ligand for LDLR type A repeats 4 and 5 in the ligand binding domain, thus displacing the bound lipoprotein (Rudenko et al., 2002). In addition, the low concentration of free calcium within the endosomal lumen can also drive lipoprotein release from LDLR by inducing structural changes to the LDLR-ligand binding domain (Zhao and Michaely, 2009). After dissociation, the released lipoprotein proceeds through the endosomal system to the lysosome for degradation liberating its component cholesterol, fatty acids and amino acids. Meanwhile, LDLR is recycled and returned to the cell surface.

Gain of function *PCSK9* mutations lead to ADH-3 and overexpression of PCSK9 in mouse liver leads to reduced levels of LDLR (Maxwell and Breslow, 2004; Park et al., 2004). At the cell surface, PCSK9 interacts with LDLR via the EGF domain A repeat and directs the receptor towards the lysosomes for degradation, in a process that requires LDLRAP-clathrin-mediated endocytosis (Lagace et al., 2006). PCSK9 may target some LDLR for degradation before it reaches the cell surface, since when intracellular Golgi to lysosomal trafficking is blocked there is a rapid increase in LDLR levels (Poirier et al., 2009). The ADH-3 mutant, D374Y increases the avidity of PCSK9-LDLR binding, even at the acidic pH of the endosomes, enhancing the rate of LDLR degradation and thus decreasing receptor recycling (Cunningham et al., 2007; Lagace et al., 2006). Loss-of-function mutations in *PCSK9* are found in some hypocholesterolemic individuals (Cohen et al., 2005; Fasano et al., 2007; Kotowski et al., 2006). Low LDL-cholesterol levels are also observed in PCSK9-null mice (Rashid et al., 2005) where decreased receptor degradation concurrent with increased receptor recycling presumably allows enhanced LDL-cholesterol clearance.

As well as the rare mutations that underlie monogenic hypercholesterolemia, common variants in *LDLR*, *APOB*, *PCSK9* and *LDLRAP1* may also modulate circulating LDL- and total cholesterol levels (Teslovich et al., 2010). For example, Linsel-Nitschke and colleagues demonstrated that the *LDLR* SNP rs2228671 was associated with LDL-cholesterol levels (-0.19mmol/l (95%CI: 0.13–0.24) per copy of the rare T allele, $P=1.5\times10^{-12}$) in a combination of German and British cohorts (Linsel-Nitschke et al., 2008). This finding is substantiated by a number of further

studies (Aulchenko et al., 2009; Sandhu et al., 2008; Teslovich et al., 2010) although the functional relevance of this marker (or of a further variant in LD) has yet to be elucidated.

A second cluster of common *LDLR* variants also demonstrate association with LDL-cholesterol levels. In contrast to that described above, the association between rs688 and LDL-cholesterol was discovered before the advent of GWAS. In a study of 719 Alberta Hutterites, Boright et al., showed a positive association between the rs688 (*LDLR* nucleotide 1773) T allele and LDL-cholesterol, with each copy of the allele conferring a 0.10mmol/l increase in LDL levels (Boright et al., 1998). Association of the rs688 T allele (and the correlated rs1529729 C allele) with increased LDL-cholesterol levels is supported by more recent GWAS data. In a Swedish population study, modest association was detected between the rs1529729 C allele and LDL-cholesterol level, with a mean level in minor allele homozygotes of 4.24±0.98mmol/l versus 4.13±0.98mmol/l in major allele homozygotes, P=0.003 (Kathiresan et al., 2008). A similar modest effect was detected in Mexican Americans (Keebler et al., 2009) and in the BRIGHT cohort of British hypertensives where each copy of the rs688 T allele increased LDL-C levels by 1.03mmol/l (P=9.25x10^{-5}) (Wallace et al., 2008). In the recent GLGC study, rs1529729 was also associated with total cholesterol levels (P =1.04x10^{-20}) (Teslovich et al., 2010).

Functional evidence supports the potential role of rs688 in modulating circulating cholesterol levels. The presence of the T allele at this position is predicted *in silico* to neutralise an *LDLR* exon 12 splicing enhancer (Zhu et al., 2007b). *In vivo*, the T allele significantly decreased splicing efficiency of *LDLR* in tissue samples from female liver (P=0.024) and *in vitro*, HepG2 cells transfected with *LDLR* exon 9-14 minigenes performed ~12% less efficient splicing when the rs688 T allele was present (Zhu et al., 2007b).

In summary, both rare variants of moderate to large effect and more common variants with smaller effect sizes contribute to the population variance in circulating cholesterol levels due to their effect upon the key players in the LDLR-mediated clearance of LDL-cholesterol

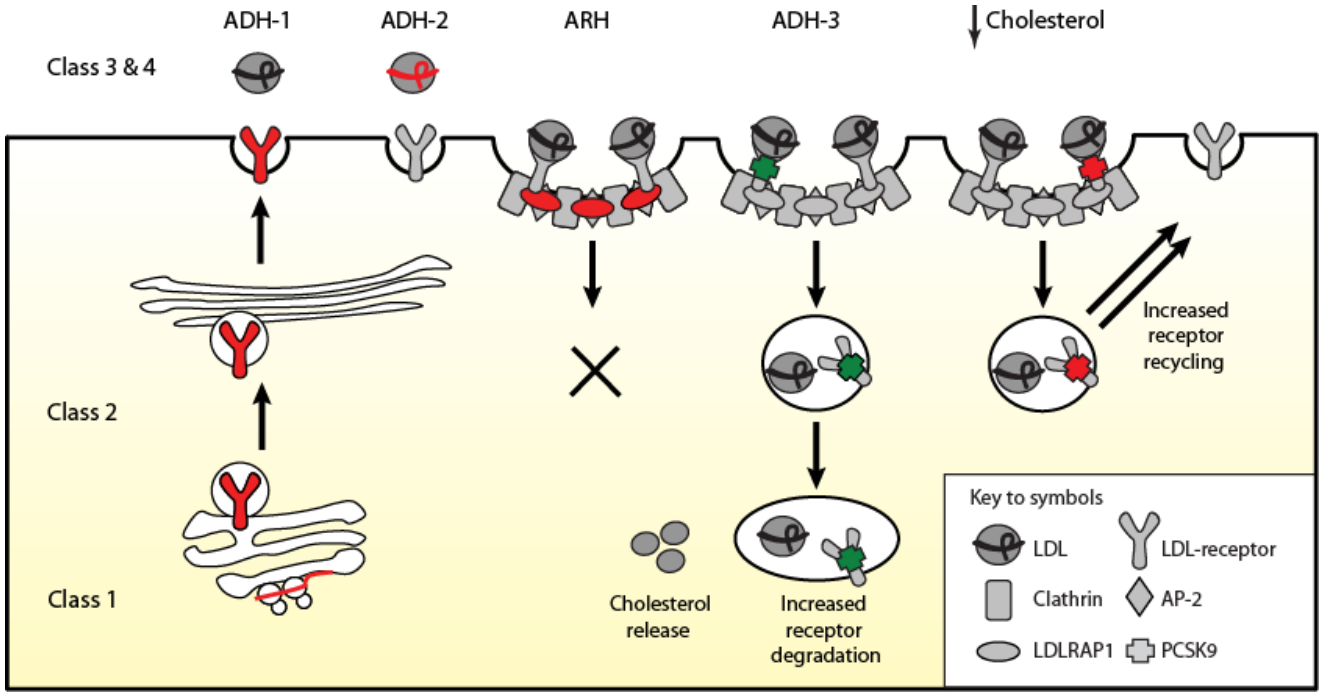


Figure 1.4) Overview of LDL Uptake and Monogenic Hypercholesterolemia

The five major classes of LDLR mutation that cause autosomal dominant hypercholesterolemia (ADH)-1 prevent: (1) production of immunologically detectable protein; (2) ER exit of encoded gene-product: complete (a) or partial (b); (3) binding of apoB100 (a) and apoE (b) ligands; (4) constitutive endocytosis, including of LDLR-apoB100 (a) and of VLDL-apoE (b); and (5) release of internalised LDLR-ligand (not shown for clarity). ADH-2 is caused by *APOB* mutations, ADH-3 from mutations of *PCSK9* (Abifadel et al., 2003) and autosomal recessive hypercholesterolemia (ARH) arises due to mutations of *LDLRAP1* (Garcia et al., 2001). Red=loss-of-function mutation, green=gain-of-function mutation. (Reproduced with permission from (Calandra et al., 2011))

1.2.9) Cholesterol Egress from the Late Endosomes/Lysosomes

Recent GWAS data suggest that common genetic variants may affect the fate of endocytosed cholesterol and in turn influence circulating cholesterol levels Specifically, a common *MLN64*/*STARD3* variant (rs11869286) was associated with plasma HDL-choelsterol levels in the recent GLGC meta-analysis (Teslovich et al., 2010).

After endocytosis, LDL-cholesterol passes to the late endosomes and lysosomes where the enzyme acid lipase hydrolyses the LDL-derived cholesterol esters to yield unesterified cholesterol (Sugii et al., 2003) which is then released to other cell compartments such as the ER, Golgi, mitochondria or plasma membrane.

Transport of cholesterol from the late/endosomes to the mitochondria is thought to involve MLN64. MLN64 is localised to the late endosomes and is able to bind cholesterol via a domain homologous to the cholesterol-binding domain of steroidogenic acute regulatory protein (which itself facilitates cholesterol import to the inner mitochondrial membrane) (Alpy et al., 2001; Miller, 2007). In normal, steroid-producing CHO cells, siRNA-mediated knock-down of MLN64, led to a ~40% decrease in production of the steroid hormone pregnenolone (a measure of cholesterol supply to the inner mitochondrial membrane). This was attributed to decreased intracellular transport of LDL-derived cholesterol (Charman et al., 2010).

Rare mutations of large effect can also influence cholesterol egress from the late endosomes/lysosomes. NPC1 and NPC2, mutations of which causes the fatal, autosomal recessive disorder, Niemann Pick Type C disease (Carstea et al., 1997; Naureckiene et al., 2000), are crucial for the cellular redistribution of unesterified cholesterol. For example, on incubation with LDL, NPC1- deficient fibroblasts accumulate unesterified cholesterol in the late endosomal/lysosomal cell compartment (Pentchev et al., 1987).

1.2.10) *De novo* Bile Acid Synthesis

Bile acids can be formed by two multi-step routes: the classic or alternative pathways. The first and rate-limiting step in the classic pathway is the hydroxylation of cholesterol to 7α-hydroxy-cholesterol catalysed by the liver-specific (cholesterol 7α-hydroxylase (CYP7A1) (Wu et al., 1999). The bile acid pool in *CYP7A1*- deficient mice is reduced to 25% of that in wild-type mice. These mice also exhibit fat

malabsorption and a concomitant increase in endogenous cholesterol synthesis by way of maintaining cholesterol homeostasis (Schwarz et al., 1996; Schwarz et al., 1998).

In humans, a *CYP7A1* frameshift mutation was identified in 9 members of one family (3 homozygotes, 6 heterozygotes) with statin-resistant hypercholesterolemia (Pullinger et al., 2002). Homozygotes had average LDL-cholesterol levels above 4.65mmol/l, whereas heterozygotes had levels intermediate to those of unaffected family members and the homozygous mutants. In contrast to the mutant mouse phenotype, an effect on cholesterol synthesis was not observed, since HMGCR activity was not altered in CYP7A1-deficient patients. Instead, the authors attribute the hypercholesterolemic phenotype to downregulation of LDLR expression brought about by hepatic accumulation of cholesterol (Pullinger et al., 2002). Recently, the common variant rs2081687, found downstream of *CYP7A1*, was associated with plasma total- and LDL-cholesterol levels in the large genome-wide association analyses of Teslovich and colleagues (Teslovich et al., 2010).

Bile acid secretion into bile is mediated by the Bile Salt Export Pump (BSEP, *ABCB11,* Figure 1.1). Mutation of BSEP causes progressive familial intrahepatic cholestasis type 2, characterised by low bile acid concentration in the bile (Strautnieks et al., 1998). In the intestine, bile acids are reabsorbed into the enterocytes by the apical sodium-dependent bile acid transporter (ASBT, *SLC10A2*) in the brush border membrane. Mutations of this gene cause primary bile acid malabsorption which is characterised by steatorrhea and reduced plasma cholesterol levels due to the disruption of enterohepatic circulation of bile acids (Oelkers et al., 1997). However, in the recent GLGC study, no *ABCB11* or *SLC10A2* SNPs were associated with plasma cholesterol levels (Teslovich et al., 2010).

1.3) Familial Combined Hyperlipidemia

Familial Combined Hyperlipidemia (FCHL) or type IIb hyperlipoproteinemia, by Fredrickson classification, is a common, complex genetic disorder with and estimated incidence of 1-6% in Western society (Goldstein et al., 1973; Hopkins et al., 2003). It was first described in 1973 with reference to the mixed pattern of lipid abnormalities displayed in the families of probands who had survived premature myocardial infarction (Goldstein et al., 1973; Nikkila and Aro, 1973). Subsequent studies have confirmed that FCHL is a highly atherogenic disorder, present in ~10-20% of

individuals with premature CHD (Austin et al., 2000; Hopkins et al., 2003). Given the common nature of FCHL and its coincidence with premature CHD, gaining an understanding of the genetic defects that underlie this condition is of vital importance.

The FCHL phenotype is characterised by elevated circulating cholesterol and/or triglyceride levels often accompanied by increased apoB and decreased HDL-cholesterol levels. Indeed, FCHL can be considered as a disorder of the overproduction of triglyceride-rich lipoproteins coupled with the impaired clearance of apoB-containing lipoproteins (Cabezas et al., 1993; Chait et al., 1980; Venkatesan et al., 1993). Probands have both raised cholesterol and triglyceride levels whilst the degree of dyslipidemia in other family members will vary and as such they may have raised plasma levels of triglyceride and/or cholesterol or may have normal lipid profile.

FCHL was originally thought to be an autosomal dominant disorder, with reduced penetrance until the age of 20 (Goldstein et al., 1973). More recently, genome-wide linkage analyses have suggested that FCHL is genetically complex (Aouizerat et al., 1999; Naoumova et al., 2003; Pajukanta et al., 1999). Despite recent expansion in our knowledge and understanding of lipid metabolism (described in the sections above), the causative genes/variants underlying the FCHL phenotype have yet to be firmly established. This is partly because linkage studies performed to date have included relatively small cohorts of FCHL families (maximum of 45 in Naoumova et al., 2003) and therefore are likely to have limited power to identify susceptibility loci.

Study of FCHL tends to focus upon the inheritance of its component traits, particularly circulating cholesterol and triglyceride levels, in affected families. Published linkage studies indeed imply that different susceptibility loci will underlie the elevated cholesterol and triglyceride levels that form the FCHL phenotype. For example, the FCHL triglyceride trait is linked to chromosome 1q21-23 (Pajukanta et al., 1998). Additionally, variation within the *APOA1/C3/A4/A5* complex is linked to the FCHL triglyceride trait (Eichenbaum-Voline et al., 2004) whilst a quantitative trait locus (QTL) on chromosome 8p23.3-22 influences cholesterol level (Naoumova et al., 2003; Yu et al., 2005). Other FCHL susceptibility loci on chromosomes 6q, 11p, 16q and 21q are suggested (Aouizerat et al., 1999; Naoumova et al., 2003; Pajukanta et al., 1999). Evidence suggests that variants in *USF1* on chromosome 1q (Lee et al., 2007; Pajukanta et al., 2004) and *WWOX* on chromosome 16q23 (Lee et al., 2008a) may explain the linkage signals detected in these chromosomal regions

however, functional causative variants remain to be identified. In the recent GLGC meta-analysis, variants in neither gene were associated with plasma lipid levels (Teslovich et al., 2010).

The aims of the studies presented in this thesis were to combine genetic and biological information to identify FCHL susceptibility loci/genes/variants in a large cohort of up to 239 white-British FCHL families.

Chapter 2 – Materials and Methods

2.1) FCHL Family Cohort Characterisation

White-British FCHL families, designated the FCHL cohort in this text, were recruited to the study by Dr. Naoumova on the basis of a proband who had: 1) fasting plasma cholesterol levels ≥95th percentile level (age- and sex-matched); 2) triglyceride levels ≥90th percentile level (age- and sex-matched) and; 3) at least one first degree relative with fasting plasma cholesterol levels ≥95th percentile level (age- and sex-matched) or triglyceride levels ≥90th percentile level (age- and sex-matched) (Naoumova et al., 2003). Percentiles from the Lipid Research Clinic (Heiss et al., 1980) were used as reference points (Appendix I). Exclusion criteria for participants were age <16 years, other genetic forms of hyperlipidemia (e.g. familial hypercholesterolemia, based upon molecular diagnosis, standard clinical signs, or diagnostic criteria) or secondary hyperlipidemia due to obesity (body mass index, BMI>30kg/m^2), type 1 or 2 diabetes (fasting glucose >7mmol/l), abnormal liver or thyroid function, excessive alcohol intake (>40 units per week), or certain medications known to interfere with lipid metabolism (such as thiazides, antiretroviral agents or beta-blockers). The research ethics committees of participating centres approved the study design. Participants gave written informed consent.

For linkage studies, a cohort comprising 182 FCHL families was utilised. Eight of these families (pedigrees 2, 4, 5, 6, 7, 10, 13 and 17) were previously included in the linkage analyses of Naoumova and colleagues (Naoumova et al., 2003). Subsequent association analyses involved an expanded cohort of 241 FCHL families. Pedigree structures are illustrated on the enclosed disk.

All lipid levels were recorded in the fasting state and were determined by automated methods using commercial kits and interassay controls as described in (Naoumova et al., 2003). Phenotyping was based upon lipid levels measured pre-treatment: patients were not taking medications known to alter circulating lipid levels.

DNA was extracted from whole blood and quantified by Tepnel Life Sciences. Quantification was carried out using a NanoDrop spectrophotometer (Thermo Fisher Scientific) or using PicoGreen reagent (Invitrogen). DNA was diluted to the appropriate concentration using Tris-EDTA buffer (Sigma Aldrich).

2.2) Microsatellite Genotyping

As part of a genome-wide linkage screen, six chromosome 21 microsatellite markers, *D21S1437, D21S2052, D21S1440, D21S2055, D21S1411* and *D21S1446* were genotyped by Marshfield Mammalian Genotyping Service using specified primers (Table 2.1). Genotype data were supplied on a DVD which also contained compressed gel image files and accompanying marker 'out' files. In order to investigate apparent non-inheritance of markers, extensive checking of genotyping calls was performed by myself, using the accompanying Imager software program.

Additional chromosome 21 microsatellite markers *D21S1909, D21S1260*, *D21S1259* and *D21S1912* were selected for genotyping using marker position and heterozygosity data gathered from the Marshfield Mammalian Genotyping Service website (URL2). Oligonucleotide primers (Table 2.1) included a 5' 'pigtail' (GTTTCTT) sequence (Brownstein et al., 1996) and one of each primer pair was labelled with 5' FAM or HEX dye. All oligonucleotides were reverse-phase cartridge purified and were obtained from Eurogentec.

Amplification reactions were carried out in a 10µl volume, in a 96-well micro-titer plate. 15ng of DNA was included in each reaction, together with 0.2 units of Phusion Taq polymerase (Finnzymes), 1X reaction buffer (Finnzymes, 7.5mM $MgCl_2$, 3% dimethyl sulfoxide (DMSO)), 200µM of each deoxynucleotide triphosphate (dNTP, A, C, G and T, Bioline) and 0.3µM of forward and reverse oligonucleotides. The remaining reaction volume was made up to 10µl with nuclease-free water. For control purposes, a blank (no DNA) reaction was included for each primer pair. Polymerase chain reactions (PCRs) were carred out on a G-storm GS4 cycler and typical cycling conditions were:

	Temperature (°C)	Time	
Initial denaturation:	98	2 minutes	
Denaturation:	98	10 seconds	30 cycles
Annealing:	63 to 65	20 seconds	
Extension:	72	30 seconds	
Final extension:	72	6 minutes	

Product specificity was determined by fractionation on a standard 1% agarose (Sigma Aldrich) 50ml minigel, made up using 1XTris-borate-EDTA (TBE, diluted from

a 5X TBE stock made up as follows: 54g Tris Base, 27.5g Boric Acid, 20ml 0.5M EDTA (pH 8.0, adjusted using NaOH pellets) dissolved in deionised water, made up to 1 litre) and stained with ethidium bromide at a concentration of 0.5 µg/ml. Product size was determined using Hyperladder I (Bioline) which was run in a parallel lane to the PCR-products.

Genotyping of PCR-products was performed by combining 1µL of PCR-product (<400bp) was with 10µL molecular biology grade formamide (Sigma Aldrich) and 0.15µL ROX 400HD (Applied Biosystems) which was subsequently loaded on to an Applied Biosystems 3730xl DNA Analyzer by the Genomics Core Facility (Medical Research Council, Clinical Sciences Centre). GeneMapper 3.7 software (Applied Biosystems) was used to determine allele size for each sample. For quality purposes, duplicate DNA samples were included on each sample plate.

Microsatellite genotype data was transferred, as a text file, to a relational database, BC|GENE (Biocomputing Platforms, Espoo, Finland). Mendelian inheritance was confirmed using PedCheck (O'Connell and Weeks, 1998), Merlin Error (Abecasis et al., 2002) was used to highlight unlikely microsatellite genotypes and the Pedwipe function was used to removed those genotype calls that could not be resolved by re-examination of the primary data.

Table 2.1) Chromosome 21 Microsatellite Primers

Marker	Study[a]	Forward primer	Reverse Primer
D21S1437	G	ATGTACATGTGTCTGGGAAGG	TTCTCTACATATTTACTGCCAACA
D21S2052	G	GCACCCCTTTATACTTGGGTG	TAGTACTCTACCATCCATCTATCCC
D21S1909	F	GATCAGCTGATTGGCCAGC	CCTGGCAGCTGTGATTGTG
D21S1440	G	GAGTTTGAAAATAAAGTGTTCTGC	CCCCACCCCTTTTAGTTTTA
D21S2055	G	AACAGAACCAATAGGCTATCTATC	TACAGTAAATCACTTGGTAGGAGA
D21S1260	F	CCCAAGGCACTGTTCCTGC	CATCCAAGGGGTTCTTCATCC
D21S1411	G	AGGATGGATAAATAGAACATAGGTAGA	CCCACTCCCAGCCTTCTAA
D21S1259	F	GACCTTGTGCCTTGGTAAGTG	GGACAAATGAGAAGGGACTGTA
D21S1912	F	CAGAGCCCACCCTGGTAAC	CTTTGCCATCAGCCCTCATAC
D21S1446	G	ATGTACGATACGTAATACTTGACAA	GTCCCAAAGGACCTGCTC

G= genome-wide screen markers, F= fine-mapping markers

2.3) Linkage Analysis

Non-parametric multipoint linkage analysis was carried out using Allegro (Gudbjartsson et al., 2000). The FCHL and component lipid traits were dichotomised with affected status being defined as serum cholesterol level ≥95th age- and sex-specific percentile level for the cholesterol trait, serum triglyceride level ≥90th age- and sex-specific percentile level for the triglyceride trait or both of these thresholds for combined hyperlipidemia trait. Input files in 'Linkage' format were exported from BC|GENE and were viewed in Microsoft Excel. Analysis settings were listed in an accompanying options file and specified multipoint linkage analysis, using both the linear and exponential model (Kong and Cox, 1997), applying equal weighting to all families and the determination of exact *P*-values to accompany the calculated NPL scores.

For quantitative trait analysis, raw triglyceride levels were first log transformed (using logarithm to the base e) in order to normalise their distribution. Both cholesterol and log triglyceride levels were adjusted, by regression, for the effects of age, sex, BMI and their interactions; affected and unaffected individuals were normalised separately. Quantitative trait linkage analyses were carried out using Merlin Regress (Abecasis et al., 2002). Input files were exported from BC|GENE. The analysis model included trait means and variances and the estimated heritability of each trait. Heritability was estimated using the variance components linkage analysis (--vc) option of Merlin.

Files for Ordered subset analysis (OSA) were prepared by myself using Microsoft. Excel. OSA (version 2.1 (Hauser et al., 2004)) was implemented by Stuart Horswell and Guiyuan Lei. Specifically, at each "marker of interest" (a genotyped microsatellite itself or one of 4 interpolated points between genotyped markers) across chromosome 21, families were ranked according to their family specific LOD scores at an index marker (any other genotyped microsatellite marker, genome-wide). Adding families consecutively, in order of rank, LOD scores were recalculated after each addition until the subset producing the maximum LOD score at the index marker was identified. The "optimal slice" option of OSA was invoked so that the subset could be drawn from families clustered at any point in the ranking, not just at the high or low extremes. Baseline and OSA LOD scores were calculated using KACLOD (Kong and Cox, 1997), under a dominant model. The significance of the increase in LOD from baseline to OSA was assessed by a fitting skewed-normal distribution. OSA data were subsequently viewed using Microsoft Excel.

2.4) Candidate Gene Annotation

Positional information for candidate genes was determined using the Ensembl genome browser (URL3). In addition to literature searches using PubMed (URL4), expression data based on expressed sequence tags (ESTs), available in UniGene (URL5) were considered.

2.5) SNP Selection and Assay Design

Non-synonymous coding SNPs were identified using Ensembl (URL3) and dbSNP (URL6) and were put forward for genotyping regardless of their minor allele frequency. Linkage disequilibrium data, derived from Phases 1 and 2 of the International HapMap Project (URL7) and specific to Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH) collection (CEU), were used to identify tag-SNPs for each gene.

AGPAT3, *CBS*, *LSS* and *UBE2G2* tag-SNPs were identified by Illumina after definition of target intervals. This SNP selection methodology employed a modified version of the ldSelect algorithm (Carlson et al., 2004), whereby the likely success of the SNP genotyping assay of a given marker using the Illumina GoldenGate genotyping platform was also incorporated into the selection process. ldSelect functions by grouping SNPs into 'bins'. Within each bin is at least one SNP whose genotype is highly correlated (r^2>0.8) with that of all other SNPs in the same bin.

SNPs residing in all other target intervals were genotyped using Sequenom MassARRAY iPLEX technology (Jurinke et al., 2002). CEU data was downloaded from the International HapMap Project (Phases 1 and 2, release 24, URL7) and uploaded into Haploview (Barrett et al., 2005). Tag-SNP selection was performed using the integrated function Tagger (de Bakker et al., 2005), using the pairwise tagging method and a minimum correlation between markers of r^2=0.8. Markers with a minor allele frequency below 1%, a genotype call rate of below 75% or a violation of Hardy Weinberg Equilibrium (*P*<0.001) in the reference dataset were excluded from the selection process.

Sequenom multiplex genotyping assays were designed with the MassARRAY assay design software using ~300bp of flanking DNA sequence from both sides of each sequence variant. Template DNA sequences were retrieved from the Ensembl database (URL3) and masked for other sequence variants. Additionally,

RepeatMasker (URL8) was used to identify high copy number repeats in the template sequences.

2.6) SNP Genotyping and Data Handling

DNA samples were supplied to Illumina in 96-well plate format for genotyping. Data were supplied on a DVD and genotype calls were made using the Illumina Bead Studio software which also allowed visualisation of cluster plots.

For Sequenom genotyping, DNA samples were transferred to a 384-well format using a Biomek Robot (Beckman Coulter). Duplicate samples and no-DNA wells were included on each DNA plate. Reverse-phase cartridge purified oligonucleotides were supplied lyophilised from Eurogentec and were subsequently rehydrated using nuclease-free water. Multiplex PCRs were carried out in 384-well plates in a volume of 5μL, which contained 2μL of template DNA (4ng/μL), 0.5U HotStar Taq polymerase (Qiagen), 1.625mM $MgCl_2$, 1.25X PCR buffer (both supplied with Taq), 500μM of each dNTP (Bioline) and a primer mix containing 100nM of each primer pair. Touchdown PCR was carried out using a GS4 G-storm thermal cycler or MJ Research tetrad, using the following cycling conditions:

	Temperature (°C)	Time	
Initial denaturation:	95	15 minutes	
Denaturation:	95	20 seconds	
Annealing:	65	30 seconds	4 cycles
Extension:	72	1 minute	
Denaturation:	95	20 seconds	
Annealing:	58	30 seconds	4 cycles
Extension:	72	1 minute	
Denaturation:	95	20 seconds	
Annealing:	53	30 seconds	38 cycles
Extension:	72	1 minute	
Final extension	72	3 minutes	

Post-PCR steps were initially carried out by Marlene Attard or Petros Takousis at Imperial College Sequenom Genotyping Service and were subsequently completed by myself, using all reagents supplied in the Sequenom iPLEX reaction kit (Sequenom). Unincorporated dNTPs were removed by 40 minutes incubation at

37°C with 0.51U Shrimp Alkaline Phosphatase (in a total volume of 7μL) followed by 5 minute heat inactivation of the enzyme at 85°C. For the extension reaction, an 'iPLEX Gold cocktail' and extension primers were added to the plate. Extension primer concentrations were determined according to their mass.

Reactions were carried out using a MJ Research tetrad under the following conditions:

	Temperature (°C)	Time (seconds)	
Initial denaturation:	94	30	
Denaturation:	94	5	40 cycles
Annealing:	52	5	40 cycles
			5 cycles
Extension:	80	5	40 cycles
Final extension:	72	180	

Products of the extension reaction were desalted by the addition of 'clean resin' and spotted onto a SpectroChip (both Sequenom), using the Sequenom MassARRAY nanodispenser. The chip was then transferred to a Bruker Daltonics Autoflex instrument for analysis by Matrix-Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) mass spectrometry. SNP genotypes were determined in MassARRAY Typer software (Sequenom) using the default clustering algorithm. The defined clusters were then checked manually to ensure correct calling of SNP genotypes.

Data obtained by Illumina and Sequenom genotyping methods were merged and transferred to BC|GENE. Mendelian inheritance patterns were checked using PedCheck (O'Connell and Weeks, 1998) and adherence to Hardy-Weinberg Equilibrium of each SNP was verified in PLINK using the pedigree founders (Purcell et al., 2007). Data from the 1000 Genomes project (URL9) was used as a reference for quality checks.

2.7) Genotype Imputation (*in silico* genotyping)

Imputation of untyped markers in the chromosome 21 interval from 41.4 to 47.7Mb was performed using MaCH (Li et al., 2010) via the BC|GENE interface. Genotyped markers with a genotyping success rate below 90% and a minor allele frequency

below 0.01 were excluded from the imputation process. These criteria match those employed in a meta-analysis of lipid traits involving the Finland-United States Investigation of NIDDM Genetics (FUSION) cohort (Willer et al., 2008).

All genotyped markers were manually aligned to the same DNA strand as the reference the data, with the underlying assumption that the rare and common alleles for A/T and G/C SNPs were the same in reference and FCHL cohort dataset. Reference datasets (Table 2.2) were pre-loaded into BC|GENE by the suppliers. Imputation model parameters were estimated using data for 239 probands, 244 spouses or both and involved 100 iterations. Imputation quality was estimated using the --mask option implemented in MaCH to mask 2% of the experimentally determined genotypes. As recommended the program authors, imputed SNPs with an r^2 value <0.5 were excluded from association analyses (URL10).

Table 2.2) Reference Datasets for Genotype Imputation

Reference Dataset	Individuals	SNPs in Interval
1000 Genomes June 2010	60 CEU	17875
HapMap Phase 2	90 CEU	5326

2.8) Multiple-Testing Correction and Power Calculations

The effective number of statistical tests was estimated using data for all family members in SNPSpDlite (URL11) under the model proposed by Li and Ji (Li and Ji, 2005; Nyholt, 2004). Input files were prepared using BC|Gene and Microsoft Excel.

Power calculations were performed in Quanto (URL12) for case-control and continuous trait outcomes, under the assumption of an additive inheritance model, an effect allele frequency of 10%, and an alpha threshold of 1.6×10^{-5}. For case-control scenarios, odds ratios ranging from 0.5 to 2.0 were considered. For continuous trait calculations, SNPs were assumed to account for 0.25-2.0% of the trait variance.

2.9) Association Analyses of Genotyped and Imputed SNPs

Fisher's exact test was used to compare allele frequencies between FCHL probands and spouses (using an R script written by Stuart Horswell). For quantitative trait association analysis, linear mixed effects models (LME) implemented in the R

package, Genome-Wide Association analyses with Family (GWAF) (Chen and Yang, 2010) were used. By this method, the R function 'lmekin' to is employed to account for within-family phenotypic correlation. Raw triglyceride levels were log-transformed. Age, sex and BMI were included in the model as covariates and an additive genetic model was specified.

Dichotomous trait association analysis was carried out using logistic regression via generalised estimating equations, in GWAF. Individuals were dichotomised for FCHL triglyceride and cholesterol trait affection status as previously described (2.2). For imputed markers, genotype dosages from the MaCH output files were used as input for the same quantitative and dichotomous analyses in GWAF.

Illustrations of SNP positions (4.2) and association results (5.2) were produced using LocusZoom (URL13) (Pruim et al., 2010) and modified in Adobe Illustrator. LD between SNPs was assessed using SNAP (URL14) (Johnson et al., 2008). Potential effects of variants were assessed using SIFT (Ng and Henikoff, 2001) and Polyphen (Sunyaev et al., 2001). Cross-species conservation was examined using VISTA (Frazer et al., 2004) in the UCSC Genome Browser (URL17).

2.10) Extraction of Global Lipid Genetics Consortium (GLGC) Data

GLGC meta-analysis data were downloaded (URL15, (Teslovich et al., 2010)) and parsed using a UNIX-based system. Specifically, data files for the total-cholesterol, LDL-cholesterol and triglyceride traits were downloaded to a Windows computer before transfer via FileZilla FTP client (URL16) to a UNIX machine. A file (snps.txt) listing all dbSNP (release 130) markers residing in the chromosome 21 interval that extends from 40Mb to 21qter was compiled using the UCSC table browser (URL17) and also transferred to the UNIX machine. The GLGC data were then parsed using the command line:

```
fgrep -w -f snps.txt TC_ONE_Europeans.tbl > TCoutput.txt
fgrep -w -f snps.txt TG_ONE_Europeans.tbl > TGoutput.txt
fgrep -w -f snps.txt LDL_ONE_Europeans.tbl > LDLoutput.txt
```

Output files were transferred back to a Windows machine using FileZilla FTP client.

2.11) Cell Culture and RNA Extraction

McArdle rat hepatoma (McA-RH) 7777 cells (strain CRL-1601) were obtained from ATCC Biomaterials (via LGC Standards) at passage 68. Cells were thawed,

according to suppliers instructions, and were then cultured in a T75 tissue culture flask, in high glucose (4500mg/litre) Dulbecco's Modified Eagle Medium (Sigma Aldrich), supplemented with L-Glutamine (2mM, Sigma Aldrich), 10% heat-inactivated fetal calf serum (BioSera) and 1X Penicillin/Streptomycin (Invitrogen) (in sum, complete culture medium) until sufficient numbers of cells had accumulated.

At around 80% confluency, cells were transferred to 12-well plates at a density of $2.5x10^5$ per well. In order to detach cells from the original culture vessel, medium was aspirated and cells were washed in phosphate buffered saline (PBS) to remove all traces of culture medium. Cells were detached from the flask surface by the addition of 3ml TrypLE Express (Invitrogen) and a short (<5 minutes) incubation at 37°C. This suspension was neutralised by the addition of 7ml of complete growth medium. Cell numbers were determined using a haemocytometer and cell counts were performed in duplicate. After 24-hours in culture, medium was aspirated and replaced with 1ml complete medium containing either 100μM or 200 μM bezafibrate dissolved in dimethyl sulfoxide (DMSO, both Sigma Aldrich), or with DMSO alone (at a final concentration of 0.1%). Cells were exposed to treatment conditions for 6 or 24 hours prior to lysis and RNA extraction using the RNeasy Mini Plus Kit (Qiagen).

To lyse cells, growth medium was removed from cell culture plate and 350uL RLT buffer (Qiagen) was added to each well. Homogenisation was carried out by pipetting lysed cell suspension in to QIAshredder spin column (Qiagen) and centrifuging at 12000rpm for 1 minute. RNA extraction from the cell homogenate was completed according to the manufacturer's instructions and RNA was eluted into a volume of 40uL RNase free water (Qiagen). RNA was quantified by duplicate measurements made using a Nanodrop (Thermo Fisher Scientific). In order to verify RNA sample quality, absorption ratios were determined (A260/A280 to indicate DNA contamination and A260/A230 to indicate the presence of organic compounds). Only those samples with an A260/A280 ratio ~2 and an A260/A230 ratio of ~1.8-2.2 were utilised in subsequent quantification reactions.

2.12) Quantitative, Real-Time Reverse Transcription (RT-q) PCR

Rat homologs of target genes within the 21q22.3 refined FCHL linkage interval were identified using Qiagen Gene Globe database. Initial RT-qPCR assays were supplied by Qiagen, with the exception of oligonucleotides specific to the three *COL18A1* alternative transcripts (Table 2.3) which were designed using Primer-

BLAST (URL18) and supplied by Eurogentec. In addition, replacement primers for non-specific assays (Table 2.4) were redesigned using Primer-BLAST (URL18) and supplied by Eurogentec. Where possible, oligonucleotides spanning intron-exon boundaries were chosen. All oligonucleotides were supplied in the lyophilised state and were resuspended in 1.1ml Tris-EDTA (TE), pH 8 (Sigma Aldrich) and stored in 200µL aliquots.

All reactions were conducted using the Qiagen QuantiFast SYBR green RT-PCR kit, in 96-well non-skirted plates with compatible optical strip caps (both Stratagene). Reactions were performed in a total volume of 25µl, containing Master Mix (1X final concentration, containing HotStarTaq Plus DNA Polymerase, QuantiFast SYBR Green RT-PCR Buffer, SYBR Green I, and ROX passive reference dye), QuantiFast reverse transcriptase mix (excluded from no-reverse transcriptase control reactions), forward and reverse primers (1µM final concentration) and template RNA (10ng, unless constructing a standard curve). For quantification reactions, three biological (different RNA samples) and three technical replicate wells (repeat experiment with same RNA sample) for each treatment condition were set up, per target gene assay and three independent experiments were performed. No DNA and no reverse transcriptase control wells were also included for each sample/target gene assay.

All reactions were carried out using an Mx3000P real-time cycler (Stratagene) and used the following conditions:

	Temperature (°C)	Time	
Reverse transcription:	50	10 minutes	
PCR initiation (hot start):	95	5 minutes	
Denaturation:	95	10 seconds	40 cycles
Combined annealing/extension:	60	30 seconds	

At the end of the amplification cycles, melt curve analysis was performed by incrementally increasing the temperature from 55ºC to 95ºC. The melt curve was generated using MxPro software (Stratagene), plotting the first negative derivative of the change in fluorescence against temperature. Where the melt curve was indicative of a non-specific PCR reaction or of the formation of primer dimers (multiple peaks were present), RT-qPCR assays were redesigned (Table 2.4).

Table 2.3) Rat *COL18A1* Transcript Specific Primers

Transcript	Human mRNA[a]	Forward Primer	Reverse Primer	Product Size (bp)
Short	NM_130445	GGACATGCTCACCAGTTTGG	GTCTTCGACTTGGGAGATCTTC	135
Medium	NM_030582	CAGATTCTCAGACAACCGAGG	GTCTTCGACTTGGGAGATCTTC	210
Long	NM_130444	CAGGAGGATGGGTACTGTGTG	GTCTTCGACTTGGGAGATCTTC	116

[a] Human RefSeq accession for COL18A1 transcripts

Table 2.4) Redesigned Primers for Non-Specific RT-qPCR Assays

Gene	Forward Primer	Reverse Primer	Product Size (bp)
AIRE	GATGAGCAGAAAGTGAGCCAG	GTCACAGCAGATGAGCTCG	132
LRRC3	CTAGGCTCTTCCGCAGGCGAAG	GCGAGCGTCCTTCCTCACGTT	74
PDE9A	CATGGACCTGTTCTGTATCGC	GCTTGACAGCCACAGGTCTC	149
SLC37A1	CCTGTGGTCTGATGCTACTGC	GACGGCAGTGGTGATCAGAG	146
TFF2	CTTGGATGCTGCTTTGACTC	ACAGTCCTCAGGGCTGATGC	142
UBASH3A	CTCACACACTCCGTCAGTGTC	CTCACCATGGCGTATGACCAG	66
CRYAA	No alternative primer sites		
PWP2	No alternative primer sites		

bp=base pairs

2.13) Relative Quantification of mRNA Levels

Before comparing target gene mRNA levels in bezafibrate-treated (test samples) and DMSO vehicle-treated cells (calibrator samples), the amplification efficiency of each set of primers was estimated, by using five concentrations of control RNA (100ng, 10ng, 1ng, 100pg and 10pg) to generate a standard curve. Primer efficiency was calculated by determining the gradient of the line of best fit generated by plotting Ct against log RNA template concentration and using the following equation:

$$E = 10^{(-1/S)} \text{ (S is the slope of the standard curve)}$$

$$\% \text{ Efficiency} = (E - 1) \times 100\%$$

Using this formula, a slope of -3.32 is equivalent to E=2 or 100% efficiency. Construction of these standard curves determined that the use of 10ng of template RNA was optimal for subsequent qRT-PCR reactions and template detection.

Transcript levels in bezafibrate-treated (test) and DMSO vehicle-treated (calibrator) samples were normalised to levels of a reference gene, cyclophilin A (*PPIA)*. In order to select a suitable house-keeping gene as a reference, 18S ribosomal RNA (*18S*), cyclophilin A (*PPIA)* were tested using a dilution series. Extremely high levels of *18S* were detected (Ct<9.0) it was deemed to be unsuitable for use as a reference gene. In contrast, *PPIA* levels (14.87≤Ct≥26.82) were within the range of test genes. *PPIA* has previously been utilised in gene expression studies following bezafibrate treatment (Chinetti-Gbaguidi et al., 2005) and where RNA was extracted from McA-RH7777 cells (Bamji-Mirza et al., 2011).

To calculate the abundance of test gene transcripts in test versus calibrator samples, relative quantification using $2^{-\Delta\Delta Ct}$ method was performed (Livak and Schmittgen, 2001). First, the Ct of the target gene was normalised to that of the reference gene (*PPIA*) in both the test and calibrator samples:

$$\Delta Ct \text{ (test)} = Ct \text{ (target, test)} - Ct \text{ (reference, test)}$$

$$\Delta Ct \text{ (calibrator)} = Ct \text{ (target, calibrator)} - Ct \text{ (reference, calibrator)}$$

Next ΔCt of the test samples was normalised to that of the calibrator:

$$\Delta\Delta Ct = \Delta Ct \text{ (test)} - \Delta Ct \text{ (calibrator)}$$

Finally, the expression ratio is calculated as:

$$2^{-\Delta\Delta Ct} = \text{normalised expression}$$

A *t*-test was used to determine whether ΔΔCt was significantly different form zero. It should be noted that the $2^{-\Delta\Delta Ct}$ calculation is based upon the assumption that the target and reference genes have similar amplification efficiencies of ~100%.

2.14) Interrogation of External Expression Datasets

The data of Schadt and colleagues (Schadt et al., 2008) were accessed using the GTEx (Genotype-Tissue Expression) eQTL Browser (URL19) using relevant chromosome and SNP search criteria. Adipose tissue eQTL data from the study of Nica et al., were searched using Genevar (URL20) (Nica et al., 2011; Yang et al., 2010). SNP-gene expression plots were also generated in Genevar. Supplementary data from the microarray study of Plaisier et al., were transferred to BioMart (URL21) in order to link the Affymetrix probe names given to their corresponding genes (Plaisier et al., 2009). Differentially expressed chromosome 21 genes were filtered from the complete dataset using Microsoft Excel.

2.15) *GPIHBP1* Sequencing

The gene structure of *GPIHBP1* was determined by alignment of its RefSeq mRNA sequence (accession number NM_178172) with the corresponding genomic interval using the UCSC genome browser (URL17). 189 FCHL patients and 48 controls were selected for sequence analysis. Primers, designed to sequence the entire *GPIHBP1* coding region plus 871bp upstream of the translation start site, were obtained from Eurogentec (Table 2.5).

10μL amplification reactions were carried out using a G-storm GS4 cycler. Each reaction used 15ng of DNA and 0.2 units of Phusion Taq polymerase (Finnzymes) as previously described (2.2). Size and specificity of the PCR-product was verified as before (2.2). In order to comply with strict laboratory regulations, I performed only the post-PCR processes. PCR-products were purified by incubation at 37°C for 20 minutes with 4μL of an Exonuclease I/Shrimp Alkaline Phosphatase (SAP) mix (0.8U of each, GE Healthcare), followed by a 15 minute inactivation step at 80°C. Direct sequencing was carried out with 1μL of this mix, using Big Dye Terminator chemistry and 3730xl sequencer (both Applied Biosystems). Sequencing reactions were completed by the Medical Research Council Clinical Sciences Centre Genomics Core Facility. Sequences were aligned and chromatograms were examined using ChromasPro (Technelysium, Australia).

Table 2.5) *GPIHBP1* Sequencing Primers

Target	Forward primer sequence	Reverse primer sequence
Upstream sequence	AGAGCCTCGACTCTGCACTC	TTGTTTCCCTCCAGGACATC
Upstream to Intron 2	CAGCACCACCTTCCACCT	AGTGTCATGGTGAGCAGCAG
Intron 1	ACTACAATGCCAGGGCACAC	GCTCTGGGCAAGCATCGTG
Intron 2	CGAGGAAGATGAGGATGAGG	AGAAACCTGGAGGTGGGTCT
Intron 2	CTGTGAGGCCCTCCTCTG	AATGAGGGTTGTGCAGGTCT
Exon 3	CAGCACAGCTTACAGGACCA	GACATTGCACAGGCTGGAC
Intron 3 toExon 4	ACGGGAACACCGGTAAGTG	GCTGCAGATGGAGGAATTC
Exon 4	GTCCAGCCTGTGCAATGTC	GGCCCCATCCTCTTCATACT
Exon 4	AGCCAGGCAGAGATGATACC	CAGATCCATCCATGACACC
Exon 4	GCAGTACCTGCTGGGCTTAG	CCCAGCTAGGCATTGTGTGT
Exon 4	CCTAGATCCCCACCAGTTCA	GCATGTGCTATGTTCACACAT

2.16) *GPIHBP1* SNP Genotyping and Quality Checks

GPIHBP1 haplotypes were generated in Haploview (Barrett et al., 2005), using patient/control sequence data. Six haplotype tagging SNPs were selected using Tagger (de Bakker et al., 2005) as previously described (Table 2.6).

SNPs were genotyped using the Sequenom Mass ARRAY iPLEX gold technology (methods described in section 2.4). Genotyping assays (PCR and extension primers, Table 2.6) were designed using the MassARRAY assay design software and reverse-phase cartridge purified oligonucleotides were obtained from Eurogentec. Post-PCR laboratory processes were performed by Marlene Attard at Imperial College Sequenom Genotyping Service. SNP genotypes were determined, by myself, using MassARRAY Typer software.

Integrity of duplicate and blank samples was confirmed. PedCheck (O'Connell and Weeks, 1998) was used to confirm Mendelian inheritance of all SNPs. Minor allele frequencies and Hardy-Weinberg Equilibrium *P*-values were calculated using maximum likelihood estimation in SOLAR, taking relatedness into account (Almasy and Blangero, 1998).

Table 2.6) *GPIHBP1* SNP Genotyping Primers for the Sequenom MassARRAY System

Variant	First PCR primer[a]	Second PCR primer[a]	Extension primer
rs61747644	ACAGCAAGAGGGCAAGCAG	ACTCAGAGTCAGGGACACAG	GGCGAGATGAAGGCGCT
rs11538389	ACCGCAGGGTTACCCTTTG	ACTCAGAGTCAGGGACACAG	GCTTGCCCTCTTGCTGT
rs11538388	AGATGACTACGACGAGGAAG	ATACCTCTGCTCCTGCCAC	CCTGTTGGTCTCCTCCTCTTC
rs10746561	AGCATGCCCCACCACACACA	ACACCTCGTGGGTGTGTGG	GTGTGTGGGGCAGGGCTC
rs10866908	TCACAGAAACCTGGAGGTGG	ACGAGGTGTCCCCTGTGAG	GAGGCCCTCCTCTGCCA
c.295+52delA	GACCTGCACAACCCTCATTG	TGGCTCTGCAGGGCTCTGTT	GGCTCTGTTCCGGGCCCC

[a] All PCR primers prefixed with: ACGTTGGAT

2.17) *GPIHBP1* Genotyping in Additional Cohorts

White British combined hyperlipidemia patients were recruited as previously described (Naoumova et al., 2003). Additional patients with combined hyperlipidemia were drawn from the University of California, San Francisco Genomic Resource in Atherosclerosis (Pullinger et al., 1995). Exclusion criteria were as described in (Naoumova et al., 2003). All cases were of Caucasian descent and were recruited on the basis of fasting plasma cholesterol levels ≥95th percentile level (age- and sex-matched) and plasma triglyceride levels ≥90th percentile level (age- and sex-matched), using percentiles from the Lipid Research Clinic as a reference (Heiss et al., 1980). White British controls were recruited from general medical practices across west London. Control groups comprised healthy individuals, who were not taking any lipid-lowering medication.

Hypertriglyceridemic cases of White-European ancestry were selected from the University of California, San Francisco Genomic Resource in Atherosclerosis (Pullinger et al., 1995) and from Lipid Clinics in Leiden/Amsterdam in the Netherlands. Cases were chosen on the basis of fasting plasma triglyceride levels >2.26mmol/l (200 mg/dl). Study participants were not taking lipid-lowering medications at the time that plasma lipid measurements were recorded. Demographically matched controls were chosen on the basis of fasting plasma triglyceride <1.70mmol/l (150 mg/dl).

Participants of the Northwick Park Heart Study II (NPHSII) were recruited as previously described (Miller et al., 1995; Talmud et al., 2002). Briefly, the study comprised 3052 men, aged 50 to 61 years, recruited from nine general medical practices, for prospective surveillance. All men were free of CHD at the start of the study. CHD events recorded were fatal and non-fatal myocardial infarction, plus coronary artery surgery and silent myocardial infarction on the follow-up ECG. Genotyping was performed by Philippa Talmud at University College London.

Participants in the Southampton Atherosclerosis Study (SAS) were recruited from patients undergoing interventional or diagnostic coronary angiography in the Wessex Cardiothoracic Unit, Southampton General Hospital as previously described (Ye et al., 2003). All individuals were males of Caucasian descent and had >50% diameter stenosis in at least one major epicardial coronary artery. Genotyping was performed by Shu Ye at Queen Mary, University of London.

2.18) *GPIHBP1* Association Analysis

Association analysis of the *GPIHBP1* SNPs with cholesterol and triglyceride levels in the FCHL families was performed using the R package GWAF, under an additive model, including age, sex and BMI as covariates (Chen and Yang, 2010). Triglyceride levels were log-transformed in order to normalise the data.

The combined hyperlipidemia and hypertriglyceridemia case-control cohorts were analysed by logistic regression in PLINK (Purcell et al., 2007), using an additive model. Since the British and UCSF combined hyperlipidemia cohorts were ascertained using the same criteria, a joint analysis of these groups was performed. Similarly, the Dutch and UCSF hypertriglyceridemia cohorts were analysed jointly.

To perform combined analysis of both the FCHL families and the case-control cohorts, the Z-method, implemented in the program METAL (Willer et al., 2010), was used. In order to accommodate the unequal numbers of cases/controls in the individual groups, the effective sample size of each study (N_{eff}) was calculated using the formula: $N_{eff} = 4/(1/N_{cases}+1/N_{ctrls})$.

SAS data were analysed using linear regression in PLINK (Purcell et al., 2007). BMI and smoking status (current, previous or never) were incorporated as covariates. In addition, data were analysed using logistic regression by dichotomising study participants according to 'young' ≤62.4 or 'old' >62.4 age of onset. Age cut-offs were defined by the median age of the cohort. Genotypic association analysis of the NPHSII data was performed by Philippa Talmud at University College London.

Chapter 3 – Evidence for an FCHL-Susceptibility Locus (Loci) on Chromosome 21q

3.1) Introduction

3.1.1) Linkage-Based Approaches in Disease Gene Mapping

Towards the end of the 20th century, linkage analysis on a genome-wide scale became feasible. Using naturally occurring DNA variations, typically microsatellites, one can track the segregation of a given marker(s) with a phenotype in pedigrees, thus identifying potential disease susceptibility loci. Such loci can then be the focus of subsequent genetic and functional assays with the ultimate aim of finding the defective gene/causative mutation. These methodologies have generated a wealth of success stories in disease gene identification, particularly when applied to Mendelian disorders such as cystic fibrosis (Riordan et al., 1989) and Tangier's disease (Brooks-Wilson et al., 1999).

Family-based linkage analyses have also facilitated the identification of genes that underlie disorders with a complex inheritance pattern, such as *BRCA1* in some cases of breast cancer (Hall et al., 1990; Miki et al., 1994), *EDNRB* in some cases of Hirschsprung's disease (Puffenberger et al., 1994) and *NOD2/CARD15* in some cases of Crohn's disease (Hugot et al., 2001; Hugot et al., 1996). Nevertheless, a significant hurdle in mapping complex traits by linkage analysis is that the phenotype under consideration is likely result from different combinations of genetic variants, found in different genomic locations, in different families/individuals, so-called genetic heterogeneity. This can dilute the linkage signal unless appropriate steps are taken to accommodate genetic heterogeneity in linkage analysis.

In the case of early onset breast cancer, the chromosome 17q region harbouring *BRCA1* was identified as a susceptibility locus by considering the cumulative LOD score in a subset of 23 extended families with 146 cases of breast cancer, ranked by age of onset (Hall et al., 1990). Thus a cumulative LOD score of greater than 5 was attained for those families with an average age of diagnosis of 48 years or below. A sharp drop to a negative cumulative LOD score was observed once the average age of diagnosis exceeded 52 years. This approach illustrated a simple means of addressing heterogeneity in linkage analysis of a complex disease, focussing upon a subset of more genetically homogenous (early-onset) families. Likewise, the

identification of the *EDNRB* Hirschsprung's disease locus was aided by linkage studies in a large inbred Old Order Mennonite community (Puffenberger et al., 1994), an isolated population, also therefore more genetically homogeneous.

The identification of *NOD2* as a Crohn's disease (CD) gene was a notable success for linkage analysis in complex trait genetics (Hugot et al., 2001; Hugot et al., 1996). Subsequently, three common *NOD2* polymorphisms (R702W, G908R and L1007PfsX2) were associated with Crohn's disease in both family-based and case-control studies (Cuthbert et al., 2002; Hampe et al., 2001; Hugot et al., 2001). In one study involving 429 British CD patients and 290 controls, R702W was present at a frequency of 9.1% in cases compared to 3.5% in controls (P=0.0001), whilst the G908R polymorphism occurred at a frequency of 3.4% of CD patients versus 0.6% of controls (P=0.0014) and L1007PfsX2 had an allele frequency of 6.6% in cases compared to 2.1% of controls (P=0.0002). A dosage effect of these polymorphisms was observed in this investigation and in a subsequent meta-analysis of 37 studies involving individuals of Caucasian (non-Jewish) descent (Cuthbert et al., 2002; Economou et al., 2004). In the meta-analysis, carriers of one copy of R702W, G908R or L1007PfsX2 had an odds ratio for CD of 2.20 (95% CI: 1.84–2.62), 2.99 (95% CI: 2.38–3.74), and 4.09 (95% CI: 3.23–5.18)] respectively (Economou et al., 2004). In carriers of two or more alleles this increased to 17.1 (95% CI: 10.7–27.2). However, whilst at least one of the three *NOD2* variants occurred in ~36% of the British CD patients (Cuthbert et al., 2002), their combined frequency of ~7.8% in a study of in 3575 healthy Caucasian controls (Hugot et al., 2007) demonstrates that they are neither necessary nor sufficient for CD development and that other genetic (and/or non-genetic) factors are involved.

3.1.2) Improving Linkage Signal Strength and Resolution by Fine-Mapping

Following a genome-wide screen, a common first-line approach in the interrogation of a linkage interval is to fine-map the region, in order to extract further genetic information from the study cohort. Computations using nuclear families indicate that an increase in marker density from 1 marker per 10cM (the standard genome-wide screen density) to a density of 1 per 2cM makes it possible to extract almost 100% of the genetic information available (Atwood and Heard-Costa, 2003; Evans and Cardon, 2004). It is argued that an increase in marker density will increase the strength of a true positive linkage signal (Wiltshire et al., 2005) and in turn improve the resolution of linkage interval. In doing so, one can reduce the number of

plausible candidate genes that warrant further investigation in the search for disease-causing variants/mutations.

In a theoretical example involving 200 nuclear families, each comprising 7 individuals, and a quantitative trait locus (QTL) explaining 20% of trait variance, an increase in marker map density from 1 per 10cM to 1 per 2cM, decreased the error in localisation of a linkage signal by ~15% (Atwood and Heard-Costa, 2003). Narrowed linkage regions (defined as peak LOD-score minus 1) were obtained in studies of bipolar disorder (interval 5cM narrower) (Schulze et al., 2003) and schizophrenia (interval 7cM narrower) (Levi et al., 2005) following fine-mapping which decreased the average marker density to ~1 per 1cM and ~1 per 1.7cM respectively.

3.1.3) Incorporating Heterogeneity into Linkage Analysis of Complex Traits

One approach in dealing with heterogeneity in linkage analysis of complex traits is the use of admixture modelling, first proposed by Smith in 1963 (Smith, 1963), in the calculation of a heterogeneity LOD (HLOD) score (Ott, 1999). In this methodology, heterogeneity is viewed as the presence of two family types, those whose disease gene is either linked or unlinked to the marker in question. Heterogeneity is modelled as a single parameter (defined as α) which represents the probability that a family falls into the 'linked' category, although is often (mis)interpreted as the proportion of families that are 'linked' (Lin and Biswas, 2004). This parameterisation means that the probability of being 'linked' is the same for all families, an assumption which is unlikely to be true in a real family-based dataset. The use of the α parameter also assumes that all families, linked and unlinked are identically distributed which in practical terms is inevitably not the case: families tend to exist in various sizes and structures (Lin and Biswas, 2004). Likewise, errors in disease model specification, required for the parametric HLOD approach, can lead to a decrease in power to detect linkage or biased estimates of α, leading Whittemore and Halpern to advise against this approach (Whittemore and Halpern, 2001).

An alternative methodology to address heterogeneity in complex trait genetics was proposed by Hauser and colleagues (Hauser et al., 2004). Ordered Subset Analysis (OSA) subsets the pedigrees studied on the basis of a family-specific, trait-related covariate, without *a priori* specification of the subset, in order to identify the subset of families that contribute maximally to the linkage signal. The result of OSA is to increase the linkage score, potentially also narrowing the linkage interval, and also to define a more homogenous sample of families for whom the same gene may underlie

their phenotype. In defining the trait-related covariate the list of candidate disease genes warranting further investigation could also be refined.

The OSA methodology has been applied to a number of datasets, including that of 346 families participating in the Genetics of Early Onset Cardiovascular Disease linkage study (Hauser et al., 2004; Shah et al., 2006). Maximum-OSA-LOD family subsets were identified based upon average family lipoprotein values (using all affected individuals in a family). For example, a subset of 120 families ranked according to descending HDL:TC ratio produced an OSA LOD of 5.10 on chromosome 3q, compared to a overall genome-wide LOD of 2.24 for early onset CAD, using all family data (P=0.008). Subsequent phenotypic characterisation of these 120 families revealed that they had significantly lower triglycerides, total and LDL-cholesterol and higher HDL-cholesterol when compared to the non-OSA subset families leading the authors to conclude that this interval exerts an effect upon early onset CAD in the absence of traditional lipid risk factors (Shah et al., 2006).

3.1.4) Previous Genome-Wide Linkage Screens for FCHL

Previous genome-wide linkage studies for FCHL have involved cohorts of British (Naoumova et al., 2003), Dutch (Aouizerat et al., 1999) and Finnish (Pajukanta et al., 1999) families. These studies highlighted a number of susceptibility loci that potentially contribute to FCHL-lipid traits, most notably the 1q21–23, 8p23.3-22 and 11p14.1–q12.1 chromosomal regions.

The chromosome 1q21-23 interval was initially identified in Finnish families through linkage analysis targeted to ten regions of the genome harbouring known lipid-metabolism genes. Assuming a dominant mode of inheritance, a two-point LOD score of 5.93 was returned between markers D1S104 and D1S1677 (Pajukanta et al., 1998). A subsequent multipoint analysis of genome-wide scan data from a total of 35 Finnish families (31 from the previous study) supported the initial findings (Pajukanta et al., 1999). Likewise, combined analysis of the Dutch and Finnish genome-wide datasets implicates the locus in the susceptibility to the FCHL-triglyceride trait (Pajukanta et al., 2003). Association analysis in 60 Finnish FCHL families (721 genotyped individuals) implicated the upstream stimulatory factor 1 (*USF1*) gene in the inheritance of FCHL lipid traits. In particular, transmission of the haplotype of rare alleles for SNPs rs3737787 and rs2073658 was reduced to men affected by the FCHL-triglyceride trait (P=0.004) therefore suggesting a protective role for this haplotype (Pajukanta et al., 2004). USF1 is a transcription factor

involved in the regulation of numerous lipid metabolism genes (Moore et al., 2003; Ribeiro et al., 1999; Salero et al., 2003). Whilst no obvious deleterious mutations have been indentified, the intronic variant rs2073658 resides on a conserved section of DNA thought to bind nuclear proteins and be important in modulating USF1 expression (Naukkarinen et al., 2005).

Linkage to chromosome 11p14.1-q12.1 was first recorded in the Dutch genome-wide screen (Aouizerat et al., 1999). Data were analysed by considering FCHL as a dichotomous trait where affected family members had either elevated plasma triglyceride (>2.3mmol/l, cholesterol (>6.5mmol/l) or apoB (>2 standard deviations above mean age-adjusted levels). In multipoint analysis, involving 18 families, a maximum LOD score of 2.6 was obtained at 62cM (Aouizerat et al., 1999). Subsequently, a genome-wide screen involving 45 white-British FCHL families, obtained a NPL score of 2.9 for the FCHL-triglyceride trait (defined as serum triglyceride levels >90th percentile age- and sex- specific values) at 49cM on chromosome 11p (Naoumova et al., 2003)

Evidence from a non-FCHL study supports the finding that the 8p23.3-22 genomic interval contains a QTL for total cholesterol level. Initial data from the British FCHL cohort generated a LOD score of 2.20 for a serum QTL at 8.3 cM (Naoumova et al., 2003). This finding was corroborated in a study of 459 families with atherogenic dyslipidemia (defined as plasma triglyceride ≥75th percentile and serum HDL-cholesterol ≤25th percentile adjusted for age, gender and population), where a LOD score of 4.05 was calculated at 15cM (Yu et al., 2005).

In summary, although evidence strongly implicates three genomic regions in the aetiology of FCHL, more susceptibility loci/disease genes remain to be identified. It is worth noting that though the 1q21-23 interval contributes to FCHL in Finnish families, there is no evidence of linkage to this interval in British FCHL families (LOD<0.3). Also, linkage analysis suggests that less than half of the British FCHL cohort contribute to the positive linkage signals detected on chromosomes 8 and 11 (Naoumova et al., 2003). Since FCHL linkage studies to date have utilised small cohorts of families (45 in the 2003 British genome-wide screen), further assessment of FCHL aetiology using a larger, better-powered FCHL cohort and analysis methods that take into account the genetic heterogeneity of the condition is an appropriate approach to take in order to identify additional FCHL susceptibility loci, with the ultimate aim of discovering causative genetic variants.

3.2) Results

3.2.1) Identification of a Chromosome 21q22.2-22.3 FCHL Susceptibility Locus

A total of 400 genome-wide microsatellite markers (screening set 16) were genotyped in 1404 individuals from 182 extended white-British FCHL families (Table 3.1) by the Marshfield Mammalian Genotyping Service. Linkage analyses were performed by considering the component lipid traits of the FCHL phenotype as dichotomous traits, consistent with previous FCHL linkage studies (Naoumova et al., 2003; Pajukanta et al., 1999). The strongest linkage signal to emerge from these analyses was found on chromosome 21q. In light of this result, the initial linkage analyses and subsequent fine-mapping of this interval are described in more detail below and further examination of this region is described in subsequent chapters of this thesis.

As part of the initial genome-wide linkage screen, six chromosome 21 microsatellite markers (markers indicated by 'G', Table 3.2) were genotyped in the FCHL families. Linkage analyses produced maximum non-parametric LOD (NPL) scores of 3.00 (P=1.8x10^{-3}) at marker D21S1411 (51cM, 44.16Mb) and 2.64 (P=5.1x10^{-3}) at marker D21S1446 (57.8cM, 48.04Mb), for the triglyceride and cholesterol traits, respectively. No evidence of linkage was found when data were dichotomised according to affection for the triglyceride and cholesterol traits combined (NPL=0.55 and 1.44 at 51cM and 57.8cM, respectively). The 1-LOD support interval of the triglyceride linkage peak, (i.e. the interval surrounding the linkage peak where NPL=maximum NPL-1) spans approximately 5.3Mb on chromosome 21q22.2-22.3. In comparison, for the cholesterol peak, the 1-LOD support interval spans around 8.6Mb.

Table 3.1) Descriptive Data Relating to 182 FCHL Cohort Families Genotyped for Chromosome 21 Mircosatellite Markers

	All Subjects	90TG Affected[a]	95TC Affected[b]
N (Male/Female)	700/704	313/228	339/220
Age (years)	44.33±15.93	46.49±15.28	47.23±14.52
BMI (kg/m^2)	25.20±3.75	26.24±3.13	25.99±3.01
Serum Triglyceride (mmol/l)	1.74	3.44	3.08
Serum Cholesterol (mmol/l)	6.37±1.63	7.36±1.64	7.85±1.25
Affected Sibling Pairs (N)		236	237
Affected Relative Pairs (N)		654	705

Age, BMI and lipid levels are mean ± standard deviation unless stated

[a] 90TG=triglyceride ≥90th percentile age- and sex-specific value

[b] 95TC=cholesterol ≥95th percentile age- and sex-specific value

[c] Median values

Table 3.2) Chromosome 21 Microsatellite Markers Genotyped in FCHL Cohort

		Position			
Marker	Study[a]	cM	Mb	Heterozygosity[b]	Size Range (bp)[c]
D21S1437	G	13.00	21.65	0.79	103-155
D21S2052	G	25.00	28.82	0.85	96-166
D21S1909	F	28.48	32.50	0.83	232-250
D21S1440	G	37.00	39.14	0.69	145-181
D21S2055	G	40.00	41.19	0.84	105-213
D21S1260	F	46.71	42.80	0.75	200-214
D21S1411	G	51.00	44.16	0.87	264-340
D21S1259	F	52.50	45.32	0.67	208-228
D21S1912	F	53.05	45.58	0.82	173-205
D21S1446	G	57.80	48.04	0.73	187-235

[a] G indicates markers genotyped in the genome-wide screen. F indicates markers typed in fine-mapping

[b] Calculated in FCHL cohort using PedStats (Wigginton and Abecasis, 2005).

[c] According to Marshfield Mammalian Genotyping Service marker information.

3.2.2) Fine-Mapping Strengthens the Evidence of Linkage for FCHL-Lipid Traits with Chromosome 21q22.2-22.3

Four additional microsatellite markers (D21S1909, D21S1260, D21S1259 and D21S1912) were genotyped, in order to extract further genetic data from the 182 FCHL cohort families (Table 3.2). Markers were selected on the basis of their distance from and interpolation between other genotyped markers and also their estimated levels of heterozygosity, thus ensuring that the most informative markers were chosen. Marker density was increased from approximately one marker per 5.93cM to one per 2.97cM, across the region of linkage.

On completion of inheritance and haplotype checks, 5501 genotypes were added to the 8890 genotypes accrued previously. Non-parametric linkage analysis using the combined dataset, comprising 10 microsatellite markers, was performed. The maximum NPL score for the FCHL-triglyceride trait was increased from 3.00 to 3.35 at marker D21S1411 (51cM, 44.16Mb, P=6x10^{-4}, 1-LOD support interval refined from ~5.3 to ~4.9Mb) and for the FCHL-cholesterol trait at marker D21S1446 (57.8cM, 48.04Mb) was increased from 2.64 to 2.97 (P =2x10^{-3}, 1-LOD support interval refined from~8.6 to ~8.3Mb) (Figure 3.1, Table 3.3). Analysis of total plasma cholesterol as a quantitative trait gave a maximum QTL LOD score of 1.97 (P =1.3x10^{-3}) at 46.7cM (Figure 3.1, Table 3.3). The maximum triglyceride QTL LOD score was 0.724 at marker D21S2055 (40cM, 41.19Mb).

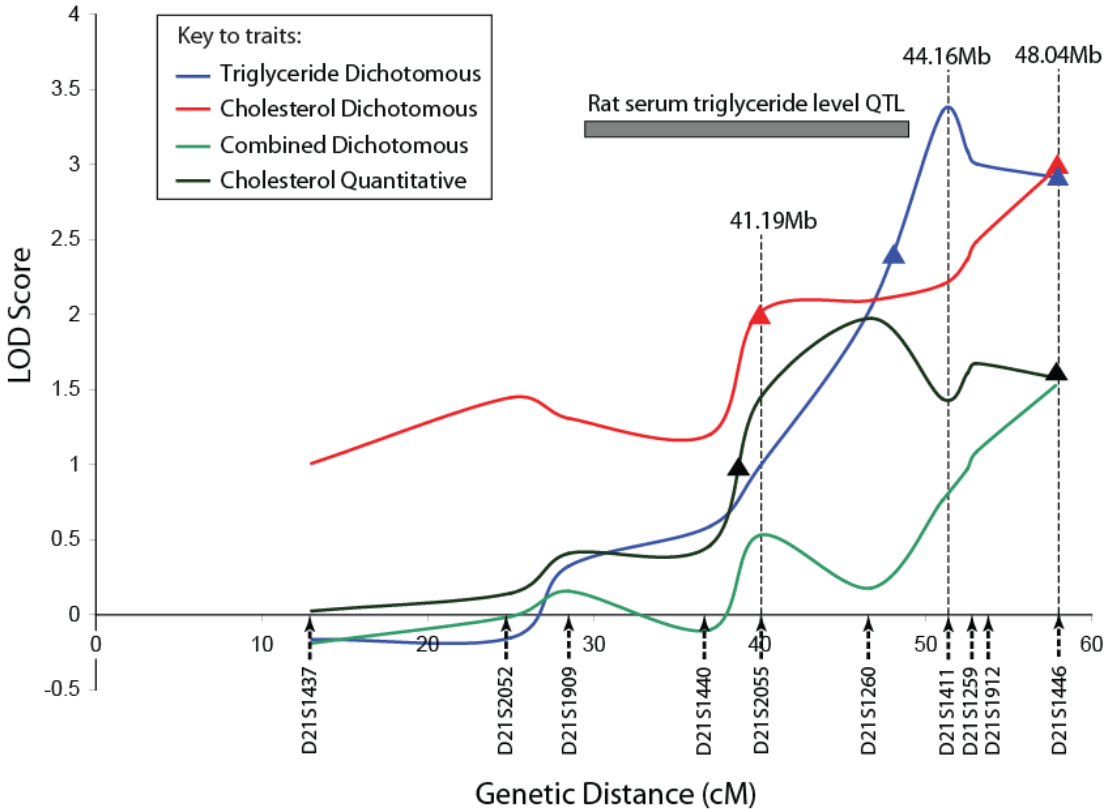


Figure 3.1) Chromosome 21 Linkage Analysis

Linkage curves are illustrated for Allegro analysis of the FCHL triglyceride, cholesterol and combined dichotomous traits. Merlin-regress analysis of cholesterol level as a quantitative trait is also shown. Non-parametric LOD scores are plotted against Kosambi centi-Morgan (cM) distance along the chromosome, with some corresponding megabase (Mb) distances also marked. 1-LOD support intervals are marked by triangles. A QTL for serum triglyceride level in the syntenic region of the rat genome is shown.

3.2.3) Heterogeneity Analysis Increases Evidence of Linkage of FCHL Lipid Traits to Chromosome 21q22.3

In order to incorporate the effects of potential heterogeneity on the 21q22.2-22.3 FCHL linkage signals, OSA was implemented. Family-specific LOD scores at a conditioning marker (any of those genotyped in the original genome-wide screen, plus 4 interpolated steps between these markers) were used as the ranking covariate. As required in the OSA methodology, LOD scores were calculated using the KACLOD approach (Kong and Cox, 1997).

The strongest OSA linkage signal for the triglyceride trait was again detected at marker D21S1411 (51cM, 44.16Mb). Specifically, the maximum OSA-LOD score was 6.26, calculated using data from a subset of 102 families (56%), indexed according to their family-specific LOD scores at marker D6S1022 on chromosome 6. This finding was supported by the calculation of six further OSA-LODs >5 at marker D21S1411, obtained using index markers at five additional chromosomal locations and overlapping family subsets (Table 3.3, Figure 3.2).

Three OSA linkage signals were detected for the FCHL-cholesterol trait (Table 3.3, Figure 3.3). At marker D21S1411 (51cM, 44.16Mb), a maximum OSA-LOD of 5.87 was obtained using a subset of 47 families (26%), ranked according to their family-specific LOD scores at ~52.2cM (42.7Mb) on chromosome 14 . In comparison, the maximum OSA-LOD score at marker D21S1446 (57.8cM, 48.04Mb) was 6.14 and was generated using a subset of 101 families (55%) ranked according to their family-specific LOD score at D1S1653 (164.1cM, 157.93Mb) on chromosome 1. This signal was supported by eight further OSA-LODs >5 at the same locus, obtained using different, but overlapping family subsets ranked at index markers on chromosomes 2, 3, 5, 6, 7, 10, 13 and 17. The third OSA-linkage signal was detected outside of the original 21q22.2-22.3 FCHL linkage interval, between markers D21S1909 and D21S1440 (~36.5Mb). Here, an OSA-LOD of 5.27 was obtained using a subset of 57 families, ranked according to their family-specific LOD-scores at marker TTCA006 (106.1cM, 75.60Mb) on chromosome 17.

Table 3.3) Chromosome 21 Loci with OSA LODs >5

			LOD Scores						
Trait[a]	Nearest Marker	Position (cM)	Baseline	OSA	Difference	*P*-Value	Index Locus[b]	Families (N)	Overlap (N)
Triglyceride	D21S1411	51.0	1.96	6.26	4.30	9.5×10^{-8}	6:31	102	
Supporting evidence:		51.0	1.96	5.85	3.90	2.4×10^{-6}	1:21	36	21
		51.0	1.96	5.81	3.85	3.4×10^{-6}	19:32	69	55
		51.0	1.96	5.72	3.77	6.3×10^{-6}	9:75	87	63
		51.0	1.96	5.48	3.52	3.4×10^{-5}	12:55	95	71
		51.0	1.96	5.40	3.44	5.9×10^{-5}	22:29	79	59
		51.6	1.88	5.25	3.37	9.2×10^{-5}	7:1	91	61
Cholesterol	D21S1446	57.8	2.22	6.14	3.92	1.0×10^{-5}	1:86	101	
Supporting evidence:		57.8	2.22	5.88	3.66	5.2×10^{-5}	2:50	84	62
		57.8	2.22	5.46	3.24	0.0005	13:19	135	81
		57.8	2.22	5.44	3.22	0.0006	10:19	109	70
		57.8	2.22	5.31	3.09	0.0011	7:30	120	80
		57.8	2.22	5.23	3.01	0.0015	3:40	143	86
		57.8	2.22	5.17	2.95	0.0020	17:54	122	82
		57.8	2.22	5.12	2.91	0.0025	6:55	112	75
		57.8	2.22	5.05	2.83	0.0035	5:17	94	68
Cholesterol	D21S1411	51.0	1.01	5.87	4.86	1.3×10^{-8}	14:19	47	
Cholesterol	D21S1909	33.6	0.50	5.27	4.77	2.6×10^{-8}	17:61	57	

[a] Dichotomous traits: triglyceride, affected=triglyceride≥90th age- and sex-matched percentile level; cholesterol, affected=cholesterol≥95th age- and sex-matched percentile level

[b] Chromosome:Marker at which families are ranked by LOD score

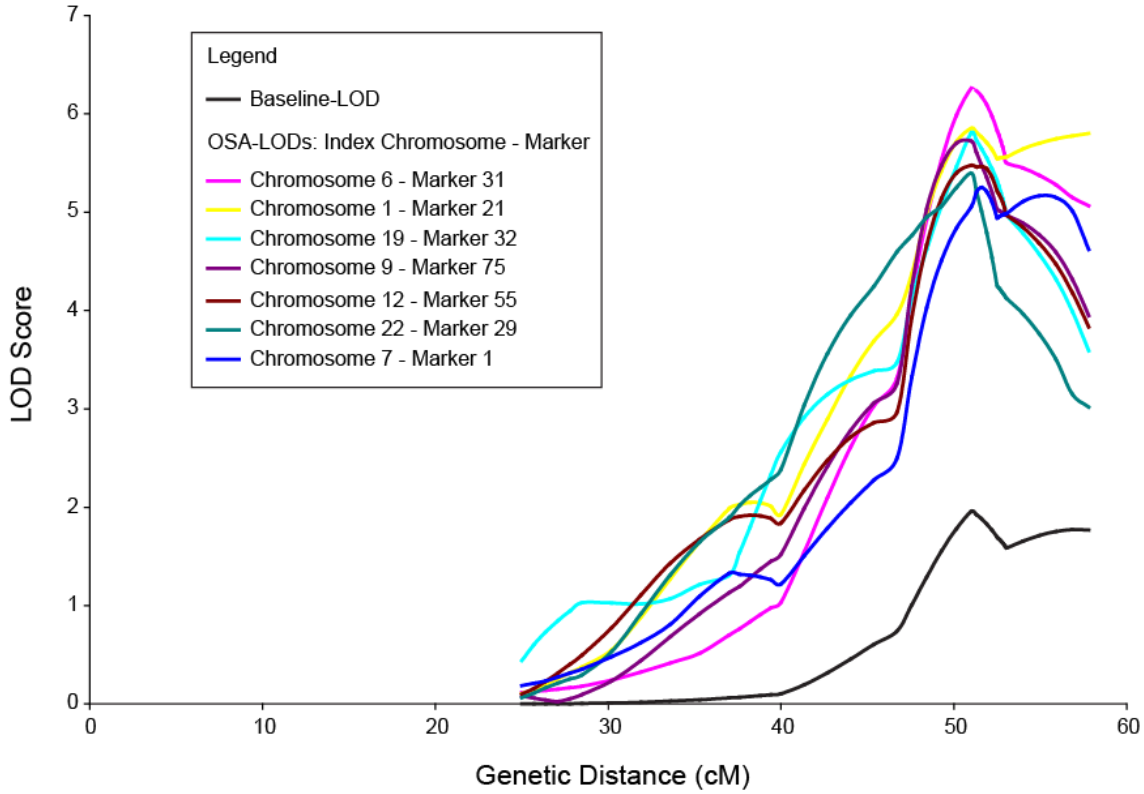


Figure 3.2) OSA Strengthens Evidence for a Triglyceride Trait Susceptibility Locus at 51cM on Chromosome 21q22.3

Subsets of FCHL cohort families, ranked at seven independent index loci produced OSA LOD scores greater than 5. Baseline LOD score is indicated by black line, subset LOD scores by other colours.

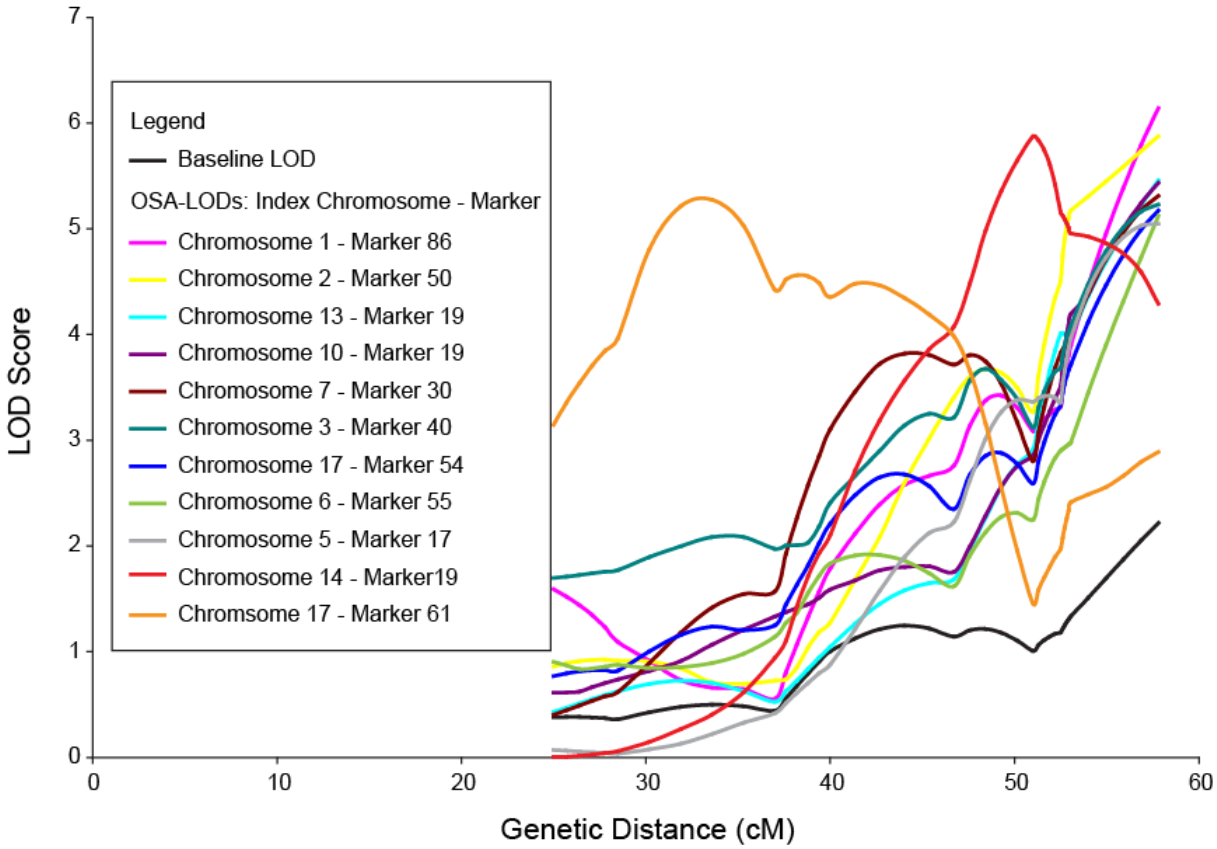


Figure 3.3) OSA Provides Evidence for Multiple FCHL-Cholesterol Trait Susceptibility Loci on Chromosome 21q

Subsets of FCHL cohort families, ranked at eleven independent index loci produced OSA LOD scores greater than 5. Baseline LOD score is indicated by black line, subset LOD scores by other colours. Peak OSA-LOD scores were focussed at 33.6, 51.0 and 57.8cM.

The results of OSA overall led to an increase in the strength of the initial genome-wide and fine-mapping linkage signals for FCHL-lipid traits on chromosome 21q22.3 as well as narrowing the 1-LOD support intervals surrounding each linkage peak (Table 3.4). The combined 1-LOD support intervals, for the 51cM cholesterol and triglyceride peaks, span ~3.8Mb, from 43.3-47.1Mb, on chromosome 21q22.3. This interval, (from this point forward referred to as the 21q22.3 FCHL linkage interval) will be the major focal point of subsequent studies described in Chapters 3, 4 and 5.

Table 3.4) Increased Strength and Specificity of Chromosome 21q22.3 Linkage Signals by Fine Mapping and Heterogeneity Analysis

		GW-screen		Fine-mapping		OSA	
Trait	Peak Marker	NPL	1-LOD support	NPL	1-LOD support	Subset LOD	1-LOD support
90TG	D21S1411	3.00	5.3	3.35	4.9	6.26	3.6
95TC	D21S1411	2.14	13.2	2.20	5.0	5.87	3.3
95TC	D21S1446	2.64	8.6	2.97	8.3	6.14	1.2

GW=genome-wide

90TG=triglyceride ≥90th percentile age- and sex-specific value

95TC=cholesterol ≥95th percentile age- and sex-specific value

3.2.4) Gene Content of the 21q22.3 FCHL Linkage Interval

Data-mining of the Ensembl genome browser (URL3) revealed that the refined 3.8Mb linkage interval contains 61 known protein-coding genes (defined as those genes with curated RefSeq protein records, Table 3.5). The interval also encodes six genes with predicted open reading frames but where evidence of protein expression and functional characterisation is currently lacking, as well as seven pseudogenes and 20 known or predicted non-coding RNAs. There are no known microRNA genes within the region.

Table 3.5) Protein Coding Genes within 21q22.3 FCHL Linkage Interval

Gene Symbol	Gene Name	Position[a]
C2CD2	C2 calcium-dependent domain containing 2	43,305,221-43,373,999
ZNF295	Zinc finger protein 295	43,406,940-43,430,496
UMODL1	Uromodulin-like 1	43,483,068-43,563,105
ABCG1	**ATP-binding cassette, sub-family G, member 1**	**43,619,799-43,720,714**
TFF3	Trefoil factor 3 (intestinal)	43,731,777-43,735,761
TFF2	Trefoil factor 2	43,766,466-43,771,237
TFF1	Trefoil factor 1	43,782,391-43,786,703
TMPRSS3	Transmembrane protease, serine 3	43,791,999-43,816,955
UBASH3A	Ubiquitin associated and SH3 domain containing A	43,824,008-43,867,791
RSPH1	Radial spoke head 1 homolog	43,892,596-43,916,464
SLC37A1	**Solute carrier family 37, member 1**	**43,916,128-44,001,550**
PDE9A	Phosphodiesterase 9A	44,073,746-44,195,619
WDR4	WD repeat domain 4	44,263,204-44,299,678
NDUFV3	NADH dehydrogenase (ubiquinone) flavoprotein 3	44,299,754-44,427,677
PKNOX1	PBX/knotted 1 homeobox 1	44,394,620-44,453,692
CBS	**Cystathionine beta synthase**	**44,473,301-44,497,053**
U2AF1	U2 small nuclear RNA auxiliary factor 1	44,513,066-44,527,697
CRYAA	Crystallin, alpha A	44,589,118-44,592,915
SIK1	**Salt-inducible kinase 1**	**44,834,395-44,847,008**
HSF2BP	Heat shock transcription factor 2 binding protein	44,949,072-45,079,374
RRP1B	Ribosomal RNA processing 1 homolog B	45,079,429-45,194,144
PDXK	Pyridoxal (pyridoxine, vitamin B6) kinase	45,138,975-45,194,151
CSTB	Cystatin B	45,192,393-45,196,326
RRP1	Ribosomal RNA processing 1 homolog	45,209,394-45,225,174
AGPAT3	**1-Acylglycerol-3-phosphate O-acyltransferase 3**	**45,285,067-45,406,417**
TRAPPC10	Trafficking protein particle complex 10	45,432,200-45,526,433
PWP2	PWP2 periodic tryptophan protein homolog	45,527,176-45,551,063
ES1	ES1	45,553,487-45,565,605
ICOSLG	Inducible T-cell co-stimulator ligand	45,642,874-45,660,849
DNMT3L	DNA (cytosine-5-)-methyltransferase 3-like	45,666,222-45,682,099
AIRE	Autoimmune regulator	45,705,721-45,718,531
PFKL	Phosphofructokinase (liver)	45,719,934-45,747,259
C21orf2	Nuclear encoded mitochondrial protein A2-YF5	45,748,827-45,759,285
TRPM2	Transient receptor potential cation channel, subfamily M, member 2	45,770,046-45,862,964
LRRC3	Leucine rich repeat containing 3	45,875,369-45,878,739
TSPEAR	Thrombospondin-type laminin G domain and EAR repeats	45,917,775-46,131,495

Gene Symbol	Gene Name	Position[a]
KRTAP10-1	Keratin associated protein 10-1	45,958,864-45,960,078
KRTAP10-2	Keratin associated protein 10-2	45,970,240-45,971,388
KRTAP10-3	Keratin associated protein 10-3	45,977,673-45,978,643
KRTAP10-4	Keratin associated protein 10-4	45,993,606-45,994,987
KRTAP10-5	Keratin associated protein 10-5	45,999,332-46,000,481
KRTAP10-6	Keratin associated protein 10-6	46,011,149-46,012,386
KRTAP10-7	Keratin associated protein 10-7	46,020,497-46,022,091
KRTAP10-8	Keratin associated protein 10-8	46,031,996-46,032,871
KRTAP10-9	Keratin associated protein 10-9	46,047,040-46,048,294
KRTAP10-10	Keratin associated protein 10-10	46,057,273-46,058,370
KRTAP10-11	Keratin associated protein 10-11	46,066,331-46,067,564
KRTAP12-3	Keratin associated protein 12-3	46,067,057-46,078,258
KRTAP12-4	Keratin associated protein 12-4	46,074,130-46,074,576
KRTAP12-2	Keratin associated protein 12-2	46,086,106-46,086,844
KRTAP12-1	Keratin associated protein 12-1	46,101,491-46,102,078
KRTAP10-12	Keratin associated protein 10-12	46,117,087-46,117,959
UBE2G2	**Ubiquitin-conjugating enzyme E2 G2**	**46,188,955-46,221,934**
SUMO3	SMT3 suppressor of mif two 3 homolog 3	46,191,374-46,238,694
PTTG1IP	Pituitary tumor-transforming 1 interacting protein	46,269,500-46,293,752
ITGB2	Integrin, beta 2	46,305,868-46,351,904
ADARB1	Adenosine deaminase, RNA-specific, B1	46,493,768-46,646,478
POFUT2	Protein O-fucosyltransferase 2	46,683,843-46,707,813
COL18A1	**Collagen, type XVIII, alpha 1**	**46,825,052-46,933,634**
SLC19A1	Solute carrier family 19, member 1	46,913,486-46,964,316
PCBP3	Poly(rC) binding protein 3	47,063,608-47,362,368

[a]Position is according to Ensembl database and is based upon human genome build 37

Candidate genes (see 4.2) are highlighted in bold.

3.3) Discussion

Combined non-parametric linkage analysis using data for ten chromosome 21 microsatellite markers, genotyped in 182 FCHL families, generated a maximum NPL score of 3.35 for the FCHL-triglyceride trait at marker D21S1411 (P=6x10^{-4}) and 2.97 for the FCHL-cholesterol trait at marker D21S1446 (P =2x10^{-3}). According to criteria set out by Lander and Kruglyak for significance threshold in linkage analysis of complex traits, these data provide suggestive evidence of linkage for FCHL-lipid traits to the chromosome 21q22.2-22.3 (Lander and Kruglyak, 1995).

Given that the peak LOD score for the FCHL-cholesterol trait is observed at the final 21q marker it is possible that this score is inflated, due to there being less identity-by-descent information at the telomere. Analysis of total plasma cholesterol as a quantitative trait gave a maximum QTL LOD score of 1.97 (P=1.3x10^{-3}) at 46.7cM (Figure 3.1, Table 3.3). Therefore, this may represent a more accurate localisation of the variant(s) within the 21q22.2-22.3 interval that increases transmission of the FCHL-cholesterol trait.

The linkage data are further supported by the results of other genome-wide linkage studies examining lipid levels in subjects with cardiovascular disease (Coon et al., 2002; North et al., 2005). In a cohort comprising 1124 hypertensive families (76% of individuals affected: systolic blood pressure ≥160 or diastolic blood pressure ≥100 mmHg or receiving to forms of treatment for existing hypertension), genome-wide linkage analysis was used to detect QTLs jointly affecting triglyceride and LDL-cholesterol levels (additive genetic correlation=0.25, $P \leq 0.001$) (North et al., 2005). A maximum empirical LOD score of 3.9 was obtained on chromosome 21q at ~54 cM (between markers D21S2055 and D21S1446). In a second investigation, a genome-wide scan for LDL-cholesterol level was performed using 2977 individuals (500 families) from the NHLBI Family Heart Study (Coon et al., 2002). A maximum multipoint LOD of 2.74 was obtained at 45.2cM (nearest genotyped marker D21S1246, 40.87Mb). Additionally, a QTL for serum triglyceride level resides in the syntenic region of the rat genome (Figure, 3.1, (Moreno et al., 2003)), thus providing further evidence for the involvement of this genomic region in modulating circulating lipid levels.

Also of potential interest, a QTL for percentage body fat was identified in a study of 260 European-American families recruited on the basis of an obese proband

(BMI≥40kg/m^2), with one obese sibling (BMI≥30kg/m^2) plus one sibling and parent of normal weight (BMI≤27kg/m^2), where a LOD score of 4.27 was calculated at 57.8cM (Li et al., 2004).

Most chromosome 21-orientated research involves Down syndrome (DS, trisomy 21), where an extra copy of whole or part of the chromosome is inherited by affected offspring. Congenital heart defects occur in around 45% of DS patients (Dolk et al., 2011; Freeman et al., 1998) however, the risk of cardiovascular disease is said to be less in adults with DS than in the general population. Murdoch and colleagues proposed that DS represents an 'atheroma-free model' (Murdoch et al., 1977). Post-mortem examination of 5 institutionalised DS patients (ages= 40, 44, 56, 60 and 66 years) in this study showed no evidence of atherosclerosis. In the same investigation, a comparison of 70 DS patients and 70 age-matched mentally-defective but non-trisomy 21 controls revealed no significant differences in serum lipid levels. In a more recent study, Draheim *et al.*, examined lipid parameters and carotid artery intima-media thickness, as a measure of atherosclerotic burden, in 52 DS patients plus 52 age-, sex- and race-matched controls (Draheim et al., 2010). The DS patients in this investigation were community-based and therefore were not subject to the controlled diet and levels of physical activity that would be present in care-facility environment. However, the results of this study still agree with those of Murdoch et al, namely the DS adults possessed lower intima-media thickness than the controls (0.43 ± 0.07 vs. 0.48 ± 0.09 mm, P<0.001). In this instance the DS patients did have slightly increased serum triglyceride levels compared to the controls (1.43± 0.62 vs. 1.17±0.60 mmol/l, P<0.048) however, this was attributed to their increased dietary fat intake (assessed by questionnaire, P=0.004) (Draheim et al., 2010).

In summary, these studies suggest that an extra copy of chromosome 21 maybe atheroprotective, which in turn begs the question as to whether this effect is mediated via circulating lipids. Whilst no significant alteration in lipid levels was observed in the small numbers of trisomy 21 patients studied here, it is possible that loss, rather than gain-of -function would have a more significant and detrimental effect on plasma triglyceride and/or cholesterol levels.

OSA (Hauser et al., 2004) was implemented in order examine the evidence for linkage of FCHL-lipid traits to chromosome 21q22.3 whilst taking into account genetic

heterogeneity. The aim of this analysis was to define more distinct LOD-score peaks for the FCHL-triglyceride and cholesterol traits, in turn reducing the number of potential FCHL candidate genes warranting further investigation.

Using chromosome 21 test markers and index markers at any genomic location, the maximum OSA LOD for the triglyceride trait was 6.26 at 51cM. For the cholesterol trait, two potential 21q22.3 susceptibility loci were identified with OSA-LODs of 5.87 at 51cM and 6.14 at 57.8cM, respectively. The 1-LOD support intervals for all three 21q22.3 linkage peaks were narrowed compared to those obtained by non-parametric linkage analysis of the whole dataset (Table 3.4). In combination, the 1-LOD support interval for the triglyceride and cholesterol OSA linkage peaks at 51cM span ~3.8Mb, a region containing 61 known protein-coding genes.

Additionally, and of potential interest, an OSA-LOD score of 5.27 for the FCHL-cholesterol trait was obtained at 33.59cM on chromosome 21q22.11-q22.13, between markers D21S1909 and D21S1440. This peak was not observed in non-parametric linkage analysis of the entire dataset, however recently two SNPs within this interval were associated with lipid levels and early-onset myocardial infarction (Kathiresan et al., 2009a; Smith et al., 2010). In a genome-wide association study of cardiovascular risk factors involving 525 unrelated participants in the Bogalusa Heart Study, rs8131349 was associated with increased total cholesterol level (P=4.6x10^{-8}) (Smith et al., 2010). Whilst in a meta-analysis, involving 12,713 cases and 12,821 age- and sex-matched controls, rs9982601 was associated with an increased risk of early-onset myocardial infarction (men ≤50 years old or women ≤60 years old) (OR=1.20, (95% CI: 1.14-1.27), P=6.4x10^{-11}) (Kathiresan et al., 2009a).

It is worth noting that since the 'optimal slice' function of OSA was used, the family-specific LOD scores at index markers were not necessarily drawn from the upper end of this distribution. This means that the index locus does not inevitably harbour a reciprocal FCHL-susceptibility locus.

Additionally, whilst 56% of the FCHL cohort (102) families contribute to the maximum triglyceride trait OSA-LOD score at D21S1411, the overlap between these families and other family subsets also producing OSA-LODs >5 around this locus is incomplete. This is because the subsets are derived from ranked LOD scores, from differing genomic regions. Since FCHL is a complex, heterogeneous disorder families are unlikely to fall into the same ranking order at all genomic loci tested. It is

therefore inappropriate to compare family subsets ranked according to different covariate index loci.

The genetic architecture of FCHL, like most complex diseases, is likely to be a balance of common and rare genetic variants (Horswell et al., 2009; Manolio et al., 2009; Schork et al., 2009). Linkage is a powerful tool in detecting regions where genetic variants in aggregate have a large phenotypic effect. Since the linkage and heterogeneity analyses presented here suggest that multiple susceptibility loci underpin the 21q22.3 FCHL-linkage signals, it seems reasonable to hypothesise that the interval could harbour both common and rare variants that in combination contribute to the raised circulating plasma cholesterol and triglyceride levels in FCHL patients

At the outset of this project, screening some or all of the 61 protein-coding genes that reside within the 21q22.3 FCHL linkage interval for rare variants/mutations, by standard sequencing methods, was both prohibitively expensive and slow, whilst 'next-generation' sequencing technology was insufficiently advanced. On the other hand, the recent advent of high-throughput SNP genotyping methodologies, sophisticated statistical techniques for genotype imputation and association analysis of large datasets, together with a readily available catalogue of common genetic variants and linkage disequilibrium measures (International HapMap Consortium, 2003), provided a cost-effective method of further interrogating the linkage interval with respect to the role of common variation. Chapters 4 and 5 describe the design, quality control and analysis of a large-scale family-based association study, targeting candidate genes within the FCHL linkage interval.

Chapter 4 – Design and Quality Control of a 21q Targeted Association Study

4.1) Introduction

Whilst linkage studies can define the approximate position of a disease gene by examining segregation patterns in families (Morton, 1955), association studies exploit recombination patterns, accumulated over many years of evolutionary history, to map causal variants on a much finer scale (Morton, 1955; Morton, 2005).

The term linkage disequilibrium (LD) refers to a non-random association between alleles at two or more loci, resulting from constrained levels of recombination. LD is a reflection of the age of the variants, the distance between them, the founder population size, selection, genetic drift and recurrent mutation (Zavattari et al., 2000). The most common measures of LD are D' or r^2 values. D' is a standardised version of the coefficient of linkage disequilibrium, D (Lewontin and Kojima, 1960). For alleles at two loci, defined A, a and B, b, with allele frequencies p_A, p_a, p_B and p_b:

$$D_{AB} = p_{AB} - p_A p_B$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (pAB) and the product of the frequencies of those alleles (pA and pB)

Likewise,

$$D_{aB} = p_{aB} - p_a p_B$$

The value of D is dependent upon the frequencies of both alleles, making it difficult to interpret. As a result, the use of D' (Lewontin, 1964) is often preferred, where:

$$D' = D / D_{max}$$

and D_{max}, is the smaller value of $p_A(1 - p_B)$ and $p_B (1 - p_A)$. As the sign of D' depends on the arbitrary labelling of alleles |D'| maybe used.

The alternative measure of LD, r^2, (Hill and Robertson, 1968) is commonly expressed as the square of the correlation coefficient between the two loci:

$$r_{AB}^2 = D_{AB}^2 / p_A p_a p_B p_b$$

This measure indicates how well one SNP may act as a surrogate for another; with an r^2 value of 1 implying that markers provide identical information. In genetic studies, the widely accepted consensus r^2 value for a marker to be a suitable proxy is 0.8 (Frazer et al., 2007).

The advent of the International HapMap Project (URL7) paved the way for sophisticated family- and population-based studies, mapping causal variants for complex phenotypes. Phase I of the HapMap project catalogued over 1 million SNPs that were genotyped in 269 individuals from four geographically diverse populations (International HapMap Consortium, 2003). Subsequent data releases extended this catalogue to include some 3 million SNPs and a further seven populations (Frazer et al., 2007; International HapMap Consortium, 2005). These data were used to define 'block'-like patterns of LD between SNPs, across the human genome. The resulting 'blocks' comprised many SNPs (30-70) however the number of common haplotypes (as defined by correspondent alleles at multiple loci) present was limited (in the range of 4 to 5.6) (International HapMap Consortium, 2005).

In concert with the aggregation of HapMap data, companies such as Illumina (Oliphant et al., 2002) and Affymetrix (Chee et al., 1996; Kennedy et al., 2003) developed SNP genotyping panels, allowing straightforward genotyping of up to 1 million SNPs, in large study cohorts. Customised versions of these panels such as the Illumina GoldenGate assay allow association studies to be targeted upon specific areas of the genome which may be useful in the setting of a family-based study where a preceding linkage screen has highlighted candidate regions of the genome where trait-related variants are likely to reside (Holt et al., 2010; Zinovieva et al., 2009). This approach can also be advantageous in limiting genotyping costs and computing power required for subsequent data analyses.

4.1.1) Advantages of Family-Based Association Study Design

Association studies may adopt a design utilising related (family-based) or unrelated (case-control/population) samples. The vast majority of studies published to date have employed a population or case-control based approach. One reason for this is due the increased time and money required to collect large numbers of family-based samples compared to sampling from unrelated individuals. However, where a cohort, such as the FCHL family cohort, has already been assembled for a previous linkage screen, this obstacle is removed.

A family-based study design has a number of advantages over those utilising unrelated individuals. Firstly, a family-based design can avoid false positive results (type 1 error) due to population stratification (Ewens and Spielman, 1995; Ewens and

Spielman, 2005). Analytical approaches that consider the transmission of variants within the family unit only, such as the transmission disequilibrium test (TDT), provide robustness to this potential confounder (Ewens and Spielman, 1995; Ewens and Spielman, 2005).

Familial cases can also have enhanced power for detecting genetic associations. For example, Antoniou and Easton (Antoniou and Easton, 2003) showed that when cases were ascertained on the basis of family history of breast cancer (mother and sister affected), the sample size of cases required to detect an associated allele with frequency 0.05, explaining 5% of the phenotypic variance was less that 500, compared to ~1000 unselected cases (Antoniou and Easton, 2003). Indeed, family-based designs may be enriched for genetic effects and are likely to involve more homogenous environmental exposures than a population based-approach (Evangelou et al., 2006). This should translate into increased ability to detect a disease pre-disposing allele, depending upon the statistical analysis method used. For example, in a genome-wide study of prostate cancer where cases with family history of the disease were utilised, the minor allele of a SNP in the promoter of *MSMB* was identified as being associated with the disease (Eeles et al., 2008). This was replicated in a subsequent study (Thomas et al., 2008). More recent functional work has demonstrated that MSMB protein levels in both tissue and urine are significantly correlated with carriage of the risk allele ($P<0.0001$) and that MSMB protein is significantly reduced in tumoriogenic versus benign prostate tissue ($P<0.0001$) (Whitaker et al., 2010).

On a more practical note, the importance of which is heightened when dealing with many thousands of genotypes, a family-based study design facilitates an extra layer of data quality control, whereby genotyping errors can be detected by checking for correct Mendelian inheritance of alleles.

It is worth noting that collaborative projects can incorporate data from a huge number of samples. For example, the first Wellcome Trust Case Control Consortium study (Wellcome Trust Case Control Consortium, 2007), utilised 2000 cases and 3000 shared controls for each of seven complex human diseases, including CHD, whilst the more recent GLGC meta-analysis (Teslovich et al., 2010) incorporated data from 46 studies, in sum involving over 100,000 individuals. However, the nature of large collaborative studies can mean that is difficult to gather uniform phenotypic data. For example, in the study of Teslovich et al., some lipid measurements were taken in the

fasting-state whilst others were not (Teslovich et al., 2010). Similarly, measurements from some individuals were obtained when they were taking lipid-lowering drugs whilst some study participants were diabetic, a condition known to affect the lipoprotein profile (Solano and Goldberg, 2006) and one of the exclusion criteria in the FCHL family cohort (Naoumova et al., 2003).

4.1.2) Quality Control Considerations in an Association Study

The genotyping effort of a large-scale association study generates many thousands of genotypes, which in the majority of cases are assigned automatically by an appropriate SNP-calling algorithm. Good quality SNP genotype data is essential so that accurate conclusions can be drawn from subsequent statistical tests. Even a small artefactual perturbation in allele calls can lead to and inflation in type 1 error rate (reviewed by (Neale and Purcell, 2008)).

Quality control can be applied at the level of the DNA sample and of the marker under consideration. Samples may be excluded upon the basis of a low call-rate, which can indicate a poor quality sample. An excessive level of heterozygosity points to potential contamination of a sample with DNA from an external source. Therefore, samples with a heterozygosity rate of more than 5 standard deviations from the mean level should be excluded (Schunkert et al., 2011).

At the marker level, quality control can begin with simple visualisation of cluster plots generated by the software that accompanies high-throughput genotyping methods, in order to ensure that discrete clusters of points representing the sample genotypes have been formed for each marker. SNPs with low genotype call-rates and those assays which fail to generate reproducible genotype calls across duplicate samples should be excluded from subsequent association analyses.

It is also recommended to apply a minor allele frequency threshold to SNPs that are tested for association. The reason for this is two-fold. Firstly, the genotype-calling methods of the Illumina and Sequenom platforms use clusters of intensity-based points in order to make a genotype call. Definition of such clusters becomes more difficult when they are composed of only a few points and thus genotype calling may become unreliable. Secondly, the statistical methods that are employed in association testing may lack power to detect association between very rare variants and a given trait (Bansal et al., 2010).

Validity of genotyping can be checked by testing for deviation from Hardy-Weinberg equilibrium (HWE) in a subsample of unrelated individuals, such as the pedigree founders (Wigginton et al., 2005). Additionally, inheritance checks can be applied to identify poorly performing genotyping assays, which consistently do not inherit in a Mendelian fashion. Recurrent inheritance problems or a deviation from HWE for a given marker(s) can be indicative of underlying copy number variation (CNV) (McCarroll and Altshuler, 2007) and therefore the regions surrounding markers identified by these checks may warrant further examination.

4.1.3) Increasing Association Study Power by Genotype Imputation

A recent advance in association analysis is the availability of sets of reference haplotypes, derived from either HapMap (URL7) or 1000 Genomes data (URL9) which can be used to impute ungenotyped markers within a study sample. The resultant dataset, comprising both directly genotyped and imputed markers, can be utilised in statistical tests for association. In generating an expanded SNP dataset one can potentially increase the power to detect association between a given marker and the phenotype under consideration. For example, in a simulation study utilising 2000 cases and controls, and a *P*-value threshold of 5×10^{-7}, imputation increased the theoretical power of Illumina 300K and Affymetrix 500K array-based studies from 0.392 and 0.363 to 0.467 and 0.450 respectively, compared to power of 0.499 for a hypothetical chip containing all HapMap SNPs (Spencer et al., 2009).

A practical example of using imputation to identify association between a SNP and particular phenotype, involves data from the Diabetes Genetics Initiative (Saxena et al., 2007) and SariNIA (Scuteri et al., 2007) cohorts, genotyped using the Affymetrix 500K array. Using these data, association between the intronic *LDLR* SNP, rs6511720 and plasma LDL-cholesterol level was initially not detected: rs6511720 is absent from the Affymetrix 500K array and is poorly tagged by the SNPs that are present (maximum $r^2=0.21$). However, genotype imputation and subsequent meta-analysis, involving these two cohorts plus data from the Finland–United States Investigation of NIDDM Genetics (FUSION) study, estimated that each copy of the rs6511720 C allele was associated with an 0.21mmol/l increase in LDL-cholesterol levels ($P=6.8\text{x}10^{-10}$) (Li et al., 2009). Crucially, imputed genotypes from 1644 population-based subjects of the KORA F3 study originally genotyped using the same Affymetrix 500K array replicated this finding ($P=4.6\times10^{-6}$, (Linsel-Nitschke et

al., 2008). Whilst the functional impact of this variant is currently unknown, these studies indicate that genotype imputation may increase power to detect such an association.

There are a number of algorithms available to perform imputation. The most widely-used are Markov-Chain Haplotyping software (MaCH, (Li et al., 2009; Li et al., 2010)) and IMPUTE (Howie et al., 2009; Marchini et al., 2007). Both of these algorithms use Hidden Markov Models and both have similar computational requirements and error rates (Marchini and Howie, 2010). For example, using the post-imputation quality measures that both algorithms generate, one can exclude potentially poorly imputed SNPs prior to association analysis. In a simulated dataset, involving 1000 cases and 1000 controls, based upon a 7Mb interval of chromosome 22 and a reference sample derived from HapMap samples, the quality measures produced by MaCH and IMPUTE were found to be highly correlated (Marchini and Howie, 2010). Despite their similarities, in a comparison of imputation programs, the user-friendliness of MaCH was noted (Nothnagel et al., 2009).

Analysis of *in silico* genotype data also requires consideration. Whilst the imputation programs generate the most likely genotype for each individual at each SNP, it is recommended that allele dosages (ranging from 0-2 and not necessarily a discrete number) are utilised to take into account the uncertain nature of the imputed data (Zheng et al., 2011). Allele dosages can also be used directly in regression-based approaches for association analysis, including GWAF (Chen and Yang, 2010) which uses data directly from MaCH.

The data presented in Chapter 3 indicate that a proportion of the FCHL cohort families harbour chromosome 21 genetic variants that segregate with their high triglyceride or cholesterol phenotype. Therefore, the hypothesis driving the investigations described in this and Chapter 5 is that a chromosome 21q gene (or genes), with an established biological link to lipid metabolism, will harbour genetic variation that contributes to the FCHL triglyceride and/or cholesterol trait. This chapter describes the design of a targeted association study, the production of a high quality SNP dataset and the acquisition of further data by genotype imputation. Additionally, the identification of potential copy number variant(s) is described.

4.2) Results

4.2.1) Identification of Candidate Genes

On the basis of published genetic and functional studies, seven plausible FCHL candidate genes, reside within the refined 3.8Mb linkage interval. In addition, *BACE2*, *LSS* and *MCM3AP*, which reside outside of the refined interval, are excellent biological candidates. Below is a summary of the published genetic, *in vitro* and *in vivo* data underpinning the candidacy of the ten genes. At this point it is worth highlighting that in mice and other rodents, the majority of cholesterol in the circulation is carried by HDL, rather the apoB-containing lipoproteins (LDL), utilised in humans (Fernandez and Volek, 2006) and therefore, in certain mouse models the linkage with a specific human lipid trait may be equivocal.

i) *BACE2*

In the 2007 Diabetes Genetics Initiative genome-wide association study (GWAS) a *BACE2* SNP, rs914186 (sixth intron), was associated with triglyceride:HDL ratio (P=7.6x10^{-5}) (Saxena et al., 2007). Biologically, BACE2 shares 45% amino acid identity and 75% similarity with BACE1 (Sun et al., 2005). Using co-immunoprecipitation and fluorescence lifetime imaging microscopy, BACE1 has been shown to interact with LRP and to facilitate its cleavage in a transfected cell line (von Arnim et al., 2005). Northern blotting showed that two *BACE2* transcripts are expressed at low levels in most adult human tissues, including liver (adipose tissue not examined, (Bennett et al., 2000)).

ii) *ABCG1*

ABCG1 is a member of the superfamily of ATP binding cassette (ABC) transporters. Human *ABCG1* mRNA levels are highest in the lung, adrenal gland, heart, spleen and liver (Klucken et al., 2000; Langmann et al., 2003).

ABCG1 was initially implicated in cholesterol efflux due to its increased expression in both human and murine-derived macrophages that were incubated with acetylated LDL and sterols, (to form lipid-loaded 'foam' cells), as well as in macrophages treated with liver X receptor agonists (Klucken et al., 2000; Tangirala et al., 2002; Venkateswaran et al., 2000). More recent studies have shown that ABCG1 effluxes cholesterol to a number of acceptors, including HDL and LDL (albeit at low levels) though not to lipid-poor ApoAI (Wang et al., 2004a).

Targeted disruption of murine *Abcg1* led to lipid accumulation in hepatocytes and in macrophages (particularly in the lung) upon administration of a high fat (21%)/high cholesterol (1.25%) diet (Kennedy et al., 2005). Specifically, levels of total cholesterol, cholesterol ester, triglyceride and phospholipids were significantly increased in the livers and lungs of Abcg*1*$^{-/-}$ mice as compared to their wild-type littermates. Conversely, transgenic mice expressing human *ABCG1* had significantly lower levels of total cholesterol, cholesterol esters, and phospholipids in their lungs and livers compared to their wild-type littermates and thus were protected from the diet-induced lipid accumulation. Despite these cellular changes in tissue lipid levels, no difference in plasma lipids between the wild-type, null or transgenic mice was observed (Kennedy et al., 2005).

In humans, levels of HDL and triglyceride are inversely correlated (Davis et al., 1980). Therefore, ABCG1 may also have an effect upon whole-body triglyceride levels. Indeed, in a study examining *Drosophila* mutants where the 5' flanking region of the *ABCG1*-homolog was disrupted through the random genomic insertion of modified transposable elements, whole-body triglyceride stores were increased by up to 1.7-fold (P<0.03, (Buchmann et al., 2007)).

iii) *SLC37A1*

SLC37A1 is highly homologous (57% identity) to a putative *Arabidopsis thaliana* glycerol-3-phosphate permease. It also shares ~30% homology with bacterial glycerol 3-phosphate permeases (Bartoloni et al., 2000). Northern blot analyses have showed highest human *SLC37A1* expression levels in bone marrow, kidney, liver and small intestine (Bartoloni et al., 2000). Based on amino acid sequence information and immunofluorescence data, it is proposed that SLC37A1 may function to transport glycerol-3-phosphate across the outer mitochondrial membrane (Bartoloni and Antonarakis, 2004) and/or into the endoplasmic reticulum (Iacopetta et al., 2010). Since the acylation of glycerol-3-phosphate is the first committed step in triglyceride/phospholipid synthesis, any alteration in glycerol-3-phosphate availability, due to mutation of *SLC37A1*, could also perturb the mitochondrial synthesis of triglyceride and in turn the incorporation ER derived triglyceride into apoB-containing lipoproteins (Figure 4.1).

iv) *CBS*

CBS catalyses the irreversible condensation of homocysteine and serine: a reaction which yields cystathionine (Figure 4.1). In humans, *CBS* mRNA is found at highest levels in the liver and pancreas (Chasse et al., 1997). In a human hepatocyte cell line cultured with 5mM homocysteine, expression levels of *SREBP-1* and the cholesterol biosynthetic genes *HMGCR* and *FDPS* were increased (Werstuck et al., 2001). Additionally, the total cell cholesterol content was ~150% that of untreated cells (Werstuck et al., 2001).

In mice, *Cbs* mRNA and protein is found predominantly in the liver and kidney (Namekata et al., 2004). $Cbs^{-/-}$ mice usually die within 2-3 weeks of birth (Watanabe et al., 1995). In a study using null mice of around this age, hepatic triglyceride levels were 11.4-fold higher in $Cbs^{-/-}$ compared to wild-type mice (P <0.0001), whilst serum triglyceride levels were also significantly increased (17.95mg/ml ±5.00(SE) versus 5.52±0.90mg/ml, P<0.04) (Namekata et al., 2004). Hepatic and serum total cholesterol levels were not significantly altered (Namekata et al., 2004).

In an alternative *in vivo* model, wild type mice were fed either a control or a high methionine, low folate (HMLF) diet for 10-weeks in order to induce hyperhomocysteinemia. After this period, the HMLF mice had higher plasma cholesterol (1.56 versus 0.91mM) and triglyceride (6.7 versus 4.8mM) levels compared to control mice. Hepatic cholesterol (0.16 versus 0.018 mg/mg protein) and triglyceride (0.69 versus 0.1 mg/mg protein) content was also higher in the HMLF mice (Werstuck et al., 2001).

In summary, these data suggest that a *CBS* loss-of-function variant could lead to the increased triglyceride and/or cholesterol levels of FCHL.

v) *SIK1*

SIK1 was first identified due to its increased expression in the adrenal glands of rats fed a high-salt diet (Wang et al., 1999). It represents a candidate for the FCHL cholesterol trait due to its role in regulating steroidogenic gene expression and thus cholesterol utilisation. *In vitro*, treatment of Y1 mouse adrenocortical tumour cells with adrenocorticotropic hormone (ACTH) led to a rapid increase in the mRNA/protein levels of *Sik1* at 1-2 hours before a return to initial levels after 8 hours. Meanwhile, mRNA levels for the steroidogenic genes cholesterol side-chain cleavage cytochrome P450 (*Cyp11a1)* and steroidogenic acute regulatory (*StAR*), rose

as *SIK1* expression fell (Lin et al., 2001). In Y1 cells overexpressing *SIK1*, the ACTH-dependent expression of steroidogenic genes *Cyp11a1*and *StAR* was repressed (Lin et al., 2001). Thus, a *SIK1* gain-of-function variant could underlie the FCHL cholesterol trait by decreasing cholesterol utilisation.

SIK1 is also implicated in *de novo* lipogenesis. Overexpression of *SIK*, delivered by adenovirus, led to reduced mRNA levels of lipogenic genes including fatty acid synthase in extracts from rat livers as well as cultured hepatocytes. *SIK1* overexpression also led to reduced hepatic triglyceride levels, *in vivo*. These effects are thought to be modulated via phosphorylation of SREBP-1c at serine 329 and/or 265/266, since when co-expressed with a mutant form of SREBP-1c, SIK1-reduction of fatty acid synthase levels was ameliorated (Yoon et al., 2009).

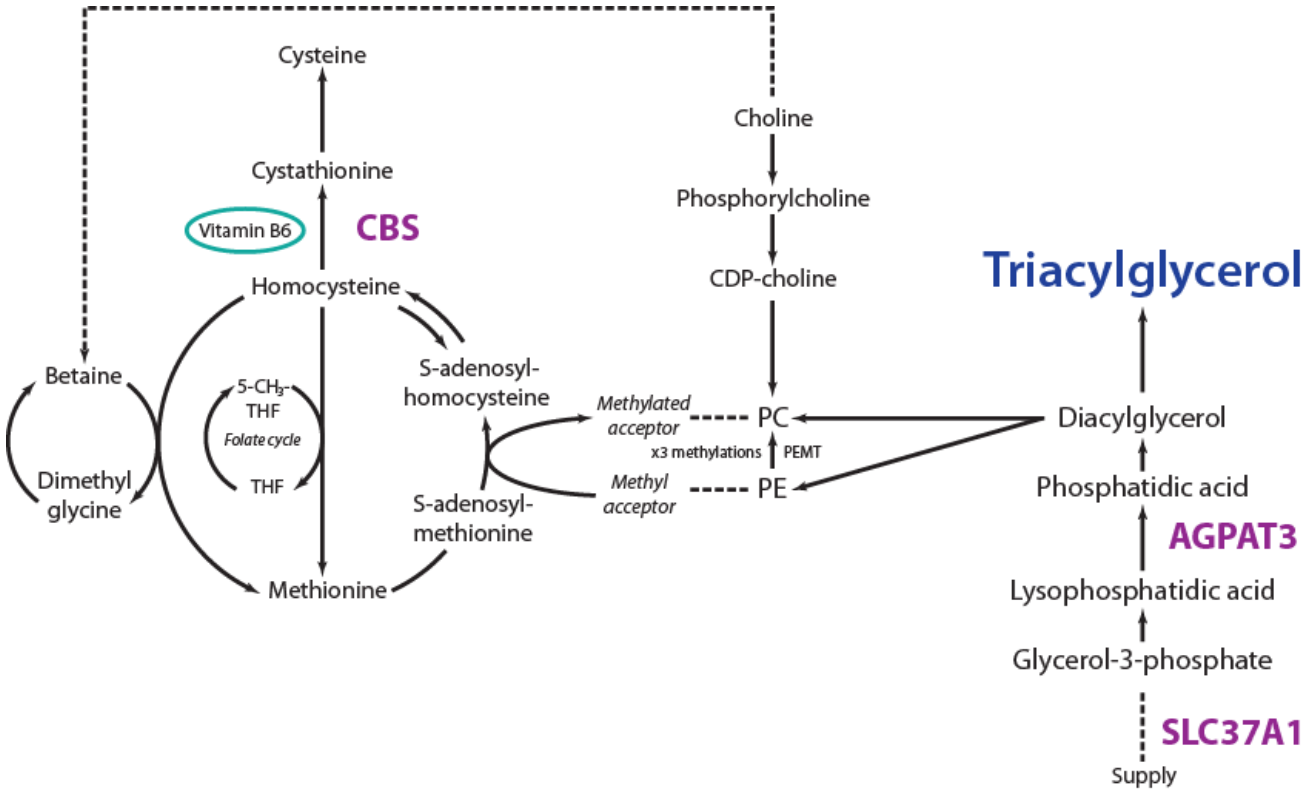


Figure 4.1) Connections between *CBS*, *SLC37A1* and *AGPAT3* and Triglyceride Metabolism

CBS (with vitamin B6 as an essential cofactor) catalyses the irreversible condensation of homocysteine and serine, thus generating cystathionine. This reaction is a branch of the methionine cycle which can also be linked to triacylglycerol and phospholipid metabolism through the methyl-donor role of S-adenosyl-methionine. Triacylgylcerol synthesis is a multi-step pathway, the first step of which involves

acylation of glycerol-3-phohphate by glycerol-3-phosphate acyltransferase to produce lysophosphatidic acid. In the next reaction, acylation of lysophosphatidic acid to produce phosphatidic acid is catalysed by a 1-acylglycerol-3-phosphate acyltransferase enzyme, such as AGPAT3. These reactions can occur in the endoplasmic reticulum or mitochondria and thus SLC37A1, a predicted glycerol-3-phosphate membrane transporter influence substrate availability for triacylgylcerol synthesis.

CDP-choline=cytidine diphosphocholine, THF=tetrahydrofolate, PEMT= phosphatidylethanolamine N-methyltransferase

vi) *AGPAT3*

The formation of triacylglycerol from the glycerol-3-phosphate backbone involves successive acylation by glycerol-3-phosphate transferases (GPATs), 1-acyl-glycerol-3-phosphate-O-acyltransferases (AGPATs) and diacylglycerol acyltransferases (1.2.4). Thus AGPATs catalyse the acylation of lysophosphatidic acid to form phosphatidic acid (Figures 1.3 and 4.1). There are at least 11 human enzymes with designated as AGPATs (Agarwal and Garg, 2010). Quantitative real-time PCR showed that *AGPAT3* has a ubiquitous expression pattern in human tissues though is found at somewhat higher levels in testis, pancreas, kidney, lung and liver (in descending order, expression in adipose moderate). *In vitro* studies showed that AGPAT3 has the expected enzymatic activity, converting lysophosphatidic acid to phosphatidic acid (Prasad et al., 2011).

vii) *COL18A1*

COL18A1 is a HSPG that exists as three different isoforms, expressed in the basement membrane of a variety of tissues including liver, kidney, adipose, retina and brain (Saarela et al., 1998; Suzuki et al., 2002). The short, medium and long protein isoforms are named according to the length of the non-collagenous domains by which they differ (NC11-303, NC11-493 and NC11-728). The NC11-728 variant contains a cysteine-rich domain, homologous to the ligand-binding domain of Frizzled proteins. Wnt/Frizzled signalling is implicated in adipogenesis ((Ross et al., 2000) and reviewed in (Christodoulides et al., 2009)). Since COL18A1 is highly expressed during adipocyte differentiation, it is believed that the NC11-728 form of COL18A1 may play an important role in this process (Errera et al., 2008; Inoue-Murayama et al., 2000). Of particular relevance to FCHL, Meex and colleagues reported that pre-adipocyte cell lines prepared from FCHL patients contain reduced levels of *COL18A1* mRNA relative to control cells (Meex et al., 2005).

Mutations of *COL18A1* cause Knobloch syndrome, a rare recessive disorder characterised by ocular defects (Sertie et al., 2000). In affected members of a large consanguineous Knobloch syndrome family, a homozygous mutation of the *COL18A1* (NC11-303 isoform) intron 1 splice acceptor sequence was identified (Sertie et al., 2000). This mutation was predicted to result in a truncated (at exon 4) protein product and in affected family members fasting triglyceride levels were higher compared to those carrying no *COL18A1* mutation (3.01±-0.58mmol/l versus 1.32±0.14mmol/l, n=7, *P*=0.02) (Bishop et al., 2010).

viii) and ix) *LSS* and *UBE2G2*

LSS, an enzyme of the cholesterol biosynthetic pathway, catalyses reactions that produce either lanosterol or oxysterols from a 2,3(*S*)-monooxidosqualene precursor (Figure 4.2). These reaction products are involved in regulating the expression of other cholesterol genes through both transcriptional and post-translational means (Figure 4.2).

UBE2G2 is an E2 ubiquitin conjugating enzyme that facilitates HMGCR degradation upon cellular accumulation of lanosterol (Figure 4.2). In humans, *UBE2G2* mRNA is expressed ubiquitously though highest levels are found in skeletal muscle (Katsanis and Fisher, 1998).

x) *MCM3AP*

In an expression profiling study where HeLa cells were grown in lipoprotein-depleted serum and 2-hydroxypropyl-β-cyclodextrin, in order to decrease cellular cholesterol levels and activate the cholesterol homeostatic system (Bartz et al., 2009), expression of *MCM3AP*, which encodes minichromosome maintenance complex component 3 associated protein, was upregulated. Specifically, six hours after treatment, *MCM3AP* mRNA levels were increased 2.36-fold. The authors did not investigate this gene further. However, it seems pertinent that the observed fold-change was similar to that recorded for the cholesterol biosynthesis gene *DHCR7* (at the 3-hour time point).

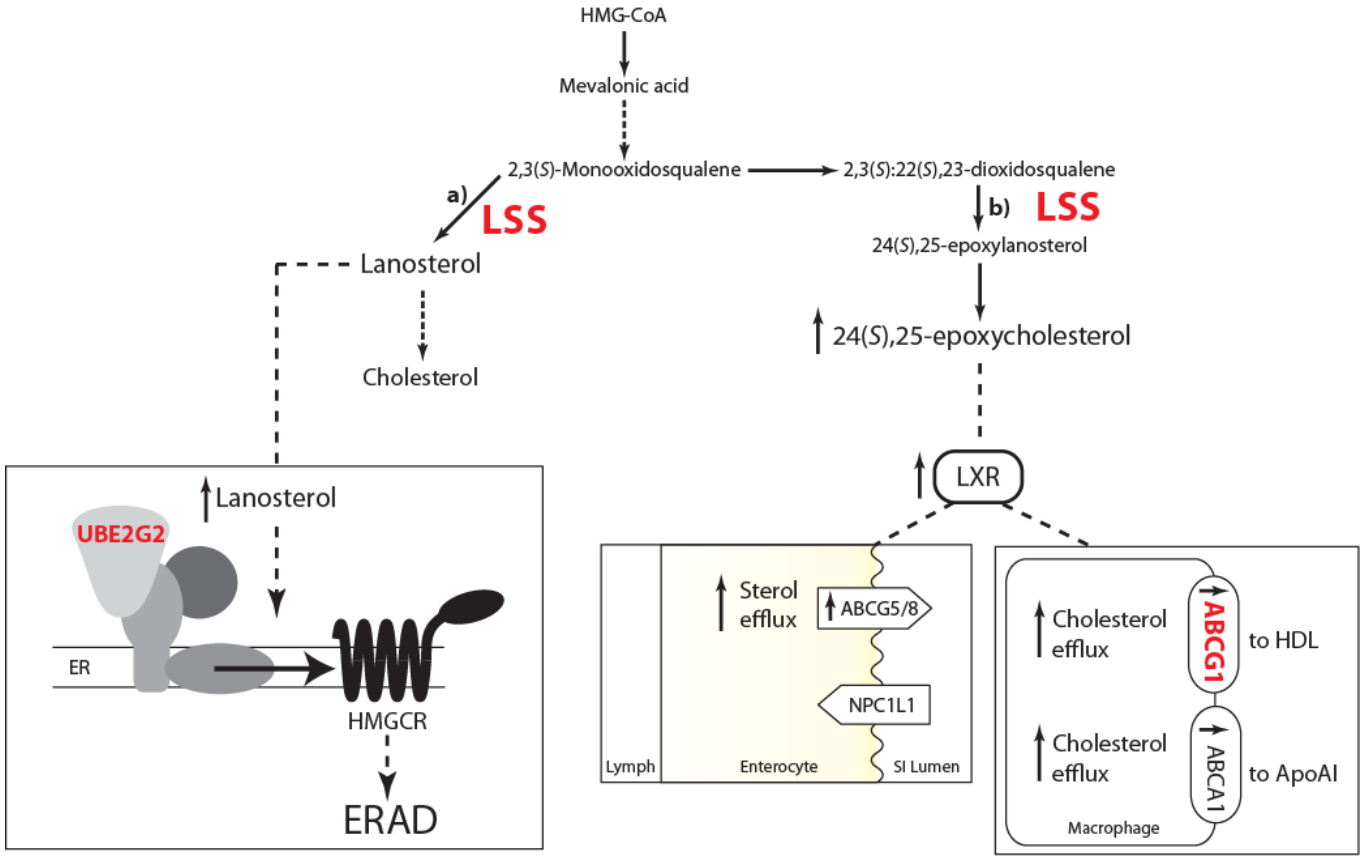


Figure 4.2) Regulation of *De novo* Cholesterol Biosynthesis by Lanosterol and UBE2G2

LSS catalyses; a) the cyclisation of 2,3(*S*)-monooxidosqualene to lanosterol or, under conditions of partial enzyme inhibition; b) the conversion of 2,3(*S*):22(*S*),23-dioxidosqualene to 24(*S*),25-epoxylanosterol and ultimately 24(*S*),25-epoxycholesterol (i.e. oxysterols), (Boutaud et al.,

1992). Lanosterol regulates cholesterol synthesis through a feedback mechanism involving endoplasmic reticulum associated protein degradation (ERAD) of HMGCR. Excess lanosterol within the cell facilitates binding of an Insig-1/Gp78/Ube2g2 complex to HMGCR. Gp78 is a membrane-bound ubiquitin E3 ligase that catalyses the addition of ubiquitin to HMGCR at Lys89 and Lys248 (Miao et al., 2010; Song et al., 2005). UBE2G2 is an E2 ubiquitin conjugating enzyme, upon which polyubiquitin chains are formed before being transferred to HMGCR by Gp78 (Li et al., 2007). Other proteins, including ubiquitin fusion degradation 1(Ufd1) protein act as cofactors to transfer the ubiquitinated HMGCR to the proteasome (Cao et al., 2007). When LSS action is shifted to oxysterol synthesis via an exogenous inhibitor such as RO01704565, selective upregulation of LXR-regulated genes involved in cholesterol efflux occurs. For example, in an animal model using miniature pigs, inhibitor treatment also led to increased hepatic (both 1.3-fold) and intestinal (1.4/1.6-fold) *ABCG5/8* mRNA levels (P<0.04) concomitant with a 37% decrease in plasma plant sterol concentrations (P<0.05) (Telford et al., 2005). Similarly, in a human macrophage cell line inhibitor treatment increased expression of the efflux-related *ABCA1*, *ABCG1* and *APOA1* by 1.9, 2.3 and 1.6-fold respectively (Beyea et al., 2007).

4.2.2) Study Design

In order to efficiently capture the haplotype diversity across the ten biological candidate genes (4.2.1) and their flanking regions (+/-50kb unless indicated in Table 4.1), SNPs were selected using the process illustrated in Figure 4.3. Figure 4.4 and Table 4.1 show the localisation of the selected SNPs. Additionally, six SNPs were genotyped on the basis of genetic data only (Table 4.1).

In second stage analysis, SNPs encompassing *TSPEAR*, a gene located immediately downstream of *UBE2G2*, were included (rs235289, Chapter 5.2). Thus, in total, 589 SNPs were genotyped using either Illumina GoldenGate (198 SNPs) or Sequenom iPLEX assays (391 SNPs), in 1739 FCHL family cohort members, drawn from 241 families (Table 4.2).

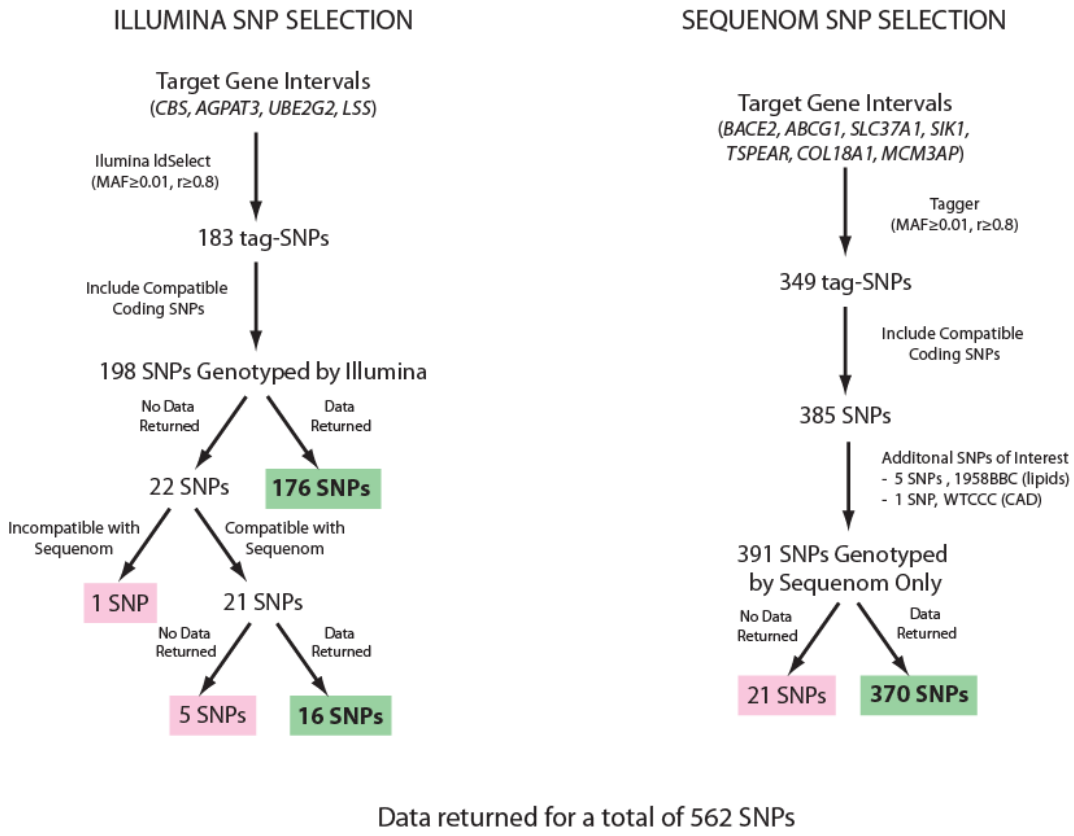


Figure 4.3) SNP Selection and Genotyping Processes Employed for Chromosome 21 Genes

A total of 589 SNPs were selected for genotyping using either Illumina Golden Gate or Sequenom MassARRAY platforms. Data were returned for 562 SNPs (shaded in green). Pink shading indicates SNPs where robust data was not returned.

Figure 4.4

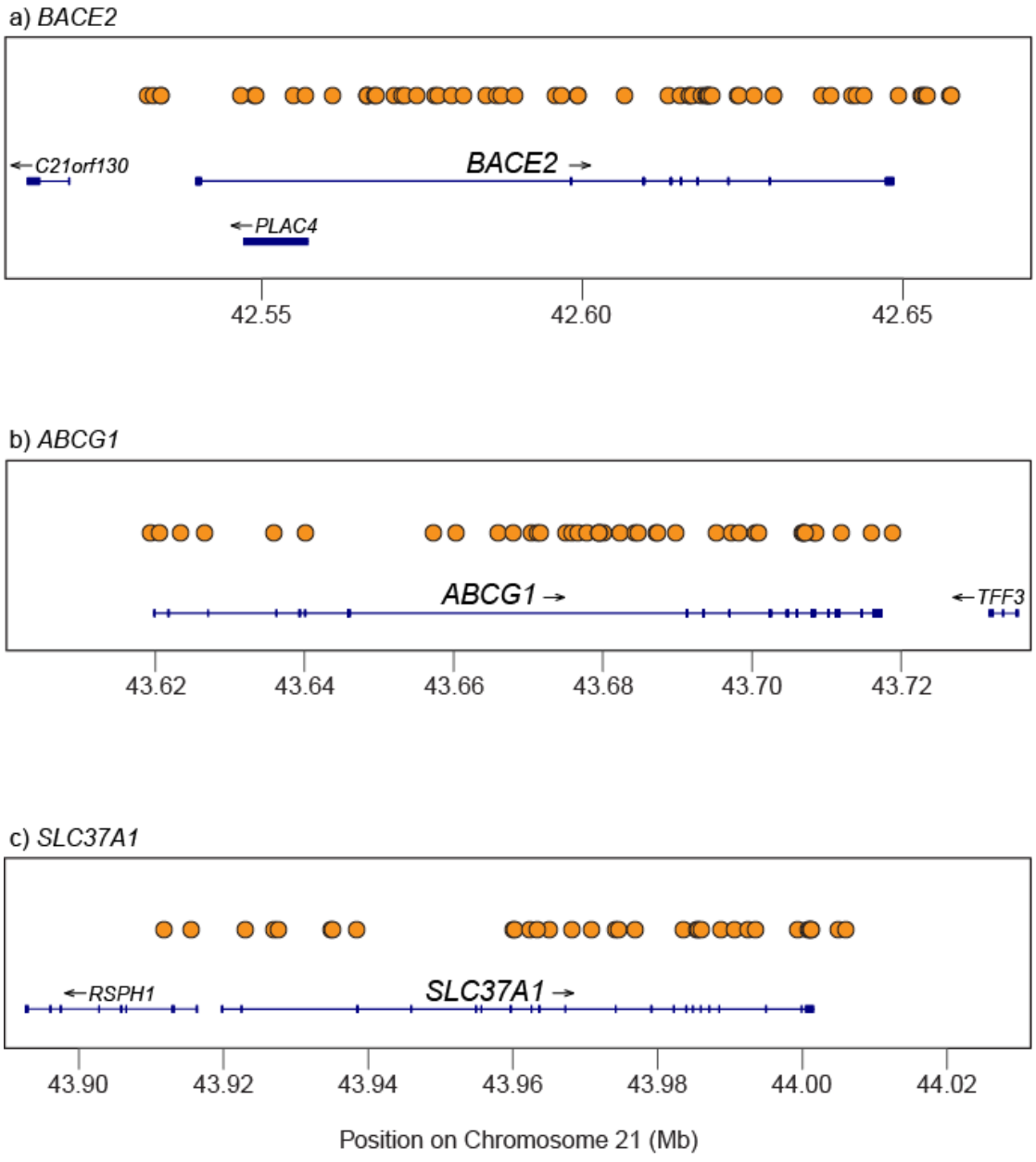
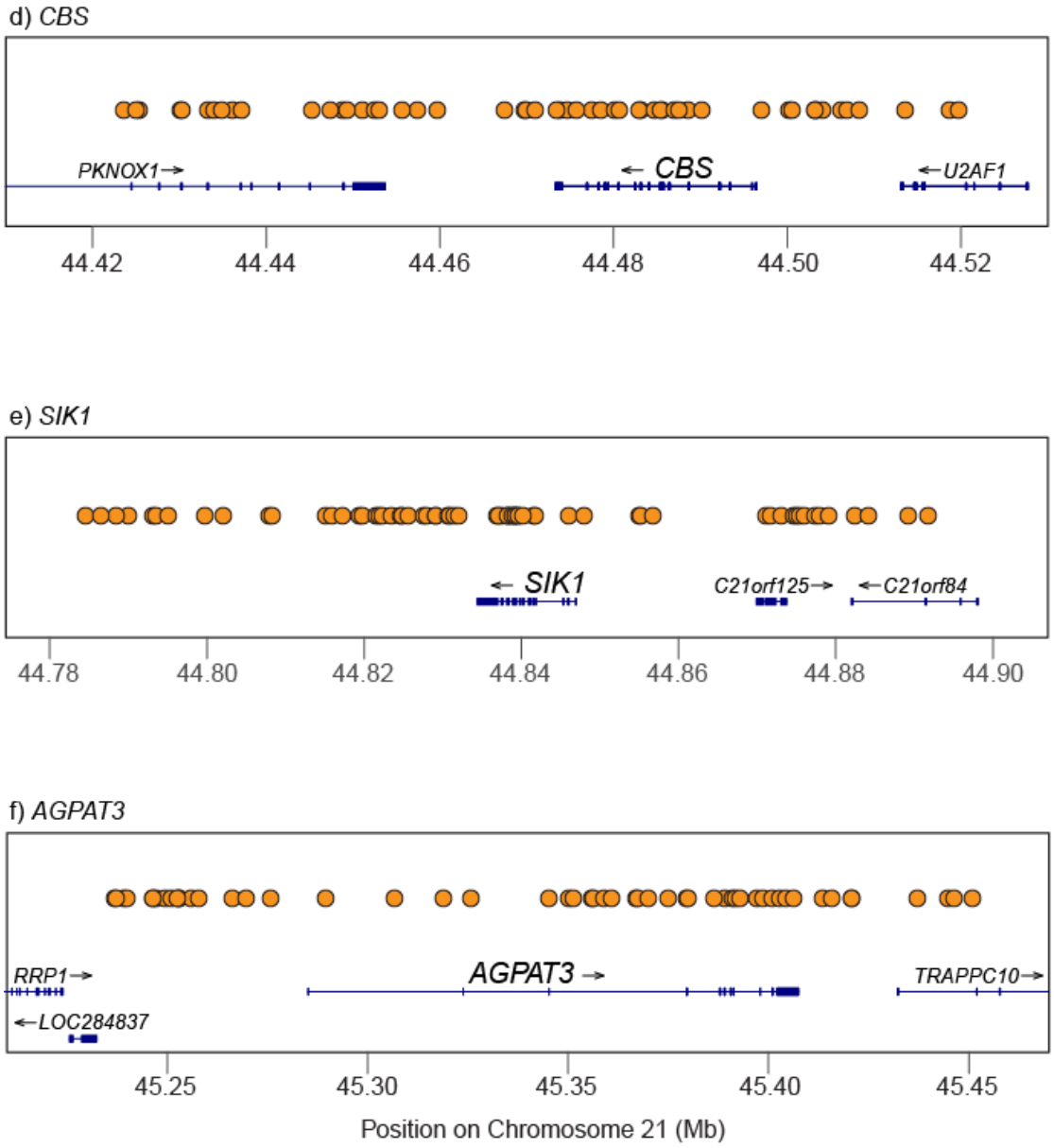

a) BACE2
C21orf130
BACE2
PLAC4
42.55
42.60
42.65
b) ABCG1
ABCG1
TFF3
43.62
43.64
43.66
43.68
43.70
43.72
c) SLC37A1
RSPH1
SLC37A1
43.90
43.92
43.94
43.96
43.98
44.00
44.02
Position on Chromosome 21 (Mb)

Figure 4.4 continued



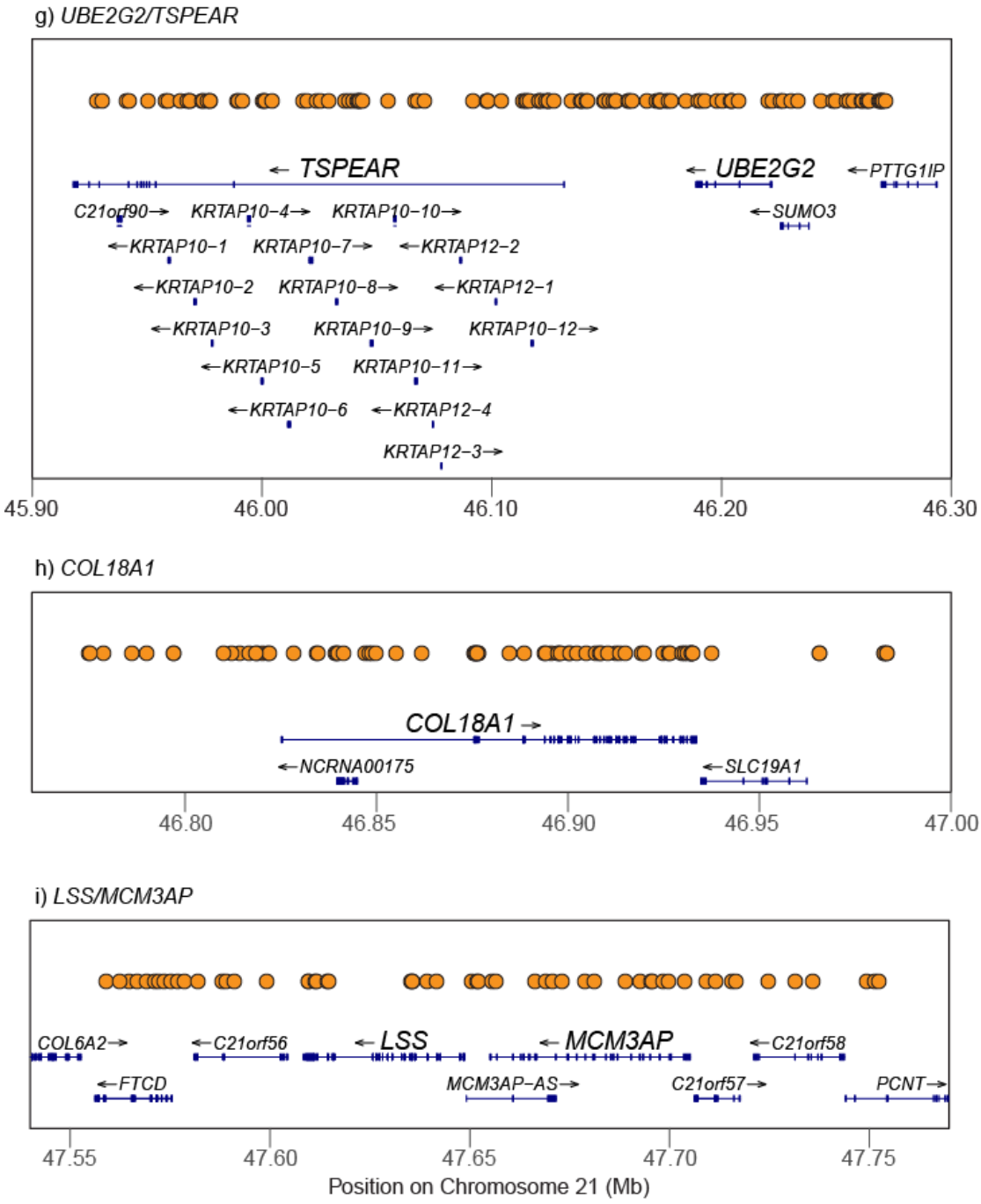


Figure 4.4) SNPs Selected to Capture Haplotype Diversity of Target Genes
Biological candidate genes (the primary targets of tag-SNPs), plus *TSPEAR* are indicated in large font. The span of each gene is shown by a horizontal blue bar whilst, exons are designated by vertical blue bars. SNPs selected for genotyping are marked by gold circles.

Table 4.1) Description of Genotyped SNPs

			Target Interval Span (Kb)		SNPs Selected		
Primary Target[a]	TSS/Position[b]	Strand	Upstream	Downstream	(Coding)[c]	Secondary Target(s)[d]	Comment
BACE2	42,539,727	+	10	10	71 (0)	*PLAC4*	rs914186 associated with triglyceride/HDL[e]
ABCG1	43,619,798	+	10	10	48 (2)	None	
SLC37A1	43,919,741	+	20	20	36 (2)	*RSPH1* (p)	
CBS	44,473,300	-	50	50	55 (5)	*PKNOX1* (p), *U2AF1* (p)	
SIK1	44,834,397	-	50	50	70 (12)	*C21ORF125*, *C21ORF84* (p)	
AGPAT3	45,285,115	+	50	50	54 (2)	*TRAPPC10* (p)	
UBE2G2	46,188,494	-	50	250	115 (0)	Multiple *KRTAP* genes, *TSPEAR* (p)	Downstream span expanded into *TSPEAR* due to initial association results
COL18A1	46,825,096	+	50	50	79 (19)	*NCRNA00175*, *SLC19A1*	
LSS/MCM3AP	47,608,359	-	100	50	55 (9)	*FTCD* (p), *C21ORF56*, *MCM3AP-AS1*, *C21ORF57*, *C21ORF58*, *PCNT* (p)	Intervals surrounding *LSS* and *MCM3AP* (+/- 50kb) overlap

rs2142090 (*DSCAM)*	41,370,504	NA	NA	NA	1	Associated with TC level in B1958BC[f] (*P* =0.01)
rs2837631	41,818,825	NA	NA	NA	1	Associated with TC level in B1958BC (*P* =0.004)
rs2837906 (*FAM3B)*	42,410,744	NA	NA	NA	1	Associated with TC level in B1958BC (*P* =0.008)
rs2158354	42,710,175	NA	NA	NA	1	Associated with TC level in B1958BC (*P* =8x10-5)
rs2155722 (*RRP1B)*	45,096,206	NA	NA	NA	1	Associated with TC level in B1958BC (*P* =0.005)
rs2838756	46,454,391	NA	NA	NA	1	Associated with CHD in WTCCC[g] (*P*=2x10^{-5})

TSS=transcriptional start site, B1958BC=British 1958 Birth Cohort, CHD=coronary heart disease, WTCCC=Wellcome Trust Case Control Consortium

[a] Primary target is either a biological candidate gene or gene/SNP identified in an alternative association study

[b] Position of most upstream TSS for genes with multiple transcripts. For single SNPs, position is listed.

[c] Total number of SNPs selected across target SNP selection interval, including tag- and coding-SNPs.

[d] Other genes either fully or partly (p) encompassed by selected tag-SNPs

[e] (Saxena et al., 2007)

[f] Unpublished WTCCC data (www.b58cgene.sgul.ac.uk/index.php)

[g] (Wellcome Trust Case Control Consortium, 2007)

Table 4.2) FCHL Family Cohort Samples Utilised for SNP Genotyping

					TG Trait[d]			TC Trait[f]		
	All Subjects	Probands	Spouses	*P*[c]	Affected	Unaffected	*P*[e]	Affected	Unaffected	*P*[g]
N (Male/Female)	866/873	173/68	86/160		380/273	452/543		415/254	428/573	
Age (years)	44.24±16.35	49.95±10.04	55.51±8.34	***	46.04±15.46	42.69±16.70	***	47.05±14.56	42.36±17.11	***
BMI (kg/m^2)	25.54±4.09	26.69±2.33	25.86±4.18	***	26.27±3.15	24.50±3.86	***	26.07±2.95	24.62±4.04	***
Serum TG (mmol/l)[a]	1.74	4.20	1.34		3.50	1.21		3.29	1.27	
Log-transformed TG[b]	0.62±0.70	1.53±0.46	0.30±0.49	***	1.30±0.50	0.16±0.40	***	1.13±0.65	0.27±0.52	***
Serum TC (mmol/l)	6.37±1.62	8.26±1.30	5.96±1.12	***	7.40±1.62	5.66±1.22	***	7.86±1.24	5.37±0.97	***

TG=triglyceride, TC= total cholesterol, BMI=body mass index

Age, BMI, log-transformed triglyceride level and cholesterol level are expressed as mean±standard deviation

[a] Median values

[b] Raw triglyceride levels were transformed (using log to the base e) in order to normalise the quantitative trait distribution

[c, e, g] Trait means were compared between probands/spouses and affecteds/unaffecteds using a student's *t*-test, ****P*<0.001

[d] TG affected=triglyceride level ≥95th percentile age- and sex-specific value

[f] TC affected=cholesterol level ≥95th percentile age- and sex-specific value

4.2.3) Sample Quality Control

As part of their genotyping service, Illumina carried out sex verification of the DNA samples. One potential sex error was identified, in a FCHL proband. Since an alternative proband could not be assigned in this family, a total of six individuals were removed from the dataset. PedCheck revealed evidence of non-Mendelian inheritance in three families, requiring the removal of eight further individuals from the dataset. Thus, 14 of the 1739 genotyped samples were excluded from the dataset following stringent quality control checks (Table 4.3).

With the Illumina dataset, all samples had a SNP genotype call rate of >0.93. Only two samples had a call rate of <0.97, indicating that all DNA samples were of good quality and as a result the same DNA sample plates were utilised for Sequenom genotyping. Mean heterozygosity across the DNA samples was 0.30±0.04 and the heterozygosity of all samples fell within 3.2 standard deviations of the mean level.

In total, after sample-specific quality checks were performed, genotypes from 1725 FCHL cohort members, from 239 families, were available for association analyses.

Table 4.3) Quality Control of Samples Prior to Association Analysis

Samples		
Remaining (N)	Excluded (N)	Reason for Exclusion
1739		
1733	6	Gender error
1725	8	Inheritance problem in family
1725	14	Total

4.2.4) Marker Quality Control

Table 4.4 shows that 87 SNPs were excluded from the dataset after quality control checks were performed.

In total, 27 SNPs failed to produce any genotype data (Figure 4.3, Table 4.4). A single marker (rs17004554) showed significant evidence of Hardy–Weinberg disequilibrium (*P*=0.0003). This marker was therefore excluded from subsequent association analysis. Four markers (rs8133035, rs8126811, rs162362 and rs162386)

displayed persistent evidence of non-Mendelian inheritance (NMI) across a number of FCHL cohort families and were also excluded from association analyses. Since recurrent non-inheritance or deviation from HWE can be attributed to potential underlying copy-number variation or poorly performing genotyping assays, these SNPs were examined more closely.

Due to its potential functional importance, the first SNP investigated was rs8133035, a putative *COL18A1* non-synonymous coding SNP. This putative SNP resides in exon 1 of the *COL18A1* long isoform (NC-728) transcript and would lead to an arginine to glutamine amino acid change at residue 427 in the protein sequence. By Sequenom genotyping the putative SNP was deemed to be heterozygous in 30 individuals from 25 families. However, due to pedigree structure, Mendelian errors were only detectable in eight offspring from seven of these families whose parents were both homozygous for the ancestral G allele. By standard Sanger DNA sequencing, the eight 'heterozygous' offspring were in fact determined to be homozygous for the G allele (Figure 4.5). Moreover, this variant is, to date, not recorded in the 1000 Genomes dataset (URL9). Thus it was judged that the cause of the apparent NMI in this instance was the poor performance of the rs8133035 genotyping assay.

Table 4.4) Quality Control of Markers Prior to Association Analysis

Markers		
Remaining (N)	Excluded (N)	Reason for Exclusion
589		
562	27	Missingness rate=100%
561	1	Hardy-Weinberg *P*-value < $1x10^{-3}$
557	4	Mendelian Inheritance Errors
544	13	Missingness Rate >20%
502	42	Minor Allele Frequency <0.01
502	87	Total

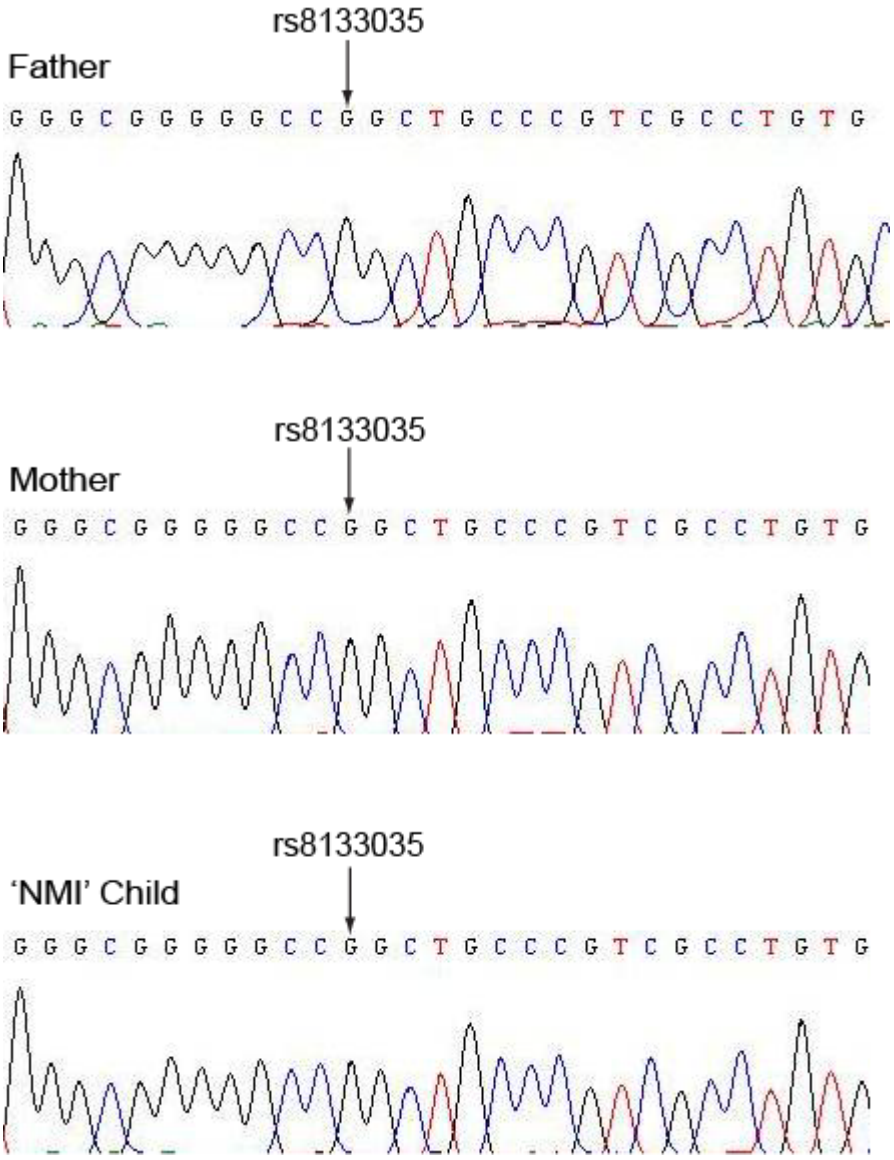


Figure 4.5) Example sequencing chromatograms from FCHL Cohort (Family 42) for the putative SNP rs8133035

Sequencing data provide no evidence for the presence of the rs8133035 minor allele in seven FCHL families where, by Sequenom genotyping, offspring were called heterozygous at this locus, despite both parents carrying two copies of the major G allele.

Three of the problem SNPs, rs162362 (NMI in six families), rs162386 (NMI in 15 families) and rs17004554 (HWE) reside in a ~21kb region upstream of *SIK1* that encompasses *C21orf125* (Figure 4.6). There was no overlap between the rs162362 and rs162386 NMI families: apparent Mendelian errors were detected in a total of 21 families. The interval also contains two SNPs (rs229353 and rs7281094) which failed to genotype totally (Figure 4.6). However, data was returned for seven more interspersed SNPs with no apparent discrepancies.

Figure 4.7 illustrates how the segregation of deletions within families can lead to apparent NMI. Closer inspection of the rs162362 NMI families reveals that for two of the six families (family 139 and 238) the observed genotypes cannot be explained by transmission of a deleted allele from one parent, whilst in the other four families (families 40, 87, 150 and 184), transmission of a deleted allele would explain the apparent NMI (Figure 4.8). It is therefore possible that a small deletion exists in these four families. Although it is worth noting that, on the basis of genotype data from the neighbouring markers, any such deletion would have to be carried upon at least three different haplotypes ('G-C-deletion-G-G' (2 families), 'A-C-deletion-A-T' and 'A-C-deletion-G-G').

In comparison, all 15 rs162386 NMI families followed the 'hemizygous-transmission' pattern (Figure 4.9) thus indicating that a distinct deletion may be present at this locus in some FCHL families. All potentially hemizygous parent-offspring pairs were heterozygous at the two neighbouring SNPs, rs17004556 and rs2838307. Based upon the positions of the primers used in genotyping these SNPs, the maximum possible size of a deletion in this interval, in these 15 families, is 929 base-pairs. This is interval is ~1.5kb downstream of *C21orf125* which is predicted to encode a long intergenic non-coding RNA.

Examination of genotype data for the surrounding SNPs in the fifteen rs162386 NMI families reveals that the potential deletion is always carried upon a genetic background of 'A-C-deletion-A-A'. A further 22 FCHL cohort families carry this haplotype. However, in these instances NMI was not detectable either due to non-transmission of the potentially deleted allele, or because the genotype of the non-deleted allele is the same as that transmitted by the other parent.

Of the 37 FCHL cohort families that might carry the putative ≤929bp deletion around rs162386, 15 are included in the family subset contributing to the maximal triglyceride trait OSA linkage signal at 51cM (total 102 families, 3.2.2). In comparison, just five potential deletion-carrying families form part of the subset of 47 families contributing to the 51cM cholesterol linkage peak, whilst 17 potential deletion-carrying families for part of the 101-family subset contributing to the cholesterol trait linkage peak at 57.8cM.

Returning to the final NMI SNP, rs8126811, it appears that this SNP (genotyped as a tag-SNP for the *TSPEAR* locus) resides in a long interspersed repeat element, which may therefore explain its aberrant genotyping.

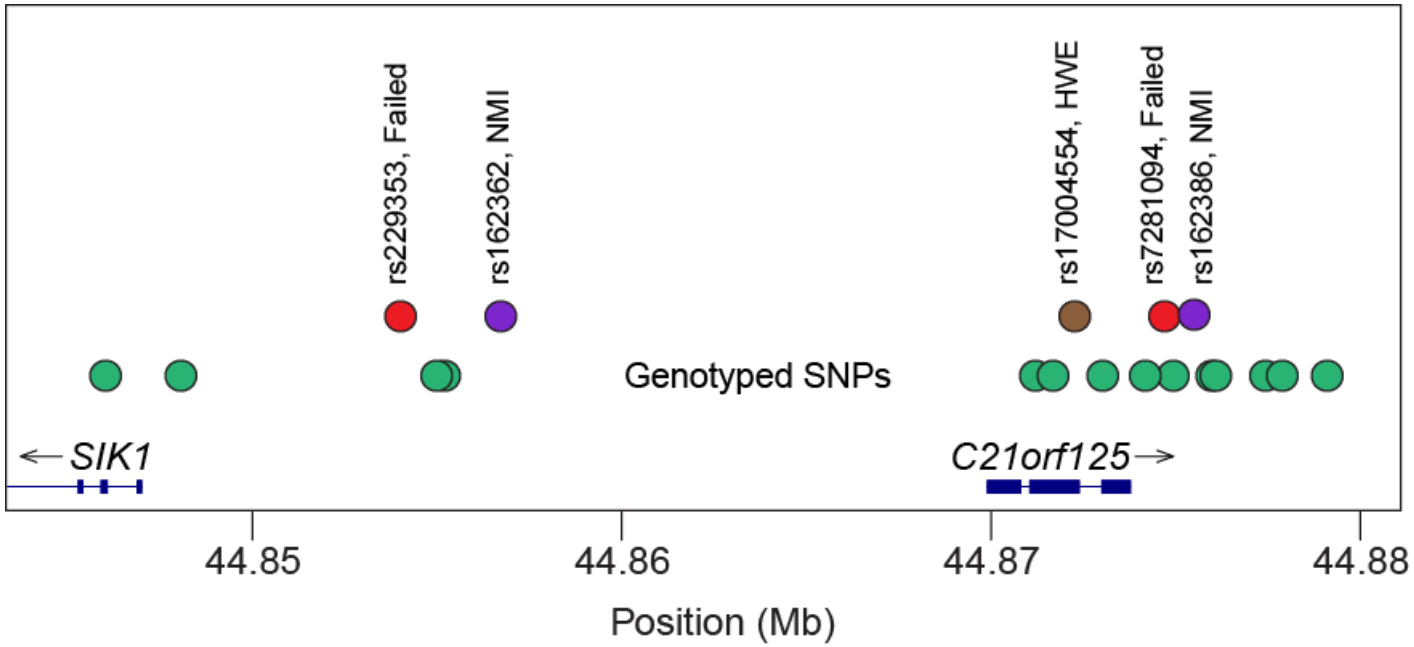


Figure 4.6) Potential CNV Locus Upstream of *SIK1*

A cluster of SNPs with genotyping problems were identified upstream of *SIK1*. Genotyped SNPs are marked by dots. Succesfully genotyped SNPs are shaded green. SNPs which failed to produce any genotype data are shaded red, those showing persistent inheritance problems are purple whilst a single SNP showing significant evidence of deviation from HWE is coloured brown.

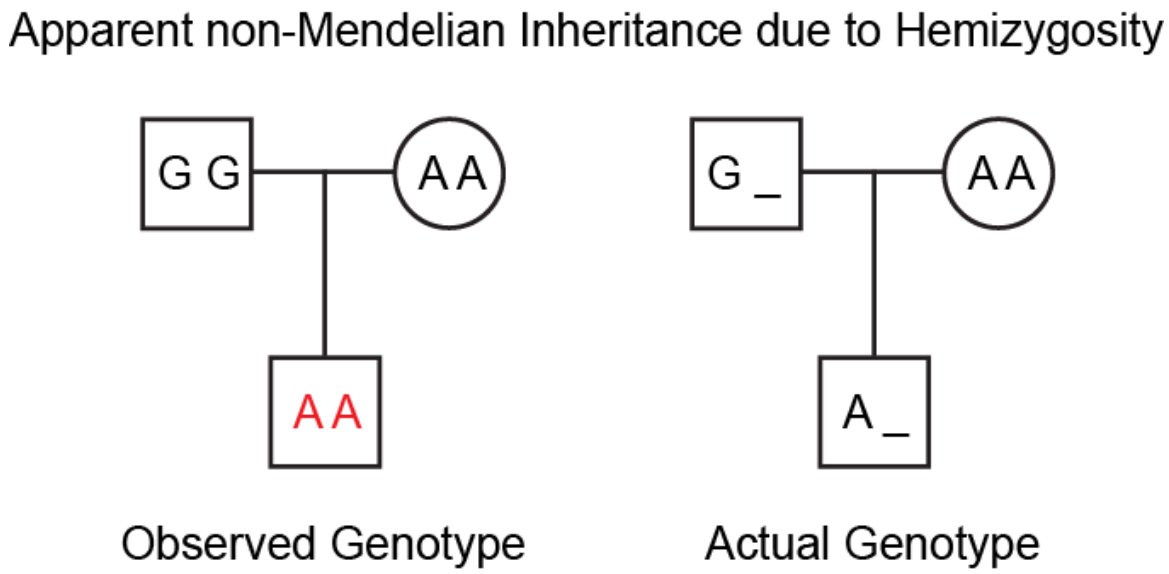


Figure 4.7) Transmission of a Deletion from Parent to Offspring May Appear as Non-Mendelian Inheritance

If a hemizygous parent transmits their deleted allele, the resulting hemizygous child will appear to be homozygous for the allele transmitted by the other parent. If this allele is different to the non-transmitted allele in the hemizygous parent then NMI will be observed.

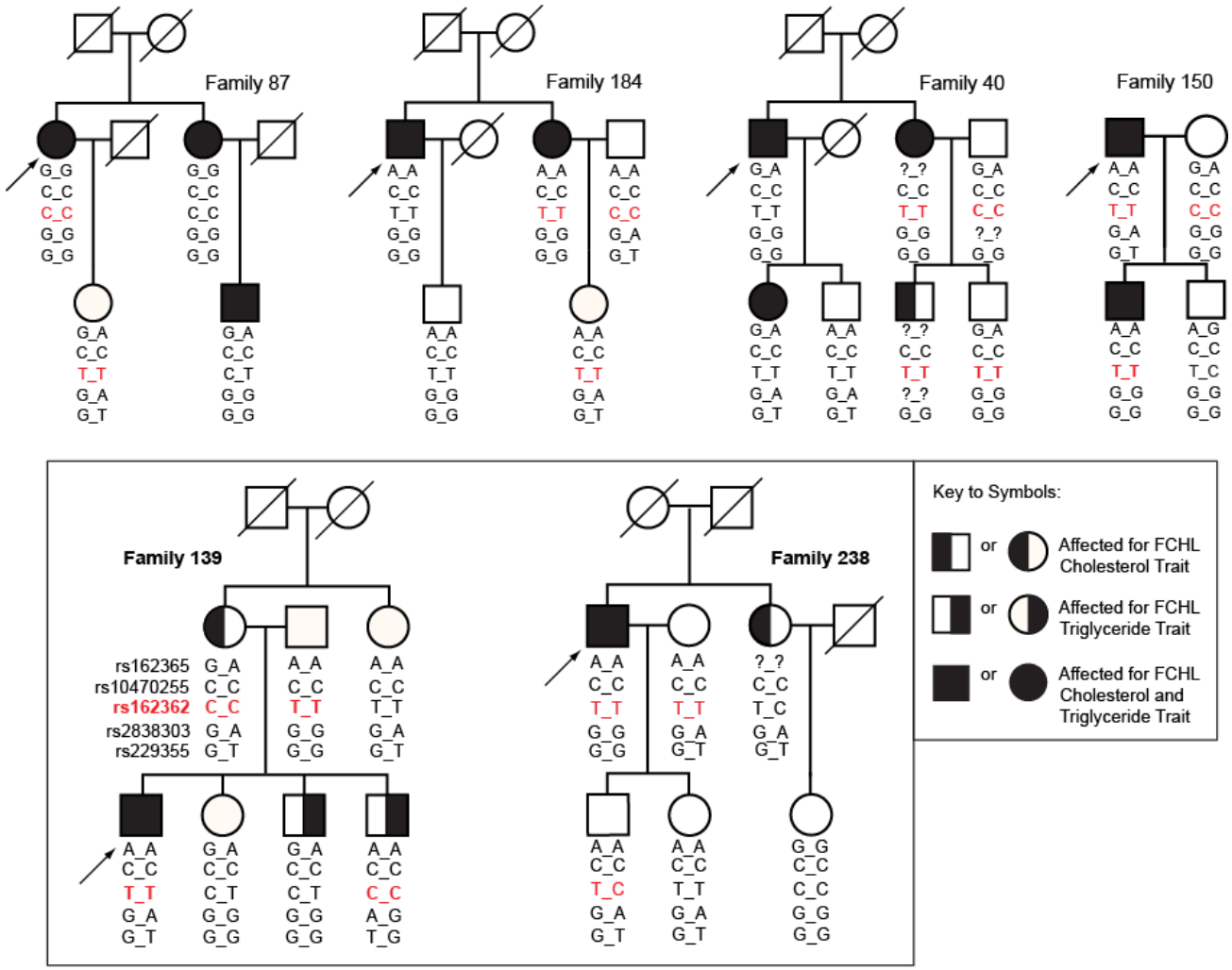
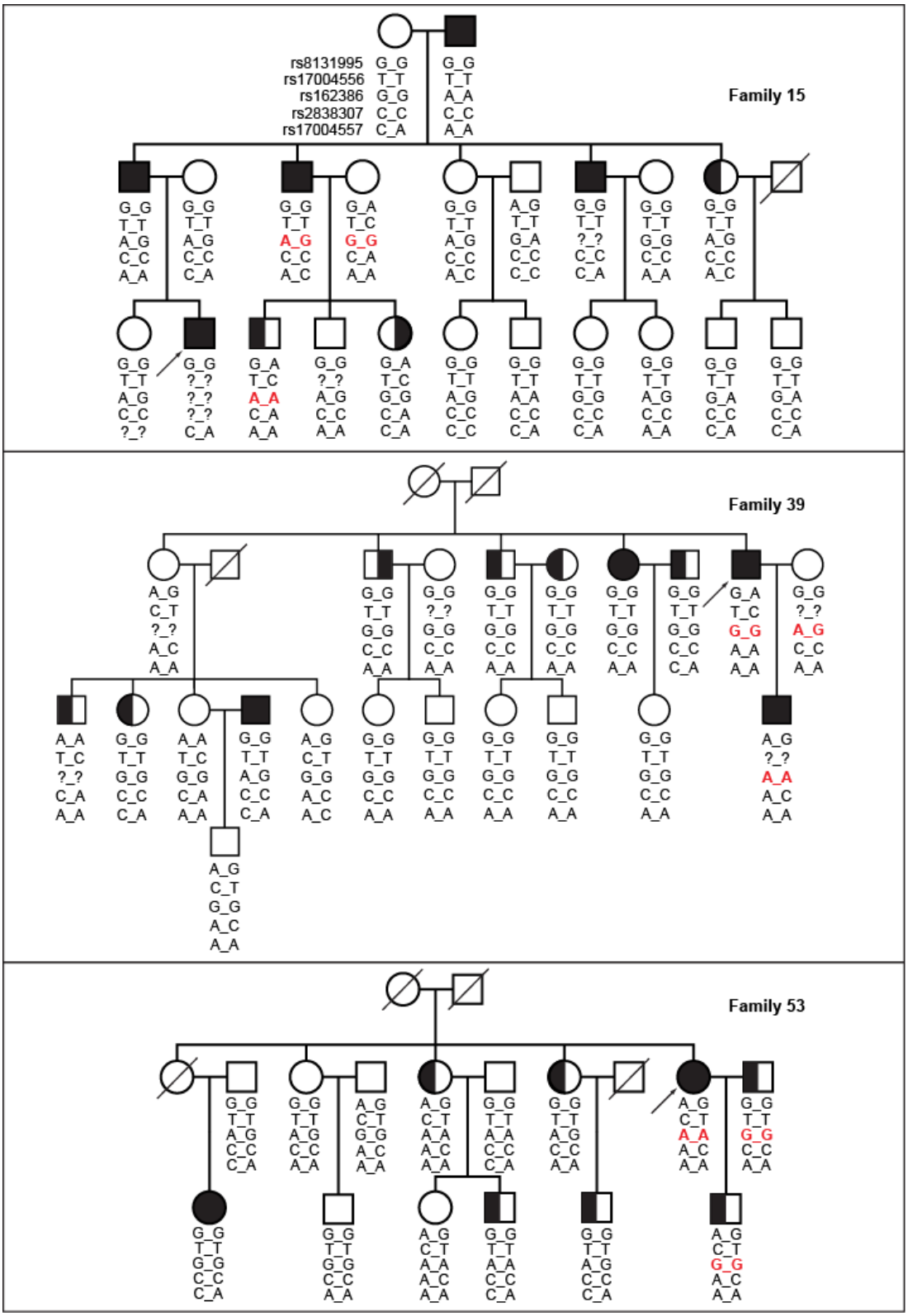
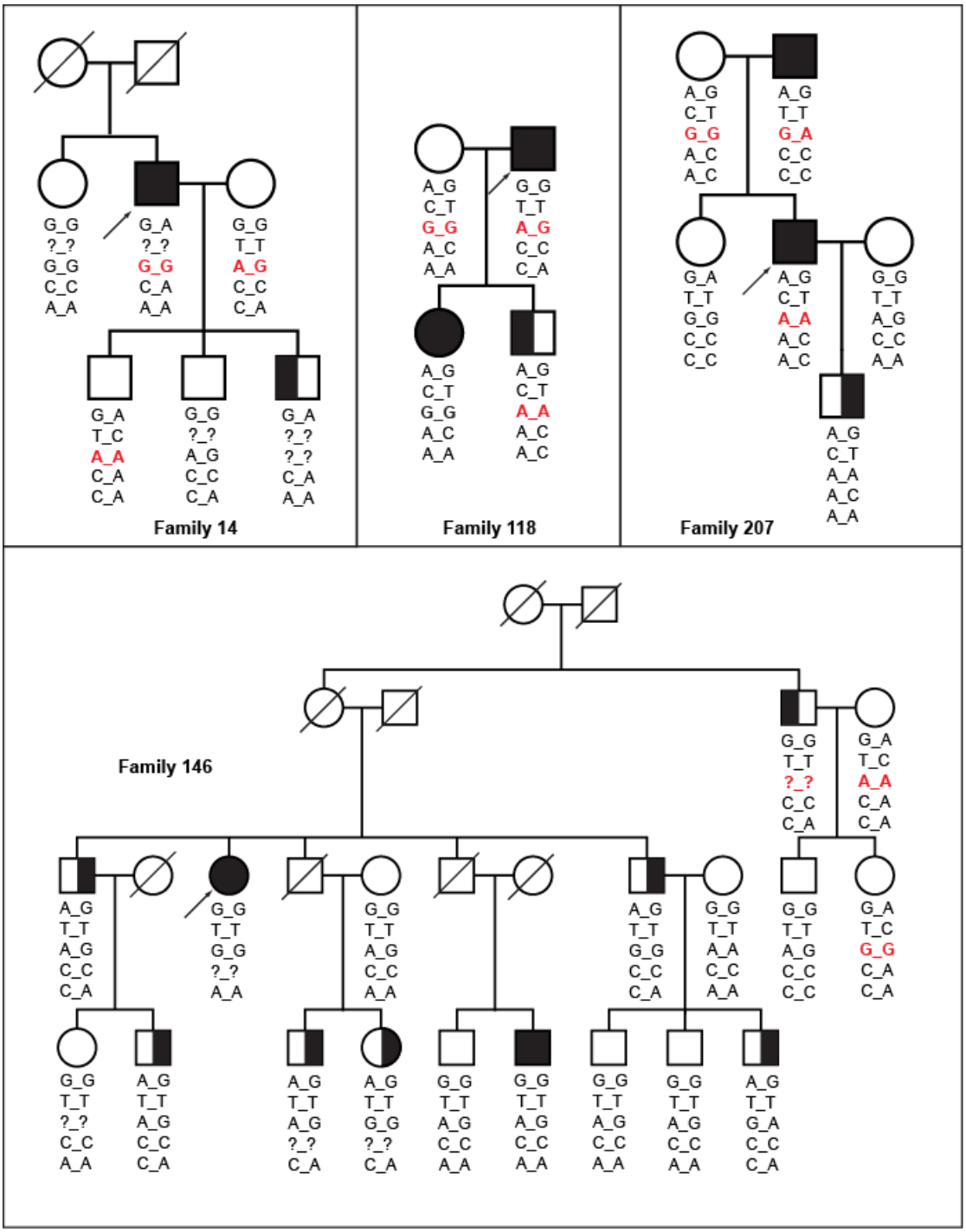

Family 87
Family 184
Family 40
Family 150
Family 139
Family 238
rs162365
rs10470255
rs162362
rs2838303
rs229355
Key to Symbols:
or
Affected for FCHL Cholesterol Trait
or
Affected for FCHL Triglyceride Trait
or
Affected for FCHL Cholesterol and Triglyceride Trait

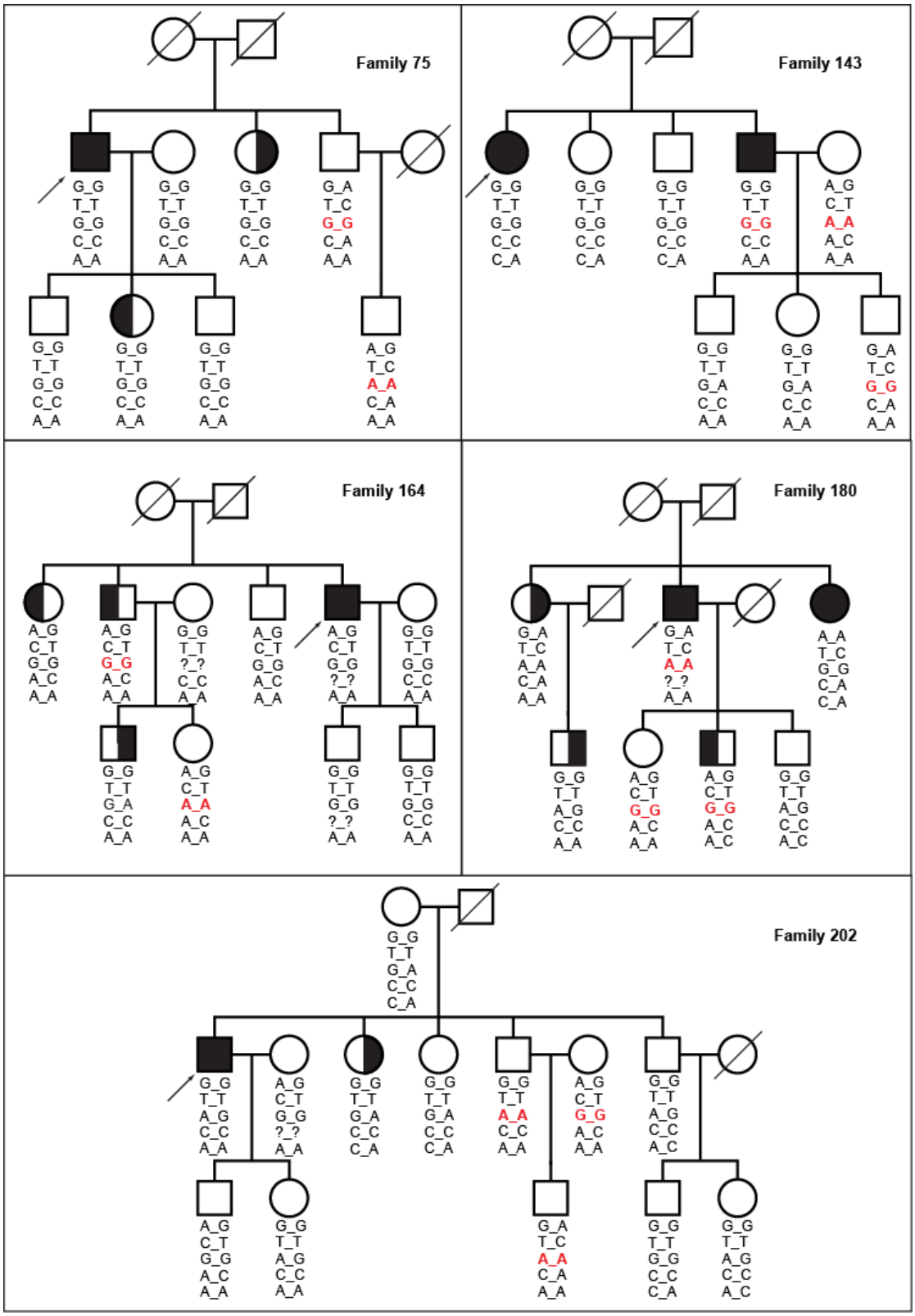
Figure 4.8) Non-Mendelian Inheritance of rs162362 in Six FCHL Families

Genotypes for rs162365, rs10470255, rs162362, rs2838303 and rs229355 are listed. Genotypes of parents and offspring where NMI was detected are highlighted in red. Missing genotypes are indicated by '?_?'. The patterns of non-Mendelian inheritance observed in families 139 and 238 (boxed) cannot be explained by transmission of a hemizygous deletion from one parent.

Figure 4.9)


rs8131995 G_G T_T A_A C_C A_A
rs17004556
rs162386
rs2838307
rs17004557
Family 15
Family 39
Family 53


G_G ?_? G_G C_C A_A
G_A ?_? G_G C_A A_A
G_G T_T A_G C_C C_A
G_A T_C A_A C_A C_A
G_G ?_? A_G C_C C_A
G_A ?_? ?_? C_A A_A
Family 14
A_G C_T G_G A_C A_A
G_G T_T A_G C_C C_A
A_G C_T G_G A_C A_A
A_G C_T A_A A_C A_C
Family 118
A_G C_T G_G A_C A_C
A_G T_T G_A C_C C_C
G_A T_T G_G C_C C_C
A_G C_T A_A A_C A_C
G_G T_T A_G C_C A_A
A_G C_T A_A A_C A_A
Family 207
Family 146
A_G T_T A_G C_C C_A
G_G T_T G_G ?_? A_A
G_G T_T A_G C_C A_A
G_G T_T ?_? C_C A_A
A_G T_T A_G C_C C_A
A_G T_T A_G ?_? C_A
A_G T_T G_G ?_? C_A
G_G T_T A_G C_C A_A
G_G T_T A_G C_C A_A
A_G T_T G_G C_C C_A
G_G T_T A_A C_C A_A
G_G T_T A_G C_C A_A
G_G T_T A_G C_C A_A
G_G T_T A_G C_C A_A
A_G T_T G_A C_C C_A
G_G T_T ?_? C_C C_A
G_A T_C A_A C_A C_A
G_G T_T A_G C_C C_C
G_A T_C G_G C_A C_A


Family 75
G_G T_T G_G C_C A_A
G_G T_T G_G C_C A_A
G_G T_T G_G C_C A_A
G_A T_C G_G C_A A_A
G_G T_T G_G C_C A_A
G_G T_T G_G C_C A_A
G_G T_T G_G C_C A_A
A_G T_C A_A C_A A_A
Family 143
G_G T_T G_G C_C C_A
G_G T_T G_G C_C C_A
G_G T_T G_G C_C C_A
G_G T_T G_G C_C A_A
A_G C_T A_A A_C A_A
G_G T_T G_A C_C A_A
G_G T_T G_A C_C A_A
G_A T_C G_G C_A A_A
Family 164
A_G C_T G_G A_C A_A
A_G C_T G_G A_C A_A
G_G T_T ?_? C_C A_A
A_G C_T G_G A_C A_A
A_G C_T G_G ?_? A_A
G_G T_T G_G C_C A_A
G_G T_T G_A C_C A_A
A_G C_T A_A A_C A_A
G_G T_T G_G ?_? A_A
G_G T_T G_G C_C A_A
Family 180
G_A T_C A_A C_A A_A
G_A T_C A_A ?_? A_A
A_A T_C G_G C_A C_A
G_G T_T A_G C_C A_A
A_G C_T G_G A_C A_A
A_G C_T G_G A_C A_C
G_G T_T A_G C_C A_C
Family 202
G_G T_T G_A C_C C_A
G_G T_T A_G C_C A_A
A_G C_T G_G ?_? A_A
G_G T_T G_A C_C C_A
G_G T_T G_A C_C C_A
G_G T_T A_A C_C A_A
A_G C_T G_G A_C A_A
G_G T_T A_G C_C A_C
A_G C_T G_G A_C A_A
G_G T_T A_G C_C A_A
G_A T_C A_A C_A A_A
G_G T_T G_G C_C C_A
G_G T_T A_G C_C A_C

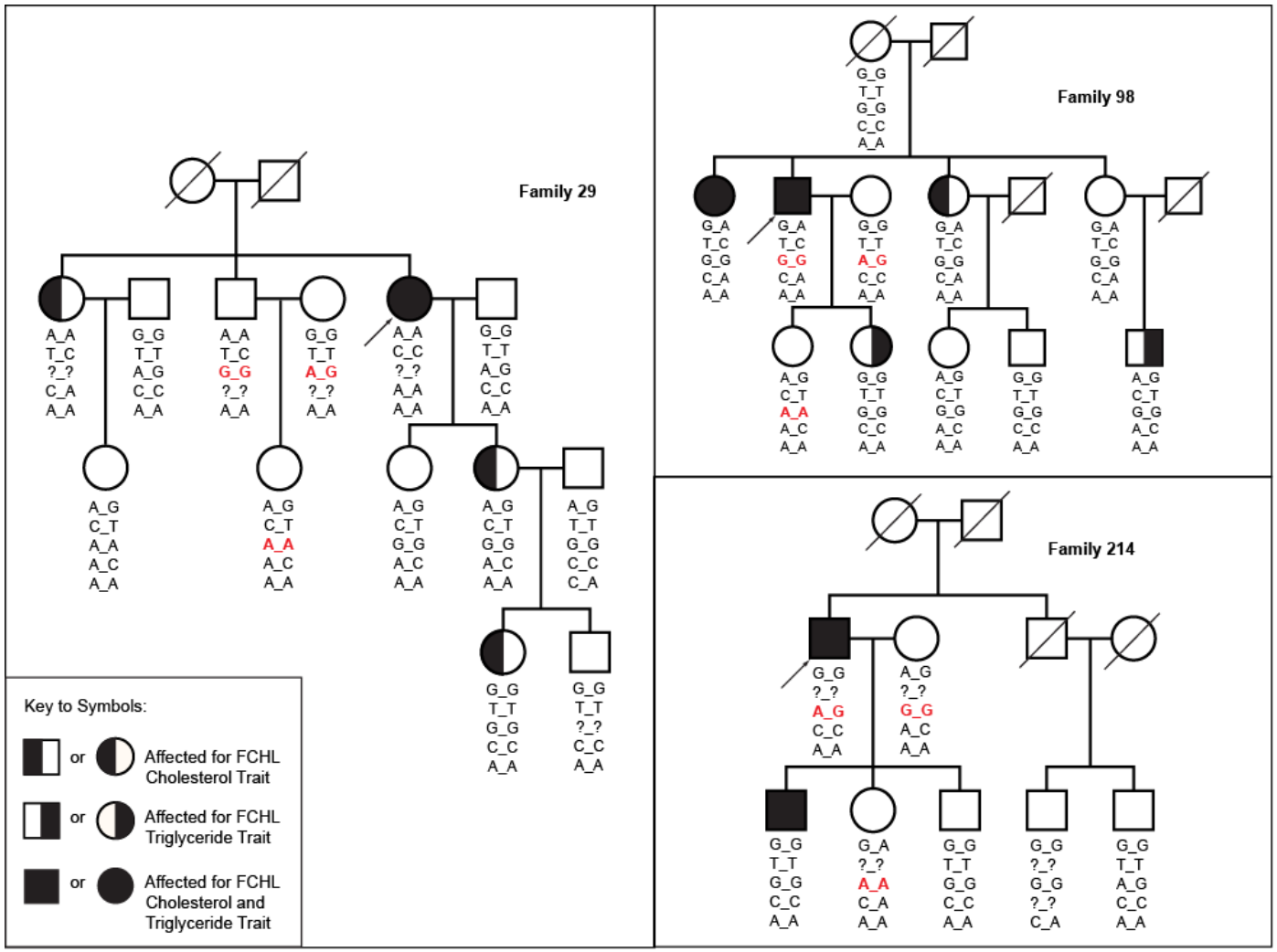

Family 29
A_A T_C ?_? C_A A_A
G_G T_T A_G C_C A_A
A_A T_C G_G ?_? A_A
G_G T_T A_G ?_? A_A
A_A C_C ?_? A_A A_A
G_G T_T A_G C_C A_A
A_G C_T A_A A_C A_A
A_G C_T A_A A_C A_A
A_G C_T G_G A_C A_A
A_G C_T G_G A_C A_A
A_G T_T G_G C_C C_A
G_G T_T G_G C_C A_A
G_G T_T ?_? C_C A_A
Key to Symbols:
or
Affected for FCHL Cholesterol Trait
or
Affected for FCHL Triglyceride Trait
or
Affected for FCHL Cholesterol and Triglyceride Trait
Family 98
G_G T_T G_G C_C A_A
G_A T_C G_G C_A A_A
G_A T_C G_G C_A A_A
G_G T_T A_G C_C A_A
G_A T_C G_G C_A A_A
G_A T_C G_G C_A A_A
A_G C_T A_A A_C A_A
G_G T_T G_G C_C A_A
A_G C_T G_G A_C A_A
G_G T_T G_G C_C A_A
A_G C_T G_G A_C A_A
Family 214
G_G ?_? A_G C_C A_A
A_G ?_? G_G A_C A_A
G_G T_T G_G C_C A_A
G_A ?_? A_A C_A A_A
G_G T_T G_G C_C A_A
G_G ?_? G_G ?_? C_A
G_G T_T A_G C_C A_A

Figure 4.9) Non-Mendelian Inheritance of rs162386 in 15 FCHL Families
Genotypes for rs8131995, rs17004556, rs162386, rs2838307 and rs17004557 are listed. Genotypes of parents and offspring where NMI was detected are highlighted in red. Missing genotypes are indicated by '?_?'. The patterns of non-Mendelian inheritance observed in all families can by explained by transmission of a hemizygous deletion from one parent which is always carried on a genetic background of A, C, 'deletion', A, A. Other family members carry this haplotype but NMI is not observed due to pedigree structure or because of the variant allele transmitted by the other parent.

Genotype call rates were determined for 557 SNPs in the 1725 FCHL family cohort members for whom quality control checks were satisfactory. The mean call rate for these 557 markers was 0.96±0.07. The removal of 13 SNPs that had a genotype call rate of below 0.8 from the dataset increased the mean call rate to 0.97±0.03.

As an additional screening procedure, 42 variants with a minor allele frequency of below 0.01 were also removed from the dataset (Table 4.5a and 4.5b). To assess whether any of these 'rare' variants were enriched within the FCHL cohort, allele frequencies were compared with those recorded in the 1000 Genomes European dataset. Of the 42 variants with a minor allele frequency <0.01, 24 were mono-allelic in the entire FCHL family cohort and 2 were found in only one individual (and were mono-allelic in the pedigree founders). Furthermore, of these 26 variants, 22 were not recorded in the European populations of the 1000 Genomes dataset (Table 4.5a) and two, rs61739237 and rs62000965, were present on only chromosome of one individual in the 1000 Genomes data (Table 4.5a). The minor allele frequencies of the outstanding two variants (rs12627414 and rs2838946) were ~4% in the 1000 Genomes dataset (Table 4.5a), perhaps indicating that the genotyping assays employed in the FCHL family cohort were unable to properly ascertain the minor alleles of these 'SNPs'.

Three low frequency variants (rs12329729, rs12627675 and rs62000960) were detected in the FCHL cohort but were not recorded in the 1000 Genomes dataset (Table 4.5a). Whilst rs12329729 and rs12627675 were found in just one and three members of the FCHL cohort respectively, the *COL18A1* non-synonymous coding variant, rs62000960 was heterozygous in nine individuals, segregating in three FCHL families, plus one lone individual. Figure 4.10 shows that the rare allele of this variant does not appear to segregate with affection status for either the FCHL-triglyceride or cholesterol trait.

In total, 13 of the 42 low frequency SNPs were present in both the FCHL and 1000 genomes datasets (Table 4.5b). The rare allele of one SNP, rs11911327 was present on 0.0177 of chromosomes from the 1000 Genomes European dataset compared to 0.0032 of chromosomes in the FCHL founders (P=0.02), which suggests that depletion of this allele could be important in FCHL. However, this possibility seems unlikely since the frequency of the rare allele was in fact higher in FCHL probands versus spouses (controls) (Table 4.5b).

In sum, these results indicate that the low frequency SNPs genotyped in this project are unlikely to contribute to the FCHL linkage signal on chromosome 21q. At the end of all quality control procedures, a final dataset of 502 SNPs, genotyped in 1725 individuals was available for statistical analysis.

Table 4.5a) Genotyped SNPs found to be Mono-allelic in FCHL Founders and/or 1000 Genomes European Ancestry Dataset

			Minor Allele Frequency		
SNP	Gene	Position	All (n=3450)	Founders (n=662)	1000G (n=566)
rs8131089	*ABCG1*	43,680,107	0	0	NR
rs17114631	*ABCG1*	43,708,490	0	0	NR
rs28653325	*SLC37A1*	43,938,500	0	0	NR
rs9976017	*CBS*	44,433,203	0	0	NR
rs5742905	*CBS*	44,483,184	0	0	NR
rs2298758	*CBS*	44,485,527	0	0	NR
rs34040148	*CBS*	44,488,631	0.0003	0	NR
rs35596465	*SNF1LK*	44,836,800	0	0	NR
rs34614061	*SNF1LK*	44,837,465	0	0	NR
rs34704670	*SNF1LK*	44,838,283	0	0	NR
rs35519605	*SNF1LK*	44,838,395	0	0	NR
rs34164089	*SNF1LK*	44,839,075	0	0	NR
rs45491503	*SNF1LK*	44,841,593	0	0	NR
rs35051468	*AGPAT3*	45,379,611	0	0	NR
rs11908904	*TSPEAR*	46,113,691	0	0	NR
rs61749002	*COL18A1*	46,875,586	0	0	NR
rs2236453	*COL18A1*	46,888,627	0	0	NR
rs62000458	*COL18A1*	46,897,730	0	0	NR
rs61738822	*COL18A1*	46,900,062	0	0	NR
rs58380358	*COL18A1*	46,912,474	0	0	NR
rs61736805	*COL18A1*	46,932,115	0.0003	0	NR
rs61735035	*COL18A1*	46,932,212	0	0	NR
rs12627414	*TSPEAR*	46,025,749	0	0	0.0389
rs61739237	*COL18A1*	46,875,683	0	0	0.0018
rs62000965	*COL18A1*	46,895,427	0	0	0.0018
rs2838946	*COL18A1*	46,908,569	0	0	0.0442
rs12329729	*COL18A1*	45,967,330	0.0003	0.0016	NR
rs12627675	*COL18A1*	46,839,075	0.0009	0.0016	NR
rs62000960	*COL18A1*	46,875,961	0.0028	0.0016	NR

n=maximum number of chromosomes genotyped, 1000G=1000 Genomes European dataset, NR=not recorded

Table 4.5b) Frequencies of SNPs with 0<MAF<0.01 in FCHL Family Cohort Founders, Probands, Spouses and in 1000 Genomes European Ancestry Dataset

			Minor Allele Frequency			
SNP	Gene	Position	Founders (n=662)	1000G (n=566)	Probands (n=478)	Spouses (n=488)
rs8133241	*BACE2*	42,561,034	0.0091	0.0106	0.0065	0.0104
rs10084555	*ABCG1*	43,623,379	0.0062	0.0035	0.0066	0.0063
rs11702801	*ABCG1*	43,670,367	0.0091	0.0035	0.0064	0.0145
rs2839481	*ABCG1*	43,687,142	0.0016	0.0018	0	0
rs884979	*SLC37A1*	43,985,452	0.0093	0.0018	0.0194	0.0106
rs34987632	*SNF1LK*	44,839,263	0.0016	0.0053	0.0044	0.0021
rs2838470	*AGPAT3*	45,437,183	0.0060	0.0088	0.0063	0.0041
rs17004674	*TSPEAR*	46,001,361	0.0030	0.0053	0	0.0042
rs11911327	*COL18A1*	46,848,156	0.0032	0.0177[a]	0.0068	0.0022
rs9976881	*COL18A1*	46,897,893	0.0078	0.0035	0	0.0064
rs2230693	*COL18A1*	46,913,477	0.0032	0.0072	0.0022	0.0043
rs9983151	*COL18A1*	46,930,442	0.0082	0.0035	0.0048	0.0067
rs2839213	*MCM3AP*	47,752,496	0.0080	0.0141	0.0044	0.0089

[a] Frequency significantly different in FCHL founders and 1000 Genomes dataset, *P*=0.02

n=maximum number of chromosomes genotyped, 1000G=1000 Genomes European dataset, NR=not recorded

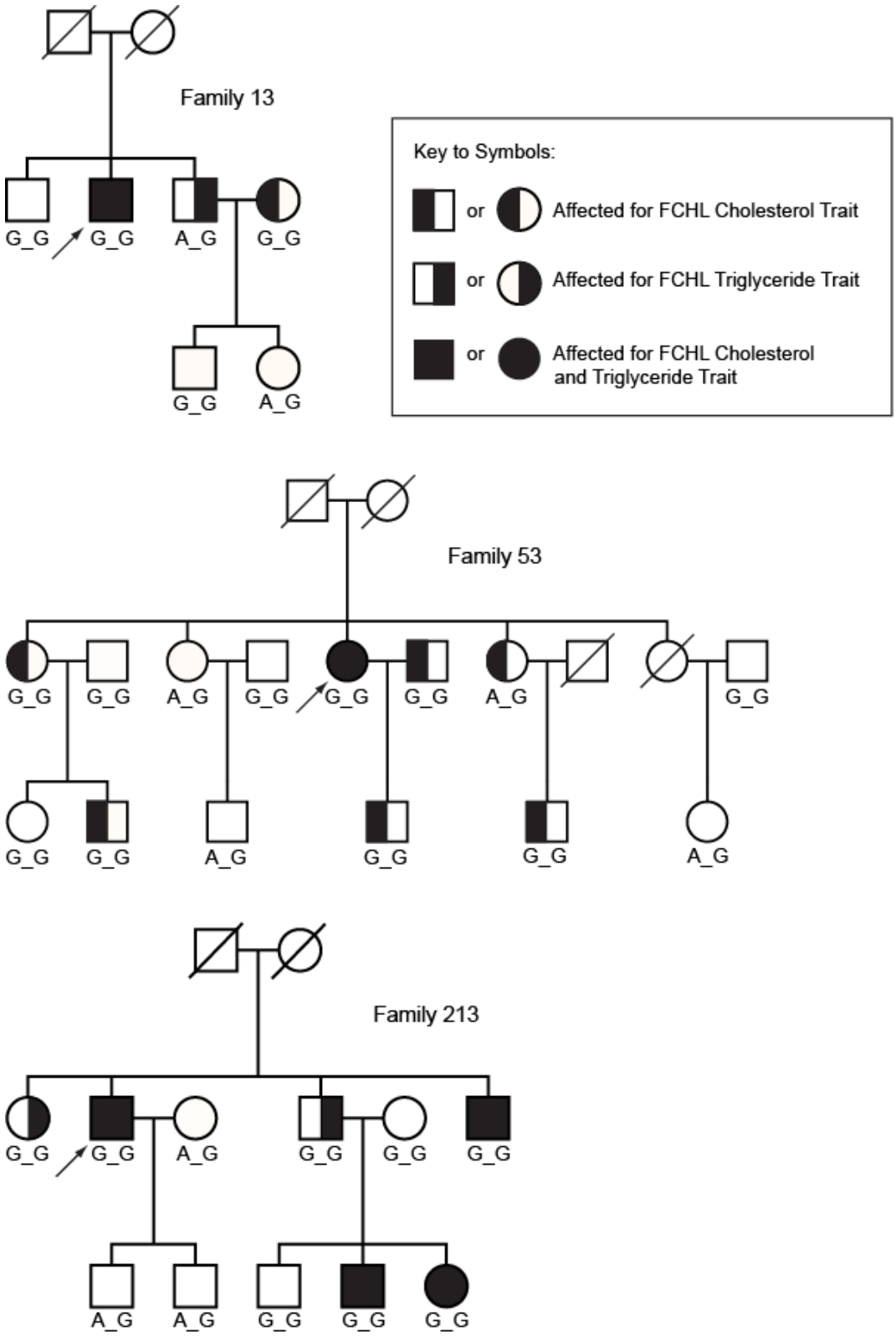


Figure 4.10) Segregation of rs62000960 in Three FCHL Families

The minor allele (A) of the *COL18A1* non-synonymous coding variant rs62000960 does not appear to segregate with lipid traits in FCHL families.

4.2.5) Coverage of Linkage Interval by Directly Genotyped SNPs

Coverage of the candidate genes and their associated target regions by the SNPs passing quality control measures was estimated using a minimum r^2 value of 0.8 and a pairwise tagging method (Table 4.6). The genotyped SNPs captured an average of 86.64±12.18% of the common variation in the candidate genes. Excluding the *SIK1*/*TSPEAR* and intervals, which had lower coverage (both 64%), 91.67±5.39% of common variation was captured.

In total, the final SNP dataset captured common genetic variation across 1.42Mb of chromosome 21q22.2-22.3, including 1.15Mb of the 3.8Mb linkage interval (~30%).

Table 4.6) Coverage of Target Genes by Genotyped SNPs Following Quality Control Measures

			Excluded SNPs						
Gene	Target Interval[a]	Total SNPs[b]	No Data	Call Rate <80%	NMI	HWE[c]	MAF<1%	Coverage[d] (%)	Span[e] (Kb)
BACE2	42,532,075-42,657,548	71	5	4	0	0	1	88	125.47
ABCG1	43,619,415-43,718,872	48	2	2	0	0	5	87	99.46
SLC37A1	43,905,327-44,006,117	36	0	1	0	0	2	97	100.79
CBS	44,423,653-44,519,655	55	0	0	0	0	4	98	96.00
SIK1	44,784,639-44,891,718	70	4	1	2	1	7	64	107.08
AGPAT3	45,237,039-45,450,717	54	3	0	0	0	2	94	213.68
TSPEAR	45,928,055-46,138,905	60	4	2	1	0	4	64	343.40
UBE2G2	46,139,210-46,271,452	55	2	0	0	0	0	94	
COL18A1	46,774,695-46,983,224	79	5	2	1	0	16	81	186.69
LSS	47,558,994-47,698,343	43	1	0	0	0	0	93[f]	144.66
MCM3AP	47,699,856-47,752,496	12	1	1	0	0	1		
	TOTAL	583	27	13	4	1	42		1417.22

[a] Co-ordinates according to human genome reference sequence build hg19 from first to last genotyped SNP of specified gene.
[b] Total number of genotyped SNPs in given gene interval
[c] Deviation from HWE with P<0.001
[d] Coverage calculated using HapMap 2 data and non-excluded SNPs.
[e] Span is calculated from first to last non-excluded in gene target interval. For the adjacent genes *TSPEAR-UBE2G2* and *LSS-MCM3AP* span incorporates both intervals.
[f] *MCM3AP* is adjacent to *LSS* and was partially covered by original *LSS* tag-SNPs. Coverage is calculated for complete interval including both genes
SNP=single nucleotide polymorphism, NMI=non-Mendelian inheritance, HWE= Hardy Weinberg Equilibrium, MAF=minor allele frequency

4.2.6) Preparation of Sample and Reference Datasets for Imputation

To prepare the sample dataset for use with MaCH, 19 markers with genotype call rate <0.90 were excluded (Table 4.7). SNPs within the sample dataset were aligned to the + strand, to ensure compatibility with reference (HapMap or 1000 Genomes) datasets. Accurate specification of the + strand was not possible for two markers (rs2837906 and rs3746917) with ~50:50 G/C allele frequencies hence these were removed from the input dataset. In total 481 SNPs were included in the input sample dataset.

When the 1000 Genomes CEU dataset was used as a reference, a further 20 markers were filtered from the sample dataset due to their absence from the June 2010 release. When the HapMap 2 reference dataset was employed, 80 non-HapMap SNPs were excluded.

Table 4.7) SNPs with Call-Rate 0.80-0.90 Excluded from Imputation Dataset

SNP	Target Gene	Position (bp)	Call Rate
rs2898441	*BACE2*	42,606,524	0.88
rs2837994	*BACE2*	42,624,124	0.86
rs6517659	*BACE2*	42,641,982	0.82
rs2006737	*BACE2*	42,657,187	0.88
rs2839538	*SLC37A1*	43,915,541	0.85
rs4920123	*SLC37A1*	43,990,667	0.89
rs2096852	*SNF1LK*	44,825,683	0.88
rs162365	*SNF1LK*	44,854,995	0.89
rs9647234	*TSPEAR*	46,098,545	0.81
rs1980978	*COL18A1*	46,818,397	0.89
rs2838906	*COL18A1*	46,821,907	0.81
rs17004772	*COL18A1*	46,849,777	0.89
rs2236458	*COL18A1*	46,894,021	0.80
rs2236469	*COL18A1*	46,902,390	0.89
rs9977268	*COL18A1*	46,907,287	0.86
rs7278425	*COL18A1*	46,926,551	0.86
rs4818835	*MCM3AP*	47,724,702	0.89
rs16979065	*MCM3AP*	47,749,456	0.87
rs2839212	*MCM3AP*	47,751,347	0.89

4.2.7) Quality Measures for Imputed Genotypes

In order to determine which reference dataset and subsample of the FCHL cohort would enable the most accurate imputation model to be built, imputation error rates were estimated for a number of possible scenarios (Table 4.8). Specifically, 2% of the input genotypes were masked at random, imputed and then compared to the experimentally determined data in order to estimate per allele and per genotype error rates. Error rates were compared for imputation models based on either 1000 Genomes or HapMap Phase 2 reference datasets and the genotypes of either 239 FCHL probands, or 244 family spouses, or both. In each instance, model estimation involved 100 iterations.

Table 4.8 shows that model estimation had the lowest error rate when using the 1000 Genomes reference dataset and when both proband and spouse data were utilised. Genotype imputation across the entire dataset was therefore performed using a model formulated using these parameters.

Following genotype imputation, the complete dataset comprised information for 17875 SNPs (461 typed plus 17414 untyped SNPs). Of the 17875 markers, 4685 (461 typed plus 4224 imputed) passed the recommended imputation quality threshold of r^2>0.5. Genotype dosage data for these SNPs were subsequently utilised in association testing.

Table 4.8) Estimated Allelic and Genotypic Error Rates for Different Imputation Model Estimation Parameters Using MaCH

Reference Dataset	Subsample of Cohort used in Model Estimation					
	Probands		Spouses		Probands & Spouses	
	AER	GER	AER	GER	AER	GER
1000G 06/2010	0.0383	0.0746	0.0382	0.0744	0.0380	0.0740
HapMap 2	0.0383	0.0748	0.0383	0.0745	0.0383	0.0746

AER=allelic error rate, GER=genotypic error rate, 1000G=1000 Genomes

4.2.8) Coverage of Linkage Interval Following Genotype Imputation

After quality control of the imputed dataset, coverage of the candidate gene regions was estimated, (using HapMap SNPs only, n=1696 due to their compatibility with the Haploview software). Based on the filtered imputed dataset, 97.78±3.07% of common variation across the candidate genes was captured (Table 4.9). Importantly, coverage if the *SIK1* and *TSPEAR* interval was increased from 64% to 97% and 98% respectively. In addition, the imputed genes covered an expanded interval of 1.67Mb compared to 1.42Mb before imputation. This constitutes ~34% of the 3.8Mb linkage interval (3.2.2) (*BACE2* and *LSS* reside outside of this region)

Table 4.9) Coverage of Candidate Genes Following Genotype Imputation

Gene Interval	Span of Imputed SNPs[a]	Total SNPs[b]	Coverage (%)
BACE2	42,515,277-42,714,288	175	90
ABCG1	43,607,858-43,727,683	125	99
SLC37A1	43,905,327-44,007,516	152	100
CBS	44,410,725-44,535,683	163	99
SIK1	44,778,836-44,923,146	145	97
AGPAT3	45,214,862-45,462,853	173	99
TSPEAR to UBE2G2	45,928,055-46,284,110	446	98
COL18A1	46,768,117-46,965,499	168	98
LSS to MCM3AP	47,558,994-47,740,905	149	100
Total	1673.63Kb	1696[c]	97.78±3.07[d]

[a] Co-ordinates according to human genome reference sequence build hg19 from first to last genotyped SNP around given gene.

[b] Total number of SNPs after imputation and SNP filtering using described quality controls.

[c] Total number of filtered imputed HapMap SNPs. Total filtered, imputed SNPs=4685.

[d] Mean±standard deviation

4.3) Discussion

When performing association analysis, incorporating large numbers of samples and/or SNPs, good quality control and an efficient study design are essential (reviewed in (Neale and Purcell, 2008)). In this investigation, ten biological candidate genes and their flanking regions were targeted and the implementation of quality measures yielded a dataset, ready for association analysis, comprising information for 502 SNPs, genotyped in 1725 individuals. This dataset encompasses 1.15Mb (~30%) of the 3.8Mb linkage interval (Chapter 3).

CNV is a form of heritable variation that is expected to contribute to common disease susceptibility (Manolio et al., 2009). Traditionally CNVs are >1kb in length, though this figure is borne out of the limitations of techniques employed for CNV detection (Zöllner and Teslovich, 2009). About 90% of CNVs identified to date have two allelic states (McCarroll et al., 2008) and around two-thirds of catalogued variants are deletions (URL22). Forms of evidence that can be used to infer the presence of copy number variable regions include the observation of apparent NMI in genotyped families (Aitman et al., 2006) and departures from HWE (Lee et al., 2008b).

In this study, four SNPs displayed recurrent problems in Mendelian inheritance and one showed a significant deviation from Hardy-Weinberg equilibrium. A potential copy number variable region of up to 929 base pairs in length was identified due to NMI of rs162386. Specifically, the patterns of inheritance in FCHL families suggest that this interval may harbour a deletion. Data from a previous study identified a much larger putative deletion (spanning ~50kb) in this region (Jakobsson et al., 2008) in two individuals. Jakobsson *et al.,* used PennCNV (Wang et al., 2007b) to predict potential CNV sites using genome-wide SNP data. These data comprised genotypes for 485 samples taken from the Human Genome Diversity Panel individuals gathered using the Illumina Hap 550 genotyping chip.

In order to determine whether the putative deletion identified in this study is in fact a true deletion, one could use PCR to amplify and sequence overlapping segments across the region. Should a deleted region be identified then a technique such as multiplex ligation dependent probe amplification (MLPA) could be used for typing in many individuals (Schouten et al., 2002). However, before further investigation is carried out, it is worth noting that the phenotypic impact of this putative CNV with regard to FCHL is questionable. Of 31 potentially hemizygous individuals in the 15 NMI families, six were FCHL probands compared to five family spouses. Twelve

individuals had cholesterol levels above the 95th age- and sex-specific percentile values and just seven had triglyceride levels above the 90th age- and sex-specific percentile values.

Recent studies have identified that hypertriglyceridemic individuals have an enrichment of rare variants in genes that are biological candidates for the disorder (Johansen et al., 2010; Wang et al., 2008). Considering these findings, the localisation and minor allele frequencies of 13 rare variants (0<MAF<0.01) genotyped in the FCHL family cohort were examined more closely. None were present solely in FCHL probands, though one non-synonymous coding variant (rs62000960), absent from current 1000 Genomes data (URL9), was found in ten FCHL family members. However, the variant in question is unlikely to play a role in FCHL aetiology since in these families it did not segregate with the FCHL triglyceride and/or cholesterol traits. One variant was present at a significantly lower frequency in FCHL families compared to individuals of European descent genotyped as part of the 1000 Genomes Project (URL9). However, as the frequency of this variant in FCHL probands was intermediate to those observed in the pedigree founders and the 1000 Genomes dataset, it too is unlikely to play a part in the aetiology of FCHL.

The imputation approach applied in this investigation is analogous to that used in a recent family-based association study of bipolar disorder, targeted to chromosome 8q24 (Zhang et al., 2010a). In their investigation, Zhang and colleagues first performed model estimation using a selection of 200 unrelated individuals from their cohort which in total comprised 3512 individuals from 737 families. In comparison, in this study, data from 239 probands, 244 spouses or both (473 unrelated individuals) were used in model estimation. Error rates for different model building scenarios were compared, with the fewest errors produced when both probands and spouses were used in model estimation. Using these parameters, genotypic and allelic error rates in the FCHL cohort were estimated to be 0.0740 and 0.0380, respectively. This allelic error rate is similar to that estimated by Li *et al* in their description of the MaCH software (Li et al., 2010). In this instance, a HapMap Phase 2 CEU reference panel comprising 872 SNPs was used to impute the genotypes of a further 992 SNPs in European populations of the Human Diversity Genome Project with an estimated allelic error rate of <0.034 (Li et al., 2010).

MACH produces an r^2 value for each marker which is equivalent to the ratio of the empirically observed dosage variance to the expected dosage variance (under a

binomial distribution, p(1 − p), where p is the estimated minor allele frequency). Using the 1000 Genomes reference dataset, 4685 of the 17875 imputed markers passed the recommended imputation quality threshold of r^2>0.5. The large number of SNPs falling below this threshold is a likely reflection of the experimental design, where no marker information was available in the sample dataset, outside of the target gene intervals.

Overall, the result of genotype imputation in the FCHL cohort was to increase the coverage of the linkage interval, without the immediate need for further direct genotyping. Prior to imputation, the filtered genotyped variants were estimated to capture an average of 0.87±0.12 of the common variation in the candidate genes. Post-imputation and quality control, this value was increased to 0.98±0.03 (significant increase, *P*=0.02). Thus, imputation was a cost-effective alternative to further direct genotyping. However, one should note that since genotypes generated *in silico* are uncertain, any imputed SNP which generated a positive association signal would require subsequent direct genotyping to confirm the result.

The efficiency of imputation is discussed by Anderson and colleagues in a comparison of data obtained by direct genotyping of 1376 samples using the Illumina HumanHap 550 genotyping chip and by a combination of genotyping with the Illumina HumanHap 300 and imputation (Anderson et al., 2008). At an r^2 of 0.8, the genomic coverage of the Illumina HumanHap 550 is 87% compared to 77% for the HumanHap 300 chip. However, when the HumanHap 300 dataset is subject to imputation using a HapMap 2 reference dataset, this coverage estimate is increased to 81%. In addition, in their study, when the sample size was weighted according to a genotyping cost per sample, (HumanHap 300 weight 1.32 vs. HumanHap 500 weight 1), Anderson *et al*. found that under multiple scenarios of minor allele frequency and sample size, the imputed dataset always provided the most power (Anderson et al., 2008). As such, it should be noted that the current study design increased coverage estimates for the target genes to 98%, above that achieved by even the larger SNP arrays that type ~1 million SNPs genome-wide (Li et al., 2010).

Chapter 5 – Testing for Association between Chromosome 21q22.2-22.3 Common Variants and FCHL Lipid Traits

5.1) Introduction

In an association study, an important consideration is how best to analyse the SNP dataset and subsequently how to interpret the results obtained. A balance must be found between maximising the power of the study to detect association, avoiding potentially false positive associations that arise due to confounding factors such as population stratification whilst avoiding false negative results arising from over-conservative correction for the multiple statistical tests performed.

One may consider two broad approaches in analysing a family-based dataset. Firstly, the traditional family-based methodologies such as the transmission disequilibrium test (TDT), the family-based associated test (FBAT) and their various extensions (Ewens and Spielman, 1995; Laird et al., 2000; Martin et al., 2000) or second, 'population-based' methods such as linear mixed effects (LME) modelling (Chen and Yang, 2010), the more-powerful quasi-likelihood score method (MQLS) (Thornton and McPeek, 2007), or standard generalised estimating equations (GEE, (Chen and Yang, 2010)).

Traditional family-based tests, like the TDT, assess transmission of alleles within a given family, using parent-offspring and/or discordant sibships. The nature of this approach means that it is robust to population stratification (Ewens and Spielman, 1995). However, family units become 'uninformative' when transmission from parents to children is ambiguous due to their homozygosity, leading to a reduction in the effective sample size and a concomitant reduction in power. Likewise, the within-family approach can be considered inefficient since it does not make use of all available information, ignoring the allelic association observed across families (Chen et al., 2009; Cupples, 2008).

The alternative 'population-based' approaches model the phenotype as a function of the genotype. In order to account for familial relatedness, a variance/covariance matrix is incorporated (Chen and Yang, 2010). This approach permits the use of information from all genotyped individuals, regardless of whether they are heterozygous for a given SNP, thus maximising study power. The counterbalance is that this type of association analysis is not robust to confounding by population

substructure. A simple method of checking for bias in association studies that might result from undetected population substructure is to produce a quantile-quantile (Q-Q) plot of observed versus expected –log10(*P*-values) for the association tests performed. This plot characterises the extent to which the observed distribution of the test statistic follows the expected (null) distribution. On a genome-wide scale and using unrelated individuals, a method such as principle components analysis (applied in the EIGENSOFT package, (Patterson et al., 2006; Price et al., 2006)) can be used to detect and correct any such stratification.

A recent simulation study has compared the power and type 1 error rates of association analysis methods applied to data from related individuals (Manichaikul et al., 2011). This study showed that 'population-based', total association methods such as LME for quantitative traits and MQLS or GEE for dichotomous traits, consistently out-performed both the within-family methods such as FBAT and GDT (generalised disequilibrium test, an extension of the TDT, (Chen et al., 2009)) and simple regression-based approaches which do not take into account the family structure. The simulations also demonstrated increased power for quantitative versus dichotomous trait analyses (Manichaikul et al., 2011).

In detail, in a simulated dataset involving 2,305 individuals across 687 families, considering a single marker with a minor allele frequency of 0.3 and a normally distributed quantitative phenotype where the disease marker accounted for 0.5% of trait variance, the power of LME at P=0.01 was 73.1% compared to 31.2% for FBAT (Manichaikul et al., 2011) When family-relatedness was ignored and simple linear regression was applied, the type 1 error rate was approximately double that of the other methods examined. Likewise, for dichotomous trait analysis under an additive model, with a simulated disease prevalence of 5%, GEE and MQLS ('population'-based) had power of 13.3% and 15.5% respectively, compared to power of 5.2% for GDT (family-based). For a more common disease, with 50% prevalence, these values rose to 47.4% for GEE and 50.9% for MQLS compared to 20.1% for GDT. Type 1 error rates were again inflated (1.5 to 2-fold) when family structure was ignored (Manichaikul et al., 2011).

Genome-Wide Association analyses with Family (GWAF) is an R package that implements both LME (quantitative trait) and GEE (dichotomous trait) (Chen and Yang, 2010). Unlike other implementations of population-based approaches for family-based studies (Thornton and McPeek, 2007), GWAF allows the incorporation

of covariates, such as body mass index, into analyses, provides estimates of effect sizes in its output files and importantly, facilitates the analysis of imputed data with the accommodation of genotype dosages (Chen and Yang, 2010; Manichaikul et al., 2011). GWAF was therefore chosen to perform association analyses using the FCHL family cohort data, the results of which are presented in this chapter.

Aside from choosing the most appropriate analysis method for a large-scale association study, a further challenge is to how to correctly interpret the significance of the results obtained. One method frequently employed is to control the family-wise error rate (the probability of observing ≥1 significant under the null hypothesis of no association), at a nominal (and indeed arbitary) level such as α_e=0.05. The simplest way to do this is to apply a Bonferroni correction to the pointwise (per-SNP) error rate (α_p), for a given α_e, such that:

$$\alpha_p = \alpha_e/n$$

where n is the number of *independent* tests performed. A genome-wide correction for 1,000,000 SNPs was proposed (Risch and Merikangas, 1996) and thus many GWAS published to date apply a significance threshold of $P<5x10^{-8}$ (Barrett et al., 2008; Ehret et al., 2011; Teslovich et al., 2010). However, recent studies have sought an alternative to the Bonferroni correction because, in studies where many, high-density SNPs are genotyped, underlying LD between markers is likely to undermine the assumption of independent tests that it requires (Johnson et al., 2010). As a result, the Bonferroni correction can be conservative and lead to an increase in type 2 error (false-negative) rate, particularly when sample sizes do not run into tens of thousands (Johnson et al., 2010). Even in a study design which employs tag-SNPs, underlying LD will exist between genotyped SNPs, dependent upon the r^2 threshold specified during the tag-SNP selection process.

Permutation testing is a favourable alternative to a Bonferroni correction due to its robustness, regardless of correlation between markers (Hoggart et al., 2008). In this approach, test statistics are calculated in both actual and permuted datasets. In the permuted datasets, the phenotype (for example case/control status) of the typed individuals is randomly shuffled to quantify the level of chance association. Using the best results from each permutation, a null distribution of test statistics can be assembled and comparisons between this and the actual data can be used to empirically define the statistical significance of the tests performed, based upon the appropriate percentile of the null distribution (Hirschhorn and Daly, 2005).

The permutation approach is computationally demanding for large datasets which can make it unsuitable for large-scale SNP data (Gao et al., 2008). Churchill and Doerge suggested that 1000 permutations are required for calculating significance thresholds at α_e =0.05, with this value rising to 10,000 for α_e =0.01 (Churchill and Doerge, 1994). In a recent simulation study, Gao *et al*, estimated that in a dataset comprising 100 SNPs genotyped in 1000 individuals, three hours computation time were required to perform 100,000 permutations. This figure extrapolated to over three years, on a standard computer, when a dataset comprising one million SNPs was considered (Gao et al., 2008). Permutation programs optimised for GWAS have recently been devised, including PRESTO (Browning, 2008) and PERMORY (Pahl and Schafer, 2010), although these currently only handle data from unrelated individuals or trios.

A further option is the calculation of the effective number of independent tests (Cheverud, 2001; Gao et al., 2008; Li and Ji, 2005; Nyholt, 2004). This process takes account of the correlation between SNPs and determines the effective number of independent SNPs upon which a Bonferroni correction can then be applied. Nyholt developed the web-based interface, SNPSpDlite (URL11), which implements the formula of Li and Ji to calculate the effective number of independent tests ($M_{eff\text{-}L}$) (Li and Ji, 2005; Nyholt, 2004). This approach is advantageous since it is directly applicable to family-based data. A recent comparison of these techniques also found that the $M_{eff\text{-}L}$ provided the best approximation of the permutation-based threshold even when considering >500 SNPs (Wen and Lu, 2011).

5.2) Results

5.2.1) Association of Directly Genotyped SNPs with FCHL Lipid Traits

Single marker association tests were performed with data for the 502 directly genotyped SNPs which passed quality control checks (4.2.3). Linear mixed effects models and generalised estimating equations were used to model family structure for the quantitative and dichotomous phenotypes respectively.

The effective number of statistical tests, estimated using the method of Li and Ni (Li and Ji, 2005; Nyholt, 2004), was 308.184. Due to the high level of correlation between cholesterol and (log) triglyceride levels in the FCHL cohort (r=0.62), data were corrected for the analysis of only one trait. Therefore, to maintain an experiment-wide error rate of 0.05, the threshold for significant association to be declared equates to $P<0.00016$.

Power calculations for the dichotomous and quantitative phenotypes (Table 5.1) demonstrate that the FCHL cohort is well-powered to detect strong association signals such as those explaining ≥1.5% of the quantitative trait variance or those with an odds ratio of ≤0.5 or >1.5 for dichotomous traits. However, the FCHL cohort has only low to moderate power to detect variants that explain ≤1% of the phenotypic variance. The power calculations also demonstrate that performing quantitative analysis using the entire cohort is potentially a more powerful approach than dichotomising individuals according to affection status for either the FCHL-triglyceride or cholesterol traits. One should also note that whilst these calculations were not subject to correction for family structure, a recent simulation study suggests that the effect of such a correction on power is minimal (<4.7% when a variant with a frequency of 30% accounted for 0.5% of the trait variance) (Manichaikul et al., 2011).

Table 5.1) Power Calculations for Different Analyses Performed using FCHL Family Cohort Association Study Data

Study Design	N	Effect	Power (%)
Dichotomous Trait	653/995[a]	OR=0.5	89.17
		OR=0.75	6.73
		OR=1.1	0.16
		OR=1.25	3.51
		OR=1.5	46.72
		OR=1.75	92.78
		OR=2.0	99.78
Quantitative Trait	1725	VE=0.25%	1.27
		VE=0.5%	8.47
		VE=1%	44.01
		VE=1.5%	78.57
		VE=2%	94.40

Power calculations were performed using QUANTO (URL12) under the assumption of an additive inheritance model, an effect allele frequency of 10%, and an alpha threshold of 1.6×10^{-5} (to correspond with the significance threshold calculated in 5.2.1).

[a] Affecteds/Unaffecteds. Numbers given are for FCHL triglyceride trait

OR=odds ratio, VE=variance explained

Quantitative and dichotomous trait analyses were performed using three groups of individuals, drawn from the FCHL cohort. Firstly, data from the entire cohort was examined, with and without inclusion of *APOE* genotype as a phenotypic covariate. Secondly, in light of the OSA data presented in Chapter 3, data from the subset of 102 families contributing to the OSA linkage signal for the FCHL-triglyceride trait at 51cM, were analysed for association with triglyceride dichotomous and quantitative traits (3.2.2). Thirdly, data for the subset of 47 families contributing to the OSA linkage signal for the FCHL-cholesterol trait at 51cM, were analysed for association with cholesterol dichotomous and quantitative traits (3.2.2).

a) Analyses using the Entire FCHL Cohort

In the entire cohort, no SNPs surpassed the threshold for a significant association, in either quantitative or dichotomous trait analyses. Figure 5.1 and Table 5.2 illustrate the results of the triglyceride trait analyses. In these analyses, the top SNP (*P*=0.0003) was rs8134486, an intronic *ABCG1* tag-SNP (Figure 5.2), which just exceeded the threshold for declaring significant quantitative trait association with triglyceride levels. In the corresponding dichotomous trait analysis the *P*-value for this variant was *P*=0.0992. Multiple species alignments of *ABCG1* sequences revealed that rs8134486 does not reside in a conserved intronic region (Figure 5.3).

In order to substantiate the evidence for association between rs8134486 and triglyceride levels, data from the recent GLGC genome-wide meta-analysis of lipid traits were examined (Teslovich et al., 2010). The GLGC study incorporated genetic and phenotypic data from 46 GWASs (~100,000 individuals) to identify loci that underpin normal variation in lipid levels. Due to its low minor allele frequency, rs8134486 was not included in the GLGC dataset and therefore, in this instance, no conclusion can be drawn about the association of this SNP with triglyceride levels at the population level.

Due to the established association between the metabolism of triglyceride-rich lipoproteins and HDL-cholesterol, the association between rs8134486 and HDL levels was also examined in the 1477 members of the FCHL cohort for whom this information was available. In this quantitative analysis, the resultant *P*-value was *P*=0.7570.

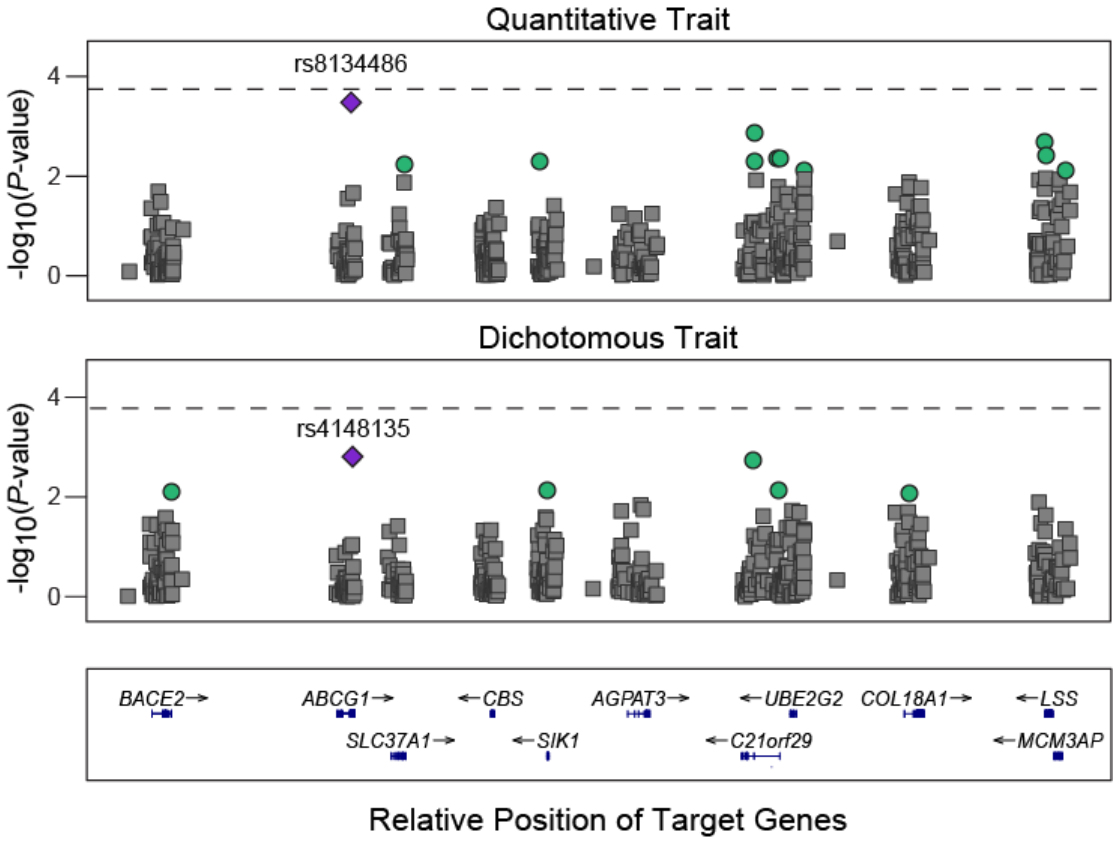


Figure 5.1) Results of Single Marker Association Tests for Directly Genotyped SNPs in the Entire FCHL Cohort – Triglyceride Traits.

To test for association between individual SNPs and serum triglyceride level, considered as a quantitative trait, linear mixed effects models were used. When considering the FCHL-triglyceride trait as a binary outcome (affected=triglyceride ≥90th percentile age- and sex-matched level), logistic regression, via generalised estimating equations was employed. The threshold for significant association is marked by a dotted line. For each analysis, the SNP giving the lowest *P*-value is labelled and is marked by a purple diamond. Other SNPs with an (uncorrected) significance level of $P<0.01$ are indicated by green circles. All other analysed SNPs are denoted by grey squares.

Table 5.2) SNP-Triglyceride Association Results from Single Marker Association Tests, (*P*<0.01 for either quantitative or dichotomous trait analysis)

SNP	Position (bp)	Candidate Gene (Relative Position)	Alleles[a]	MAF	Quantitative Trait (LME)			Dichotomous Trait[d] (GEE)	
					VE (%)[b]	β (SE)[c]	*P*	β (SE)[e]	*P*
rs7281337	42,655,459	*BACE2* (~7kb downstream)	C/G	0.15	0.08	0.0256 (0.0285)	0.3701	0.2980 (0.1128)	0.0082
rs8134486[f]	**43,695,330**	***ABCG1* (intronic)**	**T/C**	**0.01**	**0.64**	**0.2981 (0.0831)**	**0.0003**	**0.4303 (0.2994)**	**0.0992**
rs4148135[f]	43,706,995	*ABCG1* (intronic)	G/A	0.01	0.36	-0.2159 (0.0942)	0.0219	-1.0016 (0.3182)	0.0016
rs12483004	43,992,477	*SLC37A1* (intronic)	A/C	0.34	0.37	0.0589 (0.0215)	0.0060	-0.0466 (0.0774)	0.3562
rs9637180[f]	44,793,184	*SIK1* (~41.2kb downstream)	G/A	0.09	0.49	0.0983 (0.0351)	0.0051	0.0634 (0.1383)	0.1133
rs229346[f]	44,839,654	*SIK1* (intronic)	C/T	0.13	0.25	0.0482 (0.0297)	0.1047	0.2847 (0.1066)	0.0076
rs457647[g]	45,989,115	*TSPEAR* (intronic)	G/A	0.02	0.57	-0.1900 (0.0680)	0.0052	-0.5843 (0.2685)	0.0018
rs458178[g]	45,991,288	*TSPEAR* (intronic)	C/G	0.21	0.62	-0.0779 (0.0243)	0.0014	-0.1269 (0.0892)	0.0607
rs1467385[g]	46,121,817	*TSPEAR* (intronic)	A/T	0.30	0.47	0.0635 (0.0223)	0.0044	-0.0050 (0.0797)	0.0929
rs1296489[g]	46,124,881	*TSPEAR* (intronic)	T/C	0.28	0.57	-0.0619 (0.0218)	0.0045	0.0241 (0.0817)	0.0077
rs11312[g]	46,270,196	*UBE2G2* (~48.4kb upstream, 3'UTR *PTTG1IP*)	G/A	0.37	0.32	0.0545 (0.0205)	0.0080	0.0012 (0.0770)	0.0474
rs8129539	46,852,978	*COL18A1* (intronic)	C/G	0.10	0.33	0.0831 (0.0336)	0.0133	0.3130 (0.1188)	0.0084
rs2187118[h,i]	47,611,310	*LSS* (intronic)	C/T	0.41	0.68	-0.0623 (0.0202)	0.0021	-0.1130 (0.0836)	0.0498
rs35785446[i]	**47,614,553**	***LSS* (p.R614W)**	**G/A**	**0.04**	**0.51**	**0.1390 (0.0482)**	**0.0039**	**0.4984 (0.1965)**	**0.0231**
rs4818835[h,i]	47,724,702	*MCM3AP* (~18kb upstream, *C21orf58* intronic)	G/A	0.29	0.54	0.0611 (0.0230)	0.0079	0.1314 (0.0867)	0.0443

bp=base pairs, GEE=generalised estimating equations, LME=linear mixed effects model, MAF=minor allele frequency, SE=standard error, VE=variance explained

[a] Major allele/minor allele

[b] Proportion of phenotypic variance explained by SNP

[c] For quantitative trait analyses, β represents the change in triglyceride level (natural logarithm mmol/l) per copy of the minor allele

[d] Individuals were dichotomised according to affection for FCHL triglyceride trait. Affected=triglyceride levels ≥90^{th} age- and sex- specific percentile values, unaffected= triglyceride levels <90^{th} age- and sex- specific percentile values.

[e] Odds ratio for the FCHL triglyceride trait, per copy of the minor allele, corresponds to e^{β}

[f] No pairwise LD (r^2<0.05) between rs8134486/rs4148135, rs9637180/rs229346 and rs457647/rs458178/rs1467385/rs1296489

[g] r^2<0.05 between all SNPs except for rs1467385 and rs1296489 (r^2=0.11)

[h] r^2=0.26 between rs2187118 and rs4818835

[l] r^2<0.10 between rs35785446 and rs2187118/ rs35785446 and rs4818835

Results described in the text are highlighted in bold

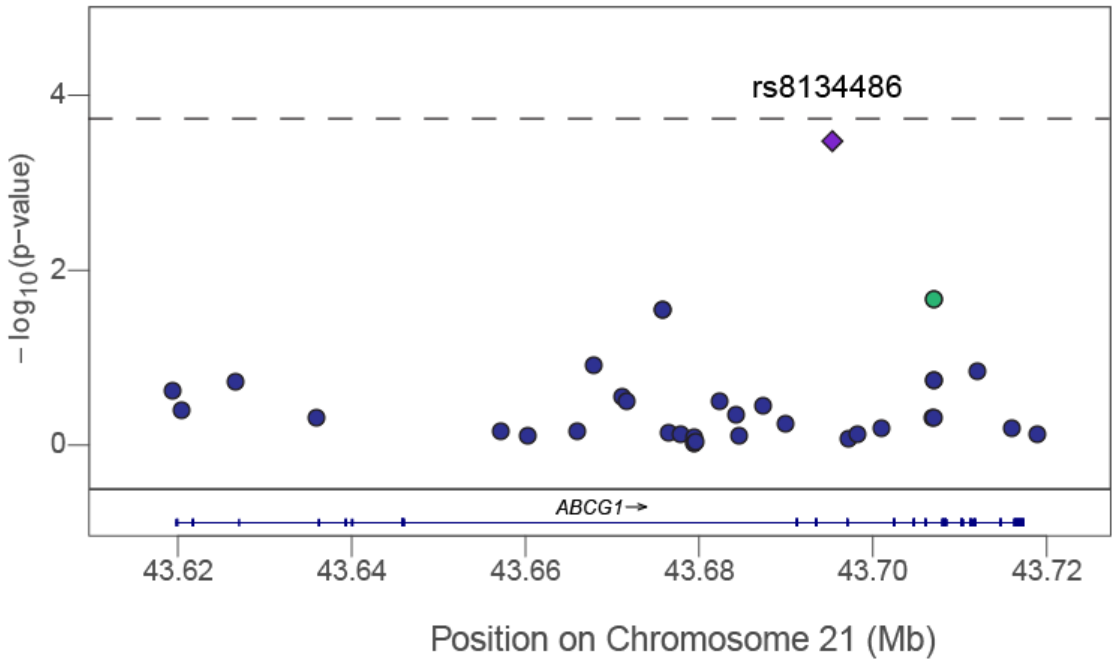


Figure 5.2) Relative Position of rs8134486 in *ABCG1*

In quantitative trait association analyses involving the entire FCHL cohort, the *ABCG1* intronic SNP, rs8134486 (purple diamond) gave the best result (*P*=0.0003). A green dot marks rs4148135 (which returned the lowest *P*-value in the dichotomous triglyceride trait analyses, not shown). All other SNPs in the *ABCG1* interval that were included in the association analyses are marked by blue dots.

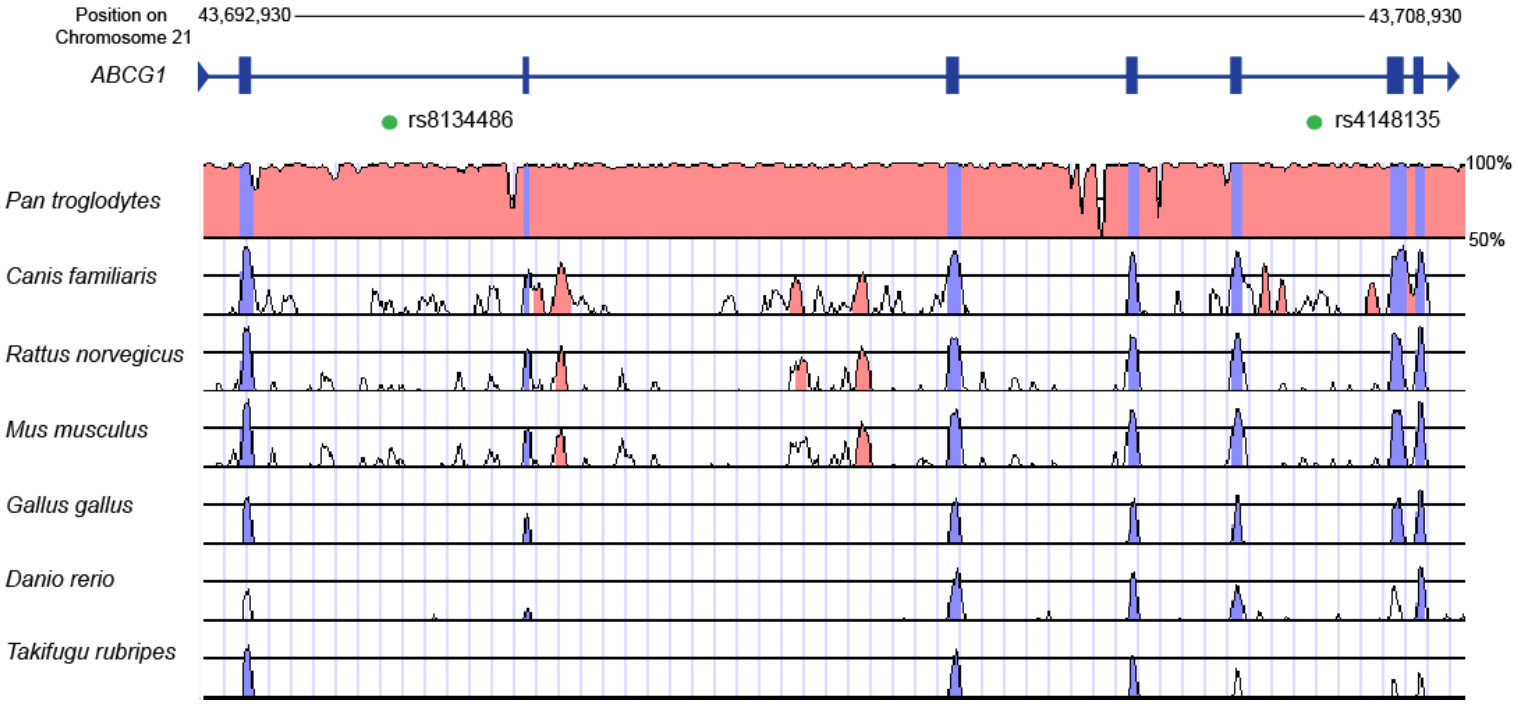


Figure 5.3) Conservation of Human *ABCG1* in other Eukaryotic Species

Comparative analysis of *ABCG1* sequences spanning 43,692,930-43,708,930 on chromosome 21 was completed using VISTA (Frazer et al., 2004) in the UCSC genome browser (URL17). Conservation scale is from 50 to 100% and conservation >75% is shaded in pink. Annotated exons are marked in blue. Intronic SNPs, rs8134486 and rs4148135 are marked by green dots.

For the cholesterol traits (Figure 5.4 and Table 5.3) the lowest *P*-value (*P*=0.0023) was obtained in the dichotomous trait analyses for rs6517654, a tag-SNP residing ~5.4kb upstream of *BACE2.* In the GLGC study, the *P*-value for association with circulating cholesterol levels was *P*=0.8563.

The non-synonymous *LSS* SNP rs35785446 (p.R614W) was nominally associated with both triglyceride (*P*=0.0079) and cholesterol (*P*=0.0074) levels. The minor allele frequency of rs35785446 in the entire FCHL cohort was 0.042, which is identical to that recorded in the 1000 Genomes pilot study (based on genotypes from 60 CEU individuals). However, the minor allele frequency in FCHL probands was 0.065 compared to 0.035 in the family spouses (*P*=0.0377).

This variant causes an arginine to tryptophan change at amino acid 614. It is predicted by SIFT (Ng and Henikoff, 2001) to be deleterious and by Polyphen (Sunyaev et al., 2001) to probably be damaging to the encoded protein. Indeed, cross-species comparisons reveal that a polar amino acid (arginine, lysine or glutamine) is present at this residue in a number of species, incorporating mammals, birds, amphibians and fungi (Figure 5.5). In contrast, the p.R614W substitution replaces this polar residue with a hydrophobic one. Structural analysis reveals that amino acid 614 forms part of an alpha helix which does not form part of the active site of the enzyme (Figure 5.6).

In order to substantiate the nominal association between rs35785446 and lipid levels, the GLGC dataset was examined. Although rs35785446 itself was not included in the GLGC study, data were available for three SNPs (rs8134875, rs12329865 and rs16979001) that are in strong LD (r^2=1, based upon 60 CEU individuals included in the 1000 Genomes pilot study) with rs35785446 (Table 5.4). These SNPs returned *P*-values of *P*=0.0968, *P*=0.0789 and *P*=0.3956, respectively for the association with plasma triglyceride levels and *P*-values of *P*=0.6774, *P*=0.6967 and *P*=0.9358 for association with plasma cholesterol levels in the GLGC dataset.

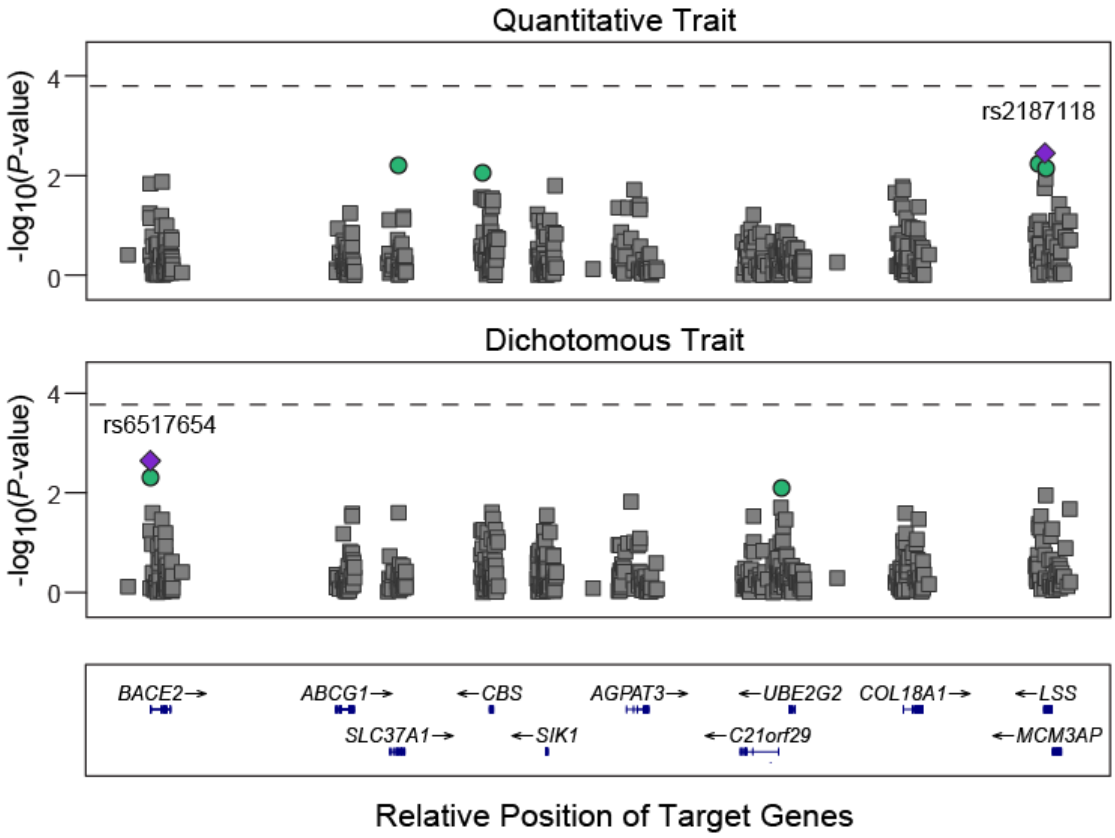


Figure 5.4) Results of Single Marker Association Tests for Directly Genotyped SNPs in the Entire FCHL Cohort – Cholesterol Traits

To test for association between individual SNPs and serum total cholesterol level, considered as a quantitative trait, linear mixed effects models were used. When considering the FCHL-cholesterol trait as a binary outcome (affected=cholesterol ≥95th percentile age- and sex-matched level), logistic regression, via generalised estimating equations was employed. The threshold for significant association is marked by a dotted line. For each analysis, the SNP giving the lowest *P*-value is labelled and is marked by a purple diamond. Other SNPs with an (uncorrected) significance level of $P<0.01$ are indicated by green circles. All other analysed SNPs are denoted by grey squares.

Table 5.3) SNP-Cholesterol Association Results from Single Marker Association Tests, (*P*<0.01for either quantitative or dichotomous analysis)

SNP	Position (bp)	Candidate (Relative Position)	Alleles[a]	MAF	Quantitative Trait (LME)			Dichotomous Trait[d] (GEE)	
					VE (%)[b]	β (SE)[c]	*P*	β (SE)[e]	*P*
rs7276900[f]	42,532,075	*BACE2* (~7.7kb upstream)	C/A	0.29	0.24	0.1121 (0.0589)	0.0572	0.2401 (0.0858)	0.0052
rs6517654[f]	**42,534,328**	***BACE2* (~5.4kb upstream)**	**G/C**	**0.22**	**0.37**	**0.1573 (0.0644)**	**0.0146**	**0.2701 (0.0885)**	**0.0023**
rs2839543	43,964,975	*SLC37A1* (intronic)	C/G	0.02	0.51	-0.5058 (0.1861)	0.0066	-0.6202 (0.2770)	0.0252
rs2839619	44,436,177	*CBS* (~37.1kb downstream, intronic *PKNOX1*)	G/A	0.50	0.47	-0.1389 (0.0532)	0.0091	-0.1492 (0.0778)	0.0551
rs235289	46,141,787	*TSPEAR/UBE2G2* (~10.3kb upstream/~47.0kb downstream)	A/G	0.20	0.13	-0.0810 (0.0654)	0.2152	-0.2167 (0.0819)	0.0081
rs4819208[g]	47,572,244	*LSS* (~36.1 kb downstream, intronic *FTCD*)	A/G	0.25	0.21	0.1698 (0.0617)	0.0059	0.1132 (0.0835)	0.1752
rs2187118[g]	47,611,310	*LSS* (intronic)	C/T	0.41	0.59	-0.1595 (0.0547)	0.0035	-0.1130 (0.0836)	0.1762
rs35785446[g]	**47,614,553**	***LSS* (p.R614W)**	**G/A**	**0.04**	**0.42**	**0.3449 (0.1289)**	**0.0074**	**0.4984 (0.1965)**	**0.0112**

bp=base pairs, GEE=generalised estimating equations, LME=linear mixed effects model, MAF=minor allele frequency, SE=standard error, VE=variance explained

[a] Major allele/minor allele

[b] Proportion of phenotypic variance explained by SNP

[c] For quantitative trait analyses, β represents change in cholesterol level (mmol/l) per copy of the minor allele

[d] Individuals were dichotomised according to affection for FCHL cholesterol trait. Affected=cholesterol levels ≥95th age- and sex- specific percentile values, unaffected= triglyceride levels <95th age- and sex- specific percentile values.

[e] Odds ratio for the FCHL cholesterol trait, per copy of the minor allele, corresponds to e^{β}

[f] r^2=0.68 between rs7276900 and rs6517654

[g] r^2<0.05 between *LSS* interval SNPs

Results described in the text are highlighted in bold

Position				UniProtID	Organism
593	LEAFACMGQTYRDGTACAEVS	**R**	ACDFLLSRQMADGGWGEDFESCEERRYLQSAQ--SQIH	P48449	*Homo sapiens*
593	LEAFACMGHTYHNGVACAEIS	**R**	ACDFLLSRQMADGGWGEDFESCKQRRYVQSAQ--SQIH	P84466	*Bos taurus*
594	LEAFACMGHTYQDGAACAEVA	**Q**	ACNFLLSQQMADGGWGEDFESCEQRRYVQSAR--SQVH	Q8BLN5	*Mus musculus*
594	LEAFACMGHIYQNRTACAEVA	**Q**	ACHFLLSRQMADGGWGEDFESCEQRRYVQSAG--SQVH	P48450	*Rattus norvegicus*
513	LEAFASMQHVYRDGVACREVA	**R**	ACQFLLSKQMADGGWGEDFESCEQRTYVQSST--SQIH	Q5ZIC3	*Gallus gallus*
591	LEAFACMGHTYKEG--CPEII	**R**	ACNFLLSHQMEDGGWGEDFESCEQRRYVQSAG--SQIH	Q0IHW7	*Xenopus tropicalis*
588	LESLSCVKDFYENS---FHSR	**R**	ACDFLVNKQEEDGGWSEGYQSCTDGIWTRHPTG-SQVV	Q96WJ0	*Pneumocystis carinii*
583	LECLRSAGETYENS---EHVR	**R**	GCEFLLSKQRKDGGWSESFQSCEQMTYIEHPTG-SQVV	B0Y4G7	*Aspergillus fumigatus*
595	LEALHTVGLDYESS---SAVK	**K**	GCDFLISKQLPDGGWSESMKGCETHSYVNGEN--SLVV	Q04782	*Candida albicans*
599	LEALHTVGETYENS---STVR	**K**	GCDFLVSKQMKDGGWGESMKSSELHSYVDSEK--SLVV	P38604	*Saccharomyces cerevisiae*
588	TGSLASAGRYYENC---PVQK	**K**	ACEFLLSKQRPDGGWSESYMACVTGVYTETES--SLVT	Q10231	*Schizosaccharomyces pombe*
700	ASALCISREIPDMAN-HPSCV	**R**	LIDFLLSHQNADGGWGEDVTASVRSLWVDNPSG-SQVV	Q964Q1	*Trypanosoma cruzi*
622	IKGMLASGKTYESS---LCIR	**K**	ACGFLLSKQLCCGGWGESYLSCQNKVYTNLPGNKSHIV	Q1G1A4	*Arabidopsis thaliana*
623	IKGLIAASKSYQES---KSIR	**R**	ACEFLLSKQLLSGGWGESYLSCELKVYTNLEGNKSHLV	Q1ERD3	*Lotus japonicus*

Figure 5.5) Cross-Species Conservation at Lanosterol Synthase Residue 614

Multiple-species alignment of LSS protein sequence based upon 14 eukaryotic organisms. Human residue 614 (arginine), which is changed to tryptophan at variant rs35785446 is highlighted in red and the corresponding residue across the different species is boxed.

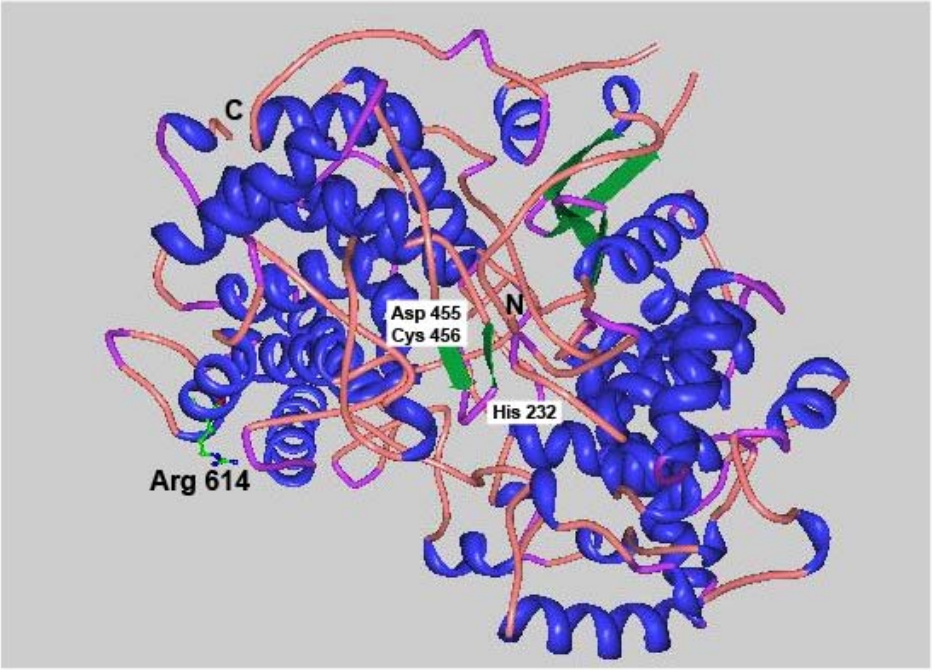


Figure 5.6) Representative Structure of Human Lanosterol Synthase

Lanosterol synthase catalyses the cyclisation of 2,3(*S*)-monooxidosqualene to produce lanosterol. The enzyme consists of two (α/α) barrel domains (blue), connected by loops (pink/purple), as well as three β-structures (green). Arginine 614 forms part of an alpha-helix. Residues important to enzymatic activity are marked. Specifically, the catalytic amino acid Asp 455 is activated by Cys 456 and Cys 533 (not shown for clarity), whilst His 232 is involved in termination of the cyclisation reaction. Amino- (N) and carboxy (C)-terminals are also shown. Adapted from (Thoma et al., 2004).

b) Analyses Adjusted for the Effect of *APOE* Isoform upon Lipid Levels

The three *APOE* isoforms (ε2/ε3/ε4) are known to influence lipid levels (Calandra et al., 2011). Therefore analyses were repeated, including *APOE* isoform as a covariate. Specifically, individuals were defined as *APOE* ε2 carriers (ε2/ ε2 homozygotes or ε2/ ε3 heterozygotes), ε3 homozygotes or ε4 carriers (ε2/ ε4 heterozygotes, ε3/ ε4 heterozygotes or ε4 homozygotes). Table 5.5 shows *APOE* genotype frequencies in the FCHL cohort, according to this classification.

Incorporating this information, no new associations with *P*-values of $P<0.01$ were detected. Moreover, the intronic *ABCG1* SNP, rs8134486, generated the lowest *P*-value in the triglyceride analyses ($P=0.0003$ with and without *APOE* adjustment). In comparison, when considering cholesterol levels, the lowest *P*-value was obtained for rs2839619 ($P=0.005$ compared to $P=0.009$ before *APOE* adjustment). This SNP resides in the fifth intron of *PKNOX1* and was included as a tag-SNP for the *CBS* locus. In the GLGC study, which did not take the *APOE* isoforms into account, the corresponding *P*-value for association with plasma cholesterol levels was $P=0.4544$.

c) Analyses using OSA-Triglyceride Families

No SNPs exceeded the calculated significance threshold of $P<0.00016$ in either the quantitative and dichotomous trait analyses based upon the 102 families contributing to the OSA-triglyceride linkage signal at 51cM (Table 5.6). Overall, the dichotomous trait analyses returned the lowest *P*-values. These were for SNPs rs4148135 ($P=0.0002$) and rs9980195 ($P=0.0009$).

Tag-SNP, rs4148135 resides in the 13th intron of *ABCG1* and is ~11.7kb downstream of rs8134486 ($P=0.0003$ for association with triglyceride levels in the entire FCHL cohort) (Figure 5.2). However, the two SNPs are not in LD ($r^2=0$ on the basis of all genotype data) and thus represent associations at two independent *ABCG1* loci. Cross-species alignments of *ABCG1* sequences provide no evidence that rs4148135 resides in a conserved section of DNA sequence (Figure 5.3). Like rs8134486, rs4148135 was not associated with HDL levels in the FCHL cohort ($P=0.8263$) and, due to low frequency, was not included in the GLGC study (Teslovich et al., 2010).

Although rs9980195 was included as a tag-SNP for the candidate gene *UBE2G2*, it resides 717bp downstream of *PTTG1IP*. In the GLGC meta-analysis, the *P*-value for association between this SNP and plasma triglyceride levels was $P=0.3476$.

d) Analyses using OSA-Cholesterol Families

Finally, quantitative and dichotomous trait association analyses were performed using data for the subset of 47 families contributing to the OSA linkage signal for the cholesterol trait at 51cM (Figure 3.3, Table 5.7). No SNPs surpassed the threshold for significant association. The SNP which achieved the lowest *P*-value (*P*=0.0004, in the dichotomous analyses) was rs2838418, which has a minor allele frequency of 4%. This SNP resides ~32.2kb upstream of *AGPAT3* and ~27.8kb downstream of *RRP*. It acts as a proxy for 12 other SNPs, of which 10 reside in within *RRP*. Examination of the GLGC dataset revealed no evidence for association of this SNP with cholesterol levels (*P*=0.8517).

Table 5.4) GLGC Data for SNPs in Total Linkage Disequilibrium with rs35785446

SNP	Position (bp)	Gene (Relative Position)	MAF[a]	*P* (TG)	*P* (TC)
rs12329865	47,583,506	*C21orf56* (intronic)	0.04	0.0968	0.6774
rs8134875	47,587,940	*C21orf56* (intronic)	0.04	0.0789	0.6967
rs16979001	47,654,301	*MCM3AP* (746bp downstream)	0.04	0.3956	0.9358

MAF=minor allele frequency, TC=total cholesterol, TG=triglyceride

[a] MAF refers to frequency in CEU individuals genotyped as part of the 1000 Genomes Project pilot study

Table 5.5) *APOE* Genotype Frequencies in the FCHL Family Cohort

Classification	Genotype	Count (N)[a]	Frequency
APOE ε2 carriers	ε2/ ε2	21	0.01
	ε2/ ε3	208	0.12
APOE ε3 homozygotes	ε3/ ε3	932	0.55
APOE ε4 carriers	ε2/ ε4	46	0.03
	ε3/ ε4	452	0.27
	ε4/ ε4	36	0.02
Missing genotypes		30	

[a] Data were examined for a total of 1725 individuals

Table 5.6) SNP-Triglyceride Association Results using the Subset of 102 Families Contributing to Linkage Signal at 51cM

SNP	Position (bp)	Candidate Gene (Relative Position)	Alleles[a]	MAF	Quantitative Trait				Dichotomous Trait[e]		
					VE(%)[b]	β (SE)[c]	*P*(Subset)	*P*(All)[d]	β (SE)[f]	*P*(Subset)	*P*(All)[d]
rs2837983[g]	42,616,757	*BACE2* (intronic)	G/A	0.21	0.09	0.0309 (0.0334)	0.3552	0.3634	0.3269 (0.1262)	0.0096	0.5738
rs914187[g]	42,619,749	*BACE2* (intronic)	C/T	0.43	0.11	0.0303 (0.0267)	0.2573	0.7530	0.2987 (0.1031)	0.0038	0.0336
rs4148082[h]	43,619,415	*ABCG1* (383bp upstream)	G/A	0.07	0.86	0.1459 (0.0542)	0.0071	0.2449	0.2368 (0.2362)	0.3161	0.1505
rs9982242[h]	43,671,120	*ABCG1* (intronic)	G/C	0.04	0.98	-0.1706 (0.0661)	0.0098	0.2852	-0.4821 (0.2726)	0.0769	0.9079
rs4148135[h]	**43,706,995**	***ABCG1* (intronic)**	**G/A**	**0.02**	**1.03**	**-0.2826 (0.1032)**	**0.0062**	**0.0219**	**-1.5344 (0.4118)**	**0.0002**	**0.0016**
rs9979100[i]	43,913,370	*SLC37A1* (~6.4kb upstream)	G/A	0.39	0.52	-0.0536 (0.028)	0.0558	0.2273	-0.3155 (0.1169)	0.0069	0.0497
rs228067[i]	43,960,077	*SLC37A1* (intronic)	C/A	0.44	0.89	-0.0739 (0.0279)	0.0081	0.1865	-0.3605 (0.1277)	0.0048	0.0385
rs884979[i]	43,985,452	*SLC37A1* (intronic)	G/A	0.02	1.02	0.2825 (0.1092)	0.0097	0.1571	1.1276 (0.3043)	0.0002	0.0379
rs4920123[i]	43,990,667	*SLC37A1* (intronic)	C/T	0.36	0.89	0.0757 (0.0291)	0.0092	0.0135	0.1762 (0.1367)	0.1974	0.2854
rs733085	45,375,174	*AGPAT3* (intronic)	G/A	0.05	0.78	0.1341 (0.0618)	0.0300	0.4149	0.7201 (0.245)	0.0033	0.0177
rs2329900[j]	46,203,641	*UBE2G2* (intronic)	G/T	0.44	0.95	-0.0741 (0.0277)	0.0074	0.1424	-0.4171 (0.1352)	0.0020	0.0188
rs2838686[j]	46,207,482	*UBE2G2* (intronic)	C/G	0.43	0.54	-0.0563 (0.0269)	0.0367	0.2785	-0.3414 (0.1316)	0.0095	0.0411
rs690626[j]	46,243,319	*UBE2G2* (~21.6kb upstream)	T/G	0.28	0.64	0.0672 (0.0299)	0.0247	0.0241	0.3440 (0.1328)	0.0096	0.0206
rs9980195[j]	**46,268,783**	***UBE2G2* (~47kb upstream)**	**G/A**	**0.47**	**1.11**	**0.0794 (0.0263)**	**0.0026**	**0.0171**	**0.4204 (0.1268)**	**0.0009**	**0.0451**
rs11312[j]	46,270,196	*UBE2G2* (~48.4kb upstream, 3' UTR *PTTG1IP*)	G/A	0.35	0.93	0.0802 (0.0278)	0.0039	0.0080	0.3798 (0.1314)	0.0038	0.0474

SNP	Position (bp)	Candidate Gene (Relative Position)	Alleles[a]	MAF	Quantitative Trait				Dichotomous Trait[e]		
					VE(%)[b]	β (SE)[c]	*P*(Subset)	*P*(All)[d]	β (SE)[f]	*P*(Subset)	*P*(All)[d]
rs84193[j]	46,270,376	*UBE2G2* (~48.6kb upstream, 3' UTR *PTTG1IP*)	G/C	0.36	0.92	0.0803 (0.0278)	0.0038	0.0116	0.3663 (0.1322)	0.0056	0.0583
rs235314[j]	46,271,452	*UBE2G2* (~49.7kb upstream, 3' UTR *PTTG1IP*)	A/G	0.46	1.14	0.081 (0.0271)	0.0028	0.0344	0.4193 (0.128)	0.0011	0.0514
rs61759533	46,875,874	*COL18A1* (p.V144I)	G/A	0.02	1.64	-0.287 (0.0927)	0.0020	0.0701	-0.9744 (0.4948)	0.0489	0.2074
rs11089051[k]	47,576,776	*LSS* (~31.6kb downstream)	G/C	0.07	1.27	-0.1598 (0.0511)	0.0018	0.0122	-0.4033 (0.2333)	0.0839	0.0128
rs2839130[k]	47,591,128	*LSS* (~17.2 kb upstream)	G/A	0.17	0.94	-0.0993 (0.0365)	0.0064	0.9460	-0.3329 (0.1497)	0.0262	0.6149

bp=base pairs, MAF=minor allele frequency, SE=standard error

[a] Major allele/minor allele

[b] Proportion of phenotypic variance explained by SNP

[c] For quantitative trait analyses, β represents change in triglyceride level (natural logarithm mmol/l) per copy of the minor allele

[d] *P*-value for same analysis performed using data from entire cohort

[e] Individuals were dichotomised according to affection for FCHL triglyceride trait. Affected=triglyceride levels ≥90th age- and sex- specific percentile values, unaffected= triglyceride levels <90th age- and sex- specific percentile values.

[f] Odds ratio for the FCHL triglyceride trait, per copy of the minor allele, corresponds to e^{β}

[g] r^2=0.09 between *BACE2* SNPs

[h] r^2=0 between *ABCG1* SNPs

[i] r^2=0.69 between rs997910 and rs228067. r2<0.10 between other *SLC37A1* SNPs

[j] *UBE2G2* SNPs in pairwise LD across interval, r^2 range=0.31-1.00

k r^2=0.32 between rs11089051 and rs2839130

Table 5.7) SNP-Cholesterol Association Results from Single Marker Tests using Subset of 47 Families Contributing to OSA Linkage Signal at 51cM

SNP	Position (bp)	Candidate Gene (Relative Position)	Alleles[a]	MAF	Quantitative Trait				Dichotomous Trait[e]		
					VE(%)[b]	β (SE)[c]	*P*(Subset)	*P*(All)[d]	β (SE)	*P*(Subset)	*P*(All)[d]
rs9636986	42,657,487	*BACE2* (~9.0kb downstream)	C/G	0.04	1.59	-0.7926 (0.3046)	0.0093	0.6194	-0.9124 (0.5074)	0.0722	0.9167
rs3788007	43,706,776	*ABCG1* (Intronic)	G/A	0.22	2.10	0.4098 (0.1506)	0.0065	0.1817	0.4775 (0.1972)	0.0155	0.0262
rs1788484	44,496,941	*CBS* (469bp upstream)	G/A	0.33	2.13	0.3299 (0.1207)	0.0063	0.2102	0.3454 (0.1564)	0.0272	0.8232
rs17004557	44,876,072	*SIK1* (~29.1kb upstream)	A/C	0.31	2.05	0.4048 (0.1293)	0.0018	0.1930	0.4594 (0.1747)	0.0086	0.1847
rs162383	44,882,415	*SIK1* (~35.4kb upstream)	T/G	0.02	0.14	0.1152 (0.1683)	0.1683	0.6148	0.962 (0.2976)	0.0012	0.8762
rs2838418	**45,252,938**	***AGPAT3* (~32.2kb upstream/ *RRP* ~27.8kb downstream)**	**G/A**	**0.04**	**2.52**	**-0.7702 (0.2897)**	**0.0078**	**0.8174**	**-1.3181 (0.3714)**	**0.0004**	**0.8120**
rs13046675	46,778,808	*COL18A1* (~46.3kb upstream)	G/A	0.25	0.79	-0.3672 (0.1381)	0.0078	0.0222	-0.1483 (0.1817)	0.4142	0.6302

bp=base pairs, MAF=minor allele frequency, SE=standard error

[a] Major allele/minor allele

[b] Proportion of phenotypic variance explained by SNP

[c] For quantitative trait analyses, β represents change in cholesterol level (mmol/l) per copy of the minor allele

[d] *P*-value for same analysis performed using data from entire cohort

[e] Individuals were dichotomised according to affection for FCHL cholesterol trait. Affected=cholesterol levels ≥95th age- and sex- specific percentile values, unaffected= triglyceride levels <95th age- and sex- specific percentile values.

[f] Odds ratio for the FCHL cholesterol trait, per copy of the minor allele, corresponds to e^{β}

[g] r^2=0 between rs17004557 and rs162383

5.2.2) Association of Imputed SNPs with FCHL-Lipid Traits

Genotype imputation and subsequent quantitative and dichotomous trait association analyses were conducted using the entire FCHL cohort. In the quantitative triglyceride trait analyses, three distinct clusters of SNPs in the first intron of *TSPEAR* provide support for signals previously detected at directly genotyped SNPs (Table 5.2). Specifically, imputed SNPs in strong LD (r^2>0.8) with the genotyped markers produced three- to eight-fold lower *P*-values in the corresponding triglyceride trait analyses (Table 5.8 and Figure 5.7). Examination of the corresponding triglyceride-level association data from the GLGC study, revealed no *P*-values below *P*=0.05 (Table 5.8).

5.2.3) Summary of Findings from Directly Genotyped and Imputed Data

Table 5.9 summarises the lowest *P*-value directly genotyped and imputed SNPs to emerge from the association analyses presented in this chapter. In brief, all but one of the variants resides close to, or within, a known gene. Four of the seven variants are rare, with a minor allele frequency of <0.05 in the FCHL cohort. One variant causes an amino acid alteration: rs35785446, which causes an arginine to tryptophan change at residue 614 in the LSS protein sequence. The lowest frequency SNPs, rs8134486 (*P*=0.0003) and rs4148135 (*P*=0.0002) were not included in the GLGC dataset. On the basis of both genotyped and imputed SNPs, three independent *TSPEAR* loci demonstrated potential association with triglyceride levels in the FCHL cohort: the lead SNPs were rs458762 (*P*=0.0004), rs34163868 (*P*=0.0009) and rs743501 (*P*=0.0005).

Table 5.8) Clusters of Imputed *TSPEAR* SNPs Associated with Triglyceride Levels in the FCHL Cohort

Directly Genotyped					Imputed[a]				
SNP	Position (bp)	MAF	*P* (FCHL)	*P* (GLGC)	SNP	Position (bp)	MAF	*P* (FCHL)	*P* (GLGC)[b]
rs458178	45,991,288	0.21	0.0014	0.8803	rs458762	45,992,260	0.22	0.0004	na
					rs382478	45,992,140	0.22	0.0004	0.8894
					rs425667	45,992,131	0.22	0.0004	0.9565
					rs465883	45,991,389	0.22	0.0004	0.7733
rs1467385	46,121,817	0.30	0.0044	0.1376	rs34163868	46,124,842	0.30	0.0009	na
					rs34800343	46,124,557	0.28	0.0010	na
					rs1108261	46,128,261	0.29	0.0018	0.0927
					rs878551	46,128,158	0.29	0.0018	0.1209
					rs11700586	46,128,983	0.29	0.0019	na
					rs34538966	46,128,638	0.28	0.0022	na
					rs998421	46,123,355	0.30	0.0036	0.0683
					rs2838643	46,118,197	0.30	0.0040	na
					rs13048736	46,118,986	0.31	0.0041	na
					rs2838642	46,118,157	0.29	0.0043	0.1426
					rs2838647	46,119,770	0.31	0.0044	0.1203
					rs13050276	46,122,724	0.30	0.0045	na

Directly Genotyped					Imputed[a]				
SNP	Position (bp)	MAF	*P* (FCHL)	*P* (GLGC)	SNP	Position (bp)	MAF	*P* (FCHL)	*P* (GLGC)[b]
rs1296489	46,124,881	0.28	0.0045		rs743501	46,127,060	0.32	0.0005	na
					rs2838652	46,124,516	0.32	0.0006	na
					rs762437	46,126,799	0.33	0.0008	na
					rs34120489	46,124,755	0.31	0.0019	na

na=not available

[a] Imputed SNPs are in strong LD (r^2>0.8) with the directly genotyped SNP listed

[b] SNPs marked 'na' were not included in the GLGC study since they are not present on the SNP genotyping panels utilised and are absent from the HapMap 2 reference panel used for imputation in the GLGC cohorts.

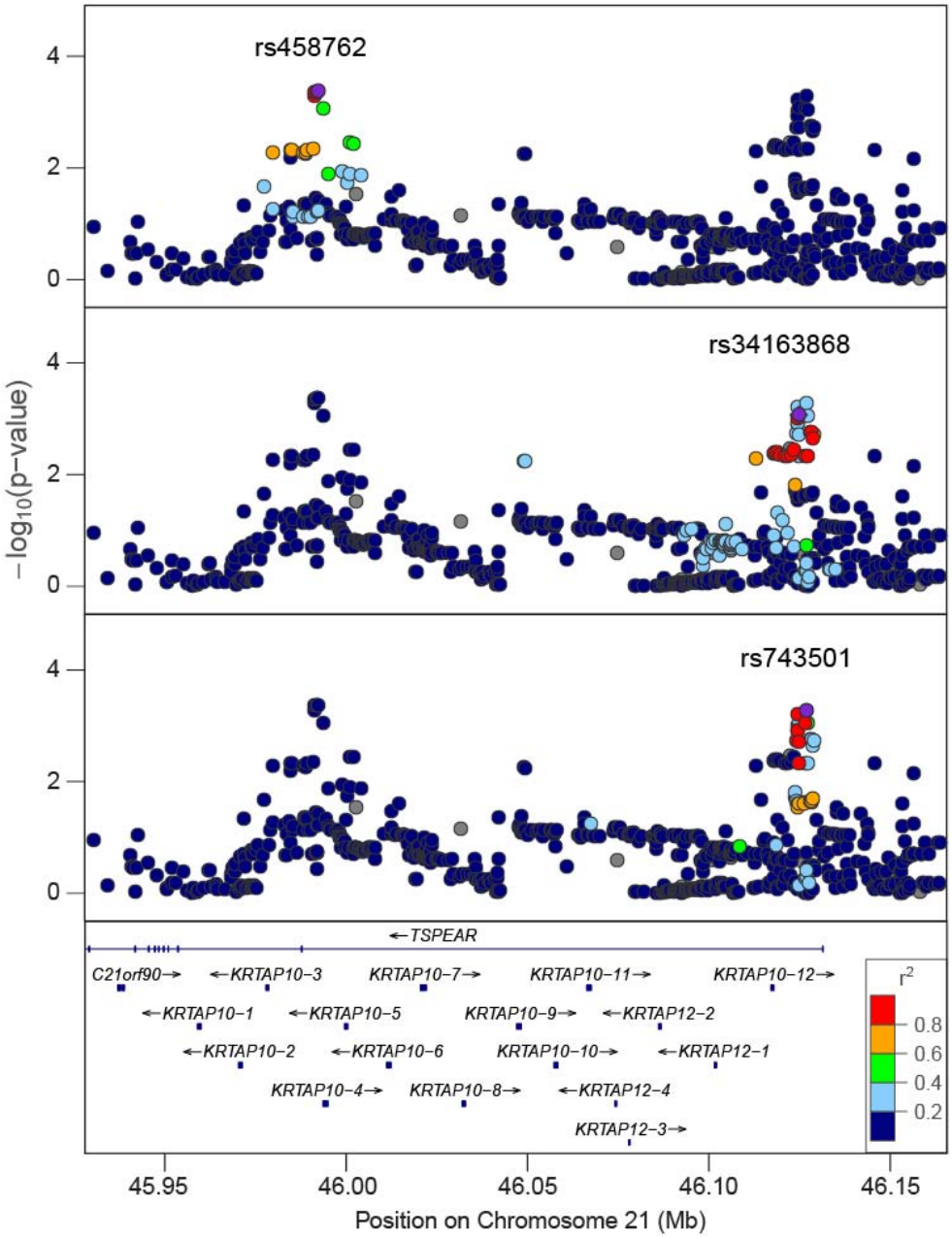


Figure 5.7) Three Clusters of *TSPEAR* SNPs Associated with Triglyceride Levels in the FCHL Cohort

One cluster of SNPs, residing in intron 2 of *TSPEAR,* is associated with triglyceride levels in the FCHL cohort (lowest *P*-value SNP=rs458762, purple dot, *P*=0.0004). Two further clusters of triglyceride associated SNPs, which are not in strong LD with each other, or the intron 2 cluster, reside in *TSPEAR* intron 1 (lowest *P*-value SNPs=rs743501 and rs34163868, purple dots, *P*=0.0005 and *P*=0.0008, respectively).

Table 5.9) Summary of Lowest *P*-Value Markers from Analyses of Directly Genotyped SNPs

SNP	Closest Gene (Relative Position)	MAF	Trait	G/I[a]	Subset	Direction of Effect	*P* (FCHL)[b]	*P* (GLGC)[c]
rs8134486	*ABCG1* (9th intron)	0.01	TG (Q)	G	Entire cohort	+	0.0003	na
rs6517654	*BACE2* (~5.4kb upstream)	0.22	TC (D)	G	Entire cohort	+	0.0023	0.8563
rs35785446	*LSS* (p.R614W)	0.04	TG (Q)	G	Entire cohort	+	0.0039	>0.07[d]
			TC (Q)	G	Entire cohort	+	0.0074	>0.67[d]
rs2838619	*PKNOX1* (5th intron)	0.496	TC (Q)	G	Entire cohort (ApoE-adjusted)	-	0.0052	0.4544
rs4148135	*ABCG1* (13th intron)	0.02	TG (D)	G	102 OSA-TG families	-	0.0002	na
rs9980195	*PTTG1IP* (717bp downstream)	0.47	TG (D)	G	102 OSA-TG families	+	0.0009	0.3476
rs2838418	*RRP* (27.8kb downstream)/ *AGPAT3* (~32.1kb upstream)	0.04	TC (D)	G	47 OSA-TC families	-	0.0004	0.8517
rs458762	*TSPEAR* (1st intron)	0.22	TG (Q)	I	Entire cohort	-	0.0004	na
rs34163868	*TSPEAR* (1st intron)	0.30	TG (Q)	I	Entire cohort	+	0.0009	na
rs743501	*TSPEAR* (1st intron)	0.32	TG (Q)	I	Entire cohort	-	0.0005	na

MAF=minor allele frequency, OSA=ordered-subset analysis, TC=(total) cholesterol, TG=triglyceride

[a] Data relates to directly genotyped (G) or imputed (I) SNPs

[b] *P*-value from LME or GEE analysis in subset of FCHL cohort specified

[c] *P*-value for corresponding analysis in GLGC meta-analysis. SNPs marked 'na' were not captured in the GLGC study

[d] *P*-values are taken from SNPs in total LD with rs35785446 (based upon CEPH individuals in 1000 Genomes Project Pilot)

5.2.4) Analysis of GLGC Data Across the 3.8Mb FCHL Linkage Interval

As a complementary approach to discover variants that might contribute to the 21q2.3 FCHL linkage signal(s), data for 3867 GLGC meta-analysis SNPs (genotyped or imputed) residing within the 3.8Mb linkage interval from were examined (Teslovich et al., 2010). No SNPs within the 3.8Mb interval, which spans from 43.3-47.1Mb on chromosome 21q22.3, were associated with plasma lipid levels at or below the declared threshold for genome-wide significance ($P<1x10^{-8}$). The ten SNPs returning the lowest *P*-values for association with plasma triglyceride and cholesterol levels are shown in Tables 5.10 and 5.11.

The lowest *P*-value SNP for association with plasma triglyceride levels in the GLGC cohorts was recorded at rs9984563 (*P*=0.0005). This SNP resides in the first intron of *PKNOX1* and was genotyped in the FCHL cohort as a tag-SNP for the *CBS* interval. In the corresponding analysis in the FCHL cohort the *P*-value for rs9984563 was *P*=0.2473.

The top-ten triglyceride-associated SNPs also include four (rs1883042, rs9647234, rs2838629, rs4818950) out of 245 SNPs located within the ~143kb first intron of *TSPEAR*. These four SNPs share moderate pairwise LD with each other (Figure 5.8) and also share moderate pairwise LD with the cluster of imputed, intron one *TSPEAR* SNPs, centred upon rs34163868, that were nominally associated with triglyceride levels in the FCHL cohort (Table 5.8, Figure 5.7, Figure 5.8).

The strongest association with plasma cholesterol levels in the GLGC study was recorded at rs926196 (*P*=0.0008). This SNP resides in the first intron of *TSPEAR* and was directly genotyped in the FCHL cohort as a *TSPEAR* tag-SNP. In the corresponding association analysis in the FCHL cohort, this SNP returned a *P*-value of *P*=0.3351.

Table 5.10) Lowest *P*-Value SNPs within the 21q22.3 FCHL Linkage Interval from the GLGC Meta-Analysis: Triglyceride Levels

SNP	Position (bp)	Relative Position	Alleles[a]	N[b]	*Z*-Score	*P* (GLGC)	*P* (FCHL)
rs9984563	**44,423,653**	***PKNOX1* (intronic)**	**T/C**	**91,949**	**3.500**	**0.0005**	**0.2473[c]**
rs884979	43,985,452	*SLC37A1* (intronic)	A/G	41,729	-3.247	0.0012	0.1571[c]
rs1883042	46,101,022	*TSPEAR* (intronic)	T/C	95,070	-3.174	0.0015	0.2067[d]
rs2122867	44,739,481	~147kb downstream *CRYAA*/ ~95kb downstream *SIK1*	T/C	84,957	-3.155	0.0016	na
rs235320	46,277,214	*PTTG1IP* (intronic)	A/G	95,070	-3.066	0.0022	0.4784[d]
rs2838412	45,249,506	~24kb downstream *RRP*/ ~36kb upstream *AGPAT3*	A/T	93,526	3.056	0.0022	0.4701[d]
rs9647234	46,098,545	*TSPEAR* (intronic)	T/C	95,070	-3.056	0.0022	0.4195[c]
rs2838629	46,097,922	*TSPEAR* (intronic)	A/G	95,070	-3.043	0.0023	0.2573[d]
rs4818950	46,057,806	*TSPEAR* (intronic)	A/G	95,070	3.024	0.0025	0.0766[d]
rs2167392	44,739,255	~147kb downstream *CRYAA*/ ~95kb downstream *SIK1*	A/G	87,227	-2.973	0.0030	na

na=not applicable (SNP not captured in FCHL cohort)

[a] Effect allele/other allele

[b] N=number of individuals included in analysis. Maximum possible N= 96,598

[c] SNP was directly genotyped in FCHL cohort

[d] SNP was imputed in FCHL cohort

Table 5.11) Lowest *P*-Value SNPs within the 21q22.3 FCHL Linkage Interval from the GLGC Meta-Analysis: Cholesterol Levels

SNP	Position (bp)	Relative Position	Alleles[a]	N[b]	*Z*-Score	*P* (GLGC)	*P* (FCHL)
rs926196	**46,115,254**	***TSPEAR* (intronic)**	**A/G**	**69,623**	**3.364**	**0.0008**	**0.3351[c]**
rs2236451	46,876,083	*COL18A1* (intronic)	A/G	98,656	3.216	0.0013	0.1635[d]
rs9985044	46,854,982	*COL18A1* (intronic)	T/C	98,656	3.176	0.0015	0.1080[c]
rs225424	43,746,695	~20kb downstream *TFF2*/ ~11kb upstream *TFF3*	A/G	98,656	3.103	0.0019	na
rs9977482	46,899,119	*COL18A1* (intronic)	T/C	98,656	-3.048	0.0023	0.1114[d]
rs9979601	46,899,678	*COL18A1* (intronic)	C/G	98,656	-3.023	0.0025	0.1050[d]
rs10775679	46,389,752	*C21orf70* (intronic)	T/C	94,354	-2.893	0.0038	na
rs365163	43,345,207	*C2CD2* (intronic)	A/G	94,472	2.877	0.0040	na
rs4818998	46,381,605	*C21orf70* (intronic)	T/C	94,354	2.866	0.0042	na
rs9974593	46,379,636	*C21orf70* (intronic)	T/C	94,354	-2.864	0.0042	na

na=not applicable (SNP not captured in FCHL cohort)

[a] Effect allele/other allele

[b] N=number of individuals included in analysis. Maximum possible N=100,184

[c] SNP was directly genotyped in FCHL cohort

[d] SNP was imputed in FCHL cohort

	rs34163868	rs1883042	rs9647234	rs2838629	rs4818950
rs34163868		0.75	0.93	1.00	0.88
rs1883042	0.36		1.00	1.00	0.66
rs9647234	0.39	0.69		1.00	0.55
rs2838629	0.22	0.34	0.50		0.67
rs4818950	0.19	0.16	0.16	0.42	

D’

r^2

Figure 5.8) Linkage Disequilibrium between *TSPEAR* SNPs

On the basis of genotype data from the 60 CEU individuals that formed part of the 1000 Genomes Pilot Study, *TSPEAR* SNPs, displaying some association with triglyceride levels in both the GLGC and FCHL cohort, are in moderate pairwise linkage disequilibrium

5.3) Discussion

Although none of the data presented in this chapter provide evidence for association between directly genotyped or imputed SNPs and lipid levels which surpassed the calculated significance threshold of $P<0.00016$, a number of the findings warrant further discussion.

Firstly, one result of potential interest was the nominal association of rs35785446, a non-synonymous coding variant that causes an arginine to tryptophan substitution in the 20th *LSS* exon, with both increased triglyceride and cholesterol levels in the FCHL cohort. The frequency of this variant was also increased in FCHL probands compared to spouse controls (5.2.1a).

Since LSS is an enzyme of the cholesterol biosynthetic pathway, one would expect any variant associated with increased lipid levels to represent a gain-of-function. To support this assumption, functional studies in miniature pigs fed an LSS *inhibitor*, plasma total cholesterol was decreased by 20% ($P<0.007$) and plasma lathosterol (which reflects cholesterol synthesis) was decreased by 57% ($P<0.001$) (Telford et al., 2005). In the same study however, plasma triglyceride levels were unchanged (Telford et al., 2005).

The substitution of a polar, hydrophilic amino acid within an alpha-helix, for one which is hydrophobic is predicted to disrupt protein structure. Indeed in the autosomal dominant condition, familial partial lipodystrophy, an arginine to tryptophan substitution in the zinc finger DNA binding domain of *PPARG* (p.R194W), prevented DNA binding and transcriptional activation by the protein, *in vitro* (Monajemi et al., 2007). Although the amino acid at LSS residue 614 does not form part of the enzyme catalytic site (Thoma et al., 2004), disruption of the protein structure could still, for example, influence substrate access to the enzyme. LSS can catalyse both the conversion of 2,3(*S*)-monooxidosqualene to lanosterol (cholesterol pathway) or the conversion of 2,3;22,23-diepoxysqualene to 24(S),25-epoxylanosterol (epoxysterol pathway). Since the end product of the epoxysterol pathway, 24(S),25-epoxycholesterol, is a potent activator of liver-X-receptors (LXRsy) (Lehmann et al., 1997), this route can lead to decreased cholesterol levels through, for example, increased cholesterol efflux via ABCG5/ABCG8 (Repa et al., 2002). In contrast, production of lanosterol would favour increased cholesterol synthesis.

A second noteworthy result was the nominal association of rs2839619 with cholesterol levels in the entire FCHL cohort (*P*=0.009 before adjustment for the effect of APOE ε2/ε3/ε4 compared to *P*=0.005 after incorporation of this covariate). This SNP was also associated with plasma total cholesterol levels in a recent GWAS involving 944 adult inhabitants of the Croatian island of Korčula (Zemunik et al., 2009). In the Croatian study, the rs2839619 G allele (frequency=0.396) was associated with increased cholesterol with an uncorrected significance of $P=7.6\times10^{-6}$ (β=0.127±0.028(SE)) and in the current study a similar magnitude of effect was observed (β=0.1495±0.0534(SE)).

Although rs2839619 was included as a tag-SNP for *CBS* in the FCHL association study, it resides in the fifth intron of the gene encoding PBX/knotted homeobox 1 (*PKNOX1)*. PKNOX1 is homeodomain transcription factor expressed in a variety of tissues including heart, liver and particularly skeletal muscle (Chen et al., 1997). It is implicated in glucose metabolism due to its interaction with and stabilising effect upon p160, a repressor of PPAR-gamma coactivator-1α (Oriente et al., 2008; Soyal et al., 2006) however, it is unclear as to how *PKNOX1* would influence circulating cholesterol levels.

In future studies, it may be worthwhile extending the SNP coverage of *PKNOX1* in the FCHL cohort. In order to capture 100% of the gene (based upon HapMap variants with a minor allele frequency of >5%) this would require six more SNPs to be genotyped. In addition, one could substantiate this finding by examining the association between rs2839619 and plasma cholesterol levels in other FCHL cohorts and by performing functional studies such as determining the impact of *PKNOX1* overexpression/silencing on cholesterol biosynthesis.

The association of a cluster of 14 SNPs, in intron one of *TSPEAR*, with plasma triglyceride levels in the FCHL cohort seems worthy of further investigation. In this cluster, the imputed SNP rs34163868 returned the lowest *P*-value of *P*=0.0009 for association with triglyceride levels in the entire FCHL cohort. Importantly, this SNP shares some pairwise LD (r^2=0.36 and r^2=0.39) with rs1883042 and rs9647234, which were the third and seventh ranking loci, in the 43.3-47.1Mb region, in the triglyceride analyses of the recent GLGC study (*P*=0.0015 and *P*=0.0022, respectively). Even though rs9647234 (genotyped) and rs1883042 (imputed) were not associated with triglyceride levels in the FCHL cohort (Table 5.10), this result

could indicate that another variant(s) this interval is involved in modulating plasma triglyceride levels.

Fourthly, the analyses also show that four of the seven lowest *P*-value, directly genotyped SNPs presented in Table 5.9, have a minor allele frequency of <5%. Two of these low-frequency SNPs, which were not included in the GLGC meta-analysis, reside in the same gene, *ABCG1.* However, caution in the interpretation of these results is required. In a recent simulation study involving 133 myocardial infarction cases and 716 controls, permutations revealed that low frequency SNPs with a minor allele frequency of 5% had significantly more false-positive associations than more common SNPs (MAF=25% and 50%) at both the $\alpha \leq 10^{-4}$ and $\alpha \leq 10^{-5}$ thresholds (*P*<0.03 and *P*<0.05, respectively) (Tabangin et al., 2009). Also, in the Framingham Heart Study 100K GWAS, the authors attributed an excess of positive results from GEE analyses in their initial results to low frequency SNPs (Cupples et al., 2007). Nonetheless, in future studies it would be important to check functional relevance of these low frequency variants.

Finally, the rationale for the inclusion of *BACE2* in this study was, in part, the association of the intronic SNP, rs914186 with triglyceride/HDL ratio in the Diabetes Genetics Initiative GWAS ($P=7.6 \times 10^{-5}$) (Saxena et al., 2007). In the FCHL cohort this finding was not replicated: *P*-values of *P*>0.79 were returned in all analyses.

Considering the lack of strong association signals between any tested SNPs and the lipid phenotypes of FCHL reported in this chapter, it is important to address two issues. Firstly, was the experimental approach justified and secondly, what factors may have contributed to the lack of observed association.

In the current study, a targeted approach, specifically focussed upon the 21q FCHL linkage interval(s) described in Chapter 3, was employed. Justification for this method, and an example directly applicable to FCHL, is provided by the targeted association study of Lee *et al*., which sought to determine the variants underlying the 16q23-q24 linkage interval for low HDL-cholesterol levels (Lee et al., 2008a). This region was previously implicated in the regulation of HDL-cholesterol levels through a genome-wide linkage screen involving Finnish and Dutch FCHL families, with a multipoint NPL of 2.7 recorded at 99.4cM/79.04Mb (Pajukanta et al., 2003). Evidence supportive of this finding was also demonstrated in Mexican Americans of the San Antonio Family Heart Study (Mahaney et al., 2003) and French-Canadian

families ascertained for low HDL-cholesterol levels (Dastani et al., 2006). In the follow-up, targeted association study of Lee *et al*., a total of 1412 coding and tag-SNPs were genotyped, across the 12.4Mb 1-LOD support interval, in 51 dyslipidemic Finnish families (33 FCHL and 17 low HDL families) (Lee et al., 2008a). Based on the most promising results from initial analyses performed using FBAT, 25 of these SNPs (15 SNPs with $P<0.01$ and 10 nearby SNPs with $P<0.05$) were then genotyped in 24 Dutch FCHL families, 28 French-Canadian low HDL families as well as 252 low HDL cases and 223 controls. Using the combined dataset, region-wide significant association was found between the WW-domain-containing oxidoreductase gene (*WWOX*) tag-SNP, rs2548861 and low HDL-C ($P=1x10^{-6}$) (Lee et al., 2008a).

Thus, the study of Lee *et al*., successfully implemented a two-stage design in which weakly associated SNPs of interest ($P<0.01$) in stage one were selected for genotyping in a larger dyslipidemic sample in stage two. Only in the combined sample was an association surpassing the calculated significance threshold identified. By analogy, the current study could be thought of as a discovery, stage one association analysis. Given the success of the Lee *et al*., approach, it would now seem prudent to genotype the lowest *P*-value SNPs, presented in Table 5.9, in other FCHL cohorts. These could include those drawn from Finnish (Pajukanta et al., 1998), Dutch (Dallinga-Thie et al., 1997) or Mexican (Huertas-Vazquez et al., 2005) populations. However, two potential complications with this approach require attention. Firstly, the FCHL diagnostic criteria differ between cohorts, thus genetic heterogeneity may be increased. For example, the Finnish cohort does not exclude people with type 2 diabetes whilst the Dutch cohort uses raw lipid levels (cholesterol >6.5mmol/l, triglyceride >2.3mmol/l and apoB >120mg/dl) rather than age- and sex-specific percentile values to determine hyperlipidemia (Shoulders et al., 2004). Secondly, by including individuals from different populations in the study group, one would need to implement the correct tests and adjustment for the presence of population stratification, in order to avoid potentially false-positive results.

With regard to the issue of power for detecting association between genotyped or imputed SNPs and FCHL-lipid traits, power calculations indicated that the FCHL cohort was well-powered to detect variants accounting for >1.5% of the total trait variance, or conferring an odds ratio of ≥1.75. Whilst compared to the population-based GWAS of Kathiresan *et al.,* where each of 11 triglyceride-associated loci explained <1% of the trait variance (~7.5% of the total variance in triglyceride level in sum) the current study may seem underpowered (Kathiresan et al., 2009b), one

should note that in a different context, the same SNPs may have a much greater phenotypic effect. For example, another recent GWAS involving 463 unrelated hypertriglyceridemic individuals and 1197 controls demonstrated that by studying the phenotypic extremes, rather than the general population, much greater effect sizes were observed. In the hypertriglyceridemic cohort, each copy of the rs964184 risk allele was associated with a 3.28-fold increase in susceptibility to hypertriglyceridemia, whereas in the general population, each copy of the same allele is associated with only a 0.2mmol/l increase in plasma triglyceride level (Johansen and Hegele, 2011; Johansen et al., 2010; Teslovich et al., 2010). It therefore seems reasonable that whilst strong association was not detected on chromosome 21, should a GWAS of FCHL be conducted, the cohort would provide sufficient power to detect some SNP-phenotype associations, under certain genetic models.

It should also be noted, that the lack of 'significant' association signals in the current study could be due to the fact that the significance threshold applied was in fact too conservative. Indeed, some investigators have questioned the validity of using *P*-values alone to determine the significance of a genetic association (Stephens and Balding, 2009; Wacholder et al., 2004; Wakefield, 2009) since despite widespread use, this approach fails to take account of the power of the test as well as the prior probability that the association between SNP and phenotype is real. Instead, these researchers suggest that a Bayesian approach should be applied.

The posterior probability of association is the Bayesian equivalent of a *P*-value and is calculated in three stages (Stephens and Balding, 2009). The Bayes factor (BF) is the ratio of the probability of the data under the alternative (H_1, association) and null hypotheses (H_0, no association) given by:

$$BF = \frac{P(Data|H_1)}{P(Data|H_0)}$$

To use the BF to measure the posterior probability of association, the prior odds are also required. Determination of prior probabilities is a complex issue (Jiang et al., 2011). On a genome-wide scale, prior probabilities of between 10^{-4} and 10^{-6} have been suggested (Wacholder et al., 2004; Wellcome Trust Case Control Consortium, 2007), based upon, for example, 1,000,000 genome-wide SNPs and a phenotype resulting from the effects of alleles at ten of these loci. However, in regions of the genome highlighted by linkage studies, prior probabilities are likely to be higher and should incorporate and functional data available (Wacholder et al., 2004). Of note, a Bayesian approach to family-based association was recently described (Naylor et al.,

2010) and future collaboration to apply this methodology to the FCHL cohort would seem a logical extension to the current study.

A further analytical consideration relates to the single-locus approach employed to examine these data. The current study examined the association of each SNP, one at a time, with the FCHL lipid phenotypes. However, this method would not be appropriate to detect an association arising from the interaction of two or more loci (Jiang et al., 2010b). In order to detect these potential epistatic effects, statistical methods that consider SNP-SNP interactions are now being developed (Hemani et al., 2011; Jiang et al., 2010b; Li et al., 2011; Zhang et al., 2010b). Conceptually, the simplest of these approaches, is the model-free exhaustive pairwise epistasis search, which considers the interaction between each SNP with all other SNPs. However, on a genome-wide scale involving, for example, 300,000 markers, such an approach would involve ~4.5×10^{10} tests, which would be both computationally demanding and present a severe multiple-testing burden (Hemani et al., 2011). One way to address these issues is to apply a two-stage approach, whereby SNPs are filtered according certain criteria prior examining interactions (Wu et al., 2010). However, Zhang et al., presented data from a study of colon cancer, whereby a two-stage approach led to significant interactions being missed (Zhang et al., 2010c). Therefore, the major aim of algorithm development is to improve analysis efficiency, without inappropriately discarding SNPs from analyses. Once algorithms are sufficiently advanced, their application to the FCHL cohort could be justified, using available SNP data from chromosomes 8, 11 and 21.

Given the targeted approach adopted in the current study, it is possible that the variants that underlie the 21q22.3 FCHL linkage signal(s) do not reside within the candidate gene intervals studied. After genotype imputation, genotype (dosage) data was available for variants encompassing ~34% of the 3.8Mb linkage interval defined in Chapter 3. Although results of the GLGC meta-analysis do not suggest stronger candidate genes than those already targeted, it is possible that the genetic architecture of the region may differ in FCHL patients, compared to that of the general population.

Should one decide to continue evaluating the role of common variants within the remaining 66% of the linkage interval in the aetiology of FCHL, a number of approaches could be feasible. The first could be immediate further genotyping. To capture the entire remaining interval at $r^2\geq0.8$ would require ~750 SNPs to be

genotyped. A more cost effective option would be to select SNPs for genotyping at a much lower density and the use imputation to 'fill in the gaps'. For example, Anderson *et al*, demonstrated that using imputation, the percentage coverage at r^2=0.8 of the lower density Illumina HumanHap 300 SNP genotyping array could be increased from 77 to 81% compared to the 87% coverage provided by the higher density but (at the time) more expensive HumanHap500 array (Anderson et al., 2008). Also, as described in Chapter 4.2 imputation increased the coverage of *SIK1* and *TSPEAR* from 64% to 97 and 98%, respectively. A third option to improve the cost effectiveness of future genotyping would be to employ a means of identifying which of the remaining genes within the interval may have a previously unsuspected role in lipid metabolism. This approach is discussed further in Chapter 6.

Finally, based on the results of the current study and the 3867 SNPs, within the linkage interval, that were included in the GLGC meta-analysis (Teslovich et al., 2010), which both focussed upon common variants with a minor allele frequency of at least 1%, it seems plausible that multiple rare variants could contribute the 21q22.3 FCHL linkage signal. In order to identify rare variants that could underlie the triglyceride and cholesterol traits of FCHL, one could sequence some, or all, of the genes within the 3.8Mb linkage interval, using next-generation sequencing. At the start of this project, next-generation sequencing technology was not sufficiently advanced to allow even one gene to be screened in sufficient numbers of FCHL patients and controls to ascertain mutations of functional importance. However, recent technological advances and cost reductions mean that such an approach is now feasible and as a result, an important future direction of this work would be to sequence the 21q22.2-22.3 FCHL candidate genes, in patient and control groups. Based upon the requirement to establish the functional relevance of any sequence variants found, priority would be given to *CBS*, *COL18A1* and *UBE2G2* since in these genes, a loss-function-mutation would be expected to lead to hyperlipidemia (4.2.1).

Chapter 6 – Identifying Further FCHL Candidate Genes Using Expression Datasets

6.1) Introduction

The integration of gene expression data with genetic data provides a powerful approach to identifying the genes/variants that underlie complex traits. In the case of chronic obstructive pulmonary disease (COPD), for example, linkage analysis was used localise a chromosome 2q susceptibility locus and then gene expression data were used to identify a single candidate susceptibility gene within the interval. Next, both family-based and case-control association analyses demonstrated significant association between *SERPINE2* SNPs and COPD phenotypes (Demeo et al., 2006).

In detail, linkage of an intermediate phenotype of COPD (the ratio of forced expiratory volume at one second to forced vital capacity) to chromosome 2q was established using 72 pedigrees, ascertained through probands with a severe, early-onset form of the disease (Palmer et al., 2003; Silverman et al., 2002). The risk gene was identified as a biological candidate by examining the mRNA levels of genes residing within the linkage interval during mouse lung development (Demeo et al., 2006). Subsequently, *SERPINE2* mRNA levels were shown to be associated with pulmonary-function measurements in lung tissue from COPD cases (n=18) and controls (n=16). Finally, family-based association analysis involving 48 SNPs across the *SERPINE2* gene demonstrated significant association with COPD-related phenotypes at 18 of these variants, five of which were replicated in a case-control cohort comprising 304 smoking-related COPD patients and 441 unaffected control smokers (Demeo et al., 2006). Association between *SERPINE2* and COPD was subsequently replicated in Norwegian (Sorheim and Gulsvik, 2008) and Chinese (An et al., 2011) case-control cohorts and in 635 families of the International COPD Genetics Network (Zhu et al., 2007a).

Another integrative approach by which expression data may be utilised to improve our understanding of complex traits is by correlating SNP genotypes with information on transcript levels. When changes in mRNA levels are correlated with local (*cis*) or more distant (*trans*) genetic variants, the variant is designated as an expression quantitative trait locus (eQTL). Importantly, recent eQTL studies have shown that SNPs associated with complex traits in GWASs are more likely to influence mRNA levels than non-associated SNPs, of equivalent frequency, from the same genotyping

panel (Greenawalt et al., 2011; Nica et al., 2010; Nicolae et al., 2010). Thus, the eQTL approach can attach functional relevance to association study findings, which is especially important for 'disease-associated' SNPs residing in non-coding regions of the genome.

One of the largest eQTL studies performed to date involved liver, adipose and stomach tissue samples collected from a patients undergoing Roux-en-Y gastric bypass (Greenawalt et al., 2011). Correlations between genome-wide SNP data and mRNA levels were determined for the 950 study participants, identifying ~25,000 *cis* and *trans* SNPs associated with expression of ~10,000 distinct genes. In addition, these data demonstrated that SNPs previously associated with complex phenotypes were also associated with expression of functionally relevant genes. For example, the HDL-cholesterol associated SNP rs1532085, was significantly associated with liver *LIPC* mRNA levels (P=1.56x10^{-26}), despite residing ~20kb upstream of the gene. Similarly, the LDL-cholesterol associated SNP rs646776 was associated with liver mRNA levels of three proximal genes (*SORT1, P*= 5.20x10^{-88}; *PSRC1, P*=3.05x10^{-86} and *CELSR2, P*=6.27x10^{-68}) (Greenawalt et al., 2011). In an earlier study which performed expression profiling using 427 human liver samples, rs599839 (r^2~0.9 with rs646776) was also associated with mRNA levels of these three genes (*SORT1, P*= 1.53x10^{-36}; *PSRC1, P*=2.17x10^{-53} and *CELSR2, P*=4.31x 10^{-23}) (Schadt et al., 2008).

A smaller scale integrative approach was recently applied to FCHL, where transcript levels in adipose tissue samples, taken from 38 Mexican FCHL cases and 32 normolipidemic controls, were correlated with the genotype of rs3737787. This SNP resides in the 3' untranslated region (UTR) of the transcription factor *USF1*, an FCHL-associated gene located within the chromosome 1q21-q23 FCHL linkage interval (Pajukanta et al., 1998; Plaisier et al., 2009). In the Mexican dataset, the transcript levels of 972 genes were correlated with rs3737787 genotype. This list of genes showed significant overlap (n=277, P=3×10^{-5}) with 2897 genes previously shown to be regulated by USF1 and with the 2189 genes (n=245, P=0.003) that were differentially expressed between the FCHL cases and control individuals. Using complex network analysis the authors then determined that a group of these genes, enriched for GO Biological Process categories Lipid Metabolic Process and Cellular Lipid Metabolic Process were associated with FCHL component traits (total cholesterol correlation=0.28, P=0.02; triglyceride correlation=0.43, P=2×10^{-4}; ApoB correlation=0.32, P=0.009) and rs3737787 genotype (correlation=0.33, P=0.006) (Plaisier et al., 2009).

Finally, expression data from animal- and cell-based models may help pinpoint disease genes. For example, a transcriptional profiling approach led to the identification of *ABCG5* as one of the genes that causes sitosterolemia (Berge et al., 2000). In their study, Berge and colleagues hypothesised that since treating mice with the LXR-agonist T0901317 led to a marked decrease in cholesterol absorption (Repa et al., 2000), this treatment would also alter the expression of the gene(s) causing sitosterolemia. The rationale underlying this hypothesis being that sitosterolemia is characterised by *excessive* sterol absorption. Thus, when they treated C57BL/6 mice with T0901317 (50 mg/kg), microarray analyses identified that hepatic and intestinal levels of a transcript later ascribed to the *ABCG5* gene were 2.5-fold higher than those of untreated mice (Berge et al., 2000).

A similar expression-based approach was employed in identifying that myosin light chain interacting protein (MYLIP), plays a role in modulating LDL-cholesterol levels, facilitating LDLR degradation through its function as an E3 ubiquitin ligase (Zelcer et al., 2009). In this study, the authors first established that treatment of HepG2 cells with the LXR-agonist GW3965 led to a rapid, dose-dependent decrease in LDLR protein levels. A similar effect was also observed when human SV589 fibroblasts were treated with GW3965. In order to investigate the mechanism mediating this reduction, Zelcer and colleagues then performed a transcriptional profiling experiment using RNA taken from BV2 microglia-like cells treated with the LXR agonist GW3695 for 24-hours. Subsequent analyses identified that levels of *MYLIP* RNA were increased 5.61-fold in treated versus control cells. This increase in mRNA levels was confirmed in primary mouse hepatocytes and macrophages, in HepG2 cells and in the spleen, intestine and adrenals of mice given a daily oral gavage of 40 mg/kg GW3956 for 3 days (Zelcer et al., 2009). Coinciding with the RNA studies of Zelcer et al., three recent GWASs have revealed significant associations between three independent *MYLIP* locus SNPs, and plasma LDL-cholesterol levels (Chasman et al., 2009; Teslovich et al., 2010; Waterworth et al., 2010).

An alternative approach to that employed in the studies described above would be to target the peroxisome proliferator-activated receptor (PPAR) subfamily nuclear receptors. In common with LXRs, PPARs form heterodimers with retinoid X receptors and bind to specific DNA response elements in target metabolic genes, activating or repressing their transcription, in response to ligand binding (Li and Glass, 2004)

The PPAR subfamily consists of PPARα, PPARβ/δ and PPARγ. PPARα, the first PPAR to be cloned, is highly expressed in oxidative tissues such as the liver, cardiac muscle and skeletal muscle and kidney and is targeted by hypolipidemic agents known as fibrates (Issemann and Green, 1990). The hypolipidemic effect of PPARα activation is modulated, in part, by increased cellular fatty acid uptake and conversion to acyl-CoA through increased transcription of the fatty acid transport protein and acyl-CoA synthetase genes (Martin et al., 1997) combined with increased peroxisomal and mitochondrial β-oxidation of fatty acids, through upregulation of genes such as carnitine palmitoyl transferase 1 (Brandt et al., 1998) and acyl CoA oxidase (Staels et al., 1995). In concert, PPARα activation also enhances the catabolism of triglyceride rich lipoproteins by decreasing *APOC3* mRNA levels in human, mouse and rat hepatocytes (Peters et al., 1997; Staels et al., 1995).

PPARβ is the least-studied PPAR. Like PPARα, it is thought to play a role in modulating lipid metabolism. In insulin-resistant rhesus monkeys, treated with a PPARβ specific agonist, serum levels of HDL were increased by 80% whilst serum triglycerides and LDL-cholesterol levels were reduced by 59% and 29%, respectively (Oliver et al., 2001). PPARβ activation also targets β-oxidation genes in adipose tissue (Wang et al., 2003). Although ubiquitously expressed, PPARβ is the most abundant PPAR isoform in skeletal muscle (Braissant et al., 1996) and is found at higher levels in oxidative type I muscles fibres compared to glycolytic type II muscle fibres (Wang et al., 2004b), thus suggesting that PPARβ maybe particularly important to lipid utilisation in skeletal muscle. Also of note for the current study, several authors have suggested that bezafibrate activates both PPARα and PPARβ ((Brown et al., 1999; Krey et al., 1997; Peters et al., 2003; Poirier et al., 2001).

PPARγ is an essential regulator of adipogenesis. Although PPARγ-null mice are embryonic lethal, a study using mice chimeric for wild-type and PPARγ-null cells showed that the null cells made no contribution to the adipose tissue (Rosen et al., 1999). This study also showed that, unlike wild-type mouse embryonic stem (ES) cells, PPARγ-null ES cells could not be differentiated into adipocytes *in vitro* (Rosen et al., 1999). PPARγ also regulates expression of a number of lipogenic genes such as adipocyte fatty acid binding protein (Tontonoz et al., 1994), phosphoenolpyruvate carboxykinase (Tontonoz et al., 1995), and *LPL* (Schoonjans et al., 1996) and is mutated in some patients with the familial partial lipodystrophy (Hegele, 2005). In addition, PPARγ is important in glucose homeostasis, regulating the expression of

the insulin dependent glucose transporter, GLUT4 (Wu et al., 1998). It is also the target of the insulin-sensitising drugs, thiazolidinediones (Lehmann et al., 1995).

In this chapter, expression data is combined with genetic data accrued in earlier chapters in order to identify further chromosome 21 genes that may warrant further investigation with respect to the lipid traits of FCHL.

6.2) Results

6.2.1) Identification of Chromosome 21q22.3 FCHL Candidate Genes by Quantification of mRNA Levels in Bezafibrate-Treated Cells

To identify genes within the chromosome 21q22.3 FCHL linkage interval that might have a previously unsuspected function in lipid metabolism, expression profiling was performed on McArdle RH7777 (McA-RH7777) cells treated with the PPARα/β agonist bezafibrate.

Based on a previous study in McA-RH7777 cells (Gbaguidi and Agellon, 2004), initial optimisation experiments were performed using two bezafibrate doses (100μM and 200μM) and two time-points (6 hours and 24 hours). Quantitative, real-time reverse transcription PCR (RT-qPCR) demonstrated that, as expected (Rakhshandehroo et al., 2007), mRNA levels of two positive control genes, monoacylglycerol lipase (*MGL)* and adipose triglyceride lipase (*PNPLA2*) mRNA, were significantly higher under all bezafibrate treatment conditions ($P<0.001$, Figure 6.1 and 6.2). For both genes, the largest fold-increase in transcript levels was observed when cells were treated with 200μM bezafibrate for a 24-hour period ($P\leq0.003$ for 200μM/24-hour treatment versus all other treatment conditions). As a result, these conditions were employed to assay the response of genes within the 21q22.3 FCHL linkage interval (n=61) to bezafibrate treatment.

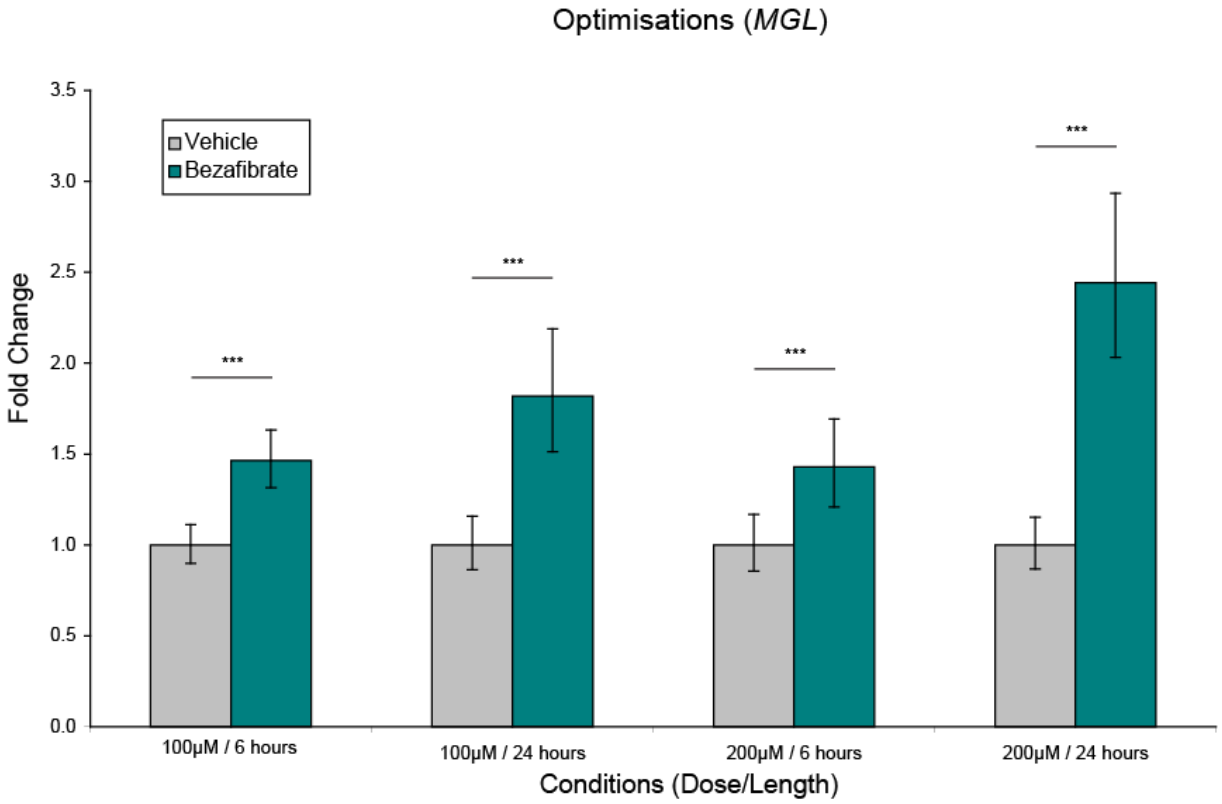


Figure 6.1) Changes in *MGL* mRNA levels in McA-RH7777 cells following different bezafibrate doses and treatment lengths

McA-RH7777 cells were treated with either 100µM or 200 µM bezafibrate or DMSO-only for a 6 or 24-hour period. All treatments significantly increased *MGL* mRNA levels, measured using SYBR green RT-qPCR (***, $P<0.001$). The 200µM/24-hour treatment was the most effective ($P\leq0.003$, not shown for clarity).

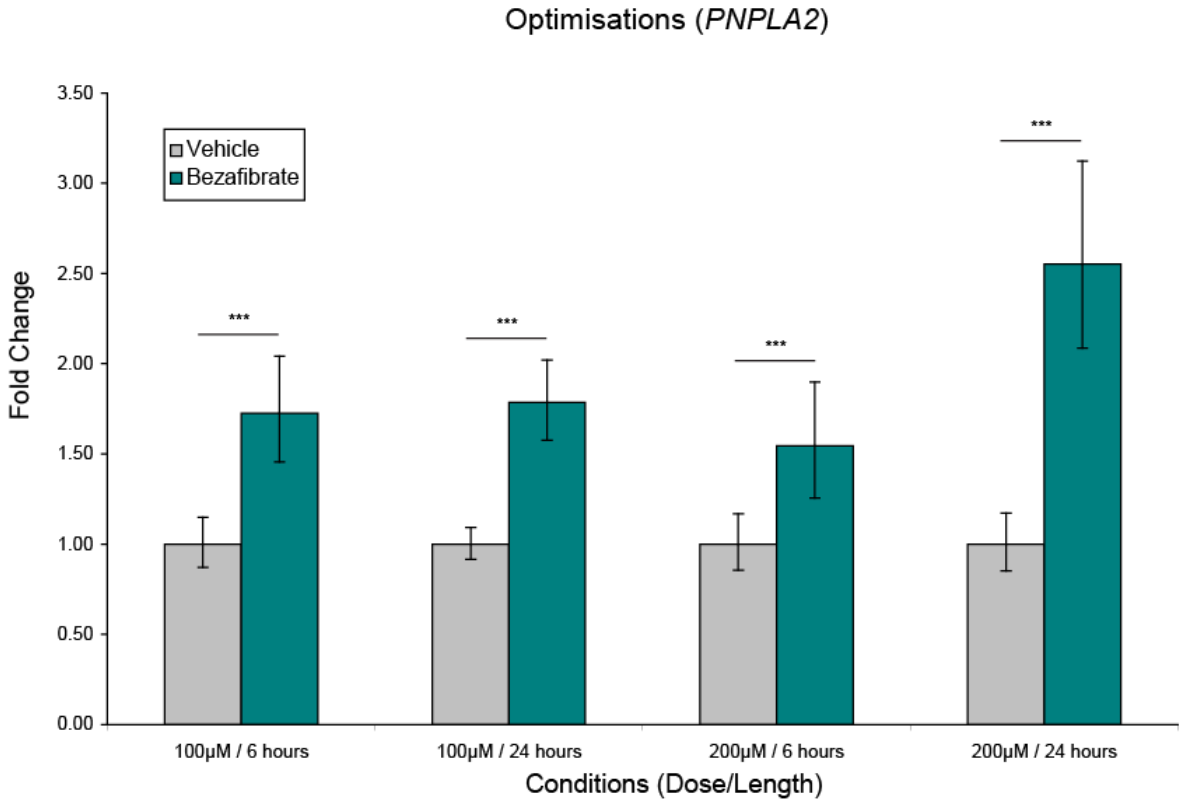


Figure 6.2) Changes in *PNPLA2* mRNA levels in McA-RH7777 cells following different bezafibrate doses and treatment lengths

McA-RH7777 cells were treated with either 100µM or 200µM bezafibrate or DMSO-only for a 6 or 24-hour period. All treatments significantly increased *PNPLA2* mRNA levels, measured using SYBR green RT-qPCR (***, $P<0.001$). The 200µM/24-hour treatment was the most effective ($P\leq0.003$, not shown for clarity).

Sixteen of the 61 protein-coding genes within the 3.8Mb linkage interval encode keratin-associated proteins, whose expression is reported to be specific to hair root cells (Shibuya et al., 2004). The transcripts of these genes were not detected in 28 other human tissues examined, including adult and fetal liver (Shibuya et al., 2004). Thus in the current study these genes were not examined.

No rat homolog was identified for four (*WDR4*, *U2AF1*, *C21ORF2*, *TSPEAR*) of the remaining 45 genes and therefore the mRNA levels of these genes could not be quantified in McA-RH7777 cells (Table 6.1). Three genes (*C2CD2*, *ZNF295* and *PCBP3*) which reside at the outer limits of the 3.8Mb linkage interval were also not examined since when the experiments were designed these genes were positioned outside of the region of interest (according to human genome build hg18) (Table 6.1). Quantification assays were performed for the remaining 38 genes. In addition, transcript-specific assays were conducted for the three known *COL18A1* transcripts (Figure 6.3).

Transcript levels for ten of the 38 genes were below reliable detection levels (defined as cycle threshold (Ct) >35 in the RT-qPCR assay; Table 6.1). In-keeping with these results, examination of data from the GenBank repository (URL5) revealed no human liver ESTs corresponding to these genes.

Non-specific RT-qPCR-products were produced for eight genes (Table 6.1, Figure 6.4). This was not resolved by redesign of the PCR primers. Consequently, the effects of bezafibrate on mRNA levels of these genes could not be examined.

Table 6.1) Genes for which mRNA Quantification in McA-RH7777 Cells was not possible

Problem With Assay	Gene	Predominant Human EST Source
No rat gene	*WDR4*	Spleen
	U2AF1	Embryonic Tissue
	C21ORF2	Ovary
	TSPEAR[a]	Testis
No expression detected	*UMODL1*	Thymus
	ABCG1[b]	CNS
	TFF1	Stomach
	TFF3	Intestine, Thyroid
	TMPRSS3	Trachea
	RSPH1	Connective tissue
	HSF2BP	Oesophagus
	ICOSLG	Brain
	DNMT3L	Placenta
	ITGB2	Spleen, Thymus
Non-specific products	*SLC37A1*	Stomach
	PDE9A	Prostate
	LRRC3	Cervix
	UBASH3A	Spleen
	TFF2	Stomach
	CRYAA	Eye
	PWP2H	Muscle
	AIRE	Thymus

EST=expressed sequence tag

[a] SNPs in first intron are nominally associated with triglyceride level in FCHL/GLGC cohort (*P*=0.0004-0.0009)

[b] Two low frequency intronic SNPs are nominally associated with triglyceride level in FCHL cohort (*P*=0.0002-0.0003)

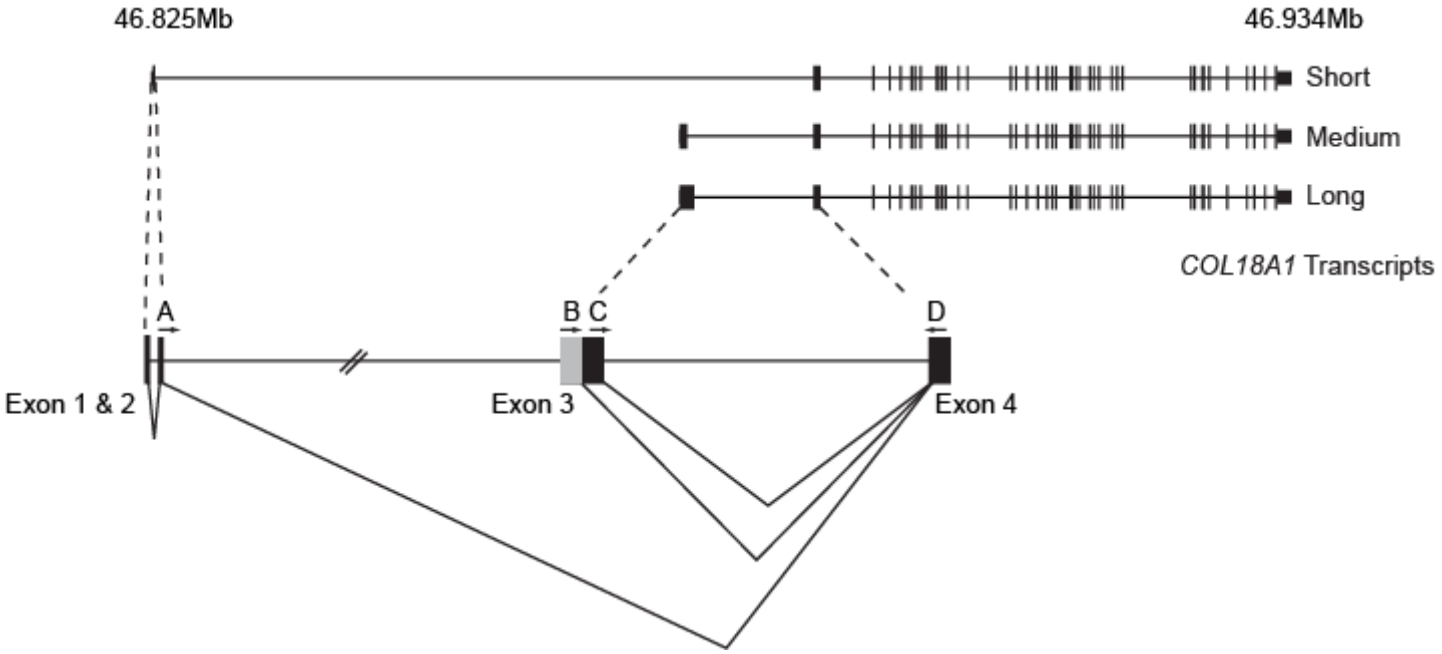


Figure 6.3) Human *COL18A1* transcripts

Transcript specific RT-qPCR primers (forward primers: A, B, C; universal reverse primer: D) were designed to detect the rat sequences homologous to the three known human *COL18A1* transcripts. The short transcript includes exon 1, 2 and 4-43. In contrast, the medium and long transcripts are spliced from exons 3, 4-43. The medium transcript contains only the first 688 base-pairs of exon 3, whereas in the long transcript, exon 3 is 1372 base pairs in length.

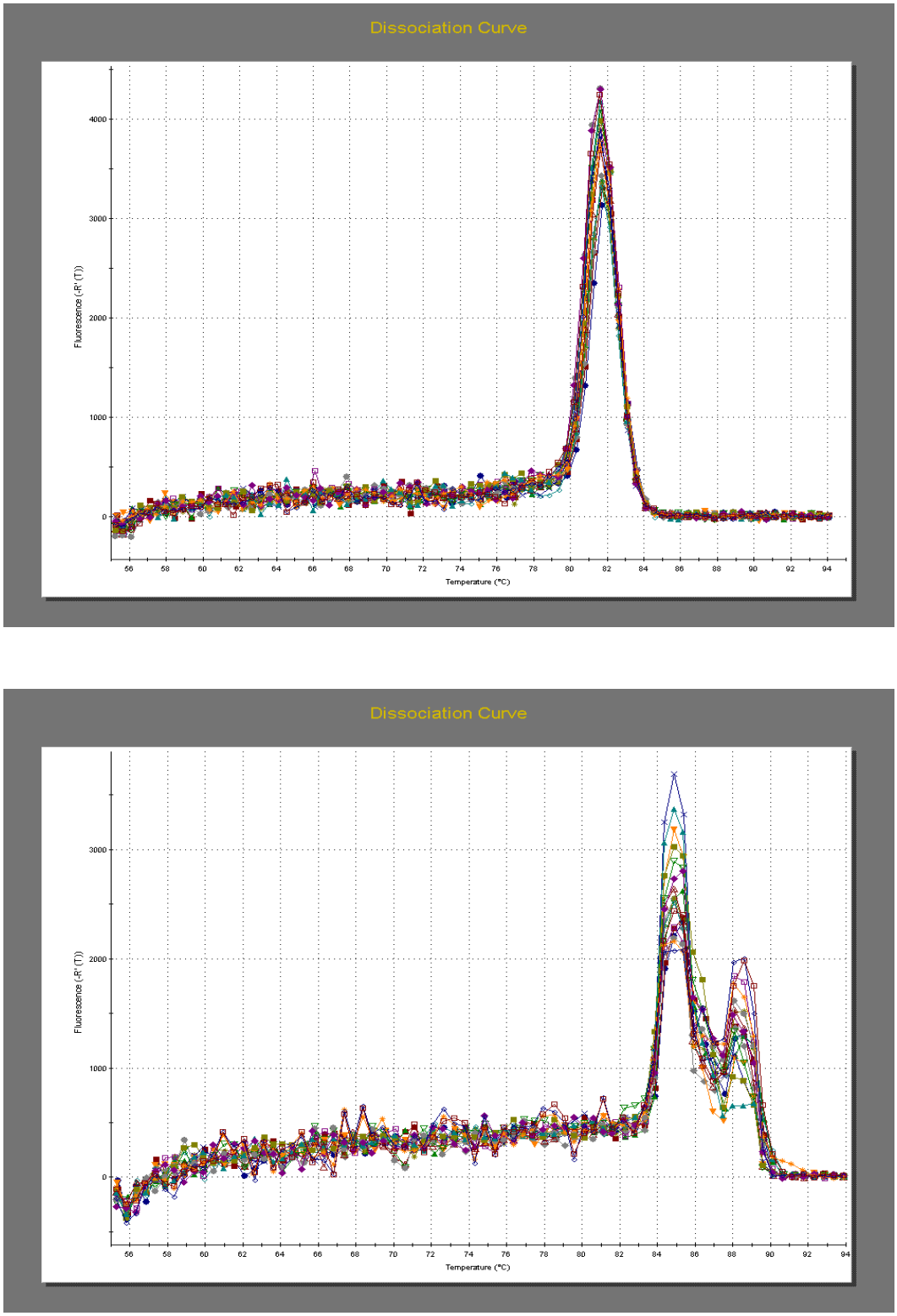


Figure 6.4) Example melt curves from RT-qPCR assays

The top panel shows the dissociation curve for a specific PCR-product whilst the bottom panel shows the result of a non-specific reaction. X-axes denote temperature in °C and y-axes represent the first negative derivative of the change in fluorescence.

With the exception of *PKNOX1*, all remaining assays (n=21), passed quality control checks (Ct<35 and approximate amplification efficiency >80% and <120%) (Table 6.2). Since *PKNOX1* dissociation curves showed no evidence of primer dimer formation, (a potential cause of inflated amplification efficiency) this assay was not excluded from subsequent quantification.

Relative quantification determined that mRNA levels of two genes were significantly altered by bezafibrate treatment (Table 6.2). Firstly, *CBS* mRNA levels in the treated cells were 0.45-fold lower than those detected in the untreated cells (*P*<0.001). Conversely, *TRPM*2 mRNA levels were 1.35 fold higher in the bezafibrate-treated cells (*P*=0.005).

CBS SNPs were included in the FCHL candidate gene association study described in Chapter 5. In the whole FCHL cohort, no *CBS* SNPs returned a *P*-value of *P*<0.01 (Tables 5.2 and 5.3). When considering solely the 47 families contributing to the OSA-cholesterol linkage peak, the CBS SNP rs1788484 returned a *P*-value of P=0.0063 for association with circulating cholesterol levels (Table 5.6). However, this SNP was not associated with cholesterol levels in the GLGC meta-analysis (*P*=0.6052). In the GLGC study, rs234706 returned the lowest *P*-value of all *CBS* SNPs for association with both plasma triglyceride (*P*=0.0096) and cholesterol levels (*P*= 0.0065) although the SNP was genotyped in only one third of the meta-analysis samples (n=33,096). In the FCHL cohort the corresponding *P*-values were *P*=0.3679 and *P*=0.6538, respectively. Therefore, current association data provide no robust evidence that *CBS* SNPs contribute to FCHL lipid traits or to plasma lipid levels in the population as a whole.

TRPM2 SNPs were not genotyped as part of the FCHL candidate gene association study. In the GLGC samples, the lowest *P*-value returned for association between a *TRPM2* SNP and plasma triglyceride levels was *P*=0.0163 for rs2838553, which resides in the second intron of the gene. In the corresponding cholesterol analyses, the lowest *P*-value was obtained for rs11911248 (*P*=0.0134) which resides in the ninth intron of the gene.

Table 6.2) Changes in mRNA Levels of 21q22.3 Genes in Response to Bezafibrate Treatment

Gene Name	Efficiency (%)[a]	Ct Range[b]	ΔΔCt	95% CI	SE ΔΔCt	Fold change[c]	95% CI	*P*-value[d]
ADARB1	101.65	26.56-27.34	-0.0726	(-0.4430, 0.2978)	0.1747	1.0516	(0.8135, 1.3594)	0.6830
AGPAT3	118.70	28.79-30.65	0.1785	(-0.2344, 0.5915)	0.1948	0.8836	(0.6637, 1.1765)	0.3730
CBS	**115.40**	**29.63-32.89**	**1.1561**	**(0.7426, 1.5696)**	**0.1951**	**0.4487**	**(0.3369, 0.5977)**	**<0.001**
COL18A1(L)	94.40	26.47-28.02	-0.1385	(-0.4325, 0.1554)	0.1387	1.1008	(0.8979, 1.3495)	0.3330
COL18A1(M)	99.50	22.17-23.68	0.1494	(-0.3590, 0.1282)	0.2313	0.9016	(0.9150, 1.2825)	0.5270
COL18A1(S)	116.00	27.54-28.81	-0.1154	(-0.3409, 0.6397)	0.1149	1.0833	(0.6418, 1.2665)	0.3300
CSTB	108.24	21.12-21.90	-0.0591	(-0.3207, 0.2026)	0.1234	1.0418	(0.8690, 1.2490)	0.6390
ES1	100.33	25.72-26.49	0.0069	(-0.2015, 0.2152)	0.0983	0.9953	(0.8614, 1.1499)	0.9450
NDUFV3	99.29	25.27-25.93	-0.0883	(-0.4777, 0.3010)	0.1837	1.0631	(0.8117, 1.3925)	0.6360
PDXK	116.39	25.06-26.35	-0.0326	(-0.5390, 0.4738)	0.2389	1.0228	(0.7201, 1.4530)	0.8940
PFKL	117.75	23.50-24.38	0.0494	(-0.2610, 0.3598)	0.1464	0.9663	(0.7792, 1.1983)	0.7400
PKNOX1	164.50	25.22-26.05	-0.0985	(-0.2695, 0.0724)	0.0797	1.0707	(0.9510, 1.2054)	0.2370
POFUT2	102.61	25.57-26.38	0.0706	(-0.1922, 0.3333)	0.1240	0.9523	(0.7937, 1.1425)	0.5770
PTTG1IP	110.73	26.75-28.62	-0.0089	(-0.4205, 0.4027)	0.1942	1.0062	(0.7564, 1.3384)	0.9640
RRP	113.07	23.17-24.02	0.0807	(-0.1448, 0.3063)	0.1064	0.9456	(0.8087, 1.1056)	0.4590
RRP1B	103.72	25.44-26.90	0.0069	(-0.2907, 0.3044)	0.1404	0.9953	(0.8098, 1.2232)	0.9620
SLC19A1	113.90	28.27-29.43	0.2285	(-0.0776, 0.5346)	0.1444	0.8535	(0.6904, 1.0552)	0.1330
SNF1LK	101.52	24.46-25.76	0.3169	(-0.0838, 0.7175)	0.1890	0.8028	(0.6081, 1.0598)	0.1130
SUMO3	97.00	22.91-23.98	0.0606	(-0.2888, 0.4099)	0.1648	0.9589	(0.7523, 1.2216)	0.7180
TRAPPC10	106.24	23.35-24.36	-0.0204	(-0.2255, 0.1848)	0.0968	1.0142	(0.8798, 1.1692)	0.8360

TRPM2	**102.96**	**29.52-33.01**	**-0.5380**	**(-0.8909, -0.1850)**	**0.1665**	**1.4519**	**(1.1368, 1.8544)**	**0.0050**
UBE2G2	111.40	23.73-24.44	0.0696	(-0.1011, 0.2404)	0.0806	0.9529	(0.8465, 1.0726)	0.4000

SE=standard error

[a] Primer efficiency was calculated from the slope of the standard curve generated using five template RNA concentrations (2.13)

[b] Range of cycle threshold values recorded across independent experiments

[c] Fold change=$2^{-\Delta\Delta Ct}$

[d] *P*-value indicates whether ΔΔCt is significantly different from 0

6.2.3) Association between rs4818835 and *C21orf57* Expression

To determine whether SNPs within the 21q22.3 linkage interval might influence gene transcript levels in a tissue pertinent to the aetiologies of FCHL, data from the study of Schadt *et al.,* which genotyped 782,476 SNPs and profiled 39,280 transcripts in 427 human liver samples, were examined (Schadt et al., 2008). Additionally, data for SNPs genotyped in the FCHL cohort residing outside the 3.8Mb linkage interval (*BACE2*, *LSS* and *MCM3AP* regions, 4.2) were considered.

In total, nine SNPs were associated with nine specific liver chromosome 21 gene expression traits at the study-defined nominal *P*-value cut-off of $5x10^{-5}$ (Table 6.3). However, the nine genes did not include *CBS* or *TRPM2* (6.2.2).

Three of the nine SNPs were directly genotyped or imputed in the FCHL cohort. One of these SNPs, rs4818835, a cis-eQTL for liver *C21orf57* transcript levels (Table 6.3), was nominally associated with triglyceride levels in the FCHL cohort (β=0.0611± 0.0230(SE), *P*=0.0079). However, in the GLGC cohort the corresponding *P*-value was *P*=0.3320 (Table 6.3).

Next, data for the nine eQTLs were examined in an additional dataset, also relevant to FCHL: the study of Nica *et al*., which performed SNP-gene expression analysis using a dataset comprising 865,544 SNPs genotyped in 166 individuals (142 twins, 24 singletons) and adipose tissue expression data for 18,170 genes (Nica et al., 2011). This study employed a matched co-twin design where data from twin-pairs were independently analysed thus allowing internal replication of potential eQTLs. In this dataset, replicating findings in the liver, rs4818835 was associated with the same gene expression trait (*C21orf57*) in both sets of twins (twin group 1, *P*=0.0042 and twin group 2, $P<1x10^{-4}$) (Figure 6.5).

Table 6.3) Chromosome 21q22.3 *cis*-eQTLs in Liver (extracted from Schadt *et al*., 2008)

SNP	Position[a]	Expression Trait	Probe Position[a]	*P*(Expression)	Association in FCHL Cohort *P*(LME_TG)	*P*(GEE_TG)	*P*(LME_TC)	*P*(GEE_TC)	GLGC Study *P*(TG)	P(TC)
rs2839501	43,805,637	*TMPRSS3*	43,791,999	$1.42x10^{-5}$	na	na	na	na	0.6427	0.0472
rs7283383	43,913,421	*SLC37A1*	43,919,742	$1.96x10^{-5}$	0.7864	0.3749	0.5756	0.3841	0.2224	0.0508
rs2839603	44,324,506	*NDUFV3*	44,313,378	$2.52x10^{-25}$	na	na	na	na	0.9867	0.9914
rs2401144	44,379,258	*PKNOX1*	44,394,643	$3.13x10^{-5}$	na	na	na	na	0.8226	0.8885
rs2329573	45,198,522	*CSTB*	45,193,831	$6.49x10^{-9}$	na	na	na	na	0.1231	0.4034
rs235343	46,249,752	*PTTG1IP*	46,269,513	$2.67x10^{-26}$	0.1899	0.1223	0.5640	0.6601	0.5168	0.4308
rs17004733	46,520,840	*C21orf2*	45,748,827	$4.23x10^{-5}$	na	na	na	na	0.3799	0.0821
rs4819035	46,520,840	*ADARB1*	46,494,515	$3.56x10^{-6}$	na	na	na	na	0.0283	0.4187
rs4818835	**46,645,715**	***C21orf57***	**47,706,267**	**$4.62x10^{-12}$**	**0.0079**	**0.0443**	**0.1706**	**0.1297**	**0.3320**	**0.0800**

na=not genotyped/imputed in FCHL cohort

[a] Base-pair position on chromosome 21

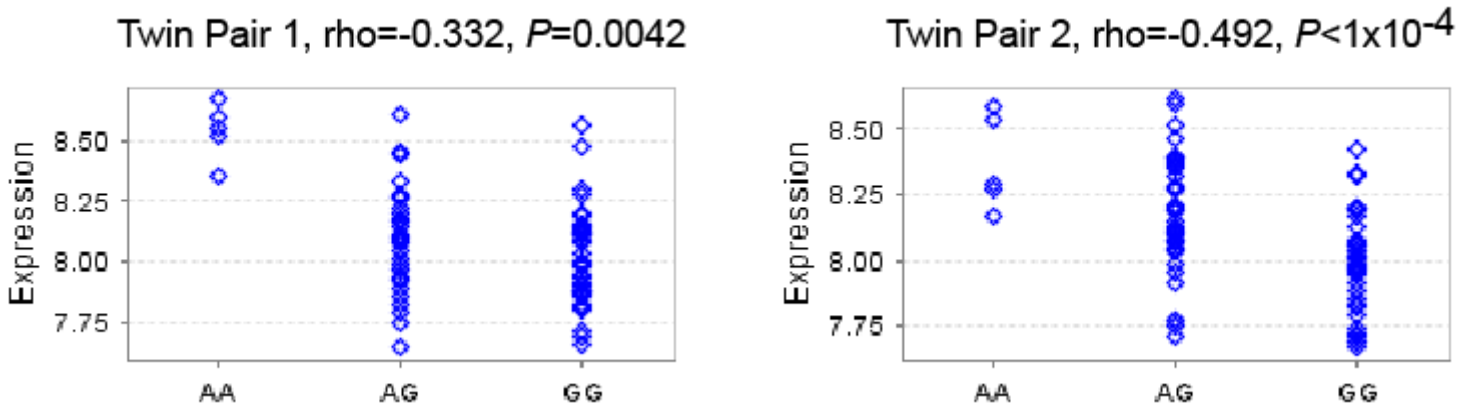


Figure 6.5) Correlation between rs4818835 Genotype and *C21orf57* Transcript Levels in Adipose Tissue from Female Twins

The rs4818835 minor allele is associated with higher *C21orf57* transcript levels in adipose tissue from twins of the Multiple Tissue Human Expression Resource (Nica et al., 2011). Association was determined using Spearman's rank correlation and *P*-values are based upon 10,000 permutations. Plot was produced using Genevar (Yang et al., 2010).

6.2.3) Analysis of Differentially Expressed Genes in Adipose Tissue from FCHL Patients and Controls

In the study of Plaisier *et al*., which compared mRNA levels in adipose tissue biopsies taken from 38 Mexican FCHL cases and 32 normolipidemic controls, transcript levels of *C21orf57* did not significantly differ between the two groups (at a significance threshold of $P<0.05$) (Plaisier et al., 2009) However, in total, five genes within the 21q22.3 FCHL linkage interval did display differential expression at $P<0.05$ (Table 6.4), including three genes (*AGPAT3*, *UBE2G2* and *COL18A1*) previously considered as FCHL candidate genes (4.2.1). Examination of the Nica et al., dataset revealed no evidence that SNPs within 100Kb of the transcription start sites of *AGPAT3, COL18A1* or *UBE2G2* appear to act as *cis*-eQTLs in adipose tissue (at $P<0.01$) in the 166 individuals studied (Nica et al., 2011). Likewise, there were no *AGPAT3, COL18A1* or *UBE2G2* *cis*-eQTLs in the liver dataset of Schadt et al., at the study defined significance threshold of $P<5x10^{-5}$ (Schadt et al., 2008)

The fourth gene for which transcript levels differed significantly between FCHL patients and controls was *PWP2,* which encodes the periodic tryptophan protein homolog. Consultation of the Nica *et al.,* adipose eQTL dataset (3.2.2) revealed that no SNPs within 100Kb of the *PWP2* transcription start site were associated ($P<0.01$) with transcript levels of the gene in the two co-twin groups (Nica et al., 2011). *PWP2* SNPs were not genotyped in the FCHL cohort and no SNPs in the *PWP2* interval returned a *P*-value of $P<0.1$ for association with plasma triglyceride levels in the GLGC study (Teslovich et al., 2010). Thus, current genetic data do not support a role for *PWP2* in modulating the FCHL triglyceride trait.

Finally, transcript levels of *ITGB2,* encoding β2-integrin, were differentially expressed in the Mexican FCHL cases and controls. No *ITGB2* cis-eQTLs were identified in the Nica *et al*., dataset (Nica et al., 2011) In the GLGC cohort, rs235305, which resides ~7kb downstream of *ITGB2,* returned the lowest *P*-value ($P=0.0057$) for association with plasma triglyceride (Teslovich et al., 2010). Therefore, on the basis of current genetic data, *ITGB2* does not represent a strong candidate for the FCHL triglyceride trait.

Table 6.4) 21q22.3 FCHL Linkage Interval Genes Differentially Expressed in Mexican FCHL Patients and Controls

Gene	TSS	Effect Estimate	SE	T-statistic	*P*[a]
AGPAT3	45,285,115	0.27	0.08	3.17	0.0023
PWP2	45,527,207	0.13	0.05	2.68	0.0091
UBE2G2	46,188,494	0.10	0.04	2.52	0.0142
ITGB2	46,305,867	0.46	0.16	2.85	0.0057
COL18A1	46,825,096	0.21	0.10	2.00	0.0489

TSS=transcription start site, SE=standard error

[a] Where mRNA levels were quantified on the basis of two microarray probes, statistics are given for the probe that returned the lowest *P*-value

6.2.4) Summary of 21q22.3 Gene Expression and Association Analyses

Table 6.5 shows that the most promising results to emerge from the data presented in this chapter involve the SNP, rs4818835. This SNP was associated with *C21orf57* transcript levels in both liver and adipose tissue (Nica et al., 2011; Schadt et al., 2008). Additionally, in Chapter 5, it was nominally associated with plasma triglyceride levels in the FCHL cohort (β=0.0611±0.0230(SE), *P*=0.0079). Unfortunately, since *C21orf57* resides (~600Kb) outside of the refined 21q22.3 linkage interval, it was not examined in the bezafibrate-treatment dataset (4.2.1). However, on the basis of the combined genetic and gene expression data, it will be important for future studies to remedy this and to further characterise the potential contribution of rs4818835/*C21orf57* to FCHL.

Table 6.5 also summarises the combined results of Chapter 4, 5 and 6, and highlights those genes which may warrant further investigation with regard to the aetiology of FCHL.

Table 6.5) 21q22.3 Genes Implicated in the Aetiology of FCHL: Biological, Genetic and Gene Expression Data

Gene	Chapter 6 Expression Datasets	Chapter 5 FCHL Cohort Association Study	Chapter 4 Candidate Gene
LSS[a]	-	rs35785446 associated with TG/TC (*P*=0.0039/0.0074)	Yes
TSPEAR[a]	-	rs34163868 associated with TG (*P*=0.0009)	No
PKNOX1[a]	**rs2401144 genotype associated with liver transcript levels**	**rs2839619 associated with TC in FCHL families (*P*=0.0091), replicating Zemunik *et al*., 2009** **rs9984563, top 21q22.3 TG-associated SNP in GLGC study (*P*=0.0005)**	**No**
CBS	**Bezafibrate treatment -> lower transcript levels**	**-**	**Yes**
TRPM2	**Bezafibrate treatment -> higher transcript levels**	**-**	**No**
C21orf57	**rs4818835 genotype associated with liver/adipose *C21orf57* transcript levels**	**rs4818835 associated with TG (*P*=0.0079)**	**No**
AGPAT3[b]	Differentially expressed in adipose tissue from FCHL patients/controls	-	Yes

COL18A1	**Differentially expressed in adipose tissue from FCHL patients/controls**	**rs8129539 associated with TG (*P*=0.0084)**	**Yes**
UBE2G2[b]	Differentially expressed in adipose tissue from FCHL patients/controls	-	Yes
PWP2	**Differentially expressed in adipose tissue from FCHL patients/controls**	**-**	**No**
ITGB2	**Differentially expressed in adipose tissue from FCHL patients/controls**	**-**	**No**

Genes/results highlighted in bold are discussed later in this chapter (6.3)

TC=total cholesterol, TG=triglyceride

[a] Discussed in Chapter 5

[b] Discussed in Chapter 4

6.3) Discussion

The culmination of data presented in this, and the preceding chapter has generated several findings that warrant further study and discussion. The significance of two of these genes, *TSPEAR* and *LSS* is discussed in Chapter 5. In this chapter, by combining gene expression data with genetic data, arguably the most important result involves *C21orf57/* rs4818835.

Notably, rs4818835 was associated with *C21orf57* transcript levels in two distinct datasets, representing human liver (Schadt et al., 2008) and adipose tissue (Nica et al., 2011) and was nominally associated with plasma triglyceride levels in the FCHL families (*P*=0.0079). Although this association does not imply causation, it does indicate that this non-coding SNP, or a variant in LD, could have a functional effect. Thus, genotyping this gene locus/SNP in further hyperlipidemic cohorts seems warranted. Future studies could also include examining *C21orf57* transcript levels in bezafibrate treated McA-RH7777 cells since, due to the position of the gene ~600Kb outside the refined 21q22.3 FCHL linkage interval, it was not included in the RT-qPCR dataset present in this chapter (6.2.1).

Data regarding *C21orf57* function are currently limited. The human gene is predicted to encode at 167 amino acid protein, designated as a putative metallopeptidase (Zhan et al., 2005). It is therefore of potential interest that serum protein levels of two other metallopeptidases, matrix metalloproteinase-2 (MMP-2) and matrix metalloproteinase-9 (MMP-9) were found to be higher in a group of 168 dyslipidemic individuals compared to a group of 179 controls (MMP-2: 1242.7±141.4 vs. 642.7±273.8ng/ml, MMP-9: 506.3±58.7 vs. 51.3±14.8ng/ml) (Derosa et al., 2009). In relation to these findings, it seems pertinent that the rare allele at SNP rs4818835 was associated with both higher *C21orf57* transcript levels in the Nica *et al*., dataset and higher triglyceride levels in the FCHL families (Nica et al., 2011).

Also in the current study, transcriptional profiling of McA-RH7777 cells revealed that the mRNA levels of two genes within the 3.8Mb FCHL linkage interval were significantly altered by treatment with the PPAR-agonist bezafibrate. The first gene, *CBS*, was previously designated an FCHL candidate gene since in mouse models of *Cbs* loss-of-function, plasma cholesterol and triglyceride levels are increased (4.2). Biologically, such a mutation would cause disruption to the methionine cycle and in turn could result in increased phospholipid production through increased availability of the methyl donor S-Adenosyl homocysteine. Paradoxically, in the current

investigation, *Cbs* mRNA levels were down-regulated 0.45-fold (95%CI: 0.34-0.60) in bezafibrate treated McA-RH7777 cells. Supporting these findings, a similar effect was observed in liver samples taken from Sprague-Dawley rats, fed a high-fat diet (48% saturated fat) and treated with the PPARα agonist WY-14,643 (30 µmol per kg, per day for four weeks) where *Cbs* mRNA levels were 0.58-fold lower in the treated versus untreated animals (N=5 per group, $P<0.05$) (Sheikh et al., 2007). Thus, in future studies it could be informative to establish both the mechanism(s) underlying this downregulation, and its downstream consequences. In addition, given the linkage of the FCHL triglyceride trait to chromosome 21q, it would seem reasonable in future studies to utilise next-generation sequencing methodologies to screen *CBS* in FCHL patients in order to identify rare variants that may have a moderate to large impact on their triglyceride levels. The rationale for this approach being that previous studies have shown that multiple rare variants influence plasma triglyceride and cholesterol levels in dyslipidemic individuals (Cohen et al., 2006; Johansen et al., 2010; Johansen et al., 2011b).

The second gene for which bezafibrate treatment led to altered transcript levels was *TRPM2.* Thus *TRPM2,* which encodes transient receptor potential cation channel melastatin 2, represents a novel FCHL candidate gene. It is expressed in multiple tissues including liver and pancreas (Nagamine et al., 1998; Perraud et al., 2001; Togashi et al., 2006) and was recently implicated in insulin secretion (Uchida et al., 2011). Following either an oral or intra-peritoneal glucose load, *Trpm2* knockout mice had significantly higher plasma glucose ($P<0.01$) and significantly lower plasma insulin ($P<0.05$) levels than wild-type animals (Uchida et al., 2011). Plasma triglyceride levels were not measured. PPARα is also implicated in the maintenance of blood glucose levels. However, in contrast to the *Trpm2*-null animals, *PPARα*-null mice develop hyperinsulinemia (Gremlich et al., 2005). In addition, treatment with the PPAR-agonist, clofibrate led to an ~31% *decrease* in insulin secretion by INS-1 insulinoma cells (Tordjman et al., 2002), whilst in humans fenofibrate treatment decreases plasma triglyceride levels without altering insulin secretion or sensitivity (Abbasi et al., 2008; Belfort et al., 2010). Therefore, on the basis of published data, it seems unlikely that PPAR-induced upregulation of *TRPM2* is directly linked to glucose homeostasis.

In rodents, fibrate administration leads to peroxisome proliferation, increased β-oxidation and in turn increased H_2O_2 production (Arnaiz et al., 1995). It may therefore be important to note that TRPM2 is thought to be primarily activated by

hydrogen peroxide (via ADP-ribose formation) (Jiang et al., 2010a). Thus, it seems feasible that the increase in *TRPM2* mRNA levels observed in bezafibrate treated McA-RH7777 cells could be a response to increased cellular hydrogen peroxide rather than the direct result of transactivation by PPARα/β.

In light of the observations outlined above, it will be important to establish if *TRPM2* could play a direct role in fatty acid and triglyceride metabolism. To do so one could, for example, examine the effect of *TRPM2* over- and under-expression on the mRNA/protein levels of genes involved in: fatty acid β-oxidation, such as *CPT1A* and *ACOX1*; fatty acid uptake, such as *CD36*; fatty acid transport such as *SLC27A1- 5*; fatty acid synthesis, such as *FAS*, on fatty acid activation genes such as the long-chain acyl CoA synthetases and triglyceride biosynthesis genes such as *DGAT1/2* and *LPIN1-3*.

A limitation of the current study, which employed McA-RH7777 as an *in vitro* model to identify possible FCHL candidate genes, was that by using a rat cell line those human genes with no known rat homolog could not be studied. To overcome this restriction, one could have utilised a human hepatoma cell line such as HepG2 or Huh7 (Higashi et al., 2002; Lalanne et al., 2005; Pan et al., 2002). However, unlike McA-RH7777, these cell lines have limited ability to secrete fully-lipidated apoB100 as VLDL particles (Meex et al., 2011).

Furthermore, the current study did not examine transcript levels of eight genes in bezafibrate-treated McA-RH7777 cells due to non-specificity of their RT-qPCR-products (Table 6.1). Such non-specific products may arise due to the amplification of multiple different length transcripts by the PCR primers used, or could indicate that the underlying DNA sequence is refractive to PCR-based amplification. In addition, mRNA levels of *C2CD2*, *ZNF295* and *PCBP3*, were not examined in the current study since, at the inception of this study, these genes were mapped outside the region of interest. In future investigations, one could extend the expression profiling experiments to include these three genes.

In future work, one may also consider targeting a different transcription factor to identify potential FCHL candidate genes. For example, the previously described studies of Berge and Zelcer and colleagues used an LXR agonist to identify genes pertinent in cholesterol absorption and LDL cholesterol clearance, respectively (Berge et al., 2000; Zelcer et al., 2009). In addition, to specifically target PPARγ one

could administer a thiazolidinedione such as rosigliatizone, although, since PPARγ is predominant in adipose tissue, this approach may be more suited to appropriately cultured 3T3-L1 cells.

Another result that warrants further discussion is the association of rs2401144 with transcript levels of *PKNOX1* in human liver samples (Schadt et al., 2008). As discussed in chapter 5, the intronic *PKNOX1* SNP rs2839619 was nominally associated with plasma total cholesterol levels in FCHL families (*P*=0.005), replicating an earlier GWAS finding (Zemunik et al., 2009). In additional, rs9984563, another *PKNOX1* intronic SNP, returned the lowest *P*-value of all SNPs within the refined linkage interval, for association with plasma triglyceride levels in the GLGC study (*P*=0.0005). Thus, although these three SNPs share limited LD (r^2<0.17 based upon 1000 Genomes CEU pilot data), these combined data suggest that *PKNOX1* variants might contribute to the lipid phenotypes of FCHL.

Currently there are too few data to determine whether *PWP2*, the mRNA levels of which were increased in the adipose tissue of FCHL patients versus normolipidemic controls, might contribute to the triglyceride trait of FCHL (Plaisier et al., 2009). This evolutionarily conserved protein is essential for 18S rRNA biosynthesis and is implicated in cell cycle progression in yeast (Dosil and Bustelo, 2004). Of note, in a previous expression array study involving adipose tissue samples from five Dutch FCHL patient and four age-, sex- and BMI-matched controls, a large subset of the differentially expressed genes detected were implicated in cell cycle progression (Eurlings et al., 2002). This is in agreement with unpublished data from our own group and whilst *PWP2* itself was not differentially expressed in British FCHL patients and controls, further investigation of the contribution of this gene to the FCHL triglyceride trait may be warranted.

In contrast to *C21orf57*, *CBS, TRPM2, PKNOX1*, *LSS* and *TSPEAR*, current data indicate that *ITGB2* variants are unlikely to contribute to the FCHL triglyceride trait. ITGB2, also known as CD18 heterodimerises with the integrin α-chains, α1 (CD11a), αM (CD11b) or αX (CD11c) to produce molecules important in cell adhesion (Notarangelo and Badolato, 2009). A recent study demonstrated that in healthy subjects fed a standardised high-fat meal (1230 calories, 40% from fat), post-prandial increases in blood triglyceride levels were correlated with increases in monocyte surface presentation of CD11c/CD18 (Gower et al., 2011). In earlier studies, correlations have also been demonstrated between oral fat load and leukocyte

expression of CD11a/CD18 (van Oostrom et al., 2004) and CD11b/CD18 (Alipour et al., 2008). Based on these results, it seems likely that the increased *ITGB2* mRNA levels observed in the Mexican FCHL patients are a consequence rather than a cause of their hypertriglyceridemia.

Finally, *COL18A1* mRNA levels were higher in adipose tissue from Mexican FCHL patients compared to that of control individuals (Plaisier et al., 2009). This finding contrasts with results from a previous study where *COL18A1* mRNA levels in five FCHL patients were 0.60-fold those of those in five age-matched controls (Meex et al., 2005). One explanation for this discrepancy could relate to which *COL18A1* transcripts were quantified by the probes/arrays used in each investigation. In their study, Plaisier *et al*., utilised Affymetrix Human Genome U133 Plus 2.0 oligonucleotide microarrays where the *COL18A1* probe in question (209082_s_at) maps to the medium and long (NC11-493 and NC11-728) isoforms of the gene but not the short (NC-303) isoform (Figure 6.3) (Plaisier et al., 2009). In comparison, Meex *et al.,* used macroarrays produced using clones representing 640 genes. Unfortunately, the *COL18A1* transcript(s) captured by the macroarray is not documented (Meex et al., 2005). It therefore remains possible that the decreased *COL18A1* expression in the Meex *et al.,* study correspond to the short (NC-303) transcript only. Supporting this possibility, a recent study demonstrated that patients from a large, consanguineous pedigree with Knobloch syndrome, who harbour a homozygous intron 1 splice acceptor mutation, predicted to specifically disrupt the short *COL18A1* isoform, have increased fasting plasma triglyceride levels compared to non-carriers (3.01±-0.58mmol/l versus 1.32±0.14mmol/l, n=7, P=0.02) (Bishop et al., 2010). In addition, *Col18a1*$^{-/-}$ mice have increased fasting triglyceride levels compared to wild-type littermate controls (1.34±0.28mmol/l in null mice vs. 0.73±0.15mmol/l in control mice, n=10, P<0.0001) (Bishop et al., 2010).

Together these data imply that *COL18A1* plays a role in triglyceride metabolism and thus it represents a prime candidate for the FCHL triglyceride trait. However, it is worth noting that in the Knobloch syndrome pedigree described above, a rare 'null' mutation resulted in only mild hypertriglyceridemia. Therefore, it is perhaps unsurprising that none of the common *COL18A1* variants examined in Chapter 5 were associated with markedly altered triglyceride levels and in future studies it would seem most appropriate to screen FCHL patients for loss-of-function *COL18A1* variants.

In conclusion, data presented in the Chapters 5 and 6 suggest that at least seven 21q22.3 genes warrant further investigation with regard to their role in the aetiology of FCHL (Table 6.5). To substantiate these data, future investigations will require further genotyping of rs35785446 (*LSS*), rs34163868 (*TSPEAR*), rs2839619 (*PKNOX1*) and rs4818835 (*C21orf57/8*) in additional hyperlipidemic cohorts. Functional characterisation of *TSPEAR, PKNOX1, TRPM2* and *C21orf57* will also be necessary to establish the potential importance of these genes in lipid/lipoprotein metabolism. Since the current and GLGC study indicates that common 21q22.3 variants alone do not explain the 21q22.3 FCHL linkage signal(s), use of next-generation sequencing technology to look for rare *CBS* or *COL18A1* variants that have a moderate to large impact on gene function would also be warranted. Finally, it is worth pointing out that in recent simulations of an FCHL linkage scan, two to three susceptibility variants, within a chromosomal region, were required to generate a LOD score in the order of that observed on chromosome 21 (Horswell et al., 2009). Thus it seems plausible that multiple variants in multiple chromosome 21q22.3 genes may contribute to the triglyceride and cholesterol traits in different subsets of FCHL families.

Chapter 7 – Association of *GPIHBP1* with Hypertriglyceridemia and CHD Risk

7.1) Introduction

In recent years, a number of large prospective studies have demonstrated that plasma triglyceride levels represent an independent risk factor for CHD (Arsenault et al., 2009; Bansal et al., 2007; Nordestgaard et al., 2007; Sarwar et al., 2007; Tirosh et al., 2007).

Tirosh et al., examined data for 13,953 apparently healthy young men (26-45 years of age) with fasting serum triglyceride levels lower than 3.39 mmol/l over a mean follow-up period of 10.5 years (Tirosh et al., 2007). Over this period, a total of 170 cases of angiography-proven CHD arose. In a multivariate model adjusted for age, BMI, HDL-cholesterol level, family history of CHD, fasting plasma glucose level, mean arterial blood pressure, physical activity, and smoking status, CHD occurrence was associated with triglyceride level. Hazard ratios for CHD increased across triglyceride level quintiles with a hazard ratio of 4.05 (95%CI: 2.68-8.61) observed for the individuals in the highest versus the lowest quintiles of the distribution (P<0.001 for trend) (Tirosh et al., 2007).

Bansal and colleagues identified an association between non-fasting plasma triglyceride levels and CHD risk (Bansal et al., 2007). Whilst lipid measurements are traditionally measured in the fasting state, it is argued that analysis of post-prandial levels might more accurately highlight a defect in the clearance of triglyceride-rich lipoproteins. In the Bansal et al., cohort of 6391 non-fasting women, 215 experienced a cardiovascular event during the median 11.4 year follow-up period (Bansal et al., 2007). In a multivariate model adjusted for age, blood pressure, smoking, use of hormone therapy, total and HDL-cholesterol, non-fasting triglyceride levels were associated with cardiovascular events. Hazard ratios for increasing tertiles of non-fasting levels were: 1 (reference lower tertile), 1.44 (95% CI: 0.90-2.29, middle tertile), and 1.98 (1.21-3.25, upper tertile) (P=0.006 for trend).

Conceptually, if a trait and a disease outcome are causally associated then one would expect that the genetic variants underlying the trait would also be associated with risk of developing the disease. Applying this rationale to circulating triglyceride levels and CHD, recent genetic data have lent support to the observed clinical

correlation. For example, the rs2954029 T allele (which resides ~40kb downstream of the *TRIB1*) was associated with lower plasma triglyceride levels in the GLGC cohort (-5.64mg/dl per copy, P=3.29x10^{-55}) and was also associated with a decreased risk of CHD in the CARDIoGRAM and COROGENE consortia (P=5x10-5) (Teslovich et al., 2010). The same variant was also associated with ischaemic heart disease and myocardial infarction in the prospective Copenhagen City Heart Study (Varbo et al., 2011).

Another example of a genetic locus associated with both plasma triglyceride and CHD risk involves the *APOA1-C3-A4-A5* locus, which is consistently associated with plasma triglyceride levels in GWAS (Kathiresan et al., 2009b; Teslovich et al., 2010). In a recent meta-analysis involving 73,252 individuals from 39 studies, each copy of the rs662799 C allele was associated with a 0.25 mmol/l (95% CI: 0.20–0.29, P=4.4x10^{-24}) increase in plasma triglyceride levels (Sarwar et al., 2010). This *APOA5* promoter variant was also associated and an odds ratio for CHD of 1.18 (95% CI: 1.11-1.26, P=2.6×10^{-7}) (Sarwar et al., 2010). These findings were replicated in the Italian study of early-onset myocardial infarction (De Caterina et al., 2011).

Meta-analyses have also demonstrated association between *LPL* and *ANGPTL3* SNPs and both plasma triglycerides and CHD (Angelakopoulou et al., 2011; Teslovich et al., 2010). Thus, published data suggest that genetic variants in genes involved in lipolytic pathways are important in hypertriglyceridemia and CHD risk.

LPL has long been established as the key protein involved in the hydrolysis of the triglyceride-core of VLDL and chylomicron particles, at peripheral sites. It is expressed in adipose tissue, skeletal muscle, cardiac muscle and mammary gland (Camps et al., 1990). LPL is synthesised within the parenchymal cells before being transferred to the luminal endothelial cell surface (Camps et al., 1990). Until recently, the means by which this transfer occurs has remained elusive (Davies et al., 2010). However, it is now established that GPIHBP1 is crucial in facilitating the movement of LPL to its site of action. At the endothelial cell surface, LPL acts as a homodimer (Olivecrona et al., 1985) and is able to 'dock' circulating triglyceride-rich lipoproteins and, activated by ApoC2, catalyse triglyceride hydrolysis, thereby releasing fatty acids which are taken up by peripheral tissues for use as energy, storage or re-esterification into biologically active lipids. In addition to ApoC2, the activity of LPL is also thought to be modulated by ApoA5, which is present in both chylomicrons and VLDL particles (O'Brien et al., 2005). Rare mutations in *LPL*, *APOC2* and *APOA5*

cause severe hypertriglyceridemia, illustrating the importance of the encoded gene products in triglyceride metabolism (Fojo and Brewer, 1992; Priore Oliva et al., 2005; Wang et al., 2007a).

GPIHBP1 was first isolated using expression cloning. Briefly, Ioka and colleagues produced a cDNA library using RNA from the livers of LDLR-deficient mice. Transient transfection of pools of these cDNAs into a CHO cell line led to the identification of a protein (GPIHBP1) that stimulated cell surface binding of HDL (Ioka et al., 2003). Subsequent investigations failed to corroborate this binding (Gin et al., 2011), however interest in GPIHBP1 was heightened by the discovery that mice *Gpihbp1*-deficient mice had milky plasma, with plasma triglyceride levels in some instances exceeding 64.5mmol/l (5000mg/dl) (Beigneux et al., 2007).

Several lines of evidence point towards an interaction between GPIHBP1 and LPL. Northern blots indicate that *GPIHBP1* mRNA levels are high in adipose tissue and the heart with detectable levels also present in skeletal muscle (Beigneux et al., 2007). Thus the expression profile of GPIHBP1 mirrors that of LPL. At the cell level, using fluorescence microscopy, Beigneux and colleagues observed murine GPIHBP1 expression exclusively in the capillary endothelial cells, colocalised with the endothelial marker CD31. Further examination by confocal microscopy showed that GPIHBP1 was found at the luminal face of the capillary cells in adipose tissue and heart, the presumed site of triglyceride hydrolysis. They also found that GPIHBP1 bound to LPL *in vitro*, using pgsA-745 CHO cells (which are unable to synthesise HSPGs, to which LPL is able to bind at the endothelial cell surface (Cheng et al., 1981)).

Davies *et al*. sought to substantiate evidence for an interaction between GPIHBP1 and LPL. Initially, using a monoclonal antibody to GPIHBP1, they observed that in wild-type mice, Gpihbp1 co-localised with LPL in the capillary endothelial cells of the adipose tissue and heart. However, in *Gpihbp1*-null mice, this localisation of LPL was lost. Instead, in the knockout mice, LPL was detected around the adipocytes and the myocytes of heart and skeletal muscle (Davies et al., 2010). These findings led the authors to consider the possibility that GPIHBP1 may be involved in the transport of LPL from its production site in the parenchymal cells to the apical surface of the capillary endothelial cells. Several lines of evidence support this proposition (Figure 7.1).

Mouse GPIHBP1 contains 228-amino acids, compared to the 184 amino-acid human protein. The protein structure consists of four domains (Figure 7.2) of which, the acidic and Ly6 domains are thought to be important in mediating GPIHBP1's ability to bind to LPL. A crystal structure GPIHBP1 is not currently available.

The acidic domain of GPIHBP1 is enriched in negatively charged amino acids and is therefore likely to mediate its electrostatic interaction with LPL and with lipoprotein particles. LPL contains clusters of positively-charged residues (Sendak and Bensadoun, 1998). Replacement of amino acids within the acidic domain of GPIHBP1 with alanine residues abolished the observed interaction between GPIHBP1 and LPL (Gin et al., 2008). ApoB and ApoA5 molecules that form part of chylomicron and VLDL particles also contain clusters of positively charged amino acids (Cladaras et al., 1986; Nilsson et al., 2007; Talmud, 2007).

In order to investigate the role of the cysteine-rich Ly6 domain in GPIHBP1 function, Beigneux and colleagues systematically mutated all ten cysteine residues of the Ly6 domain to alanine and observed the effect of these mutations upon trafficking of GPIHBP1 to the endothelial cell surface and upon GPIHBP1 binding to LPL. Mutated proteins were successfully trafficked and expressed on the endothelial cell surface however they were, in all cases, unable to bind to LPL (Beigneux et al., 2009b).

To date, sequencing studies in humans have established that homozygous/ compound heterozygous *GPIHBP1* mutations cause chylomicronemia in a small number of patients with this condition (Table 7.1). Four of the mutations alter residues in the Ly6 domain and abolish the binding of GPIHBP to LPL *in vitro*.

A screen of 160 unrelated chylomicronemia patients also identified one individual who was homozygous for the G56R mutation (Wang and Hegele, 2007). In contrast to other *GPIHBP1* mutations, plasma triglyceride levels in the three other family members who were G56R heterozygotes were above the 95th percentile value for their age and sex (Table 7.1). This phenotype markedly contrasts with that of Ly6 domain mutation heterozygotes who do not have elevated triglyceride levels (Beigneux, 2010; Wang and Hegele, 2007). The functional impact of the G56R mutation *in vivo* is unknown. Mutant G56R and wild-type GPIHBP1 expressed in a mutant CHO cell line were judged to have the same localisation and be of equal abundance at the cell surface whilst the mutation had no observable effect on LPL binding (Gin et al., 2007).

The contribution of *GPIHBP1* mutations to triglyceride levels in FCHL is currently unknown. Therefore the aims of the current study were to screen FCHL patients for *GPIHBP1* mutations and subsequently to examine the impact of common *GPIHBP1* variants upon triglyceride levels in FCHL families.

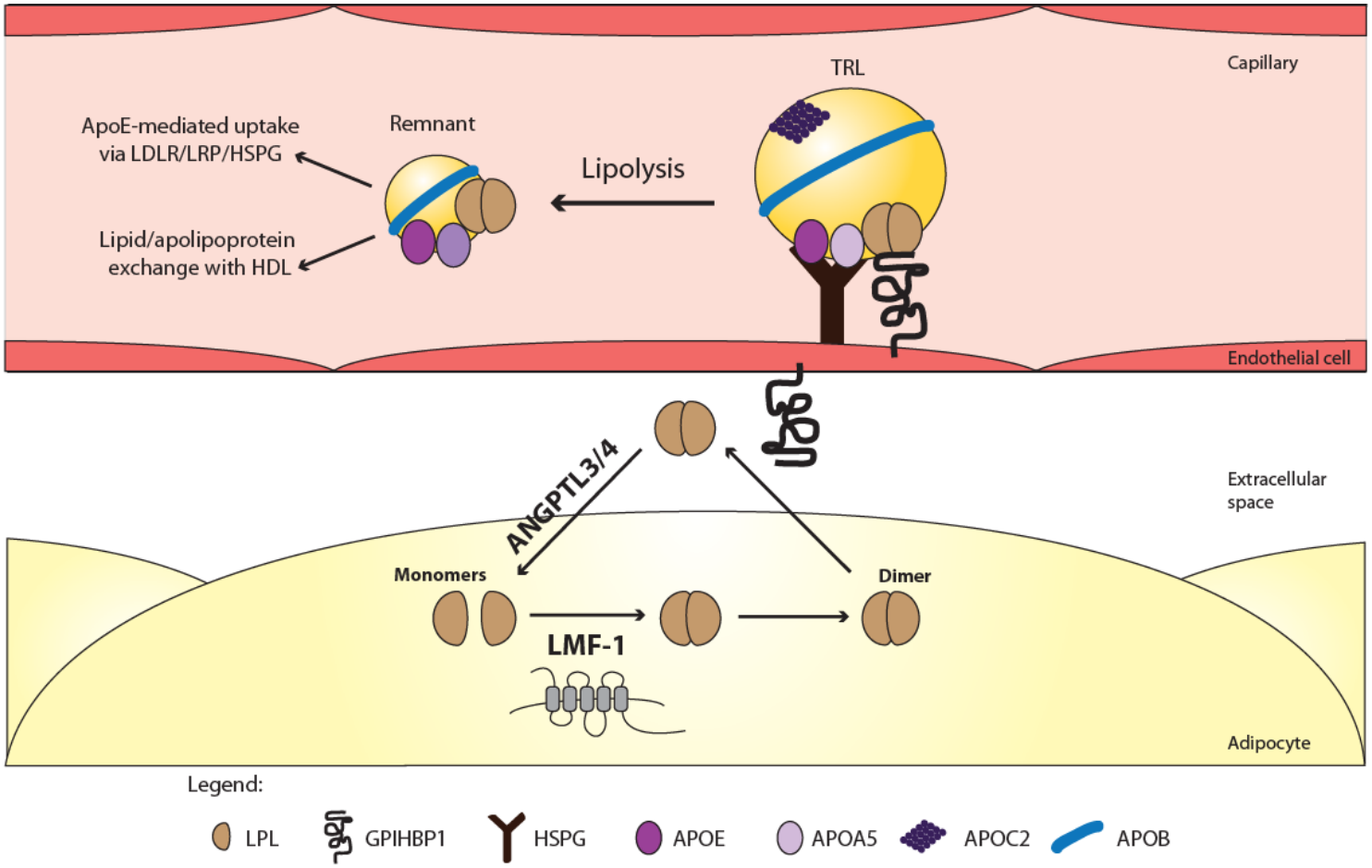

Capillary
TRL
ApoE-mediated uptake
via LDLR/LRP/HSPG
Remnant
Lipolysis
Lipid/apolipoprotein
exchange with HDL
Endothelial cell
ANGPTL3/4
Extracellular
space
Monomers
Dimer
LMF-1
Adipocyte
Legend:
LPL
GPIHBP1
HSPG
APOE
APOA5
APOC2
APOB

Figure 7.1)

Figure 7.1) Lipolysis of Triglyceride-Rich Lipoproteins by Lipoprotein Lipase

Lipoprotein lipase is essential for lipolysis of the triglyceride-rich lipoproteins (TRLs). LPL is synthesised in the parenchymal cells of adipose, heart and skeletal muscle. LPL undergoes maturation with the assistance of LMF-1, an ER-protein with five membrane-spanning domains, to produce active head-to-tail homodimers (Osborne et al., 1985; Peterfy et al., 2007). Conversely, ANGPTL4 promotes LPL inactivation, by promoting the conversion of homodimers to inactive monomers (Sukonina et al., 2006).

GPIHBP1 is thought to facilitate the transport of LPL from the basolateral to apical endothelial cell surface. In rat heart microvessel endothelial cells, GPIHBP1 can be released from both cell surfaces by cleavage of its GPI-anchor (Davies et al., 2010). The protein is also capable of moving across endothelial cells and in doing so is able to transport LPL from the basolateral to apical cell surface (Davies et al., 2010). Once at the apical cell surface LPL, stabilised by heparan sulphate proteoglycans, can catalyse the hydrolysis of the triglyceride-rich core of chylomicron and VLDL particles. This hydrolysis generates free fatty acids that may be taken up by peripheral cells for use as energy in muscle cells or for storage in adipose tissue. In addition, the resultant remnant particles undergo an exchange of lipids/lipoproteins with HDL particles and ultimately are cleared from the circulation, in the liver via LDLR.

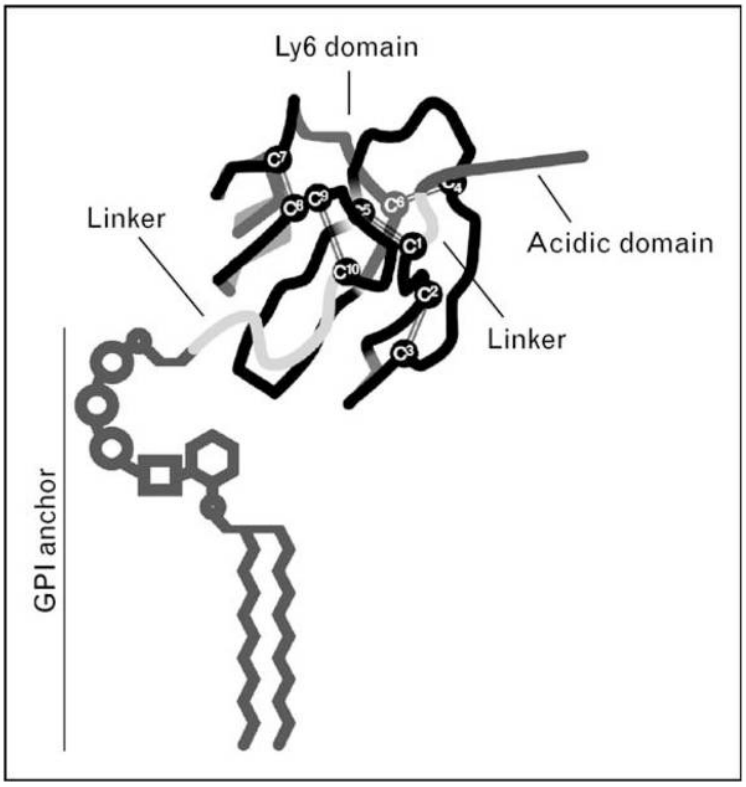


Figure 7.2) Schematic illustration of GPIHBP1 structure from (Beigneux, 2010) Human GPIHBP1 comprises four protein domains: a signal peptide (not shown), a N-terminal acidic domain, spanning amino acids 25 to 50, a lymphocyte antigen 6 (Ly6) domain from residue 65 to135 and a hydrophobic C-terminal motif that is replaced by the addition of a GPI-anchor, which itself serves to tether mature GPIHBP1 to the cell surface. In addition there are two linker sequences (Beigneux et al., 2009c; Ioka et al., 2003). The acidic domain is enriched in negatively charged amino acids such that 21 of the 26 consecutive residues are aspartate or glutamate whilst the Ly6 domain contains ten cysteine residues which, based upon the structure of other Ly6 proteins including UPAR and CD59, are thought to form disulphide bonds, producing a three-fingered structural motif. Both the acidic and Ly6 domain are important for GPIHBP1-LPL binding (see main text).

Table 7.1) Known *GPIHBP1* Mutations

Subjects (Study)	Mutation	Subjects[a]	Lipid measurements (mmol/l) TG	TC	LDL-C	HDL-C	LPL mass[b] (ng/ml)	LPL activity[b] (mU/ml)	Binding to LPL
60 chylomicronemic patients (Beigneux et al., 2009a)	Q115P	**Homozygous male[c]**	**8.40**	**3.10**	**2.10**	**0.23**	**36**	**40**	Mutant protein reaches cell surface but does not bind LPL
One chylomicronemic family (Olivecrona et al., 2010)	C65S/ C68G	**Compound heterozygous daughter**	**18.10**	**4.90**	**nd**	**0.12**	**56**	**6.3**	Mutant protein reaches cell surface but does not bind LPL
		Compound heterozygous son	**9.10**	**3.20**	**nd**	**0.31**	**52**	**11**	
		Compound heterozygous daughter	**29.30**	**7.50**	**nd**	**0.14**	**42**	**7**	
		Heterozygous father	0.73	5.70	nd	1.63	1075	261	
		Heterozygous mother	0.85	4.50	nd	1.58	984	130	
One chylomicronemic family (Franssen et al., 2010)	C65Y	**Homozygous son**	**17.78**	**3.26**	**2.15**	**0.50**	**<50**	**<10**	Mutant protein reaches cell surface but does not bind LPL
		Heterozygous father	1.66	4.68	3.26	0.92	>400	>150	
		Heterozygous mother		5.61	3.26	1.89	~200	~100	
160 chylomicronemic patients[d] (Wang and Hegele, 2007)	G56R	**Homozygous female**	**80.10**	**22.40**	**nd**	**0.55**	**nd**	**normal**	nd
		Homozygous brother	48.00	17.20	nd	0.46	nd	normal	
		Heterozygous brother	2.55	5.12	2.87	1.09	nd	nd	
		Heterozygous brother	2.90	6.30	3.93	1.05	nd	nd	
		Normal spouse	1.10	4.27	2.62	1.15	nd	nd	
		Heterozygous daughter	6.50	7.50	nd	0.75	nd	nd	

Five chylomicronemic patients (Coca-Prieto et al., 2010)	C68Y	Homozygous female	15.78	6.59	nd	0.34	Very low	Very low	nd
One chylomicronemic family (Rios et al., 2011)	Whole-gene deletion	**Homozygous male**	**420.5[e]**	**27.31**	**nd**	**nd**	**Very low**	**nd**	**nd**
		Homozygous aunt	11.08	nd	nd	nd	Very low	nd	
		Heterozygous mother	1.51	nd	nd	nd	nd	nd	
		Heterozygous father	2.34	nd	nd	nd	Normal	nd	

HDL-C=high-density lipoprotein cholesterol, LDL-C=low-density lipoprotein cholesterol, LPL=lipoprotein lipase, nd=not determined, TC=total cholesterol, TG=triglyceride

[a] Probands marked in bold

[b] LPL mass/activity measured after intra-venous injection of heparin

[c] Proband adopted, therefore preventing studies on other family members

[d] Other family members screened after mutation in proband identified in initial screen

[e] Level outside detection limits of assay

7.2) Results

7.2.1) Identification of *GPIHBP1* sequence variants

Examination of genomic data repositories (URL3, URL17) revealed evidence for a single human *GPIHBP1* transcript. This transcript is 2281bp in length and comprises 4 exons, including a 75bp 5' UTR and a 1651bp 3' UTR. The encoded protein is predicted to be 184 amino acids in length.

The entire coding region of *GPIHBP1*, plus 871bp upstream of the transcription start site was sequenced in 189 FCHL patients and 48 GP controls (Table 7.2). The selected patients had the most severe hypertriglyceridemia in the FCHL cohort (median=5.1mmol/l, range=3.42-49.70) and 187 of these individuals had triglyceride levels above the 95th age-sex-specific percentile (determined according to Heiss, et al., 1980).

A total for 40 variants were identified (Table 7.3). These included a tri-allelic tandem repeat element in intron 2 of the gene. This element consists of up to 3 repeats of a 28 or 29 base pair sequence variant (Figure 7.3). One non-synonymous variant, rs11538389, which is predicted to replace the encoded cysteine with phenylalanine, at residue 14 in the protein sequence was found. This variant is predicted by SIFT (Ng et al., 2001) not to damage the protein structure. Three rare synonymous coding SNPs (c.259C>G, rs138215873 and rs142959160) were identified in FCHL probands only and only in the heterozygous state. The remaining 37 variants detected were present in both probands and control individuals. Collectively these results imply that deleterious variants in the *GPIHBP1* coding sequence are not a major cause of the FCHL triglyceride trait.

Table 7.2) Unrelated FCHL Patients and Control Subjects used for *GPIHBP1* Resequencing

	Unrelated Patients	Matched Controls
N	189	48
Male/Female	138/51	28/20
Age (years)	49.4±10.19	52.9±8.77
TG (mmol/l)	5.10 (3.42-49.70)	1.14 (0.53-6.03)[a]
TC (mmol/l)	8.34±1.58	5.76±1.09
HDL (mmol/l)	1.00±0.25	1.69 ±0.64
BMI (kg/m^2)	26.9±2.60	24.8±3.9

Values are mean ± standard deviation except for TG which is median (range)

[a] Two controls with triglyceride levels above the 95th age- and sex-specific percentile value

Table 7.3) *GPIHBP1* Variants in 189 Severely Hypertriglyceridemic FCHL Patients and 48 Controls

	SNP ID	Location	Position (Effect)[a]	MAF (Patients)	MAF (Controls)
1	rs34407803	5' flanking region	-584G>A	0.363	0.461
2	rs34039115	5' flanking region	-561C>T	0.366	0.461
3	rs72691625	5' flanking region	-469G>A	0.196	0.092
4	rs11136268	5' flanking region	-370G>A	0.134	0.145
5	rs67963553	5' flanking region	-283G>C	0.194	0.092
6	rs61747644	Exon 1	c.12C>T (p.L4L)	0.216	0.091
7	rs11538389	Exon 1	c.41G>T (p.C14F)	0.104	0.057
8	rs13253453	Intron 1	c.52+127G>C	0.150	0.103
9	rs74880494	Intron 1	c.52+140G>A	0.104	0.083
10	VNTR	Intron 1	c.52+276-362		
			- Full sequence (Repeat unit 1+2+3)	-	0.218
			- Repeat unit 1 + 3	0.126	0.141
			- Repeat unit 3	0.39	-
11	rs11538388	Exon 2	c.138G>T (p.V46V)	0.387	0.385
12	rs75794120	Intron 2	c.181+527C>T	0.119	0.144
13	rs58301047	Intron 2	c.182-408G>C	0.190	0.080
14	rs10746561	Intron 2	c.182-387T>C	0.164	0.132
15	rs10866908	Intron 2	c.182-333G>C	0.130	0.167
16	rsxxxxxxx[b]	Intron 2	c.182-318G>A	0.003	0

	SNP ID	Location	Position (Effect)[a]	MAF (Patients)	MAF (Controls)
17	rsxxxxxxx[b]	Intron 2	c.182-275C>T	0.003	0
18	rs11136269	Intron 2	c.182-251C>G	0.392	0.452
19	rsxxxxxxx[b]	Exon 3	c.258C>G (p.G86G)	0.003	0
20	rs138215873	Exon 3	c.285C>T (p.H95H)	0.019	0
21	rs142959160	Exon 3	c.294C>T (p.T98T)	0.008	0
22	rs56046179	Intron 3	c.295+27C>T	0.105	0.048
23	rsxxxxxxx[b]	Intron 3	c.295+52delA	0.022	0.012
24	rs78632677	Intron 3	c.296-50C>T	0.003	0
25	rs112271883	Exon 4 (3' UTR)	c.*253C>T	0.066	0.034
26	rs12545739	Exon 4 (3' UTR)	c.*274C>T	0.282	0.182
27	rsxxxxxxx[b]	Exon 4 (3' UTR)	c.*275G>A	0.019	0.045
28	rs80113655	Exon 4 (3' UTR)	c.*380C>T	0.226	0.093
29	rs72691630	Exon 4 (3' UTR)	c.*394G>A	0.226	0.093
30	rsxxxxxxx[b]	Exon 4 (3' UTR)	c.*590_591insG	0.117	0.107
31	rsxxxxxxx[b]	Exon 4 (3' UTR)	c.*591G>C	0.024	0.080
32	rs111633453	Exon 4 (3' UTR)	c.*592G>C	0.209	0.091
33	rs11136270	Exon 4 (3' UTR)	c.*653A>G	0.225	0.125
34	rs11989781	Exon 4 (3' UTR)	c.*657C>T	0.142	0.167
35	rs71516016	Exon 4 (3' UTR)	c.*766delT	0.182	0.108
36	rs67149914	3' flanking region	+184C>T	0.426	0.379
37	rs66854753	3' flanking region	+200T>C	0.426	0.379

	SNP ID	Location	Position (Effect)[a]	MAF (Patients)	MAF (Controls)
38	rs13248015	3' flanking region	+247G>A	0.008	0.017
39	rs61059595	3' flanking region	+303T>C	0.293	0.107
40	rs56925758	3' flanking region	+309G>A	0.196	0.074

MAF=minor allele frequency

[a] Position relative to translation start site. Effect on transcript is given for coding SNPs

[b] Variant not currently listed in dbSNP (version 135)

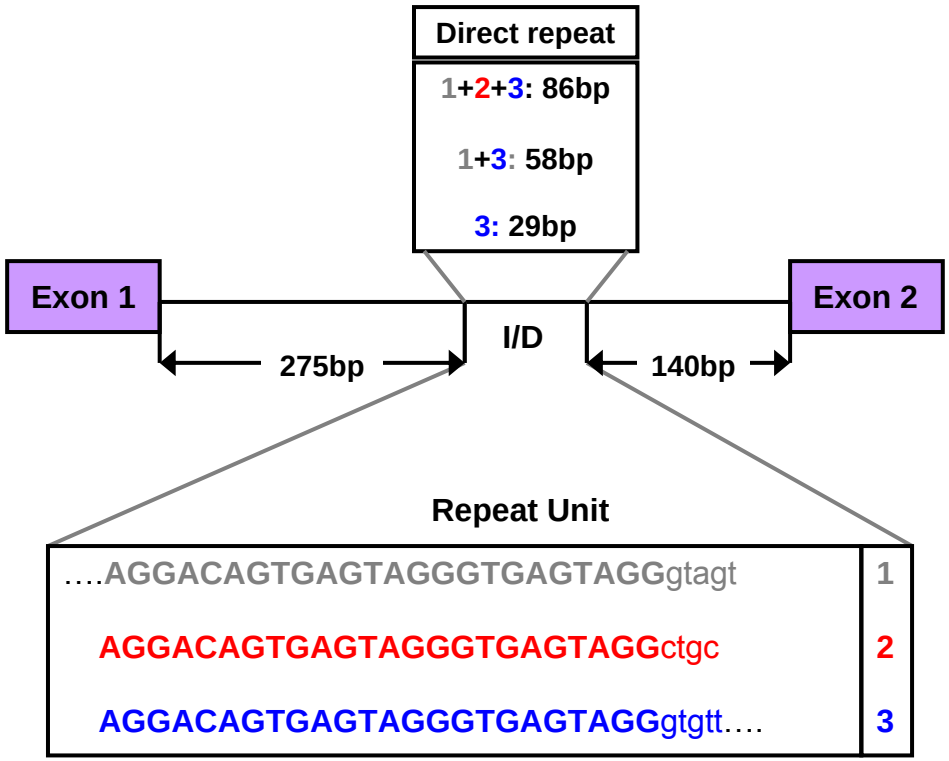


Figure 7.3) *GPIHBP1* Intron 1 Tandem Repeat Unit

A tri-allelic repeat element is found in the first intron of *GPIHBP1*. This repeat element consists of up to three repeats of a 28/29 base-pair sequence (of which a core 24 base-pairs are identical). In the resequenced FCHL patients and controls, the single repeat-unit form is in complete linkage disequilibrium with the minor allele of rs11538388.

7.2.2) Association of rs11538388 with Hyperlipidemia and CHD

Using patient and control genotype data from the resequencing dataset, six major haplotype groups, with frequencies greater than 0.01, were defined (Table 7.4). Based on these haplotypes, six haplotype-tagging SNPs were selected for genotyping in 1725 members of the FCHL cohort (Table 7.5). Of note, the single-unit form of the intron 2 tandem repeat element is in complete linkage disequilibrium with the T allele of SNP rs11538388 in these data (Table 7.4).

In quantitative trait association analyses (Table 7.6), rs11538388 returned a *P*-value of *P*=0.0035 for association with triglyceride levels and of *P*=0.0026 for association with total cholesterol levels. Specifically, the rs11538388 T allele was associated with lower triglyceride and cholesterol levels in the FCHL cohort. In addition, the rs11538388 T allele was present at a lower frequency in FCHL probands versus family spouses (0.40 vs. 0.49, *P*=0.0094) (Table 7.7).

Table 7.4) Major *GPIHBP1* Haplotypes in the FCHL Cohort and Control Resequencing Dataset

Haplotype	SNPs							Frequency[a]	
	rs61747644	rs11538389	VNTR	rs11538388	rs10746561	rs10866908	c.295+52delA	Patients	Controls
1	**T**	G	Full	G	C	G	A	0.199	0.078
2	C	**T**	Full	G	C	G	A	0.102	0.051
3	C	G	3	**T**	C	G	A	0.391	0.593
4	C	G	Full	G	**T**	G	A	0.150	0.105
5	C	G	1+3	G	C	**C**	A	0.117	0.143
6	C	G	Full	G	C	G	**T**	0.018	0.012

[a] Frequencies are derived from haplotypes with a frequency of >1% in the resequencing data

Table 7.5) Characteristics of 1725 Individuals from the FCHL Cohort Genotyped for *GPIHBP1* Haplotype-Tagging SNPs

				TG Trait[c]		TC Trait[d]	
	All Subjects	Probands	Spouses	Affected	Unaffected	Affected	Unaffected
N (Male/Female)	858/867	171/68	86/158	376/271	448/538	412/253	423/567
Age (years)	44.28±16.35	50.02±10.06	55.59±8.32	46.17±15.44	42.67±16.72	47.05±14.60	42.41±17.12
BMI (kg/m^2)	25.56±4.09	26.67±2.34	25.90±4.17	26.27±3.15	24.52±3.87	26.06±2.94	24.64±4.05
Serum TG (mmol/l)[a]	1.75	4.21	1.34	3.50	1.21	3.28	1.27
	(0.32-49.7)	(1.35-20.4)	(0.42-5.49)	(1.21-49.7)	(0.32-2.79)	(0.49-49.7)	(0.32-7.88)
Log-transformed TG[b]	0.63±0.70	1.53±0.46	0.30±0.49	1.31±0.50	0.16±0.40	1.13±0.65	0.28±0.52
Serum TC (mmol/l)	6.37±1.62	8.26±1.30	5.97±1.12	7.41±1.62	5.67±1.22	7.86±1.24	5.38±0.97

TG=triglyceride, TC= total cholesterol, BMI=body mass index

Age, BMI, log-transformed triglyceride level and cholesterol level are expressed as mean±standard deviation.

[a] Median values (range)

[b] Raw triglyceride levels were transformed (using log to the base e) in order to normalise the quantitative trait distribution.

[c] TG affected=triglyceride level ≥95th percentile age- and sex-specific value

[d] TC affected=cholesterol level ≥95th percentile age- and sex-specific value

Table 7.6) Quantitative Trait Association Analyses for *GPIHBP1* Haplotype-Tagging SNPs in FCHL Families

				Triglyceride[b]		Total Cholesterol	
SNP	Alleles[a]	MAF (SE)	HWE-*P*	β (SE)	*P*	β (SE)	*P*
rs61747644	C/T	0.1900 (0.0102)	0.4511	0.0268 (0.0252)	0.2875	0.1239 (0.0673)	0.0657
rs11538389	G/T	0.0975 (0.0077)	0.3454	0.0133 (0.0330)	0.6868	0.0081 (0.0887)	0.9272
rs11538388	**G/T**	**0.4536 (0.0134)**	**0.4966**	**-0.0603 (0.0207)**	**0.0035**	**-0.1661 (0.0553)**	**0.0026**
rs10746561	C/T	0.1157 (0.0084)	0.4419	0.0108 (0.0317)	0.7324	0.0320 (0.0846)	0.7050
rs10866908	G/C	0.0988 (0.0078)	0.2300	0.0396 (0.0325)	0.2237	0.0295 (0.0876)	0.7368
c.295+52delA	A/T	0.0185 (0.0035)	0.6133	0.1050 (0.0703)	0.1353	0.4409 (0.1889)	0.0196

Age, sex and BMI were included as covariates in the analyses

HWE=Hardy-Weinberg Equilibrium, MAF=minor allele frequency, SE=standard error

[a] Major/minor allele

[b] Raw triglyceride levels were transformed (using log to the base e) in order to normalise the quantitative trait distribution

Table 7.7) Allele Frequencies of *GPIHBP1* Haplotype-Tagging SNPs in FCHL Probands and Family Spouses

		Minor Allele Frequency		
SNP	Alleles[a]	Probands	Spouses	*P*[b]
rs61747644	C/T	0.19	0.16	0.1735
rs11538389	G/T	0.13	0.09	0.0612
rs11538388	**G/T**	**0.40**	**0.49**	**0.0094**
rs10746561	C/T	0.11	0.12	0.9156
rs10866908	G/C	0.11	0.10	0.8313
c.295_52delA	A/T	0.02	0.02	0.6445

[a] Major/minor allele

[b] Significance was generated using Fisher's exact test

To substantiate the stage one data, in stage two of the study, rs11538388 was genotyped in four additional datasets (Table 7.8). Case-control association analysis was performed using logistic regression. The British and UCSF combined hyperlipidemia (CHL) cohorts were both ascertained using the same criteria, thus these datasets were combined and analysed jointly (cases, n=280, controls, n=784). Similarly, the UCSF and Dutch hypertriglyceridemic (HTG) case-control cohorts were analysed jointly (cases, n=308, controls, n=362).

Neither CHL nor HTG were associated with rs11538388 at a *P*-value of $P<0.05$ (Table 7.9). However, the direction of effect was the same as that observed in the FCHL families. Thus, in these cohorts the rs11538388 T allele was protective against development of the hyperlipidemic phenotype, with odds ratios of 0.831 and 0.879, respectively (Table 7.9).

The combined analysis of the family-based and case-control cohorts, rs11538388 returned a *P*-value of $P=2.98\times10^{-4}$) for association with hypertriglyceridemia (Table 7.9). After applying a Bonferroni correction for seven tests (six SNPs tested in stage one and one SNP in the combined analysis) the resultant *P*-value was $P=2.09\times10^{-3}$.

To extend the investigation, the association between rs11538388 and CHD was examined in two independent cohorts: 1) 715 men from the Southampton Atherosclerosis study (SAS), all with CHD diagnosed by a coronary angiogram; and 2) 2394 men from the Northwick Park Heart Study II (NPHSII), a prospective study where participants were followed for up to 15 years to record CHD events.

In the SAS cohort, rs11538388 was associated with delayed onset of overt CHD (Table 7.10). Specifically, by logistic regression, dichotomising individuals according to age of diagnosis above or below the median value (62.4 years), the rs11538388 decreased the odds of diagnosis in the 'young' age group (OR=0.7627, $P=0.0146$). Likewise, in the NPHSII, the frequency of the rs11538388 T allele was higher in those men who did not develop CHD during the follow-period (controls, frequency=0.46), compared to those who did (cases, frequency=0.42), albeit not significantly ($P=0.1435$) (Table 7.11). Those individuals carrying at least one copy of the T allele had a decreased prospective CHD risk (hazard ratio=0.76; 95% CI: 0.58-1.00; $P=0.048$) (Table 7.11). In summary, data from the SAS and NPHSII suggest that the rs11538388 T allele is protective against the development of high plasma triglyceride levels and CHD.

Table 7.8) Characteristics of Stage Two Combined Hyperlipidemia and Hypertriglyceridemia Cohorts

	Combined Hyperlipidemia Cohorts[c]				Hypertriglyceridemia Cohorts[c]			
	UK	Controls	UCSF	Controls	UCSF	Controls	Dutch	Controls
N (M/F)	85/37	279/267	68/90	47/191	106/52	69/113	118/32	89/91
Age (Years)	49.21±9.74	52.7±8.24	46.98±16.44	39.48±15.18	50.21±12.88	52.01±18.37	50.84±11.97	50.23±16.24
BMI (kg/m^2)	26.37±2.35	25.25±4.04	25.75±3.04	23.97±4.64	26.17±2.50	24.78±4.54	na	na
TG (mmol/l)[a]	3.90	1.21	3.53	0.86	12.00	0.97	12.21	1.01
logTG[b]	1.50±0.48	0.22±0.50	1.36±0.45	-0.15±0.34	2.56±0.72	-0.05±0.35	2.70±0.85	0.00±0.29
TC (mmol/l)	8.16±1.62	5.74±1.08	9.34±2.45	4.50±0.80	9.94±4.86	4.77±0.73	7.79±3.67	5.18±0.83

na=not available, TG=triglyceride, TC=total cholesterol, BMI=body mass index

Age, BMI, log-transformed triglyceride level and cholesterol level are expressed as mean±standard deviation.

[a] Median values

[b] Raw triglyceride levels were transformed (using log to the base e) in order to normalise the quantitative trait distribution.

[c] All individuals were of Caucasian background. Exclusion criteria were: age below 16 years and possible secondary hyperlipidemia due to Type 1 or 2 Diabetes or obesity (BMI>30).

Table 7.9) Association of rs11538388 in FCHL Families, Combined Hyperlipidemia and Hypertriglyceridemia Cases and Controls

Stage 1	MAF	β	*Z*	*P*
FCHL Families	0.40/0.49[a]	-0.0603	-2.919	0.0035
Stage 2		OR	*Z*	*P*
CHL Cases & Controls	0.427/0.475[b]	0.831	-1.879	0.0602
HTG Cases & Controls	0.425/0.458[b]	0.879	-1.159	0.2466
Stage 1& 2			*Z*	*P*
Combined Analysis			-3.617	2.98×10^{-4}

CHL=combined hyperlipidemia, HTG=hypertriglyceridemia, MAF=minor allele frequency

[a] Minor allele freuqency in FCHL probands/spouses

[b] Minor allele frequency in cases/controls

Table 7.10) Association of rs11538388 with Age of Diagnosis in the Southampton Atherosclerosis Study

SNP	Effect Allele	Frequency	Age of Diagnosis[a]	N	Test[b]	OR	β	*P*
rs11538388	T	0.451	All	715	Linear regression	-	0.9474	0.0701
		0.425	Early onset<62.4 years	366	Logistic regression	0.7627	-0.1177	0.0146
		0.477	Late onset≥ 62.4 years	367				

[a] For case-control analyses, individuals were dichotomised according to cases=late-onset, age≥62.4 years (median age), controls=early-onset, age<62.4 years (median age)

[b] BMI and smoking status were incorporated as covariates

Table 7.11) Association of rs11538388 with Prospective CHD Risk in the Northwick Park Heart Study II

a) Allelic association

SNP	Alleles	Cases	Controls	OR[a]	*P*[c]
rs11538388	G	0.577	0.540	0.8631	0.1435
	T	0.423	0.460		

b) Genotypic Association (analyses performed by Philippa Talmud)

		Frequency			
SNP	Genotype	Cases	Controls	HR (95% CI)[b]	*P*
rs11538388	G/G	0.357	0.297	1	
	G/T	0.438	0.486	0.74 (0.56-0.99)	0.1300
	T/T	0.204	0.217	0.81 (0.57-1.16)	
	GG	0.357	0.297	1	0.0480
	G/T+T/T	0.643	0.703	0.76 (0.58-1.00)	

95% CI= 95% confidence intervals, HR=hazard ratio, OR=odds ratio

[a] OR not age-adjusted

[b] HR obtained from Cox regression models with adjustment for age

[c] Frequencies were compared using Fisher's exact test

7.3) Discussion

Clinical and genetic data suggest that variants in genes involved in the lipolytic processing of triglyceride-rich lipoproteins may influence both circulating triglyceride levels and the risk of developing CHD (Angelakopoulou et al., 2011; Sarwar et al., 2010; Teslovich et al., 2010; Varbo et al., 2011). GPIHBP1 is thought to facilitate the transport of the essential lipolytic enzyme, LPL, from the basolateral to apical endothelial cell surface (Davies et al., 2010). Since *GPIHBP1* loss-of-function mutations explain a small number of cases of chylomicronemia, the current study sought to examine the contribution of *GPIHBP1* variants to the lipid phenotypes of FCHL.

The data presented in this chapter indicate that a common *GPIHBP1* variant, rs11538388, is associated with plasma triglyceride levels in FCHL families, with the rare T allele conferring an additive, protective effect. In addition, the rs11538388 minor allele frequency is depleted in FCHL probands compared to spouse controls (0.40 versus 0.49, *P*=0.0094). Data also suggest that the protective effect extends to a decreased risk of CHD in carriers of the minor allele (Tables 7.10 and 7.11). In other words, the common allele at this SNP locus predisposes to CHD, potentially through impaired post-prandial catabolism of the triglyceride-rich core of circulating lipoproteins.

Guay et al., have recently described an association between another *GPIHBP1* SNP (rs72691625) and plasma triglyceride levels (Guay et al., 2011). In their study, the rare A allele at this locus was present at frequency of 0.24 in 278 hypertriglyceridemic individuals (triglyceride>2mmol/l) compared to a frequency of 0.16 in 263 'normolipidemic' individuals (triglyceride>2mmol/l, *P*<0.041) (Guay et al., 2011). In the FCHL cohort, rs61747644 was genotyped and acted as a proxy for rs72691625 (r^2=1 based on patient/control sequencing data). As shown in Table 7.7, the frequency of the rare allele at the rs61747644 locus in the FCHL spouse controls was 0.16, the same as observed in the controls of the Guay et al., study (Guay et al., 2011). However, the frequency of the rs61747644 rare allele was not significantly higher in the FCHL probands (0.19, *P*=0.1735) and in the FCHL families, quantitative analyses showed no significant association between rs61747644 and plasma triglyceride levels (*P*=0.2875). In contrast, in quantitative analyses involving the entire Guay et al., dataset, (a total of 541 individuals) rs72691625 had an additive effect upon plasma triglyceride levels (triglyceride: GG, 2.15mmol/l; GA, 2.43mmol/l; AA, 3.01mmol/l, *P*=0.029) (Guay et al., 2011).

In the recent GLGC meta-analysis, no common *GPIHBP1* variants were significantly associated with plasma triglyceride levels (Teslovich et al., 2010), Thus, rs11136269 which is in total LD with rs11538388 (r^2=1, in CEU individuals) returned a *P*-value of *P*=0.3795 for association with plasma triglyceride levels and of *P*=0.4857 for association with plasma cholesterol levels. The discrepancy between these results might reflect differences in the genetic backgrounds of the study cohorts. The GLGC study involved data from 46 individual GWAS which comprised non-ascertained, family- and population-based samples as well as diabetes, CHD, obesity, metabolic syndrome and hypertension case-control cohorts (Teslovich et al., 2010). In contrast, the FCHL cohort represents a highly ascertained and well-characterised patient group who are likely to be enriched for genetic variants (rare and/or common) that impact upon many of the biological process that regulate circulating lipid levels. A recent investigation demonstrated that the rates of triglyceride turnover and storage in the adipose tissue of FCHL patients differ significantly in comparison to those in both obese and non-obese subjects (Arner et al., 2011). Therefore, it seems plausible that the genetic variants underlying these traits and in turn plasma triglyceride levels will differ between FCHL patients and many of the subjects analysed in the GLGC study.

Given that the FCHL phenotype is likely to arise from different combinations of rare and common variants in different families (Horswell et al., 2009), it is also probable that the effect of rs11538388 on circulating triglyceride levels may only be apparent in presence of a further, as yet undetermined genetic variant(s). This multi-locus effect is similar to that observed in dysbetalipoproteinemia where ApoE ε2 homozygotes require a 'second hit' for phenotypic presentation (Groenewegen et al., 1994). Consistent with this, it is now known that interplay between LDLR and ApoE isoforms can influence the clearance of cholesterol-rich lipoproteins (reviewed in (Calandra et al., 2011)).

Since GPIHBP1 serves a vital role in facilitating LPL-mediated hydrolysis of triglyceride-rich lipoproteins (Beigneux et al., 2007; Davies et al., 2010), the data presented in this chapter suggest that the rare allele at SNP site rs11538388 equates to gain-of-function variant, whereby lipolysis and thus post-prandial clearance of triglyceride-rich lipoproteins is expedited, leading to decreased plasma triglyceride levels and CHD risk. A potential mechanism for the rs11538388 gain-of-function effect could involve the tandem repeat element that occurs in the first intron of

GPIHBP1. As shown in Figure 4.2, this element consists of up to three repeats of a 28 or 29-base pair sequence. The short form of this element, comprising only one copy of the repeat sequence, is in total LD with the rs11538388 rare (T) allele (in the FCHL cohort). Importantly, this short form represents a deletion as compared the current human genome assembly (hg19) and based upon analyses using the Neural Network splice site prediction tool (URL23, Reese et al., 1997), this deletion abolishes four predicted splice donor sites (Figure 4.3). Each of these splice donor sites is scored higher (more likely to be a true positive) by the prediction program than the only currently known *GPIHBP1* exon/intron 1 splice site (0.94, 0.99, 0.93 and 0.99 versus 0.78). At present there are no EST data to support the presence of multiple *GPIHBP1* transcripts. However, it remains possible that in carriers of the rs11538388 rare allele, pre-mRNA processing generates more copies of the primary *GPIHBP1* transcript than in homozygotes for the common allele.

Supporting the proposition of differential pre-mRNA processing, a similar effect was observed following an association study involving 2346 individuals from the Micronesian Island of Kosrae where the major alleles several *HMGCR* variants in LD with rs3846662 (r^2=0.82-0.93) were associated with decreased plasma LDL-cholesterol levels (Burkhardt et al., 2008). In cells expressing the major (A) allele at SNP site rs3846662, increased numbers of *HMGCR* transcripts lacking exon 13 of the gene were produced. This exon skipping event is expected to produce a protein that lacks 53 amino acids from its catalytic domain. The authors hypothesise that this would lead to decreased HMGCR enzymatic activity, decreased cellular cholesterol synthesis and in turn increased cellular uptake of LDL-cholesterol from the circulation, thus supporting their association study results (Burkhardt et al., 2008).

An alternative explanation for the putative gain-of-function effect of the rs11538388 minor allele might be that the intron one repeat element alters the regulation of *GPIHBP1* transcription, either by acting as a regulatory domain that recruits transcriptional activators or repressors, or by adopting a DNA structure that inhibits gene transcription. Similar phenomena are thought to underlie Friedreich's ataxia (Kim et al., 2011; Sakamoto et al., 2001) and may contribute to the association between an intronic tandem repeat element in the serotonin transporter gene and various neurological conditions (Klenova et al., 2004).

The current study identified only one non-synonymous *GPIHBP1* variant, rs11538389, which was present in both FCHL patients and unaffected individuals.

This variant alters the 14th amino acid of the protein from cysteine to phenylalanine, a change which *in silico* is predicted not to damage the encoded protein (Ng and Henikoff, 2001). However, in future work it would be prudent to verify this prediction *in vitro*. In this regard, it is worth noting that the minor allele frequency of rs11538389 was slightly higher in FCHL probands compared to spouses (0.13 versus 0.09, *P*=0.0612). Although there was no association between this SNP and plasma triglyceride or cholesterol levels in the FCHL families, this may relate to a lack of power to detect association between a relatively rare variant of small effect size and plasma lipid levels, in the FCHL cohort. Further genotyping of this variant in additional FCHL families may therefore be warranted.

Finally, by sequencing *GPIHBP1* in 189 severely hypertriglyceridemic FCHL patients and 48 controls, just three synonymous variants were identified in the patient group only. Three further variants that were unique to patient group were intronic. These data suggest that rare *GPIHBP1* coding variants do not to contribute to the FCHL triglyceride trait. This finding is in agreement with the recently published study of Johansen et al., which found no enrichment of rare *GPIHBP1* coding variants in 413 hypertriglyceridemic patients, as compared to 324 population-based controls (Johansen et al., 2011c).

To conclude, the data presented in this chapter demonstrate, for the first time, that a common genetic variant in *GPIHBP1* is associated with hypertriglyceridemia and CHD risk. Thus, these genetic data provide further support for the clinically observed correlation between plasma triglyceride levels and CHD (Arsenault et al., 2009; Bansal et al., 2007; Tirosh et al., 2007), an effect potentially mediated by a delay in the catabolism of circulating atherogenic lipoproteins.

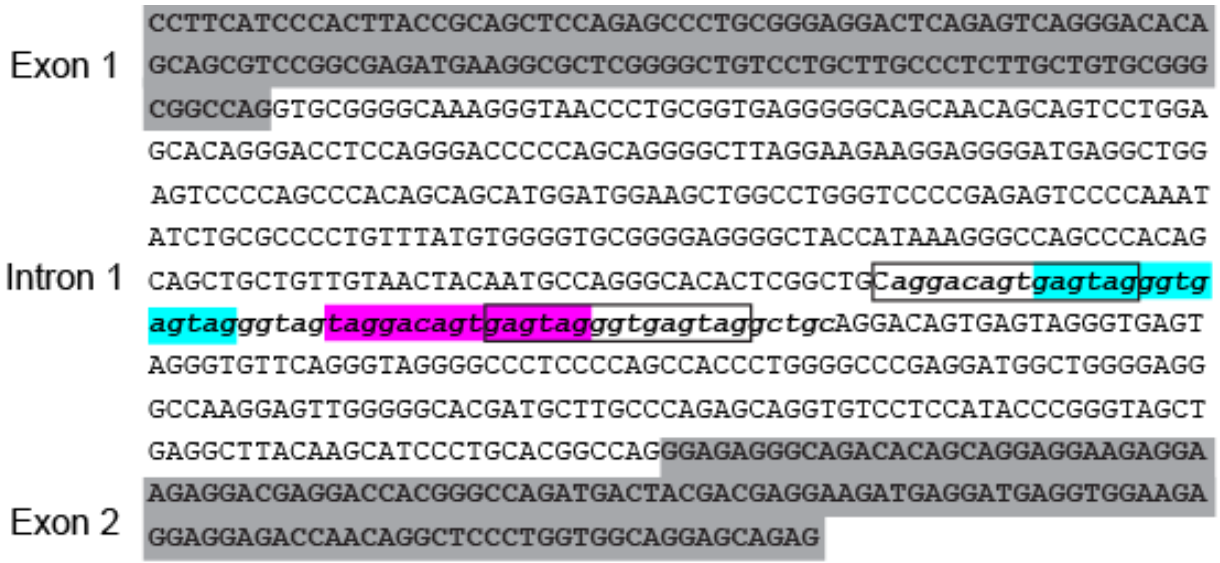


Figure 4.3) Predicted Splice Donor Sites in *GPIHBP1* Intron 1

The short form of the *GPIHBP1* intron 1 tandem repeat represents a deletion of 57 base pairs compared to the human genome reference sequence (small case italics). Deletion of this sequence removes four predicted splice donor sites (boxed or pink or blue shading). Exons are marked by grey boxes.

Concluding Remarks

Human plasma contains a diverse assortment of lipid species (Quehenberger et al., 2010) which, in part, may reflect the vast range of metabolic processes that contribute to whole-body lipid homeostasis. Indeed, historical genetic studies of Mendelian disorders and more recent genome-wide investigations have served to demonstrate that a heterogeneous array of genetic variants, with varying frequencies and effect sizes, contribute to variation in circulating triglyceride and cholesterol levels and that they mediate the observed phenotypes through a range of molecular and cellular mechanisms (Table 1.1, (Calandra et al., 2011; Johansen et al., 2011a)).

In this thesis, a multi-pronged approach, combining linkage, association and gene-expression data, was employed to investigate the contributions of biological candidates and of previously unexpected genes to the aetiology of FCHL. Association analyses focussed upon common variants within the 21q22.3 genomic interval because evidence from genome-wide linkage analyses and subsequent fine-mapping involving 182 FCHL families strongly suggests that this interval harbours variants that increase the transmission of both the FCHL-triglyceride and FCHL-cholesterol traits. In support of this, the region was implicated in modulating plasma triglyceride and LDL-cholesterol levels in earlier genome-wide linkage screens involving families with cardiovascular disease (Coon et al., 2002; North et al., 2005).

The most interesting results to emerge from the culmination of data presented in Chapters 5 and 6 relate to *C21orf57*, *CBS, COL18A1*, *PKNOX1*, *TRPM2* and *TSPEAR*. Existing data for *CBS* (Namekata et al., 2004; Werstuck et al., 2001) and *COL18A1* (Bishop et al., 2010) suggest that loss-of-function mutations in these genes could underlie the FCHL-triglyceride trait, and potentially the cholesterol trait. Thus in future studies it will be important to establish, using targeted next-generation sequencing, in a large number of FCHL patients, whether these genes harbour multiple variants that have a moderate to large impact on plasma lipid levels in FCHL families. A recent study suggests that it would be sensible to incorporate family-specific linkage information when selecting FCHL patients for discovery phase sequence analysis (Shi et al., 2011).

It is unknown whether gain- or loss-of-function *C21orf57*, *PKNOX1*, *TRPM2* and *TSPEAR* variants would impact on lipid metabolism, and indeed whether such

variants are contributory to FCHL. Thus, future studies of these genes would initially focus upon determining their sites of expression and examine the effects of gene silencing upon, for example, the expression of other genes involved in the secretion of apoB-containing lipoproteins.

One potential limitation of the targeted association study design adopted in this thesis was that common variants in 32 of the 61 protein-coding genes within the refined linkage interval were not considered. However, it is important to note that data from the GLGC study, published whilst this investigation was underway, revealed no significant association between common variants within these 32 genes and plasma triglyceride or cholesterol levels (Teslovich et al., 2010). Furthermore, consultation of the Mouse Genome Database (Blake et al., 2011) reveals no phenotypic evidence that mutation of any of these genes lead to abnormal lipid metabolism.

A further limitation of the gene expression analyses performed in this thesis was the sole use of bezafibrate treatment in McA-RH7777 cells to identify potential FCHL candidate genes. Since this approach would be most likely to only impact cellular pathways involved in fatty acid metabolism, future studies could involve *in vitro* treatment with a statin or LXR-agonist to modulate cholesterol metabolism. Additionally, it may be informative to extend future studies to a human cell line in order to consider species differences in gene regulation.

The final chapter of this thesis also considered the contribution of variants in the excellent FCHL candidate gene, *GPIHBP1,* to hypertriglyceridemia and CHD. Although rare *GPIHBP1* mutations are known to cause rare cases of chylomicronemia, they were found not to contribute to the elevated lipid levels of FCHL. However, the minor allele of the common *GPIHBP1* variant, rs11538388 was associated with lower triglyceride levels and with decreased CHD risk. This protective effect may be conferred by an intronic repetitive element in LD with rs11538388. Future studies could examine this possibility using a minigene approach similar to that of Burkhardt and colleagues (Burkhardt et al., 2008).

In sum, the data presented in this thesis implicate both obvious biological candidates and previously unsuspected genes in the aetiology of FCHL. Future investigations will need to harness the power of next-generation sequencing to examine the contribution of variants of moderate to large effect, in one or more 21q22.3 gene(s), to the lipid abnormalities of FCHL.

Appendix I) Lipid Percentiles (from Heiss et al., 1980)

Female Triglyceride Age-Specific Percentile Levels (mmol/l)

Age (years)	5%	10%	25%	50%	75%	90%	95%
0-4	0.38	0.43	0.51	0.67	0.87	1.09	1.27
5-9	0.36	0.40	0.50	0.62	0.80	1.02	1.19
10-14	0.42	0.50	0.61	0.79	1.02	1.29	1.48
15-19	0.44	0.50	0.60	0.77	0.98	1.29	1.50
20-24	0.45	0.53	0.68	0.92	1.22	1.60	1.87
25-29	0.45	0.52	0.65	0.88	1.21	1.61	1.95
30-34	0.46	0.52	0.65	0.87	1.20	1.65	2.00
35-39	0.46	0.53	0.68	0.91	1.28	1.81	2.20
40-44	0.53	0.61	0.73	1.00	1.38	1.93	2.37
45-49	0.53	0.61	0.78	1.06	1.50	2.09	2.59
50-54	0.61	0.69	0.87	1.14	1.60	2.18	2.70
55-59	0.63	0.73	0.92	1.22	1.68	2.30	2.92
60-64	0.64	0.72	0.92	1.22	1.68	2.28	2.72
65-69	0.68	0.75	0.94	1.27	1.78	2.30	2.73
70+	0.68	0.78	0.96	1.26	1.70	2.30	2.63

Male Triglyceride Age-Specific Percentile Levels (mmol/l)

Age (years)	5%	10%	25%	50%	75%	90%	95%
0-4	0.32	0.37	0.45	0.57	0.76	0.95	1.12
5-9	0.34	0.37	0.45	0.57	0.73	0.96	1.14
10-14	0.36	0.42	0.51	0.67	0.88	1.15	1.42
15-19	0.42	0.48	0.61	0.78	1.03	1.36	1.68
20-24	0.50	0.56	0.71	0.97	1.35	1.87	2.28
25-29	0.52	0.61	0.79	l.07	1.54	2.26	2.82
30-34	0.56	0.65	0.85	1.18	1.69	2.42	3.02
35-39	0.61	0.70	0.92	1.28	1.93	2.85	3.64
40-44	0.62	0.72	0.97	1.38	1.97	2.81	3.63
45-49	0.65	0.77	1.01	1.40	1.97	2.87	3.71
50-54	0.65	0.77	0.98	1.40	2.04	2.84	3.63
55-59	0.65	0.76	0.98	1.35	1.93	2.67	3.25
60-64	0.65	0.77	0.98	1.35	1.92	2.67	3.30
65-69	0.64	0.72	0.94	1.27	1.69	2.36	3.03
70+	0.65	0.71	0.94	1.26	1.69	2.40	2.93

Female Cholesterol Age-Specific Percentile Levels (mmol/l)

Age (years)	5%	10%	25%	50%	75%	90%	95%
0-4	2.87	3.07	3.56	4.00	4.41	4.84	5.12
5-9	3.23	3.43	3.74	4.17	4.58	5.00	5.20
10-14	3.17	3.35	3.69	4.05	4.46	4.87	5.15
15-19	3.07	3.25	3.58	3.97	4.41	4.89	5.20
20-24	3.20	3.43	3.82	4.35	4.87	5.48	5.84
25-29	3.33	3.58	3.97	4.43	4.94	5.53	5.87
30-34	3.35	3.61	4.05	4.51	5.02	5.69	6.10
35.39	3.61	3.79	4.20	4.69	5.25	5.84	6.28
40-44	3.76	3.97	4.38	4.92	5.51	6.07	6.48
45-49	3.89	4.12	4.56	5.15	5.79	6.43	6.87
50-54	4.17	4.43	4.92	5.51	6.15	6.82	7.30
55-59	4.33	4.64	5.12	5.84	6.35	7.07	7.53
60-64	4.43	4.69	5.17	5.79	6.46	7.12	7.58
65-69	4.38	4.69	5.25	5.79	6.46	7.10	7.61
70+	4.28	4.61	5.10	5.74	6.43	7.07	7.38

Male Cholesterol Age-Specific Percentile Levels (mmol/l)

Age(years)	5%	10%	25%	50%	75%	90%	95%
0-4	2.92	3.20	3.51	3.87	4.38	4.76	5.20
5-9	3.10	3.33	3.71	4.07	4.48	4.89	5.20
10-14	3.05	3.25	3.58	3.97	4.43	4.87	5.17
15-19	2.89	3.07	3.38	3.74	4.23	4.69	5.05
20-24	3.17	3.33	3.74	4.23	4.76	5.23	5.58
25-29	3.41	3.66	4.07	4.56	5.17	5.82	6.25
30-34	3.50	3.79	4.28	4.87	5.46	6.12	6.51
35-39	3.74	4.02	4.51	5.05	5.71	6.38	6.92
40-44	3.87	4.17	4.66	5.20	5.84	6.41	6.87
45-49	4.05	4.33	4.82	5.38	6.00	6.61	7.07
50-54	4.05	4.33	4.79	5.38	6.00	6.69	7.10
55-59	4.00	4.28	4.84	5.43	6.02	6.71	7.07
60-64	4.07	4.38	4.82	5.38	6.02	6.64	7.06
65-69	4.05	4.35	4.87	5.38	5.97	6.61	7.02
70+	3.87	4.15	4.66	5.25	5.87	6.41	6.97

Appendix II) List of Suppliers

Applied Biosystems/Invitrogen
(Life Technologies Ltd.)
3 Fountain Drive
Inchinnan Business Park
Paisley
PA4 9RF, UK

Beckman Coulter (UK) Ltd.
Oakley Court
Kingsmead Business Park
London Road
High Wycombe
HP11 1JU, UK

Biocomputing Platforms Ltd.
Innopoli 2
Tekniikantie 14
FI-02150 Espoo, Finland

Bioline Reagents Ltd.
Unit 16 The Edge Business Centre
Humber Road
London
NW2 6EW, UK

Biosera Ltd.
1 Acorn House
The Broyle
Ringmer
East Sussex, UK

Bruker Daltonics UK
Banner Lane
Coventry
CV4 9GH, UK

Eurogentec Ltd.
Old Headmasters House
Unit 1, Building 1
Forest Business Centre
Fawley Road, Fawley
Southampton
SO45 1FJ, UK

Finnzymes/
Thermo Fisher Scientific UK Ltd.
Bishop Meadow Road
Loughborough
Leicestershire
LE11 5RG, UK

Illumina, Inc.
5200 Research Place
San Diego
CA 92122, USA

LGC Standards
Queens Road
Teddington
Middlesex
TW11 0LY, UK

Qiagen
Qiagen House
Fleming Way
Crawley
West Sussex
RH10 9NQ, UK

Sequenom (Europe)
Mendelssohnstrasse 15D
D-22761
Hamburg, Germany

Sigma-Aldrich Company Ltd.
The Old Brickyard
New Road
Gillingham
Dorset
SP8 4XT, UK

Stratagene
Agilent Technologies, Inc.
5301 Stevens Creek Boulevard
Santa Clara,
CA 95051-7201, USA

Technelysium Pty Ltd.
403/132 Alice St
Brisbane
Queensland 4000, Australia

Tepnel Life Sciences
(Gen-Probe Life Sciences Ltd.)
Heron House
Oaks Business Park
Crewe Road
Wythenshawe
Manchester
M23 9HZ, UK

Appendix III) List of URLs

URL1	http://www.bhf.org.uk/heart-health/statistics/mortality.aspx
URL2	http://research.marshfieldclinic.org/genetics/GeneticResearch/compMaps.asp
URL3	http://www.ensembl.org/
URL4	http://www.ncbi.nlm.nih.gov/pubmed
URL5	http://www.ncbi.nlm.nih.gov/unigene
URL6	http://www.ncbi.nlm.nih.gov/snp/
URL7	http://hapmap.ncbi.nlm.nih.gov/
URL8	http://www.repeatmasker.org/
URL9	http://browser.1000genomes.org/index.html
URL10	http://genome.sph.umich.edu/wiki/MaCH
URL11	http://gump.qimr.edu.au/general/daleN/SNPSpDlite/
URL12	http://hydra.usc.edu/gxe/
URL13	http://csg.sph.umich.edu/locuszoom/
URL14	http://www.broadinstitute.org/mpg/snap/ldsearchpw.php
URL15	http://www.sph.umich.edu/csg/abecasis/public/lipids2010/
URL16	http://filezilla-project.org/
URL17	http://genome.ucsc.edu/
URL18	http://www.ncbi.nlm.nih.gov/tools/primer-blast/
URL19	http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi
URL20	http://www.sanger.ac.uk/resources/software/genevar/
URL21	http://www.ensembl.org/biomart/martview
URL22	http://projects.tcag.ca/variation/
URL23	http://www.fruitfly.org/seq_tools/splice.html

Appendix IV) Names and Addresses of Supervisors and Collaborators

Professor Carol Shoulders
Centre for Endocrinology, William Harvey Research Institute,
John Vane Science Centre, Charterhouse Square, Barts & The London School of Medicine and Dentistry, Queen Mary University of London,
London, EC1M 6BQ

Mr Stuart Horswell
Bioinformatics and BioStatistics, Cancer Research UK,
44 Lincoln's Inn Fields, Holborn,
London, WC2A 3LY

Dr Geesje Dallinga-Thie
Department of Vascular and Experimental Vascular Medicine G1.113,
AMC, Meibergdreef 9, 1105AZ, Amsterdam, The Netherlands

Professor Phillipa Talmud
Centre for Cardiovascular Genetics, Rayne Building, University College London,
5 University St, London, WC1E 6JF

Dr Andrew Walley
Department of Genomics of Common Disease, School of Public Health,
Imperial College London, Room E303, Burlington-Danes Building,
Hammersmith Hospital, Du Cane Road, London, W12 0NN

Professor Shu Ye
Centre for Clinical Pharmacology, William Harvey Research Institute,
John Vane Science Centre, Charterhouse Square, Barts & The London School of Medicine and Dentistry, Queen Mary University of London,
London, EC1M 6BQ

References

Abbasi, F., Chen, Y. D., Farin, H. M., Lamendola, C., and Reaven, G. M. (2008). Comparison of three treatment approaches to decreasing cardiovascular disease risk in nondiabetic insulin-resistant dyslipidemic subjects. Am J Cardiol *102*, 64-69.

Abecasis, G. R., Cherny, S. S., Cookson, W. O., and Cardon, L. R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet *30*, 97-101.

Abifadel, M., Varret, M., Rabes, J. P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D.*, et al.* (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat Genet *34*, 154-156.

Agarwal, A. K., Arioglu, E., De Almeida, S., Akkoc, N., Taylor, S. I., Bowcock, A. M., Barnes, R. I., and Garg, A. (2002). AGPAT2 is mutated in congenital generalized lipodystrophy linked to chromosome 9q34. Nat Genet *31*, 21-23.

Agarwal, A. K., and Garg, A. (2010). Enzymatic activity of the human 1-acylglycerol-3-phosphate-O-acyltransferase isoform 11: upregulated in breast and cervical cancers. J Lipid Res *51*, 2143-2152.

Aitman, T. J., Dong, R., Vyse, T. J., Norsworthy, P. J., Johnson, M. D., Smith, J., Mangion, J., Roberton-Lowe, C., Marshall, A. J., Petretto, E.*, et al.* (2006). Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature *439*, 851-855.

Alipour, A., van Oostrom, A. J., Izraeljan, A., Verseyden, C., Collins, J. M., Frayn, K. N., Plokker, T. W., Elte, J. W., and Castro Cabezas, M. (2008). Leukocyte activation by triglyceride-rich lipoproteins. Arterioscler Thromb Vasc Biol *28*, 792-797.

Almasy, L., and Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet *62*, 1198-1211.

Alpy, F., Stoeckel, M. E., Dierich, A., Escola, J. M., Wendling, C., Chenard, M. P., Vanier, M. T., Gruenberg, J., Tomasetto, C., and Rio, M. C. (2001). The steroidogenic acute regulatory protein homolog MLN64, a late endosomal cholesterol-binding protein. J Biol Chem *276*, 4261-4269.

Altmann, S. W., Davis, H. R., Jr., Zhu, L. J., Yao, X., Hoos, L. M., Tetzloff, G., Iyer, S. P., Maguire, M., Golovko, A., Zeng, M.*, et al.* (2004). Niemann-Pick C1 Like 1 protein is critical for intestinal cholesterol absorption. Science *303*, 1201-1204.

An, L., Yang, T., Zhang, Y., Lin, Y., Zhang, H., Jiao, X., Hua, L., Dai, H., and Wang, C. (2011). Association of SERPINE2 gene with the risk of chronic obstructive

pulmonary disease and spirometric phenotypes in northern Han Chinese population. Mol Biol Rep.

Anderson, C. A., Pettersson, F. H., Barrett, J. C., Zhuang, J. J., Ragoussis, J., Cardon, L. R., and Morris, A. P. (2008). Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms. Am J Hum Genet *83*, 112-119.

Angelakopoulou, A., Shah, T., Sofat, R., Shah, S., Berry, D. J., Cooper, J., Palmen, J., Tzoulaki, I., Wong, A., Jefferis, B. J.*, et al.* (2011). Comparative analysis of genome-wide-association studies signals for lipids, diabetes, and coronary heart disease: Cardiovascular Biomarker Genetics Collaboration. Eur Heart J.

Antoniou, A. C., and Easton, D. F. (2003). Polygenic inheritance of breast cancer: Implications for design of association studies. Genet Epidemiol *25*, 190-202.

Aouizerat, B. E., Allayee, H., Cantor, R. M., Davis, R. C., Lanning, C. D., Wen, P. Z., Dallinga-Thie, G. M., de Bruin, T. W., Rotter, J. I., and Lusis, A. J. (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.

Arnaiz, S. L., Travacio, M., Llesuy, S., and Boveris, A. (1995). Hydrogen peroxide metabolism during peroxisome proliferation by fenofibrate. Biochim Biophys Acta *1272*, 175-180.

Arner, P., Bernard, S., Salehpour, M., Possnert, G., Liebl, J., Steier, P., Buchholz, B. A., Eriksson, M., Arner, E., Hauner, H.*, et al.* (2011). Dynamics of human adipose lipid turnover in health and metabolic disease. Nature *478*, 110-113.

Arsenault, B. J., Rana, J. S., Stroes, E. S., Despres, J. P., Shah, P. K., Kastelein, J. J., Wareham, N. J., Boekholdt, S. M., and Khaw, K. T. (2009). Beyond low-density lipoprotein cholesterol: respective contributions of non-high-density lipoprotein cholesterol levels, triglycerides, and the total cholesterol/high-density lipoprotein cholesterol ratio to coronary heart disease risk in apparently healthy men and women. J Am Coll Cardiol *55*, 35-41.

Atwood, L. D., and Heard-Costa, N. L. (2003). Limits of fine-mapping a quantitative trait. Genet Epidemiol *24*, 99-106.

Aulchenko, Y. S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I. M., Pramstaller, P. P., Penninx, B. W., Janssens, A. C., Wilson, J. F., Spector, T.*, et al.* (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet *41*, 47-55.

Austin, M. A., McKnight, B., Edwards, K. L., Bradley, C. M., McNeely, M. J., Psaty, B. M., Brunzell, J. D., and Motulsky, A. G. (2000). Cardiovascular disease mortality in

familial forms of hypertriglyceridemia: A 20-year prospective study. Circulation *101*, 2777-2782.
Bainton, D., Miller, N. E., Bolton, C. H., Yarnell, J. W., Sweetnam, P. M., Baker, I. A., Lewis, B., and Elwood, P. C. (1992). Plasma triglyceride and high density lipoprotein cholesterol as predictors of ischaemic heart disease in British men. The Caerphilly and Speedwell Collaborative Heart Disease Studies. Br Heart J *68*, 60-66.
Balk, E. M., Lichtenstein, A. H., Chung, M., Kupelnick, B., Chew, P., and Lau, J. (2006). Effects of omega-3 fatty acids on serum markers of cardiovascular disease risk: a systematic review. Atherosclerosis *189*, 19-30.
Bamji-Mirza, M., Sundaram, M., Zhong, S., Yao, E. F., Parks, R. J., and Yao, Z. (2011). Secretion of triacylglycerol-poor VLDL particles from McA-RH7777 cells expressing human hepatic lipase. J Lipid Res *52*, 540-548.
Bang, H. O., Dyerberg, J., and Sinclair, H. M. (1980). The composition of the Eskimo food in north western Greenland. Am J Clin Nutr *33*, 2657-2661.
Bansal, S., Buring, J. E., Rifai, N., Mora, S., Sacks, F. M., and Ridker, P. M. (2007). Fasting compared with nonfasting triglycerides and risk of cardiovascular events in women. Jama *298*, 309-316.
Bansal, V., Libiger, O., Torkamani, A., and Schork, N. J. (2010). Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet *11*, 773-785.
Barlowe, C. (2002). COPII-dependent transport from the endoplasmic reticulum. Curr Opin Cell Biol *14*, 417-422.
Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21*, 263-265.
Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., Brant, S. R., Silverberg, M. S., Taylor, K. D., Barmada, M. M.*, et al.* (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet *40*, 955-962.
Bartoloni, L., and Antonarakis, S. E. (2004). The human sugar-phosphate/phosphate exchanger family SLC37. Pflugers Arch *447*, 780-783.
Bartoloni, L., Wattenhofer, M., Kudoh, J., Berry, A., Shibuya, K., Kawasaki, K., Wang, J., Asakawa, S., Talior, I., Bonne-Tamir, B.*, et al.* (2000). Cloning and characterization of a putative human glycerol 3-phosphate permease gene (SLC37A1 or G3PP) on 21q22.3: mutation analysis in two candidate phenotypes, DFNB10 and a glycerol kinase deficiency. Genomics *70*, 190-200.
Bartz, F., Kern, L., Erz, D., Zhu, M., Gilbert, D., Meinhof, T., Wirkner, U., Erfle, H., Muckenthaler, M., Pepperkok, R., and Runz, H. (2009). Identification of cholesterol-regulating genes by targeted RNAi screening. Cell Metab *10*, 63-75.

Beigneux, A. P. (2010). GPIHBP1 and the processing of triglyceride-rich lipoproteins. Clin Lipidol *5*, 575-582.
Beigneux, A. P., Davies, B. S., Gin, P., Weinstein, M. M., Farber, E., Qiao, X., Peale, F., Bunting, S., Walzem, R. L., Wong, J. S.*, et al.* (2007). Glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1 plays a critical role in the lipolytic processing of chylomicrons. Cell Metab *5*, 279-291.
Beigneux, A. P., Franssen, R., Bensadoun, A., Gin, P., Melford, K., Peter, J., Walzem, R. L., Weinstein, M. M., Davies, B. S., Kuivenhoven, J. A.*, et al.* (2009a). Chylomicronemia with a mutant GPIHBP1 (Q115P) that cannot bind lipoprotein lipase. Arterioscler Thromb Vasc Biol *29*, 956-962.
Beigneux, A. P., Gin, P., Davies, B. S., Weinstein, M. M., Bensadoun, A., Fong, L. G., and Young, S. G. (2009b). Highly conserved cysteines within the Ly6 domain of GPIHBP1 are crucial for the binding of lipoprotein lipase. J Biol Chem *284*, 30240-30247.
Beigneux, A. P., Weinstein, M. M., Davies, B. S., Gin, P., Bensadoun, A., Fong, L. G., and Young, S. G. (2009c). GPIHBP1 and lipolysis: an update. Curr Opin Lipidol *20*, 211-216.
Belfort, R., Berria, R., Cornell, J., and Cusi, K. (2010). Fenofibrate reduces systemic inflammation markers independent of its effects on lipid and glucose metabolism in patients with the metabolic syndrome. J Clin Endocrinol Metab *95*, 829-836.
Bennett, B. D., Babu-Khan, S., Loeloff, R., Louis, J. C., Curran, E., Citron, M., and Vassar, R. (2000). Expression analysis of BACE2 in brain and peripheral tissues. J Biol Chem *275*, 20647-20651.
Berge, K. E., Tian, H., Graf, G. A., Yu, L., Grishin, N. V., Schultz, J., Kwiterovich, P., Shan, B., Barnes, R., and Hobbs, H. H. (2000). Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. Science *290*, 1771-1775.
Berge, K. E., von Bergmann, K., Lutjohann, D., Guerra, R., Grundy, S. M., Hobbs, H. H., and Cohen, J. C. (2002). Heritability of plasma noncholesterol sterols and relationship to DNA sequence polymorphism in ABCG5 and ABCG8. J Lipid Res *43*, 486-494.
Bertolini, S., Cassanelli, S., Garuti, R., Ghisellini, M., Simone, M. L., Rolleri, M., Masturzo, P., and Calandra, S. (1999). Analysis of LDL receptor gene mutations in Italian patients with homozygous familial hypercholesterolemia. Arterioscler Thromb Vasc Biol *19*, 408-418.
Beyea, M. M., Heslop, C. L., Sawyez, C. G., Edwards, J. Y., Markle, J. G., Hegele, R. A., and Huff, M. W. (2007). Selective up-regulation of LXR-regulated genes ABCA1,

ABCG1, and APOE in macrophages through increased endogenous synthesis of 24(S),25-epoxycholesterol. J Biol Chem *282*, 5207-5216.
Bhattacharya, S., and Redgrave, T. G. (1981). The content of apolipoprotein B in chylomicron particles. J Lipid Res *22*, 820-828.
Bishop, J. R., Passos-Bueno, M. R., Fong, L., Stanford, K. I., Gonzales, J. C., Yeh, E., Young, S. G., Bensadoun, A., Witztum, J. L., Esko, J. D., and Moulton, K. S. (2010). Deletion of the basement membrane heparan sulfate proteoglycan type XVIII collagen causes hypertriglyceridemia in mice and humans. PLoS One *5*, e13919.
Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E., and Eppig, J. T. (2011). The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. Nucleic Acids Res *39*, D842-848.
Boren, J., Lee, I., Zhu, W., Arnold, K., Taylor, S., and Innerarity, T. L. (1998). Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective apo-B100. J Clin Invest *101*, 1084-1093.
Boright, A. P., Connelly, P. W., Brunt, J. H., Morgan, K., and Hegele, R. A. (1998). Association and linkage of LDLR gene variation with variation in plasma low density lipoprotein cholesterol. J Hum Genet *43*, 153-159.
Bosner, M. S., Lange, L. G., Stenson, W. F., and Ostlund, R. E., Jr. (1999). Percent cholesterol absorption in normal women and men quantified with dual stable isotopic tracers and negative ion mass spectrometry. J Lipid Res *40*, 302-308.
Boutaud, O., Dolis, D., and Schuber, F. (1992). Preferential cyclization of 2,3(S):22(S),23-dioxidosqualene by mammalian 2,3-oxidosqualene-lanosterol cyclase. Biochem Biophys Res Commun *188*, 898-904.
Braissant, O., Foufelle, F., Scotto, C., Dauca, M., and Wahli, W. (1996). Differential expression of peroxisome proliferator-activated receptors (PPARs): tissue distribution of PPAR-alpha, -beta, and -gamma in the adult rat. Endocrinology *137*, 354-366.
Brandt, J. M., Djouadi, F., and Kelly, D. P. (1998). Fatty acids activate transcription of the muscle carnitine palmitoyltransferase I gene in cardiac myocytes via the peroxisome proliferator-activated receptor alpha. J Biol Chem *273*, 23786-23792.
Brooks-Wilson, A., Marcil, M., Clee, S. M., Zhang, L. H., Roomp, K., van Dam, M., Yu, L., Brewer, C., Collins, J. A., Molhuizen, H. O*., et al.* (1999). Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. Nat Genet *22*, 336-345.
Brown, A. J., Sun, L., Feramisco, J. D., Brown, M. S., and Goldstein, J. L. (2002). Cholesterol addition to ER membranes alters conformation of SCAP, the SREBP escort protein that regulates cholesterol metabolism. Mol Cell *10*, 237-245.

Brown, P. J., Winegar, D. A., Plunket, K. D., Moore, L. B., Lewis, M. C., Wilson, J. G., Sundseth, S. S., Koble, C. S., Wu, Z., Chapman, J. M.*, et al.* (1999). A ureido-thioisobutyric acid (GW9578) is a subtype-selective PPARalpha agonist with potent lipid-lowering activity. J Med Chem *42*, 3785-3788.
Browning, B. L. (2008). PRESTO: rapid calculation of order statistic distributions and multiple-testing adjusted P-values via permutation for one and two-stage genetic association studies. BMC Bioinformatics *9*, 309.
Buch, S., Schafmayer, C., Volzke, H., Becker, C., Franke, A., von Eller-Eberstein, H., Kluck, C., Bassmann, I., Brosch, M., Lammert, F.*, et al.* (2007). A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. Nat Genet *39*, 995-999.
Buchmann, J., Meyer, C., Neschen, S., Augustin, R., Schmolz, K., Kluge, R., Al-Hasani, H., Jurgens, H., Eulenberg, K., Wehr, R.*, et al.* (2007). Ablation of the cholesterol transporter adenosine triphosphate-binding cassette transporter G1 reduces adipose cell size and protects against diet-induced obesity. Endocrinology *148*, 1561-1573.
Burkhardt, R., Kenny, E. E., Lowe, J. K., Birkeland, A., Josowitz, R., Noel, M., Salit, J., Maller, J. B., Pe'er, I., Daly, M. J.*, et al.* (2008). Common SNPs in HMGCR in micronesians and whites associated with LDL-cholesterol levels affect alternative splicing of exon13. Arterioscler Thromb Vasc Biol *28*, 2078-2084.
Cabezas, M. C., de Bruin, T. W., Jansen, H., Kock, L. A., Kortlandt, W., and Erkelens, D. W. (1993). Impaired chylomicron remnant clearance in familial combined hyperlipidemia. Arterioscler Thromb *13*, 804-814.
Calandra, S., Priore Oliva, C., Tarugi, P., and Bertolini, S. (2006). APOA5 and triglyceride metabolism, lesson from human APOA5 deficiency. Curr Opin Lipidol *17*, 122-127.
Calandra, S., Tarugi, P., Speedy, H. E., Dean, A. F., Bertolini, S., and Shoulders, C. C. (2011). Mechanisms and genetic determinants regulating sterol absorption, circulating LDL levels, and sterol elimination: implications for classification and disease risk. J Lipid Res *52*, 1885-1926.
Camps, L., Reina, M., Llobera, M., Vilaro, S., and Olivecrona, T. (1990). Lipoprotein lipase: cellular origin and functional distribution. Am J Physiol *258*, C673-681.
Cao, J., Wang, J., Qi, W., Miao, H. H., Wang, J., Ge, L., DeBose-Boyd, R. A., Tang, J. J., Li, B. L., and Song, B. L. (2007). Ufd1 is a cofactor of gp78 and plays a key role in cholesterol metabolism by regulating the stability of HMG-CoA reductase. Cell Metab *6*, 115-128.

Carlson, C. S., Eberle, M. A., Rieder, M. J., Yi, Q., Kruglyak, L., and Nickerson, D. A. (2004). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet *74*, 106-120.
Carstea, E. D., Morris, J. A., Coleman, K. G., Loftus, S. K., Zhang, D., Cummings, C., Gu, J., Rosenfeld, M. A., Pavan, W. J., Krizman, D. B.*, et al.* (1997). Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. Science *277*, 228-231.
Cefalu, A. B., Noto, D., Arpi, M. L., Yin, F., Spina, R., Hilden, H., Barbagallo, C. M., Carroccio, A., Tarugi, P., Squatrito, S.*, et al.* (2009). Novel LMF1 nonsense mutation in a patient with severe hypertriglyceridemia. J Clin Endocrinol Metab *94*, 4584-4590.
Chait, A., Albers, J. J., and Brunzell, J. D. (1980). Very low density lipoprotein overproduction in genetic forms of hypertriglyceridaemia. Eur J Clin Invest *10*, 17-22.
Charman, M., Kennedy, B. E., Osborne, N., and Karten, B. (2010). MLN64 mediates egress of cholesterol from endosomes to mitochondria in the absence of functional Niemann-Pick Type C1 protein. J Lipid Res *51*, 1023-1034.
Chasman, D. I., Pare, G., Mora, S., Hopewell, J. C., Peloso, G., Clarke, R., Cupples, L. A., Hamsten, A., Kathiresan, S., Malarstig, A.*, et al.* (2009). Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. PLoS Genet *5*, e1000730.
Chasse, J. F., Paul, V., Escanez, R., Kamoun, P., and London, J. (1997). Human cystathionine beta-synthase: gene organization and expression of different 5' alternative splicing. Mamm Genome *8*, 917-921.
Chatterton, J. E., Phillips, M. L., Curtiss, L. K., Milne, R., Fruchart, J. C., and Schumaker, V. N. (1995). Immunoelectron microscopy of low density lipoproteins yields a ribbon and bow model for the conformation of apolipoprotein B on the lipoprotein surface. J Lipid Res *36*, 2027-2037.
Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., and Fodor, S. P. (1996). Accessing genetic information with high-density DNA arrays. Science *274*, 610-614.
Chen, H., Rossier, C., Nakamura, Y., Lynn, A., Chakravarti, A., and Antonarakis, S. E. (1997). Cloning of a novel homeobox-containing gene, PKNOX1, and mapping to human chromosome 21q22.3. Genomics *41*, 193-200.
Chen, M. H., and Yang, Q. (2010). GWAF: an R package for genome-wide association analyses with family data. Bioinformatics *26*, 580-581.
Chen, W. J., Goldstein, J. L., and Brown, M. S. (1990). NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. J Biol Chem *265*, 3116-3123.

Chen, W. M., Manichaikul, A., and Rich, S. S. (2009). A generalized family-based association test for dichotomous traits. Am J Hum Genet *85*, 364-376.
Cheng, C. F., Oosta, G. M., Bensadoun, A., and Rosenberg, R. D. (1981). Binding of lipoprotein lipase to endothelial cells in culture. J Biol Chem *256*, 12893-12898.
Cheverud, J. M. (2001). A simple correction for multiple comparisons in interval mapping genome scans. Heredity *87*, 52-58.
Chinetti-Gbaguidi, G., Rigamonti, E., Helin, L., Mutka, A. L., Lepore, M., Fruchart, J. C., Clavey, V., Ikonen, E., Lestavel, S., and Staels, B. (2005). Peroxisome proliferator-activated receptor alpha controls cellular cholesterol trafficking in macrophages. J Lipid Res *46*, 2717-2725.
Chrast, R., Saher, G., Nave, K. A., and Verheijen, M. H. (2011). Lipid metabolism in myelinating glial cells: lessons from human inherited disorders and mouse models. J Lipid Res *52*, 419-434.
Christodoulides, C., Lagathu, C., Sethi, J. K., and Vidal-Puig, A. (2009). Adipogenesis and WNT signalling. Trends Endocrinol Metab *20*, 16-24.
Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. Genetics *138*, 963-971.
Cladaras, C., Hadzopoulou-Cladaras, M., Nolte, R. T., Atkinson, D., and Zannis, V. I. (1986). The complete sequence and structural analysis of human apolipoprotein B-100: relationship between apoB-100 and apoB-48 forms. Embo J *5*, 3495-3507.
Coca-Prieto, I., Kroupa, O., Gonzalez-Santos, P., Magne, J., Olivecrona, G., Ehrenborg, E., and Valdivielso, P. (2010). Childhood-onset chylomicronaemia with reduced plasma lipoprotein lipase activity and mass: identification of a novel GPIHBP1 mutation. J Intern Med *270*, 224-228.
Cohen, J., Pertsemlidis, A., Kotowski, I. K., Graham, R., Garcia, C. K., and Hobbs, H. H. (2005). Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet *37*, 161-165.
Cohen, J. C., Pertsemlidis, A., Fahmi, S., Esmail, S., Vega, G. L., Grundy, S. M., and Hobbs, H. H. (2006). Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc Natl Acad Sci U S A *103*, 1810-1815.
Coleman, R. A., and Lee, D. P. (2004). Enzymes of triacylglycerol synthesis and their regulation. Prog Lipid Res *43*, 134-176.
Conklin, D., Gilbertson, D., Taft, D. W., Maurer, M. F., Whitmore, T. E., Smith, D. L., Walker, K. M., Chen, L. H., Wattler, S., Nehls, M., and Lewis, K. B. (1999). Identification of a mammalian angiopoietin-related protein expressed specifically in liver. Genomics *62*, 477-482.

Coon, H., Eckfeldt, J. H., Leppert, M. F., Myers, R. H., Arnett, D. K., Heiss, G., Province, M. A., and Hunt, S. C. (2002). A genome-wide screen reveals evidence for a locus on chromosome 11 influencing variation in LDL cholesterol in the NHLBI Family Heart Study. Hum Genet *111*, 263-269.
Cunningham, D., Danley, D. E., Geoghegan, K. F., Griffor, M. C., Hawkins, J. L., Subashi, T. A., Varghese, A. H., Ammirati, M. J., Culp, J. S., Hoth, L. R.*, et al.* (2007). Structural and biophysical studies of PCSK9 and its mutants linked to familial hypercholesterolemia. Nat Struct Mol Biol *14*, 413-419.
Cupples, L. A. (2008). Family study designs in the age of genome-wide association studies: experience from the Framingham Heart Study. Curr Opin Lipidol *19*, 144-150.
Cupples, L. A., Arruda, H. T., Benjamin, E. J., D'Agostino, R. B., Sr., Demissie, S., DeStefano, A. L., Dupuis, J., Falls, K. M., Fox, C. S., Gottlieb, D. J.*, et al.* (2007). The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports. BMC Med Genet *8 Suppl 1*, S1.
Cuthbert, A. P., Fisher, S. A., Mirza, M. M., King, K., Hampe, J., Croucher, P. J., Mascheretti, S., Sanderson, J., Forbes, A., Mansfield, J.*, et al.* (2002). The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. Gastroenterology *122*, 867-874.
Dallinga-Thie, G. M., van Linde-Sibenius Trip, M., Rotter, J. I., Cantor, R. M., Bu, X., Lusis, A. J., and de Bruin, T. W. (1997). Complex genetic contribution of the Apo AI-CIII-AIV gene cluster to familial combined hyperlipidemia. Identification of different susceptibility haplotypes. J Clin Invest *99*, 953-961.
Dastani, Z., Quiogue, L., Plaisier, C., Engert, J. C., Marcil, M., Genest, J., and Pajukanta, P. (2006). Evidence for a gene influencing high-density lipoprotein cholesterol on chromosome 4q31.21. Arterioscler Thromb Vasc Biol *26*, 392-397.
Davies, B. S., Beigneux, A. P., Barnes, R. H., 2nd, Tu, Y., Gin, P., Weinstein, M. M., Nobumori, C., Nyren, R., Goldberg, I., Olivecrona, G.*, et al.* (2010). GPIHBP1 is responsible for the entry of lipoprotein lipase into capillaries. Cell Metab *12*, 42-52.
Davis, C. E., Gordon, D., LaRosa, J., Wood, P. D., and Halperin, M. (1980). Correlations of plasma high-density lipoprotein cholesterol levels with other plasma lipid and lipoprotein concentrations. Circulation *62*, IV24-30.
Davis, H. R., Jr., Zhu, L. J., Hoos, L. M., Tetzloff, G., Maguire, M., Liu, J., Yao, X., Iyer, S. P., Lam, M. H., Lund, E. G.*, et al.* (2004). Niemann-Pick C1 Like 1 (NPC1L1) is the intestinal phytosterol and cholesterol transporter and a key modulator of whole-body cholesterol homeostasis. J Biol Chem *279*, 33586-33592.

Davis, R. C., Ben-Zeev, O., Martin, D., and Doolittle, M. H. (1990). Combined lipase deficiency in the mouse. Evidence of impaired lipase processing and secretion. J Biol Chem *265*, 17960-17966.
de Bakker, P. I., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., and Altshuler, D. (2005). Efficiency and power in genetic association studies. Nat Genet *37*, 1217-1223.
De Caterina, R., Talmud, P. J., Merlini, P. A., Foco, L., Pastorino, R., Altshuler, D., Mauri, F., Peyvandi, F., Lina, D., Kathiresan, S.*, et al.* (2011). Strong association of the APOA5-1131T>C gene variant and early-onset acute myocardial infarction. Atherosclerosis *214*, 397-403.
Deckelbaum, R. J., Granot, E., Oschry, Y., Rose, L., and Eisenberg, S. (1984). Plasma triglyceride determines structure-composition in low and high density lipoproteins. Arteriosclerosis *4*, 225-231.
Demeo, D. L., Mariani, T. J., Lange, C., Srisuma, S., Litonjua, A. A., Celedon, J. C., Lake, S. L., Reilly, J. J., Chapman, H. A., Mecham, B. H.*, et al.* (2006). The SERPINE2 gene is associated with chronic obstructive pulmonary disease. Am J Hum Genet *78*, 253-264.
Derosa, G., Maffioli, P., D'Angelo, A., Salvadeo, S. A., Ferrari, I., Fogari, E., Gravina, A., Mereu, R., Palumbo, I., Randazzo, S., and Cicero, A. F. (2009). Evaluation of metalloproteinase 2 and 9 levels and their inhibitors in combined dyslipidemia. Clin Invest Med *32*, E124-132.
Dolk, H., Loane, M., and Garne, E. (2011). Congenital heart defects in Europe: prevalence and perinatal mortality, 2000 to 2005. Circulation *123*, 841-849.
Dosil, M., and Bustelo, X. R. (2004). Functional characterization of Pwp2, a WD family protein essential for the assembly of the 90 S pre-ribosomal particle. J Biol Chem *279*, 37385-37397.
Draheim, C. C., Geijer, J. R., and Dengel, D. R. (2010). Comparison of intima-media thickness of the carotid artery and cardiovascular disease risk factors in adults with versus without the Down syndrome. Am J Cardiol *106*, 1512-1516.
Economou, M., Trikalinos, T. A., Loizou, K. T., Tsianos, E. V., and Ioannidis, J. P. (2004). Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. Am J Gastroenterol *99*, 2393-2404.
Eeles, R. A., Kote-Jarai, Z., Giles, G. G., Olama, A. A., Guy, M., Jugurnauth, S. K., Mulholland, S., Leongamornlert, D. A., Edwards, S. M., Morrison, J.*, et al.* (2008). Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet *40*, 316-321.

Ehret, G. B., Munroe, P. B., Rice, K. M., Bochud, M., Johnson, A. D., Chasman, D. I., Smith, A. V., Tobin, M. D., Verwoert, G. C., Hwang, S. J.*, et al.* (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature *478*, 103-109.

Eichenbaum-Voline, S., Olivier, M., Jones, E. L., Naoumova, R. P., Jones, B., Gau, B., Patel, H. N., Seed, M., Betteridge, D. J., Galton, D. J.*, et al.* (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.

Emmerich, J., Beg, O. U., Peterson, J., Previato, L., Brunzell, J. D., Brewer, H. B., Jr., and Santamarina-Fojo, S. (1992). Human lipoprotein lipase. Analysis of the catalytic triad by site-directed mutagenesis of Ser-132, Asp-156, and His-241. J Biol Chem *267*, 4161-4165.

Errera, F. I., Canani, L. H., Yeh, E., Kague, E., Armelin-Correa, L. M., Suzuki, O. T., Tschiedel, B., Silva, M. E., Sertie, A. L., and Passos-Bueno, M. R. (2008). COL18A1 is highly expressed during human adipocyte differentiation and the SNP c.1136C > T in its "frizzled" motif is associated with obesity in diabetes type 2 patients. An Acad Bras Cienc *80*, 167-177.

Eurlings, P. M., Van Der Kallen, C. J., Geurts, J. M., Kouwenberg, P., Boeckx, W. D., and De Bruin, T. W. (2002). Identification of differentially expressed genes in subcutaneous adipose tissue from subjects with familial combined hyperlipidemia. J Lipid Res *43*, 930-935.

Evangelou, E., Trikalinos, T. A., Salanti, G., and Ioannidis, J. P. (2006). Family-based versus unrelated case-control designs for genetic associations. PLoS Genet *2*, e123.

Evans, D. M., and Cardon, L. R. (2004). Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps. Am J Hum Genet *75*, 687-692.

Ewens, W. J., and Spielman, R. S. (1995). The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet *57*, 455-464.

Ewens, W. J., and Spielman, R. S. (2005). What is the significance of a significant TDT? Hum Hered *60*, 206-210.

Eyster, K. M. (2007). The membrane and lipids as integral participants in signal transduction: lipid signal transduction for the non-lipid biochemist. Adv Physiol Educ *31*, 5-16.

Fahmi, S., Yang, C., Esmail, S., Hobbs, H. H., and Cohen, J. C. (2008). Functional characterization of genetic variants in NPC1L1 supports the sequencing extremes strategy to identify complex trait genes. Hum Mol Genet *17*, 2101-2107.

Fasano, T., Cefalu, A. B., Di Leo, E., Noto, D., Pollaccia, D., Bocchi, L., Valenti, V., Bonardi, R., Guardamagna, O., Averna, M., and Tarugi, P. (2007). A novel loss of function mutation of PCSK9 gene in white subjects with low-plasma low-density lipoprotein cholesterol. Arterioscler Thromb Vasc Biol *27*, 677-681.
Fernandez, M. L., and Volek, J. S. (2006). Guinea pigs: a suitable animal model to study lipoprotein metabolism, atherosclerosis and inflammation. Nutr Metab (Lond) *3*, 17.
Figarella, C., De Caro, A., Leupold, D., and Poley, J. R. (1980). Congenital pancreatic lipase deficiency. J Pediatr *96*, 412-416.
Fisher, R. M., Humphries, S. E., and Talmud, P. J. (1997). Common variation in the lipoprotein lipase gene: effects on plasma lipids and risk of atherosclerosis. Atherosclerosis *135*, 145-159.
FitzPatrick, D. R., Keeling, J. W., Evans, M. J., Kan, A. E., Bell, J. E., Porteous, M. E., Mills, K., Winter, R. M., and Clayton, P. T. (1998). Clinical phenotype of desmosterolosis. Am J Med Genet *75*, 145-152.
Fojo, S. S., and Brewer, H. B. (1992). Hypertriglyceridaemia due to genetic defects in lipoprotein lipase and apolipoprotein C-II. J Intern Med *231*, 669-677.
Franssen, R., Young, S. G., Peelman, F., Hertecant, J., Sierts, J. A., Schimmel, A. W., Bensadoun, A., Kastelein, J. J., Fong, L. G., Dallinga-Thie, G. M., and Beigneux, A. P. (2010). Chylomicronemia with low postheparin lipoprotein lipase levels in the setting of GPIHBP1 defects. Circ Cardiovasc Genet *3*, 169-178.
Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M.*, et al.* (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851-861.
Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. Nucleic Acids Res *32*, W273-279.
Fredrickson, D. S., and Lees, R. S. (1965). A System for Phenotyping Hyperlipoproteinemia. Circulation *31*, 321-327.
Freeman, S. B., Taft, L. F., Dooley, K. J., Allran, K., Sherman, S. L., Hassold, T. J., Khoury, M. J., and Saker, D. M. (1998). Population-based study of congenital heart defects in Down syndrome. Am J Med Genet *80*, 213-217.
Gao, X., Starmer, J., and Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol *32*, 361-369.
Garuti, R., Jones, C., Li, W. P., Michaely, P., Herz, J., Gerard, R. D., Cohen, J. C., and Hobbs, H. H. (2005). The modular adaptor protein autosomal recessive

hypercholesterolemia (ARH) promotes low density lipoprotein receptor clustering into clathrin-coated pits. J Biol Chem *280*, 40996-41004.
Gbaguidi, G. F., and Agellon, L. B. (2004). The inhibition of the human cholesterol 7alpha-hydroxylase gene (CYP7A1) promoter by fibrates in cultured cells is mediated via the liver x receptor alpha and peroxisome proliferator-activated receptor alpha heterodimer. Nucleic Acids Res *32*, 1113-1121.
Gin, P., Beigneux, A. P., Davies, B., Young, M. F., Ryan, R. O., Bensadoun, A., Fong, L. G., and Young, S. G. (2007). Normal binding of lipoprotein lipase, chylomicrons, and apo-AV to GPIHBP1 containing a G56R amino acid substitution. Biochim Biophys Acta *1771*, 1464-1468.
Gin, P., Beigneux, A. P., Voss, C., Davies, B. S., Beckstead, J. A., Ryan, R. O., Bensadoun, A., Fong, L. G., and Young, S. G. (2011). Binding preferences for GPIHBP1, a glycosylphosphatidylinositol-anchored protein of capillary endothelial cells. Arterioscler Thromb Vasc Biol *31*, 176-182.
Gin, P., Yin, L., Davies, B. S., Weinstein, M. M., Ryan, R. O., Bensadoun, A., Fong, L. G., Young, S. G., and Beigneux, A. P. (2008). The acidic domain of GPIHBP1 is important for the binding of lipoprotein lipase and chylomicrons. J Biol Chem *283*, 29554-29562.
Goldstein, J. L., Schrott, H. G., Hazzard, W. R., Bierman, E. L., and Motulsky, A. G. (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.
Gower, R. M., Wu, H., Foster, G. A., Devaraj, S., Jialal, I., Ballantyne, C. M., Knowlton, A. A., and Simon, S. I. (2011). CD11c/CD18 expression is upregulated on blood monocytes during hypertriglyceridemia and enhances adhesion to vascular cell adhesion molecule-1. Arterioscler Thromb Vasc Biol *31*, 160-166.
Graf, G. A., Li, W. P., Gerard, R. D., Gelissen, I., White, A., Cohen, J. C., and Hobbs, H. H. (2002). Coexpression of ATP-binding cassette proteins ABCG5 and ABCG8 permits their transport to the apical surface. J Clin Invest *110*, 659-669.
Graf, G. A., Yu, L., Li, W. P., Gerard, R., Tuma, P. L., Cohen, J. C., and Hobbs, H. H. (2003). ABCG5 and ABCG8 are obligate heterodimers for protein trafficking and biliary cholesterol excretion. J Biol Chem *278*, 48275-48282.
Greenawalt, D. M., Dobrin, R., Chudin, E., Hatoum, I. J., Suver, C., Beaulaurier, J., Zhang, B., Castro, V., Zhu, J., Sieberts, S. K.*, et al.* (2011). A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. Genome Res *21*, 1008-1016.

Gremlich, S., Nolan, C., Roduit, R., Burcelin, R., Peyot, M. L., Delghingaro-Augusto, V., Desvergne, B., Michalik, L., Prentki, M., and Wahli, W. (2005). Pancreatic islet adaptation to fasting is dependent on peroxisome proliferator-activated receptor alpha transcriptional up-regulation of fatty acid oxidation. Endocrinology *146*, 375-382.

Groenewegen, W. A., Krul, E. S., Averna, M. R., Pulai, J., and Schonfeld, G. (1994). Dysbetalipoproteinemia in a kindred with hypobetalipoproteinemia due to mutations in the genes for ApoB (ApoB-70.5) and ApoE (ApoE2). Arterioscler Thromb *14*, 1695-1704.

Guay, S. P., Gaudet, D., and Brisson, D. (2011). The g.-469G>A polymorphism in the GPIHBP1 gene promoter is associated with hypertriglyceridemia and has an additive effect on the risk conferred by LPL defective alleles. Nutr Metab Cardiovasc Dis.

Gudbjartsson, D. F., Jonasson, K., Frigge, M. L., and Kong, A. (2000). Allegro, a new computer program for multipoint linkage analysis. Nat Genet *25*, 12-13.

Gylling, H., Hallikainen, M., Pihlajamaki, J., Agren, J., Laakso, M., Rajaratnam, R. A., Rauramaa, R., and Miettinen, T. A. (2004). Polymorphisms in the ABCG5 and ABCG8 genes associate with cholesterol absorption and insulin sensitivity. J Lipid Res *45*, 1660-1665.

Gylling, H., and Miettinen, T. A. (2002). Inheritance of cholesterol metabolism of probands with high or low cholesterol absorption. J Lipid Res *43*, 1472-1476.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., and King, M. C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. Science *250*, 1684-1689.

Hampe, J., Cuthbert, A., Croucher, P. J., Mirza, M. M., Mascheretti, S., Fisher, S., Frenzel, H., King, K., Hasselmeyer, A., MacPherson, A. J.*, et al.* (2001). Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet *357*, 1925-1928.

Hauser, E. R., Watanabe, R. M., Duren, W. L., Bass, M. P., Langefeld, C. D., and Boehnke, M. (2004). Ordered subset analysis in genetic linkage mapping of complex traits. Genet Epidemiol *27*, 53-63.

Hayashi, K., Hayashi, M., Jalkanen, M., Firestone, J. H., Trelstad, R. L., and Bernfield, M. (1987). Immunocytochemistry of cell surface heparan sulfate proteoglycan in mouse tissues. A light and electron microscopic study. J Histochem Cytochem *35*, 1079-1088.

He, G., Gupta, S., Yi, M., Michaely, P., Hobbs, H. H., and Cohen, J. C. (2002). ARH is a modular adaptor protein that interacts with the LDL receptor, clathrin, and AP-2. J Biol Chem *277*, 44044-44049.

Hegele, R. A. (2005). Lessons from human mutations in PPARgamma. Int J Obes (Lond) *29 Suppl 1*, S31-35.
Heiss, G., Tamir, I., Davis, C. E., Tyroler, H. A., Rifkand, B. M., Schonfeld, G., Jacobs, D., and Frantz, I. D., Jr. (1980). Lipoprotein-cholesterol distributions in selected North American populations: the lipid research clinics program prevalence study. Circulation *61*, 302-315.
Heller, D. A., de Faire, U., Pedersen, N. L., Dahlen, G., and McClearn, G. E. (1993). Genetic and environmental influences on serum lipid levels in twins. N Engl J Med *328*, 1150-1156.
Hemani, G., Theocharidis, A., Wei, W., and Haley, C. (2011). EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. Bioinformatics *27*, 1462-1465.
Herscovitz, H., Hadzopoulou-Cladaras, M., Walsh, M. T., Cladaras, C., Zannis, V. I., and Small, D. M. (1991). Expression, secretion, and lipid-binding characterization of the N-terminal 17% of apolipoprotein B. Proc Natl Acad Sci U S A *88*, 7313-7317.
Higashi, Y., Itabe, H., Fukase, H., Mori, M., Fujimoto, Y., Sato, R., Imanaka, T., and Takano, T. (2002). Distribution of microsomal triglyceride transfer protein within sub-endoplasmic reticulum regions in human hepatoma cells. Biochim Biophys Acta *1581*, 127-136.
Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. Theoretical and Applied Genetics *38*, 226-231.
Hirschhorn, J. N., and Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. Nat Rev Genet *6*, 95-108.
Ho, K. J., and Taylor, C. B. (1970). Control mechanisms of cholesterol biosynthesis. Arch Pathol *90*, 83-92.
Hobbs, H. H., Russell, D. W., Brown, M. S., and Goldstein, J. L. (1990). The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. Annu Rev Genet *24*, 133-170.
Hoggart, C. J., Clark, T. G., De Iorio, M., Whittaker, J. C., and Balding, D. J. (2008). Genome-wide significance for dense SNP and resequencing data. Genet Epidemiol *32*, 179-185.
Holt, R., Barnby, G., Maestrini, E., Bacchelli, E., Brocklebank, D., Sousa, I., Mulder, E. J., Kantojarvi, K., Jarvela, I., Klauck, S. M.*, et al.* (2010). Linkage and candidate gene studies of autism spectrum disorders in European populations. Eur J Hum Genet *18*, 1013-1019.
Hopkins, P. N., Heiss, G., Ellison, R. C., Province, M. A., Pankow, J. S., Eckfeldt, J. H., and Hunt, S. C. (2003). Coronary artery disease risk in familial combined

hyperlipidemia and familial hypertriglyceridemia: a case-control comparison from the National Heart, Lung, and Blood Institute Family Heart Study. Circulation *108*, 519-523.
Hopkins, P. N., Wu, L. L., Hunt, S. C., and Brinton, E. A. (2005). Plasma triglycerides and type III hyperlipidemia are independently associated with premature familial coronary artery disease. J Am Coll Cardiol *45*, 1003-1012.
Horswell, S. D., Ringham, H. E., and Shoulders, C. C. (2009). New technologies for delineating and characterizing the lipid exome: prospects for understanding familial combined hyperlipidemia. J Lipid Res *50 Suppl*, S370-375.
Howie, B. N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet *5*, e1000529.
Hu, F. B., Rimm, E. B., Stampfer, M. J., Ascherio, A., Spiegelman, D., and Willett, W. C. (2000). Prospective study of major dietary patterns and risk of coronary heart disease in men. Am J Clin Nutr *72*, 912-921.
Hua, X., Nohturfft, A., Goldstein, J. L., and Brown, M. S. (1996). Sterol resistance in CHO cells traced to point mutation in SREBP cleavage-activating protein. Cell *87*, 415-426.
Huertas-Vazquez, A., Aguilar-Salinas, C., Lusis, A. J., Cantor, R. M., Canizales-Quinteros, S., Lee, J. C., Mariana-Nunez, L., Riba-Ramirez, R. M., Jokiaho, A., Tusie-Luna, T., and 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.
Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M.*, et al.* (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature *411*, 599-603.
Hugot, J. P., Laurent-Puig, P., Gower-Rousseau, C., Olson, J. M., Lee, J. C., Beaugerie, L., Naom, I., Dupas, J. L., Van Gossum, A., Orholm, M.*, et al.* (1996). Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature *379*, 821-823.
Hugot, J. P., Zaccaria, I., Cavanaugh, J., Yang, H., Vermeire, S., Lappalainen, M., Schreiber, S., Annese, V., Jewell, D. P., Fowler, E. V.*, et al.* (2007). Prevalence of CARD15/NOD2 mutations in Caucasian healthy people. Am J Gastroenterol *102*, 1259-1267.
Humphries, S. E., Nicaud, V., Margalef, J., Tiret, L., and Talmud, P. J. (1998). Lipoprotein lipase gene variation is associated with a paternal history of premature

coronary artery disease and fasting and postprandial plasma triglycerides: the European Atherosclerosis Research Study (EARS). Arterioscler Thromb Vasc Biol *18*, 526-534.
Iacopetta, D., Lappano, R., Cappello, A. R., Madeo, M., De Francesco, E. M., Santoro, A., Curcio, R., Capobianco, L., Pezzi, V., Maggiolini, M., and Dolce, V. (2010). SLC37A1 gene expression is up-regulated by epidermal growth factor in breast cancer cells. Breast Cancer Res Treat *122*, 755-764.
Inoue-Murayama, M., Sugimoto, Y., Niimi, Y., and Aso, H. (2000). Type XVIII collagen is newly transcribed during bovine adipogenesis. Differentiation *65*, 281-285.
International HapMap Consortium (2003). The International HapMap Project. Nature *426*, 789-796.
International HapMap Consortium (2005). A haplotype map of the human genome. Nature *437*, 1299-1320.
Ioka, R. X., Kang, M. J., Kamiyama, S., Kim, D. H., Magoori, K., Kamataki, A., Ito, Y., Takei, Y. A., Sasaki, M., Suzuki, T.*, et al.* (2003). Expression cloning and characterization of a novel glycosylphosphatidylinositol-anchored high density lipoprotein-binding protein, GPI-HBP1. J Biol Chem *278*, 7344-7349.
Issemann, I., and Green, S. (1990). Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. Nature *347*, 645-650.
Jakobsen, M. U., O'Reilly, E. J., Heitmann, B. L., Pereira, M. A., Balter, K., Fraser, G. E., Goldbourt, U., Hallmans, G., Knekt, P., Liu, S.*, et al.* (2009). Major types of dietary fat and risk of coronary heart disease: a pooled analysis of 11 cohort studies. Am J Clin Nutr *89*, 1425-1432.
Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. R., VanLiere, J. M., Fung, H. C., Szpiech, Z. A., Degnan, J. H., Wang, K., Guerreiro, R.*, et al.* (2008). Genotype, haplotype and copy-number variation in worldwide human populations. Nature *451*, 998-1003.
Jensen, M. K., Rimm, E. B., Rader, D., Schmidt, E. B., Sorensen, T. I., Vogel, U., Overvad, K., and Mukamal, K. J. (2009). S447X variant of the lipoprotein lipase gene, lipids, and risk of coronary heart disease in 3 prospective cohort studies. Am Heart J *157*, 384-390.
Ji, Z. S., Brecht, W. J., Miranda, R. D., Hussain, M. M., Innerarity, T. L., and Mahley, R. W. (1993). Role of heparan sulfate proteoglycans in the binding and uptake of apolipoprotein E-enriched remnant lipoproteins by cultured cells. J Biol Chem *268*, 10160-10167.

Jiang, L. H., Yang, W., Zou, J., and Beech, D. J. (2010a). TRPM2 channel properties, functions and therapeutic potentials. Expert Opin Ther Targets *14*, 973-988.
Jiang, X., Barmada, M. M., Cooper, G. F., and Becich, M. J. (2011). A bayesian method for evaluating and discovering disease loci associations. PLoS One *6*, e22075.
Jiang, X., Barmada, M. M., and Visweswaran, S. (2010b). Identifying genetic interactions in genome-wide data using Bayesian networks. Genet Epidemiol *34*, 575-581.
Johansen, C. T., and Hegele, R. A. (2011). Genetic bases of hypertriglyceridemic phenotypes. Curr Opin Lipidol *22*, 247-253.
Johansen, C. T., Kathiresan, S., and Hegele, R. A. (2011a). Genetic determinants of plasma triglycerides. J Lipid Res *52*, 189-206.
Johansen, C. T., Wang, J., Lanktree, M. B., Cao, H., McIntyre, A. D., Ban, M. R., Martins, R. A., Kennedy, B. A., Hassell, R. G., Visser, M. E.*, et al.* (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet *42*, 684-687.
Johansen, C. T., Wang, J., Lanktree, M. B., McIntyre, A. D., Ban, M. R., Martins, R. A., Kennedy, B. A., Hassell, R. G., Visser, M. E., Schwartz, S. M.*, et al.* (2011b). An increased burden of common and rare lipid-associated risk alleles contributes to the phenotypic spectrum of hypertriglyceridemia. Arterioscler Thromb Vasc Biol *31*, 1916-1926.
Johansen, C. T., Wang, J., McIntyre, A. D., Martins, R. A., Ban, M. R., Lanktree, M. B., Huff, M. W., Peterfy, M., Mehrabian, M., Lusis, A. J.*, et al.* (2011c). Excess of Rare Variants in Non-GWAS Candidate Genes in Patients with Hypertriglyceridemia. Circ Cardiovasc Genet.
Johnson, A. D., Handsaker, R. E., Pulit, S. L., Nizzari, M. M., O'Donnell, C. J., and de Bakker, P. I. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics *24*, 2938-2939.
Johnson, R. C., Nelson, G. W., Troyer, J. L., Lautenberger, J. A., Kessing, B. D., Winkler, C. A., and O'Brien, S. J. (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics *11*, 724.
Jones, B., Jones, E. L., Bonney, S. A., Patel, H. N., Mensenkamp, A. R., Eichenbaum-Voline, S., Rudling, M., Myrdal, U., Annesi, G., Naik, S.*, et al.* (2003). Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders. Nat Genet *34*, 29-31.
Jurevics, H., and Morell, P. (1995). Cholesterol for synthesis of myelin is made locally, not imported into brain. J Neurochem *64*, 895-901.

Jurinke, C., van den Boom, D., Cantor, C. R., and Koster, H. (2002). Automated genotyping using the DNA MassArray technology. Methods Mol Biol *187*, 179-192.
Kannel, W. B., Garcia, M. J., McNamara, P. M., and Pearson, G. (1971). Serum lipid precursors of coronary heart disease. Hum Pathol *2*, 129-151.
Karpe, F., Steiner, G., Uffelman, K., Olivecrona, T., and Hamsten, A. (1994). Postprandial lipoproteins and progression of coronary atherosclerosis. Atherosclerosis *106*, 83-97.
Kathiresan, S., Manning, A. K., Demissie, S., D'Agostino, R. B., Surti, A., Guiducci, C., Gianniny, L., Burtt, N. P., Melander, O., Orho-Melander, M.*, et al.* (2007). A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. BMC Med Genet *8 Suppl 1*, S17.
Kathiresan, S., Melander, O., Anevski, D., Guiducci, C., Burtt, N. P., Roos, C., Hirschhorn, J. N., Berglund, G., Hedblad, B., Groop, L.*, et al.* (2008). Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med *358*, 1240-1249.
Kathiresan, S., Voight, B. F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P. M., Anand, S., Engert, J. C., Samani, N. J., Schunkert, H.*, et al.* (2009a). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet *41*, 334-341.
Kathiresan, S., Willer, C. J., Peloso, G. M., Demissie, S., Musunuru, K., Schadt, E. E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T.*, et al.* (2009b). Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet *41*, 56-65.
Katsanis, N., and Fisher, E. M. (1998). Identification, expression, and chromosomal localization of ubiquitin conjugating enzyme 7 (UBE2G2), a human homologue of the Saccharomyces cerevisiae ubc7 gene. Genomics *51*, 128-131.
Keebler, M. E., Sanders, C. L., Surti, A., Guiducci, C., Burtt, N. P., and Kathiresan, S. (2009). Association of blood lipids with common DNA sequence variants at 19 genetic loci in the multiethnic United States National Health and Nutrition Examination Survey III. Circ Cardiovasc Genet *2*, 238-243.
Kennedy, G. C., Matsuzaki, H., Dong, S., Liu, W. M., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J.*, et al.* (2003). Large-scale genotyping of complex DNA. Nat Biotechnol *21*, 1233-1237.
Kennedy, M. A., Barrera, G. C., Nakamura, K., Baldan, A., Tarr, P., Fishbein, M. C., Frank, J., Francone, O. L., and Edwards, P. A. (2005). ABCG1 has a critical role in mediating cholesterol efflux to HDL and preventing cellular lipid accumulation. Cell Metab *1*, 121-131.

Kersten, S., Mandard, S., Tan, N. S., Escher, P., Metzger, D., Chambon, P., Gonzalez, F. J., Desvergne, B., and Wahli, W. (2000). Characterization of the fasting-induced adipose factor FIAF, a novel peroxisome proliferator-activated receptor target gene. J Biol Chem *275*, 28488-28493.
Kim, E., Napierala, M., and Dent, S. Y. (2011). Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia. Nucleic Acids Res *39*, 8366-8377.
Klenova, E., Scott, A. C., Roberts, J., Shamsuddin, S., Lovejoy, E. A., Bergmann, S., Bubb, V. J., Royer, H. D., and Quinn, J. P. (2004). YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders. J Neurosci *24*, 5966-5973.
Klett, E. L., Lee, M. H., Adams, D. B., Chavin, K. D., and Patel, S. B. (2004a). Localization of ABCG5 and ABCG8 proteins in human liver, gall bladder and intestine. BMC Gastroenterol *4*, 21.
Klett, E. L., Lu, K., Kosters, A., Vink, E., Lee, M. H., Altenburg, M., Shefer, S., Batta, A. K., Yu, H., Chen, J.*, et al.* (2004b). A mouse model of sitosterolemia: absence of Abcg8/sterolin-2 results in failure to secrete biliary cholesterol. BMC Med *2*, 5.
Klucken, J., Buchler, C., Orso, E., Kaminski, W. E., Porsch-Ozcurumez, M., Liebisch, G., Kapinsky, M., Diederich, W., Drobnik, W., Dean, M.*, et al.* (2000). ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a regulator of macrophage cholesterol and phospholipid transport. Proc Natl Acad Sci U S A *97*, 817-822.
Knott, T. J., Rall, S. C., Jr., Innerarity, T. L., Jacobson, S. F., Urdea, M. S., Levy-Wilson, B., Powell, L. M., Pease, R. J., Eddy, R., Nakai, H., and et al. (1985). Human apolipoprotein B: structure of carboxyl-terminal domains, sites of gene expression, and chromosomal localization. Science *230*, 37-43.
Koishi, R., Ando, Y., Ono, M., Shimamura, M., Yasumo, H., Fujiwara, T., Horikoshi, H., and Furukawa, H. (2002). Angptl3 regulates lipid metabolism in mice. Nat Genet *30*, 151-157.
Kong, A., and Cox, N. J. (1997). Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet *61*, 1179-1188.
Koster, A., Chao, Y. B., Mosior, M., Ford, A., Gonzalez-DeWhitt, P. A., Hale, J. E., Li, D., Qiu, Y., Fraser, C. C., Yang, D. D.*, et al.* (2005). Transgenic angiopoietin-like (angptl)4 overexpression and targeted disruption of angptl4 and angptl3: regulation of triglyceride metabolism. Endocrinology *146*, 4943-4950.

Kotowski, I. K., Pertsemlidis, A., Luke, A., Cooper, R. S., Vega, G. L., Cohen, J. C., and Hobbs, H. H. (2006). A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. Am J Hum Genet *78*, 410-422.
Krey, G., Braissant, O., L'Horset, F., Kalkhoven, E., Perroud, M., Parker, M. G., and Wahli, W. (1997). Fatty acids, eicosanoids, and hypolipidemic agents identified as ligands of peroxisome proliferator-activated receptors by coactivator-dependent receptor ligand assay. Mol Endocrinol *11*, 779-791.
Kulinski, A., Rustaeus, S., and Vance, J. E. (2002). Microsomal triacylglycerol transfer protein is required for lumenal accretion of triacylglycerol not associated with ApoB, as well as for ApoB lipidation. J Biol Chem *277*, 31516-31525.
Kuriyama, M., Yoshidome, H., Nakahara, K., Nakagawa, H., Fujiyama, J., Take, H., and Osame, M. (1999). Blood dolichols in a patient with abetalipoproteinaemia. Ann Clin Biochem *36 (Pt 2)*, 176-179.
Lagace, T. A., Curtis, D. E., Garuti, R., McNutt, M. C., Park, S. W., Prather, H. B., Anderson, N. N., Ho, Y. K., Hammer, R. E., and Horton, J. D. (2006). Secreted PCSK9 decreases the number of LDL receptors in hepatocytes and in livers of parabiotic mice. J Clin Invest *116*, 2995-3005.
Laird, N. M., Horvath, S., and Xu, X. (2000). Implementing a unified approach to family-based tests of association. Genet Epidemiol *19 Suppl 1*, S36-42.
Lalanne, F., Lambert, G., Amar, M. J., Chetiveaux, M., Zair, Y., Jarnoux, A. L., Ouguerram, K., Friburg, J., Seidah, N. G., Brewer, H. B., Jr.*, et al.* (2005). Wild-type PCSK9 inhibits LDL clearance but does not affect apoB-containing lipoprotein production in mouse and cultured cells. J Lipid Res *46*, 1312-1319.
Lander, E., and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet *11*, 241-247.
Langmann, T., Mauerer, R., Zahn, A., Moehle, C., Probst, M., Stremmel, W., and Schmitz, G. (2003). Real-time reverse transcription-PCR expression profiling of the complete human ATP-binding cassette transporter superfamily in various tissues. Clin Chem *49*, 230-238.
Law, A., and Scott, J. (1990). A cross-species comparison of the apolipoprotein B domain that binds to the LDL receptor. J Lipid Res *31*, 1109-1120.
Lee, J. C., Weissglas-Volkov, D., Kyttala, M., Dastani, Z., Cantor, R. M., Sobel, E. M., Plaisier, C. L., Engert, J. C., van Greevenbroek, M. M., Kane, J. P.*, et al.* (2008a). WW-domain-containing oxidoreductase is associated with low plasma HDL-C levels. Am J Hum Genet *83*, 180-192.
Lee, J. C., Weissglas-Volkov, D., Kyttala, M., Sinsheimer, J. S., Jokiaho, A., de Bruin, T. W., Lusis, A. J., Brennan, M. L., van Greevenbroek, M. M., van der Kallen, C. J.*, et*

al. (2007). USF1 contributes to high serum lipid levels in Dutch FCHL families and U.S. whites with coronary artery disease. Arterioscler Thromb Vasc Biol *27*, 2222-2227.
Lee, M. H., Lu, K., Hazard, S., Yu, H., Shulenin, S., Hidaka, H., Kojima, H., Allikmets, R., Sakuma, N., Pegoraro, R.*, et al.* (2001). Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption. Nat Genet *27*, 79-83.
Lee, S., Kasif, S., Weng, Z., and Cantor, C. R. (2008b). Quantitative analysis of single nucleotide polymorphisms within copy number variation. PLoS One *3*, e3906.
Lehmann, J. M., Kliewer, S. A., Moore, L. B., Smith-Oliver, T. A., Oliver, B. B., Su, J. L., Sundseth, S. S., Winegar, D. A., Blanchard, D. E., Spencer, T. A., and Willson, T. M. (1997). Activation of the nuclear receptor LXR by oxysterols defines a new hormone response pathway. J Biol Chem *272*, 3137-3140.
Lehmann, J. M., Moore, L. B., Smith-Oliver, T. A., Wilkison, W. O., Willson, T. M., and Kliewer, S. A. (1995). An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). J Biol Chem *270*, 12953-12956.
Leigh, S. E., Foster, A. H., Whittall, R. A., Hubbart, C. S., and Humphries, S. E. (2008). Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. Ann Hum Genet *72*, 485-498.
Levi, A., Kohn, Y., Kanyas, K., Amann, D., Pae, C. U., Hamdan, A., Segman, R. H., Avidan, N., Karni, O., Korner, M.*, et al.* (2005). Fine mapping of a schizophrenia susceptibility locus at chromosome 6q23: increased evidence for linkage and reduced linkage interval. Eur J Hum Genet *13*, 763-771.
Lewington, S., Whitlock, G., Clarke, R., Sherliker, P., Emberson, J., Halsey, J., Qizilbash, N., Peto, R., and Collins, R. (2007). Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. Lancet *370*, 1829-1839.
Lewontin, R. C. (1964). The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. Genetics *49*, 49-67.
Lewontin, R. C., and Kojima, K. (1960). The evolutionary dynamics of complex polymorphisms. Evolution *14*, 458-472.
Li, A. C., and Glass, C. K. (2004). PPAR- and LXR-dependent pathways controlling lipid metabolism and the development of atherosclerosis. J Lipid Res *45*, 2161-2173.
Li, J., Horstman, B., and Chen, Y. (2011). Detecting epistatic effects in association studies at a genomic level based on an ensemble approach. Bioinformatics *27*, i222-229.

Li, J., and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity *95*, 221-227.
Li, W., Tu, D., Brunger, A. T., and Ye, Y. (2007). A ubiquitin ligase transfers preformed polyubiquitin chains from a conjugating enzyme to a substrate. Nature *446*, 333-337.
Li, W. D., Dong, C., Li, D., Zhao, H., and Price, R. A. (2004). An obesity-related locus in chromosome region 12q23-24. Diabetes *53*, 812-820.
Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype imputation. Annu Rev Genomics Hum Genet *10*, 387-406.
Li, Y., Willer, C. J., Ding, J., Scheet, P., and Abecasis, G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol *34*, 816-834.
Lichtenstein, A. H., Appel, L. J., Brands, M., Carnethon, M., Daniels, S., Franch, H. A., Franklin, B., Kris-Etherton, P., Harris, W. S., Howard, B.*, et al.* (2006). Summary of American Heart Association Diet and Lifestyle Recommendations revision 2006. Arterioscler Thromb Vasc Biol *26*, 2186-2191.
Lichtenstein, L., Berbee, J. F., van Dijk, S. J., van Dijk, K. W., Bensadoun, A., Kema, I. P., Voshol, P. J., Muller, M., Rensen, P. C., and Kersten, S. (2007). Angptl4 upregulates cholesterol synthesis in liver via inhibition of LPL- and HL-dependent hepatic cholesterol uptake. Arterioscler Thromb Vasc Biol *27*, 2420-2427.
Lin, S., and Biswas, S. (2004). On modeling locus heterogeneity using mixture distributions. BMC Genet *5*, 29.
Lin, X., Takemori, H., Katoh, Y., Doi, J., Horike, N., Makino, A., Nonaka, Y., and Okamoto, M. (2001). Salt-inducible kinase is involved in the ACTH/cAMP-dependent protein kinase signaling in Y1 mouse adrenocortical tumor cells. Mol Endocrinol *15*, 1264-1276.
Linsel-Nitschke, P., Gotz, A., Erdmann, J., Braenne, I., Braund, P., Hengstenberg, C., Stark, K., Fischer, M., Schreiber, S., El Mokhtari, N. E.*, et al.* (2008). Lifelong reduction of LDL-cholesterol related to a common variant in the LDL-receptor gene decreases the risk of coronary artery disease--a Mendelian Randomisation study. PLoS One *3*, e2986.
Liscum, L., Finer-Moore, J., Stroud, R. M., Luskey, K. L., Brown, M. S., and Goldstein, J. L. (1985). Domain structure of 3-hydroxy-3-methylglutaryl coenzyme A reductase, a glycoprotein of the endoplasmic reticulum. J Biol Chem *260*, 522-530.
Liu, J., Afroza, H., Rader, D. J., and Jin, W. (2010). Angiopoietin-like protein 3 inhibits lipoprotein lipase activity through enhancing its cleavage by proprotein convertases. J Biol Chem *285*, 27561-27570.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.
Mackay, J., and Mensah, G. A. (2004). The Atlas of Heart Disease and Stroke: Geneva: World Health Organisation).
Magana, M. M., Lin, S. S., Dooley, K. A., and Osborne, T. F. (1997). Sterol regulation of acetyl coenzyme A carboxylase promoter requires two interdependent binding sites for sterol regulatory element binding proteins. J Lipid Res *38*, 1630-1638.
Magana, M. M., and Osborne, T. F. (1996). Two tandem binding sites for sterol regulatory element binding proteins are required for sterol regulation of fatty-acid synthase promoter. J Biol Chem *271*, 32689-32694.
Mahaney, M. C., Almasy, L., Rainwater, D. L., VandeBerg, J. L., Cole, S. A., Hixson, J. E., Blangero, J., and MacCluer, J. W. (2003). A quantitative trait locus on chromosome 16q influences variation in plasma HDL-C levels in Mexican Americans. Arterioscler Thromb Vasc Biol *23*, 339-345.
Mahley, R. W., and Huang, Y. (2007). Atherogenic remnant lipoproteins: role for proteoglycans in trapping, transferring, and internalizing. J Clin Invest *117*, 94-98.
Manichaikul, A., Chen, W. M., Williams, K., Wong, Q., Sale, M. M., Pankow, J. S., Tsai, M. Y., Rotter, J. I., Rich, S. S., and Mychaleckyj, J. C. (2011). Analysis of family- and population-based samples in cohort genome-wide association studies. Hum Genet.
Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A.*, et al.* (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753.
Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. Nat Rev Genet *11*, 499-511.
Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet *39*, 906-913.
Martin, E. R., Monks, S. A., Warren, L. L., and Kaplan, N. L. (2000). A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet *67*, 146-154.
Martin, G., Schoonjans, K., Lefebvre, A. M., Staels, B., and Auwerx, J. (1997). Coordinate regulation of the expression of the fatty acid transport protein and acyl-CoA synthetase genes by PPARalpha and PPARgamma activators. J Biol Chem *272*, 28210-28217.

Maxwell, K. N., and Breslow, J. L. (2004). Adenoviral-mediated expression of Pcsk9 in mice results in a low-density lipoprotein receptor knockout phenotype. Proc Natl Acad Sci U S A *101*, 7100-7105.
McCarroll, S. A., and Altshuler, D. M. (2007). Copy-number variation and association studies of human disease. Nat Genet *39*, S37-42.
McCarroll, S. A., Kuruvilla, F. G., Korn, J. M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M. H., de Bakker, P. I., Maller, J. B., Kirby, A.*, et al.* (2008). Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet *40*, 1166-1174.
Medeiros, A. M., Alves, A. C., Francisco, V., and Bourbon, M. (2010). Update of the Portuguese Familial Hypercholesterolaemia Study. Atherosclerosis *212*, 553-558.
Meex, S. J., Andreo, U., Sparks, J. D., and Fisher, E. A. (2011). Huh-7 or HepG2 cells: which is the better model for studying human apolipoprotein-B100 assembly and secretion? J Lipid Res *52*, 152-158.
Meex, S. J., van der Kallen, C. J., van Greevenbroek, M. M., Eurlings, P. M., El Hasnaoui, M., Evelo, C. T., Lindsey, P. J., Luiken, J. J., Glatz, J. F., and de Bruin, T. W. (2005). Up-regulation of CD36/FAT in preadipocytes in familial combined hyperlipidemia. Faseb J *19*, 2063-2065.
Miao, H., Jiang, W., Ge, L., Li, B., and Song, B. (2010). Tetra-glutamic acid residues adjacent to Lys248 in HMG-CoA reductase are critical for the ubiquitination mediated by gp78 and UBE2G2. Acta Biochim Biophys Sin (Shanghai) *42*, 303-310.
Miettinen, T. A., Strandberg, T. E., and Gylling, H. (2000). Noncholesterol sterols and cholesterol lowering by long-term simvastatin treatment in coronary patients: relation to basal serum cholestanol. Arterioscler Thromb Vasc Biol *20*, 1340-1346.
Miettinen, T. A., Tilvis, R. S., and Kesaniemi, Y. A. (1989). Serum cholestanol and plant sterol levels in relation to cholesterol metabolism in middle-aged men. Metabolism *38*, 136-140.
Miettinen, T. A., Tilvis, R. S., and Kesaniemi, Y. A. (1990). Serum plant sterols and cholesterol precursors reflect cholesterol absorption and synthesis in volunteers of a randomly selected male population. Am J Epidemiol *131*, 20-31.
Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., and et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science *266*, 66-71.
Miller, G. J., Bauer, K. A., Barzegar, S., Foley, A. J., Mitchell, J. P., Cooper, J. A., and Rosenberg, R. D. (1995). The effects of quality and timing of venepuncture on

markers of blood coagulation in healthy middle-aged men. Thromb Haemost *73*, 82-86.
Miller, W. L. (2007). Steroidogenic acute regulatory protein (StAR), a novel mitochondrial cholesterol transporter. Biochim Biophys Acta *1771*, 663-676.
Monajemi, H., Zhang, L., Li, G., Jeninga, E. H., Cao, H., Maas, M., Brouwer, C. B., Kalkhoven, E., Stroes, E., Hegele, R. A., and Leff, T. (2007). Familial partial lipodystrophy phenotype resulting from a single-base mutation in deoxyribonucleic acid-binding domain of peroxisome proliferator-activated receptor-gamma. J Clin Endocrinol Metab *92*, 1606-1612.
Moore, M. L., Park, E. A., and McMillin, J. B. (2003). Upstream stimulatory factor represses the induction of carnitine palmitoyltransferase-Ibeta expression by PGC-1. J Biol Chem *278*, 17263-17268.
Moreno, C., Dumas, P., Kaldunski, M. L., Tonellato, P. J., Greene, A. S., Roman, R. J., Cheng, Q., Wang, Z., Jacob, H. J., and Cowley, A. W., Jr. (2003). Genomic map of cardiovascular phenotypes of hypertension in female Dahl S rats. Physiol Genomics *15*, 243-257.
Morton, N. E. (1955). Sequential tests for the detection of linkage. Am J Hum Genet *7*, 277-318.
Morton, N. E. (2005). Linkage disequilibrium maps and association mapping. J Clin Invest *115*, 1425-1430.
Murdoch, J. C., Rodger, J. C., Rao, S. S., Fletcher, C. D., and Dunnigan, M. G. (1977). Down's syndrome: an atheroma-free model? Br Med J *2*, 226-228.
Murota, K., and Storch, J. (2005). Uptake of micellar long-chain fatty acid and sn-2-monoacylglycerol into human intestinal Caco-2 cells exhibits characteristics of protein-mediated transport. J Nutr *135*, 1626-1630.
Musunuru, K., Pirruccello, J. P., Do, R., Peloso, G. M., Guiducci, C., Sougnez, C., Garimella, K. V., Fisher, S., Abreu, J., Barry, A. J.*, et al.* (2010). Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. N Engl J Med *363*, 2220-2227.
Nagamine, K., Kudoh, J., Minoshima, S., Kawasaki, K., Asakawa, S., Ito, F., and Shimizu, N. (1998). Molecular cloning of a novel putative Ca2+ channel protein (TRPC7) highly expressed in brain. Genomics *54*, 124-131.
Namekata, K., Enokido, Y., Ishii, I., Nagai, Y., Harada, T., and Kimura, H. (2004). Abnormal lipid metabolism in cystathionine beta-synthase-deficient mice, an animal model for hyperhomocysteinemia. J Biol Chem *279*, 52961-52969.
Naoumova, R. P., Bonney, S. A., Eichenbaum-Voline, S., Patel, H. N., Jones, B., Jones, E. L., Amey, J., Colilla, S., Neuwirth, C. K., Allotey, R.*, et al.* (2003).

Confirmed locus on chromosome 11p and candidate loci on 6q and 8p for the triglyceride and cholesterol traits of combined hyperlipidemia. Arterioscler Thromb Vasc Biol *23*, 2070-2077.
Narcisi, T. M., Shoulders, C. C., Chester, S. A., Read, J., Brett, D. J., Harrison, G. B., Grantham, T. T., Fox, M. F., Povey, S., de Bruin, T. W., and et al. (1995). Mutations of the microsomal triglyceride-transfer-protein gene in abetalipoproteinemia. Am J Hum Genet *57*, 1298-1310.
Naukkarinen, J., Gentile, M., Soro-Paavonen, A., Saarela, J., Koistinen, H. A., Pajukanta, P., Taskinen, M. R., and Peltonen, L. (2005). USF1 and dyslipidemias: converging evidence for a functional intronic variant. Hum Mol Genet *14*, 2595-2605.
Naureckiene, S., Sleat, D. E., Lackland, H., Fensom, A., Vanier, M. T., Wattiaux, R., Jadot, M., and Lobel, P. (2000). Identification of HE1 as the second gene of Niemann-Pick C disease. Science *290*, 2298-2301.
Naylor, M. G., Weiss, S. T., and Lange, C. (2010). A Bayesian approach to genetic association studies with family-based designs. Genet Epidemiol *34*, 569-574.
Neale, B. M., and Purcell, S. (2008). The positives, protocols, and perils of genome-wide association. Am J Med Genet B Neuropsychiatr Genet *147B*, 1288-1294.
Nemeth, A., Myrdal, U., Veress, B., Rudling, M., Berglund, L., and Angelin, B. (1995). Studies on lipoprotein metabolism in a family with jejunal chylomicron retention. Eur J Clin Invest *25*, 271-280.
Ng, P. C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. Genome Res *11*, 863-874.
Nica, A. C., Montgomery, S. B., Dimas, A. S., Stranger, B. E., Beazley, C., Barroso, I., and Dermitzakis, E. T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet *6*, e1000895.
Nica, A. C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., Travers, M., Potter, S., Grundberg, E., Small, K.*, et al.* (2011). The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet *7*, e1002003.
Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., and Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet *6*, e1000888.
Nierman, M. C., Rip, J., Kuivenhoven, J. A., van Raalte, D. H., Hutten, B. A., Sakai, N., Kastelein, J. J., and Stroes, E. S. (2005). Carriers of the frequent lipoprotein lipase S447X variant exhibit enhanced postprandial apoprotein B-48 clearance. Metabolism *54*, 1499-1503.

Nikkila, E. A., and Aro, A. (1973). Family study of serum lipids and lipoproteins in coronary heart-disease. Lancet *1*, 954-959.
Nilsson, S. K., Lookene, A., Beckstead, J. A., Gliemann, J., Ryan, R. O., and Olivecrona, G. (2007). Apolipoprotein A-V interaction with members of the low density lipoprotein receptor gene family. Biochemistry *46*, 3896-3904.
Nohturfft, A., Brown, M. S., and Goldstein, J. L. (1998). Sterols regulate processing of carbohydrate chains of wild-type SREBP cleavage-activating protein (SCAP), but not sterol-resistant mutants Y298C or D443N. Proc Natl Acad Sci U S A *95*, 12848-12853.
Nordestgaard, B. G., Abildgaard, S., Wittrup, H. H., Steffensen, R., Jensen, G., and Tybjaerg-Hansen, A. (1997). Heterozygous lipoprotein lipase deficiency: frequency in the general population, effect on plasma lipid levels, and risk of ischemic heart disease. Circulation *96*, 1737-1744.
Nordestgaard, B. G., Benn, M., Schnohr, P., and Tybjaerg-Hansen, A. (2007). Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. Jama *298*, 299-308.
North, K. E., Miller, M. B., Coon, H., Martin, L. J., Peacock, J. M., Arnett, D., Zhang, B., Province, M., Oberman, A., Blangero, J.*, et al.* (2005). Evidence for a gene influencing fasting LDL cholesterol and triglyceride levels on chromosome 21q. Atherosclerosis *179*, 119-125.
Notarangelo, L. D., and Badolato, R. (2009). Leukocyte trafficking in primary immunodeficiencies. J Leukoc Biol *85*, 335-343.
Nothnagel, M., Ellinghaus, D., Schreiber, S., Krawczak, M., and Franke, A. (2009). A comprehensive evaluation of SNP genotype imputation. Hum Genet *125*, 163-171.
Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet *74*, 765-769.
O'Brien, P. J., Alborn, W. E., Sloan, J. H., Ulmer, M., Boodhoo, A., Knierman, M. D., Schultze, A. E., and Konrad, R. J. (2005). The novel apolipoprotein A5 is present in human serum, is associated with VLDL, HDL, and chylomicrons, and circulates at very low concentrations compared with other apolipoproteins. Clin Chem *51*, 351-359.
O'Connell, J. R., and Weeks, D. E. (1998). PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet *63*, 259-266.
Oelkers, P., Kirby, L. C., Heubi, J. E., and Dawson, P. A. (1997). Primary bile acid malabsorption caused by mutations in the ileal sodium-dependent bile acid transporter gene (SLC10A2). J Clin Invest *99*, 1880-1887.

Oliphant, A., Barker, D. L., Stuelpnagel, J. R., and Chee, M. S. (2002). BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques *Suppl*, 56-58, 60-51.
Olivecrona, G., Ehrenborg, E., Semb, H., Makoveichuk, E., Lindberg, A., Hayden, M. R., Gin, P., Davies, B. S., Weinstein, M. M., Fong, L. G.*, et al.* (2010). Mutation of conserved cysteines in the Ly6 domain of GPIHBP1 in familial chylomicronemia. J Lipid Res *51*, 1535-1545.
Olivecrona, T., Bengtsson-Olivecrona, G., Osborne, J. C., Jr., and Kempner, E. S. (1985). Molecular size of bovine lipoprotein lipase as determined by radiation inactivation. J Biol Chem *260*, 6888-6891.
Oliver, W. R., Jr., Shenk, J. L., Snaith, M. R., Russell, C. S., Plunket, K. D., Bodkin, N. L., Lewis, M. C., Winegar, D. A., Sznaidman, M. L., Lambert, M. H.*, et al.* (2001). A selective peroxisome proliferator-activated receptor delta agonist promotes reverse cholesterol transport. Proc Natl Acad Sci U S A *98*, 5306-5311.
Oriente, F., Fernandez Diaz, L. C., Miele, C., Iovino, S., Mori, S., Diaz, V. M., Troncone, G., Cassese, A., Formisano, P., Blasi, F., and Beguinot, F. (2008). Prep1 deficiency induces protection from diabetes and increased insulin sensitivity through a p160-mediated mechanism. Mol Cell Biol *28*, 5634-5645.
Osborne, J. C., Jr., Bengtsson-Olivecrona, G., Lee, N. S., and Olivecrona, T. (1985). Studies on inactivation of lipoprotein lipase: role of the dimer to monomer dissociation. Biochemistry *24*, 5606-5611.
Ostlund, R. E., Jr., Bosner, M. S., and Stenson, W. F. (1999). Cholesterol absorption efficiency declines at moderate dietary doses in normal human subjects. J Lipid Res *40*, 1453-1458.
Ott, J. (1999). Analysis of Human Genetic Linkage, Third edition edn (Baltimore: Johns Hopkins University Press).
Pahl, R., and Schafer, H. (2010). PERMORY: an LD-exploiting permutation test algorithm for powerful genome-wide association testing. Bioinformatics *26*, 2093-2100.
Pajukanta, P., Allayee, H., Krass, K. L., Kuraishy, A., Soro, A., Lilja, H. E., Mar, R., Taskinen, M. R., Nuotio, I., Laakso, M.*, et al.* (2003). Combined analysis of genome scans of dutch and finnish families reveals a susceptibility locus for high-density lipoprotein cholesterol on chromosome 16q. Am J Hum Genet *72*, 903-917.
Pajukanta, P., Lilja, H. E., Sinsheimer, J. S., Cantor, R. M., Lusis, A. J., Gentile, M., Duan, X. J., Soro-Paavonen, A., Naukkarinen, J., Saarela, J.*, et al.* (2004). Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). Nat Genet *36*, 371-376.

Pajukanta, P., Nuotio, I., Terwilliger, J. D., Porkka, K. V., Ylitalo, K., Pihlajamaki, J., Suomalainen, A. J., Syvanen, A. C., Lehtimaki, T., Viikari, J. S.*, et al.* (1998). Linkage of familial combined hyperlipidaemia to chromosome 1q21-q23. Nat Genet *18*, 369-373.

Pajukanta, P., Terwilliger, J. D., Perola, M., Hiekkalinna, T., Nuotio, I., Ellonen, P., Parkkonen, M., Hartiala, J., Ylitalo, K., Pihlajamaki, J.*, et al.* (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.

Palmer, L. J., Celedon, J. C., Chapman, H. A., Speizer, F. E., Weiss, S. T., and Silverman, E. K. (2003). Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. Hum Mol Genet *12*, 1199-1210.

Pan, M., Liang Js, J. S., Fisher, E. A., and Ginsberg, H. N. (2002). The late addition of core lipids to nascent apolipoprotein B100, resulting in the assembly and secretion of triglyceride-rich lipoproteins, is independent of both microsomal triglyceride transfer protein activity and new triglyceride synthesis. J Biol Chem *277*, 4413-4421.

Parhofer, K. G., Barrett, P. H., Bier, D. M., and Schonfeld, G. (1992). Lipoproteins containing the truncated apolipoprotein, Apo B-89, are cleared from human plasma more rapidly than Apo B-100-containing lipoproteins in vivo. J Clin Invest *89*, 1931-1937.

Park, S. W., Moon, Y. A., and Horton, J. D. (2004). Post-transcriptional regulation of low density lipoprotein receptor protein by proprotein convertase subtilisin/kexin type 9a in mouse liver. J Biol Chem *279*, 50630-50638.

Paterniti, J. R., Jr., Brown, W. V., Ginsberg, H. N., and Artzt, K. (1983). Combined lipase deficiency (cld): a lethal mutation on chromosome 17 of the mouse. Science *221*, 167-169.

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

Pentchev, P. G., Comly, M. E., Kruth, H. S., Tokoro, T., Butler, J., Sokol, J., Filling-Katz, M., Quirk, J. M., Marshall, D. C., Patel, S., and et al. (1987). Group C Niemann-Pick disease: faulty regulation of low-density lipoprotein uptake and cholesterol storage in cultured fibroblasts. Faseb J *1*, 40-45.

Perraud, A. L., Fleig, A., Dunn, C. A., Bagley, L. A., Launay, P., Schmitz, C., Stokes, A. J., Zhu, Q., Bessman, M. J., Penner, R.*, et al.* (2001). ADP-ribose gating of the calcium-permeable LTRPC2 channel revealed by Nudix motif homology. Nature *411*, 595-599.

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

Peters, J. M., Aoyama, T., Burns, A. M., and Gonzalez, F. J. (2003). Bezafibrate is a dual ligand for PPARalpha and PPARbeta: studies using null mice. Biochim Biophys Acta *1632*, 80-89.

Peters, J. M., Hennuyer, N., Staels, B., Fruchart, J. C., Fievet, C., Gonzalez, F. J., and Auwerx, J. (1997). Alterations in lipoprotein metabolism in peroxisome proliferator-activated receptor alpha-deficient mice. J Biol Chem *272*, 27307-27312.

Peterson, J., Ayyobi, A. F., Ma, Y., Henderson, H., Reina, M., Deeb, S. S., Santamarina-Fojo, S., Hayden, M. R., and Brunzell, J. D. (2002). Structural and functional consequences of missense mutations in exon 5 of the lipoprotein lipase gene. J Lipid Res *43*, 398-406.

Plaisier, C. L., Horvath, S., Huertas-Vazquez, A., Cruz-Bautista, I., Herrera, M. F., Tusie-Luna, T., Aguilar-Salinas, C., and Pajukanta, P. (2009). A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. PLoS Genet *5*, e1000642.

Poirier, H., Degrace, P., Niot, I., Bernard, A., and Besnard, P. (1996). Localization and regulation of the putative membrane fatty-acid transporter (FAT) in the small intestine. Comparison with fatty acid-binding proteins (FABP). Eur J Biochem *238*, 368-373.

Poirier, H., Niot, I., Monnot, M. C., Braissant, O., Meunier-Durmort, C., Costet, P., Pineau, T., Wahli, W., Willson, T. M., and Besnard, P. (2001). Differential involvement of peroxisome-proliferator-activated receptors alpha and delta in fibrate and fatty-acid-mediated inductions of the gene encoding liver fatty-acid-binding protein in the liver and the small intestine. Biochem J *355*, 481-488.

Poirier, S., Mayer, G., Poupon, V., McPherson, P. S., Desjardins, R., Ly, K., Asselin, M. C., Day, R., Duclos, F. J., Witmer, M.*, et al.* (2009). Dissection of the endogenous cellular pathways of PCSK9-induced low density lipoprotein receptor degradation: evidence for an intracellular route. J Biol Chem *284*, 28856-28864.

Polisecki, E., Peter, I., Simon, J. S., Hegele, R. A., Robertson, M., Ford, I., Shepherd, J., Packard, C., Jukema, J. W., de Craen, A. J.*, et al.* (2010). Genetic variation at the NPC1L1 gene locus, plasma lipoproteins, and heart disease risk in the elderly. J Lipid Res *51*, 1201-1207.

Porter, F. D., and Herman, G. E. (2011). Malformation syndromes caused by disorders of cholesterol synthesis. J Lipid Res *52*, 6-34.

Prasad, S. S., Garg, A., and Agarwal, A. K. (2011). Enzymatic activities of the human AGPAT isoform 3 and isoform 5: localization of AGPAT5 to mitochondria. J Lipid Res *52*, 451-462.
Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet *38*, 904-909.
Priore Oliva, C., Pisciotta, L., Li Volti, G., Sambataro, M. P., Cantafora, A., Bellocchio, A., Catapano, A., Tarugi, P., Bertolini, S., and Calandra, S. (2005). Inherited apolipoprotein A-V deficiency in severe hypertriglyceridemia. Arterioscler Thromb Vasc Biol *25*, 411-417.
Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., Boehnke, M., Abecasis, G. R., and Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336-2337.
Puffenberger, E. G., Kauffman, E. R., Bolk, S., Matise, T. C., Washington, S. S., Angrist, M., Weissenbach, J., Garver, K. L., Mascari, M., Ladda, R., and et al. (1994). Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. Hum Mol Genet *3*, 1217-1225.
Pullinger, C. R., Eng, C., Salen, G., Shefer, S., Batta, A. K., Erickson, S. K., Verhagen, A., Rivera, C. R., Mulvihill, S. J., Malloy, M. J., and Kane, J. P. (2002). Human cholesterol 7alpha-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype. J Clin Invest *110*, 109-117.
Pullinger, C. R., Hennessy, L. K., Chatterton, J. E., Liu, W., Love, J. A., Mendel, C. M., Frost, P. H., Malloy, M. J., Schumaker, V. N., and Kane, J. P. (1995). Familial ligand-defective apolipoprotein B. Identification of a new mutation that decreases LDL receptor binding affinity. J Clin Invest *95*, 1225-1234.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., and Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.
Qiao, Q., Tervahauta, M., Nissinen, A., and Tuomilehto, J. (2000). Mortality from all causes and from coronary heart disease related to smoking and changes in smoking during a 35-year follow-up of middle-aged Finnish men. Eur Heart J *21*, 1621-1626.
Quehenberger, O., Armando, A. M., Brown, A. H., Milne, S. B., Myers, D. S., Merrill, A. H., Bandyopadhyay, S., Jones, K. N., Kelly, S., Shaner, R. L.*, et al.* (2010). Lipidomics reveals a remarkable diversity of lipids in human plasma. J Lipid Res *51*, 3299-3305.

Rakhshandehroo, M., Sanderson, L. M., Matilainen, M., Stienstra, R., Carlberg, C., de Groot, P. J., Muller, M., and Kersten, S. (2007). Comprehensive analysis of PPARalpha-dependent regulation of hepatic lipid metabolism by expression profiling. PPAR Res *2007*, 26839.

Ranheim, T., Kulseth, M. A., Berge, K. E., and Leren, T. P. (2006). Model system for phenotypic characterization of sequence variations in the LDL receptor gene. Clin Chem *52*, 1469-1479.

Rashid, S., Curtis, D. E., Garuti, R., Anderson, N. N., Bashmakov, Y., Ho, Y. K., Hammer, R. E., Moon, Y. A., and Horton, J. D. (2005). Decreased plasma cholesterol and hypersensitivity to statins in mice lacking Pcsk9. Proc Natl Acad Sci U S A *102*, 5374-5379.

Repa, J. J., Berge, K. E., Pomajzl, C., Richardson, J. A., Hobbs, H., and Mangelsdorf, D. J. (2002). Regulation of ATP-binding cassette sterol transporters ABCG5 and ABCG8 by the liver X receptors alpha and beta. J Biol Chem *277*, 18793-18800.

Repa, J. J., Turley, S. D., Lobaccaro, J. A., Medina, J., Li, L., Lustig, K., Shan, B., Heyman, R. A., Dietschy, J. M., and Mangelsdorf, D. J. (2000). Regulation of absorption and ABC1-mediated efflux of cholesterol by RXR heterodimers. Science *289*, 1524-1529.

Ribeiro, A., Pastier, D., Kardassis, D., Chambaz, J., and Cardot, P. (1999). Cooperative binding of upstream stimulatory factor and hepatic nuclear factor 4 drives the transcription of the human apolipoprotein A-II gene. J Biol Chem *274*, 1216-1225.

Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L., and et al. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science *245*, 1066-1073.

Rios, J. J., Shastry, S., Jasso, J., Hauser, N., Garg, A., Bensadoun, A., Cohen, J. C., and Hobbs, H. H. (2011). Deletion of GPIHBP1 causing severe chylomicronemia. J Inherit Metab Dis.

Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. Science *273*, 1516-1517.

Roitelman, J., Olender, E. H., Bar-Nun, S., Dunn, W. A., Jr., and Simoni, R. D. (1992). Immunological evidence for eight spans in the membrane domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase: implications for enzyme degradation in the endoplasmic reticulum. J Cell Biol *117*, 959-973.

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

Romeo, S., Yin, W., Kozlitina, J., Pennacchio, L. A., Boerwinkle, E., Hobbs, H. H., and Cohen, J. C. (2009). Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. J Clin Invest *119*, 70-79.

Rosen, E. D., Sarraf, P., Troy, A. E., Bradwin, G., Moore, K., Milstone, D. S., Spiegelman, B. M., and Mortensen, R. M. (1999). PPAR gamma is required for the differentiation of adipose tissue in vivo and in vitro. Mol Cell *4*, 611-617.

Rosengren, A., and Wilhelmsen, L. (1997). Physical activity protects against coronary death and deaths from all causes in middle-aged men. Evidence from a 20-year follow-up of the primary prevention study in Goteborg. Ann Epidemiol *7*, 69-75.

Ross, S. E., Hemati, N., Longo, K. A., Bennett, C. N., Lucas, P. C., Erickson, R. L., and MacDougald, O. A. (2000). Inhibition of adipogenesis by Wnt signaling. Science *289*, 950-953.

Roy, C. C., Levy, E., Green, P. H., Sniderman, A., Letarte, J., Buts, J. P., Orquin, J., Brochu, P., Weber, A. M., Morin, C. L., and et al. (1987). Malabsorption, hypocholesterolemia, and fat-filled enterocytes with increased intestinal apoprotein B. Chylomicron retention disease. Gastroenterology *92*, 390-399.

Rudenko, G., Henry, L., Henderson, K., Ichtchenko, K., Brown, M. S., Goldstein, J. L., and Deisenhofer, J. (2002). Structure of the LDL receptor extracellular domain at endosomal pH. Science *298*, 2353-2358.

Russell, D. W. (2003). The enzymes, regulation, and genetics of bile acid synthesis. Annu Rev Biochem *72*, 137-174.

Russell, D. W., Brown, M. S., and Goldstein, J. L. (1989). Different combinations of cysteine-rich repeats mediate binding of low density lipoprotein receptor to two different proteins. J Biol Chem *264*, 21682-21688.

Saarela, J., Rehn, M., Oikarinen, A., Autio-Harmainen, H., and Pihlajaniemi, T. (1998). The short and long forms of type XVIII collagen show clear tissue specificities in their expression and location in basement membrane zones in humans. Am J Pathol *153*, 611-626.

Sakamoto, N., Larson, J. E., Iyer, R. R., Montermini, L., Pandolfo, M., and Wells, R. D. (2001). GGA*TCC-interrupted triplets in long GAA*TTC repeats inhibit the formation of triplex and sticky DNA structures, alleviate transcription inhibition, and reduce genetic instabilities. J Biol Chem *276*, 27178-27187.

Salen, G., Horak, I., Rothkopf, M., Cohen, J. L., Speck, J., Tint, G. S., Shore, V., Dayal, B., Chen, T., and Shefer, S. (1985). Lethal atherosclerosis associated with abnormal plasma and tissue sterol composition in sitosterolemia with xanthomatosis. J Lipid Res *26*, 1126-1133.
Salen, G., Tint, G. S., Shefer, S., Shore, V., and Nguyen, L. (1992). Increased sitosterol absorption is offset by rapid elimination to prevent accumulation in heterozygotes with sitosterolemia. Arterioscler Thromb *12*, 563-568.
Salero, E., Gimenez, C., and Zafra, F. (2003). Identification of a non-canonical E-box motif as a regulatory element in the proximal promoter region of the apolipoprotein E gene. Biochem J *370*, 979-986.
Sandhu, M. S., Waterworth, D. M., Debenham, S. L., Wheeler, E., Papadakis, K., Zhao, J. H., Song, K., Yuan, X., Johnson, T., Ashford, S.*, et al.* (2008). LDL-cholesterol concentrations: a genome-wide association study. Lancet *371*, 483-491.
Sarwar, N., Danesh, J., Eiriksdottir, G., Sigurdsson, G., Wareham, N., Bingham, S., Boekholdt, S. M., Khaw, K. T., and Gudnason, V. (2007). Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. Circulation *115*, 450-458.
Sarwar, N., Sandhu, M. S., Ricketts, S. L., Butterworth, A. S., Di Angelantonio, E., Boekholdt, S. M., Ouwehand, W., Watkins, H., Samani, N. J., Saleheen, D.*, et al.* (2010). Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. Lancet *375*, 1634-1639.
Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I., Chen, H., Roix, J. J., Kathiresan, S., Hirschhorn, J. N., Daly, M. J.*, et al.* (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science *316*, 1331-1336.
Schadt, E. E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P. Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C.*, et al.* (2008). Mapping the genetic architecture of gene expression in human liver. PLoS Biol *6*, e107.
Schaefer, J. R., Sattler, A. M., Hackler, B., Kurt, B., Hackler, R., Maisch, B., and Soufi, M. (2004). Hyperlipidemia in patients with apolipoprotein E 2/2 phenotype: apolipoprotein A5 S19W mutation as a cofactor. Clin Chem *50*, 2214.
Schonfeld, G., Lin, X., and Yue, P. (2005). Familial hypobetalipoproteinemia: genetics and metabolism. Cell Mol Life Sci *62*, 1372-1378.
Schoonjans, K., Peinado-Onsurbe, J., Lefebvre, A. M., Heyman, R. A., Briggs, M., Deeb, S., Staels, B., and Auwerx, J. (1996). PPARalpha and PPARgamma activators direct a distinct tissue-specific transcriptional response via a PPRE in the lipoprotein lipase gene. Embo J *15*, 5336-5348.

Schork, N. J., Murray, S. S., Frazer, K. A., and Topol, E. J. (2009). Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev *19*, 212-219.
Schouten, J. P., McElgunn, C. J., Waaijer, R., Zwijnenburg, D., Diepvens, F., and Pals, G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res *30*, e57.
Schroepfer, G. J., Jr. (2000). Oxysterols: modulators of cholesterol metabolism and other processes. Physiol Rev *80*, 361-554.
Schulze, T. G., Chen, Y. S., Badner, J. A., McInnis, M. G., DePaulo, J. R., Jr., and McMahon, F. J. (2003). Additional, physically ordered markers increase linkage signal for bipolar disorder on chromosome 18q22. Biol Psychiatry *53*, 239-243.
Schunkert, H., Konig, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., Preuss, M., Stewart, A. F., Barbalic, M., Gieger, C*., et al.* (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet *43*, 333-338.
Schwarz, M., Lund, E. G., Setchell, K. D., Kayden, H. J., Zerwekh, J. E., Bjorkhem, I., Herz, J., and Russell, D. W. (1996). Disruption of cholesterol 7alpha-hydroxylase gene in mice. II. Bile acid deficiency is overcome by induction of oxysterol 7alpha-hydroxylase. J Biol Chem *271*, 18024-18031.
Schwarz, M., Russell, D. W., Dietschy, J. M., and Turley, S. D. (1998). Marked reduction in bile acid synthesis in cholesterol 7alpha-hydroxylase-deficient mice does not lead to diminished tissue cholesterol turnover or to hypercholesterolemia. J Lipid Res *39*, 1833-1843.
Scuteri, A., Sanna, S., Chen, W. M., Uda, M., Albai, G., Strait, J., Najjar, S., Nagaraja, R., Orru, M., Usala, G*., et al.* (2007). Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet *3*, e115.
Sendak, R. A., and Bensadoun, A. (1998). Identification of a heparin-binding domain in the distal carboxyl-terminal region of lipoprotein lipase by site-directed mutagenesis. J Lipid Res *39*, 1310-1315.
Sertie, A. L., Sossi, V., Camargo, A. A., Zatz, M., Brahe, C., and Passos-Bueno, M. R. (2000). Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). Hum Mol Genet *9*, 2051-2058.
Sever, N., Yang, T., Brown, M. S., Goldstein, J. L., and DeBose-Boyd, R. A. (2003). Accelerated degradation of HMG CoA reductase mediated by binding of insig-1 to its sterol-sensing domain. Mol Cell *11*, 25-33.

Shah, S. H., Kraus, W. E., Crossman, D. C., Granger, C. B., Haines, J. L., Jones, C. J., Mooser, V., Huang, L., Haynes, C., Dowdy, E.*, et al.* (2006). Serum lipids in the GENECARD study of coronary artery disease identify quantitative trait loci and phenotypic subsets on chromosomes 3q and 5q. Ann Hum Genet *70*, 738-748.
Sheikh, K., Camejo, G., Lanne, B., Halvarsson, T., Landergren, M. R., and Oakes, N. D. (2007). Beyond lipids, pharmacological PPARalpha activation has important effects on amino acid metabolism as studied in the rat. Am J Physiol Endocrinol Metab *292*, E1157-1165.
Shi, G., Simino, J., and Rao, D. C. (2011). Enriching rare variants using family-specific linkage information. BMC Proceedings *5*, S82
Shibuya, K., Obayashi, I., Asakawa, S., Minoshima, S., Kudoh, J., and Shimizu, N. (2004). A cluster of 21 keratin-associated protein genes within introns of another gene on human chromosome 21q22.3. Genomics *83*, 679-693.
Shoulders, C. C., Jones, E. L., and Naoumova, R. P. (2004). Genetics of familial combined hyperlipidemia and risk of coronary heart disease. Hum Mol Genet *13 Spec No 1*, R149-160.
Siddiqi, S. A., Gorelick, F. S., Mahan, J. T., and Mansbach, C. M., 2nd (2003). COPII proteins are required for Golgi fusion but not for endoplasmic reticulum budding of the pre-chylomicron transport vesicle. J Cell Sci *116*, 415-427.
Silverman, E. K., Palmer, L. J., Mosley, J. D., Barth, M., Senter, J. M., Brown, A., Drazen, J. M., Kwiatkowski, D. J., Chapman, H. A., Campbell, E. J.*, et al.* (2002). Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. Am J Hum Genet *70*, 1229-1239.
Simons, K., and Toomre, D. (2000). Lipid rafts and signal transduction. Nat Rev Mol Cell Biol *1*, 31-39.
Smith, C. A. (1963). Testing for Heterogeneity of Recombination Fraction Values in Human Genetics. Ann Hum Genet *27*, 175-182.
Smith, E. N., Chen, W., Kahonen, M., Kettunen, J., Lehtimaki, T., Peltonen, L., Raitakari, O. T., Salem, R. M., Schork, N. J., Shaw, M.*, et al.* (2010). Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. PLoS Genet *6*.
Solano, M. P., and Goldberg, R. B. (2006). Management of dyslipidemia in diabetes. Cardiol Rev *14*, 125-135.
Song, B. L., Javitt, N. B., and DeBose-Boyd, R. A. (2005). Insig-mediated degradation of HMG CoA reductase stimulated by lanosterol, an intermediate in the synthesis of cholesterol. Cell Metab *1*, 179-189.

Sorheim, I. C., and Gulsvik, A. (2008). Genetics of chronic obstructive pulmonary disease: a case-control study in Bergen, Norway. Clin Respir J *2 Suppl 1*, 129-131.
Soutar, A. K., and Naoumova, R. P. (2007). Mechanisms of disease: genetic causes of familial hypercholesterolemia. Nat Clin Pract Cardiovasc Med *4*, 214-225.
Soyal, S., Krempler, F., Oberkofler, H., and Patsch, W. (2006). PGC-1alpha: a potent transcriptional cofactor involved in the pathogenesis of type 2 diabetes. Diabetologia *49*, 1477-1488.
Spencer, C. C., Su, Z., Donnelly, P., and Marchini, J. (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet *5*, e1000477.
Staels, B., Vu-Dac, N., Kosykh, V. A., Saladin, R., Fruchart, J. C., Dallongeville, J., and Auwerx, J. (1995). Fibrates downregulate apolipoprotein C-III expression independent of induction of peroxisomal acyl coenzyme A oxidase. A potential mechanism for the hypolipidemic action of fibrates. J Clin Invest *95*, 705-712.
Stahl, A., Hirsch, D. J., Gimeno, R. E., Punreddy, S., Ge, P., Watson, N., Patel, S., Kotler, M., Raimondi, A., Tartaglia, L. A., and Lodish, H. F. (1999). Identification of the major intestinal fatty acid transport protein. Mol Cell *4*, 299-308.
Stanford, K. I., Bishop, J. R., Foley, E. M., Gonzales, J. C., Niesman, I. R., Witztum, J. L., and Esko, J. D. (2009). Syndecan-1 is the primary heparan sulfate proteoglycan mediating hepatic clearance of triglyceride-rich lipoproteins in mice. J Clin Invest *119*, 3236-3245.
Stephens, M., and Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. Nat Rev Genet *10*, 681-690.
Strautnieks, S. S., Bull, L. N., Knisely, A. S., Kocoshis, S. A., Dahl, N., Arnell, H., Sokal, E., Dahan, K., Childs, S., Ling, V.*, et al.* (1998). A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. Nat Genet *20*, 233-238.
Stremmel, W., Lotz, G., Strohmeyer, G., and Berk, P. D. (1985). Identification, isolation, and partial characterization of a fatty acid binding protein from rat jejunal microvillous membranes. J Clin Invest *75*, 1068-1076.
Sudhop, T., Lutjohann, D., Kodal, A., Igel, M., Tribble, D. L., Shah, S., Perevozskaya, I., and von Bergmann, K. (2002). Inhibition of intestinal cholesterol absorption by ezetimibe in humans. Circulation *106*, 1943-1948.
Sugii, S., Reid, P. C., Ohgami, N., Du, H., and Chang, T. Y. (2003). Distinct endosomal compartments in early trafficking of low density lipoprotein-derived cholesterol. J Biol Chem *278*, 27180-27189.

Sukonina, V., Lookene, A., Olivecrona, T., and Olivecrona, G. (2006). Angiopoietin-like protein 4 converts lipoprotein lipase to inactive monomers and modulates lipase activity in adipose tissue. Proc Natl Acad Sci U S A *103*, 17450-17455.
Sun, X., Wang, Y., Qing, H., Christensen, M. A., Liu, Y., Zhou, W., Tong, Y., Xiao, C., Huang, Y., Zhang, S.*, et al.* (2005). Distinct transcriptional regulation and function of the human BACE2 and BACE1 genes. Faseb J *19*, 739-749.
Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., 3rd, Kondrashov, A. S., and Bork, P. (2001). Prediction of deleterious human alleles. Hum Mol Genet *10*, 591-597.
Suzuki, O. T., Sertie, A. L., Der Kaloustian, V. M., Kok, F., Carpenter, M., Murray, J., Czeizel, A. E., Kliemann, S. E., Rosemberg, S., Monteiro, M.*, et al.* (2002). Molecular analysis of collagen XVIII reveals novel mutations, presence of a third isoform, and possible genetic heterogeneity in Knobloch syndrome. Am J Hum Genet *71*, 1320-1329.
Sviridov, D. D., Safonova, I. G., Talalaev, A. G., Repin, V. S., and Smirnov, V. N. (1986). Regulation of cholesterol synthesis in isolated epithelial cells of human small intestine. Lipids *21*, 759-763.
Tabangin, M. E., Woo, J. G., and Martin, L. J. (2009). The effect of minor allele frequency on the likelihood of obtaining false positives. BMC Proc *3 Suppl 7*, S41.
Talmud, P. J. (2007). Rare APOA5 mutations--clinical consequences, metabolic and functional effects: an ENID review. Atherosclerosis *194*, 287-292.
Talmud, P. J., Hawe, E., Robertson, K., Miller, G. J., Miller, N. E., and Humphries, S. E. (2002). Genetic and environmental determinants of plasma high density lipoprotein cholesterol and apolipoprotein AI concentrations in healthy middle-aged men. Ann Hum Genet *66*, 111-124.
Tangirala, R. K., Bischoff, E. D., Joseph, S. B., Wagner, B. L., Walczak, R., Laffitte, B. A., Daige, C. L., Thomas, D., Heyman, R. A., Mangelsdorf, D. J.*, et al.* (2002). Identification of macrophage liver X receptors as inhibitors of atherosclerosis. Proc Natl Acad Sci U S A *99*, 11896-11901.
Tarugi, P., Averna, M., Di Leo, E., Cefalu, A. B., Noto, D., Magnolo, L., Cattin, L., Bertolini, S., and Calandra, S. (2007). Molecular diagnosis of hypobetalipoproteinemia: an ENID review. Atherosclerosis *195*, e19-27.
Telford, D. E., Lipson, S. M., Barrett, P. H., Sutherland, B. G., Edwards, J. Y., Aebi, J. D., Dehmlow, H., Morand, O. H., and Huff, M. W. (2005). A novel inhibitor of oxidosqualene:lanosterol cyclase inhibits very low-density lipoprotein apolipoprotein B100 (apoB100) production and enhances low-density lipoprotein apoB100 catabolism through marked reduction in hepatic cholesterol content. Arterioscler Thromb Vasc Biol *25*, 2608-2614.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J.*, et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707-713.

Thoma, R., Schulz-Gasch, T., D'Arcy, B., Benz, J., Aebi, J., Dehmlow, H., Hennig, M., Stihle, M., and Ruf, A. (2004). Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. Nature *432*, 118-122.

Thomas, G., Jacobs, K. B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A.*, et al.* (2008). Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet *40*, 310-315.

Thornton, T., and McPeek, M. S. (2007). Case-control association testing with related individuals: a more powerful quasi-likelihood score test. Am J Hum Genet *81*, 321-337.

Tigyi, G. (2010). Aiming drug discovery at lysophosphatidic acid targets. Br J Pharmacol *161*, 241-270.

Tint, G. S., Irons, M., Elias, E. R., Batta, A. K., Frieden, R., Chen, T. S., and Salen, G. (1994). Defective cholesterol biosynthesis associated with the Smith-Lemli-Opitz syndrome. N Engl J Med *330*, 107-113.

Tirosh, A., Rudich, A., Shochat, T., Tekes-Manova, D., Israeli, E., Henkin, Y., Kochba, I., and Shai, I. (2007). Changes in triglyceride levels and risk for coronary heart disease in young men. Ann Intern Med *147*, 377-385.

Togashi, K., Hara, Y., Tominaga, T., Higashi, T., Konishi, Y., Mori, Y., and Tominaga, M. (2006). TRPM2 activation by cyclic ADP-ribose at body temperature is involved in insulin secretion. Embo J *25*, 1804-1815.

Tolleshaug, H., Goldstein, J. L., Schneider, W. J., and Brown, M. S. (1982). Posttranslational processing of the LDL receptor and its genetic disruption in familial hypercholesterolemia. Cell *30*, 715-724.

Tontonoz, P., Hu, E., Devine, J., Beale, E. G., and Spiegelman, B. M. (1995). PPAR gamma 2 regulates adipose expression of the phosphoenolpyruvate carboxykinase gene. Mol Cell Biol *15*, 351-357.

Tontonoz, P., Hu, E., Graves, R. A., Budavari, A. I., and Spiegelman, B. M. (1994). mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer. Genes Dev *8*, 1224-1234.

Tordjman, K., Standley, K. N., Bernal-Mizrachi, C., Leone, T. C., Coleman, T., Kelly, D. P., and Semenkovich, C. F. (2002). PPARalpha suppresses insulin secretion and induces UCP2 in insulinoma cells. J Lipid Res *43*, 936-943.

Tremblay, A. J., Lamarche, B., Lemelin, V., Hoos, L., Benjannet, S., Seidah, N. G., Davis, H. R., Jr., and Couture, P. (2011). Atorvastatin increases intestinal expression of NPC1L1 in hyperlipidemic men. J Lipid Res *52*, 558-565.
Uchida, K., Dezaki, K., Damdindorj, B., Inada, H., Shiuchi, T., Mori, Y., Yada, T., Minokoshi, Y., and Tominaga, M. (2011). Lack of TRPM2 impaired insulin secretion and glucose metabolisms in mice. Diabetes *60*, 119-126.
Vaca, G., Vazquez, A., Magana, M. T., Ramirez, M. L., Davalos, I. P., Martinez, E., Marin, B., and Carrillo, G. (2011). Mutational analysis of the LDL receptor and APOB genes in Mexican individuals with autosomal dominant hypercholesterolemia. Atherosclerosis *218*, 391-396.
van Oostrom, A. J., Rabelink, T. J., Verseyden, C., Sijmonsma, T. P., Plokker, H. W., De Jaegere, P. P., and Cabezas, M. C. (2004). Activation of leukocytes by postprandial lipemia in healthy volunteers. Atherosclerosis *177*, 175-182.
Varbo, A., Benn, M., Tybjaerg-Hansen, A., Grande, P., and Nordestgaard, B. G. (2011). TRIB1 and GCKR polymorphisms, lipid levels, and risk of ischemic heart disease in the general population. Arterioscler Thromb Vasc Biol *31*, 451-457.
Venkatesan, S., Cullen, P., Pacy, P., Halliday, D., and Scott, J. (1993). Stable isotopes show a direct relation between VLDL apoB overproduction and serum triglyceride levels and indicate a metabolically and biochemically coherent basis for familial combined hyperlipidemia. Arterioscler Thromb *13*, 1110-1118.
Venkateswaran, A., Laffitte, B. A., Joseph, S. B., Mak, P. A., Wilpitz, D. C., Edwards, P. A., and Tontonoz, P. (2000). Control of cellular cholesterol efflux by the nuclear oxysterol receptor LXR alpha. Proc Natl Acad Sci U S A *97*, 12097-12102.
von Arnim, C. A., Kinoshita, A., Peltan, I. D., Tangredi, M. M., Herl, L., Lee, B. M., Spoelgen, R., Hshieh, T. T., Ranganathan, S., Battey, F. D.*, et al.* (2005). The low density lipoprotein receptor-related protein (LRP) is a novel beta-secretase (BACE1) substrate. J Biol Chem *280*, 17777-17785.
Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., and Rothman, N. (2004). Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst *96*, 434-442.
Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with P-values. Genet Epidemiol *33*, 79-86.
Wallace, C., Newhouse, S. J., Braund, P., Zhang, F., Tobin, M., Falchi, M., Ahmadi, K., Dobson, R. J., Marcano, A. C., Hajat, C.*, et al.* (2008). Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. Am J Hum Genet *82*, 139-149.

Wang, H., and Eckel, R. H. (2009). Lipoprotein lipase: from gene to obesity. Am J Physiol Endocrinol Metab *297*, E271-288.
Wang, J., Ban, M. R., Zou, G. Y., Cao, H., Lin, T., Kennedy, B. A., Anand, S., Yusuf, S., Huff, M. W., Pollex, R. L., and Hegele, R. A. (2008). Polygenic determinants of severe hypertriglyceridemia. Hum Mol Genet *17*, 2894-2899.
Wang, J., Cao, H., Ban, M. R., Kennedy, B. A., Zhu, S., Anand, S., Yusuf, S., Pollex, R. L., and Hegele, R. A. (2007a). Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). Arterioscler Thromb Vasc Biol *27*, 2450-2455.
Wang, J., and Hegele, R. A. (2007). Homozygous missense mutation (G56R) in glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1 (GPI-HBP1) in two siblings with fasting chylomicronemia (MIM 144650). Lipids Health Dis *6*, 23.
Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., Hakonarson, H., and Bucan, M. (2007b). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res *17*, 1665-1674.
Wang, L., Jiang, Z. G., McKnight, C. J., and Small, D. M. (2010). Interfacial properties of apolipoprotein B292-593 (B6.4-13) and B611-782 (B13-17). Insights into the structure of the lipovitellin homology region in apolipoprotein B. Biochemistry *49*, 3898-3907.
Wang, N., Lan, D., Chen, W., Matsuura, F., and Tall, A. R. (2004a). ATP-binding cassette transporters G1 and G4 mediate cellular cholesterol efflux to high-density lipoproteins. Proc Natl Acad Sci U S A *101*, 9774-9779.
Wang, Y. X., Lee, C. H., Tiep, S., Yu, R. T., Ham, J., Kang, H., and Evans, R. M. (2003). Peroxisome-proliferator-activated receptor delta activates fat metabolism to prevent obesity. Cell *113*, 159-170.
Wang, Y. X., Zhang, C. L., Yu, R. T., Cho, H. K., Nelson, M. C., Bayuga-Ocampo, C. R., Ham, J., Kang, H., and Evans, R. M. (2004b). Regulation of muscle fiber type and running endurance by PPARdelta. PLoS Biol *2*, e294.
Wang, Z., Takemori, H., Halder, S. K., Nonaka, Y., and Okamoto, M. (1999). Cloning of a novel kinase (SIK) of the SNF1/AMPK family from high salt diet-treated rat adrenal. FEBS Lett *453*, 135-139.
Watanabe, M., Osada, J., Aratani, Y., Kluckman, K., Reddick, R., Malinow, M. R., and Maeda, N. (1995). Mice deficient in cystathionine beta-synthase: animal models for mild and severe homocyst(e)inemia. Proc Natl Acad Sci U S A *92*, 1585-1589.

Waterworth, D. M., Ricketts, S. L., Song, K., Chen, L., Zhao, J. H., Ripatti, S., Aulchenko, Y. S., Zhang, W., Yuan, X., Lim, N.*, et al.* (2010). Genetic variants influencing circulating lipid levels and risk of coronary artery disease. Arterioscler Thromb Vasc Biol *30*, 2264-2276.
Weiss, L. A., Pan, L., Abney, M., and Ober, C. (2006). The sex-specific genetic architecture of quantitative traits in humans. Nat Genet *38*, 218-222.
Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661-678.
Wen, S. H., and Lu, Z. S. (2011). Factors affecting the effective number of tests in genetic association studies: a comparative study of three PCA-based methods. J Hum Genet *56*, 428-435.
Werstuck, G. H., Lentz, S. R., Dayal, S., Hossain, G. S., Sood, S. K., Shi, Y. Y., Zhou, J., Maeda, N., Krisans, S. K., Malinow, M. R., and Austin, R. C. (2001). Homocysteine-induced endoplasmic reticulum stress causes dysregulation of the cholesterol and triglyceride biosynthetic pathways. J Clin Invest *107*, 1263-1273.
Whitaker, H. C., Kote-Jarai, Z., Ross-Adams, H., Warren, A. Y., Burge, J., George, A., Bancroft, E., Jhavar, S., Leongamornlert, D., Tymrakiewicz, M.*, et al.* (2010). The rs10993994 risk allele for prostate cancer results in clinically relevant changes in microseminoprotein-beta expression in tissue and urine. PLoS One *5*, e13363.
Whittemore, A. S., and Halpern, J. (2001). Problems in the definition, interpretation, and evaluation of genetic heterogeneity. Am J Hum Genet *68*, 457-465.
Wigginton, J. E., and Abecasis, G. R. (2005). PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. Bioinformatics *21*, 3445-3447.
Wigginton, J. E., Cutler, D. J., and Abecasis, G. R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet *76*, 887-893.
Willer, C. J., Li, Y., and Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190-2191.
Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R., Heath, S. C., Timpson, N. J., Najjar, S. S., Stringham, H. M.*, et al.* (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet *40*, 161-169.
Williams, K. J. (2008). Molecular processes that handle -- and mishandle -- dietary lipids. J Clin Invest *118*, 3247-3259.
Wiltshire, S., Morris, A. P., McCarthy, M. I., and Cardon, L. R. (2005). How useful is the fine-scale mapping of complex trait linkage peaks? Evaluating the impact of

additional microsatellite genotyping on the posterior probability of linkage. Genet Epidemiol *28*, 1-10.
Wittrup, H. H., Tybjaerg-Hansen, A., and Nordestgaard, B. G. (1999). Lipoprotein lipase mutations, plasma lipids and lipoproteins, and risk of ischemic heart disease. A meta-analysis. Circulation *99*, 2901-2907.
Wu, J., Devlin, B., Ringquist, S., Trucco, M., and Roeder, K. (2010). Screen and clean: a tool for identifying interactions in genome-wide association studies. Genet Epidemiol *34*, 275-285.
Wu, Z., Martin, K. O., Javitt, N. B., and Chiang, J. Y. (1999). Structure and functions of human oxysterol 7alpha-hydroxylase cDNAs and gene CYP7B1. J Lipid Res *40*, 2195-2203.
Wu, Z., Xie, Y., Morrison, R. F., Bucher, N. L., and Farmer, S. R. (1998). PPARgamma induces the insulin-dependent glucose transporter GLUT4 in the absence of C/EBPalpha during the conversion of 3T3 fibroblasts into adipocytes. J Clin Invest *101*, 22-32.
Xu, J., Nakamura, M. T., Cho, H. P., and Clarke, S. D. (1999). Sterol regulatory element binding protein-1 expression is suppressed by dietary polyunsaturated fatty acids. A mechanism for the coordinate suppression of lipogenic genes by polyunsaturated fats. J Biol Chem *274*, 23577-23583.
Yabe, D., Brown, M. S., and Goldstein, J. L. (2002). Insig-2, a second endoplasmic reticulum protein that binds SCAP and blocks export of sterol regulatory element-binding proteins. Proc Natl Acad Sci U S A *99*, 12753-12758.
Yahagi, N., Shimano, H., Hasty, A. H., Amemiya-Kudo, M., Okazaki, H., Tamura, Y., Iizuka, Y., Shionoiri, F., Ohashi, K., Osuga, J.*, et al.* (1999). A crucial role of sterol regulatory element-binding protein-1 in the regulation of lipogenic gene expression by polyunsaturated fatty acids. J Biol Chem *274*, 35840-35844.
Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. L., and Russell, D. W. (1984). The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. Cell *39*, 27-38.
Yang, T., Espenshade, P. J., Wright, M. E., Yabe, D., Gong, Y., Aebersold, R., Goldstein, J. L., and Brown, M. S. (2002). Crucial step in cholesterol homeostasis: sterols promote binding of SCAP to INSIG-1, a membrane protein that facilitates retention of SREBPs in ER. Cell *110*, 489-500.
Yang, T. P., Beazley, C., Montgomery, S. B., Dimas, A. S., Gutierrez-Arcelus, M., Stranger, B. E., Deloukas, P., and Dermitzakis, E. T. (2010). Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics *26*, 2474-2476.

Ye, S., Dunleavey, L., Bannister, W., Day, L. B., Tapper, W., Collins, A. R., Day, I. N., and Simpson, I. (2003). Independent effects of the -219 G>T and epsilon 2/ epsilon 3/ epsilon 4 polymorphisms in the apolipoprotein E gene on coronary artery disease: the Southampton Atherosclerosis Study. Eur J Hum Genet *11*, 437-443.

Yoon, Y. S., Seo, W. Y., Lee, M. W., Kim, S. T., and Koo, S. H. (2009). Salt-inducible kinase regulates hepatic lipogenesis by controlling SREBP-1c phosphorylation. J Biol Chem *284*, 10446-10452.

Yu, L., Li-Hawkins, J., Hammer, R. E., Berge, K. E., Horton, J. D., Cohen, J. C., and Hobbs, H. H. (2002). Overexpression of ABCG5 and ABCG8 promotes biliary cholesterol secretion and reduces fractional absorption of dietary cholesterol. J Clin Invest *110*, 671-680.

Yu, Y., Wyszynski, D. F., Waterworth, D. M., Wilton, S. D., Barter, P. J., Kesaniemi, Y. A., Mahley, R. W., McPherson, R., Waeber, G., Bersot, T. P.*, et al.* (2005). Multiple QTLs influencing triglyceride and HDL and total cholesterol levels identified in families with atherogenic dyslipidemia. J Lipid Res *46*, 2202-2213.

Zavattari, P., Deidda, E., Whalen, M., Lampis, R., Mulargia, A., Loddo, M., Eaves, I., Mastio, G., Todd, J. A., and Cucca, F. (2000). Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. Hum Mol Genet *9*, 2947-2957.

Zelcer, N., Hong, C., Boyadjian, R., and Tontonoz, P. (2009). LXR regulates cholesterol uptake through Idol-dependent ubiquitination of the LDL receptor. Science *325*, 100-104.

Zemunik, T., Boban, M., Lauc, G., Jankovic, S., Rotim, K., Vatavuk, Z., Bencic, G., Dogas, Z., Boraska, V., Torlak, V.*, et al.* (2009). Genome-wide association study of biochemical traits in Korcula Island, Croatia. Croat Med J *50*, 23-33.

Zeng, B. J., Mortimer, B. C., Martins, I. J., Seydel, U., and Redgrave, T. G. (1998). Chylomicron remnant uptake is regulated by the expression and function of heparan sulfate proteoglycan in hepatocytes. J Lipid Res *39*, 845-860.

Zhan, C., Fedorov, E. V., Shi, W., Ramagopal, U. A., Thirumuruhan, R., Manjasetty, B. A., Almo, S. C., Fiser, A., Chance, M. R., and Fedorov, A. A. (2005). The ybeY protein from Escherichia coli is a metalloprotein. Acta Crystallogr Sect F Struct Biol Cryst Commun *61*, 959-963.

Zhang, H., Henderson, H., Gagne, S. E., Clee, S. M., Miao, L., Liu, G., and Hayden, M. R. (1996). Common sequence variants of lipoprotein lipase: standardized studies of in vitro expression and catalytic function. Biochim Biophys Acta *1302*, 159-166.

Zhang, P., Xiang, N., Chen, Y., Sliwerska, E., McInnis, M. G., Burmeister, M., and Zollner, S. (2010a). Family-based association analysis to finemap bipolar linkage peak on chromosome 8q24 using 2,500 genotyped SNPs and 15,000 imputed SNPs. Bipolar Disord *12*, 786-792.
Zhang, X., Huang, S., Zou, F., and Wang, W. (2010b). TEAM: efficient two-locus epistasis tests in human genome-wide association study. Bioinformatics *26*, i217-227.
Zhang, X., Pan, F., Xie, Y., Zou, F., and Wang, W. (2010c). COE: a general approach for efficient genome-wide two-locus epistasis test in disease association study. J Comput Biol *17*, 401-415.
Zhao, Z., and Michaely, P. (2009). The role of calcium in lipoprotein release by the low-density lipoprotein receptor. Biochemistry *48*, 7313-7324.
Zheng, J., Li, Y., Abecasis, G. R., and Scheet, P. (2011). A comparison of approaches to account for uncertainty in analysis of imputed genotypes. Genet Epidemiol *35*, 102-110.
Zhu, G., Warren, L., Aponte, J., Gulsvik, A., Bakke, P., Anderson, W. H., Lomas, D. A., Silverman, E. K., and Pillai, S. G. (2007a). The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. Am J Respir Crit Care Med *176*, 167-173.
Zhu, H., Tucker, H. M., Grear, K. E., Simpson, J. F., Manning, A. K., Cupples, L. A., and Estus, S. (2007b). A common polymorphism decreases low-density lipoprotein receptor exon 12 splicing efficiency and associates with increased cholesterol. Hum Mol Genet *16*, 1765-1772.
Zinovieva, E., Bourgain, C., Kadi, A., Letourneur, F., Izac, B., Said-Nahal, R., Lebrun, N., Cagnard, N., Vigier, A., Jacques, S.*, et al.* (2009). Comprehensive linkage and association analyses identify haplotype, near to the TNFSF15 gene, significantly associated with spondyloarthritis. PLoS Genet *5*, e1000528.
Zöllner, S., and Teslovich, T. M. (2009). Using GWAS data to identify copy number variants contributing to common complex diseases. Statistical Science *24*, 530-546.