The Genetic Architecture of Familial Hypercholesterolaemia

Marta Futema

University College London

A thesis submitted in accordance with the regulations of the University College London for the degree of Doctor of Philosophy

Centre for Cardiovascular Genetics
UCL Genetics Institute

“Nothing in life is to be feared,

it is only to be understood.

Now is the time to understand more,

so that we may fear less.”

Maria Skłodowska-Curie

Declaration

I, Marta Futema confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Regarding the Oxford FH cohort, I would like to acknowledge Prof. Andrew Neil and Dr. Mary Seed on behalf of the Simon Broome Familial Hypercholesterolaemia Register Group for providing the patients samples, the clinical information and expertise.

Regarding the cohort of FH patients of South-Asian origin, samples were provided by Mr Darren Harvey, Dr. Anjly Jain and Dr. Devi Nair (Royal Free Hospital, London), Sister Emma McGowan and Dr. Michael Khan (University Hospitals Coventry and Warwickshire NHS Trust), Dr. Julian Barth (Leeds General Infirmary), and Dr. Manjeet Mehta (Kokilaben Dhirubhai Ambani Hospital and Medical Research Institute).

Dr. Emma Ashton and Dr. Nicholas Lench from the North East Thames Regional Genetics Service (Great Ormond Street Hospital) kindly provided their advice and equipment for MLPA experiments.

The replication cohorts for LDL-C gene score analysis were provided by Albert Wiegman and Joep Defesche (Academic Medical Center, Amsterdam, the Netherlands), Rob Hegele and Mary Aderayo Bamimore (University of Western Ontario, London, Canada), and Dr. Euridiki Drogari and Dr. Vasiliki Mollaki (University of Athens Medical School, “Aghia Sophia” Children's Hospital, Athens, Greece).

The UK10K consortium, including the project production team, was responsible for the library preparation, sequence capture and bioinformatics of the whole exome sequencing FH cohort and the controls. The full list of the members can be found at: http://www.uk10k.org/consortium.html.

Dr. Vincent Plagnol was involved in the whole exome sequencing data processing and analysis.

Ms KaWah Li did some of the 12 SNPs genotyping for the LDL-C gene score analysis.

Collaborators who provided FH samples for whole exome sequencing include: Dr. Stefano Bertolini (University of Genoa, Italy), Dr. Sebastiano Calandra (University of Modena and Reggio Emilia, Modena, Italy), Dr. Olivier S Descamps (Centre de Recherche Médicale de Jolimont, Haine St-Paul, Belgium), Dr. Colin Graham (Belfast Health and Social Care Trust / City Hospital, Belfast, Northern Ireland, UK), Dr. Fredrik Karpe (University of Oxford, Churchill Hospital, Oxford, UK), Dr. Eran Leitersdorf (Hadassah Hebrew University Medical Centre, Jerusalem, Israel), Dr. Handrean Soran (Central Manchester University Hospitals Foundation Trust, Manchester, UK), Dr. Frank Van Bockxmeer (Royal Perth Hospital, Wellington St, Perth, Australia).

Acknowledgements

Firstly, I would like to acknowledge all the patients who participated in this study. It is through their willingness to donate their samples for research that new medical and scientific knowledge has been gained.

I would like to thank my teachers, whose courage, commitment and passion have inspired me over the years. I have been extremely fortunate to learn from you. Especially I would like to express my sincere thanks to my supervisor Professor Steve Humphries, whose guidance throughout my project helped me to bring out the best in myself. I would like to also convey my gratitude to Vincent Plagnol and Ros Whittall, who generously shared their time and expertise. I extend my thanks to all members of the Cardiovascular Genetics lab, who make a friendly and exciting team of scientists always keen to share their knowledge.

I am extremely grateful to the Medical Research Council and Gen-Probe Incorporated, who have supported me financially during the project. The research carried out in this thesis was also partially founded by the British Heart Foundation and the Wellcome Trust (the UK10K project).

I thank my family and friends for their constant support. Special thanks to my parents who managed to provide stability and support throughout my life. Your love and care gave me strength and courage in fulfilling my ambitions. Many thanks to Artus, my husband, who was always there for me and whose positive attitude was extremely motivating in difficult moments.

Publications resulting from the work in this thesis:

1. **Futema M**, Whittall RA, Kiley A, Steel LK, Cooper JA, Badmus E, Leigh SE, Karpe F, Neil HAon behalf of the Simon Broome Register Group, and Humphries SE. Analysis of the frequency and spectrum of mutations recognised to cause familial hypercholesterolaemia in routine clinical practice in a UK specialist hospital lipid clinic. ***Atherosclerosis***. 2013;229(1):161-8.
2. **Futema M**, Plagnol V, Whittall RA, Neil HA on behalf of the Simon Broome Register Group, Humphries SE and the UK10K Consortium. Use of targeted exome sequencing as a diagnostic tool for Familial Hypercholesterolaemia. ***J Med Genet***. 2012 Oct;49(10):644-9.

Other publications produced during the thesis:

1. Talmud PJ, Shah S, Whittall R, **Futema M**, Howard P, Cooper JA, Harrison SC, Li K, Drenos F, Karpe F, Neil HA, Descamps OS, Langenberg C, Lench N, Kivimaki M, Whittaker J, Hingorani AD, Kumari M, Humphries SE. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. ***Lancet***. 2013 Apr 13;381(9874):1293-301.
2. Pećin I, Whittall R, **Futema M**, Sertić J, Reiner Z, Leigh SE, Humphries SE. Mutation detection in Croatian patients with familial hypercholesterolemia. ***Ann Hum Genet***. 2013 Jan;77(1):22-30.

Manuscripts can be found at the back of the thesis.

Abstract

Familial Hypercholesterolaemia (FH) is a common autosomal dominant disorder of the defective plasma clearance of LDL-cholesterol (LDL-C). Mutations in three genes, *LDLR*/*APOB*/*PCSK9*, can be detected in 60-90% of definite FH patients. DNA-based testing for FH mutations has important clinical utility and is recommended by the UK and European guidelines to identify affected relatives. This thesis aimed to determine the frequency and spectrum of FH mutations in two independent cohorts of FH patients (from one Oxford lipid clinic, and of Indian background). The FH mutation spectrum was shown to be highly heterogeneous and the mutation detection rate was significantly dependent on the pre-treatment total cholesterol and triglyceride levels. This project also validated the findings that a proportion of clinically diagnosed FH patients have a polygenic cause of hypercholesterolaemia due to an accumulation of common mild LDL-C-raising alleles by analysing LDL-C gene score in 88 mutation negative and 21 mutation positive FH patients, and by replicating the results in further 231 FH patients. A high-throughput DNA sequencing method was assessed as a novel diagnostic tool for detection of FH mutations, and compared it with the currently used methods. This highlighted the need for updating the current FH mutation screening methods as well as the need for more efficient bioinformatics for the next generation sequencing data analysis. Lastly, whole exome sequencing of 125 definite FH patients with no mutations detected in known genes was performed to identify novel monogenic causes of FH. Variants in two genes, *CH25H* and *INSIG2*, were identified as potential novel FH mutations. Overall, the results of this thesis demonstrate the heterogeneous FH aetiology and help to understand the genetic architecture of the disease.

Contents

List of Figures

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

CHAPTER SEVEN

List of Tables

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

CHAPTER SEVEN

Abbreviations

ABCA1	ATP-binding cassette transporter 1
ANOVA	Analysis of Variance
Apo	Apolipoprotein
ARMS	Amplification Refractory Mutation System
BHF	British Heart Foundation
CAD	Coronary Artery Disease
CE	Cholesterol Esters
CETP	Cholesterylester Transfer Protein
CH25H	25-cholesterol hydroxylase
CHD	Coronary Heart Disease
CNV	Copy Number Variant
CVD	Cardiovascular Disease
DFH	Definite Familial Hypercholesterolaemia
DLCN	Dutch Lipid Clinic Network
DNA	Deoxyribose Nucleic Acid
FDB	Familial Defective Apolipoprotein B
FFA	Free Fatty Acids
FH	Familial Hypercholesterolaemia
GLGC	Global Lipids Genetic Consortium
GWAS	Genome Wide Association Study
HDL	High Density Lipoprotein
HMGCR	3-Hydroxy-3-Methylglutaryl-CoA Reductase
HPS2-THRIVE	Heart Protection Study 2-Treatment of HDL to Reduce the Incidence of Vascular Events
HRM	High Resolution Melting
IDL	Intermediate Density Lipoprotein
IGV	Integrative Genomic Viewer
INSIG2	Insulin Induced gene 2
LCAT	Lecithin-cholesterol Acyltransferase

LDL	Low Density Lipoprotein
LDL-C	Low Density Lipoprotein Cholesterol
LDLR	Low Density Lipoprotein Receptor
LDLRAP1	Low Density Lipoprotein Receptor Apadptor Protein 1
LPL	Lipoprotein Lipase
LRP	LDL-receptor Related Protein
M	Molar
MADGE	Microplate Array Diagonal Gel Electrophoresis
MEDPED	Make Early Diagnoses and Prevent Early Deaths study
MI	Myocardial Infarction
MLPA	Multiplex Ligation-dependent Probe Amplification
mM	Milimolar
nM	Nanomolar
NTC	No Template Control
ox-LDL	Oxydised Low Density Lipoprotein
PCR	Polymerase Chain Reaction
PCSK9	Proprotein Convertase Subtilisin/Kexin type 9
PFH	Possible Familial Hypercholesterolaemia
RFLP	Restriction Fragment Polymorphism
RNA	Ribonucleic Acid
SB	Simon Broome
SBBHF	Simon Broome British Heart Fundation study
SCAP	Sterol Regulatory Element Binding Protein Cleavage-Activating Protein
SD	Standard Deviation
SMR	Standardized Mortality Ratio
SNP	Single Nucleotide Polymorphism
SR-B1	Scavenger Receptor Class B Type 1
SREBP	Sterol Regulatory Element Binding Protein
TG	Triglyceride
Tx	Tendon Xanthoma
VLDL	Very Low Density Lipoprotein
WHII	Whitehall Phase II Study

WHO	World Health Organisation

I. CHAPTER ONE - General Introduction

1.1. Lipid metabolism

1.1.1. Lipids and Lipoproteins

Lipids are water-insoluble, structurally diverse, organic molecules characterized by a hydrocarbon backbone, which can be utilized or synthesized from the dietary fats. Several lipid subgroups can be distinguished, depending on their chemical composition. Triglyceride (TG) is an ester derived from glycerol and three fatty acids, stored in the adipose tissue, which is important for energy supply. TG, commonly named as fat, is the source of free fatty acids, which are the major substrate for the energy production (saturated fatty acids), precursors for tissue hormones (unsaturated fatty acids) (Bhathena 2006), or act as ligands for transcription factors, influencing gene expression (Lin et al. 1999, Hostetler et al. 2005, Hostetler et al. 2006). Cholesterol, the crucial component of cellular membranes, is the major sterol in human body, which also functions as a precursor for steroid hormones. Cholesterol regulates its own metabolism and biosythesis by inducing the activity of sterol regulatory element-binding protein (SREBP). Other lipid subgroups include lipophilic vitamins (vitamins A, D, E, and K), which also function as ligands for transcription factors and are important for regulation of genes involved in energy metabolism (McEwan 2009), and phospho- and sphingolipids, which are structural membrane components.

In order to be transported in the plasma, hydrophobic lipids are assembled in lipoprotein particles. Lipoproteins have a hydrophobic inner core containing

TG and cholesterol esters (CE), covered by a membrane-like structure formed by hydrophilic phospholipids, unesterified cholesterol and proteins called apolipoproteins. The role of apolipoproteins is to solubilize and stabilize lipids in the aqueous environment of plasma. Lipoproteins are divided into five classes, according to their origin and density: chylomicrons, very low density lipoproteins (VLDL), intermediate density lipoproteins (IDL), low density lipoproteins (LDL) and high density lipoproteins (HDL). The plasma lipoproteins differ in their composition, as shown in Table 1.

Lipoprotein trafficking can be divided in three pathways: the exogenous pathway, during which dietary cholesterol and fatty acids are absorbed, the endogenous pathway, when lipoproteins synthesized in the liver are distributed to peripheral tissues, and the reverse cholesterol pathway, which mediates transport of excess cholesterol from the periphery to the liver, summarized in Figure 1.

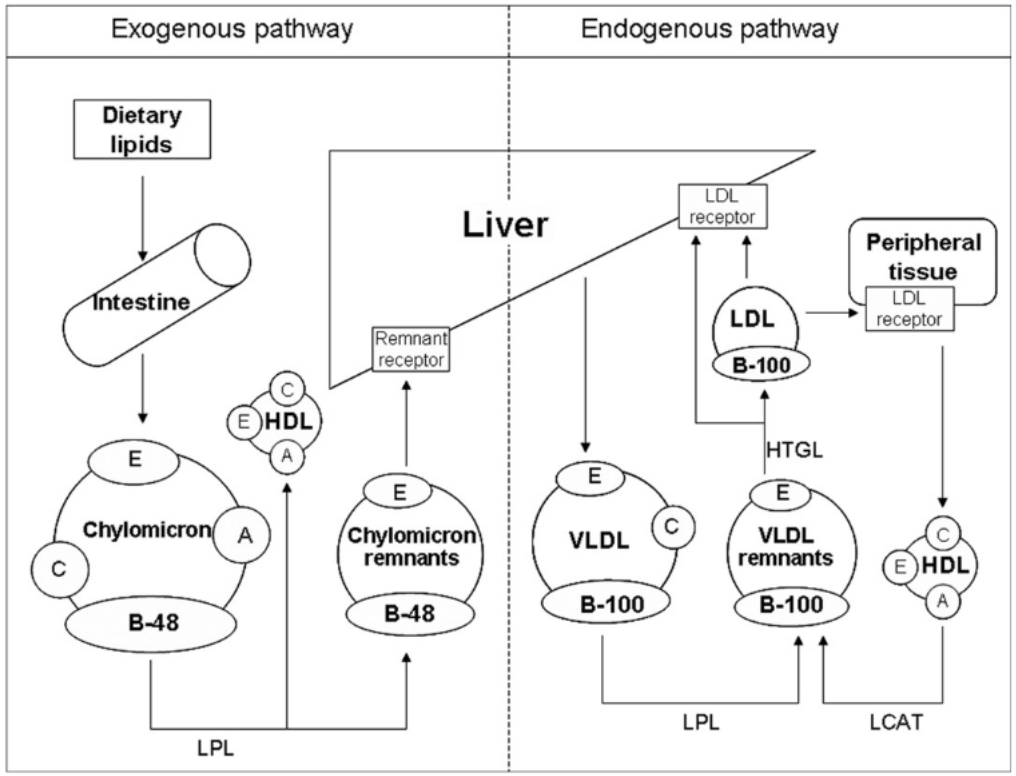
Table 1

Composition of the plasma lipoproteins (% total) (taken from (Feher and Richmond 1997)).

	Chylomicrons	**VLDL**	**IDL**	**LDL**	**HDL**
protein	2	10	18	25	55
TG	85	50	26	10	4
cholesterol	1	7	12	8	2
cholesterol ester	3	13	22	37	15
phospholipid	9	20	22	20	24

Figure 1

Schematic representation of lipid metabolism. E= Apolipoprotein E, C= Apolipoprotein C, A= Apolipoprotein A, B-48=Apolipoprotein B-48, B-100=Apolipoprotein B-100, LPL=lipoprotein lipase, LCAT=lecithin-cholesterol acyltransferase, HTGL=hepatic triglyceride lipase (adapted from (Goldstein et al. 1983)).



1.1.2. Exogenous pathway

Exogenous transport of cholesterol and fats begins with the absorption of dietary lipids in the intestine, during which TGs are hydrolyzed to monoacylglycerols and fatty acids. After lipid diffusion across the microvillus membrane, cholesterol and fatty acids are re-esterified and packaged with apolipoprotein B-48 (ApoB-48), several apolipoproteins A (ApoA-I, ApoA-II, ApoA-IV) and other lipids to form chylomicrons. After their synthesis in the intestinal wall the chylomicrons enter the blood circulation via the lymphatic system where they acquire apolipoproteins E and C (ApoE, ApoC-I, ApoC-II, ApoC-III) from HDL-C. ApoC-II activates lipoprotein lipase (LPL), which rapidly hydrolyses TGs to produce free fatty acids (FFAs). The FFAs leave chylomicrons and are transported to different tissues such as muscle and adipose tissue. The remaining chylomicrons, much reduced in size, contain the core remnants rich in cholesterol esters, ApoB-48 and ApoE, and the surface remnants, mainly composed of phospholipids and Apo-C. The latter is transferred to HDL particles, while the core remnants are taken up by the liver via the LDL-receptor, an ApoB/E specific receptor, and via the LDL-receptor related protein (LRP) (Choi et al. 1991, Rohlmann et al. 1998).

1.1.3. Endogenous pathway

To maintain and regulate the energy supply for tissues, the liver is able to synthesize TGs from fatty acids produced from acetate of dietary carbohydrates (Gibbons 1990). TGs, which can also be derived from the

chylomicron remnants produced during the exogenous pathway, are packaged together with cholesterol, cholesterol ester, phospholipid, apolipoproteins B-100 and A to form VLDL. Newly assembled VLDL particles enter the circulation where, similarly to the chylomicrons, acquire more apolipoprotein (ApoC and ApoE) from HDL, which activates LPL to hydrolyze TGs and to release FFA. The cholesterylester transfer protein (CETP) controls the exchange of VLDL lipids with mature HDL, which enriches VLDL in cholesterol esters and forms the intermediate density lipoprotein (IDL). Further loss of TGs and phospholipids from IDL particles (mainly transferred to HDL) is accompanied by a loss of apoliporoteins C and E, which leads to an increase in the lipid density of the particles. Apolipoprotein B-100 (ApoB-100) becomes the major carrier of the assembled LDL particles, which acts as a ligand for the LDL-receptors located on the surface of the liver cells or any other cells requiring cholesterol supply. Once bound to the receptor, LDL is internalized and undergoes lysosomal hydrolysis.

1.1.4. Reverse cholesterol transport

Excess cholesterol in peripheral tissues can be transferred back to the liver in high density lipoprotein (HDL) particles, in the metabolic pathway of reverse cholesterol transport (Glomset 1968). Apolipoprotein A-I produced by the liver circulates as pre-beta HDL particles to peripheral cells, where it picks up cholesterol and phospholipids. The ATP-binding cassette (ABC) transporter 1 (ABCA1), described as the gatekeeper of reverse cholesterol transport pathway (Oram and Lawn 2001), transports cellular cholesterol and phospholipids to the cell surface where they can bind apolipoproteins (Oram

and Yokoyama 1996). While in the circulation, HDL accumulates more lipids, which reduces its density. It can also acquire ApoE, which enables it to bind to the ApoE receptors on hepatocytes. Lecithin-cholesterol acyltransferase (LCAT) is an enzyme that esterifies the cholesterol into cholesterol esters that accumulate in HDL particle (Roheim 1986). Cholesterol esters can be exchanged for TG from TG-rich lipoproteins by CETP or transferred to the liver by the scavenger receptor SR-B1 or ABC transporters (Acton et al. 1996). Once in the liver cholesterol can be recycled to form lipoproteins, used for bile acid production or secreted directly into the bile.

1.1.5. Cholesterol homeostasis

The concentration of cellular cholesterol is controlled by the regulated transport of the ER membrane-bound transcription factor, called the sterol regulatory element-binding protein (SREBP), to the Golgi complex (Radhakrishnan et al. 2007). Once in the Golgi, SREBP is proteolytically processed to release its active fragments, which are imported to the nucleus to switch on the transcription of 3-Hydroxy-3-Methylglutaryl-CoA Reductase (*HMGCR)* and Low Density Lipoprotein Receptor (*LDLR)*, and other sterol-regulated genes (Goldstein et al. 2006). *HMGCR* encodes the hydroxymethyl-glutaryl CoA reductase, the rate-limiting enzyme of cellular cholesterol biosynthesis, whereas *LDLR* codes for the LDL-receptor responsible for the uptake of LDL-C from plasma. Thus a regulation of the expression of SREBP's target genes enables the cell to maintain a constant level of membrane cholesterol, which is important especially during periods of rapid cell growth. The activity of SREBP is regulated by the end product,

cholesterol, and by oxysterols, which are derivatives of cholesterol (Kandutsch et al. 1977, Brown and Goldstein 1974b, Kandutsch et al. 1978). Cholesterol and oxysterols can suppress cholesterol biosynthesis and cholesterol uptake by binding to Insulin-Induced Gene (INSIG) proteins or to Sterol Regulatory Element Binding Protein Cleavage-Activating Protein (SCAP, cholesterol only), the escort protein of SREBP (Radhakrishnan et al. 2007). The excess of cholesterol or oxysterols initiates the formation of an INSIG/SCAP complex, which blocks the transport of SREBPs to the Golgi, hence *HMGCR* and *LDLR* are no longer transcribed (Goldstein et al. 2006). Inhibition of SREBP prevents toxic accumulation of cholesterol.

1.2. Atherosclerosis

Atherosclerosis derives its name from Greek words ‘sclerosis’ (hardening) and ‘athere‘ (gruel). The condition remains to be the leading cause of mortality and morbidity in developed countries (World Health Organization, http://www.who.int/gho/publications/world_health_statistics/2012/en/). The process of atherogenesis is driven by many environmental and genetic factors. One of the major risk factors is hyperlipidemia, especially increased levels of LDL-C in the plasma. Retention of LDL-C in the vessel wall leads to oxidation, which promotes the recruitment of monocytes and lymphocytes and an increased production of cytokines and growth factors. Monocytes and lymphocytes convert to macrophages and take up the oxidized LDL, which leads to formation of foam cells (Figure 2(b)). Foam cells form a lesion of yellow appearance, also called the fatty streak, which activates inflammation by recruiting further inflammatory cells. The lesion’s growth is caused by the

infiltration and proliferation of smooth muscle cells (SMCs) and synthesis of extracellular matrix containing collagen, elastin and proteoglycans. Some of the macrophages go through apoptosis and die, releasing more lipids to the core of the lesion forming a so-called necrotic core (Figure 2(c)). Rupture of the atherosclerotic plaque causes release of its core components to the lumen, which triggers thrombosis and occlusion of the artery (Figure 2(d)), which can lead to myocardial infarction resulting in sudden death.

CVD risk reduction by LDL-C-lowering therapy is well established (Baigent et al. 2005). In contrast, HDL-C has been thought to be atheroprotective (Figure 3), thus low levels of HDL-C were considered to be the risk factor for atherosclerosis (Di Angelantonio et al. 2009). However, results of a recent mendelian randomization study demonstrated that a gene score including 14 genetic variants associated with high levels of HDL-C was not associated with reduced risk of myocardial infarction, suggesting that HDL-C is not a causative factor for CHD (Voight et al. 2012). In addition, the Heart Protection Study 2-Treatment of HDL to Reduce the Incidence of Vascular Events (HPS2-THRIVE) showed that raising HDL-C with niacin treatment has no benefits or can be even harmful for patients at risk of CHD (http://www.sciencedaily.com-/releases/2013/03/130311101827.htm accessed on 20 Aug 2013). Other CVD risk factors can be divided into the modifiable risk factors, such as high blood pressure, smoking, lack of physical activity, diet, obesity, diabetes mellitus (including abnormal lipids as mentioned before), and the non-modifiable risk factors, including age, gender and genetic predisposition, which is reflected by the family history in majority of the affected patients (Boer et al. 1999). Other risk factors with a strong genetic

component include homocysteine (Rodgers and Conn 1990), haemostatic factors (Kannel 2005), lipoprotein(a) (Djurovic and Berg 1997) and systemic inflammation (Ross 1999). In the majority of patients, the multistep process of atherosclerosis development is likely to be caused by a combination of different risk factors, as well as gene-gene and gene-environment interactions (Do et al. 2011), however currently the robust statistical evidence to support this from single studies and their subsequent replication in larger studies is lacking.

Figure 2

Stages in the development of atherosclerotic lesion (taken from (Libby et al. 2011)) and the normal artery (a).

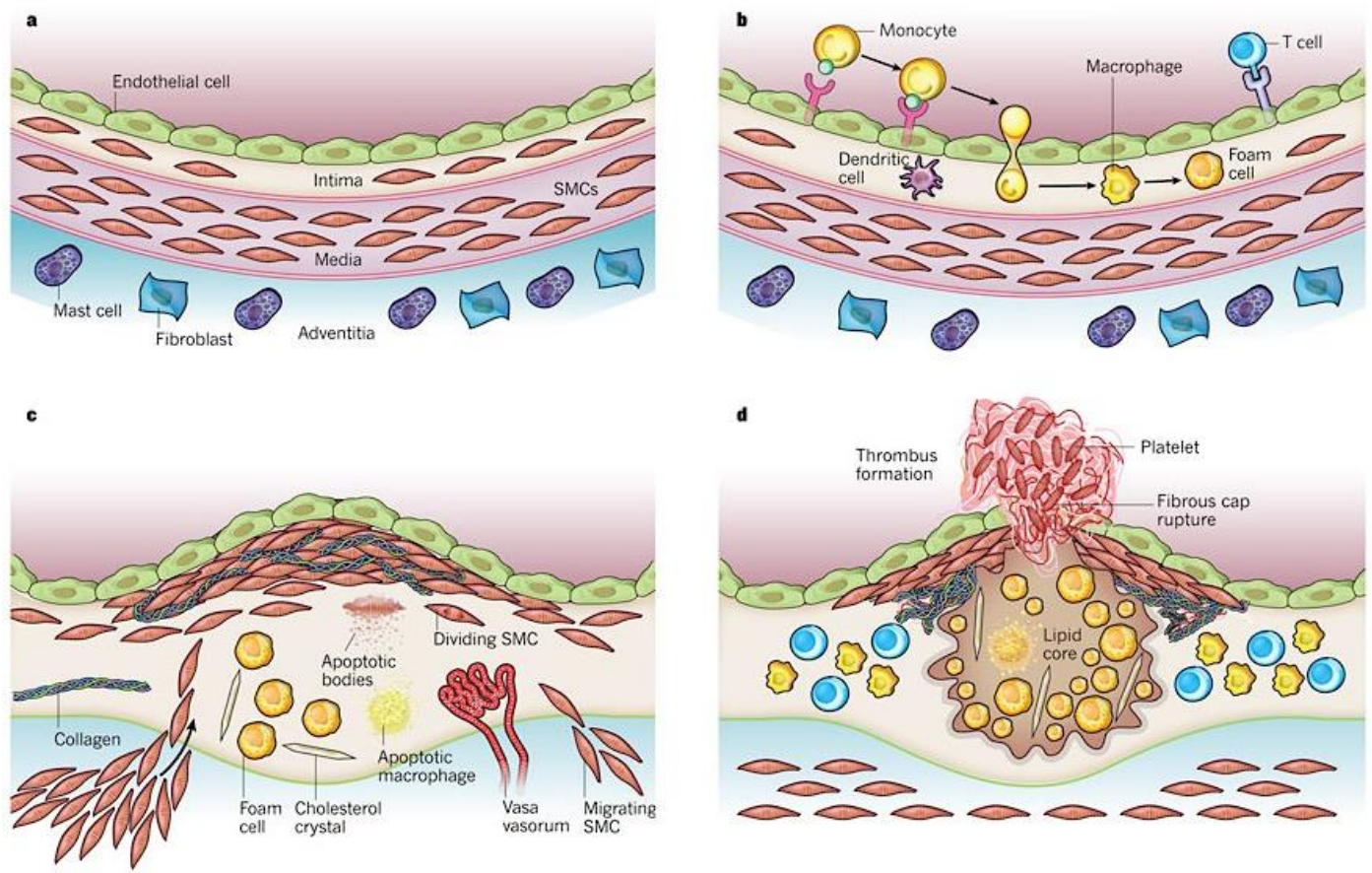
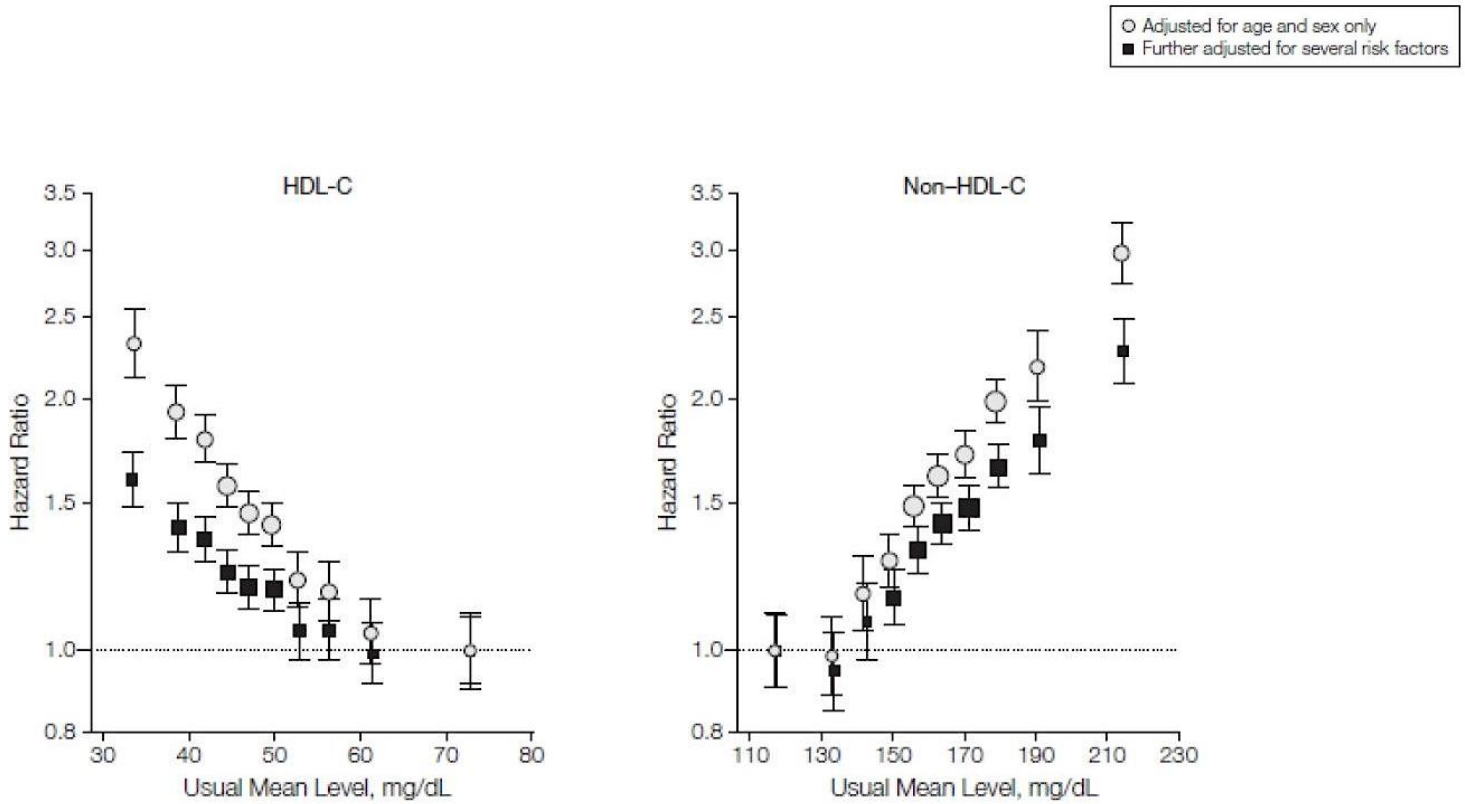


Figure 3

Hazard ratios for CHD across quantiles of HDL-C and non-HDL-C plasma levels, based on 302,403 participants (including 12,785 cases). Adapted from (Di Angelantonio et al. 2009).



Familial hypercholesterolaemia (FH) is a monogenic cause of atherosclerosis, which starts at a very early age, and eventually results in CHD. It was demonstrated that, if untreated, 30% of women and 50% of men patients with heterozygous FH will develop CHD by the age of 55 (Slack 1969). FH patients have five to eight times higher than average risk of developing CHD (Marks et al. 2003). Homozygous FH patients develop CHD very early in life and die before the age 20 if no treatment is provided.

1.3. Overview of the molecular pathogenesis of FH

Early work by the 1985 Nobel Prize laureates, Michael S. Brown and Joseph L. Goldstein, significantly broadened the understanding of lipid metabolism and the effect of its inborn errors. Extensive studies of the receptor-mediated endocytosis of low-density lipoprotein particles (LDL-C) explained the basis of cholesterol regulation within a cell system (Goldstein and Brown 1974). Because two thirds of human plasma cholesterol is contained within LDL particles, formation of atherosclerotic plaques was correctly linked with increased levels of LDL-C. Therefore answering the question: 'What determines blood LDL-C levels in an individual?' was crucial in order to prevent atherosclerosis and premature coronary heart disease (CHD). Brown and Goldstein's biochemical and structural studies resulted in the description of the sequential steps of LDL-C uptake by a cell surface receptor, called the LDL-receptor (Brown and Goldstein 1979). They discovered that LDL-receptors are located in clathrin-coated pits. Binding of the LDL-C particle to its receptor initiates internalisation of LDL-LDL-R complexes within coated vesicles, which rapidly fuse with lysosomes. This is followed by the

lysosomal hydrolysis of LDL-C, resulting in the release of unesterified cholesterol to the cytoplasm while the receptor recycles to the surface. Understanding of the LDL-C pathway became the foundation for the investigation of FH pathogenesis, which in 1974 led to identification of the first FH mutations at the *LDLR* locus (Brown and Goldstein 1974a).

Mutation in one of the two inherited genes, the heterozygous form of FH, leads to an approximately two-fold increase in plasma cholesterol level. However, the homozygous FH is characterised by much greater cholesterol, even four- to five-fold the normal level. This occurs when an FH mutation is passed to the offspring from both his/her mother as well as from the father. An individual with two identical mutations, the true homozygote, or with two different mutations (the compound heterozygous FH) suffers extreme lipid levels from a very young age and require aggressive lipid-lowering treatment to prolong his/hers life.

1.4. Clinical diagnosis of FH

There are currently three FH diagnostic algorithms used in lipid clinics around the world. In the UK, the Simon Broome Heart Research Trust, which was originally founded by Mrs. Katherine Broome, widow of Simon Broome, who died prematurely of heart disease, created the Simon Broome register of patients suffering from FH. On the basis of the Simon Broome register, its Steering Committee ('Risk of fatal coronary heart disease in familial hypercholesterolaemia. Scientific Steering Committee on behalf of the Simon Broome Register Group' 1991) developed the UK's clinical diagnostic criteria for FH. The criteria distinguish two types of the disease: definite FH (DFH),

and possible FH (PFH). Both DFH and PFH patients should have total cholesterol (TC) greater than 7.5 mmol/l, and/or LDL-cholesterol (LDL-C) greater than 4.9 mmol/l (Table 2). The criteria also account for the difference in cholesterol levels between adults and children, thus the pre-treatment cholesterol cutoffs are appropriately lowered for patients under the age of 16 (Table 2). Positive family history of myocardial infarction (MI) is also one of the criteria, although the specificity of this criterion depends on the prevalence to MI in the studied population. In addition, definite FH patients or their relatives (first or second degree) have observable lipid deposition in the tendons forming tendon xanthomata, as shown in Figure 4, although these are becoming increasingly uncommon due to improvements in early diagnosis and treatment. Presence of a variant known or predicted to affect function in one of the known FH genes (discussed in section 1.11) also classifies a patient as DFH (Marks et al. 2003). Approximately two thirds of all FH patients in the UK's lipid clinics fall under the diagnosis of possible FH (Hadfield et al. 2008).

Table 2

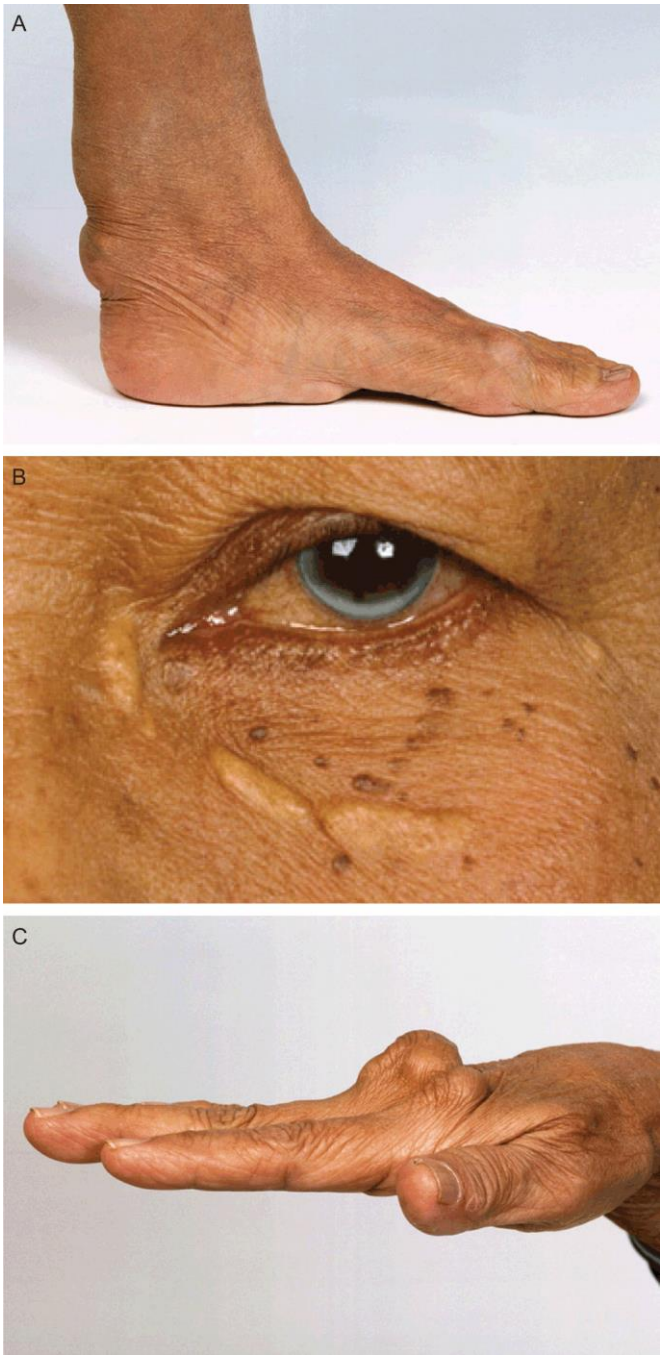
Simon Broome diagnostic criteria for FH (Marks et al. 2003).

Criteria	
A	Plasma cholesterol measurements of either: TC >7.5 mmol/l (adult) or >6.7 mmol/l (child <16 years) LDL-C >4.9 mmol/l (adult) or >4.0 mmol/l (child <16 years)
B	Presence of tendon xanthomata in patient or in a relative (first or second degree)
C	DNA-based evidence of a mutation in an FH gene
D	Family history of MI in a second degree relative <50 years or in a first degree relative <60 years of age
E	Family history of plasma TC >7.5 mmol/l ina first or second degree relative

Diagnosis	**Criteria required**
Definite FH	A + B or C
Possible FH	A +D or A + E

Figure 4

Tendon xanthoma shown in A. is the clinical feature of DFH; xanthelasmas and arcus corneae (B) and xanthomata of the extensor tendons (C) are also commonly observed in older FH patients, however they are not diagnostic features for FH (all taken from (Liyanage et al. 2011))



A similar FH diagnostic system has been developed by the Dutch Lipid Clinic Network (DLCN). The algorithm uses numerical scores and classifies patients into three sub-groups (Marks et al. 2003). The definite FH patients have a score greater than 8, the probable FH - between 6 and 8, and the possible FH - between 3 and 5 points. No diagnosis is made if the score is below 3. Similarly to the Simon Broome FH criteria, DLCNC score is based on a family history of premature CHD, LDL-C levels, presence of tendon xanthoma/ corneal arcus, and DNA analysis, as summarized in Table 3. A possible advantage of the DLCN score over the Simon Broome FH criteria is that the DLCN does not require tendon xanthoma to be present for the diagnosis of DFH patients (if a mutation has not been identified). As shown in Table 3, a definite FH diagnosis can be made on the basis of an extreme LDL-C measurement only.

Table 3

The FH diagnostic criteria, as defined by the Dutch Lipid Clinic Network. The DLCN score distinguishes three forms of FH: possible (3-5 points), probable (6-8 points), and definite (>8 points)

	Dutch lipid clinic network diagnosis of FH	Points
	Family history	
A	First degree relative with known premature (<55 years men; <60 years women) coronary and vascular disease	1
B	First-degree relative with known LDL-cholesterol >95th percentile and/or2	
A	First degree relative with tendon xanthomata and/or arcus cornealis	2
B	Children below 18 years with LDL cholesterol >95th percentile	
	Clinical history	
A	Patient has premature (<55 years men; <60 years women) coronary artery disease	2
B	Patient has premature (<55 years men; <60 years women) cerebral or peripheral vascular disease	1
	Physical Examination	
A	Tendon xanthomata	6
B	Arcus cornealis below 45 years	4
	Laboratory analysis	
A	LDL-cholesterol > 8.5 mmol/l	8
B	LDL-cholesterol 6.5-8.4 mmol/l	5
C	LDL-cholesterol 5.0-6.4 mmol/l	3
D	LDL-cholesterol 4.0-4.9 mmol/l	1
	(HDL-cholesterol and triglycerides are normal)	
	DNA analysis	
A	Functional mutation	8

An early approach to diagnose FH was developed by the MEDPED study (Make Early Diagnoses and Prevent Early Deaths), in the USA (Williams et al. 1993). The MEDPED criteria were based on age- and sex- adjusted total cholesterol levels, however it has been demonstrated that in order to avoid a high false positive rate the diagnosis of FH should generally not be made using total or LDL cholesterol levels alone. The MEDPED cholesterol cut-offs for each age group are shown in Table 4.

The sensitivity and specificity in identifying FH patients with a mutation using all three clinical diagnostic algorithms was compared in a study of 408 patients, which showed that the Simon Broome criteria and the DLCN score did not differ much, but both performed better than the MEDPED diagnosis (Damgaard et al. 2005).

Table 4

MEDPED program TC cut-offs for diagnosis of FH in relatives of a patient and their comparison with the general population (Williams et al. 1993).

	Total cholesterol (mmol/L)		
Age (years)	**First degree relative**	**Second degree relative**	**General population (Utah)**
<20	5.7	5.9	7.0
20–29	6.2	6.5	7.5
30–39	7	7.2	8.8
>40	7.5	7.8	9.3

1.5. Genetic diagnosis of FH

A DNA-based test can give a definite diagnosis of FH and it is particularly useful at identifying affected relatives of a patient, especially those at a young age. Because of the difference in the FH mutation spectrum between populations, different screening methods have been applied. Methods used in the UK, a country with high genetic heterogeneity, originally included the commercially available Amplification Refractory Mutation System (ARMS) kit (Elucigene FH20, Gen-Probe, Oxford, UK), which was designed to screen for the most common UK FH variants including18 *LDLR* mutations, the *APOB* p.R3527Q, and the *PCSK9* p.D374Y. Taylor et al. reported that the ARMS assay detected over 44% of the UK FH mutations (Taylor et al. 2010b), with the *APOB* p.R3527Q being the most commonly observed accounting for 12% of all detected mutations.

Because of the wide spectrum of FH mutations in the UK, high resolution melting (HRM) of the entire coding region of *LDLR*, followed by Sanger sequencing, has been used to screen patients for less common mutations (Whittall et al. 2010b). The method was shown to be effective, however its sensitivity decreases in some regions depending on the nucleotide composition of the fragment. An additional method was developed to test for large deletions and duplications within the *LDLR* gene (Taylor et al. 2009). The multiplex ligation-dependent probe amplification (MLPA) (MRC-Holland) is a rather laborious and costly method, but it allows the detection of large deletions and insertions, which account for up to 10% of all UK FH mutations (Futema et al. 2013). Another commercially available FH diagnostic kit, which

test for over 250 FH mutations is LIPOchip (Progenica Biopharma, Spain), was shown to detect an FH mutation in 66% of over 800 screened Spanish FH patients (Alonso et al. 2009). Although Elucigene FH20 and LIPOchip have been widely used as a simple diagnostic tool, a recent review funded by The National Institute for Health Research Health Technology Assessment programme highlighted the need for novel comprehensive FH mutation detection methods, and demonstrated that the off-the-shelf diagnostic kits may be less cost-effective in comparison to the novel emerging methods (Sharma et al. 2012).

Next generation sequencing (NGS) has certainly revolutionised molecular diagnostic labs, with the costs of custom design targeted deep sequencing methods, provided by for example Illumina or Agilent, steadily decreasing. Novel NGS approaches, which include PCR-free methods such as the TruSeq DNA PCR-free sample preparation kit from Illumina, can provide reliable results in two days, however their sensitivity and specificity is still being tested.

The development of more efficient and cheaper screening methods for FH will lead to higher mutation detection rate. However, another reason for low mutation detection rate in some of the lipid clinics could be due to misdiagnosis of the monogenic hypercholesterolaemia with the polygenic form of the disease. A recent report demonstrated that at least 20% of the patients diagnosed with FH but with no detected mutation may have had high cholesterol due to polygenic factors (Talmud et al. 2013). Finally, the detection rate can be low due to a presence of novel FH loci, located outside of the screened regions.

Table 5

FH mutations included in the ARMS test. The list contains cDNA position starting from ATG.

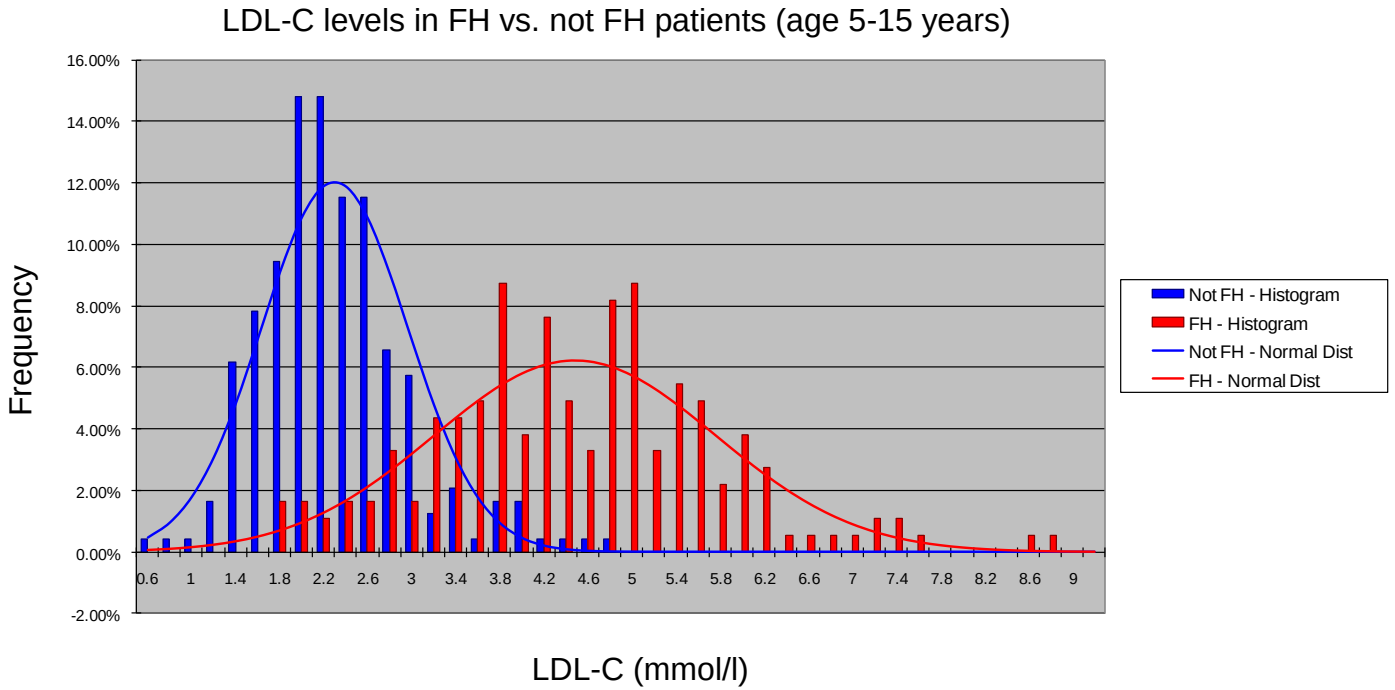
Mutation and protein change
c.259T>G (p.W87G)
c.301G.A (p.E101K)
c.313+1G>A
c.551G>A (p.C184Y)
c.654_656delTGG (p.G218del)
c.680_681delAC (p.D227Gfs*12)
c.662A>G (p.D221G)
c.681C>G (p.D227E)
c.682G>T (p.E228X)
c.932_933delAA (p.K311Rfs*20)
c.1048C>T (p.R350X)
c.1150C>T (p.Q384X)
c.1285G>A (p.V429M)
c.1436T>C (p.L479P)
c.1444G>C (p.D482H)
c.1444G>A (p.D482N)
c.2029T>C (p.C677R)
c.2054C>T (p.P685L)
APOB c.10580G>A (p.R3527Q)
PCSK9 c.1120G>T (p.D374Y)

1.6. Clinical utility of DNA testing

DNA testing for FH mutations can provide an unequivocal diagnosis in comparison to diagnosis based on clinical observations and lipid measurements. This has been confirmed in several studies (Soutar and Naoumova 2007, Varret et al. 2008, Campagna et al. 2008). The genetic diagnosis overcomes the overlap problem of LDL-C levels between the affected and not affected individuals. It was reported that the LDL-C overlap in children (5-15 years old) leads to false negative rate of FH diagnosis in 15% of patients (Figure 5), and it gets worse with age (Starr et al. 2008).

Figure 5

LDL-C levels in FH patients (with genetic diagnosis) vs. not FH patients aged 5-15 years. The LDL-C overlap leads to 8% mutation negative individuals being falsely diagnosed with FH based on their LDL-C level, and 15% of mutation positive patients not being diagnosed with FH based on their LDL-C level. Taken from (Starr et al. 2008).



One of the main benefits of identifying the disease-causing variant is that it allows for diagnosis of affected relatives. In a family with a monogenic disease like FH, 50% of the first-degree relatives are expected to carry the mutation. For comparison, a study showed that only 25.6% of relatives of FH patients without a detected mutation were given the FH diagnosis, which suggests a polygenic rather than monogenic cause of hypercholesterolaemia (Humphries et al. 2006a). The DNA testing has been recommended by the National Institute for Health and Clinical Excellence (NICE) as a tool to identify affected relatives in cascade testing, as discussed in the next section.

Genetic diagnosis of FH in children can help to implement life-style changes, such as diet, exercise, no smoking, at an early age, which can reduce the risk of developing premature CHD. The use of statins in pediatric patients was demonstrated to be beneficial (Eiland and Luttrell 2010, Arambepola et al. 2007, Avis et al. 2007), although its long-term efficacy have not been established. The need for a long-term randomized controlled trial in children has been highlighted (Vuorio et al. 2010).

Furthermore, identification of an FH mutation is beneficial for the patients monitoring and for the management of LDL-C-lowering therapy (Humphries et al. 2006a, Umans-Eckenhausen et al. 2001).

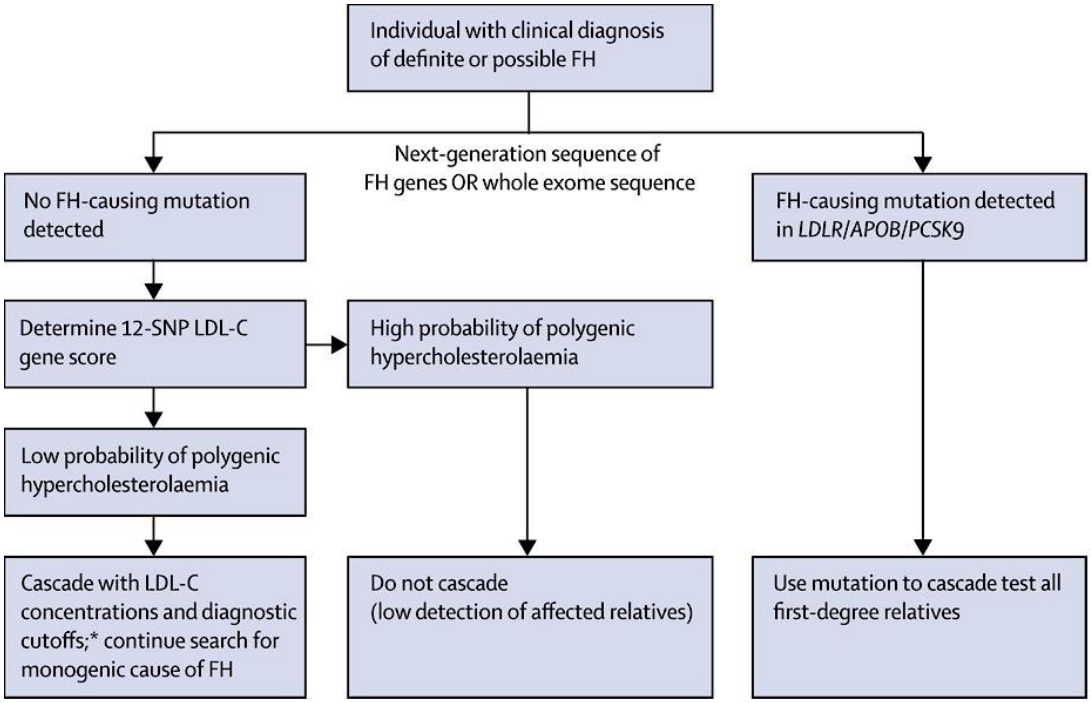
1.7. Cascade testing to identify FH patients

In 2008, the National Institute for Health and Clinical Excellence (NICE) published guidelines and recommendations on the diagnosis and management of FH in the UK (CG71) (Wierzbicki et al. 2008). The guidelines

state that healthcare professionals should offer a referral for cascade testing to every FH patient, in order to identify affected relatives of the patient. This should include at least the first- and second- and, when possible, third-degree biological relatives. Cascade testing is of high public health interest and its goal is to reduce morbidity and mortality from heart disease in persons with FH through early diagnosis and effective disease management. A combination of DNA testing (for families where the causal mutation is already identified), which provides unequivocal diagnosis (Campagna et al. 2008), and measurement of LDL-C levels (in the remaining families) has been demonstrated as the most cost-effective current approach (Nherera et al. 2011). This strategy has been successfully applied in the Netherlands in families of familial hypercholesterolaemia probands with a detected mutation, thus allowing statin treatment of family members at risk of early coronary heart disease (Umans-Eckenhausen et al. 2001). A recent case-control study on the use of LDL-C SNPs score to identify patients with polygenic hypercholesterolaemia (Talmud et al. 2013) suggested to implement the LDL-C SNPs score calculations (discussed in section 2.3) before beginning the cascade testing in the mutation negative patients (Figure 6). The proposed strategy needs to be tested further, however it is predicted to increase the cost-effectiveness of the cascade testing. Despite clear benefits of the early identification of FH patients recently highlighted by the HEART UK charity (http://heartuk.org.uk/files/uploads/documents/HUK_HealthEconomics_FINAL2012_2702.pdf, accessed on 20 Aug 2013), cascade testing has not been widely implemented in England.

Figure 6

The proposed diagnostic workflow for cascade testing in patients with FH (taken from (Talmud et al. 2013)). *as recommended by NICE (Leren et al. 2004)

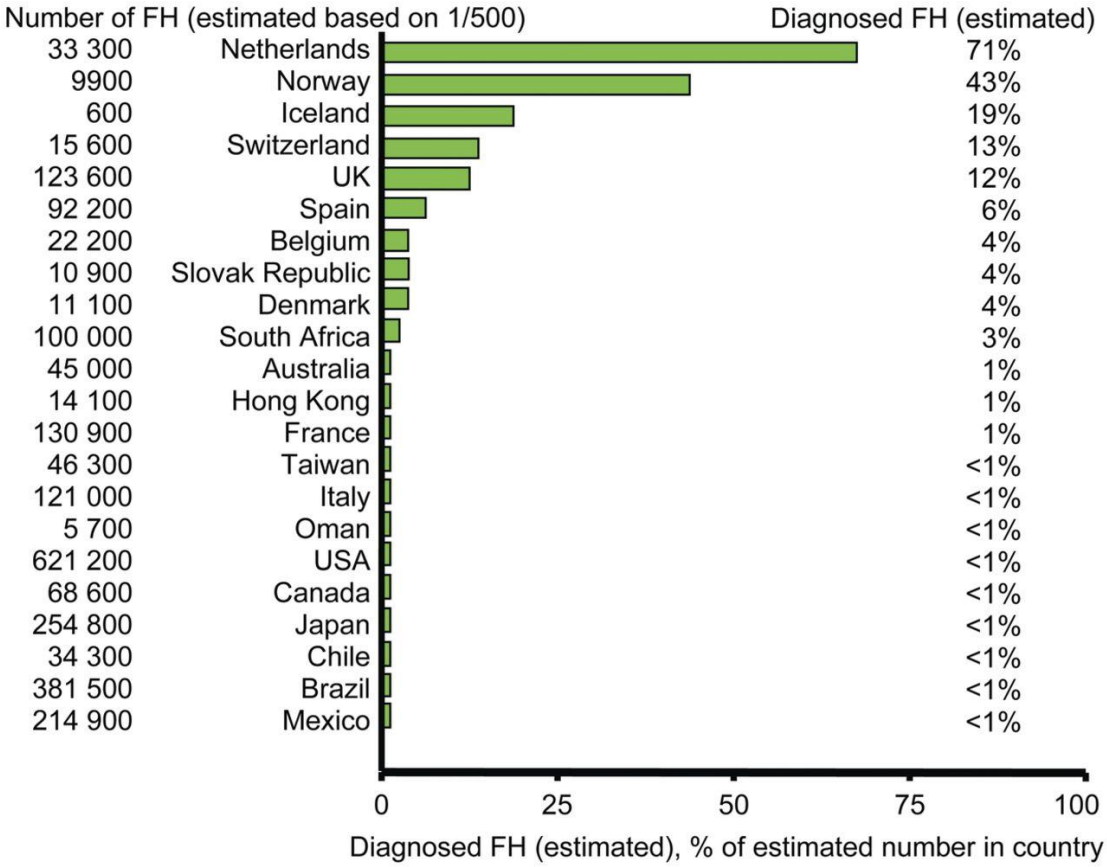


1.8. Problem of FH underdiagnosis

In 2000, a prospective study in an Oxford based patients registry, highlighted the problem of FH underdiagnosis (Neil et al. 2000), which was further confirmed in a census estimating that only 15% of FH patients in the UK were identified (Marks et al. 2004). A recent report of the European Atherosclerosis Society demonstrated that in the majority of studied countries the proportion of FH patients being diagnosed is below 1% (Nordestgaard et al. 2013). The highest proportion of diagnosed patients was reported in the Netherlands (71%), followed by Norway (43%), Iceland (19%), Switzerland (13%), the UK (12%) and Spain (6%) (Figure 7). Authors estimated the proportions based on the FH prevalence of 1 in 500, however the differences in the frequency between some countries were not accounted for. The report suggested that there are between 14 and 34 million individuals affected by FH worldwide, which highlights the extent of FH underdiagnosis. Early diagnosis should lead to the introduction of statin treatment, patient education and advice on diet and life-style as well as cascade testing to identify affected relatives. Well-operated cascade testing system would lead to an increase in the efficiency of CHD prevention and safe in the UK only an estimate of £6.9 million per year from avoiding cardiovascular events, as stated by the HEART UK charity (http://heartuk.org.uk/files/uploads/documents/HUK_HealthEconomics_FINAL2012_2702.pdf, accessed on 20 Aug 2013).

Figure 7

Estimated per cent of individuals diagnosed with familial hypercholesterolaemia in different countries/territories, as a fraction of those theoretically predicted based on a frequency of 1/500 in the general population. As most countries do not have valid nationwide registries for familial hypercholesterolaemia, several values in this figure represent informed estimates from clinicians/scientists with recognized expertise in and knowledge of familial hypercholesterolaemia in their respective countries. Numbers were provided by Michael Livingston, Steve E. Humphries (UK), Olivier S. Descamps (Belgium). (Taken from (Nordestgaard et al. 2013).



1.9. Treatment of FH

The most common therapy for FH patients includes daily use of statins, which are lipid-lowering compounds, which stimulate expression of the LDL-receptor gene (*LDLR*) by inhibiting HMG-CoA (hydroxymethylglutaryl co-enzyme A) reductase activity. Although use of statins in FH patients was shown to be the most effective (see the next paragraph), alternative medications are available, when a patient does not tolerate the statin treatment. These include bile acid sequestrants, fibrates, nicotic acid, all in conjunction with low-fat diet. Lifestyle changes, in particular quitting of smoking, are advised to every FH patient. A combined FH therapy is also in use. Ezetimibe is a recently introduced lipid-lowering drug, which acts by decreasing cholesterol absorption in the intestine (Toth and Davidson 2005). Ezetimibe can be used in combination with statins, which can result in additional cholesterol reduction (Morrone et al. 2012), and it has been recommended as an alternative and additional therapy by the NICE guidelines (2007), although a recent study of over 3,800 patients showed no significant mortality benefit over the use of statins alone (Patel et al. 2013).

The efficacy of statins in heterozygous FH patients was assessed by the Simon Broome Scientific Steering Committee in a study of over 3,000 FH individuals, who were followed up for 26 years, until the end of 2006 (Neil et al. 2008). The standardized mortality ratio (SMR), which was expressed as a percentage (%), was compared between different age groups of patients before and after the statin treatment was introduced in 1992. The authors reported that the statins, when used as the primary prevention, reduced the

SMR for CHD over 59% in the middle age group (40-59) of participants (Table 6). There was also a beneficial effect in the younger age group; however the result may be imprecise because of the small sample size hence large confidence intervals. In addition, it was shown that the secondary prevention reduced the number of deaths from the CHD in the whole group by 25%, with a larger effect in women. Overall, the study validated the significant impact of the early identification and treatment of heterozygous FH patients, and confirmed the importance of the early identification of FH patients for example by cascade testing (discussed in section 1.7.).

Table 6

The effect of the use of statins (primary prevention) in heterozygote FH patients (SMR for CHD, SMR=100 for the reference population) [adapted from (Neil, Cooper et al. 2008)].

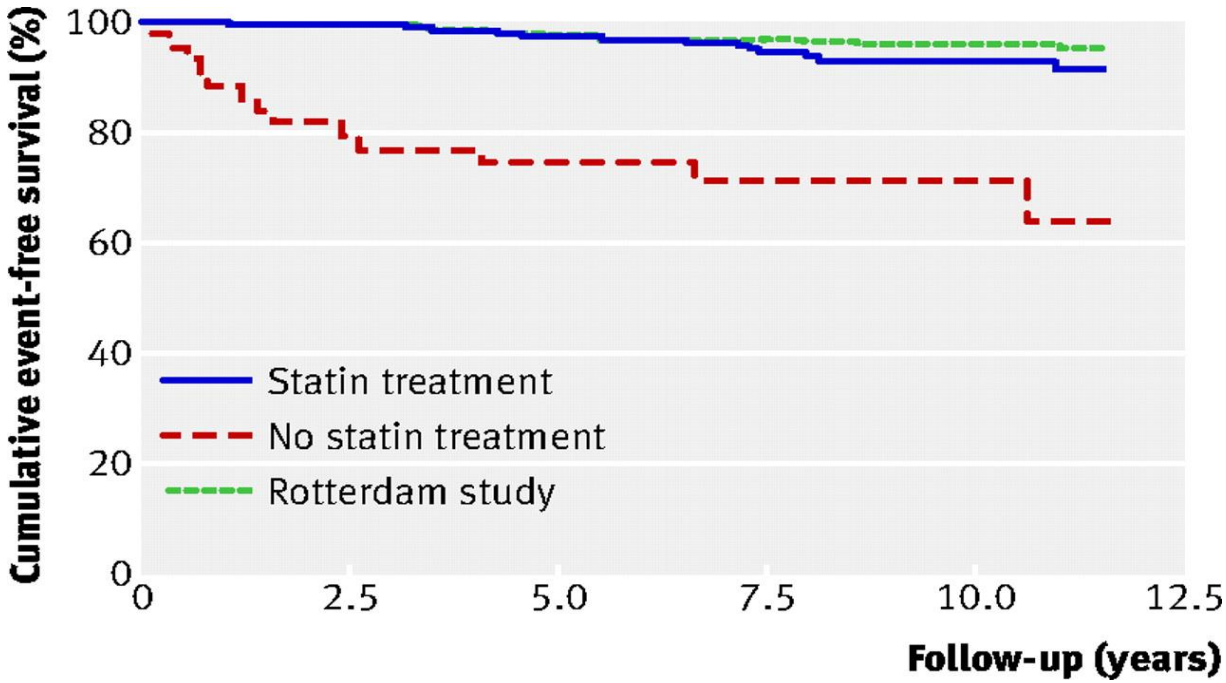
Age groups (years)	**1980-1991**			**1992-2006**		
	SMR	**95% CI**	***P*-value**	**SMR**	**95% CI**	***P*-value**
20-39	3750	773-10 959	<0.001	1153	238-3372	<0.01
40-59	342	148-674	<0.01	141	75-242	0.28

Similar findings were reported by Versmissen et al., who mimicked a statin clinical trial by following up around 2,000 FH patients, for an average of 8.5 years, before and after the statin treatment (Versmissen et al. 2008). The authors observed a 76% reduction in the risk of CHD in treated versus

untreated FH patients, which led to a significant improvement of the event-free survival. Even more, when the risk of myocardial infraction (MI) in statin-treated FH patients (older than 55) was compared with the data of a general population (the Rotterdam study), the event-free survival of FH individuals was not different from the general population (Figure 8).

Figure 8

Kaplan-Meier curve of cumulative MI-free survival among FH patients (older than 55) according to the statin treatment (the Rotterdam study represents the general population) (Versmissen et al. 2008).



Although most patients respond well to currently available treatments, the development of novel pharmacotherapies specific to particular defects in the LDL-C pathway is on its way. Inhibition of PCSK9 by human monoclonal antibodies (REGN727 and AMG 145), which prevent the degradation of LDL-receptors, has been recently demonstrated as a safe, well-tolerated and effective way of lowering plasma LDL-C (Koren et al. 2012, Stein et al. 2012b, Dias et al. 2012, Raal et al. 2012). The promising outcomes of PCSK9 targeting led to further developments of the antisense oligonucleotide (ASO) (Graham et al. 2007), locked nucleic acid (LNA)-ASO (Gupta et al. 2010), siRNAs (Frank-Kamenetsky et al. 2008), and mimetic peptides (Shan et al. 2008), all aiming to silence the activity of PCSK9. The safety and efficacy of these new therapeutic approaches is currently being tested.

Targeting ApoB as a novel FH treatment has also shown some signs of success. The overproduction of ApoB in FH patients can be lowered by an antisense oligonucleotide (known as Mipomersen) or by microsomal transfer protein (MTP) inhibitor (known as Lomitapide). A recent clinical trial showed that Mipomersen can effectively reduce ApoB-containing lipoproteins in FH patients, when used in combination with statins (Stein et al. 2012a) and in statin-intolerant patients (Visser et al. 2012). The use of an MTP inhibitor also resulted in a significant LDL-C reduction, despite an increased hepatic fat accumulation in some of the treated individuals (Cuchel et al. 2007), Lomitapide together with Mipomersen have recently been approved by the Food and Drug Administration (FDA) for use as an adjunct to diet and statin therapy in homozygous FH patients.

The emerging novel therapies target specific defects in the cholesterol metabolism, thus knowing the precise mutation, hence the defect in the pathway, may be required for choosing the effective therapy. For example, patients carrying a mutation affecting the LDL-receptor recycling process are less likely to respond to the PCSK9 inhibitors. The same prediction applies to homozygous FH patients with a null mutation in the *LDLR* gene.

1.10. FH frequency

The estimated frequency of FH in most populations is 1 in 500 for heterozygous individuals, and about 1 in 1 million for homozygotes (Goldstein et al. 1973). However significant differences have been observed between populations. A recent study on a general population in Denmark population using the DLCN diagnostic score above 5 (the lowest point for probable FH) found an FH frequency of 1 in 137, which is much higher than commonly perceived (Benn et al. 2012). A study performed on a cohort of Afrikaners living in the region of Johannesburg, in South Africa, revealed a much higher prevalence of the disease, estimating its heterozygous frequency to be around 1 in 100 (Seftel et al. 1980). Similar observations were made by Khachadurian et al (Khachadurian and Uthman 1973), who reported an even higher heterozygote frequency (1 in 85) in the Christian Lebanese population, with projected occurrence of homozygotes higher than 1in 100,000. On the other hand, Japanese citizens appear to have much lower incidence of FH, calculated to be approximately 1 in 900 individuals (Mabuchi et al. 1977). These clear differences in FH frequency are mainly due to founder effects, which occur when a small cohort is genetically isolated during evolution,

leading to a so-called population bottleneck. Such isolation can be caused by for example linguistic, geographical or religious barriers. The frequency of FH in South-Indian population remains unexamined. A study of Indians living in South Africa suggested the prevalence to be even higher than 1 in 100 but this would be due to a founder effect specific to the Indians in South Africa who derived from 150,000 South and East Indian immigrants, who settled there between 1860 and 1911 (Rubinsztein et al. 1993). However, the number of deaths from CVD is steeply increasing in India (Yusuf et al. 2001, Setia et al. 2012) and there is an urgent need for identification of FH mutations in the Indian population (Setia et al. 2012).

When molecular diagnoses of FH were developed it became clear that in populations with a higher prevalence, FH is caused by a much smaller number of *LDLR* mutations. For instance, almost 90% of FH patients in the North Karelian region of Finland carry the same mutation, which has been first seen in Finland around 400 years ago (Vuorio et al. 1997). Several other cohorts with higher FH frequency are caused by just a few mutations including western Scotland (Lee et al. 1998), northwestern Greece (Miltiadous et al. 2001), people in central and southern Tunisia (Jelassi et al. 2009) and Iceland (Gudnason et al. 1997). Countries with a much lower population stratification, such as those in central Europe, have in contrast a much bigger variety of *LDLR* mutations (Dedoussis et al. 2004).

Knowing the FH frequency and spectrum of the mutations can help to design a simple and cost-effective strategy to identify affected individuals, specially in isolated populations. Over 1,200 unique *LDLR* variants known to cause FH have been logged into a database maintained by Dr Sarah Leigh, at UCL,

which can be accessed at http://www.ucl.ac.uk/fh (Leigh et al. 2008).

1.11. FH genes

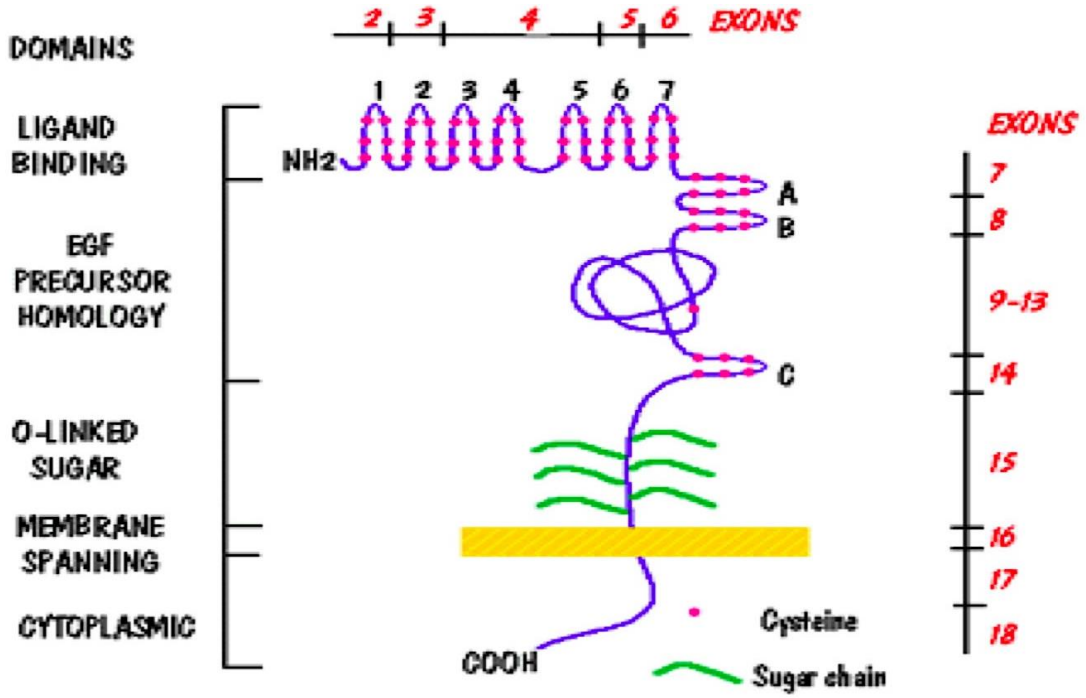
1.11.1. Low Density Lipoprotein Receptor gene (*LDLR*)

The first identified FH mutations were found to affect the function of the low-density lipoprotein receptor (LDL-R), encoded by the *LDLR* gene. *LDLR,* which is located on chromosome 19p13.2, was called the "mosaic of exons shared with different proteins" (Sudhof et al. 1985). It comprises of 18 exons that are transcribed and translated into several distinct domains, which together form a cell surface receptor for the LDL particle, as shown in Figure 9. Exon one (not shown in Figure 9) codes for a 21-amino-acid-long signal domain, which is cleaved from the mature protein during its synthesis in the endoplasmic reticulum. Exons 2-6 code for the cysteine-rich region, which once translated forms disulfide bridges and functions as the ligand-binding domain for ApoB- the major protein of LDL-C (discussed later). Exons 7-14 are highly similar to the human epidermal growth factor (EGF) precursor gene, and the domain encoded by this region acts during receptor recycling, when lipoproteins dissociate from the LDL-R in the endosome. Davis et al. (Davis et al. 1987) also shown that the EGF precursor homology domain positions the ligand-binding domain of LDL-R to the cellular surface and therefore enables it to bind LDL-C. Exon 15 encodes 58 amino acids, mainly threonine and serine, which function as an attachment site for O-linked carbohydrate chains. However, a study on cultured hamster fibroblasts suggested that the O-linked carbohydrate chains domain is not essential for

the normal functioning of LDL-R (Davis et al. 1986). The LDL-R membrane-spanning domain is composed of 22 hydrophobic amino acids, and is encoded by exon 16 and the 5' of exon 17. The final 50 amino acids function to localise the receptor into coated pits on the cell surface, and are encoded by exons 17(3') and 18 (Chen et al. 1990).

Figure 9

Domain structure of the human LDLR protein and its relation to the exon organisation of the gene [adapted from (Hobbs et al. 1990)]. The signal peptide encoded by exon 1 of *LDLR* is not shown.



To date, over 1,200 *LDLR* mutations have been identified (Usifo et al. 2012) and they account for the majority of all FH cases. The distribution of the reported FH variants across the gene is not even, as shown in Figure 10A. The highest frequency of FH mutations was observed in exon 4, which could be explained by its large size but also by the frequent occurrence of CpG sequences, which are highly mutable (Hobbs et al. 1990, Day et al. 1997b, Leigh et al. 2008). Austin et al. suggested that selection bias could also be one of the reasons for the higher number of exon 4 mutations (Austin et al. 2004). They concluded that patients with mutations in the ligand-binding domain of *LDLR* are overrepresented in the lipid clinic, due to the highly detrimental effect of such variants on the gene function and therefore more severe FH phenotype.

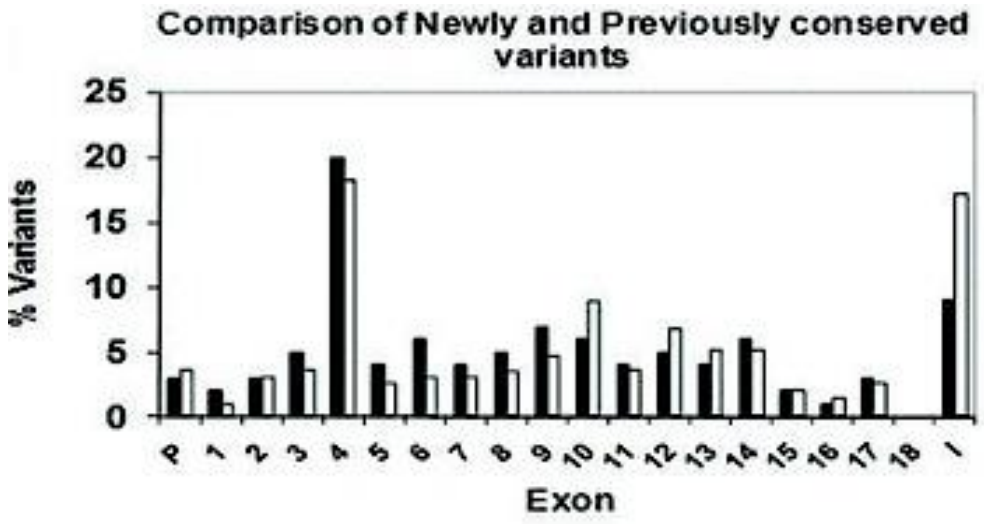
On the other hand, the FH-causing mutation frequency in exons 15 and 16 are relatively low. Considering conservation scores for each of the *LDLR* exons, exons 15 and 16 are the least conserved (apart from the signal peptide domain encoded by exon 1) along the gene, as shown in Figure 10B, which suggests their minor importance in the whole functionality of LDL-R.

Although the *LDLR* gene has been studied extensively, the functional effect of every detected base change must still be thoroughly evaluated. Several bioinformatics tools are available to predict the effect of nonsynonymous variants (e.g. PolyPhen, SIFT, Mutation Taster), however the assessment of intronic changes or variants that may affect the splicing process require further investigation, including *in vitro* and family co-segregation studies.

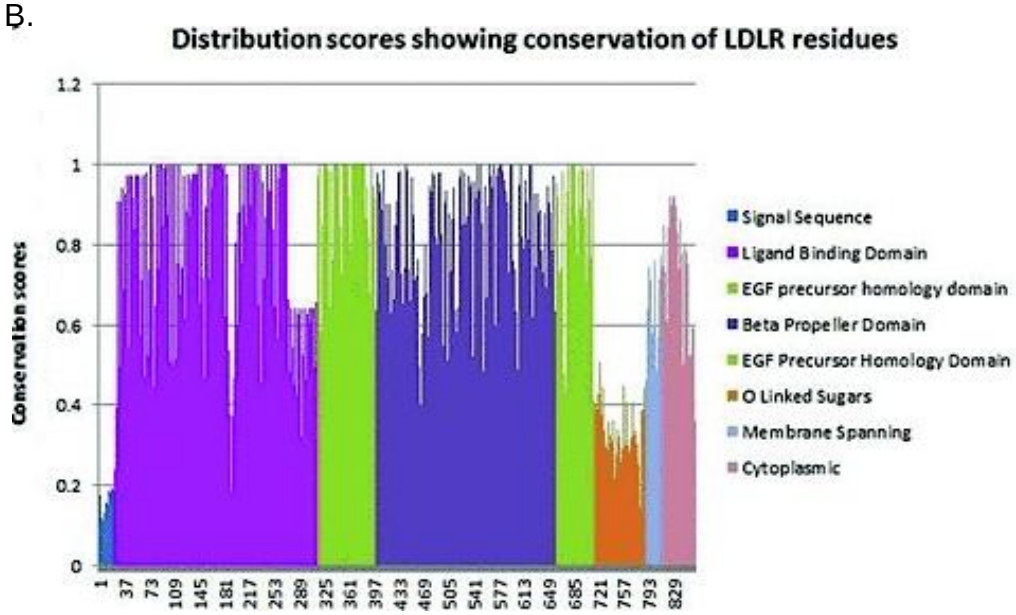
Figure 10

A. Percentage of reported *LDLR* variants in 2008 (Leigh et al. 2008) (black bars) and in 2012 (Usifo et al. 2012) (white bars). There was an increase in the prevelance of mutations reported in the promoter, exons 10, 12, 13 and 16, and in the intronic variants between the UCL FH database updates. B. Conservation scores for all *LDLR* domains, based on the allignments of functionally similar proteins with the same domain types. (Both taken from (Usifo et al. 2012))

A.



B.

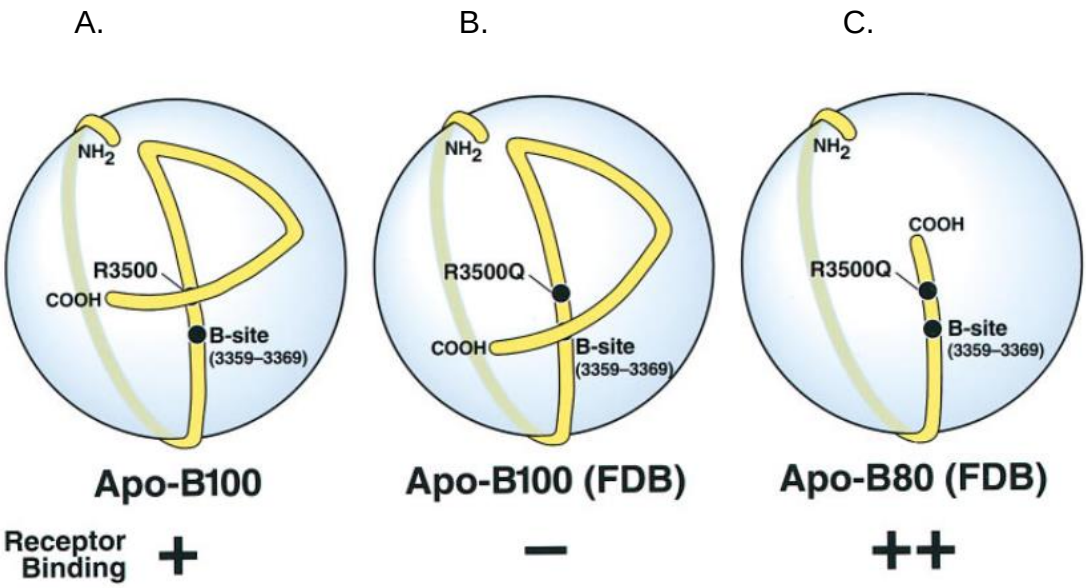


1.11.2. Apolipoprotein B gene (*APOB*)

The second locus found to be causing FH is the *APOB* gene, located on chromosome 2. The gene encodes apolipoprotein B (ApoB), which, once synthesised in the liver, is the major apolipoprotein of VLDL, IDL, and LDL particles. ApoB also functions as a ligand to the LDL receptor. A thorough study by Boren and colleagues identified the LDL-R binding site in ApoB, which is located between residues 3,386-3,396 (Boren et al. 1998). The interaction between the receptor and the ligand is known to be modulated by the carboxyl tail of ApoB, as demonstrated in Figure 11. The *APOB* defect was first discovered in FH patients with normal LDL-R activity, but with reduced ApoB affinity to the LDL receptor (Innerarity et al. 1987). Soria et al. (Soria et al. 1989), followed by further studies (Ludwig and McCarthy 1990), which demonstrated that the Familial Defective Apolipoprotein B (FDB), also called Type B Familial Hypercholesterolaemia, occurs due to G to A transition at nucleotide 10580, in exon 26 of *APOB*, which leads to the substitution of Glutamine 3527 to Arginine (p.R3527Q). The mutation accounts for approximately 6-10% of all FH defects in the UK and it is mainly limited to individuals of European descent. It was hypothesised that the mutation originated from an European ancestor about 7,000 years ago and since then it has been moved across the continent due to migration (Myant et al. 1997). Carriers of the p.3527Q are usually characterised by slightly lower pre-treatment TC and LDL-C than carriers of any *LDLR* mutation, but the risk of developing premature CHD (at the age of 50) was estimated to be 40% for males and 20% for females (Tybjaerg-Hansen and Humphries 1992).

Figure 11

Schematic representation of A. Apo-B100 (full length) wrapped around the LDL particle and the carboxyl tail of ApoB as the modulator of LDL-receptor binding. B.FH mutation p.R3527Q (the figure shows old numbering without the signal peptide (27 amino acids)) is predicted to disrupt the interaction of p.R3527 residue with the carboxyl tail, which blocks the LDL-R-binding site (B-site) and ApoB can no longer bind to LDL-receptor. C. Deletion of the carboxyl modulator site allows Apo-B80 (80% of the full length) enhances binding of ApoB to the receptor despite the FH mutation p.R3527Q being present. Adapted from (Boren et al. 1998).



APOB is a very large and polymorphic gene, encoding 4,563 amino acids. Thus the majority of FH studies examined only a portion of exon 26, which is known to influence the binding of ApoB to the LDL-receptor (Figure 11). A recent study, which applied next generation sequencing (NGS) methodology, has identified two novel *APOB* variants, p.R3059C and p.K3394N, which were demonstrated to co-segregate with the FH phenotype in a family (Motazacker et al. 2012). However, these variants are predicted to be very rare as to date no further reports confirming the mutations have been published. Several studies associated the *APOB* variant p.R3531C with FH (Pullinger et al. 1995, Rabès et al. 1997, Wenham et al. 1997), however its effect was not consistent within examined families and the variant was later concluded as not FH-causing (Rabès et al. 2000). Furthermore, some rare *APOB* variants, which lead to premature truncation of the peptide, were found to lower lipid levels in FH patients who were found to carry an FH-causing mutation at a different locus. FH patients with a nonsense *APOB* variant had 56% lower LDL-C levels when compared to FH patients without premature stop codons in *APOB* (Huijgen et al. 2012).

1.11.3. Proprotein Convertase Subtilisin/Kexin type 9 gene (*PCSK9*)

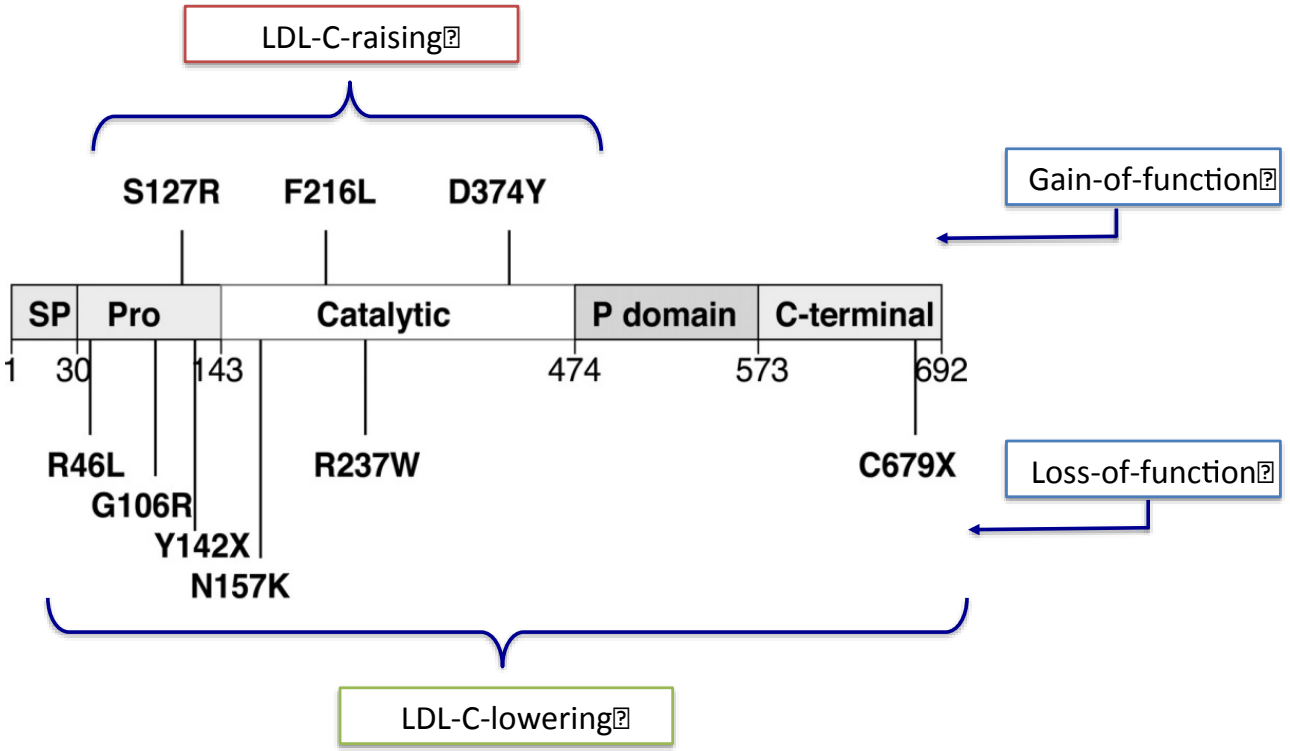
A third FH locus, the *PCSK9* gene located on chromosome 1p32, was found by linkage analysis of a large Utah family with strong FH history, but with no *LDLR* or *APOB* mutation detected (Abifadel et al. 2003). *PCSK9*, also known as *NARC1*-neural apoptosis regulated convertase, encodes proprotein convertase subtilisin/kexin type 9, which enhances degradation of LDL-Rs

during endocytotic recycling back to the cell surface (Qian et al. 2007). PCSK9 residues between 367 and 381 and its catalytic site (residues 153-155) bind directly to the N terminal of the EGF-like domain A (Figure 9) of LDL-R (Kwon et al. 2008).In 2003, a co-segregation study of a families with autosomal dominant hypercholesterolaemia identified a *PCSK9* variant p.D374Y as a novel cause of FH (Abifadel et al. 2003), which was followed by functional analyses (Sun et al. 2005). P.374Y was later found to increase the affinity of PCSK9 for the LDL-R, and therefore behaved as a 'gain-of-function' mutation (Kwon et al. 2008). Patients with the p.374Y mutation have more severe form of the disease than those with *LDLR* or *APOB* mutations, so that their pre-treatment cholesterol is even higher and their response to statin treatment is not so effective (Naoumova et al. 2005). Approximately 2% of all UK FH patients with a detected mutation have the *PCSK9* variant in the UK population (Humphries et al. 2006c). University College London hosts a publically available database of published *PCSK9* variants (www.ucl.ac.uk/fh), which contains nearly 100 independent *PCSK9* variations, some of those are shown in Figure 7. A quarter of the reported variants are predicted to be loss-of-function, half are common polymorphisms and 23 seem to be gain-of-function, however further studies are needed to establish their effect (Leigh et al. 2009). Loss-of-function variants are known to have a beneficial effect on LDL-C levels (Hooper et al. 2007, Yue et al. 2006). A study found that two of the nonsense variants, p.Y142* and p.C679*, which can reduce LDL-C by 40%, are much more common in the African American population (combined frequency 2%) in comparison to Americans of European background (frequency less than 0.1%) (Cohen et al. 2005). Positive selection pressure

has been suggested to be the reason for such high frequency of the loss-of-function variants in individuals of African ancestry.

Figure 12

Summary of functional variants identified in *PCSK9* and their position in relation to gene's domains. SP=signal peptide, Pro=prosegment (both cleaved-off in the ER). Adapted from (Berge et al. 2006)



1.11.4. Low Density Lipoprotein Receptor Adaptor Protein 1 gene (*LDLRAP1*)

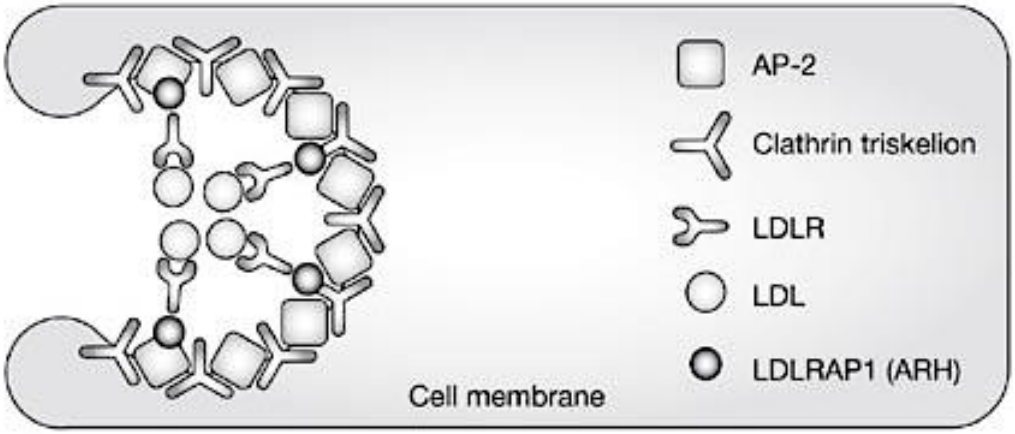
In the mid-nineties, a condition of very similar characteristics to the homozygote FH was observed, however the inheritance pattern did not agree with the dominant model (Zuliani et al. 1999). The condition, called Autosomal Recessive Hypercholesterolaemia (ARH), occurs when a patient inherits two recessive variants, which in 2001 were mapped to the *ARH* gene, later named Low Density Lipoprotein Receptor Adaptor Protein 1 (*LDLRAP1)* (Garcia et al. 2001). *LDLRAP1* encodes an adaptor protein, which is necessary for internalisation of the LDL-LDLR complex and for efficient binding of LDL to its receptor (Michaely et al. 2004). Because of the recessive model of ARH, the disease is relatively rare, and *LDLRAP1* mutations are more common in founder populations, such as those of Sardinian origin.

The *LDLRAP1* gene, originally named *ARH* (Autosomal Recessive Hypercholesterolaemia), was linked with the rare pathogenesis of hypercholesterolaemia inherited in a recessive manner (Garcia et al. 2001). Mutations in *LDLRAP1* are more common in the Sardinian population and the majority of patients are characterized by bulky xanthomas from early childhood, although their plasma LDL-C appears to be lower than in patients with homozygous FH (Soutar et al. 2003). The gene is located on the short arm of chromosome 1 and it codes for the adaptor protein characterized by the presence of the phosphotyrosine-binding domain (PTB), which is known to bind NPXY motifs found in the cytoplasmic domain of signaling receptors (Forman-Kay and Pawson 1999). LDLRAP1 interacts with the cytoplasmic

domain of the LDL-receptor, which promotes the receptor endocytosis via the clathrin-coated pits on the cell surface, demonstrated in Figure 13.

Figure 13

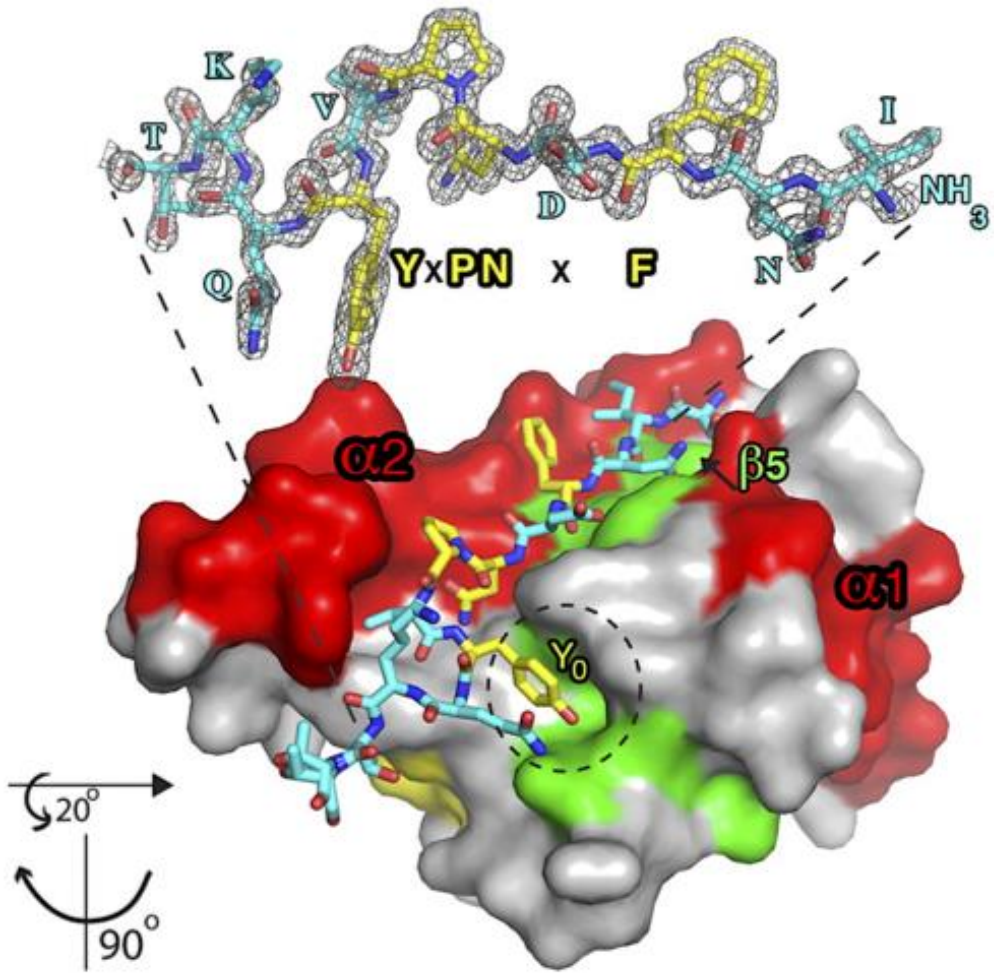
Schematic overview of LDLRAP1 function. The LDL receptor adaptor protein functions to engage the LDL receptor with the clathrin-coated pit machinery for endocytosis (taken from (Soutar and Naoumova 2007)).



In a recent study, it has been thoroughly demonstrated that the LDLRAP1 interacts with the "$I_{-7}xF_{-5}xNPxY0QK_{+2}$" domain of the LDL-R tail (Dvir et al. 2012). The domain is located between residues p.I821 (named as I_{-7}) and p.K830 (named as K_{+2}), and the crystal structure of the LDLRAP1-LDL-R complex is shown in Figure 14. The presented model will allow for analysis of point mutations in LDLRAP1 as well as mutations affecting the *LDLR* cytoplasmic domain, which are of ambiguous effect.

Figure 14

The structure of PTB domain of LDLRAP1 and LDL-R tail complex. LDL-R tail shown as cyan sticks with the core LDLRAP1-interacting residues in yellow.



1.11.5. Other loci causing FH

Since an FH-causing mutation can be detected in only 60-90% of definite FH patients (Graham et al. 2005, Taylor et al. 2010b, Futema et al. 2013) several studies on patients without a reported mutation were conducted in order to find a monogenic cause for hypercholesterolaemia in the unexplained FH cases. In 2010, a study of a large French family with FH led to linkage of 16q22.1 region with the disease (Marques-Pinheiro et al. 2010). The genomic region containing 154 genes, and was named *HCHOLA4*, however after the extensive analysis the causal variant was not identified. A year later, family linkage analysis of three pedigrees mapped an FH locus to 8q24.22 cytoband in one of the studied families (Cenarro et al. 2011). The identified region contains approximately 30 genes, however the authors could not pinpoint the functional variant. No linkage was found in the remaining two families. Shortly after, another study reported linkage of FH with two different chromosomal regions, 3q25-26 and 21q22, in one Chinese FH pedigree (Wang et al. 2011). The authors selected two genes, *ABCA1* and *LSS*, for further sequencing however no mutations were identified. All three studies are summarized in Table 7.

Table 7

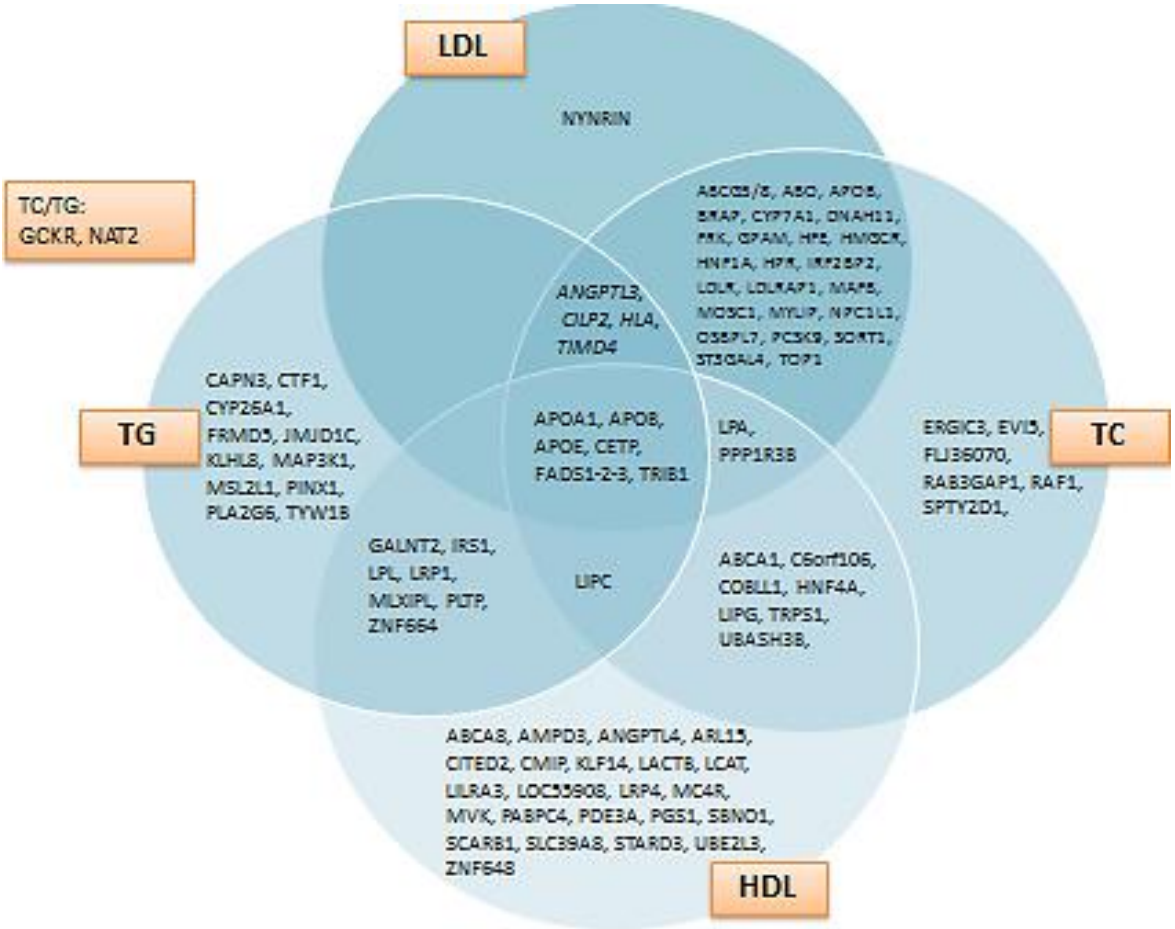
Summary of genome-wide linkage studies of families with clinical diagnosis of FH.

Publication	Number of studied families	Number of studied members of the family	Locus	Follow up
Marques-Pinheiro A., et al. (2010)	1	30 individuals over 4 generations (10 affected, 20 unaffected)	16q22.1 (*HCHOLA4*)	57 (out of 154) genes sequenced
Cenarro A., et al. (2011)	3	Family 1: 21 individuals over 3 generations (10 affected, 11 unaffected)	8q24.22	no genes sequenced
		Family 2: 9 individuals (5 affected, 4 unaffected)	No linkage detected	
		Family 3: 6 individuals (4 affected, 2 unaffected)	No linkage detected	
Wang X., et al. (2011)	1	17 individuals over 4 generations (11 affected, 6 unaffected)	3q25.1-26.1 and 21q22.3 (possible multiplicative interaction between the loci)	*ABCG1* and *LSS* sequenced (out of 92 and 113 genes, respectively)

Analysis of loci identified by genome wide association studies (GWAS) for lipid traits might also help to explain some of the FH cases. In 2010, GWAS meta-analysis of over 100,000 individuals reported association of 95 loci with lipids in European population (Teslovich et al. 2010). Associated loci are shown in Figure 15. This was followed by a larger study of over 180,000 individuals using genome-wide and custom genotyping arrays (Willer et al. 2013a). Single Nucleotide Polymorphisms (SNPs) located in the regions of *LDLR*/*APOB*/*PCSK9*/*LDLRAP1* showed significant association with LDL-C, thus one could hypothesise that other genes, within which common variants were having a moderate effect on LDL-C, may contain rare FH-causing variants. To date, no reports on the analysis of rare variants in the GWAS hits for the LDL-C have been published.

Figure 15

Venn diagram of genes associated with lipids as reported in meta-GWAS on over 100,000 individuals (Teslovich et al. 2010).



1.12. Polygenic Hypercholesterolaemia

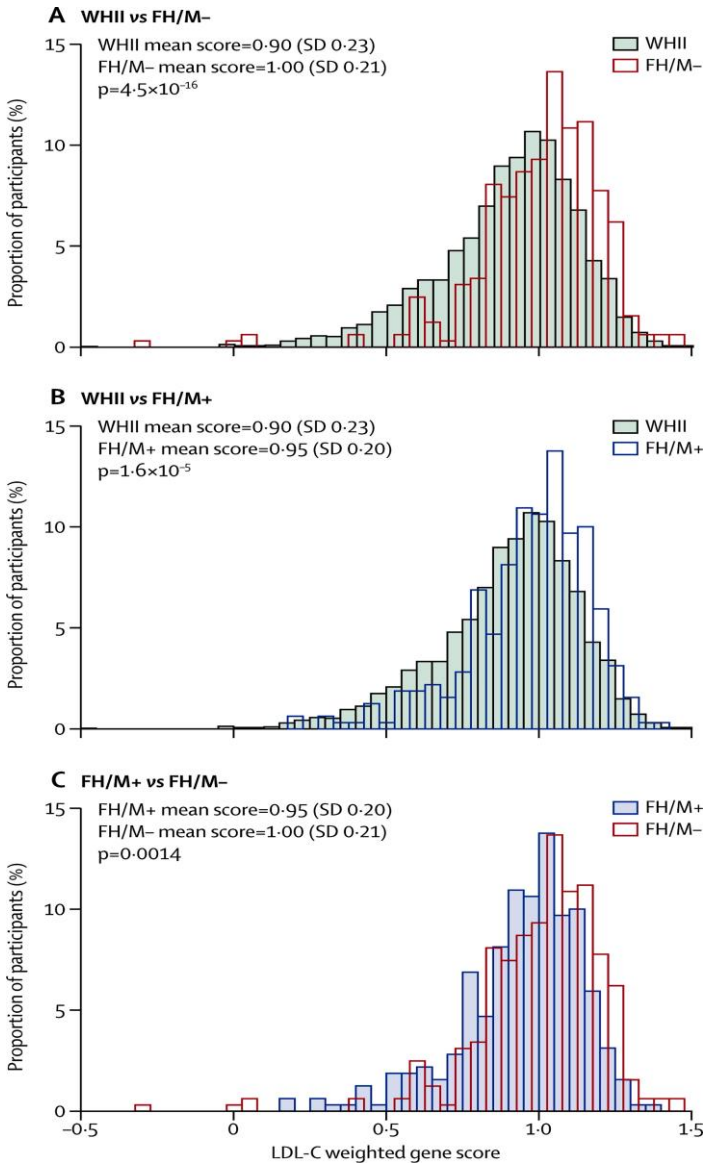
Polygenic hypercholesterolaemia is thought to be a condition of uncertain aetiology. It may occur in the absence of a positive family history (sporadic), it may be associated with a familial component of unclear mode of inheritance or it may be interpreted as being the result of the interaction of multiple genes with a small effect (polygenic). High cholesterol may also result from one or more environmental factors (e.g. high saturated fat/cholesterol diet, obesity, pregnancy) interacting with a genetic predisposing factor or susceptibility gene (multifactorial).

The Global Lipid Genetic Consortium (GLGC) meta-analysis of genome-wide association studies identified several loci where common variants affect LDL-C concentration (Teslovich et al. 2010) and results of another study (Talmud et al. 2009) showed that a proportion of individuals carrying several LDL-C-raising single nucleotide polymorphisms (SNPs) have LDL-C concentrations that exceed the diagnostic LDL-C threshold for FH of 4·9 mmol/l. Using this information a recent report demonstrated that individuals clinically diagnosed with the familial form of hypercholesterolaemia are likely to express the same phenotype due to polygenic causes (Talmud et al. 2013). Authors selected only the lead SNP from each locus reported by Teslovich et al., and if a SNP was associated with more than one lipid fraction, they included it if LDL-C was the lead trait for the SNP. They composed a score using the weighted sum of the LDL-C-raising allele for each of the 640 (321- mutation negative, 319-mutation positive) studied FH patients and for 3,020 controls from the Whitehall II study (WHII). The major observation was that the mutation

negative patients with clinical diagnosis of FH had significantly higher SNP score than the control population (Figure 16(A)). They also found that mutation positive FH patients had a higher SNP score than the controls (Figure 16(B)), which may explain heterogeneity in LDL-C cholesterol between patients with the same mutation. The difference between the mutation negative and mutation positive FH patients was also significant (Figure 16(C)). High SNPs score explained hypercholesterolaemia in at least 20% of patients with no identified mutation. The overlap of clinical features of both forms of hypercholesterolaemia highlights the importance and clinical utility of genetic testing for FH, discussed in section 1.7.

Figure 16

Distribution of weighted LDL-C gene scores(A) Whitehall II controls (WHII) versus patients with familial hypercholesterolaemia without a known mutation (FH/M–). (B) WHII controls versus patients with familial hypercholesterolaemia with a known mutation (FH/M+). (C) FH/M– versus FH/M+. In both patients with definite familial hypercholesterolaemia (DFH) and those with possible familial hypercholesterolaemia (PFH), patients without a detected mutation had a significantly higher mean weighted LDL-C gene score than did those with a detected mutation (taken from (Talmud et al. 2013)



II. CHAPTER TWO – Materials and Methods

2.1. Patient selection criteria

2.1.1. Oxford FH

The Oxford FH cohort comprised individuals who attended sequentially the Oxford Lipid Clinic, in England over the period 2009-2011. All participants were Caucasian, aged 18 or over, and were diagnosed with either definite FH (DFH) or possible FH (PFH) using the Simon Broome clinical diagnostic criteria ('Risk of fatal coronary heart disease in familial hypercholesterolaemia. Scientific Steering Committee on behalf of the Simon Broome Register Group' 1991, Neil et al. 2004), or as having unclassified hypercholesterolaemia (UH) which was defined as a total cholesterol and/or LDL-C concentration above the Simon Broome criteria cut off (respectively >7.5 mmol/l and/or >4.9 mmol/l) but with no family history of early CHD or with no such family history that could be elicited. The Simon Broome diagnostic criteria for FH exclude subjects with a triglyceride level of > 4.5mmol/l and none of the patients exceeded this level. There were a total of 289 patients in the cohort, of which 272 probands were apparently unrelated.

The Simon Broome British Heart Foundation study (SBBHF) of 409 individuals was used for the replication of the FH clinical diagnosis methods comparison between the Simon Broome FH criteria and the Dutch Lipid Clinic Network (DLCN) score, and for the testing of the mutation detection association with TC and TG quartiles. This was a cross-sectional comparison of white patients aged 18 years or more with treated DFH with and without

clinically documented CHD recruited from clinics in London, Oxford and Manchester. Recruitment methods, inclusion and exclusion and diagnostic criteria have been described previously (Neil et al. 2004). The cohort consisted of 328 FH-mutation positive (FH/M+) and 81 FH-mutation negative (FH/M-) patients.

2.1.2. South-Asian FH

Blood samples of FH patients with South-Asian origin were collected across several lipid clinics in England, including the Royal Free Hospital and the Newham Hospital in London, the University Hospital of Coventry and Warwickshire, and the Leeds Teaching Hospital, over the period of three years (2010-2013). In total, 34 samples (33 probands) were collected, of whom four presented with or had a family history of tendon xanthomas. The Simon Broome diagnostic criteria LDL-C cutoff of 4.9 mmol/l for adults was lowered to 4.2 mmol/l.

In addition to the samples collected in the UK, eight members of a consanguineous hypercholesterolaemic family from Kokilaben Dhirubhai Ambani Hospital in Mumbai, India, were studied. The extreme TC and LDL-C levels observed in the family proband - nine years old boy, were suggestive of homozygous FH.

2.1.3. Whole exome sequencing cohort

The whole exome sequencing cohort was comprised of 125 unrelated patients diagnosed using the UK Simon Broome criteria as DFH on the basis of the presence or family history of tendon xanthomas. The majority of patients (n=65) were originally from the Simon Broome British Heart Foundation study (SBBHF), which used the High Resolution Melting (HRM) method to screen the entire promoter and coding region, including splice sites, of the *LDLR* gene (Whittall et al. 2010b). SBBHF patients were tested for presence of the *APOB* mutation, p.(R3527Q), using a restriction enzyme digest (Mamotte and van Bockxmeer 1993), and the entire coding region of the *PCSK9* was examined by HRM (Humphries et al. 2006b). Fragments with a shifted melting curve were analysed further by direct sequencing. Screening for large rearrangements within the *LDLR* gene was done using the MLPA SALSA P062 *LDLR* kit from MRC-Holland (Amsterdam) (Taylor et al. 2009). The final cohort also included 14 DFH patients with no detectable mutation selected from the Oxford FH cohort, studied in Chapter Three. The FH mutation screening methods were as described in the 'Molecular genetic analysis' section.

The remaining 45 samples sequenced in this study were provided by several collaborators. These included: 16 from Australia (PI: Frank Van Bockxmeer), 10 from Israel (PI: Eran Leitersdorf), 10 from Nothern Ireland (PI: Colin Graham), five from Great Ormond Street Diagnostic Laboratory (PI: Nicholas Lench), three from Modena in Italy (PI: Sebastiano Calandra) and one from Cairo Montenotte in Italy (PI: Stefano Bertolini). The initial FH mutation

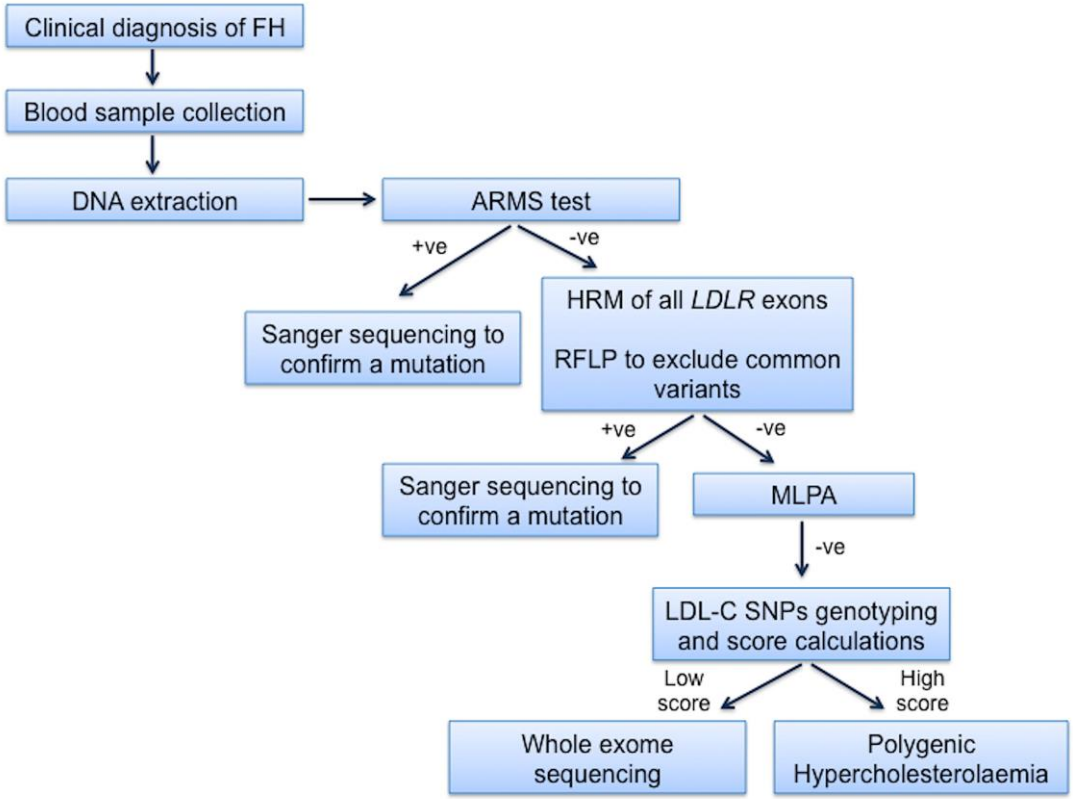
screening methods varied between the labs and are summarised in Appendix Table I.

2.2. Molecular genetic analysis

The FH mutation testing workflow overview is shown in Figure 17.

Figure 17

The FH mutation screening overview



2.2.1. Whole blood DNA extraction

The whole blood DNA extraction method was adapted from Miller et al. "salting-out" method (Miller et al. 1988)

Solutions:

Reagent A: 0.32M sucrose, 5mM $MgCl_2$, 10mM Tris-HCl pH 7.5, 1% Triton-X-100, in deionised water.

Reagent B: 10mM Tris-HCl pH 8.2, 0.4M NaCl, 2mM Na_2EDTA pH 8.0, 10X 1% SDS, in deionised water.

TE Buffer: 10mM Tris, 1mM EDTA, in deionised water, pH7.6

Blood samples were collected into 4.5 ml tubes, containing anticoagulation agent EDTA.

Cell Lysis: 3-5 ml of blood was thoroughly mixed with 20 ml of ice-cold reagent A, and centrifuged at 10,000 rpm (13,000 g) for 10 minutes, at 4°C. The supernatant was discarded into a waste tube without disturbing the pellet. The pellet was resuspended and washed in 20 ml of 'reagent A', and the centrifugation step repeated.

Nuclear Lysis: The pellet was resuspended in 2 ml of 'reagent B'.

Deproteinisation: 1 ml of 5 M sodium perchlorate was added, mixed thoroughly and incubated on a shaker for 15 minutes.

Extraction: 2 ml of ice-cold chloroform was added, mixed and centrifuged at 10,000 rpm (13,000 g) for 3 minutes, to separate the DNA into upper aqueous phase.

Precipitation: The aqueous phase was transferred, without disturbing the organic phase, to a fresh 30 ml polypropylene tube. 10 ml of ice-cold 100% ethanol was slowly added to the tube and incubated on a bench for 3 minutes. The tube was then inverted several times to precipitate the DNA.

Washing: Precipitated DNA was removed from the solution with a sterile Pasteur tip, washed briefly in 70% ethanol, and placed in a nuclease-free tube containing 0.5-1.0 ml of TE buffer (pH 7.6). Samples were then incubated overnight at 37°C to dissolve.

Concentration and purity of each sample were measured using the Nanodrop ND8000 spectrophotometer, supplied by Labtech International. The ratios of absorbance at 260/280 nm and 260/230 nm, indicating the purity of a sample, were accepted at values 1.8-2.0 and 1.8-2.0, respectively.

DNA was standardised to concentration of 30ng/µl and stored at 4°C.

2.2.2. Amplification Refractory Mutation System (ARMS)

Patients' DNA were first screened for the 20 most common UK mutations with a commercially available Elucigene^{TM} FH20 (Gen-Probe Life Sciences, UK) Amplification Refractory Mutation System (ARMS) kit. The ARMS assay is based on a principle of multiplex allele-specific PCR, allowing to test for the presence of 18 of the most common *LDLR* mutations, one *PCSK9* (p.(Asp374Tyr)) and one *APOB* (p.(Arg3527Gln)) mutation in only three reactions. The full list of the mutations is shown in Table 8.

Table 8

The list of FH mutations tested by each ARMS primer mix (*LDLR* residue numbering includes the signal peptide).

PRIMER MIX A	PRIMER MIX B	PRIMER MIX C
p.D374Y (*PCSK9*)	p.R3527Q (*APOB*)	p.W87G
p.P685L	p.IVS3+1G>A	p.V429M
p.L479P	p.D482H	p.D227E
p.R350X	p.G218del	p.C677R
p.E228X	p.D227GfsX12	p.K311RfsX20
p.D221G	p.Q384X	p.C184Y
p.E101K		p.D482N

Reaction set-up

Enzyme Dilution (for 10 reactions, for each Primer Mix: A, B, and C):

21 µl sterile deionised H_2O

6 µl dilution buffer

30 µl loading dye

3 µl AmpliTaq Gold

Reaction Mix (for 10 reactions):

55 µl enzyme dilution

165 µl TA/TB or TC primer mix

Loading a PCR plate:

20 µl reaction mix (TA, TB or TC)

2 µl 30ng/µl DNA

3 µl sterile deionised H_2O

Cycling conditions on a standard thermocycler

94°C for 20 minutes (Taq activation)

35 cycles of:

94°C for 30 seconds (denaturation)

61°C for 1 minute (annealing)

72°C for 1 minute (elongation)

72°C for 30 minutes (final elongation)

Agarose gel electrophoresis

Visualisation of PCR products was achieved with 2% Agarose gel electrophoresis (120V for 60 minutes), with addition of Ethidium Bromide (5g/l), in 1 x TBE buffer (1x, 18mM Tris-borate and 4mM EDTA, pH8). Loading dye for PCR products was Bromophenol Blue.

2.2.3. High Resolution Melting (HRM)

High Resolution Melting (HRM) is a post-PCR method, based on thermal denaturation of double stranded PCR products. It enables the detection of genetic variations (including small deletions or insertions) by comparing melting patterns of DNA fragments. HRM was shown to be a sensitive and

robust screening method for FH mutations, which is able to sense all four classes of DNA base pair changes (Whittall et al. 2010b, Liew et al. 2004), although the class four DNA changes (A>T or T>A variants) were shown to be the most difficult to detect (Whittall et al. 2010b).

Reaction set-up

Most of the HRM reactions in this project were performed using AccuMelt™ HRM SuperMix (Quanta Biosciences). Each 20 µl reaction contained:

10 µl AccuMelt HRM SuperMix (containing the Syto9 saturating dye at 2X concentration)

0.8 mM Forward/Reverse Primer

2 µl DNA template at 30 ng/µl

Nuclease-free H_2O up to 20 µl

Exon 4.4 was amplified using LightScanner® Master Mix (BioFire Diagnostics, Inc.), which incorporated the LCGreen® saturating dye formulated at 2.5X concentration. Each 10 µl reaction contained:

4 µl Master Mix (2.5X)

0.8 mM Forward/Reverse Primer

2 µl DNA template at 30 ng/µl

Nuclease-free H_2O up to 10 µl

Cycling conditions

HRM was performed on Rotor-Gene^{6000} realtime rotary analyser. Reaction cycling conditions and HRM temperature gradient are presented in Table 9. Primer sequences for each fragment are shown in Appendix Table II.

Table 9

A summary of PCR cycling conditions and the HRM melting temperature range for the *LDLR* promoter, each *LDLR* exon, *APOB* exon 26 and *PCSK9* exon 7.

Gene (Exon)	Enzyme activation	Amplification conditions	N° of cycles	HRM T Gradient
LDLR (Promoter, 1, 3, 4.1, 4.2, 4.3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18)	95°C 5 min	95°C 5 sec 60°C 10 sec 70°C 20 sec	40 x	80-94°C
LDLR (2)	95°C 5 min	95°C 5 sec 55°C 10 sec 70°C 20 sec	40 x	80-94°C
LDLR (4.4)	95°C 5 min	95°C 10 sec 66°C 10 sec 72°C 20 sec	45 x	80-94°C
LDLR (12)	95°C 5 min	95°C 10 sec 56°C 20 sec 70°C 30 sec	40 x	80-94°C
APOB (26)	95°C 5 min	95°C 5 sec 60°C 10 sec 70°C 20 sec	40 x	85-95°C
PCSK9 (7)	95°C 5 min	95°C 10 sec 66°C 10 sec 72°C 20 sec	45 x	80-94°C

2.2.4. Restriction Fragment Length Polymorphism (RFLP)

Restriction Fragment Length Polymorphism (RFLP) was used to genotype commonly occurring polymorphisms in *LDLR* exons. Table 10 contains a list of all common *LDLR* SNPs, restriction enzymes, and amplification primers used for the genotyping.

PCR amplification of each fragment was performed in 30 µl reaction:

		Volume	Concentration
30ng/µl	DNA	2 µl	60ng/30µl
10x	Polmix	3 µl	1x
100mM	dNTPs	0.24 µl	0.8 mM
100pmoles/µl	Primers	0.12/0.12 µl	12 pmol/30µl
50mM	Mg2+	1.2 µl	2 mM
5u/µl	Taq	0.12 µl	0.6 units
	H2O	25.2 µl	

Cycling conditions on a standard thermocycler

Heated Lid

96^{o}C for 3 minutes

35 cycles of:

94^{o}C for 30 seconds

60°C for 40 seconds*

65°C for 10 minutes

72°C for 30 minutes

16°C for 2 minutes

*****Exon 2 annealing temperature was 57°C

PCR products were checked on 2% agarose gel.

Table 10

Commonly occurring *LDLR* SNPs, the RFLP primers and enzymes used to distinguish between the alternative alleles.

***LDLR* exon**	**SNP ID**	**SNP position**	**Enzyme used for RFLP**	**Primer direction**	**Amplification Primers (5' to 3')**	
2	rs2228671	c.81C>T	HhaI	Forward	TGATCGGATGACATTTCTGG	*forced site
				Reverse	GGCACTGGAACTCGTTTCTTGC	
7	rs12710260	c.1060+10C>G	SmaI	Forward	GGCGAAGGGATGGGTAGGGG	
				Reverse	CCCCTACCCATCCCTTCGCC	
8	rs11669576	c.1171G>A	StuI	Forward	CATTGGGGAAGAGCCTCCCC	
				Reverse	GCCTGCAAGGGGTGAGGCCG	
$10^{5'}$	rs5930	c.1413G>A	BsmAI	Forward	AGATGAGGGCTCCTGGTGC	
				Reverse	GCCCTTGGTATCCGCAACAGAGACA	
11	rs5929	c.1617C>T	AciI	Forward	TCCTCCCCCGCCCTCCAGCC	
				Reverse	CTGGGACGGCTGTCCTGCG	
12	rs1799898	c.1725C>T	BsmAI	Forward	CTCCTCAGTGGCCGTCT	*forced site
				Reverse	ATCCGCCACCTAAGTGCTTG	
12	rs688	c.1773C>T	HincII	Forward	GCACGTGACCTCTCCTTATCCACTTG	
				Reverse	CACCTAAGTGCTTGCATCTCGTACG	
13	rs5925	c.1959C>T	AvaII	Forward	GTCATCTTCCTTGCTGCCTGTTTAG	
				Reverse	GTTTCCACAAGGAGGTTTCAAGGTT	
15	rs5927	c.2232A>G	MspI	Forward	GAAGGGCCTGCAGGCACGTGGCACT	
				Reverse	CTAGGGAGGGCCCAGTCTTT	

Restriction enzyme digest (13 µl)

		Volume	Concentration
	PCR product	8 µl	
10x	Buffer	1.3 µl	1x
	Restriction Enzyme	As required	3u
100x	BSA	0.13 µl	1x
	H20	Make up to 13 µl	

The incubation temperature and time were dependent on the enzyme.

2.2.5. Microtitre Array Diagonal Gel Electrophoresis (MADGE)

The DNA fragments produced by restriction enzyme digest were separated using electrophoresis on a non-denaturing polyacrylamide gel, using Microtitre Array Diagonal Gel Electrophoresis (MADGE). Utilisation of this technique makes it possible to electrophorese the entire 96 wells of a standard PCR plate on a single gel, by allowing the samples to run diagonally. Use of MADGE was an essential part of the project as it allowed high throughput screening of the PCR product, and allowed the 96 well DNA array format to be retained throughout the screening process (Day et al. 1996). Standard 7.5% MADGE gels were made up in batches for economy of time.

MADGE consists of an open arrangement of 8x12 wells each 2mm deep. The wells are arranged at an angle of 71.2^{o} to the short axis of the array, but perpendicular to

the long-axis of the Perspex formers used. Thus the maximum track length is 26.5mm allowing sufficient travel for genotype resolution.

Glass plates of appropriate size (160 x 100 x 2mm) were rigorously cleaned and hand dried. Once dry, 5 drops of γ-methacryloxypropyltrimethoxysilane ("sticky" silane) were spread across the plates and left to air-dry. Silane was used to ensure that the MADGE gel would adhere to the glass plates.

Gel solution

5mls of 10xTBE (10x Tris-Boric acid-Ethylenediaminetetraacetic acid contained: 0.9M Tris, 0.9M Ortho-boric acid, 0.2M ethylenediaminetetraacetic acid)

12.2mls 30% acrylamide-bisacrylamide (in a ratio of 19:1)

32.5mls of distilled dH_2O

150μl N'-tetramethylethylenediamine (TEMED)

150μl 25% ammonium per sulphate (APS)

Polymerisation was initiated by the addition of APS, which was added last (as the combined solution begins to set within thirty seconds of mixing) before the solution was mixed and quickly poured into the three-dimensional former. A glass plate was then gently placed over the mould (silanised side facing downwards) taking care not to trap any air bubbles. This was then left for fifteen minutes to set, using a small weight to ensure that the glass did not slip whilst the gel was setting. Excess gel was trimmed from the edges of the MADGE former before the glass plate and attached gel were then prized

away from the plastic former. MADGE gels were stored in a plastic Stuart box containing neat TBE (1x concentration).

Gel staining and loading

Prior to loading a gel with digested PCR product, each gel was stained with Ethidium Bromide (EtBr). This was achieved by placing them individually in a Stuart box, shielded from direct light, containing 100ml of 1x TBE and 10μl EtBr for 20 minutes.

Whilst the gels were being stained, the PCR-digest product was prepared for loading onto the MADGE gel. 2μl of formamide dye (98% formamide, 10mmol/l EDTA, 0.025% xylene cyanol, 0.025% bromophenol blue) was added to each well of a new, round-bottomed, loading tray, followed by 5μl of each digested sample, using a multi-channel pipette to pick up the samples from the microtitre plates. After placing the stained MADGE gel into an electrophoresis tank containing 1000ml of 1xTBE, a multi-channel pipette was then used to transfer 5μl of this digest/dye mixture to the wells of the gel. At all times the samples were kept in the same layout as on the PCR tray, allowing each sample to be easily identified without being re-labelled. The gel was electrophoresed at 150 volts for 40 minutes.

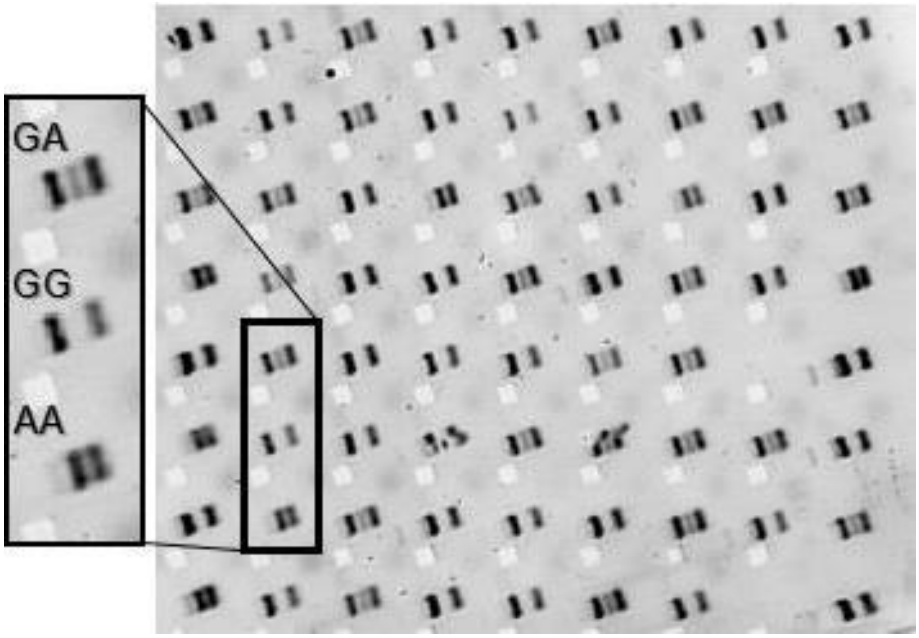
Following electrophoresis the gel was viewed and photographed under ultraviolet light using the UVP Gel Documentation System. Care was once more taken to ensure the correct orientation of the MADGE under UV. This was ensured by always placing the well corresponding to grid-reference “A1” in the bottom left-hand corner of the UV-viewing box. The image produced was used for genotyping.

Genotyping quality control

All genotyping was performed in a double blind fashion using both positive and negative controls. Two individuals rechecked the results at the time of MADGE imaging and during data entry into the computer database. Any apparent genotype differences were resolved by repeat PCR. An example of RFLP and MADGE gel results are shown in Figure 18.

Figure 18

A typical MADGE gel image demonstrating BsmAI RFLP digests of rs5930 polymorphism located in exon 10 *LDLR*. Different fragment lengths indicate different genotypes of the SNP.



2.2.6. PCR product purification

Prior to the Sanger sequencing of fragments that showed a shift in the melting during the HRM, PCR products were purified to remove any remaining enzyme or dye from the reaction mix. The DNA purification was performed using the Illustra GFX PCR DNA and Gel Band Purification kit from GE Healthcare Life Sciences. Briefly, the PCR volume was mixed with 500 µl of the Capture Buffer (type 3) and loaded onto the assembled GFX MicroSpin column and collection tube. The sample was then centrifuged at 16,000 x g for 30 seconds. The flow through was discarded and the column was placed back inside the collection tube. Next, 500 µl of the Wash Buffer (type 1) containing 70% ethanol was added onto the column and the centrifugation at 16,000 x g for 30 seconds was repeated. The collection tube was discarded and the column was placed on a fresh DNase-free 1.5 ml microcentrifuge tube. The PCR product was eluted from the column with 30 µl of sterile nuclease-free water by another centrifugation step, 16,000 x g for 1 minute. DNA concentration of the eluted product was measured on the NanoDrop ND8000 spectrophotometer and aliquoted to 1 ng/µl per 100 bp of product in volume of 15 µl, for Sanger sequencing.

2.2.7. Sanger sequencing

Every variant detected by ARMS and/or a melting curve shift on HRM was confirmed by standard Sanger sequencing (in the forward and reverse

direction), service provided by Source Bioscience LifeSciences. Primers for the fragment amplification and sequencing are shown in Appendix Table X. All produced sequencing reads were then analysed manually on FinchTV (Geospiza Inc.).

2.2.8. Multiplex Ligation-dependent Probe Amplification (MLPA)

Multiplex Ligation-dependent Probe Amplification (MLPA) SALSA P062-C1 LDLR kit (MRC-Holland, Amsterdam) was used to screen for large deletions and insertions in the *LDLR* gene of definite FH patients (Taylor et al. 2009). The kit contains probes for all 18 *LDLR* exons, producing fragments of different length, as well as 15 reference probes located on different chromosomes. The experiment requires DNA samples to be aliquoted to standard concentration, which in this case was 60 ng/µl.

DNA denaturation

5 µl of each 60 ng/µl DNA sample was loaded into 96-well AB-0800 plate. DNA was then denatured on thermocycler for 5 minutes at 98°C (with a heated lid) and cooled to 25°C.

Hybridisation

3 µl of the hybridisation mix containing (per reaction):

1.5 µl MLPA buffer

1.5 µl probemix

was loaded into each denatured sample and mixed by pipetting up and down. Reactions were incubated for 1 minute at 95°C, and then for 16-20 hours at 60°C.

Ligation

Ligase mix contained (per reaction):

3 µl Ligase-65 buffer A

3 µl Ligase-65 buffer B

25 µl dH_2O

1 µl Ligase-65

32 µl of the ligase mix was mixed into each reaction during a pause at 54°C on the thermocycler. This step was then followed by 15 minutes incubation at 54°C (ligation), and 5 minutes heat inactivation of Ligase enzyme at 98°C. Completed ligation reactions were cooled down to 15°C.

Amplification

PCR buffer mix (per reaction):

4 µl SALSA PCR buffer

26 µl dH_2O

30 µl of the PCR buffer mix was loaded into each reaction well on a new plate and mixed with 10 µl of ligation product. The plate was then incubated at 60°C on a thermocycler and each reaction was mixed with 10 µl of Polymerase master mix, which contained (per reaction):

2 µl SALSA PCR-primers

2 µl SALSA enzyme dilution buffer

5.5 µl dH_2O

0.5 µl SALSA Polymerase

PCR cycling conditions:

60°C pause

35 cycles:

95°C for 30 seconds

60°C for 30 seconds

72°C for 60 seconds

72°C for 20 minutes

15°C pause

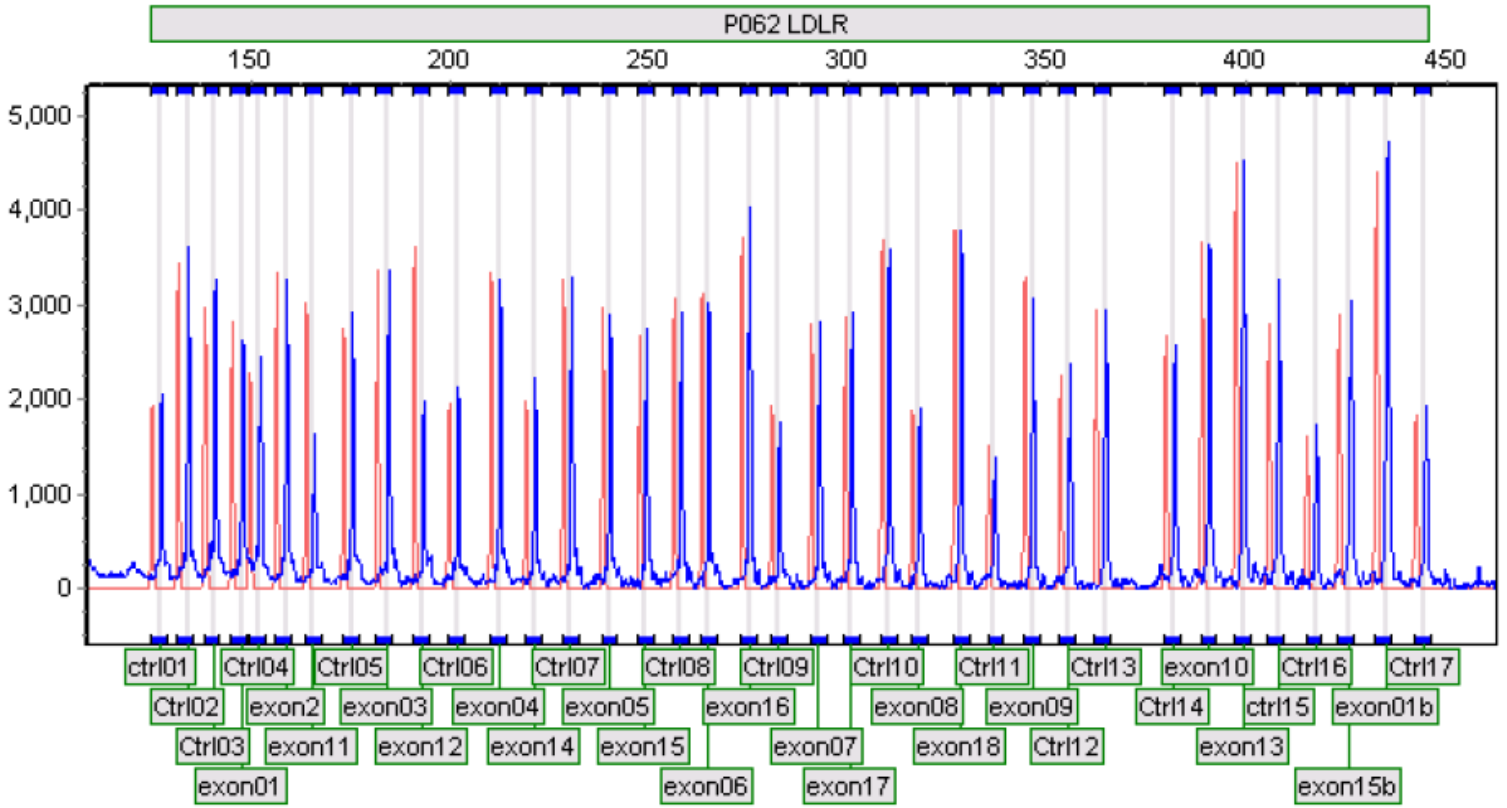
Fragment separation

1 µl of each PCR product was mixed with 12.5 µl of Hi-Di Formamide and 0.5 µl of Size Standard. Fragments were separated on ABI-3730XL with 36 cm capillary length.

Results, normalized with synthetic control, were analysed on GeneMarker® software. A typical MLPA result graph is shown in Figure 19.

Figure 19

A typical MLPA results can be evaluated by comparing a peak patterns of separated fragments between the tested samples (vlue peaks) and the reference data (red peaks). Each peak represents a probe for a specific region of *LDLR* or for a control (located at outside of the gene). Probes are labeled in grey boxes under the graph.



2.3. Global Lipids Genetic Consortium gene score calculation

The GLGC reported SNPs that were significantly associated with LDL-C, with a p-value cut-off of less than $5x10^{-8}$ (Teslovich et al. 2010). For the purpose of the gene score calculation only the lead SNP from each locus was selected and if a SNP was associated with more than one lipid fraction it was only used in this score calculation if LDL-C was the lead trait (most strongly associated trait) for the SNP. For each individual, the LDL-C-specific gene scores were calculated using the weighted sum of the risk allele (i.e. the LDL-C-raising allele). The weights used were the corresponding per-risk-allele beta coefficients reported by the GLGC. Although the GLGC consortium included healthy subjects and CAD cases, any individuals taking statins were excluded from the estimates of the effect size associated with the SNPs. Since the beta-coefficients reported by GLGC are for the effect of each minor allele, where the effect was LDL-lowering, the other allele was considered the risk allele and the absolute effect size was used as the weight. *APOE* is a major determinant of LDL-C levels and the effect of the *APOE* haplotype was included in the LDL score calculation. For *APOE* the previously reported effect of the *APOE* haplotypes on LDL-C from a meta-analysis in 61,463 healthy participants was used (Bennet et al. 2007). Based on the latter study, using **ε**3/**ε**3 individuals as reference, the *APOE* haplotype was scored as follows: **ε**2**ε**2=-0.9, **ε**2**ε**3=-0.4, **ε**2**ε**4=-0.2, **ε**3/**ε**3=0, **ε**3**ε**4=0.1 and **ε**4**ε**4=0.2. A total of 10 SNPs plus two *APOE* SNPs were used for the LDL-C score calculation, shown in Table 11.

Table 11

The top LDL-rising SNPs and their effects (as reported by the GLGC) used for the LDL-C gene score genotyping and calculation.

SNP ID	Nearest gene	Risk Allele	Beta coefficient (mmol/l)
rs2479409	*PCSK9*	G	0.051978278
rs629301	*CELSR2*	T	0.146108094
rs1367117	*APOB*	A	0.104732351
rs4299376 (rs6544713)	*ABCG8*	T	0.071114559
rs3757354	*MYLIP*	C	0.036979571
rs1800562	*HFE*	G	0.057408844
rs1564348	*SLC22A1*	T	0.01448151
rs4055111 (rs11220462)	*ST3GAL4*	G	0.050426687
rs8017377	*NYNRIN*	A	0.029480217
rs6511720	*LDLR*	G	0.180760279

2.3.1. LDL-rising SNPs genotyping

2.3.1.1. KASP genotyping

The KASP (Kbiosciences, UK Hoddesdon, Herts, UK) genotyping was used for genotyping of all SNPs shown in Table 11. Probes for two of the SNPs, rs4299376 (in the *ABCG5*/*ABCG8* region) and rs11220462 (in the *ST3GAL4* region), could not be designed because of an occurrence of other polymorphisms in the surrounding region, therefore proxies (SNPs in high Linkage Disequilibrium (LD)) were used instead, rs6544731 and rs4055111, respectively.

The method utilises a form of competitive allele-specific PCR. Bi-allelic discrimination was achieved by the competitive binding of two allele-specific primers, each with a unique tail sequence that corresponded with two universal FRET (fluorescence resonant energy transfer) cassettes. Each allele-specific primer was labelled, one with FAM dye and the other with HEX dye, which enabled the discrimination between alleles. The reverse primer was universal (common) for both forward allele-specific primers.

The genotyping reactions were carried out in a 384-well plate format. 5 ng of DNA was transferred from a standardized array, using the Biomek 2000 Laboratory Automation Workstation, into a 384-well plate and centrifuged at 3000rpm for 1 minute before being air-dried. Each plate contained at least two no-template-control (NTC) wells.

KASP Master Mix (5 µl per each reaction):

2.5 µl KASP Reaction Mix (2X)

0.07 µl KASP-by-Design (KBD) assay

2.5 µl dH_2O

Cycling conditions on a standard thermocycler:

94°C for 15 minutes (hot-start enzyme activation)

10 cycles (dropping 0.8°C per each cycle):

94°C for 20 seconds

65-57°C for 60 seconds

26 cycles:

94°C for 20 seconds

57°C for 60 seconds

2.3.1.2. TaqMan Genotyping

The *APOE* **ε**2/**ε**3/**ε**4 haplotype is made up of two SNPs (rs429358 and rs7412) and resulting in different isoforms of the ApoE protein, is an important genetic determinant of LDL-C levels. Off-the-shelf TaqMan probes (Life Technologies, Carlsbad, California, US) for the *APOE* genotyping were available.

Each ABI TaqMan Genotyping Assay (40x) contained two target SNP specific primers for the region of interest and two allele-specific TaqMan MGB probes. Each allele specific probe has three main elements: a reporter dye (at the 5' end), a minor groove binder (MGB), and a non-fluorescent quencher (NFQ, at the 3' end). Two commonly used reporter dyes, VIC and FAM, enabled to distinguish between each allele. The MGB was added in order to increase the

melting temperature for a given probe length, which increases the efficiency of allelic discrimination. The NFQ does not fluoresce and will quench the fluorescent signal in any reporter dye linked to the 5' end for as long as the probe remains intact.

ABI TaqMan Genotyping reactions were carried out in a 384-well plate format. 5 ng of DNA was transferred from a standardized array, using the Biomek 2000 Laboratory Automation Workstation, into a 384-well plate and centrifuged at 3000rpm for 1 minute before being air-dried. Each plate contained at least two no-template-control (NTC) wells.

The TaqMan genotyping reaction was carried out as per the TaqMan Genotyping mastermix manufacturer's instructions. Each plate was centrifuged at 3000rpm for 1 minute and PCR reactions were carried out using the 7900HT Fast Real-Time PCR system.

Cycling conditions:

95°C for 10 minutes (AmpliTaq Gold activation)

40 cycles:

92°C for 15 seconds

60°C for 60 seconds

2.3.1.3. Genotyping signal detection

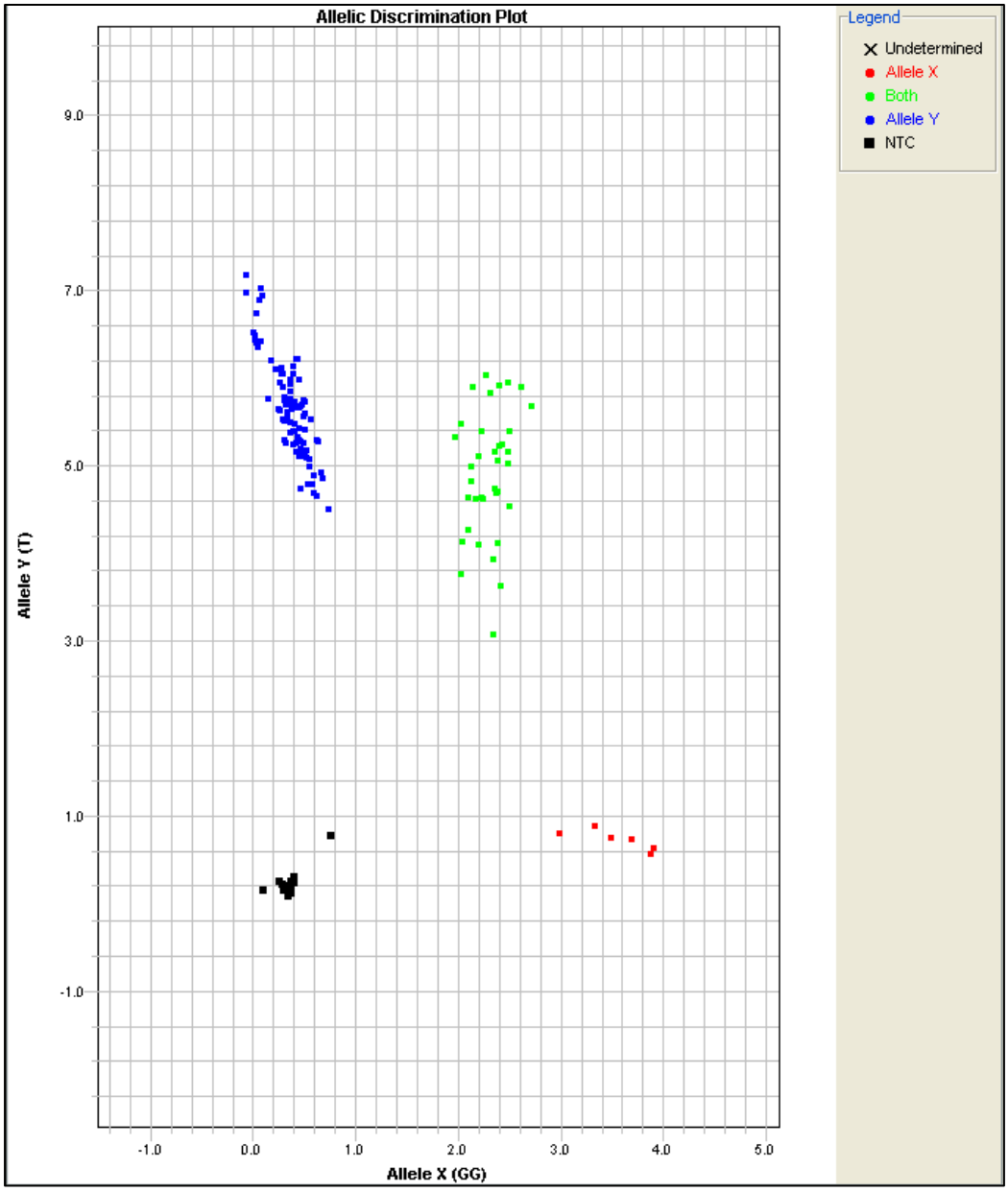
The standard allelic discrimination program assigned genotypes automatically to a plate coordinate in a text output file and in the form of an allelic discrimination plot based on the fluorescent signal detected. The light source

from the cycler excites the reporter dyes by fluorescence resonance energy transfer. The dyes used in KASP or TaqMan genotyping have significantly different excitation and emission wavelengths, which allows them to be differentially detected. An example of the allelic discrimination plot is shown in Figure 20. Each dot on the scatterplot represents one well of the plate (either DNA sample or an NTC). Samples that did not reach the threshold fluorescent signal level for any of the genotypes (undefinable samples) were marked by an ‘X’. Results of each SNP were checked manually by study personnel using SNPviewer software.

The distribution of each genotype frequency was checked using the Hardy-Weinberg equation.

Figure 20

Allelic discrimination plot. This plot represents KASP genotyping results for rs629301A>C SNP in part of the Oxford FH study. Samples with a significant FAM signal are called homozygous for the A (blue cluster). Samples with an even distribution of HEX and FAM signals are heterozygotes (green cluster). Homozygotes for the rare allele are those with a significant HEX signal (red cluster). NTCs are represented by black squares.



2.3.2. Healthy comparison group

A healthy comparison group of white men and women was drawn from the UK Whitehall II study (WHII) (Marmot et al. 1991). This study recruited 10,308 participants (70% men) between 1985 and 1989 from 20 London-based Civil service departments. Clinical measurements were taken at 5-year intervals. Phase 3 (1991-1993) provided the first comprehensive phenotyping and is considered the baseline phase, and lipid measurements from this phase were used in this analysis. Blood samples for DNA extraction were collected from 5066 individuals at a follow-up screening in 2003-2004. From those that had DNA available, 642 subjects who developed either T2D (N=383) or CHD (N=193) or both (N=66) over phase 3-7 were selected and matched with two healthy participants (total 1926), and an additional 1482 subjects were selected at random for genotyping with the Metabochip platform (Illumina Inc, San Diego, CA, USA), a custom Illumina iSelect genotyping array comprising approximately 200,000 SNPs at regions identified by meta-analyses of GWAS for cardio-metabolic endpoints and traits. After removal of samples for quality control purposes (outliers, low call rate, gender ambiguity, cryptic relatedness, non-Caucasians) and for missing data for the SNPs required for generating LDL score, 3020 individuals were used as controls for this analysis. These samples were a subset of a previous random selection of 5059 WHII individuals that were genotyped using the Illumina Cardiochip platform.

2.3.3. Replication cohorts

Several FH cohorts provided by close collaborators were used to replicate the LDL-C gene score findings. The replication cohort included 78 adult FH individuals from Canada (39 patients with an identified FH mutation and 39 mutation negative patients). Two groups of FH children were also studied, which included 130 patients from Greece (70 mutation positive and 60 mutation negative), and 23 patients from Holland (all mutation negative). A cohort of FH patients with South-Asian background (studied in Chapter Four, 26 mutation negative, and 8 mutation positive) was also genotyped for the 12 SNPs.

2.4. Whole exome sequencing

2.4.1. Genomic DNA purification

Prior to the whole exome sequencing each genomic DNA sample was purified using a chloroform-based method. Briefly, 500 µl of each sample was mixed with 20 mg/ml Proteinase-K and incubated for 2 hours at 55°C. 1 µl of RNaseA was then added, followed by an incubation at 37°C for 15 minutes. 510 µl of chloroform was thoroughly mixed with the sample and centrifuged at 16,000 x g for 3 minutes. The DNA-containing top phase was divided (250 µl each) between two fresh 1.5 ml microcentrifuge tubes, whereas the remaining pellet (containing the chloroform) was discarded with care using the hazardous waste facility. 25 µl of NaAc (3M solution) was mixed with each pellet. DNA was then precipitated by adding 825 µl of ice-cold 100% absolute

ethanol and by slowly inverting the tube. DNA was spooled, washed briefly in 70% ethanol and dissolved in appropriate volume of TE buffer (100-250 µl). Genomic DNA samples were aliquoted to concentration of 100 ng/µl in volume of 50 µl. DNA purity was thoroughly assessed by repeated readings on NanoDrop ND8000 spectrophotometer, and only samples with absorbance readings of 260/280 nm = 1.8 and 260/230 = 2.0 were sent for exome sequencing.

Prior to the sequencing, samples went through several quality control steps and cherry-picked for sequence capture.

2.4.2. Library preparation and sequence capture

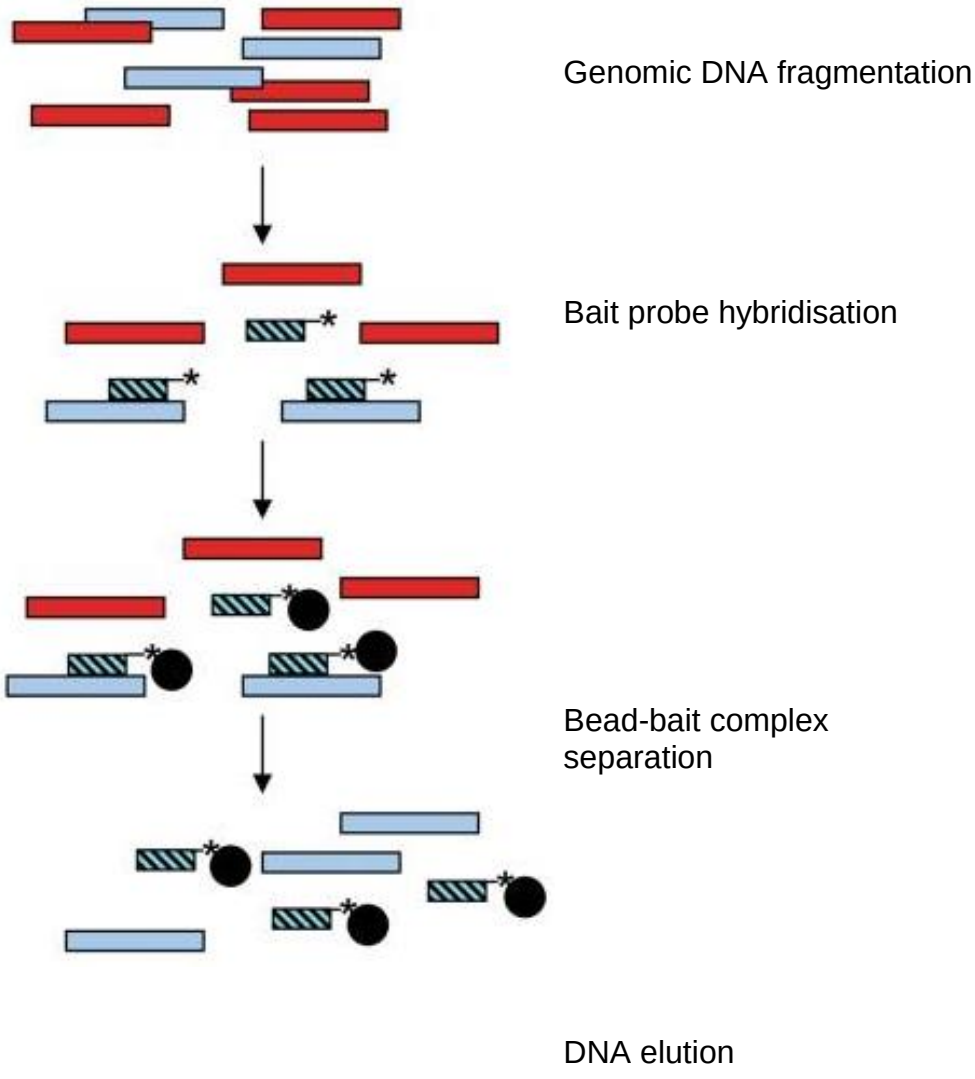
The library preparation and deep sequencing was performed at the Wellcome Trust Sanger Institute in Cambridge, as a part of the UK10K sequencing project (http://www.uk10k.org). The methods used were as below:

DNA (1-3 µg) was shared to 100-400 bp fragments using Covaris E210 or LE220 (Covaris, Woburn, MA, USA). Shared DNA was subjected to Illumina paired-end DNA library preparation and enriched for target sequences (Agilent Technologies; Human All Exon 50 Mb – ELID S02972011) according to manufacturer's recommendations (Agilent Technologies; SureSelectXT Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing). The set of target regions for the human exome was based on the GENCODE project annotations, which provide the most recent human protein-coding region with alternatively transcribed variants, non-coding loci with transcript evidence, and pseudogenes (Coffey et al. 2011). The exome capture was performed by a liquid-phase hybridization method, summarized

in Figure 21. Enriched libraries were sequenced using the HiSeq 2000 platform (Illumina) as paired-end 75 base reads according to manufacturer's protocol.

Figure 21

An overview of the insolution (a liquid-phase) sequence capture. Fragmented genomic DNA was hybridised with biotinylated bait probes (blue/black boxes with an asterisk) in solution. Streptavidin beads (black circles) were used to separate the bead-bait complexes, which were then washed followed by the captured DNA (light blue fragments) elution. Adapted from (Teer and Mullikin 2010).



2.4.3. Whole exome sequencing control cohort

A cohort of 1,926 samples, with no lipid abnormalities, was used as a comparison group for the rare variant association gene-based burden test (see 'Gene-based burden test' section). The control sample consisted of eight groups of different rare diseases (with no abnormal cholesterol metabolism), which were a part of the UK10K project and therefore were sequenced in parallel with the FH cohort using exactly the same methods. Variants called in the controls were processed together with the cases using the same pipeline.

2.4.4. Variant calling and informatics

All the exome sequencing data were produced and annotated together with the controls at the Wellcome Trust Sanger Institute in Cambridge as a part of the UK10K sequencing project. The variant calling pipelines were as follows:

2.4.4.1. Single-sample calling release

To improve raw alignment BAMs for SNP calling, calls were realigned around known (1000 Genomes pilot (Abecasis et al. 2010)) indels, and the base quality scores was recalibrated using GATK. BAQ tags were added using SAMtools calmd. BAMs were merged to sample level and duplicates were marked using Picard. Variants (SNPs and Indels) were called on each sample individually with both SAMtools mpileup (version 0.1.17) and GATK UnifiedGenotyper (version 1.3-21), restricted to exon bait regions plus or

minus 100 bp window. Various quality filters were applied to each of the call sets separately (Table 12). Calls were then merged, giving preference to GATK information when possible. Calls were annotated with 1000 Genomes allele frequencies, dbSNP132 rs IDs and earliest appearance in dbSNP. Functional annotation was added using Ensembl Variant Effect Predictor (version 2.2) against Ensembl 64 and included coding consequence predictions, SIFT, PolyPhen and Condel annotations, and GERP and Grantham Matrix scores.

Table 12

Quality filters settings used for the filtering of the whole exome sequencing calls.

Filters for mpileup variants				
all	DP	<4	MinDP	Minimum depth at locus
all	DP	>2000	MaxDP	Maximum depth at locus
all	MQ	<=25	MQual	Minimum mapping quality
snp	GQ	<=25	GQual	Minimum genotype quality
snp	QUAL	<=30	MinQ	Minimum quality at locus
indel	GQ	<=60	iGQual	Minimum indel genotype quality
indel	QUAL	<=60	iMinQ	Minimum quality at indel locus
Filters for GATK variants				
snp	QUAL	<30	MinQ	Minimum quality at locus
snp	QD	<5	MinQD	Minimum quality by depth at locus
snp	HRun	>5	MaxHRun	Max homopolymer run length
snp	SB	>-0.1	StrBias	Max strand bias
snp	FS	>60	MaxFS	Max Fishers p-value
indel	FS	>200	iMaxFS	Max indel Fishers p-value
indel	QD	<2	iMinQD	Minimum indel quality by depth
indel	ReadPosRankSum	<-20	iMinRank	Minimum Z-score
indel	InbreedingCoeff	<-0.8	iMinInbreed	Minimum inbreeding coefficient

2.4.4.2. Multi-sample calling release

Calls were made from sets of merged BAMs containing multiple samples per BAM. This set of merged BAMs contained the data for all 4,732 exome samples in this UK10K release. A BCF file was created with SAMtools mpileup, calculating genotype likelihoods for every site in the bait (+/-100bp) regions file, then variants (SNPs and Indels) were called by bcftools.
Variant Quality Score Recalibration (VQSR) was used to filter SNPs. The GATK (v1.6-13-g91f02df) UnifiedGenotyper was used to recall the sites/alleles discovered by SAMtools in order to generate annotations to be used for recalibration. The GATK VariantRecalibrator was then used to model the variants, followed by GATK ApplyRecalibration, which assigns VQSLOD (variant quality score log odds ratio) values to the variants.
Variants, based on their VQSLOD score, were assigned to 'truth sensitivity' tranches, e.g. s truth sensitivity of 99% means that with variants above this minimum VQSLOD score, one retains 99% of truth sites.
Each VCF contained the following filters applied to SNPs:
- LowQual,Description="Low quality variant according to GATK (GATK)"
- MinVQSLOD,Description="Minimum VQSLOD score [SNPs:-1.8768, truth sensitivity 99.48]"
All SNP sites which do not fail these filters will be marked as PASS in the VCF.
Indels were left-aligned using GATK LeftAlignVariants. The following soft filters were applied to indels using vcf-annotate:

- StrandBias,Description="Min P-value for strand bias (INFO/PV4) [0.0001]"
- BaseQualBias,Description="Min P-value for baseQ bias (INFO/PV4) [1e-100]"
- EndDistBias,Description="Min P-value for end distance bias (INFO/PV4) [0.0001]"
- GapWin,Description="Window size for filtering adjacent gaps [3]"
- MaxDP,Description="Maximum read depth (INFO/DP or INFO/DP4) [8000000]"
- MinAB,Description="Minimum number of alternate bases (INFO/DP4) [2]">
- MinDP,Description="Minimum read depth (INFO/DP or INFO/DP4) [16000]"
- MinMQ,Description="Minimum RMS mapping quality for SNPs (INFO/MQ) [10]"
- Qual,Description="Minimum value of the QUAL field [10]"
- RefN,Description="Reference base is N []"

The calls were annotated using vcf-annotate. Variant consequence annotations are called using the Ensembl Variant Effect Predictor (http://www.ensembl.org/info/docs/variation/vep/index.html), v2.4 against Ensembl 66. This provides coding consequence predictions and SIFT, PolyPhen and Condel annotations. Ensembl also provides GERP conservation scores. Grantham matrix values come from a simple lookup.

2.4.5. Filtering of variants

Variants were flagged as *somewhat-rare* and *rare* according to their frequency in publicly available databases including 1000 Genomes and NHLBI Exome Sequencing Project (ESP6500) (http://evs.gs.washington.edu/EVS/). The *somewhat-rare* flag was used to filter out all variants with frequency higher than 0.5%, whereas the *rare* flag was more stringent, selecting variants that were novel (i.e. absent from the above databases). In addition to the frequency filters, a *functional* flag was added, which prioritised variants that are most likely to affect a protein's function, i.e. non-synonymous SNVs, stop gain SNVs, stop loss SNVs, frameshift deletions and insertions, and splice site variants.

Variants within Tier 1 genes (*LDLR*, *APOB*, *PCSK9*, and *LDLRAP1*) were assessed on the bases of their frequency in 1000 Genomes and 6500ESP, and manually by looking at their annotations in the UCL FH mutation database (Usifo et al. 2012). Sanger sequencing was used to confirm all called mutations. Samples with known FH mutations and therefore an explained cause were removed from further analysis.

The Tier 2 gene list consisted of genes associated with LDL-C as a lead trait in the largest (at the time) available Global Lipid Genetic Consortium (GLGC) meta-analysis of genome-wide association studies (Teslovich et al. 2010), which included: *APOE*, *SORT1* (*CELSR2*, *PSRC1*), *ABCG5/ABCG8*, *MYLIP*, *HFE*, *LPA*, *PLEC*, *ABO*, *ST3GAL4*, *NYNRIN*, *OSBPL7*, *TOP1*, *PLEC1*. *PLEC1* appeared to be highly polymorphic in both groups, showing an excess of variants in the controls as opposed to cases, therefore variants in genes

sharing the LD block with *PLEC1* were also considered (LD block is shown in Appendix Figure X). These included: *PARP10*, *GRINA*, *SPATC1*, *OPLAH*, *EXOSC4*, *GPAA1*, *KIAA1875*, *CYC1*, *SHARPIN*, *MAF1*.

Genes that encode proteins, which are directly involved in the lipid metabolism were also closely analysed, by looking for variants with strong predicted effect on the gene function. These included: *SREBP1*, *SREBP2*, *HMGCR*, *SCAP*, *INSIG1*, *INSIG2*, *NPC1L1*, *CETP*, *LPL*, *LIPC*, *ANGPTL4*.

In the past few years, several loci have been demonstrated to be associated with FH in family linkage studies, however none was able to map the association to a specific gene (Cenarro et al. 2011, Marques-Pinheiro et al. 2010, Wang et al. 2011). In this study, genes located in the associated regions were also analysed for loss-of-function variants.

2.4.6. Copy Number Variant (CNV) calling

The ExomeDepth package for R (freely available at the Comprehensive R Archive Network (CRAN)) was used to call Copy Number Variants (CNVs) (Plagnol et al. 2012). ExomeDepth is based on an algorithm designed to compare the read count data from exome sequencing experiments across a batch of sequenced samples. The model builds an optimised reference set from the input data, which increases the CNV detection power. The ExomeDepth approach has been demostrated to overcome biases associated with sequence capture and high throughput sequencing when used on samples sequenced at the same time in the same manner (Plagnol et al. 2012). The package uses the February 2009 human genome reference sequence (GRCh37, or hg19) for exon annotations, which was produced by

the Genome Reference Consortium (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/). The CNV calling was performed on BAM files in batches of 24 or 25 samples, since they were sequenced at different time points of the project. Common CNVs were annotated using published data from Conrad et al. (Conrad et al. 2010). CNVs called by ExomeDepth can be visualised using a plot function, which shows the ration between observed and expected read depth, as well as the 95% confidence interval generated for the studied batch of samples. Observed/expected ratios of around 1.5 indicate heterozygous duplication, ratios above 2.0 suggests a homozygous CNV. Deletions are characterised by values smaller than 0.5.

The R script for the CNV calling using the ExomeDepth package can be found at https://github.com/marta10/PhD_Thesis_MF/blob/master/cnvs_calling.R.

In addition to ExomeDepth, CNVs were also called using an alternative method – ConVex, which was provided by the UK10K project. ConVex analysis is also based on the read depth information, which was compared to a reference set generated for 50 random whole exome samples from the UK10K project. The algorithm detects CNVs with the log2 ratio of depth over median depth, using an error-weighted score and the Smith-Waterman algorithm.

2.5. In silico mutation analysis

2.5.1. Integrative Genomic Viewer (IGV)

Integrative Genomic Viewer (IGV) is a visualisation tool of large-scale sequencing genomic or exomic data sets (Robinson et al. 2011), which supports wide variety of data types, including genomic annotations and sequence alignments, such as BAM files. IGV was freely available from: http://www.broadinstitute.org/igv/. The tool was used for a visual assessment of a candidate variant call quality, for example by analysing the depth of coverage, read clonality, or forward/reverse primer bias.

2.5.2. PolyPhen-2

Polymorphism Phenotyping 2 (PolyPhen-2) is a web-based tool that predicts possible impact of an amino acid substitution on the structure and function of a human protein by using 11 predictive features, such as PSIC score, the sequence identity to the closest homologue, CpG context, and structural features (Adzhubei et al. 2010). PolyPhen can be accessed via http://genetics.bwh.harvard.edu/pph2/. The mutation prediction algorithm classifies variants into three groups: *Benign*, *Possibly Damaging* and *Probably Damaging*.

2.5.3. SIFT

Sorting Intolerant From Tolerant (SIFT) is a web-based amino acid change prediction tool, which can be found at http://sift.jcvi.org/ (Ng and Henikoff 2003). SIFT examines a conservation of the studied amino acid change in the protein family it belongs to, the positive/negative charge, its hydrophobic or

hydrophilic character, and summarizes the data into probability score resulting in the outcome of either: *Tolerated* or *Not Tolerated*.

2.5.4. Mutation Taster

Mutation Taster is a free web-based tool designed to assess the effect of a single nucleotide change and/or deletion/insertion in a human gene (Schwarz et al. 2010). A studied variant is evaluated according to its evolutionary conservation, its ability to cause splice-site alterations, the protein features that it may affect, as well as interactions with other molecules. Results are assessed by naive Bayes classifier, which predicts the pathogenicity of a given variation. A variant can be predicted as *Disease Causing* or as a *Polymorphism*.

2.5.5. HOPE

Have (y)Our Protein Explained (HOPE) is a Web server that can be accessed at http://www.cmbi.ru.nl/hope/ . HOPE is a fully automated program that analyses DNA base pair changes (including insertions and deletions), by considering not only amino acid properties and conservation, but also interactions of the sequence with other molecules (e.g. ligands, metals, hydrogen bonds, disulphide bridges), modifications that can affect post-translational processes (e.g. phosphorylation), splice site effects, and functionality of structural domains (Venselaar et al. 2010). In addition, the report produced by HOPE contains visualisation of each amino acid change in

3D-structure of the analysed protein (if available in Protein Data Bank, PDB), which can help to understand the effect of a particular change.

2.5.6. Human Splicing Finder (HSF)

Human Splicing Finder (HSF) version 2.4.1 was used for assessment of single base pair changes that occur in the region in or around intronic/exonic boundaries and have ambiguous effect on the amino acid composition of a given protein. HSF is freely available on http://www.umd.be/SSF/. The tool uses an algorithm, which calculates the strength of a potential splice site and the strength of splicing branch points and compares them between the wild type and the mutated state (Desmet et al. 2009).

2.6. Statistical analysis

All statistical analyses were carried out using R (R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0).

2.6.1. Oxford FH

Matched pre- and post-treatment LDL-C values were available for 104 patients (69 mutation negative, 35 mutation positive). Concentrations of serum cholesterol, LDL-C, HDL-C and triglyceride were not normally distributed, and were presented as geometric means with an approximate standard deviation. Matched pre-treatment TC and TG values were available for 159 patients (62 mutation positive). Dutch scores were calculated using

the weights for diagnostic traits as described in Table 3. SBBHF study subjects were scored zero for arcus cornealis since there was no information available for this item. SBBHF LDL-C measures were post-treatment and we therefore estimated pre-treatment values for use in the score assuming a reduction of 40% in LDL-C with treatment. The ability of the criteria to discriminate between mutation carriers and non-carriers was assessed by the area under the ROC curve using combined data from both studies. Areas were compared using the method described by Delong (DeLong et al. 1988). Dutch scores were adjusted for study differences before construction of the ROC curves. A p value of <0.05 was taken as significant.

2.6.2. South-Asian FH

Matched pre- and post-treatment LDL-C values were available for 19 patients (2 mutation positive). Concentrations of serum cholesterol, LDL-C, HDL-C and triglyceride were not normally distributed, and were presented as geometric means with an approximate standard deviation. Matched pre-treatment TC and TG values were available for 24 patients (6 mutation positive). A p value of <0.05 was taken as significant.

2.6.3. LDL-C gene score association testing

LDL-C gene score for each individual was calculated by summing the weights of each of the ten risk alleles the individual was carrying and by adding the weight of the observed ApoE isoform. The scores were compared by Welch Two-sample T Test between the mutation negative FH and the mutation positive FH, the mutation negative and the healthy control, and the healthy control and the mutation positive FH. ANOVA test was used to test the LDL-C gene score distribution difference between the three groups together.

2.6.4. Whole exome sequencing gene-based burden test

Analysis of the rare variant association with FH cohort was performed using a gene-based burden test. A gene-based burden test groups variants into genes, which increases the power of detecting the association in comparison to a single-variant association tests. The increase in power is a great

advantage of the burden test however it depends on the number of associated variants, the number of neutral variants diluting the signal, and the direction of the effect of the associated gene. In this study variants were flagged by their frequency as *somewhat-rare* or *rare* (see 'Filtering of variants' section) and by their function (as *functional*) in cases and controls. Flagged variants were then counted per gene in the cases and controls. The imbalance of *functional somewhat-rare* or *functional rare* variants was assessed using a binomial probability test (pbinom) in R. The R script for the association testing can be found at: https://github.com/marta10/PhD_Thesis_MF/blob/master/association_updated_aug_2013.sh .

III. CHAPTER THREE - Analysis of the frequency and spectrum of FH mutations in patients from Oxford

3.1. Introduction

The clinical phenotype of FH is known to be due to mutations in three genes encoding proteins involved in the uptake of LDL-C from the plasma, *LDLR*, *APOB* and *PCSK9*. Mutations in *LDLRAP1* are known to cause recessive form of FH. In the UK, the Simon Broome Register criteria are used for the clinical diagnosis of FH, whereas other European countries may use a score developed by the Dutch Lipid Clinic Network (DLCN) ('Risk of fatal coronary heart disease in familial hypercholesterolaemia. Scientific Steering Committee on behalf of the Simon Broome Register Group' 1991, Marks et al. 2003). To date, there are over 1,200 different *LDLR* mutations reported (Usifo et al. 2012) but only one common *APOB* (c.10580G>A, p.(Arg3527Gln)) and one *PCSK9* (c.1120G>T, p.(Asp374Tyr)) (Humphries et al. 2006a). The spectrum of FH mutations in Europe varies between countries, from Greece with only six mutations, which account for 60% of FH in the country, to the Netherlands with one of the most heterogeneous spectrum (Glynou et al. 2008, Dedoussis et al. 2004). In the UK there are over 200 different mutations reported (Leigh et al. 2008), which is similar to other western countries. *LDLR* mutations include mainly single nucleotide changes, which alter the amino acid composition of the mature protein, affect the correct splicing of the transcript, or binding of key transcription factors, if located in the promoter region (Khamis A 2012). Large deletions and insertions account for approximately 5-6% of all FH genetic defects (Taylor et al. 2010b, Leigh et al. 2008).

Statin drug therapy significantly reduces the morbidity and mortality from premature coronary disease in FH, particularly if affected individuals are identified and treated in childhood or early adulthood (Neil et al. 2008, Gill et al. 2012, Nherera et al. 2011). The UK National Institute for Health and Clinical Excellence (NICE) guidelines published in 2008 recommended that all FH patients be offered a DNA test to confirm the diagnosis and that identified mutations should be used as the basis for cascade testing of first-degree relatives of index cases. Patients newly identified by such screening can then be offered treatment to reduce the risk of premature cardiac events (Wierzbicki et al. 2008). DNA testing for FH has also been shown to complement cholesterol measurement in the management of affected individuals (Humphries et al. 2006c).

This study aimed to assess the frequency and spectrum of mutations recognised to cause FH among patients attending the Oxford Lipid Clinic. The *LDLRAP1* gene was not analysed since all patients had the dominant pattern of FH inheritance. The frequency of specific mutations in the UK differs between areas, with p.(Glu101Lys) being the most common in Manchester (Webb et al. 1992), p.(Arg350*) in South of England (Day et al. 1997a), and p.(Cys184Tyr) in Glasgow (Lee et al. 1998). This study examined whether there are any specific mutations that occur with an unexpected frequency among patient attending the Oxford lipid clinic, which is a specialist clinic with a catchment population of over 620,000 people (Neil et al. 2000). The correlation between the measured pre-treated cholesterol, pre-treated triglycerides and the mutation detection rate was also assessed to test the

hypothesis that the individuals carrying a FH mutation have higher pre-treatment cholesterol levels and lower triglyceride level compared to those with no mutation. The likelihood of identifying mutation carriers was compared using two different clinical diagnostic criteria: the Simon Broome criteria and the DLCN score. In addition, the study examined whether the effectiveness of lipid-lowering therapy varied between patients with different genetic causes of FH.

3.2. Results

3.2.1. Patients characteristics

In total, 289 (272 probands) FH patients were screened for FH mutations in three genes, *LDLR*, *APOB* and *PCSK9*. Characteristics of the patients recruited for the study is shown in Table 13. The majority (52%) of individuals had the clinical diagnosis of PFH, with 23% being DFH. 26% of the patients could not be classified (UH) due to a lack of family history of early CHD or the patient was unaware of the family history. There was no significant difference in age or in the male/female ratio between the groups. The mean pre-treatment cholesterol differed significantly between groups ($p<0.0001$), with DFH having the highest TC (9.79 mmol/l) and LDL-C (6.93 mmol/l) levels, while PFH and UH groups had similar pre-treatment TC and LDL-C levels. The highest mean pre-treatment TG levels (2 mmol/l) was observed in the UH patients, and it was significantly different between the groups ($p=0.004$), however similar to PFH ($p=0.162$). There was no difference in the HDL-C distribution between the groups.

A well-phenotyped cohort studied by Neil et al. (Neil et al. 2004) was used for the comparison and replication. The SBBHF study comprised 409 DFH patients, diagnosed using the Simon Broome criteria for FH, of whom the majority (80%) had a reported FH-causing mutation. The baseline characteristics of this cohort are shown in Table 14. Briefly, mutation positive patients were significantly younger (p=0.01), had higher pre- and post-treatment TC and LDL-C ($p=3x10^{-5}$, p=0.03, $p=4x10^{-5}$). The post-treatment TG levels were significantly lower in the mutation carriers (1.25 mmol/l vs. 1.62 mmol/l, $p=3.136x10^{-6}$). The replication SBBHF patients were younger than the DFH Oxford FH participants. There was no significant difference in pre-treatment TC levels between the groups, however the SBBHF patients had significantly higher post-treatment TC, TG and LDL-C (Table 15). These differences are likely to be due to the different lipid-lowering treatment that was available for the cohorts. The replication cohort is historic and post-treatment lipid levels were recorded before Atorvastatin treatment was approved in 1996. There was no difference in the male/female ratio between the cohorts.

Table 13

Baseline characteristics of the Oxford FH cohort. Lipid levels (in mmol/l) are shown as geometric means (with approximate standard deviation), because they were not normally distributed. P value between two groups was calculated on log transformed data using Welch Two Sample t-test. The overall p values were obtained using ANOVA test.

Variable	**DFH (N=65)**	***p* value (DFH vs. PFH)**	**PFH (N=150)**	***p* value (PFH vs. UH)**	**UH (N=74)**	***p* value (overall)**
Gender						
% Male (N)	47.7 (34)		46.7 (70)		58.1 (43)	0.25
Mean age (SD)	58.0 (12.7)		53.8 (14.1)		54.2 (14.9)	0.12
Pre-treatment TC	9.79 (1.66)	$7.6x10^{-06}$	8.71 (1.27)	0.391	8.47 (1.92)	<0.0001
Pre-treatment HDL-C	1.38 (0.35)	0.302	1.46 (0.40)	0.634	1.39 (0.39)	0.62
Pre-treatment TG	1.22 (0.56)	0.008	1.53 (0.72)	0.115	2.00 (0.81)	0.004
Pre-treatment LDL-C	6.93 (1.61)	0.005	6.12 (1.15)	0.074	5.45 (1.72)	0.0002
Post-treatment TC	5.95 (1.01)	0.003	5.46 (1.13)	0.604	5.36(1.17)	0.008
Post-treatment HDL-C	1.31 (0.36)	0.05	1.44 (0.32)	0.866	1.40 (0.46)	0.1
Post-treatment TG	1.12 (0.52)	0.188	1.25 (0.60)	0.848	1.21 (0.38)	0.42
Post-treatment LDL-C	4.1 (1.04)	$7.7x10^{-06}$	3.29 (1.00)	0.596	3.65 (1.57)	0.0001
Detected mutations (in probands only)						
LDLR (%)	43(71.7)		30(21.1)		9(12.9)	
APOB (%)	1(1.7)	$1.94x10^{-09}$	8(5.6)	0.062	1(1.4)	<0.0001
None (%)	16(26.7)		104(73.2)		60(85.7)	

Table 14

Baseline characteristics of the SBBHF study used for the replication and comparison with the original results in Oxford FH cohort. Lipid levels (in mmol/l) are shown as geometric means (with approximate standard deviation), because they were not normally distributed.

Variable	**Mutation -ve (N=81)**	**Mutation +ve (N=328)**	**Total (N=409)**	**P value**
% male	50.6% (41)	51.5% (169)	51.3% (210)	0.88
Age	52.4 (13.6)	48.3 (13.5)	49.1 (13.6)	0.01
Pre-treatment TC	9.22 (1.71)	10.35 (1.88)	10.13 (1.90)	0.00003
Post-treatment TC	6.39 (1.41)	6.74 (1.29)	6.67 (1.32)	0.03
Post-treatment LDL-C	4.06 (1.27)	4.67 (1.24)	4.54 (1.28)	0.00004
Post-treatment HDL-C	1.37 (0.33)	1.34 (0.37)	1.34 (0.36)	0.49
Post-treatment Triglycerides	1.62 (0.76)	1.25 (0.55)	1.32 (0.60)	3.136×10^{-6}

Table 15

Baseline characteristics comparison of the Oxford definite FH patients and the SBBHF replication study. Lipid levels (in mmol/l) are shown as geometric means (with approximate standard deviation). P value between two groups was calculated on log transformed data using Welch Two Sample t-test.

Variable	**Oxford DFH (N=65)**	**SBBHF (N=409)**	***p* value**
Gender			
% Male (N)	47.7 (34)	51.3 (210)	0.99
Mean age (SD)	58.0 (12.7)	49.1 (13.6)	1.2×10^{-7}
Pre-treatment TC	9.79 (1.66)	10.13 (1.90)	NS
Post-treatment TC	5.95 (1.01)	6.67 (1.32)	5.5×10^{-6}
Post-treatment HDL-C	1.31 (0.36)	1.34 (0.36)	NS
Post-treatment TG	1.12 (0.52)	1.32 (0.60)	0.03
Post-treatment LDL-C	4.1 (1.04)	4.54 (1.28)	0.01

NS = not significant

3.2.2. Mutation spectrum

A FH-causing variant was found in 101 individuals, of which the most frequently observed was *APOB* p.(Arg3527Gln), present in 11 individuals (10 probands). There were 54 different *LDLR* mutations, which were found in 90 patients and accounted for 89% of all observed mutations. The most commonly observed mutations in *LDLR* were c.301G>A (p.(Glu101Lys)) present in six probands, c.259T>G (p.(Trp87Gly)), c.313+1G>A, c.680_681delAC (p.(Asp227Glyfs*12)), c.681C>G (p.(Asp227Glu)), c.1116_1119dupGGGT (p.(Glu374fs*8)), and c.2054C>T (p.(Pro685Leu)), all observed in three FH probands. Most of the changes occurred in exons 4 and 10, which are the longest *LDLR* exons. There were no FH-causing variants found in exons 1, 12, 16 and 18 of the *LDLR*. The MLPA analysis of *LDLR* detected large gene rearrangements in 11 probands, which accounted for over 10% of all Oxford FH mutations. All observed mutations are summarised in Table 16. There were no patients with the *PCSK9* p.(Asp374Tyr) mutation in this cohort. There were no compound heterozygotes for mutations in *LDLR* and *APOB*.

Table 16

All FH mutations found in the Oxford FH cohort, ordered by their frequency. Variants marked in bold were novel. Predicted consequence: A=amino acid change, B=splicing defect, C=peptide truncation due to a frame shift, D=large in frame indel, E=premature stop codon formation, F=gene expression effect, G=structure/folding defect, H=peptide elongation.

Gene	Base pair change	Predicted amino acid change	Number of carriers (number of probands)	Predicted consequence
APOB	c.10580G>A	p.(Arg3527Gln)	11(10)	A
LDLR	c.301G>A	p.(Glu101Lys)	6(6)	A
LDLR	c.313+1G>A	N/A	4(4)	B
LDLR	c.681C>G	p. (Asp227Glu)	4(3)	A
LDLR	c.1444G>A	p.(Asp482Asn)	4(3)	A
LDLR	c.259T>G	p.(Trp87Gly)	3(3)	A
LDLR	c.680_681delAC	p.(Asp227Glyfs*12)	3(3)	C
LDLR	c.1116_1119dupGGGT	p.(Gly374fs*8)	3(3)	C
LDLR	c.2054C>T	p.(Pro685Leu)	3(3)	A
LDLR	c.191-?_1186+?del	N/A	3(2)	D
LDLR	**c.2187_2197del**	**p.(Leu729Leufs*39)**	**3(2)**	**C**
LDLR	c.2089G>C	p.(Ala697Pro)	3(1)	A
LDLR	c.662A>G	p.(Asp221Gly)	2(2)	A

LDLR	c.682G>T	p.(Glu228*)	2(2)	E
LDLR	c.695-?_817+?del	N/A	2(2)	D
LDLR	c.912C>G	p.(Asp304Glu)	2(2)	A
LDLR	c.1846-?_2140+?del	N/A	2(2)	C
LDLR	**c.667_693del**	**p.(Lys223_Cys231del)**	**2(1)**	**G**
LDLR	**c.-121T>C**	**N/A**	**1(1)**	**F**
LDLR	c.68-?_1186+?dup	N/A	1(1)	D
LDLR	c.118delA	p.(Ile40Serfs*166)	1(1)	C
LDLR	c.139G>A	p.(Asp47Asn)	1(1)	A
LDLR	c.191-?_1186+?dup	N/A	1(1)	D
LDLR	c.191-?_313+?del	N/A	1(1)	D
LDLR	**c.361T>A**	**p.(Cys121Ser)**	**1(1)**	**A**
LDLR	c.502G>A	p.(Asp168Asn)	1(1)	A
LDLR	c.621C>T	p.(Gly207Gly)	1(1)	A
LDLR	**c.629T>A**	**p.(Ile210Asn)**	**1(1)**	**A**
LDLR	c.654_656delTGG	p.(Gly219del)	1(1)	A
LDLR	**c.859G>A**	**p.(Gly287Ser)**	**1(1)**	**A**
LDLR	c.889A>C	p.(Asn297His)	1(1)	A
LDLR	**c.898A>T**	**p.(Arg300*)**	**1(1)**	**E**
LDLR	c.933delA	p.(Lys311fs*59)	1(1)	C
LDLR	c.1048C>T	p.(Arg350*)	1(1)	E
LDLR	c.1049G>C	p.(Arg350Pro)	1(1)	A
LDLR	c.1135T>C	p.(Cys379Arg)	1(1)	A
LDLR	c.1215C>G	p.(Asn405Lys)	1(1)	A

LDLR	**c.1230G>T**	**p.(Arg410Ser)**	**1(1)**	**A**
LDLR	c.1246C>T	p.(Arg416Trp)	1(1)	A
LDLR	c.1285G>A	p.(Val429Met)	1(1)	A
LDLR	c.1369_1372dupGACA	p.(Arg458fs*10)	1(1)	C
LDLR	**c.1379_1402delinsCAGCTTGACCCGC**	**p.(His460Profs*3)**	**1(1)**	**C**
LDLR	c.1436T>C	p.(Leu479Pro)	1(1)	A
LDLR	c.1466A>G	p.(Tyr489Cys)	1(1)	A
LDLR	c.1567G>A	p.(Val523Met)	1(1)	A
LDLR	c.1574A>T	p.(Asp525Val)	1(1)	A
LDLR	c.1587-?_1845+?del	N/A	1(1)	C
LDLR	**c.1587-?_1845+?dup**	**N/A**	**1(1)**	**C**
LDLR	c.1646G>A	p.(Gly549Asp)	1(1)	A
LDLR	c.1694G>C	p.(Gly565Ala)	1(1)	A
LDLR	c.1897C>T	p.(Arg633Cys)	1(1)	A
LDLR	c.1941_1944delins1922_1934	N/A	1(1)	C
LDLR	**c.2098G>A**	**p.(Asp700Asn)**	**1(1)**	**A**
LDLR	c.2393_2401del	p.(Leu799_Phe801del)	1(1)	A
LDLR	**c.2476C>A**	**p.(Pro826Thr)**	**1(1)**	**A**

3.2.3. Novel *LDLR* variants

There were 12 novel *LDLR* variants, summarized in Table 17, which were not reported on the UCL FH database. These included one promoter variant (c.-121T>C), which was further studied and proved to affect a transcription factor binding site and reduced luciferase activity by 50±8% suggesting strongly that this variant is FH-causing (publication in revision, Khamis A, et al.). Six non-synonymous changes were identified of which two were located in exon 4 of the gene, one nonsense mutation (c.898A>T, p.(Arg300*)), three small rearrangements, of which two lead to a frame shift and premature termination, and one large gene rearrangement - duplication of exon 11, predicted to cause frame shift and premature termination. Using *in silico* tools, all novel variants were predicted to be protein damaging suggesting that they are likely to be pathogenic.

Deletion of nine amino acids in c.667_693del (p.(Lys223_Cys231del)) was assessed using conservation and structure analysis (Usifo et al. 2012). This region includes the highly conserved D-x-S-D-E motif (residues 224-228) in exon 4. The secondary structure of this region and the coordination of a calcium cation are crucial for pH dependent recycling of the LDL-R peptide. Therefore, deletion of residues 227 and 228, which are directly involved in calcium coordination and removal of the disulphide bridge formed between residues 231and 216, is very likely to have a pathogenic effect (shown in Figure 22). All novel mutations were submitted to the UCL FH database (Usifo et al. 2012).

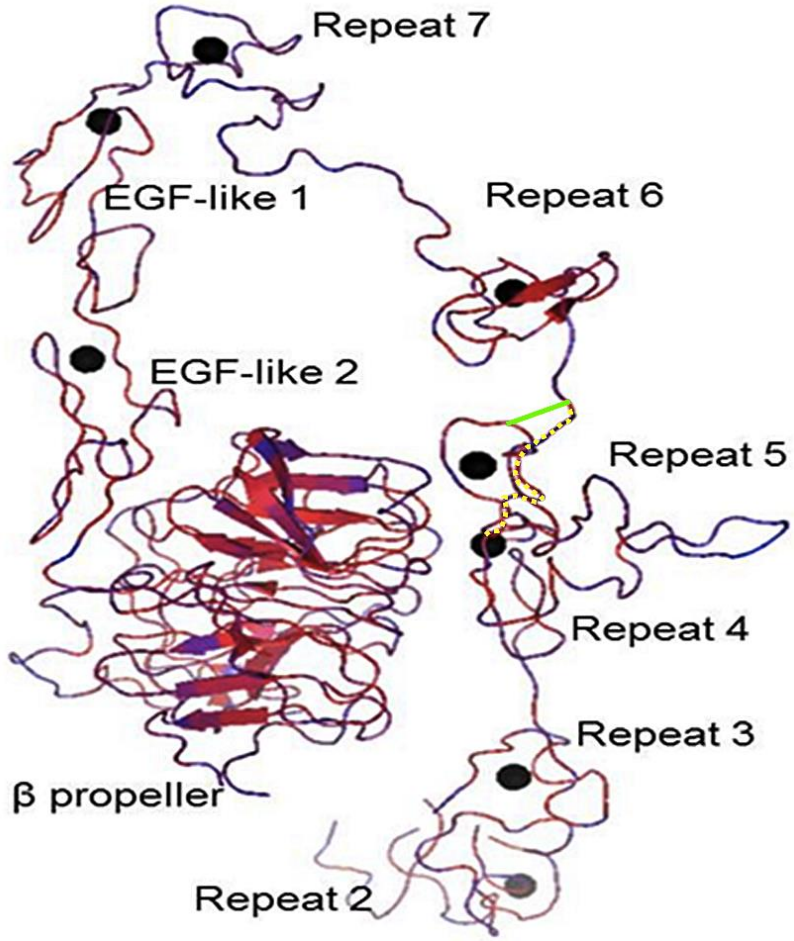
Table 17

Novel *LDLR* variants identified in the Oxford Lipid Clinic patients and the *in silico* prediction of their effect.

Mutation type/Exon	Variant position	PolyPhen	SIFT	Mutation Taster	Conclusion
Promoter					
	c.-121T>C	NA	NA	D	Transcription factor binding site disruption (publication in preparation)
Missense					
4	c.361T>A (p.(Cys121Ser))	D	D	D	FH-causing
4	c.629T>A (p.(Ile210Asn))	D	D	D	FH-causing
6	c.859G>A (p.(Gly287Ser))	D	D	D	FH-causing
9	c.1230G>T (p.(Arg410Ser))	D	D	D	FH-causing
14	c.2098G>A (p.(Asp700Asn))	P	D	D	FH-causing
17	c.2476C>A (p.(Pro826Thr))	D	D	D	FH-causing
Nonsense					
6	c.898A>T (p.(Arg300*))	NA	NA	D	Formation of premature stop codon
Small rearrangements					
4	c.667_693del (p.(Lys223_Cys231del))	NA	NA	NA	Deletion of 9 highly conserved residues
10	c.1379_1402delinsCAGCTTGACCCGC (p.(His460Profs*3))	NA	NA	NA	Frame shift → premature stop codon
15	c.2187_2197del (p.(Leu729Leufs*39))	NA	NA	D	Frame shift → premature stop codon
Large rearrangements					
11	c.1587-?_1845+?dup	NA	NA	NA	Frame shift → premature stop codon

Figure 22

3D structure of LDL-R extracellular domains (Protein Database 1N7D, http://www.ebi.ac.uk/pdbsum/1N7D) viewed using Jmol, overlaid with conservation scores. Red indicates high conservation, purple moderate conservation, blue poor conservation. Calcium cations are shown as black dots. Residues deleted in p.(Lys223_Cys231) are shown by yellow dotted line, disulphide bridge broken by the deletion shown as green line.



3.2.4. Mutation detection rate and patients' lipid levels

An FH-causing mutation was detected in 73% DFH and in 27% PFH patients (probands), with 10 mutations found in the unclassified hypercholesterolaemia group (14% of probands) and overall this difference in detection rate was highly statistically significant ($p < 0.0001$). The mutation detection rate was significantly different by pre-treatment TC levels quartile ($p=9.83 \times 10^{-5}$), and in the top quartile (pre-treated TC of 10.0-15.0mmol/l) an FH mutation was found in 74% of patients, as shown in Table 18. The mutation detection rate also differed significantly depending on pre-treatment TG levels ($p=0.0005$). Individuals with the lowest TG levels (0.4-1.0mmol/l) had the highest detection rate (60%), which decreased to 20% for patients in the top quartile (2.16-4.3 mmol/l). When combined, the mutation detection rate was 100% in individuals whose TG levels were in the lowest quartile and TC levels in the top quartile compared to a less than 5% detection rate in those with TG levels in the highest quartile and TC levels in the lowest quartile, demonstrated in Figure 23. These findings were replicated in the SBBHF cohort, where the highest mutation detection rate of 90% was observed in patients with pre-treatment TC above 11.6 mmol/l, which decreased to 68% in those with TC equal to or below 8.7 mmol/l, confirming the association of pre-treatment TC levels with the mutation detection rate ($p=0.0001$). Because pre-treatment TG levels were not available, the post-treatment TG levels were used and this also showed high association ($p=1.8 \times 10^{-6}$) (Table 18).

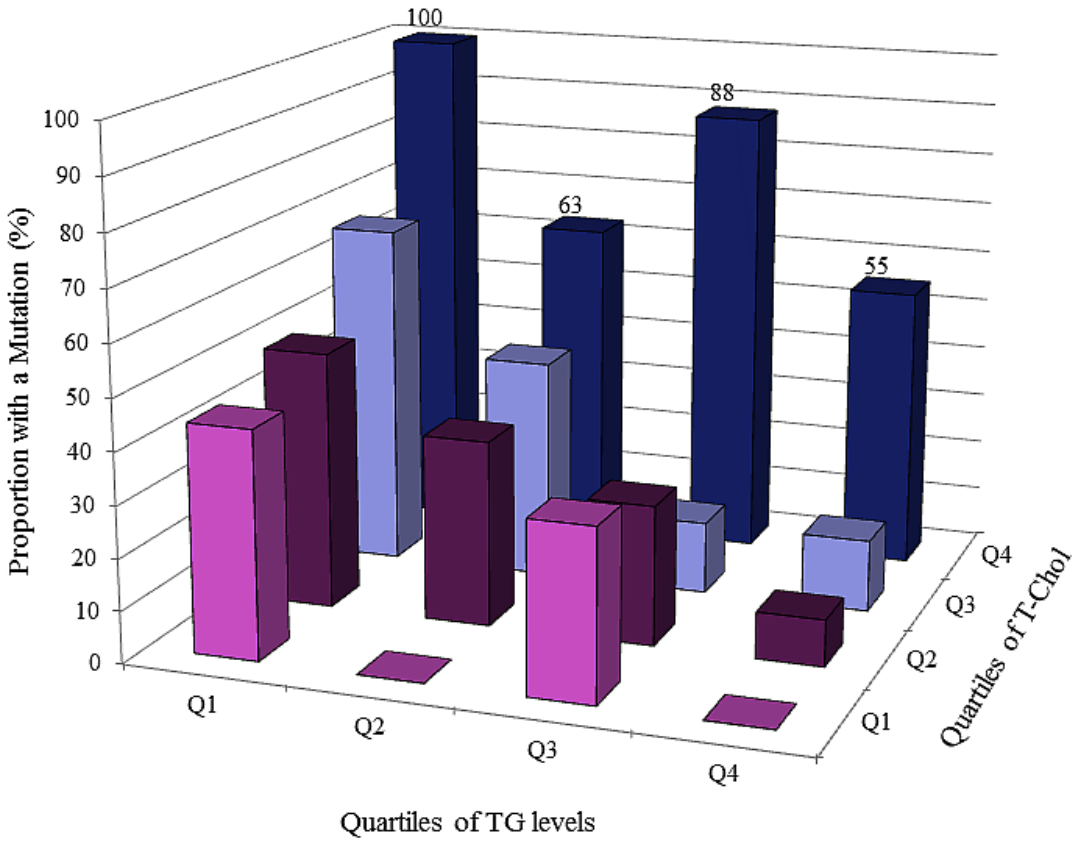
Table 18

FH mutation detection rate by quartile of pre-treatment TC and pre-treatment TG. (Pre-treatment TG levels for SBBHF study were not available).

		Oxford FH study			
Quartile of Pre-treatment TC (mmol/l)	**N**	**Mutation +ve (%)**	**Quartile of Pre-treatment TG (mmol/l)**	**N**	**Mutation +ve (%)**
Q1 <= 8.0	40	10 (25)	Q1 <= 1.0	42	25 (60)
Q2 8.1-8.7	41	12 (29)	Q2 1.10-1.32	38	15 (40)
Q3 8.8-9.9	44	15 (34)	Q3 1.33-2.15	39	14 (36)
Q4 >10.0	34	25 (74)	Q4 2.16-4.30	40	8 (20)
P value (trend)		$p = 9.83 \times 10^{-5}$			0.000458
		Replication SBBHF study			
Quartile of pre-treatment TC (mmol/l)	**N**	**Mutation +ve (%)**	**Quartile of post-treatment TG (mmol/l)***	**N**	**Mutation +ve (%)**
Q1 <=8.7	76	52 (68.4)	Q1 <1	105	92 (87.6)
Q2 8.8-10.1	82	66 (80.5)	Q2 1-1.3	123	109 (88.6)
Q3 10.2-11.6	69	62 (89.9)	Q3 1.4-1.8	90	70 (77.8)
Q4 >11.6	72	65 (90.3)	Q4 >1.8	91	57 (62.6)
P value (trend)		0.0001			$P= 1.80 \times 10^{-6}$

Figure 23

FH mutation detection rate by quartiles of pre-treatment TC and pre-treatment TG levels combined in Oxford FH patients.



The specificity and sensitivity of correctly identifying mutation carriers by using the Simon Broome FH criteria was compared to the Dutch Lipid Clinic Network (DLCN) score. 220 Oxford FH patients had data available to allow the DLCN Score to be calculated. As shown in Table 19 the mutation detection rate significantly increased (p=0.004) with rising DLCN Score. The number of DFH individuals diagnosed using the Simon Broome criteria was significantly correlated with the DLCN score (2.6×10^{-7}), indicating high similarity of both methods. When we compared the discriminatory power of the two methods using the A_{ROC} statistic, both methods performed well (ROC values for SB= 0.73 (95% CI = 0.68-0.77) vs. DLCN = 0.72 (95% CI = 0.69-0.77), p values for difference = 0.68), confirming that the sensitivity and specificity of both clinical diagnostic approaches were not different (Figure 24).

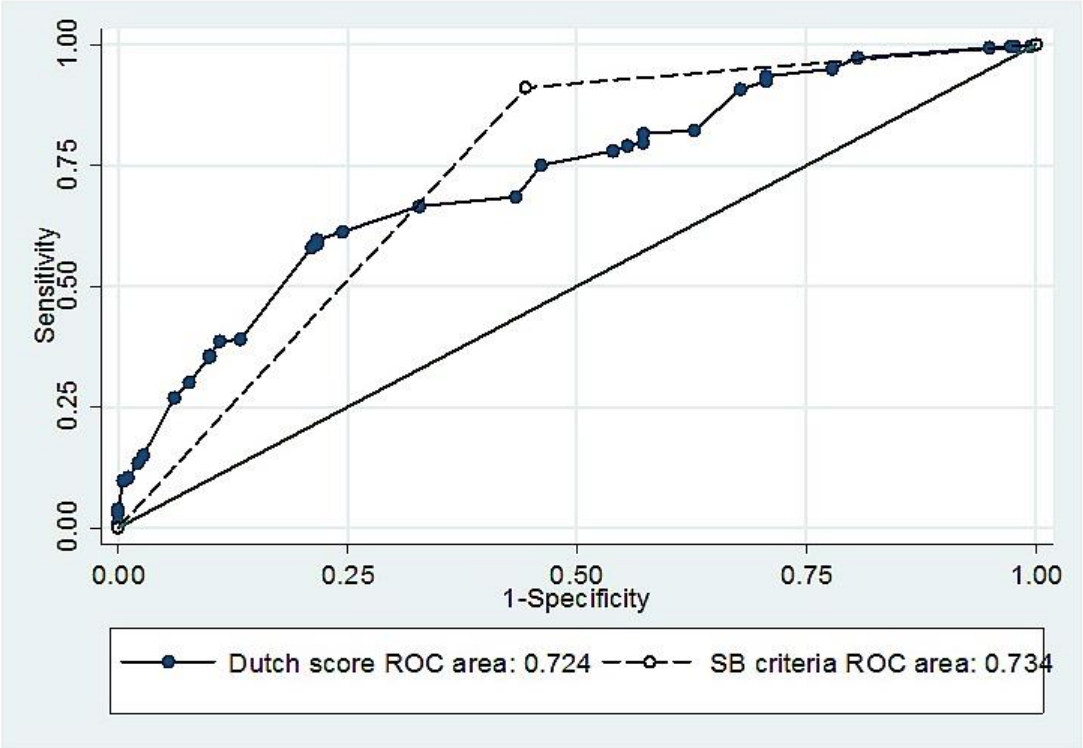
Table 19

The FH mutation detection rate in patients diagnosed using the DLCN score, with the percentage of DFH diagnosed by the Simon Broome FH criteria in each score group.

DLCN Score	Patient Count	Mutation +ve (%)	DFH (%)
< 3	13	3 (23)	2 (15)
3 - 5	69	19 (28)	9 (13)
6 - 8	49	19 (39)	8 (16)
> 8	89	48 (54)	45 (51)
P value for the trend (Fisher's exact)		0.004	2.6×10^{-7}

Figure 24

Receiver Operating Characteristic (ROC) curve comparing the accuracy of the Simon Broome FH criteria (SB) vs. the Dutch Lipid Clinic Network criteria (Dutch score) in Oxford Lipid Clinic patients and the SBBHF study combined. Chi square test p value of the area under the curve was not significant (p > 0.68).



ROC

	Observations	Area	Std. Err.	95% Conf.	Interval
DLCN score	485	0.724	0.0233	0.678	0.769
SB criteria	485	0.734	0.0203	0.694	0.773

3.2.5. Mutation carriage and lipid levels pre- and post-treatment

The mean level of pre-treatment TC in patients with any detected FH-causing mutation was significantly higher than in those with no mutation ($p=2.15 \times 10^{-08}$, Welch Two Sample t-test) (Figure 25). The highest pre-treatment TC was seen in the patients with any *LDLR* mutation (9.81 mmol/l, SD=1.52), followed by the familial defective ApoB (FDB) patients (9.12 mmol/l, SD=0.85), while the mutation-negative patients had the lowest pre-treatment TC (8.47 mmol/l, SD=1.36), as shown in Figure 26. The difference between the TC levels in *LDLR* and *APOB* mutation carriers was significantly different (p=0.03) as well as the overall difference between the three groups (ANOVA test, $p=3.31 \times 10^{-8}$). This result confirms previous findings (Humphries et al. 2006c).

Figure 25

Mean pre-treatment and post-treatment TC levels in mutation negative (-ve) vs. mutation positive (+ve) patients. P values shown from Welch Two Sample t-test. NS = not significant.

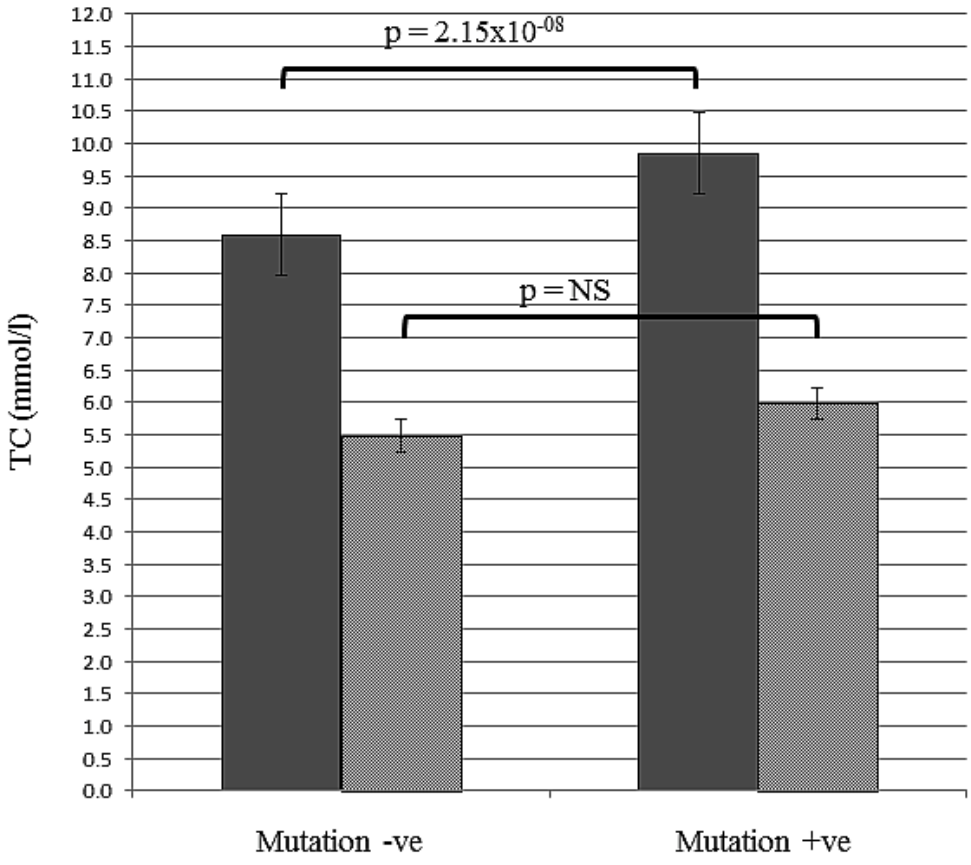
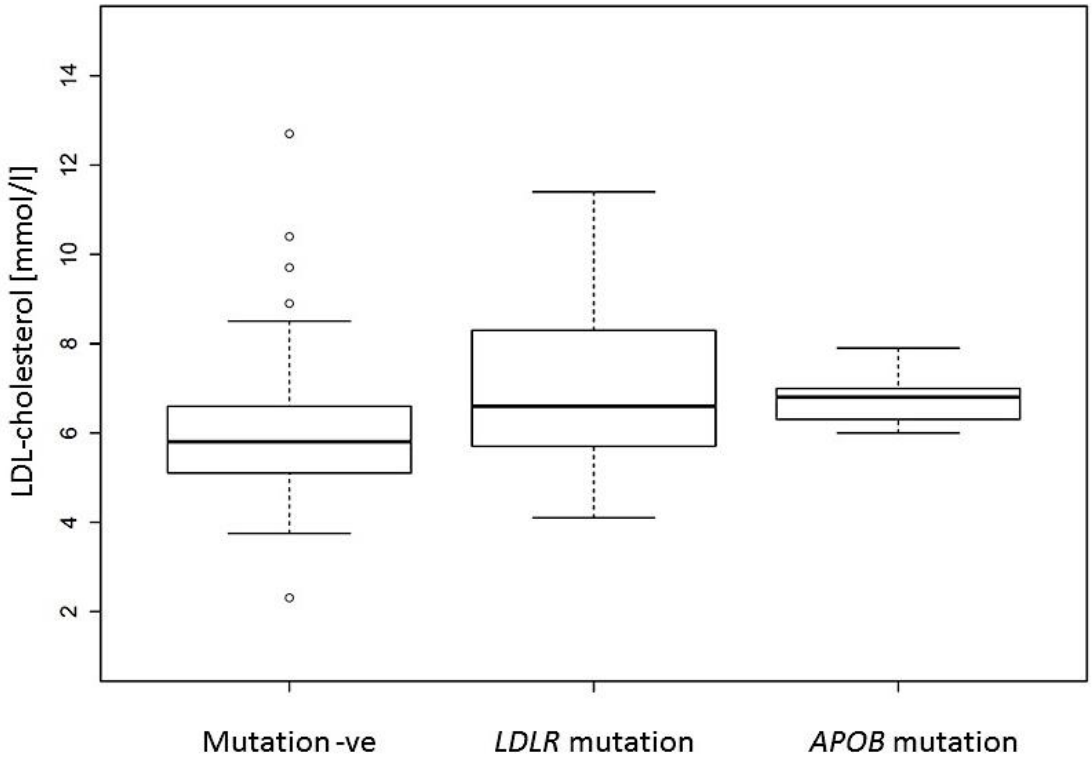


Figure 26

Pre-treatment LDL cholesterol in patients with mutation in *LDLR* or *APOB* genes, or with no mutation identified. The difference between the groups was significant (ANOVA $p = 7.76 \times 10^{-05}$).



The TC level was significantly reduced by treatment in both mutation positive and mutation negative groups by at least 35% (see Table 20). The UK guidelines on management of FH dyslipidaemias recommend that the treatment goal for FH should be to reduce LDL-C by at least 50% from baseline (Wierzbicki et al. 2008), while guidelines on management of dyslipidaemias recommend that LDL-C in high risk subjects should be reduced below 2.5 mmol/l and for individuals with CVD (very high risk) below 1.8 mmol/l (Catapano et al. 2011). Data on pre- and post-treatment LDL-C was available for 104 individuals, and the 50% LDL-C reduction was achieved in 47% of the patients. This was similar to the study from 2008, when 64% patients had 45% reduction in LDL-C (Hadfield et al. 2008), and to the findings of the National Audit of the Management of Familial Hypercholesterolaemia 2010, when 44% of patients achieved 50% reduction (http://www.rcplondon.ac.uk/resources/audits/FH).

Post-treatment LDL-C values were available for 176 patients of whom 26 (14.8%) had LDL-C reduced below 2.5 mmol/l and 1 patient (0.6%) had LDL-C below 1.8 mmol/l. The pre- and post-treatment LDL-C distribution is shown in Figure 27.

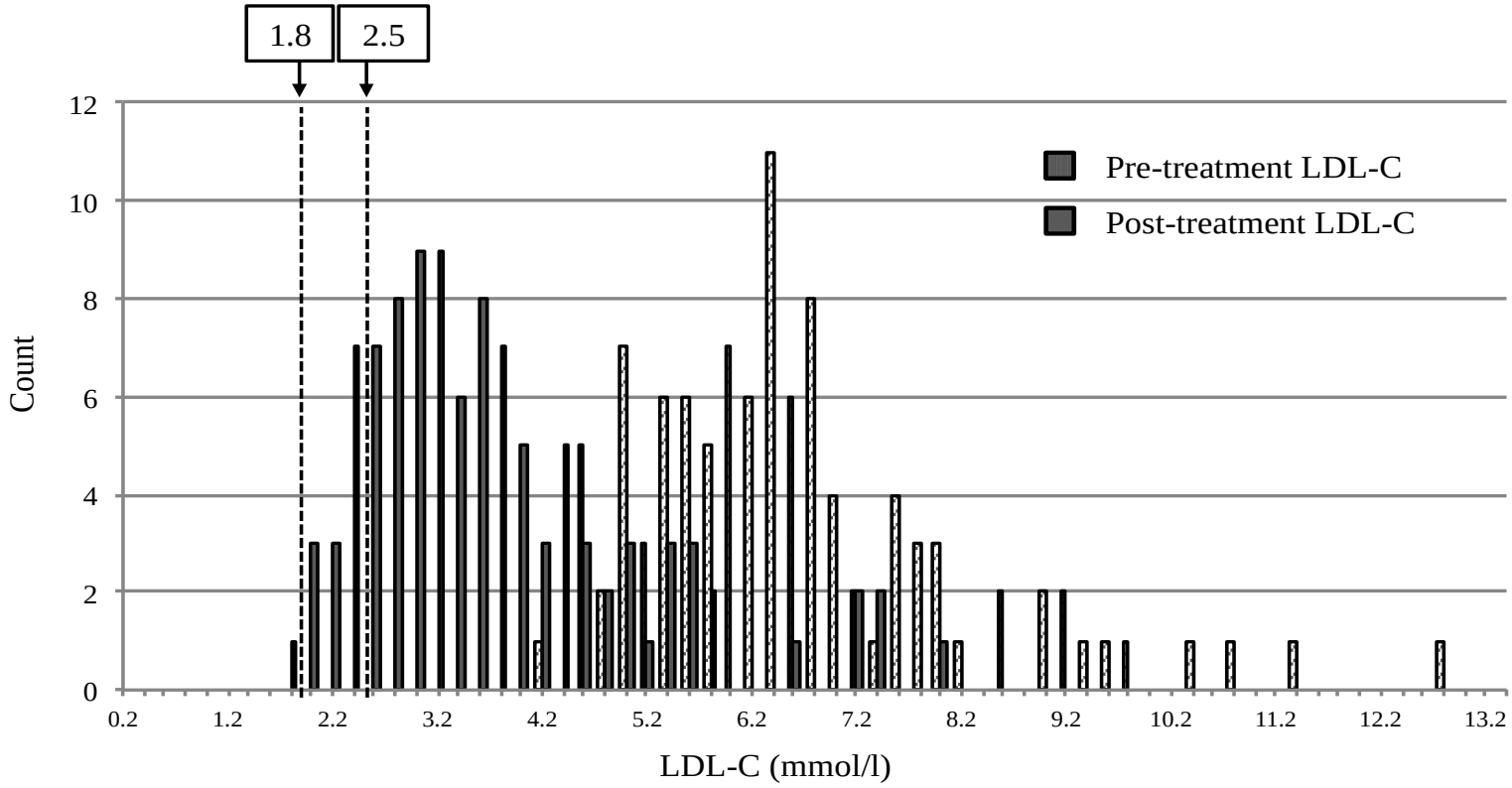
Table 20

Total cholesterol reduction after treatment in patients with or without an identified FH mutation, in DFH, PFH and UH.

	DFH	**PFH**	**UH**
Mutation -ve	-37.5% (-47.0 to -26.3)	-38.1% (-41.3 to -34.7)	-35.2% (-39.5 to -30.6)
Mutation +ve	-39.9% (-44.0 to -35.5)	-41.8% (-45.3 to -38.0)	-41.0% (-52.2 to -27.3)
All	-39.3% (-43.1 to -35.3)	-39.1% (-41.6 to -36.4%)	-36.5% (-40.5 to -32.2)

Figure 27

LDL-C levels distribution before and after treatment. Dashed lines at 1.8 mmol/l and 2.5 mmol/l indicate the recommended post-treatment levels for FH individuals with CVD and with high risk of CVD, respectively.



3.3. Discussion

This study has identified marked genetic heterogeneity among patients with heterozygous familial hypercholesterolaemia attending a single UK lipid clinic with a catchment population of more than 620,000 people in the Oxfordshire region. Amongst 272 unrelated patients 54 different *LDLR* mutations in 90 subjects were identified, which is comparable with the 107 different mutations identified in 232 subjects in the UK Cascade pilot project who were from five geographically dispersed sites (Taylor et al. 2010b). There were 12 previously unreported mutations found in the Oxford cohort, which accounted for 22% of all *LDLR* mutations found in this study. This is also substantially higher than the 7% novel variants found in the previous UK study (Taylor et al. 2010b). While this may be a result of the increased sensitivity of the current mutation detection methods, it may also reflect the genetic heterogeneity of this Oxford sample. Overall, the mutation spectrum was similar to the rest of the UK, however the most commonly occurring *LDLR* mutation in Oxford was c.301G>A (p.(Glu101Lys), which accounted for 6% of all observed mutations, as opposed to 1% in the whole of the UK study (Taylor et al. 2010b). The frequency of gross deletions/duplications within the *LDLR* was higher (10%) than previously reported (5-6%), however this was not statistically different (Fisher's exact test, p = 0.06). The *PCSK9* (p.(Asp374Tyr)) mutation, which is associated with a higher CHD risk than other FH-causing mutations (Humphries et al. 2006a), was not observed in the Oxford cohort.

A striking finding was the importance of high pre-treatment cholesterol and low triglyceride levels as predictors of the likelihood of detection of an FH-causing mutation. The detection rate of 73% in DFH compared to the 27% in PFH in patients attending the Oxford clinic was similar to that reported previously in UK patients (Humphries et al. 2006c, Heath et al. 2001, Graham et al. 1999, Damgaard et al. 2005).

Additionally, this study examined the utility of the Dutch Lipid Clinic Network (DLCN) score, in identifying patients with a high or low probability of carrying an FH-causing mutation. In patients with a DLCN score indicative of Definite FH (>8), Probable FH (>5 and <8) and Possible FH (>3 and <5), the mutation detection rates were 54% vs. 39% vs. 28%, respectively, and using the AROC statistic the two approaches were not significantly different in discrimination. Combining pre-treatment TC and TG levels gave the highest likelihood of finding patients carrying FH mutations, with 100% mutation detection rate in patients with pre-treated TC above 10.0 mmol/l and with TG below 1.0 mmol/l, compared to >5% detection rate in those with pre-treated TC below 8.0 mmol/l and with TG above 2.15 mmol/l.

A proportion of the studied patients (26%) did not fulfil the Simon Broome criteria for FH as there was either no family history of hypercholesterolaemia or premature CHD or, alternatively, the family history was unknown or incomplete. There were 10 FH-causing mutations found among probands of the unclassified hypercholesterolaemic individuals (14%), and the identification of an FH-causing mutation therefore changed the diagnosis to DFH, which would have warranted cascade testing of their first-degree relatives. This finding supports the clinical utility of DNA testing, even in

individuals with a low probability of being affected. We can speculate as to the likely genetic cause of the elevated cholesterol and triglyceride levels seen in the mutation negative subjects examined in this study. As recently demonstrated (Talmud et al. 2013) and discussed further in Chapter Three, in a significant proportion of mutation negative patients with a clinical diagnosis of FH, the elevated LDL-C can be explained by the “polygenic” contribution of 12 common LDL-C-raising variants, in genes identified through Genome Wide Association studies. Similarly it has been demonstrated that the frequencies of common triglyceride-associated variants are also significantly higher in groups of patients with different Fredrickson classification forms of hypertriglyceridaemia (Johansen et al. 2011) compared with controls. Thus, it is likely that the elevated levels of lipids seen in the Oxford no-mutation group have a polygenic and not a monogenic explanation. Several of the mutation negative DFH patients from Oxford were genotyped for the 12 LDL-C-rising SNPs in Chapter Five.

A limitation of the study was that pre-treatment TC measurements were not available for 31 out of 289 participants, and pre-treatment TG levels were not recorded for 130 individuals. This was because some patients who met the criteria in other respects (e.g. personal family history of premature CHD or tendon xanthomas) were referred to the lipid clinic whilst on diet or lipid-lowering medication, and it would not have been ethical to stop medication to collect data for this study. Although both the SB and DLCN clinical diagnostic methods performed well, the clinical utility of the Simon Broome diagnostic criteria may be questioned since they require tendon xanthomata to be

present for a diagnosis of DFH, and these are increasingly uncommon as patients are diagnosed at a younger age. Although the methodology used for the mutation screening is appropriate and was shown to be sensitive and robust (Whittall et al. 2010a), a recent report exposed some limitations to the current methods, which included false negative calling due to human error in data and sample handling (Futema et al. 2012), which is discussed further in Chapter Six.

The UK 2008 NICE guidelines for the management of FH recommends a target reduction in LDL-C with treatment of greater than 50% from baseline, which was achieved in less than half of patients in this study. Furthermore, reduction of LDL-C below the ESC/EAS guideline of 2.5 mmol/l for moderate risk subjects (Catapano et al. 2011) was achieved in only about 15% of the patients, and only one patient had post-treatment LDL-C below the 1.8 mmol/l target for individuals at a very high risk. The mechanism of action of statins drives the expression of the *SREBP* transcription factor, which activates the expression of *LDLR* but also activates expression of *PCSK9*, which results in greater degradation of LDL-receptors. Inadequate LDL-C reduction in most patients highlights the need for more effective lipid-lowering drug therapy, and novel treatments, such as PCSK9 inhibition, may offer new therapeutic opportunities to achieve this (Raal et al. 2012).

IV. CHAPTER FOUR - Mutation detection, spectrum and frequency in FH patients of South-Asian background

4.1. Introduction

The World Health Organization (WHO) has recently reported that the leading cause of death in India is cardiovascular disease (CVD) ((WHO) 2011), and the number of deaths from CVD is steeply increasing with 17.8% rise between 1990-2004 (Yusuf et al. 2001, Setia et al. 2012). FH is a monogenic risk factor for premature CVD, which can be identified by DNA testing. The urgent need for identification of FH mutations in the Indian population has been recently highlighted and the importance of the cascade testing to identify affected relatives has been emphasized (Setia et al. 2012). The FH mutation frequency and spectrum in the Indian population is highly unexplored with only a handful of studies reporting mutations characteristic to patients of Indian origin (Kotze et al. 1997, Ashavaid et al. 2000, Taylor et al. 2010b). This study aimed to assess the FH mutation frequency and spectrum in hypercholesterolaemic patients of Indian origin. Patients' samples were collected across several lipid clinics in England. A number of reports demonstrated that individuals of Indian origin manifest CVD at lower levels of risk factors, including plasma lipid levels, than other populations (Mohan et al. 2001, Goel et al. 2003). It has been suggested that variations in cultural and environmental factors such as diet, physical activity, and body weight may be responsible for the difference (Mohan et al. 2001). For the above reason, the Simon Broome criteria LDL-C cutoff for FH (4.9 mmol/l) has been lowered to 4.2 mmol/l during the patient's recruitment for this study. Thus this study aimed to assess if South-Asian patients carrying an FH-causing mutation present with lower plasma LDL-C levels than individuals of European descent.

The second part of this study reports a case of homozygous FH in a large consanguineous family from Mumbai, India. The aim of this investigation was to test the severely affected proband who was diagnosed with multiple xanthomas and seven of his relatives for mutations in FH genes. Genetic characterization of the family members would provide an early diagnosis and help to introduce the appropriate treatment.

4.2. Results – South-Indian patients living in England

4.2.1. Patients characteristics

The objective of this study was to test 100 hypercholesterolaemic patients from Indian Sub-continent for FH mutations. However, only 34 samples were collected across several lipid clinics in England and referred for the molecular genetic testing. 32 of the patients were adults, with mean age of 45 years. None of the patients presented with tendon xanthoma thus according to the Simon Broome clinical criteria for FH all patients were diagnosed as possible FH (PFH). One of the patients was seven years old at the time of the study and therefore was excluded from the overall lipid analysis of adult patients. Her pre-treatment lipid values were: TC = 7.6mmol/l, LDL-C = 5.8 mmol/l, HDL-C = 1.4 mmol/l and TG = 0.9 mmol/l. The patient's elevated cholesterol was not treated at the time of the study because of the young age of the patient. In addition, a 17 years old female (patient's ID 2011) was referred as a suspected homozygous FH case. She presented with pre-treatment lipid levels of TC = 16.9mmol/l, LDL-C = 14mmol/l, HDL-C = 2mmol/l, and pre-

treatment TG = 1.9 mmol/l, with family history of hypercholesterolaemia in both parents and was also excluded from the cohort baseline characteristics analysis. Overall, the South-Asian patients' cohort had significantly lower pre-treatment TC (7.68mmol/l ±0.9) than the DFH (9.79mmol/l, $p=8x10^{-12}$), PFH (8.71mmol/l, $p=6x10^{-6}$) and UH (8.47mmol/l, $p=5x10^{-3}$) patients from the Oxford FH study. The pre-treatment LDL-C of the South-Asian cohort (5.33mmol/l ±1.1) was similar to patients categorised as unclassified hypercholesterolaemia in the Oxford FH cohort, however significantly lower than the DFH (6.93mmol/l, $p=8x10^{-6}$) and PFH (6.12mmol/l, $p=2x10^{-3}$). Baseline characteristics of South-Asian patients in comparison to the Oxford FH cohort are shown in Table 21.

Table 21

Baseline characteristics of the South-Asian FH cohort in comparison to the Oxford FH. Lipid levels (in mmol/l) are shown as geometric means (with approximate standard deviation), because they were not normally distributed. P value between two groups was calculated on log transformed data using Welch Two Sample t-test. Significant p values are in bold.

Variable	South-Asian FH (n=32)	Oxford DFH (n=65)	South-Asian vs. Oxford DFH	Oxford PFH (n=150)	South-Asian vs. Oxford PFH	Oxford UH (n=74)	South-Asian vs. Oxford UH
Gender							
% Male (N)	44% (14)	48% (34)		47% (70)		58% (43)	
Mean age (SD)	45 (14)	58 (13)		54 (14)		54 (15)	
Pre-treatment TC	7.68 (0.9)	9.79 (1.66)	**$8x10^{-12}$**	8.71 (1.27)	**$6x10^{-6}$**	8.47 (1.92)	**$5x10^{-3}$**
Pre-treatment HDL-C	1.37 (0.4)	1.38 (0.35)	0.91	1.46 (0.40)	0.32	1.39 (0.39)	0.92
Pre-treatment TG	1.79 (0.7)	1.22 (0.56)	**$5x10^{-3}$**	1.53 (0.72)	0.23	2.00 (0.81)	0.53
Pre-treatment LDL-C	5.33 (1.1)	6.93 (1.61)	**$8x10^{-6}$**	6.12 (1.15)	**$2x10^{-3}$**	5.45 (1.72)	0.47
Post-treatment TC	5.47 (1.3)	5.95 (1.01)	0.15	5.46 (1.13)	0.91	5.36(1.17)	0.71
Post-treatment HDL-C	1.48 (0.5)	1.31 (0.36)	0.11	1.44 (0.32)	0.53	1.40 (0.46)	0.76
Post-treatment TG	1.53 (0.8)	1.12 (0.52)	**0.03**	1.25 (0.60)	0.12	1.21 (0.38)	0.10
Post-treatment LDL-C	3.13 (1.2)	4.1 (1.04)	**$3x10^{-3}$**	3.29 (1.00)	0.62	3.65 (1.57)	0.45

4.2.2. Mutation spectrum

South-Asian patients were screened for mutations in the promoter and coding region of *LDLR*, for the *APOB* p.R3527Q mutation and for the *PCSK9* p.D374Y mutation. The *LDLRAP1* gene was not analysed since all patients had a dominant pattern of FH inheritance. Eight patients were found to have an FH mutation, all of which were located in the *LDLR* gene. Three of the patients were heterozygous for the c.2054C>T (p.P685L) allele, also called Gujarati FH, whereas patient 2011 was found to be homozygous for the variant. The mutation is positioned within the EGF precursor homology domain repeat C and it has been shown to affect the intracellular transport of the LDL receptor as well as the ligand binding (Soutar et al. 1989). Another mutation located in the EGF precursor homology domain, found in one of the patients, was a non-synonymous change at c.1285G>A, which mutates the Valine 429 residue to Methionine. The p.M429 allele, also known as FH Afrikaner-2, has been demonstrated to increase the rate of LDL receptor degradation (Leitersdorf et al. 1989). Two nonsense mutations were identified, one in exon 3 of the gene (c.285C>A → p.C95*), and one in exon 9 (c.1257C>G → p.Y419*), which was found in the seven years old patient of Indian Punjabi origin. Both variants predicted to cause a formation of the truncated peptide have been reported before (Hirayama et al. 1998, Zakharova et al. 2001, Van Gaal et al. 2001). There was one splice site mutation found in intron 10 of *LDLR*. The c.1587-1G>A variant mutates the first nucleotide of the 3' splice acceptor, and it has been reported in a

Japanese cohort (Yu et al. 2002). All identified mutations and their *in silico* predictions are summarised in Table 22.

Table 22

Summary of FH mutations identified in the South-Asian cohort.

Mutation type/Exon	Variant position	PolyPhen	SIFT	Mutation Taster	Number of carriers
Nonsense					
3	c.285C>A (p.(Cys95*))	NA	NA	Disease causing	1
9	c.1257C>G (p.(Tyr419*))	NA	NA	Disease causing	1
Missense					
9	c.1285G>A (p.(Val429Met)	Probably damaging	Not tolerated	Disease causing	1
14	c.2054C>T (p.(Pro685Leu)	Probably damaging	Not tolerated	Disease causing	3 + 1 homozygote
Splicing					
intron 10	c.1587-1G>A	NA	NA	Disease causing	1

4.2.3. Mutation carriage and patients' lipid levels

An FH mutation was detected in 17.6% of 34 Indian patients. The mutation detection rate was similar to the Unclassified Hypercholesterolaemia cohort (14.3%, p= 0.8) studied in Chapter One. The mean pre-treatment TC and LDL-C values in the mutation positive patients were higher than in the mutation negative (8.46 vs. 7.51 mmol/l, and 6.26 vs. 5.11 mmol/l, respectively), however the difference was not statistically different. Baseline characteristics of both groups are summarized in Table 23. In comparison to the FH patients from Oxford, the pre-treatment LDL-C level in the mutation positive patients did not differ between the cohorts, however it was significantly lower in the mutation negative Indians (p=0.002) Figure 28. All but one mutation positive patient had the pre-treatment TC and LDL-C levels above the Simon Broome FH criteria cutoff (>7.5 mmol/l and >4.9 mmol/l, respectively). One of the patients (22 years old) who was identified to carry the p.P685L mutation had pre-treatment TC and LDL-C levels below the defined cutoff, TC = 6.7 mmol/l and LDL-C = 4.6 mmol/l. Lipid measures were not available for the carrier of the p.C95* (c.285C>A) mutation.

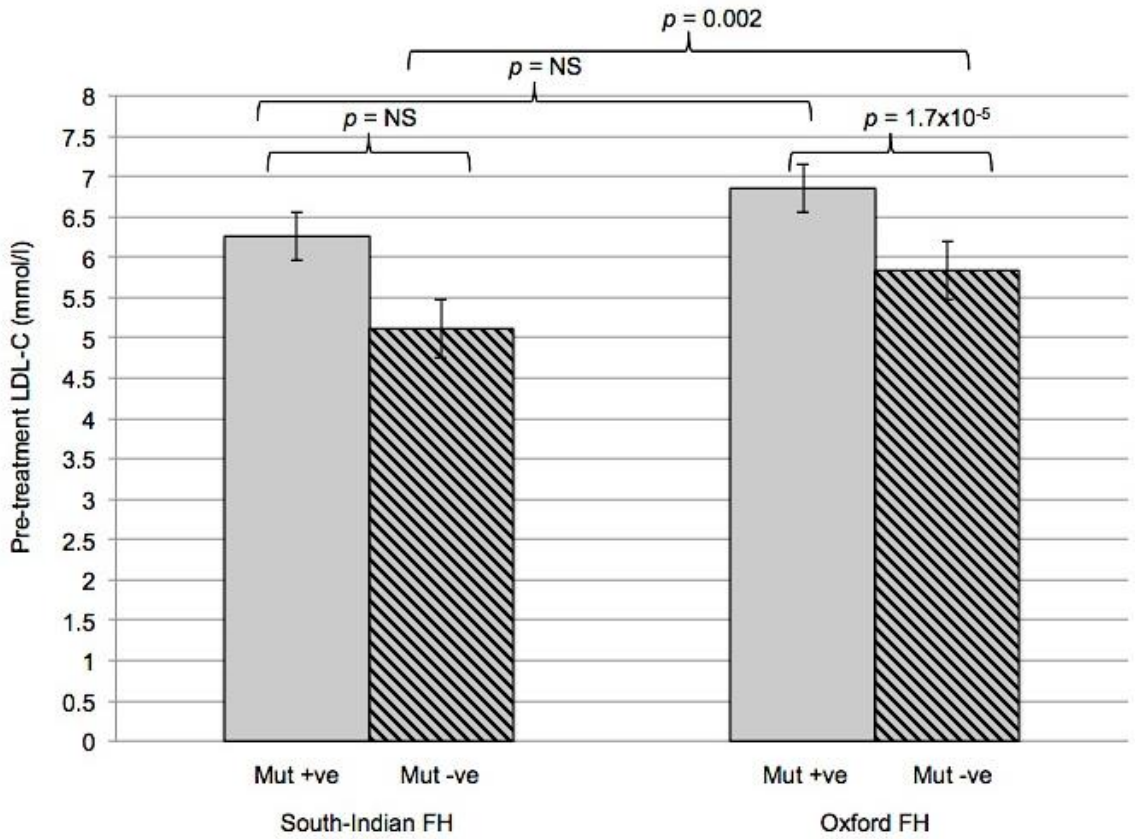
Table 23

Baseline characteristics of South-Asian FH cohort by presence of a mutation. Lipid levels (in mmol/l) are shown as geometric means (with approximate standard deviation), because they were not normally distributed. P value between two groups was calculated on log-transformed data using Welch Two Sample t-test.

Variable	Mutation positive (n=6)	Mutation negative (n=26)	p value
Gender			
% Male (N)	50% (2)	48% (12)	
Mean age (SD)	37 (14.1)	47 (13.3)	0.17
Pre-treatment TC	8.46 (1.6)	7.51 (0.6)	0.23
Pre-treatment HDL-C	1.31 (0.3)	1.38 (0.4)	0.70
Pre-treatment TG	1.50 (0.6)	1.89 (0.7)	0.30
Pre-treatment LDL-C	6.26 (1.6)	5.11 (0.8)	0.14
Post-treatment TC	6.13 (1.9)	5.38 (1.2)	0.57
Post-treatment HDL-C	1.53 (0.1)	1.47 (0.5)	0.63
Post-treatment TG	1.44 (0.3)	1.54 (0.8)	0.71
Post-treatment LDL-C	3.84 (1.8)	3.04 (1.1)	0.53

Figure 28

Pre-treatment LDL-C levels in mutation positive and mutation negative FH patients of Indian origin in comparison to FH patients from Oxford. NS= not significant



Using the Oxford FH study quartile cutoffs for pre-treatment TC and TG, the mutation detection rate in Indian FH patients increased with each pre-treatment TC quartile and decreased with increasing pre-treatment TG level, as shown in Table 24, which agreed with the findings reported in Chapter One.

Table 24

FH mutation detection rate in South-Indian FH cohort by quartile of pre-treatment TC and pre-treatment TG levels (the quartiles used came from the Oxford FH study from Chapter Three). P values were calculated using logistic regression.

Quartile of pre-treatment TC (mmol/l)		N	Mutation +ve (%)	Quartile of pre-treatment TG (mmol/l)		N	Mutation +ve (%)
Q1	<=8.0	14	2 (14)	Q1	<=1.0	2	1 (50)
Q2	8.1-8.7	5	1 (20)	Q2	1.10-1.32	3	1 (33)
Q3	8.8-9.9	2	1 (50)	Q3	1.33-2.15	11	2 (18)
Q4	>10	1	1 (100)	Q4	2.16-4.30	6	1 (17)
p value			0.03				0.3

4.3. Results – Consanguineous family from Mumbai, India.

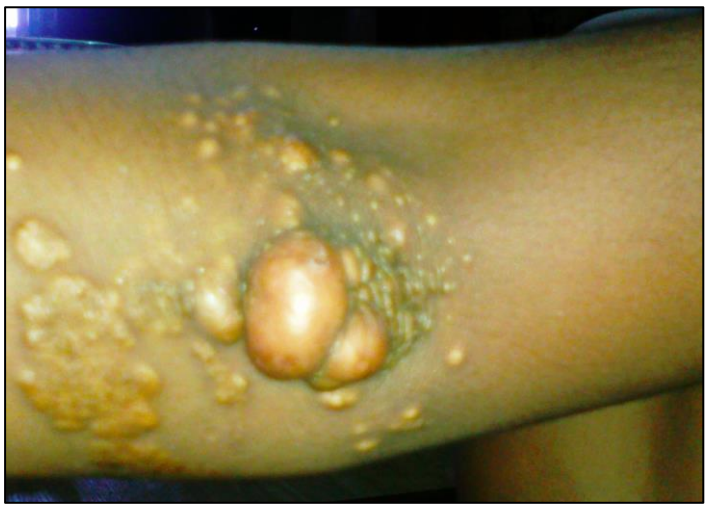
4.3.1. Family characteristics

In addition to the samples collected in England, DNA from a consanguineous Indian (from Mumbai) family of eight was collected. The family proband (referred here as FH13), a nine years old boy was diagnosed with severe cholesterol levels (TC = 22.1mmol/l and LDL-C = 19mmol/l) and multiple xanthomas around his elbow and at the back of his legs, as shown in Figure 29. Both of his parents were said to be hypercholesterolaemic. An extensive family history of myocardial infarction, coronary bypass surgeries and sudden cardiac deaths was reported in relatives of both parents. Available baseline characteristics of the family are shown in Table 25.

Figure 29

Images of physical examination of FH13 patient, nine years old boy with severe plasma cholesterol levels (TC = 22.1mmol/l and LDL-C = 19mmol/l). Both A and B show a typical of severe DFH or homozygous FH physical manifestation of cholesterol deposition within the skin. A: multiple xanthomas on the patient's elbow; B: eruptive xanthoma on the back of both legs and on the buttocks of the patient.

A.



B.

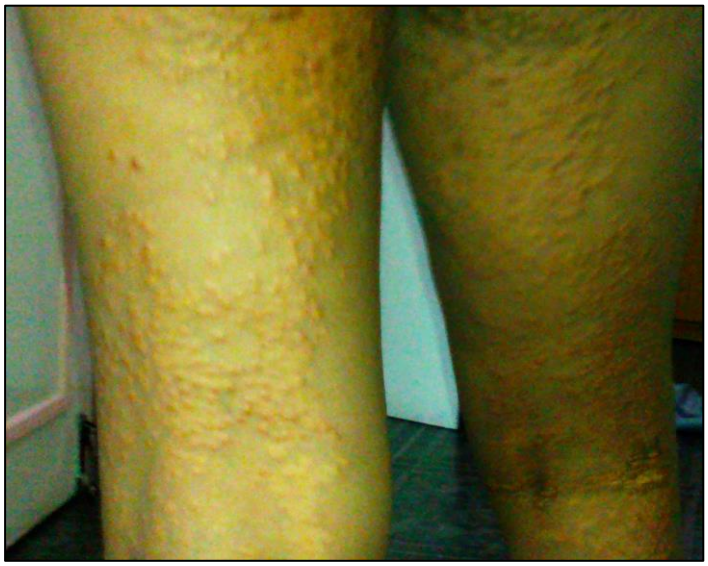


Table 25

Baseline characteristics of the Mumbai consanguineous family. Other measurements were not available.

Relation	Gender	Age (years)	Sample ID	TC mmol/l	LDL-C mmol/l
Proband	M	9	FH13	22.1	19
Mother	F	35	FH10	7.4	5.7
Father	M	38	FH4	7.7	N/A
Father's brother	M	42	FH11	8.5	6.2
Cousin	F	12	FH12	N/A	N/A
Father's brother	M	48	FH14	20	N/A
Cousin	M	23	FH15	5.4	3.63
Cousin	M	17	FH16	7.8	6.7

N/A = not available

4.3.2. FH mutation spectrum

Mutation testing of the consanguineous family members identified two variants in the *LDLR* gene. HRM detected changes of the melting temperature in fragments of exons 7 and 11 *LDLR* in several family members, as shown in **Figure 30**. Exons 7 and 11 were examined further by Sanger sequencing. A previously unreported variant in exon 7 (absent from the UCL FH mutation database and other publically available databases), c. 948C>G, was identified in four relatives, however it was absent from the proband and proband's parents. The variant changes Asparagine to Lysine at the position 316, which was predicted as 'probably damaging' by PolyPhen, 'not tolerated' by SIFT and 'disease causing' by Mutation Taster. The other variant was located in intron 11, within the consensus splice site at the position c.1705+1G>C, which is predicted to alter the protein splicing as shown in **Figure 31**. The intronic mutation has been previously observed in a Japanese family, where it reduced the LDL-C uptake by 66%, in comparison to a control (Hattori et al. 2002). Family members FH11 and FH14 were heterozygous for both variants, whereas FH12 and FH15 were found to carry only the exon 7 variant. The c.1705+1G>C mutation on its own was identified in FH4 (proband's father), FH10 (proband's mother), and FH16. The family proband FH13 was found to be homozygous for the splice site mutation and didn't carry the p.N316K in exon 7.

Figure 30

Normalised melting curve of fluorescence (F, Y axis) and temperature (T, X axis) of A: exon 7 *LDLR.* Samples FH11, FH14 and FH15 in addition to the c. 948C>G are also heterozygous for rs12710260 SNP; B: exon 11 *LDLR*. Sample FH4 in addition to the c.1705+1G>C is also heterozygous for rs5929 SNP. Insets ‘i’ and ‘ii’ represent difference graphs for the fragments with deltaF/deltaT (on Y axis) and T (on X axis).

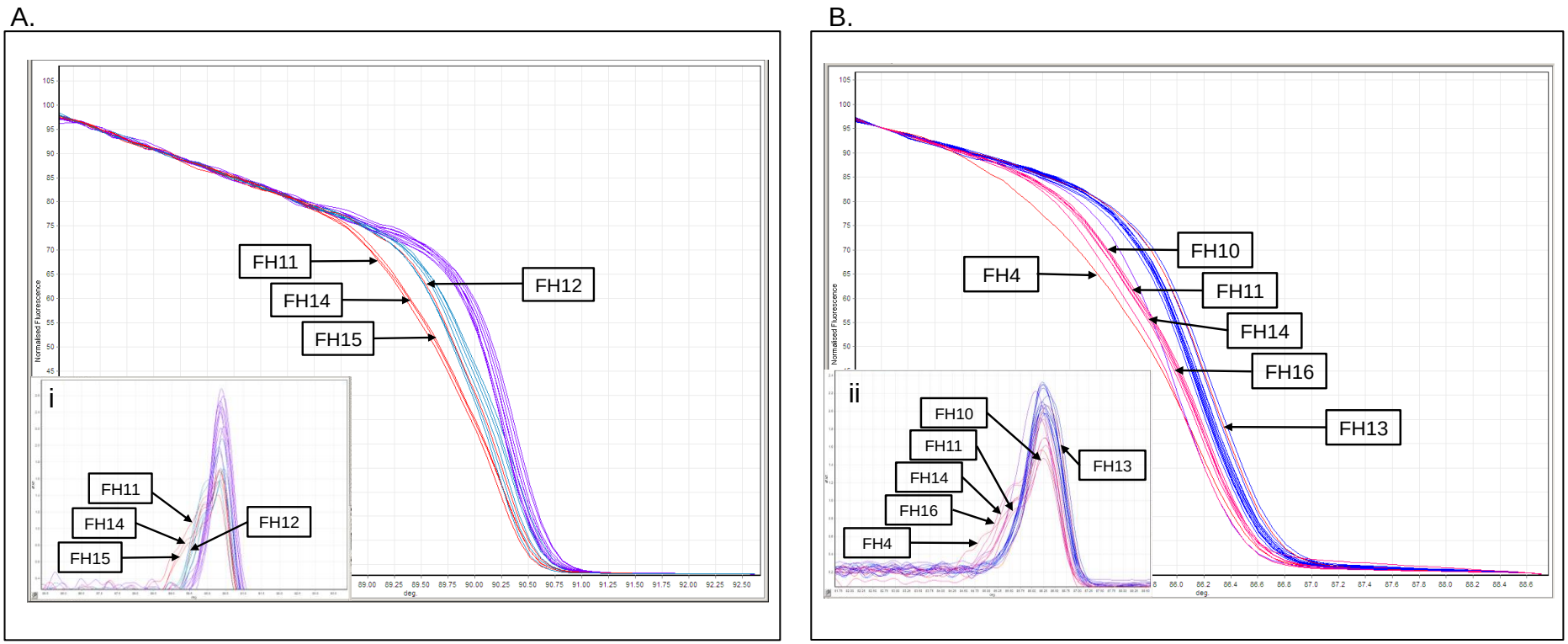


Figure 31

Splice site analysis of c.1705+1G>C variant in intron 11 of *LDLR*. Scores for each predicted donor and acceptor splice site were calculated using the Human Splicing Finder. Marked in blue are the naturally occurring (wild type) splice sites for *LDLR* exon 11. Marked in red are changes in splice site score when c.1705+1G is mutated to C. When mutated, the donor splice site is no longer recognised, which may lead to exon skipping of the entire exon, which will cause a frame shift.

Predicted Wild Type Donor splice site

Score	Exon	Intron
1.00	accctag**gt**atgttc	

Predicted Donor splice site for c.1705+1G>C

Score	Exon	Intron
None		

Predicted Wild Type Acceptor splice site

Score	Intron	Exon
0.56	ggatcctcccccgccctcc**ag**	cctcacagctattctctgtc
0.73	ccccgccctccagcctcac**ag**	ctattctctgtcctcccacc
0.50	tattctctgtcctcccacc**ag**	cttcatgtactggactgact
0.57	tcaccctaggtatgttcgc**ag**	gacagccgtcccagccaggg
0.89	tttgtttttttctttttttg**ag**	atgaggtcttggtctgtcac

Predicted Acceptor splice site for c.1705+1G>C

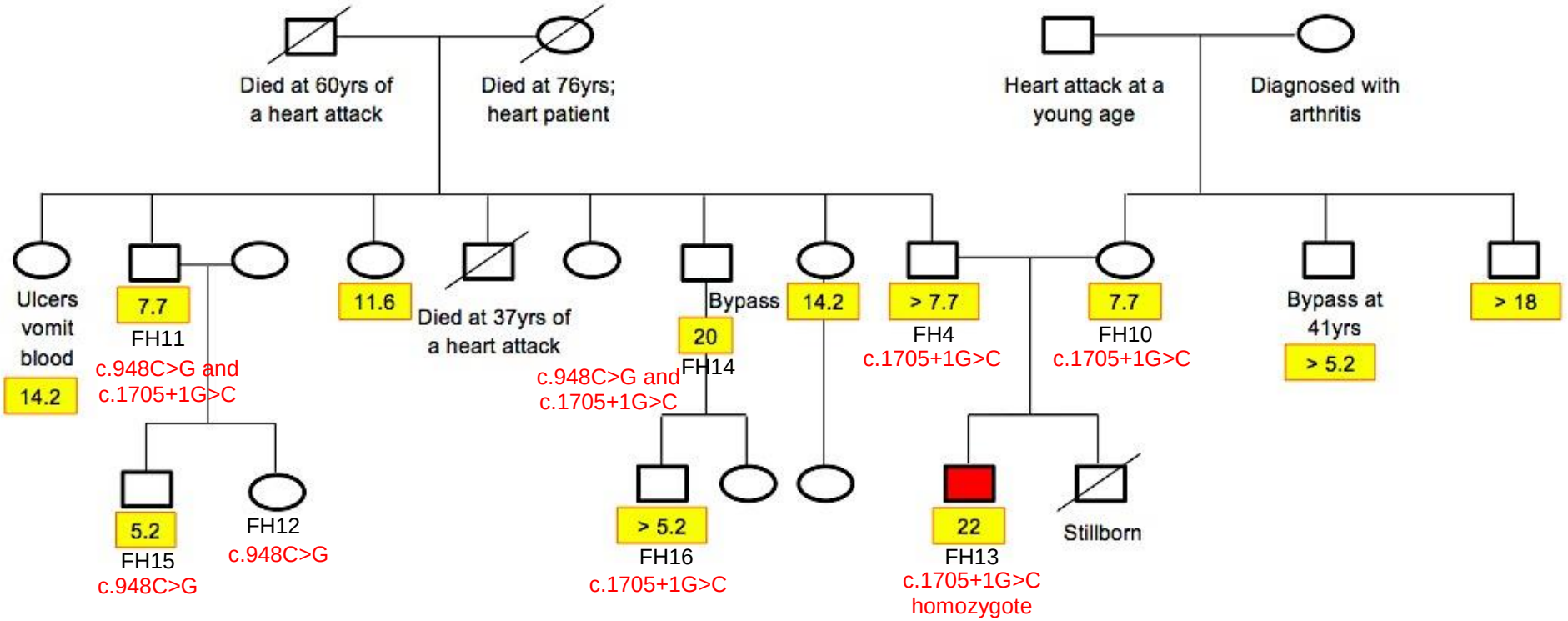
Score	Intron	Exon
0.56	ggatcctcccccgccctcc**ag**	cctcacagctattctctgtc
0.73	ccccgccctccagcctcac**ag**	ctattctctgtcctcccacc
0.50	tattctctgtcctcccacc**ag**	cttcatgtactggactgact
0.80	tcaccctaggtatgttcgc**ag**	gacagccgtcccagccaggg
0.89	tttgtttttttctttttttg**ag**	atgaggtcttggtctgtcac

4.3.3. Mutation carriage and lipid levels

Pre-treatment TC and LDL-C measurements were available for seven members of the Mumbai consanguineous family, whereas post-treatment values were not available at all. A summary of the mutation carriage and lipid levels within the family is presented in Figure 32. One of the proband's cousins, FH15, was found to carry the novel variant in exon 7 (c.948C>G), however exhibited normal plasma cholesterol levels. Two family members, brothers FH11 and FH14, were compound heterozygotes for both identified variants, but only FH14 presented with extreme pre-treatment TC levels, characteristic of homozygous FH. The family doctor was contacted to confirm the initial pre-treatment cholesterol measurements for FH15 and FH11 patients, whose clinical information did not agree with the molecular diagnosis. Further clinical assessment was requested for the family members, however due to a difficult access to the patients to date no tests have been carried out.

Figure 32

A family pedigree of a nine years old homozygous FH proband (in red) from Mumbai. Available values of total cholesterol (in mmol/L) are shown in yellow boxes.



4.4. Discussion

This study of FH mutations in patients from the Indian Sub-Continent identified five different FH-causing variants in the *LDLR* gene, in eight out of 34 studied patients. There were no *APOB* and no *PCSK9* FH mutations observed in this cohort. A heterozygous mutation was detected in 17.6% of unrelated patients, which was low is comparison to the 73% rate in the DFH and 27% in the PFH patients from Oxford (studied in Chapter One), however similar to the 14.3% mutation detection rate in UH group from Oxford. One of the reasons for the low mutation detection rate could be the sample size of this study. The initial target of collecting 100 unrelated samples has not been achieved in the time of my PhD course, which remained to be the main limitation of the analysis. The TC/LDL-C cutoff used for the clinical diagnosis of FH could also be the reason for low mutation detection rate. The Simon Broome FH criteria cutoffs have been lowered from 7.5 mmol/l to 6.7 mmol/l for TC, and from 4.9 mmol/l to 4.2 mmol/l for LDL-C, because of the evidence for lower levels of CVD risk factors in Indian populations (Mohan et al. 2001, Goel et al. 2003). However, this study demonstrated that there was no difference between pre-treatment LDL-C levels of the mutation positive patients of Indian origin and the mutation positive FH patients from the Oxford cohort, thus the Simon Broome clinical diagnostic criteria for FH are suitable for patients of Indian background. As shown in Chapter One, the mutation detection rate highly depends on the pre-treatment TC and TG levels. In this study, 13 of the recruited patients did not reach the Simon Broome 7.5 mmol/l diagnostic cutoff for the pre-treatment TC. Furthermore, pre-treatment TC

level was not available for five of the participants. The mean pre-treatment TG in Indians (1.53 ± 0.8 mmol/l) was higher than the levels observed in the Oxford cohorts (1.12 ± 0.52 in DFH, 1.25 ± 0.6 in PFH and 1.21 ± 0.38 in UH), although the difference was statistically significant only between the Oxford DFH and the Indian FH (p=0.03). This study also confirmed the finding from Chapter One, of the importance of high pre-treatment cholesterol and low triglyceride levels as predictors of the likelihood of detection of an FH-causing mutation, since the proportion of patients carrying an FH mutation increased with higher pre-treatment TC and decreased with higher pre-treatment TG. The most commonly observed mutation in the South-Indian cohort was the p.Pro685Leu (found in 43% of probands), which is known as the Gujarati FH mutation. However overall, the mutation spectrum in South-Indians living in the UK differed from the spectrum reported in Indians from South Africa (Rubinsztein et al. 1993). None of the detected mutations was observed in the studied South Africans of Indian origin. Although this study did not identify any novel mutations, none of them were previously associated with Indian populations (apart from the p.Pro685Leu). Two of the variants (c.285C>A (p.Cys95*) and c.1587-1G>A) were initially found in Japanese patients, whereas c.1257C>G (p.Tyr419*) was reported in Belgium and c.1285G>A (p.Val429Met) is an Afrikaner FH mutation. The difference in FH mutation spectrum could be the result of a lack of reports on South-Indian FH but it may also reflect the genetic heterogeneity of the Indian population living in the UK.

One of the limitations of this study is a low patients recruitment rate. Although reasons for patients refusal to participate in this study were not recorded,

previous genetic studies demonstrated reduced participation among racial minorities (Sterling et al. 2006) mainly due to concerns about discrimination, confidentiality, or lack of confidence in researchers. A weakness of the study was that pre-treatment TC measurements were not available for 5 out of 34 participants, and pre-treatment TG levels were not recorded for 10 individuals. This was because some patients who met the criteria in other respects (e.g. personal family history of premature CHD) were referred to the lipid clinic whilst on diet or lipid-lowering medication, and it would not have been ethical to stop medication to collect data for this study. Another limitation to this project was that patients were not screened for large rearrangements in *LDLR* because of a high cost of the test.

Although the need for development of a sensitive, reliable and inexpensive kit for a routine search for FH-causing variants specific to the Indian population has been emphasized (Ashavaid et al. 2000), many more studies on this ethnic group will be required since no common FH mutations have been identified, apart from the p.Pro685Leu. Results of this study suggest high genetic heterogeneity of FH mutation in patients of Indian descent, which may limit the development of a population-specific FH mutation detection assay.

The consanguineous family from Mumbai was found to carry two predicted FH-causing variants, in different parts of the *LDLR* gene. The severely affected family proband was homozygous for a splice site mutation in intron 11 of the gene, c.1705+1G>C. This variant has been previously observed in a Japanese family, where a heterozygous proband manifested with multiple

xanthomas. Proband's parents were both heterozygous for the mutation, with cholesterol levels about 3 fold lower (~7.7 mmol/l) in comparison to the proband. Four other family members were identified to carry a novel variant, p.Asn316Lys, located in exon 7 of *LDLR*, which codes for the first part of the EGF precursor homology domain. SIFT, PolyPhen and the Mutation Taster *in silico* tools consistently predicted the variant to have a damaging effect on the protein. However the effect of p.Asn316Lys on lipid levels in the family was not consistent. The 23 years old proband's cousin, a carrier of the p.Asn316Lys variant, had plasma lipids within the normal range (TC=5.4 mmol/l, LDL-C=3.63 mmol/l), whereas the cousin's father who was found to be a compound heterozygote for the p.Asn316Lys and c.1705+1G>C, manifested heterozygous FH lipid levels (TC=7.7 mmol/l), which was about 3 fold lower than his brother (TC=20 mmol/l) – also compound heterozygote. Therefore the data suggests that the variant p.Asn316Lys is benign, however the extreme TC level in FH14 remains unexplained thus further investigation is required. Patient FH14 is likely to carry another LDL-rising variant outside of the screened region. Sanger sequencing of the entire *PCSK9* coding region in FH11 and FH14 patients did not identify any LDL-lowering or LDL-raising alleles. If resources were available, whole exome sequencing of both samples would be recommended as an informative experiment to identify additional causes.

V. CHAPTER FIVE - LDL-C SNPs score analysis for polygenic hypercholesterolaemia

5.1. Introduction

The FH mutation detection rate in DFH patients varies between 60-90% (Graham et al. 2005, Taylor et al. 2010b, Futema et al. 2013), suggesting that there are other genetic causes, located outside of the currently screened regions, which are yet to be identified. Recent report provided new evidence that a substantial proportion of individuals with a clinical diagnosis of FH inherit a combination of small-effect changes in several genes (polygenic) rather than a large-effect mutation in a single gene (monogenic) (Talmud et al. 2013). The proposed molecular test for polygenic hypercholesterolaemia uses findings of the Global Lipid Genetic Consortium (GLGC) meta-analysis of genome-wide association studies, which identified several loci where common variants affect LDL-C concentration (Teslovich et al. 2010), and results of another study that showed that a proportion of individuals carrying several LDL-C-raising single nucleotide polymorphisms (SNPs) have LDL-C concentrations that exceed the Simon Broome FH diagnostic LDL-C threshold of 4.9 mmol/l (Talmud et al. 2009).

In this chapter, LDL-C SNPs scores, derived from 12 common LDL-C-raising SNPs in 11 genes (two in *APOE*), were calculated for patients with European ancestry who had the clinical diagnosis of definite familial hypercholesterolaemia and were included in the whole exome sequencing project, as discussed in Chapter Seven, and compared with those of healthy men and women of European ancestry from the UK Whitehall II (WHII) cohort study. The aim was to test the hypothesis that familial hypercholesterolaemia can be caused by an accumulation of common small-effect LDL-C-raising

alleles. Identifying patients with low LDL-C SNPs score would help to increase the chance of finding novel monogenic causes of FH when analysing the whole exome sequencing data (Chapter Seven). In addition, this study intended to validate the findings using three independent cohorts from Canada, Greece and Holland, which in total included 109 mutation positive and 122 mutation negative FH patients.

5.2. Results

5.2.1. Healthy comparison cohort

A healthy comparison group of white men and women was drawn from the UK Whitehall II study (WHII), discussed in section 2.3.2. Genotypes of 3020 individuals were used for the final LDL-C gene score comparison. Baseline characteristics of the WHII study are shown in Table 26.
Frequencies of the 12 SNPs used for the score calculations in the FH cohort and the WHII study are shown in Table 27.

Table 26

Baseline characteristics of the LDL-C score control cohort, the WHII study. Pre-treatment TG levels were not available.

	WHII (n=3020)
% male	76% (2308/3020)
Age (years)	49.0 ± 6.0
Pre-treatment TC (mmol/l)	6.4 ± 1.1
Pre-treatment LDL-C (mmol/l)	4.4 ± 1.0
Post-treatment LDL-C (mmol/l)	4.4 ± 1.0
Post-treatment HDL-C (mmol/l)	1.4 ± 0.4
Post-treatment TG (mmol/l)	1.4 ± 1.1

Table 27

LDL-C gene score calculations with observed minor allele frequencies in the studied cohort and the WHII control. Weights for risk alleles were acquired from (Teslovich et al. 2010), and from (Bennet et al. 2007) for *APOE* genotypes.

	Chromosome	Gene	Minor allele	Common allele	Weight for score calculation	Minor allele frequency	
						FH (n=109)	WHII controls (n=3020)
rs2479409	1	*PCSK9*	G*	A	0.052	0.38	0.35
rs629301	1	*CELSR2*	G	T*	0.15	0.14	0.21
rs1367117	2	*APOB*	A*	G	0.1	0.38	0.33
rs4299376	2	*ABCG8*	G*	T	0.071	0.42	0.32
rs1564348	6	*SLC22A1*	C	T*	0.014	0.25	0.17
rs1800562	6	*HFE*	A	G*	0.057	0.06	0.07
rs3757354	6	*MYLIP*	T	C*	0.037	0.18	0.21
rs11220462	11	*ST3GAL4*	A*	G	0.05	0.16	0.13
rs8017377	14	*NYNRIN*	A*	G	0.029	0.52	0.48
rs6511720	19	*LDLR*	T	G*	0.18	0.04	0.13
rs429358	19	*APOE*	C	T	..	0.2	0.15
rs7412	19	*APOE*	T	C	..	0.01	0.08
ε2ε2	19	*APOE*	..	..	-0.9	..	..
ε2ε3	19	*APOE*	..	..	-0.4	..	..
ε2ε4	19	*APOE*	..	..	0.2	..	..
ε3ε3	19	*APOE*	..	..	0	..	..
ε3ε4	19	*APOE*	..	..	0.1	..	..
ε4ε4	19	*APOE*	..	..	0.2	..	..

*Risk allele (LDL-rising)

The mean weighted LDL-C gene score of the WHII participants was 0.90 (SD 0.23), and WHII individuals in the bottom decile (decile 1) of the LDL-C gene score distribution had a mean LDL-C concentration of 3.76 mmol/l (0.95), whereas those in the top decile (decile 10) had a mean LDL-C concentration of 4.9 mmol/l (0.99), the diagnostic cutoff point for familial hypercholesterolaemia. Participants from WHII in decile 10 of the LDL-C score distribution had a much higher likelihood than those in decile 1 of having observed LDL-C concentrations above the UK NICE recommended diagnostic threshold of 4.9 mmol/l, with 146 (50%) of 292 individuals in decile 10 having observed LDL-C concentrations of greater than the Simon Broome FH criteria cutoff, compared with 35 (12%) of 295 individuals in decile 1, as demonstrated in Table 28. Thus the top decile cutoff of LDL-C weighted score for the WHII population (1.16) was used as a lowest threshold for polygenic hypercholesterolaemia.

Table 28

LDL-C gene scores in WHII cohort by deciles and its correlation with LDL-C level.

	LDL-C weighted score in WHII controls		Measured LDL-C (mmol/L) in WHII controls, mean (SD)	WHII controls with LDL-C >4.9 mmol/L
	Mean (SD)	Range		Measured
Decile 1	0.43 (0.14)	-0.5 to 0.58	3.76 (0.95)	36/299 (12%)
Decile 2	0.66 (0.04)	0.58 to 0.73	3.99 (0.88)	43/296 (15%)
Decile 3	0.77 (0.03)	0.73 to 0.81	4.21 (0.96)	71/300 (24%)
Decile 4	0.85 (0.02)	0.81 to 0.88	4.34 (0.95)	85/298 (29%)
Decile 5	0.91 (0.02)	0.88 to 0.93	4.36 (0.94)	80/300 (27%)
Decile 6	0.96 (0.01)	0.94 to 0.98	4.48 (0.91)	96/298 (32%)
Decile 7	1.00 (0.01)	0.98 to 1.02	4.50 (1.00)	102/295 (35%)
Decile 8	1.05 (0.02)	1.02 to 1.08	4.56 (0.93)	96/292 (33%)
Decile 9	1.12 (0.02)	1.08 to 1.16	4.68 (1.05)	120/294 (41%)
Decile 10	1.23 (0.06)	1.16 to 1.46	4.90 (0.99)	148/295 (50%)

*decile 1 used as reference

5.2.2. Familial Hypercholesterolaemia cohort

LDL-C SNPs score calculations were performed for the whole exome sequencing cohort (studied further in Chapter Seven), which contained 14 DFH mutation negative samples from the Oxford FH study (Chapter Three). Complete genotypes of all 12 LDL-C-rising SNPs were obtained for 109 samples. There were 15 samples for which the score could not be attained because of insufficient DNA. Out of the 109 genotyped FH samples 21 were found to carry an FH mutation, missed by the initial mutation screening methods, which was revealed by the whole exome sequencing analysis. Mutation positive samples were treated separately from the 88 mutation negative FH patients. Overall, 29 of the genotyped mutation negative individuals had a gene score above the 1.16 cutoff. The distribution of the scores in the mutation negative FH patients and the WHII study is shown in Figure 33. In the mutation positive patients' group two samples had high gene score, of whom one had an *LDLR* mutation in exon 11 (c.1690A>C (p.N564H)) found on the same allele as a 9bp deletion in exon 17 (c.2393_2401del9 (p.L799_V801del)), which together have been demonstrated as not fully-penetrant (Mozas et al. 2000), the other was positive for a deletion of a consensus splice site at the 5' of exon 5, c.695-6_698del, which has not been examined *in vivo* to confirm its likely effect on splicing.

The mean LDL-C gene score for the FH mutation negative group was 1.08, which was significantly higher than 0.9 for the WHII study ($p < 2.2x10^{-16}$), and 0.96 for the FH mutation positive group ($p = 0.006$), as shown in Figure 2. The

overall difference between the groups was significant (ANOVA, $p = 1.33 \times 10^{-12}$).

Figure 33

Distribution of the LDL-C SNPs score in mutation negative DFH patients (in purple) and in the healthy WHII cohort (in grey). Red line indicates the LDL-C score top decile cutoff for WHII.

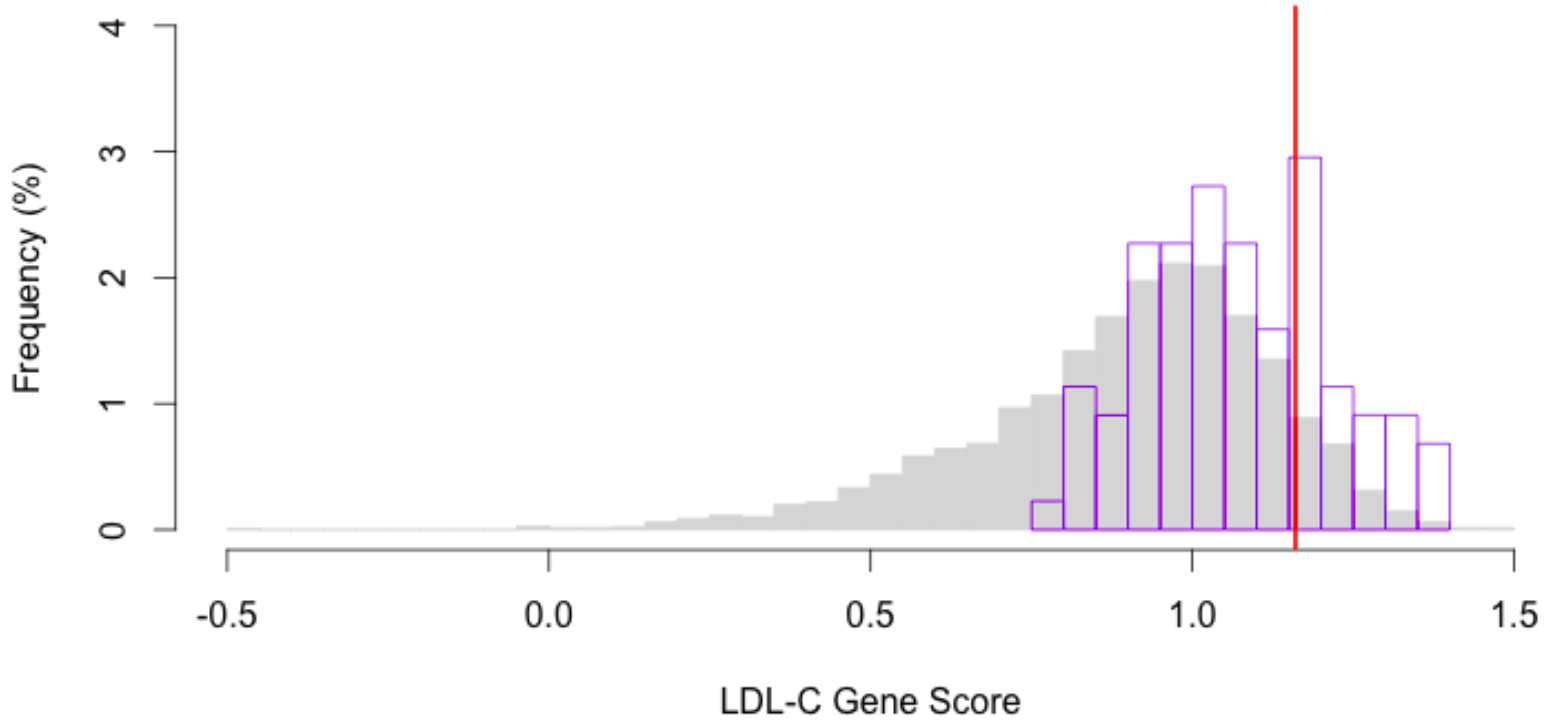
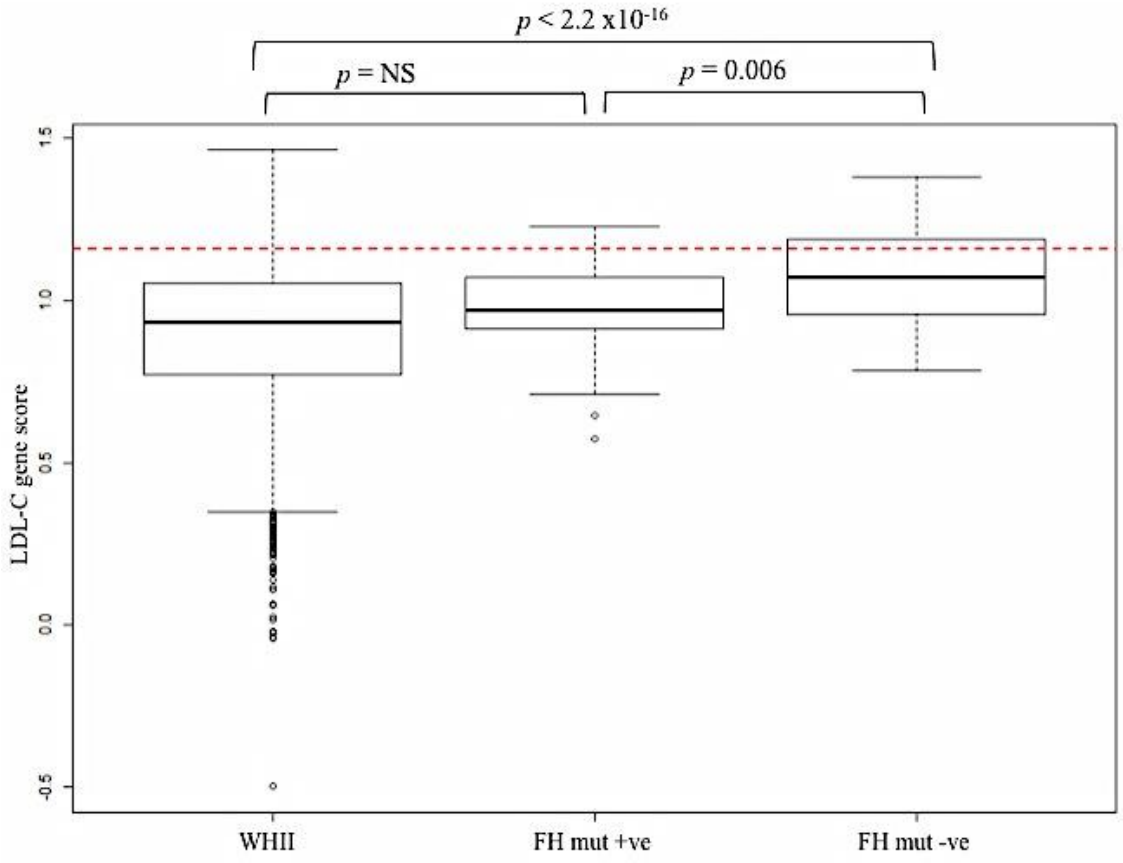


Figure 34

Comparison of the LDL-C SNPs score between the control WHII population (n=3020), the FH mutation positive individuals (n=21), and the FH mutation negative individuals (n=83) in a standard boxplot (the minimum, lower quartile, median, upper quartile and maximum). The overall difference between the groups was highly significant (ANOVA, $p < 2.2 \times 10^{-16}$). Dashed red line indicates the top decile cutoff for the WHII cohort (=1.16).



NS = not significant

5.2.3. FH replication cohort

For validation of the LDL-C SNPs score association three independent cohorts of FH patients were studied. The replication cohort included 78 adult FH individuals from Canada, composed of 39 patients with an identified FH mutation and 39 mutation negative patients. Two groups of FH children were also studied, which included 130 patients from Greece and 23 patients from Holland. Baseline characteristics of the validation cohorts are shown in Table 29. There was no tendon xanthomas observed in children cohorts. Only one person of the mutation negative adults was diagnosed with definite FH, using the Simon Broome criteria.

The highest mean LDL-C SNPs score was observed in the mutation negative FH children from the Dutch cohort (1.04), followed by the mutation negative FH children cohort from Greece (1.00). The LDL-C gene score difference between the mutation negative FH patients from Canada and the healthy WHII population was not statistically different although the mutation negative FH individuals had the highest mean score in this cohort. The summary of the findings in the individual validation cohorts is shown in Table 30.

Table 29

Baseline characteristics of the LDL-C score replication cohorts. A. FH Canada adults cohort; B. FH Greece children cohort; C. FH Holland children cohort.

A.	Canada FH M+	Canada FH M-
Number	39	39
Percent male	43.2	43.2
Age (years)	47.5±11.7	48.3±11.9
Percent with tendon xanthoma	20.5	2.6 (1 person)
Total cholesterol (mmol/L)	9.35±2.43	7.40±2.13
LDL-C (mmol/L)	7.36±2.17	5.24±2.03
HDL-C (mmol/L)	1.25±0.35	1.24±0.34
Triglyceride (mmol/L)	1.73±1.02	2.05±0.99

B.	Greece FH M+	Greece FH M-
Number	70	60
Percent male	61.4	55
Age (years)	6.9±0.45	8.7±0.38
Percent with tendon xanthoma	0	0
Total cholesterol (mmol/L)	7.9±0.16	6.4±0.09
LDL-C (mmol/L)	5.8±0.17	4.2±0.09
HDL-C (mmol/L)	1.5±0.04	1.7±0.04
Triglyceride (mmol/L)	±	±

C.	Holland FH M+	Holland FH M-
Number	0	23
Percent male	NA	43.5
Age (years)	NA	11.7±3.6
Percent with tendon xanthoma	NA	0
Total cholesterol (mmol/L)	NA	6.15±0.72
LDL-C (mmol/L)	NA	4.32±0.68
HDL-C (mmol/L)	NA	1.36±0.35
Triglyceride (mmol/L)	NA	0.85±0.50

M+ = mutation positive

M- = Mutation negative

NA = not applicable

Table 30

Summary of LDL-C gene score results in the replication FH cohorts.

		Replication FH cohort		
		Greece	**Holland**	**Canada**
FH/M-	N	60	23	39
	Mean score	1.00	1.04	0.96
FH/M+	N	70	0	39
	Mean score	0.93	NA	0.92
WHII	N	3020	3020	3020
	Mean score	0.90	0.90	0.90
t-test	FH/M- vs. FH/M+	0.02	NA	0.50
	FH/M- vs. WHII	3.318×10^{-6}	0.0002	0.14
	FH/M+ vs. WHII	0.20	NA	0.60

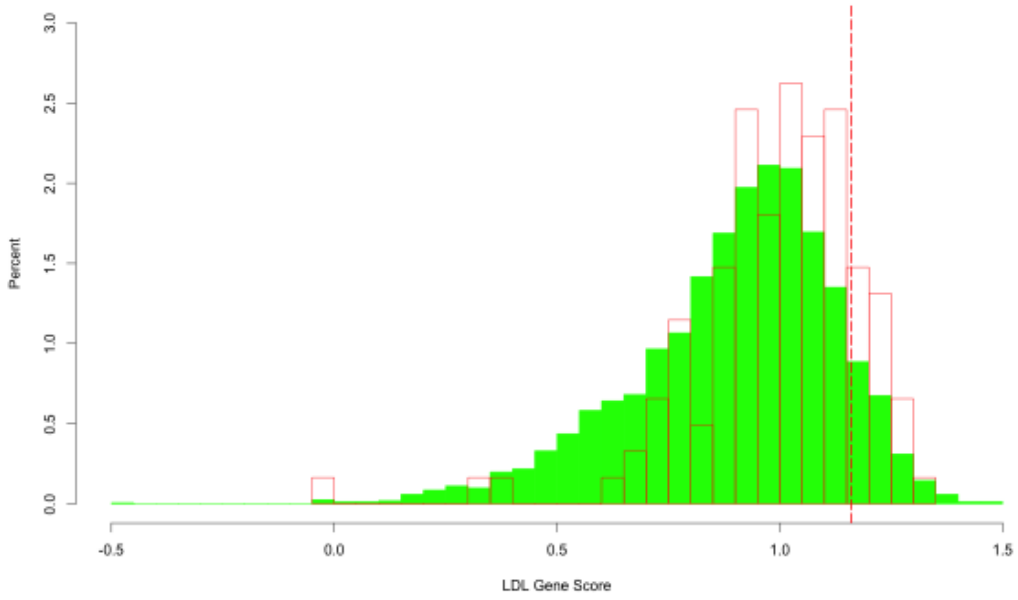
NA = not applicable

There were 122 mutation negative and 109 mutation positive FH patients in the validation studies, when combined. Mutation negative FH individuals had significantly higher LDL-C gene score than the healthy WHII population (0.99 vs.0.89, $p < 4.8 \times 10^{-7}$). The mutation negative patients group had also higher gene score than mutation positive FH patients (0.99 vs. 0.93, $p < 0.009$). Although the FH mutation positive patients had higher mean LDL-C gene score than the WHII population the difference was not statistically different. The LDL-C SNPs score distribution in the replication study is shown in Figure 35. In the mutation negative patients group there were three individuals with relatively low LDL-C gene score, which suggests that they do carry an FH variant with strong effect, which has not been yet identified. The boxplot shown in Figure 36 summarises the difference between groups in the replication cohort.

Figure 35

Distribution of the LDL-C SNPs score in A. mutation negative FH patients from the replication cohorts combined (in red); B. mutation positive FH patients from the replication cohorts combined (in purple) and in the healthy WHII cohort (in green). Dashed red line indicates the LDL-C score top decile cutoff (1.16) for WHII.

A.



B.

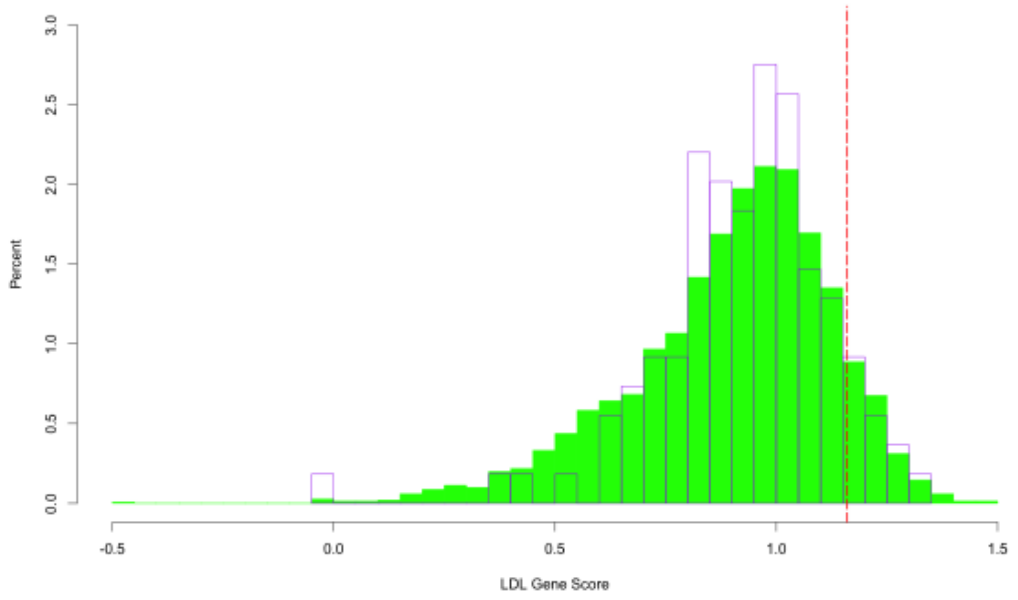
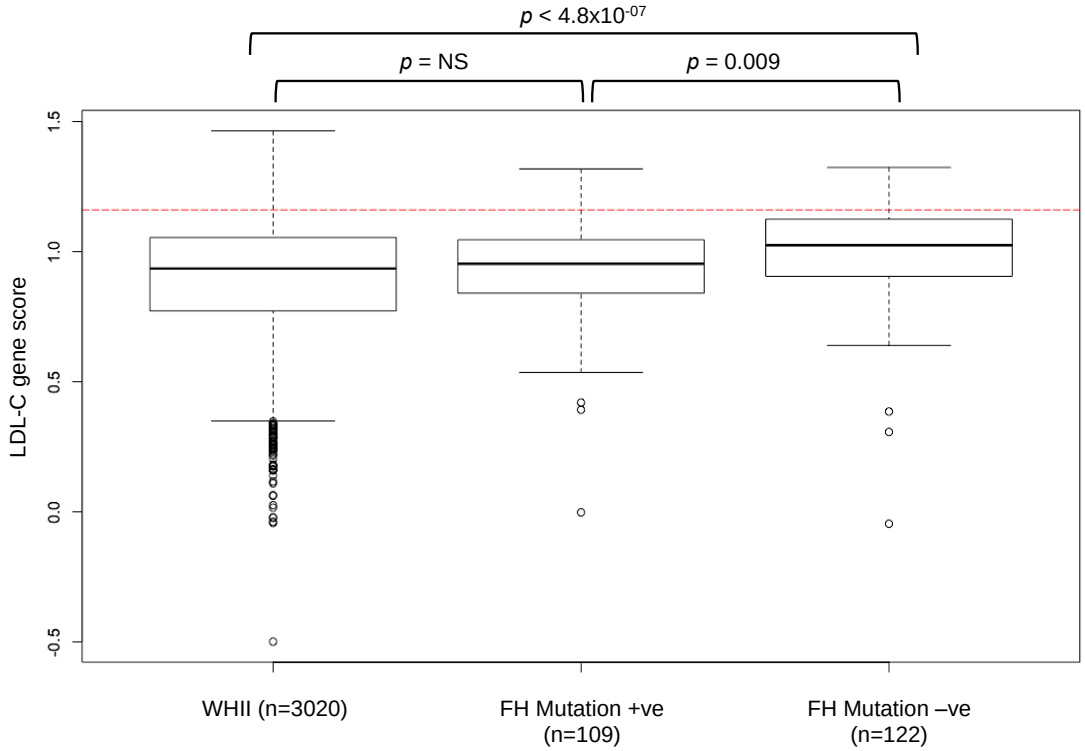


Figure 36

Comparison of the LDL-C SNPs score between the control WHII population (n=3020), the FH mutation positive individuals from replication cohorts combined (n=109), and the FH mutation negative individuals from the replication cohorts combined (n=122) in a standard boxplot (the minimum, lower quartile, median, upper quartile and maximum). Dashed red line indicates the top decile cutoff for the WHII cohort (=1.16).



NS = not significant

5.3. Discussion

The major finding of this chapter supports results reported in the original study (Talmud et al. 2013) demonstrating that the phenotype of familial hypercholesterolaemia can be caused by a collection of common genetic variants of mild LDL-C-raising effect. Using the top decile of LDL-C gene score for the healthy cohort as a cutoff, the cumulative effect of risk alleles was shown to be the likely cause of hypercholesterolaemia in a significant proportion (25%) of the examined patients. A high gene score was also observed in two patients with considerably mild *LDLR* mutations, which demonstrates that common polymorphisms can contribute to the presentation of an individual carrying a mild FH mutation with LDL-C levels above the diagnostic threshold. However, it is possible that the LDL-C gene score cutoff of 1.16 for polygenic hypercholesterolaemia is too stringent. Thus using the 9^{th} decile cutoff of 1.08, in which a 41% of WHII individuals had LDL-C above the 4.9 mmol/l (mean LDL-C = 4.68 ±1.05mmol/l) FH diagnostic level, could be more appropriate. By doing so, the phenotype would be explained in an additional nine mutation-negative patients.

VI. CHAPTER SIX - Use of targeted exome sequencing as a diagnostic tool for Familial Hypercholesterolaemia.

6.1. Introduction

The clinical phenotype of FH is known to be due to mutations in three genes encoding proteins involved in the clearance of LDL-C from the plasma, *LDLR*, *APOB* and *PCSK9*. There are over 1,200 different *LDLR* mutations (Usifo et al. 2012) but only one common *APOB* (c.10580G>A, p.R3527Q) and one *PCSK9* (c.1120G>T, p.D374Y) mutation, reported in the UK population (Humphries et al. 2006a). While *LDLR*/*APOB*/*PCSK9* mutations cause a dominant pattern of inheritance, an autosomal recessive hypercholesterolaemia (ARH) has also been observed. The locus for ARH was mapped to a chromosome one gene, the *LDLRAP1*, in which both homozygous and compound heterozygous mutations can be found (Garcia et al. 2001).

In most laboratories FH mutation screening includes use of commercially available kits, designed to test for the most common mutations, such as Elucigene™ FH20 (Gen-Probe Life Sciences, UK) and LIPOchip® (Progenica Biopharma, Spain), and SALSA MLPA probemix (MRC-Holland) for large gene rearrangements (deletions or duplications), which account for 4-5% of all FH mutations (Taylor et al. 2010a). However, in the UK, due to the highly heterogeneous nature of the population this approach is not fully effective, and many patient samples require screening the *LDLR* promoter and coding regions, splice sites, and splice branch points for causative mutations, and in the diagnostic laboratory this is currently performed using Sanger sequencing. Because of the time and labour of these methods, there has been interest in next generation sequencing (NGS) technology for the diagnosis of genetic

disorders. However, whether NGS is ready for clinical use has been questioned (Desai and Jere 2012). Main limitations of the technology include the requirement for complex data analysis, significant computing infrastructure with respect to data analysis and storage and legal and ethical issues associated with incidental findings from acquiring whole exome data. In the research laboratory a four-phased approach is used to screen FH patients to identify the causative mutation, using the commercially available ARMS (Amplification Refractory Mutation System) kit, which tests for 20 the most common UK mutations, followed by High Resolution Melting (HRM) to detect changes within the coding region and splice sites of FH genes, followed by Multiplex Ligation-dependent Probe Amplification (MLPA), for the detection of large *LDLR* gene rearrangements, and finally Sanger sequencing (Graham et al. 2005, Whittall et al. 2010a, Humphries et al. 2006c). Using standard molecular diagnostic techniques, an FH-causing mutation can be detected in 20-30% of PFH patients and in 60-80% of DFH patients (Taylor et al. 2010b). This chapter presents results of is a pilot study, based on the analysis of the first UK10K project data release, which included whole exome sequencing results for 48 unrelated FH patients. The aim of this project was to identify and demonstrate advantages and limitations of the whole exome sequencing approach as a tool for molecular diagnosis of FH. In addition, the sensitivity problems of the current FH mutation screening methods were discussed.

6.2. Results

6.2.1. Patients characteristics

48 unrelated familial hypercholesterolaemia patients were selected from the Simon Broome FH register (Neil et al. 2004). Patients were diagnosed using the UK Simon Broome criteria as DFH on the basis of the presence or history of tendon xanthomas. The entire promoter and coding region, including splice sites, of the *LDLR* gene were screened by the High Resolution Melting method, as previously described (Whittall et al. 2010a), on the Rotor-Gene (6000) realtime rotary analyser. Patients were screened for presence of the *APOB* mutation, p.(R3527Q), using a restriction enzyme digest (Mamotte and van Bockxmeer 1993), and the entire coding region of the *PCSK9* were examined by HRM (Humphries et al. 2006c). Fragments with a shifted melting curve were analysed further by direct sequencing. Screening for large rearrangements within the *LDLR* gene was done using the MLPA (Taylor et al. 2009) SALSA P062 *LDLR* kit from MRC-Holland (Amsterdam). 195 non-FH Caucasian samples, sequenced in parallel with the FH cohort, as a part of the UK10K rare disease arm project (http://www.uk10k.org/studies/rarediseases.html), were used as controls. None of these subjects had disorders known to affect plasma lipid levels.

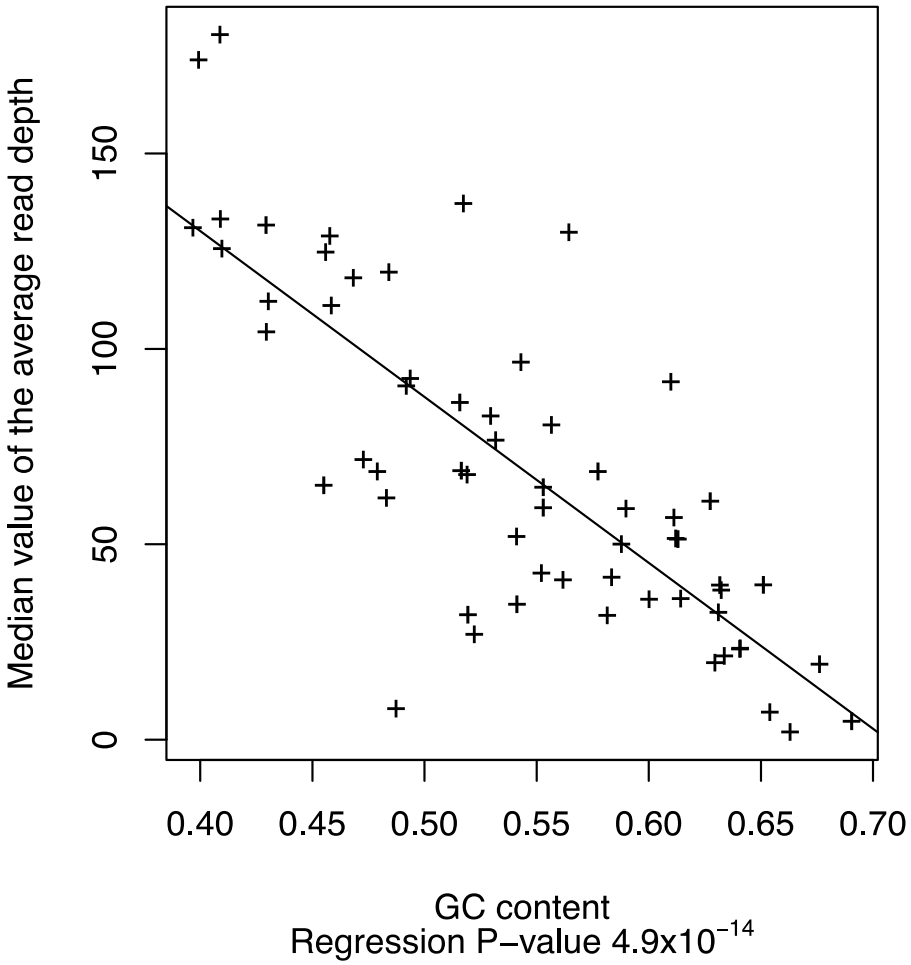
6.2.2. Gene coverage

Overall, the mean read depth for the whole exome sequence was 72x, with 78.9% of the exome covered at least 20x. 55.8% of the targeted sequence

was covered 50x or more. A highly significant negative correlation between the G/C content and the exome depth was observed in the FH genes ($p=4.9x10^{-14}$), as shown in Figure 37, which was reflected in poor coverage of promoter regions.

Figure 37

The negative correlation of the median read depth and the GC content for each targeted exon of the four FH genes (*LDLR*, *APOB*, *PCSK9* and *LDLRAP1*).



6.2.3. *LDLR* analysis

The average read depth of *LDLR* exons varied from 136x, for exon 12, to 4x, for exon 18, as shown in Figure 38. Using a 16x coverage threshold, which would give a 99% probability of observing a rare allele at least 3 times, all except exon 1 and 18 showed adequate coverage. Exons 3 and 4 contain the largest number of reported FH-causing mutations (Usifo et al. 2012) and both of these exons are well covered (mean depth 92 and 57, respectively).

As shown in Table 31, in 14 out of the 48 samples, a FH-unique variant in the *LDLR* was called, in 11 of these the variant has been previously reported to be FH causing (Usifo 2012).

Figure 38

The exonic coverage of FH genes: *APOB*, *LDLR*, *LDLRAP1* and *PCSK9* in a standard boxplot (the minimum, lower quartile, median, upper quartile and maximum for each gene exon). The horizontal dashed line indicates the 16x coverage, when the probability of observing a rare allele at least 3 times is 99%. The additional Y-axis describes GC content (0.4= 40%, 0.5= 50%, etc.) for a given exon, shown as ◇.

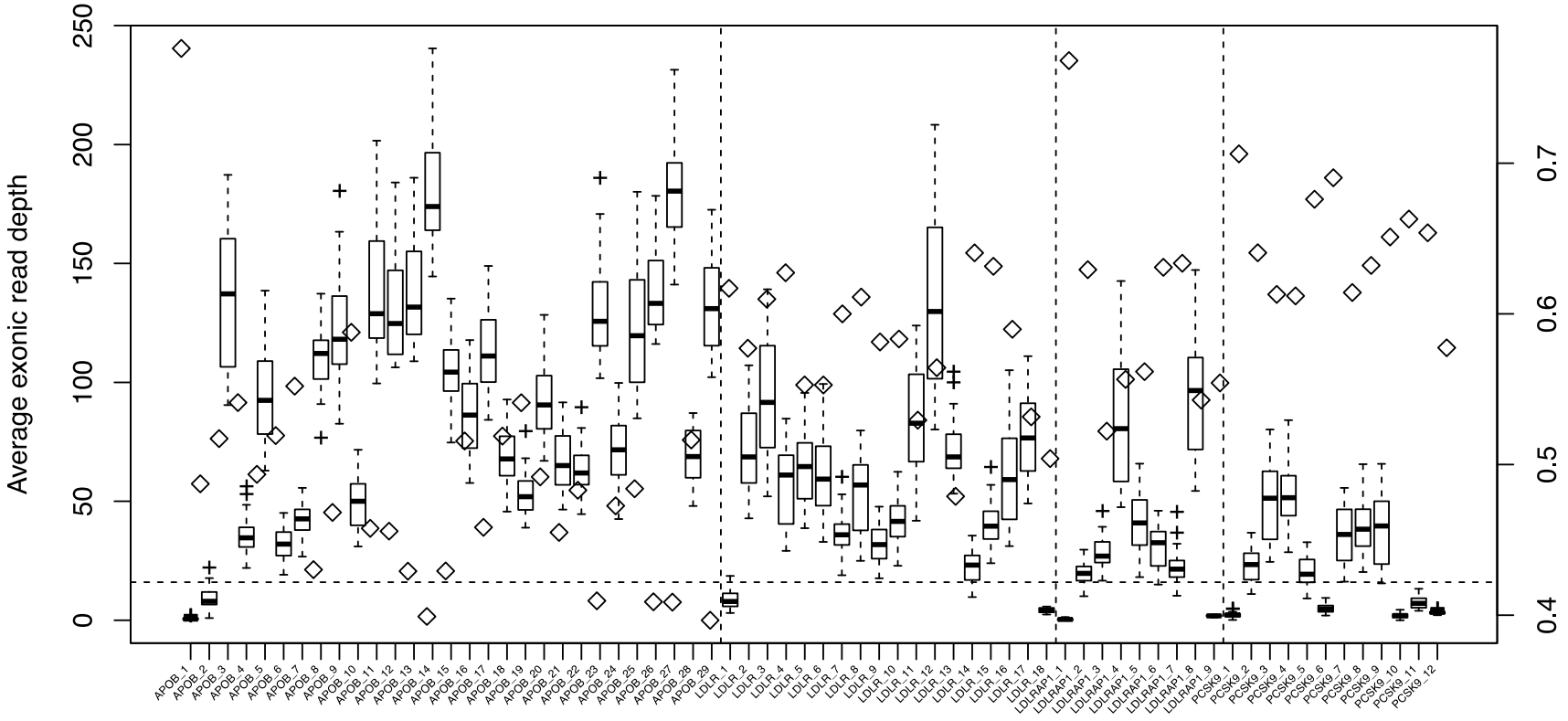


Table 31

Summary of pathogenic single nucleotide changes and small deletions/insertions in the FH genes. 'Depth' refers to the coverage depth; 'Quality' values are Phred-like quality scores generated by SAMtools.

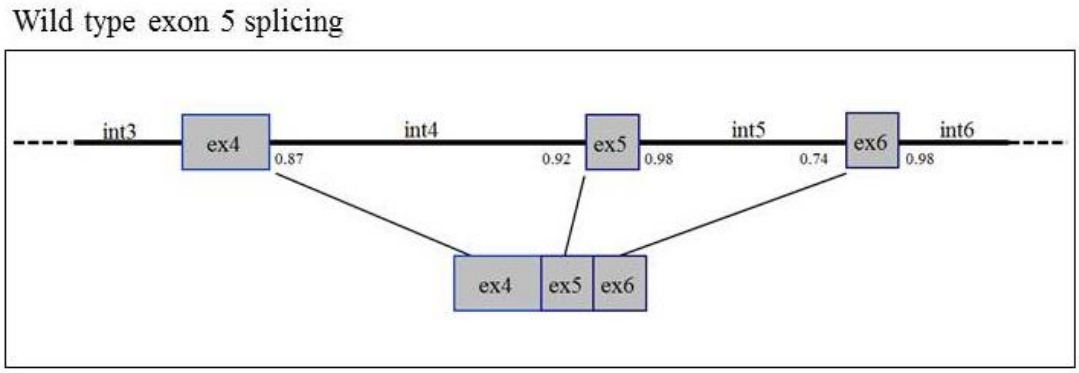
Gene	No. samples	Nucleotide change	Functional effect	Depth	Quality	Comments
LDLR						
	1	c.326G>A	p.(C109Y)	43	506	known FH mutation
	1	c.1690A>C	p.(N564H)	36	343	known FH mutation
	1	c.1823C>T	p.(P608L)	82	1214	known FH mutation
	1	c.2054C>T	p.(P685L)	20	135	known FH mutation
	1	c.2479G>A	p.(V827I)	65	749	known FH mutation
	2	c.682G>T	p.(E228X)	13	155	known FH mutation
	1	c.1048C>T	p.(R350X)	60	816	known FH mutation
	1	c.1150C>T	p.(Q384X)	20	275	known FH mutation
	1	c.1685G>A	p.(W562X)	41	701	known FH mutation
	1	c.2140+1G>A	Splicing	22	258	known FH mutation
	1	c.695-6_698del	Splicing	36	1543	novel
	2	c.1776_1778del	p.(G592del)	148	2634	novel
APOB						
	1	c.10277G>A	p.(A3426V)	192	2785	novel
	2	c.10580C>T	p.(R3527Q)	161	2144	known FH mutation
PCSK9						
	1	c.1027G>C	p.(D343H)	51	44	false positive
	1	c.1028A>C	p.(D343A)	50	201	false positive
LDLRAP1						
	1	c.432_433insA	p.(A145KfsX26)	90	2186	heterozygous

All of the *LDLR* variants were confirmed to be correct by Sanger sequencing of duplicate DNA samples (not shown). The variants included five different missense mutations and five nonsense mutations. The two novel *LDLR* variants included: c.695-6_698del and c.1776_1778del (p.(G592del)). The c.695-6_698del is predicted to cause a frameshift and a premature stop codon by altering *LDLR* splicing, as demonstrated in Figure 39. The deletion of Glycine at residue 592 is predicted to disrupt packaging of the LDL-R propeller blades in the EGF domain, which could affect displacement of the ligand from the ligand-binding region, shown in Figure 40. Of these 14 mutations, all should have been detected by the initial screening protocol, except for c.695-6_698del, where the change was located in the primer sequence used for PCR.

Figure 39

In silico, splice site predictions using BDGP prediction tools (http://www.fruitfly.org/seq_tools/splice.html) (Reese et al. 1997). A: Wild type *LDLR* scores for the acceptor/donor splice sites. B: *LDLR* mutation (c.695-6_698del) removes the exon 5 acceptor site and potentially activates an upstream cryptic site (c.695-457/8).

A.



B.

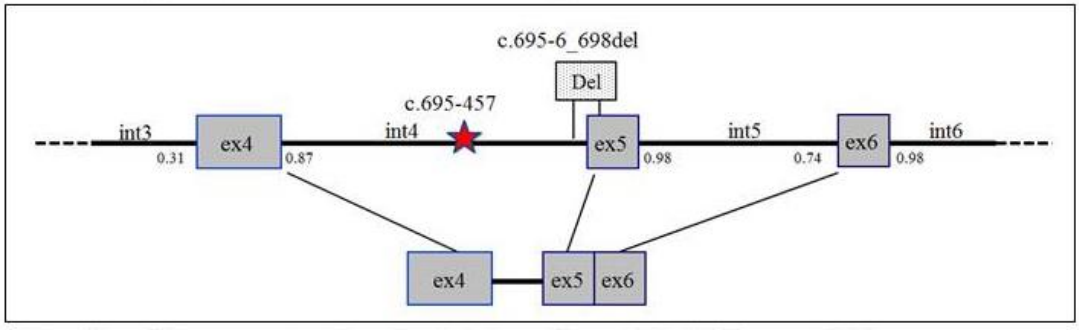
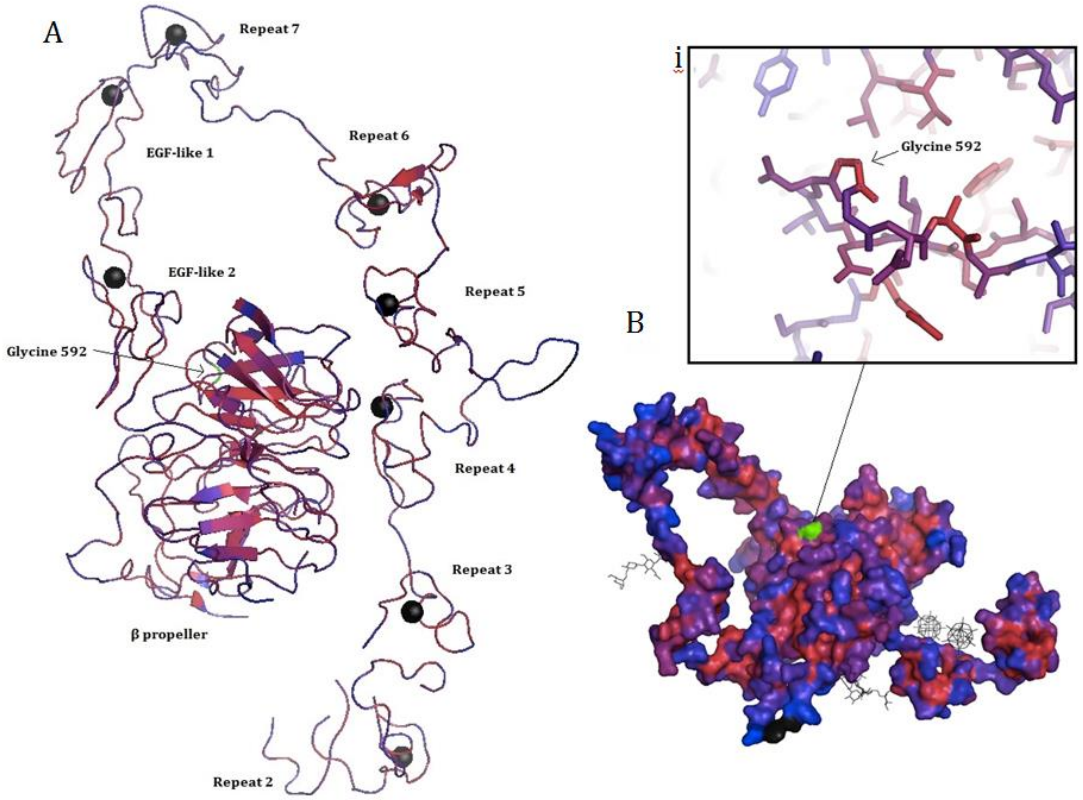


Figure 40

Conservation score calculated over the entire available crystal structure of the LDL-R protein. Red shows high conservation, purple- moderate conservation, blue- poor conservation and black- no conservation. A. LDL-R protein displaying the calcium molecules (black spheres) and labeled regions including Glycine 592. B. Glycine 592 is on the surface of the protein and is highly conserved (see inset ‘i’).

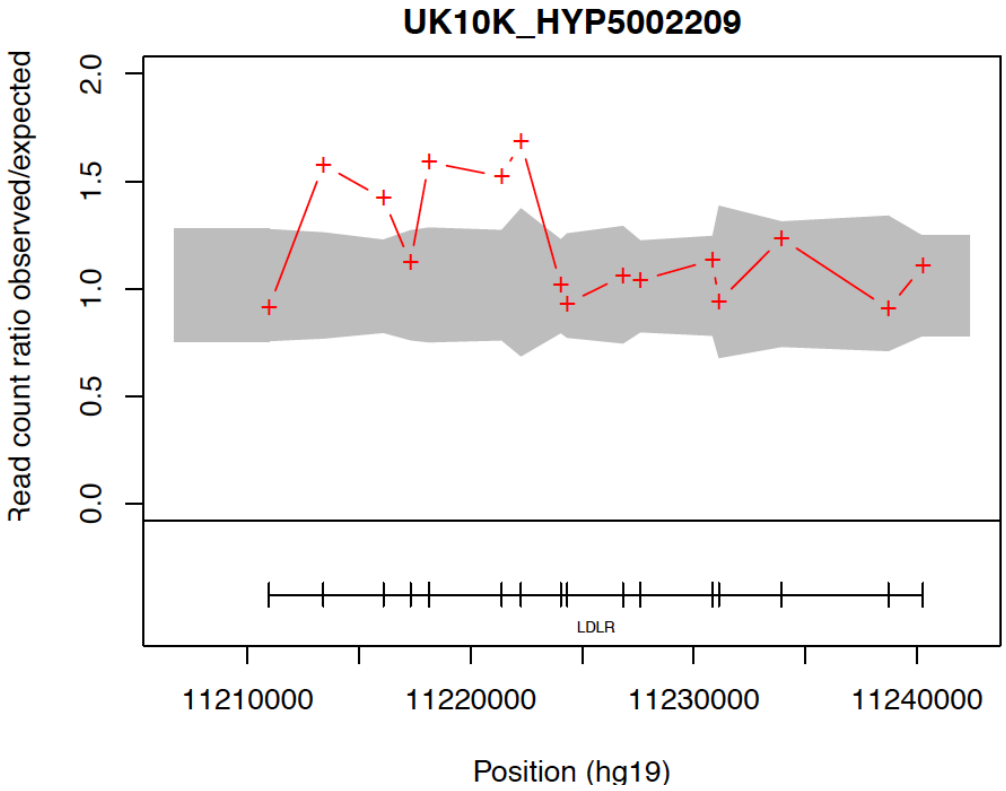


Copy Number Variations (CNVs) calling was performed using the ExomeDepth package, which identified one deletion of exons 11 to 12 (c. 1587-?_1845+?del), and two duplications, of exons 3 to 8 (191-?_1186+?dup), and exons 13 to 15 (c.1846-?_2311+?dup), as shown in Figure 41. These variants affect a substantial part of the gene and are predicted to cause a frameshift. All CNVs were confirmed by MLPA experiment, of which results are presented in Figure 42.

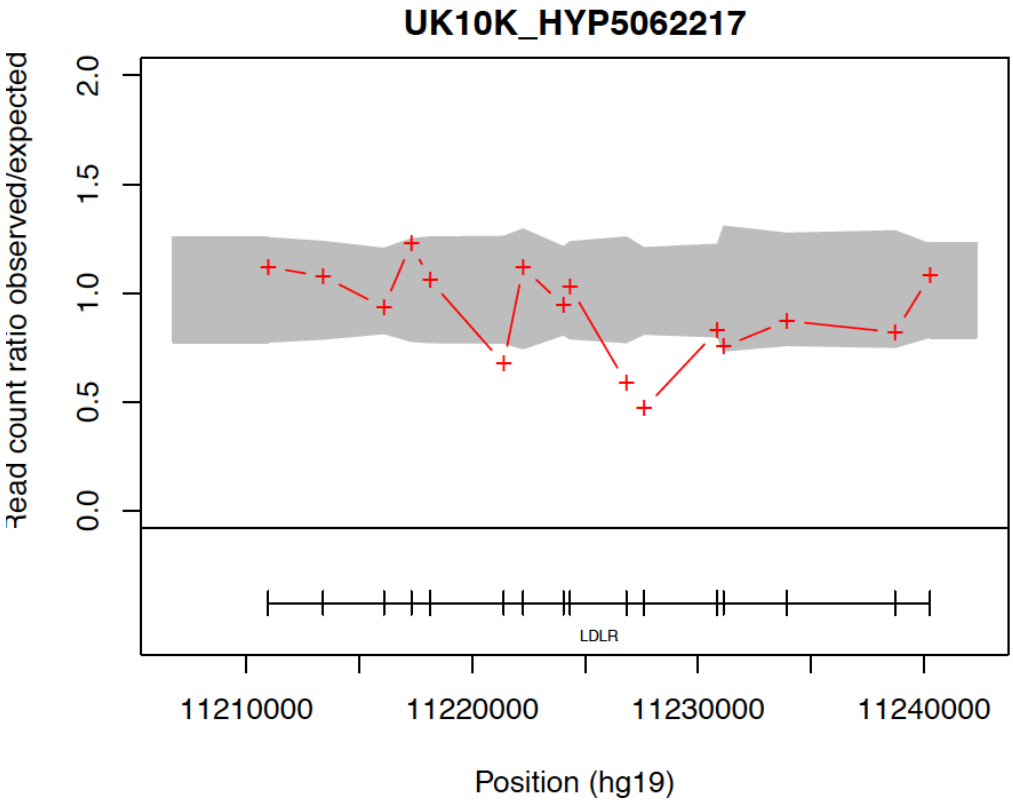
Figure 41

Copy Number Variants (CNVs) in the *LDLR* gene. A: Heterozygous duplication of exons 3 to 8. B: Heterozygous deletion of exons 11 and 12. C: Heterozygous duplication of exons 13 to 15. All identified by ExomeDepth in the exome sequencing data. The crosses show the ratio of observed/expected number of reads for the test sample. The grey shaded region shows the estimated 99% confidence interval for this observed ratio in the absence of CNV call. The presence of contiguous exons with read count ratio located outside of the confidence interval is indicative of a heterozygous deletion or duplication in a sample. Exons 1 and 18 were excluded from the analysis (not shown on the graph) as they did not reach the threshold of 100 for the total number of reads.

A.



B.


UK10K_HYP5062217
Read count ratio observed/expected
0.0
0.5
1.0
1.5
2.0
LDLR
11210000
11220000
11230000
11240000
Position (hg19)

C.

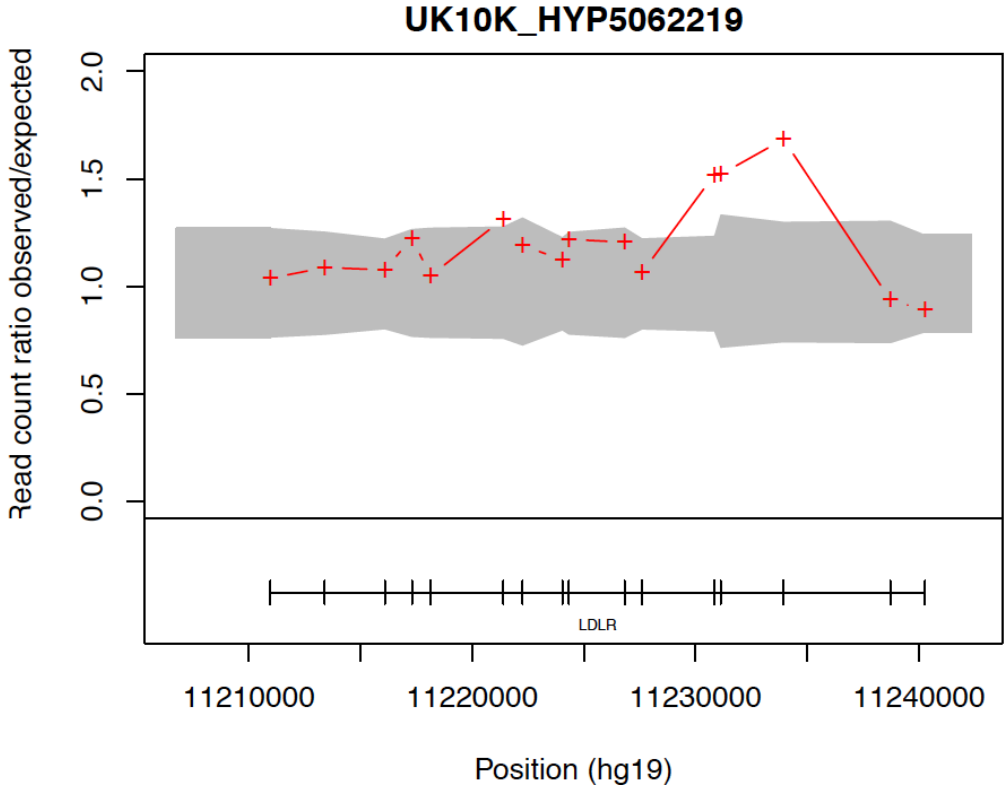
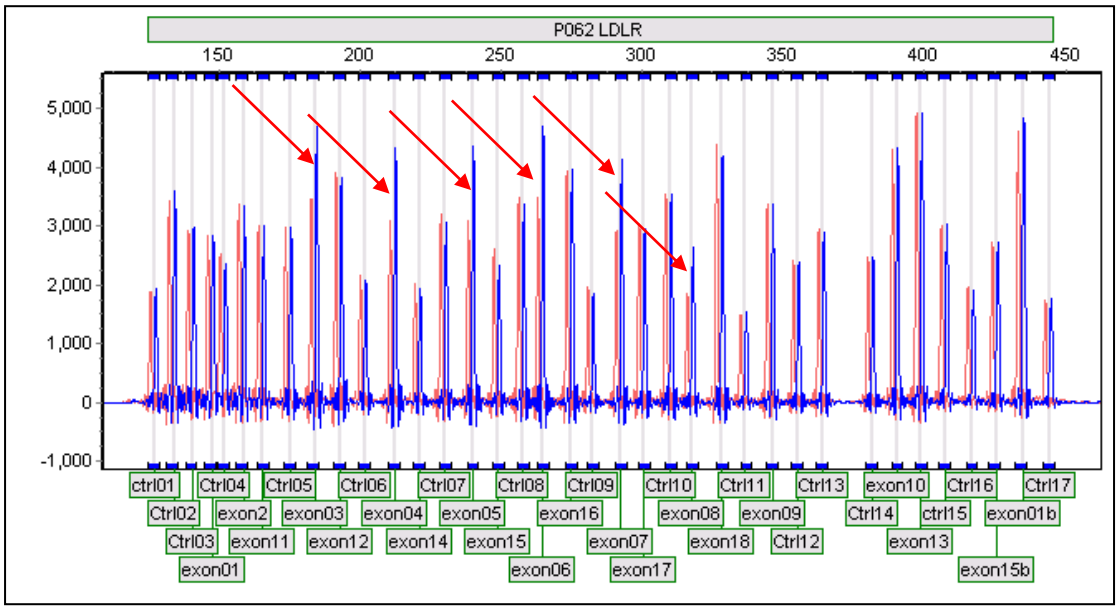

UK10K_HYP5062219
Read count ratio observed/expected
0.0
0.5
1.0
1.5
2.0
LDLR
11210000
11220000
11230000
11240000
Position (hg19)

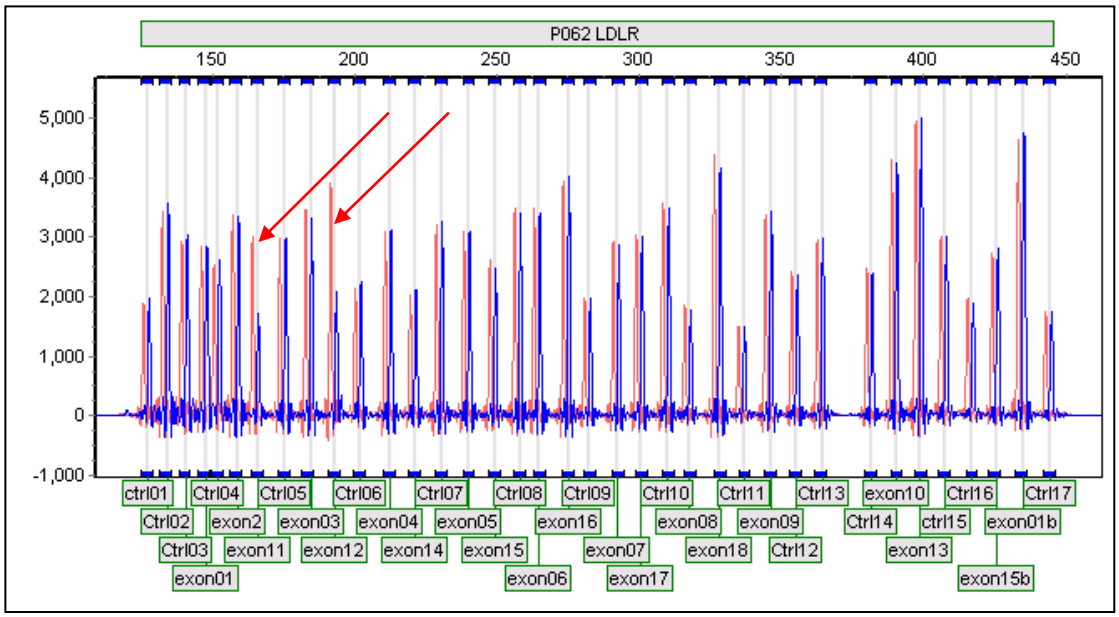
Figure 42

MLPA results quantified by the fluorescence peak heights for the tested sample (blue) and normalised control (red). Red arrows mark peaks, in which the sample/control difference was significant. A: Heterozygous duplication of exons 3 to 8 in sample HYP5002209. B: Heterozygous deletion of exons 11 and 12 in sample HYP5062217. C: Heterozygous duplication of exons 13 to 15 in sample HYP5062219.

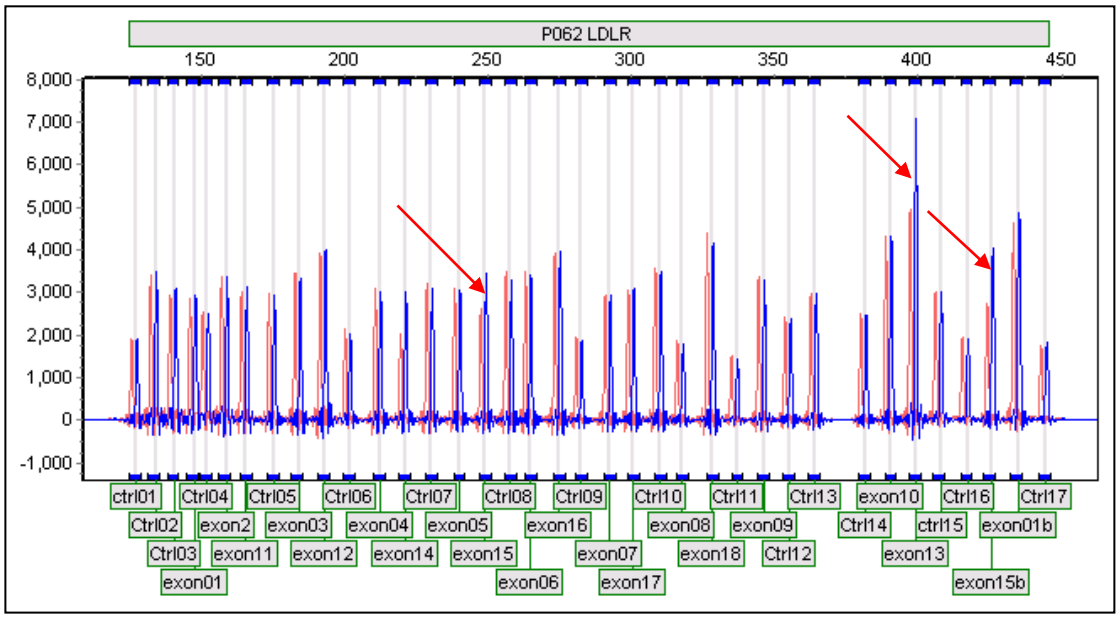
A.



B.


P062 LDLR
150
200
250
300
350
400
450
5,000
4,000
3,000
2,000
1,000
0
-1,000
ctrl01
Ctrl04
Ctrl05
Ctrl06
Ctrl07
Ctrl08
Ctrl09
Ctrl10
Ctrl11
Ctrl13
exon10
Ctrl16
Ctrl17
Ctrl02
exon2
exon03
exon04
exon05
exon16
exon08
exon09
Ctrl14
ctrl15
exon01b
Ctrl03
exon11
exon12
exon14
exon15
exon07
exon18
Ctrl12
exon13
exon01
exon06
exon17
exon15b

C.


P062 LDLR
150
200
250
300
350
400
450
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
-1,000
ctrl01
Ctrl04
Ctrl05
Ctrl06
Ctrl07
Ctrl08
Ctrl09
Ctrl10
Ctrl11
Ctrl13
exon10
Ctrl16
Ctrl17
Ctrl02
exon2
exon03
exon04
exon05
exon16
exon08
exon09
Ctrl14
ctrl15
exon01b
Ctrl03
exon11
exon12
exon14
exon15
exon07
exon18
Ctrl12
exon13
exon01
exon06
exon17
exon15b

6.2.4. *APOB* analysis

The mean read depth of *APOB* exons was 93. Exons 26 and 29 were covered on average 135x, whereas exon 1 was covered only once (Figure 38). Two individuals were found to carry the FH-causing *APOB* mutation in exon 26, the p.(R3527Q), which were confirmed by Sanger sequencing. There were no nonsense or frameshift mutations observed in the gene sequence. One novel non-synonymous variant was found in exon 26 of *APOB*, the p.(A3426V), which was unique to the FH cohort. The variant was predicted as 'Tolerated', 'Benign' and 'Polymorphism' by SIFT, PolyPhen and Mutation Taster, respectively. Four other FH-unique non-synonymous variants were observed outside of the ligand-binding domain, as shown in . The functional impact of these variants as predicted by PolyPhen/SIFT/Mutation Taster was not consistent and whether or not these are FH causing is unclear. There were no CNVs found within the *APOB* gene.

Table 32

Summary of novel *APOB* variants, located outside of exons 26 and 29. 'Depth' refers to the depth coverage; 'Quality' values are Phred-like quality scores generated by SAMtools. D- damaging; B-benign; T-tolerated; N-polymorphism.

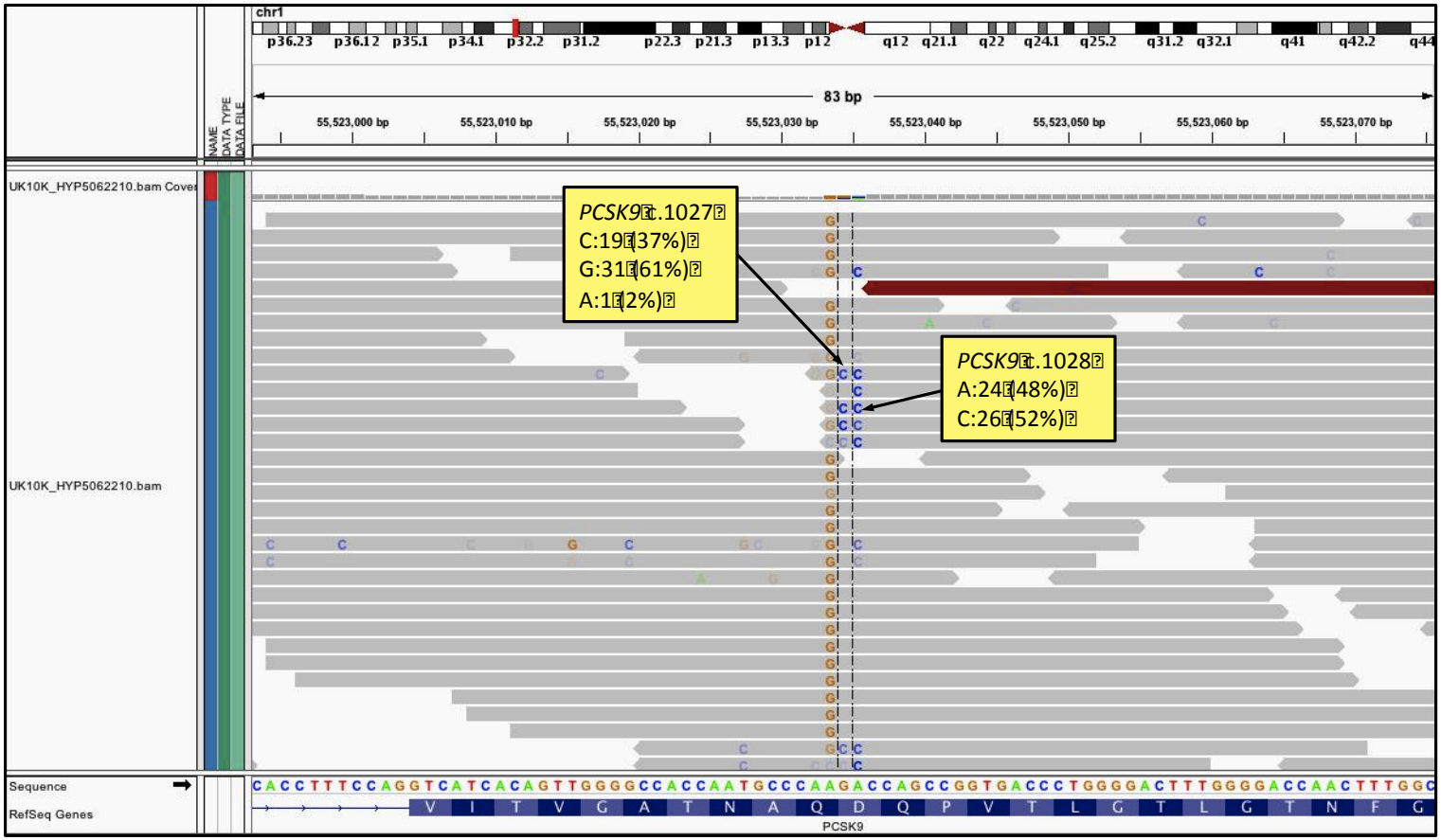
Nucleotide change	Functional effect	Depth	Quality	PolyPhen	SIFT	Mutation Taster
c.148C>T	p.(R50W)	139	181	D	D	N
c.1199G>A	p(.R400H)	35	100	B	D	N
c.2938G>A	p.(A980T)	29	198	B	T	N
c.3931A>C	p.(K1311Q)	68	170	B	D	N

6.2.5. *PCSK9* analysis

The mean read depth of the *PCSK9* exons was 23. Of these, only 58% of the gene coding sequence had the mean coverage higher than 16, whereas exons 1, 6, 10, 11, and 12 were covered 4x on average (Figure 38). Exon 7, where the common UK FH-causing mutation (c.1120G>T, p.(D374Y)) occurs, was covered 36x. Two novel non-synonymous variants were called by the exome sequencing, c.1027G>C and c.1028A>C, both present in the same sample. However, despite the high read depth (51x and 50x), and the high number of read count for the novel alleles (19 and 26), presented in Figure 43, the Sanger sequencing did not confirm the variants. There were no CNVs observed in *PCSK9*

Figure 43

Intergrative Genomic Viewer image of the coverage of *PCSK9* exon 7 region containing two false positive variants c.1027G>C and c.1028A>C. These artifacts were probably created during the amplification step in the sequence capture process.



6.2.6. *LDLRAP1* analysis

The average read depth of *LDLRAP1* was 36, with all except exons 1 and 9, covered above the 16x threshold (Figure 38). The *LDLRAP1* variant analysis was performed using a homozygosity-based strategy, and the presence of compound heterozygote variants was also assessed. There were no homozygous or compound heterozygous functional changes within the gene in any of the individuals. One patient was found to be heterozygous for a known Sicilian/Sardinian ARH mutation, the c.432_433insA, p.(A145KfsX26), which is a frame shift mutation resulting in a truncated peptide formation (Garcia et al. 2001, Barbagallo et al. 2003). Further analysis of this sample showed no other pathogenic variants in known FH genes, which could contribute to the phenotype. CNV calling did not detect large rearrangements in the *LDLRAP1*.

6.3. Discussion

The exome sequencing results exposed sensitivity problems with the current FH mutation screening methods used in our research laboratory. Overall, the standard variant detection process already in place (HRM, MLPA and Sanger sequencing) did not detect 17 *LDLR* mutations (including 3 CNVs) and 2 *APOB* mutations. Although the HRM has proved to be efficient at detecting FH variants (Whittall et al. 2010a), its sensitivity decreases in some gene regions, depending on the nucleotide composition of the fragment. Re-examining previous results for the samples with a *LDLR* or *APOB* variant

called by the NGS, we observed that most of the variants showed a melting curve shift during the HRM assay, but Sanger sequencing of the identified gene region did not detect any heterozygous changes in the sequence despite being repeated several times (i.e. only the predicted wild-type sequence was obtained). After the exome sequencing, the Sanger sequencing was repeated on a duplicate DNA sample and the predicted mutations were confirmed to be present, validating the exome sequencing and variant calling. Although Sanger sequencing is considered to be the gold standard mutation-detection method, a combination of PCR artefacts and the human error aspect in the protocol appears to be the main reason for the false negative calling in the original screening.

Two novel variants in *LDLR* were identified, a deletion of 10 bp on boundary of intron 4 and exon 5, which is predicted to cause a frameshift resulting in a premature stop codon by altering *LDLR* splicing, and a three bp deletion which deletes Glycine at residue 592, which is predicted to disrupt packaging of the LDL-R propeller blades in the EGF domain. Both of these variants are not found in dbSNP, the 1000 Genomes or the NHLBI Exome Sequencing Project, and are highly likely to be FH-causing, although further work is required to confirm this.

APOB codes for one of the largest human proteins, which is the major component of the LDL-C responsible for binding to the LDL-receptor (Innerarity et al. 1987). The actual binding site for the receptor, the B-site (residues 3386–3396), has been mapped to a region encoded by exon 26 of

the *APOB*, which is the longest coding exon known (7572 bps) (Boren et al. 1998). In addition, the C-terminus encoded by exon 29 of the gene was proposed to function as a modulator of the receptor binding (Boren et al. 1998). Therefore our variant analysis strategy prioritised novel variants located in exons 26 and 29 of the gene, as these are more likely to cause the FH phenotype. In this study, there was only one novel variant identified in the exon 26 of *APOB*, the c.10277G>A (p.(A3426V)), which was not observed previously by the dbSNP, the 1000 Genomes or the NHLBI Exome Sequencing Project. The variant was not present in the 195 non-FH exomes from the UK10K project, which were processed using the same pipeline, increasing the likelihood that it is in fact disease causing. The novel p.(A3426V) variant is located near to the LDL-receptor-binding site (B-site), and close to the known FH mutation p.(R3527Q), and although it does not alter the charge at the site, it may produce a conformational change affecting the LDL-R/ApoB interaction. This requires further experiments since the current *in silico* prediction tools are not able to assess protein-protein interactions. We will also examine whether or not the variant co-segregates with the disease. Four other novel *APOB* variants were identified in this group of patients in the N-terminal part of the protein. Although these variants are less likely to influence LDL clearance from the blood, since the N-terminal region of the protein is not involved in interacting with the LDL-R, some of the variants are predicted as damaging by Polyphen or SIFT. Future work includes the assessment of novel identified variants, which will involve family co-segregation and functional assays.

There were no FH-causing mutations identified in the control cohort at this stage of the study. Further data releases, analysed in Chapter Seven, which included whole exome sequencing data for almost 2,000 controls revealed several FH-causing variants, which included two *APOB* p.R3527Q mutations and six previously published *LDLR* variants, which would be considered as pathogenic. This finding demonstrates that WES has a potential for the identification and reporting of incidental findings, as recently recommended by the American College of Medical Genetics and Genomics (Green et al. 2013). Most of the sequencing data generated for Mendelian disorders are focused on the exome, which constitutes around 1% of the whole human genome. Prediction tools for the analysis of non-synonymous changes are well established and widely used to estimate the deleteriousness of amino acid changes. However, since the majority of human variations are located in the non-coding regions ('A map of human genome variation from population-scale sequencing' 2010), concern about the bias towards variants in the protein-coding sequence was highlighted (Cooper and Shendure 2011). Proving the pathogenic effect of promoter variants requires use of functional assays. To date, there are 13 *LDLR* promoter variants predicted to be causal (in revision (Khamis A 2012)). Disappointingly, but not surprisingly given they were not targeted, the exome sequencing data generated by the SureSelect Human All Exon (Agilent) assay, had negligible coverage of the gene promoter regions, which can lead to false negative conclusions. Further updates of the human exome capture assay should include coverage of the *LDLR* promoter sequences, which can cause autosomal dominant disease by altering gene regulation.

The SureSelect Human All Exons capture assay is a standard product, which proved to be efficient at detecting mutations within the *LDLR* and the *APOB* genes. In this sample, 78.9% of exome bases were covered at least 20 times, which is in line with the product description ~80%. For both, *LDLR* and *APOB*, the majority of the coding sequence was covered more than the 16x threshold to achieve an estimated 99% chance of seeing a real variant (present in a heterozygous state) of at least 3 times, and overall 19 mutations, were found by high throughput DNA sequencing, which had been missed by conventional methods in our research laboratory. This indicates increased sensitivity for NGS, which can be due both to the method used and to the reduced human intervention and the highly automated protocol. However, as with many PCR-based methods, exome sequencing has some limitations when it comes to amplification of highly repetitive regions or sequences rich in GC content. A highly significant negative correlation between the G/C content and the exome depth was observed in the FH genes ($p=4.9x10^{-14}$), as shown in Figure 37. Specifically, only 58% of the *PCSK9* gene was covered more than 16x, producing unreliable results for variant calling in a significant proportion of the gene's coding region. If a read depth threshold of 30x was considered to be required for complete certainty of variant calling, at which the sensitivity to detect heterozygous variants was shown to be 100% (Choi et al. 2009), exons that would be insufficiently covered would also include exons 1, 14 and 18 of *LDLR*, exons 2 and 5 of *PCSK9*, and exons 2, 3 and 7 of *LDLRAP1*. Thus, although the quality of the produced data is good, validation of called variants in poorly covered regions is still necessary. There were two novel non-synonymous *PCSK9* variants called by the exome sequencing, which were

not confirmed by the capillary sequencing. Despite their sufficient coverage (51x and 50x), the variants were wrongly called due to inadequate variant calling settings. In later data releases, the variant calling was optimized and both variants did not pass filtering criteria. Applying more stringent filters to the raw data increases the specificity of the calling. However it may also lead to false negative results, since not all of the exome's regions are equally covered. Newer versions of the SureSelect assay show markedly improved coverage of exons that were previously poorly covered (unpublished data), so we can expect the sensitivity of exome sequencing to improve.

The Agilent SureSelect assay was efficient in capturing the exon-intron junctions, covering on average 80-100bps of the intronic regions. This was an advantage over our current screening protocol, and enabled us to detect a novel variant, the c.695-6_698del, which is partially positioned on the annealing site for the sequencing primer routinely used in our lab.

The methodology behind the ExomeDepth package (Plagnol et al. 2012) proved to be robust and enabling the use of the exome data, which are composed of short sequence reads for exonic regions, for the detection of large gene rearrangements, which are known in the *LDLR* to be usually due to intronic Alu sequence misparing (Lehrman et al. 1987a, Lehrman et al. 1987b). The method was shown to allow identification of heterozygous CNVs within the *LDLR* gene, which were missed by the currently used MLPA. However, in order to maximise the sensitivity and to minimise the noise created by technical variability between samples, CNV analysis by ExomeDepth requires quality data of well-matched exomes (>6 samples), i.e. sequenced under the exact same conditions.

The greater time efficiency of the exome sequencing is a significant advantage over the current screening methods. Although each called variant currently needs to be individually confirmed by Sanger sequencing before a mutation report can be prepared, analysing a number of patients in parallel, in a short period of time, is likely to be an efficient way forward for screening of heterogeneous FH patients. More importantly, limited use of manual checks and human intervention reduce the issues of possible human error. The cost efficiency of NGS is also increasing. The development of novel approaches of gene-targeted sequencing, using Illumina MiSeq platform, reduces not only the costs of sequencing itself but also the time spent on data analysis and computer storage requirements. The possibility of designing custom amplicons for each disease, recently offered by Illumina TruSeq Custom Amplicon or Agilent HaloPlex products, will also improve the capture of promoters and other regulatory regions, which could be omitted in WES.

VII. CHAPTER SEVEN - Whole exome sequencing of no *LDLR*/*APOB*/*PCSK9* mutation definite Familial Hypercholesterolaemia patients.

7.1. Introduction

The importance of identifying an FH-causing variant, which has clinical utility in providing an unequivocal diagnosis (Humphries et al. 2008), has been emphasized by the National Institute of Clinical Excellence (NICE), which in 2008 recommended cascade testing using DNA information for finding the affected relatives of a patient (Wierzbicki et al. 2008). The risk of early CHD can be significantly reduced by statin treatment (Neil et al. 2008), and genetic information has been demonstrated to complement the management of treated patients (Humphries et al. 2006c).

However, an FH-causing mutation can be detected in 63-87% of DFH patients, thus in a proportion of patients the cause of high LDL-C remains unexplained. The majority of FH patients inherit a mutation in *LDLR* (Usifo et al. 2012). The *APOB* variant (c.10580G>A, p.(Arg3527Gln)) accounts for 5-12% of UK FH cases (Graham et al. 2005, Taylor et al. 2010b, Futema et al. 2013) whereas a gain-of-function mutation in *PCSK9*, (c.1120G>T, p.(Asp374Tyr)), can be found in roughly 1.7% of FH patients (Taylor et al. 2010b). In the past few years, several loci have been reported to co-segregate with FH in family linkage studies, however to date this has not led to the identification of a specific causal gene (Cenarro et al. 2011, Marques-Pinheiro et al. 2010, Wang et al. 2011). It is likely that there are novel FH mutations located in unknown genes influencing lipid metabolism and that their discovery may contribute to the identification of novel treatment targets. In order to find novel causes of FH the whole exomes of 125 unrelated DFH patients were sequenced at a high depth. The expected frequency of an FH-causing mutation in a novel gene was lower than the gain-

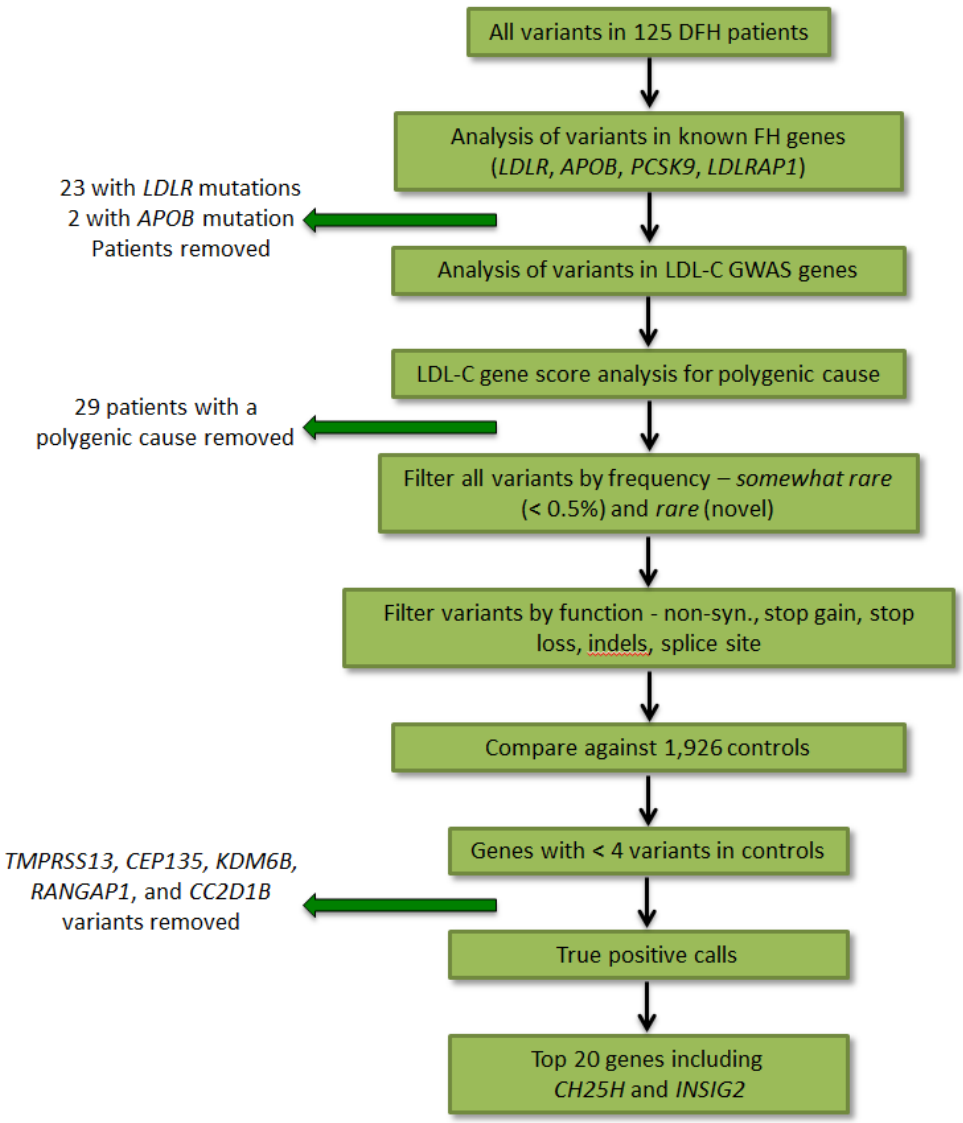
of-function mutation in *PCSK9* (1.7%), reasoning that a higher frequency would give researchers more power and led to the discovery earlier on. Patients were also assessed for polygenic hypercholesterolaemia, which is discussed in Chapter Three.

7.2. Results

The overall workflow of the whole exome sequencing data analysis is shown in Figure 44. All copy number variants called in the FH cohort can be found here: https://github.com/marta10/PhD_Thesis_MF/tree/master/cnvs_results .

Figure 44

Novel FH gene discovery pipeline. To increase the chance of detecting true FH-causing variants with a strong effect and reducing the noise, samples with a mutation in *LDLR* or *APOB* (apart from novel *APOB* variants of unknown effect) or those with a high LDL-C gene score (discussed in Chapter Five) were removed from the analysis. Remaining variants were filtered by their frequency and functional effect and compared against controls. Genes with more than 4 rare functional variants in controls were filtered out. Remaining variants were manually assessed and false positive calls were removed.



7.2.1. Variants in known FH genes

Firstly, variants in known FH genes were analysed (Figure 44). The coverage of *LDLR*, *APOB*, *PCSK9* and *LDLRAP1* were discussed in Chapter Six. For the *LDLR* ten individuals were carrying a missense mutation, five a nonsense mutation, three had small deletions, and two individuals had intronic changes known to affect splicing Table 33. Two of the variants were novel, c.695-6_698del and c.1776_1778del, and their effect was discussed in Chapter Four (Figure 39 and Figure 40). Analysis with ExomeDepth package for CNVs identified two large duplications and one deletion within the *LDLR* region, discussed in Chapter Four (Figure 41). For *APOB* two individuals carried the known FH mutation, c.10580G>A (p.R3527Q), and several *rare* and cases-unique *APOB* variants, distributed across different gene exons, were identified (Table 34). Because *APOB* is highly polymorphic the overall number of rare variants was not significantly different in comparison to controls. *PCSK9* had the lowest mean read depth (18x), with four exons (1, 5, 9, and 10) covered less than 10x. There were no FH-causing variants identified in this gene. Samples with an explained cause (FH mutation positive) were removed from further analysis.

Table 33

Summary of the identified *LDLR* mutations and their *in silico* predicted effect, including calculated LDL-C gene scores for the mutations carriers (presented in bold are the gene scores that are above the top decile cutoff for the control population).

Mutation type/Exon	Mutation	Gene Score	PolyPhen	SIFT	Mutation Taster
Missense					
4	c.326G>A (p.C109Y)	1.03	Probably damaging	Not tolerated	Disease Causing
4	2X c.502G>C (p.D168H)	0.91, N/A	Probably damaging	Not tolerated	Disease Causing
4	c.681C>G (p.D227E)	N/A	Probably damaging	Not tolerated	Disease Causing
9	2X c.1196C>A (p.A399D)	1.03 and 1.07	Possibly damaging	Not tolerated	Disease Causing
11	c.1690A>C (p.N564H)[1]	**1.17**	Probably damaging	Tolerated	Disease Causing
12	c.1823C>T (p.P608L)	1.09	Probably damaging	Not tolerated	Disease Causing
14	c.2054C>T (p.P685L)	0.97	Probably damaging	Not tolerated	Disease Causing
17	c.2479G>A (p.V827I)	0.92	Probably damaging	Not tolerated	Disease Causing
Nonsense					
4	2X c.682G>T (p.E228*)	0.78 and 0.84	NA	NA	Disease Causing
7	c.1048C>T (p.R350*)	1.11	NA	NA	Disease Causing
8	c.1150C>T (p.Q384*)	0.65	NA	NA	Disease Causing
11	c.1685G>A (p.W562*)	0.95	NA	NA	Disease Causing
Indels					
5	c.695-6_698del (Figure 39)	**1.23**	NA	NA	Disease Causing
12	2X c.1776_1778del p.G592del (Figure 40)	N/A	NA	NA	Disease Causing
Intronic					

intron14	c.2140+1G>A	0.58	NA	NA	Disease Causing
intron9	c.1359-31_1359-23 delinsCGGCT	0.92	NA	NA	NA
Large rearrangements					
3_8	c.191-?_1186+?dup	1.03	10kb in frame duplication, peptide elongation		
11_12	c.1587-?_1845+?del	N/A	4kb out of frame deletion, truncated protein		
13_15	c.1846-?_2311+?dup	0.92	7kb out of frame duplication, truncated peptide		

N/A - not available

NA - not applicable

1 – carrier of this variant also has a deletion in exon 17 of *LDLR* c.2393_2401del9 (p.L799_V801del))

Table 34

All *rare APOB* variants identified in the FH cases (updated Table 32), including *in silico* predictions of their effect and LDL-C gene scores for the corresponding variant carriers. Using *in silico* mutation prediction tools (PolyPhen2, SIFT, Mutation Taster) the variant located in exon 3 of *APOB* (c.148C>T (p.R50W) has been predicted to be pathogenic by all three algorithms. The mutant Tryptophan is bigger than the wild type Arginine and it is predicted to cause a loss of hydrogen bonds in the core of the protein, which may result in an incorrect folding. Other variants include c.598G>A (p.A200T), c.1199G>A (p.R400H), and c.G2700G>T (p.Q900H) in both cases the mutant differs in size and hydrophobicity from the wild type residue, which may affect the folding of the protein as well as the hydrophobic interactions within the protein's core. The novel c.10277C>T (p.A3426V) variant is located near to the LDL-receptor-binding site (Boren et al. 1998), and although it has been predicted as benign/tolerated/polymorphism by the *in silico* tools, it may affect the LDL-R/ApoB interaction.

Exon	Variant	Gene Score	PolyPhen	SIFT	MutationTaster	ID
3	c.148C>T(p.R50W)	0.83	Probably Damaging	Not Tolerated	Disease Causing	HYP5062228
6	c.598G>A (p.A200T)	0.98	Possibly Damaging	Not Tolerated	Polymorphism	HYP5269576
10	c.1199G>A(p.R400H)	N/A	Benign	Not Tolerated	Polymorphism	HYP5159267
18	c.G2700G>T (p.Q900H)	1.19	Probably Damaging	Not Tolerated	Polymorphism	HYP5358899
26	c.10277C>T (p.A3426V) and c.6639_6641delTGA (p.2213_2214delD)	1.17	Benign NA	Tolerated NA	Polymorphism Disease Causing	HYP5002222
	2 X c.G10580G>A (p.R3527Q)	0.71 and 1.01	Probably Damaging	Not Tolerated	Disease Causing	HYP5062226 and HYP5062216

7.2.2. LDL-C gene score analysis

Out of 109 FH samples (21 mutation positive, 88 mutation negative, remaining samples had insufficient DNA), genotyped for all 12 SNPs, 31 had a gene score above the 1.16 cutoff (results are demonstrated in Chapter Five), within which two samples, in addition to the high gene score, had an LDLR mutation, one in exon 11 (c.1690A>C (p.N564H)) found on the same allele as a 9bp deletion in exon 17 (c.2393_2401del9 (p.L799_V801del)), which have been demonstrated as not fully-penetrant (26). The other was a deletion of a consensus splice site at the 5' of exon 5, c.695-6_698del, which has not been examined in vivo to confirm its likely effect on splicing.

The mean LDL-C gene score for the FH mutation negative group was 1.08, which was significantly higher than 0.90 for the WHII study ($p < 2.2x10-16$), and 0.96 for the FH mutation positive group ($p = 0.006$). The overall difference between the groups was significant (ANOVA, $p = 1.33x10-12$). Individuals with a gene score above the top decile cut-off for the WHII subjects (>1.16), were considered to have polygenic hypercholesterolaemia and excluded from further analysis

7.2.3. Variants in GLGC LDL-associated genes

Next, variants in genes identified through GWAS as being involved in determining levels of LDL-C in healthy individuals (Teslovich et al. 2010) were examined. A burden test on all *rare functional* variants in any gene singly or in all 25 genes combined showed no obvious candidate for a novel FH locus

(Table 35). In addition, there were no loss-of-function variants (i.e. premature stop codon formation, loss of a stop codon, frameshift indels, CNVs) observed in these genes.

Chromosomal regions previously associated with FH in family linkage studies were also analysed (Cenarro et al. 2011, Marques-Pinheiro et al. 2010, Wang et al. 2011). There was no association of *rare functional* variants in any gene located within the associated loci. The p vales for each of the loci are shown in Appendix Table III.

Table 35

Top *p* values of the *rare functional* variant association between cases and controls in the GLGC candidate genes.

Gene Name	Variant count in Cases (n=71)	Controls (n=1926)	*p* value
MTTP	3	12	0.04
NYNRIN	3	18	0.09
KIAA1875	2	11	0.14
CYC1	1	4	0.23
HFE	1	4	0.23
TOP1	1	4	0.23
PVRL2	1	6	0.31
SCAP	2	23	0.37
LIPG	1	10	0.44
ABCG8	1	17	0.62
LIPC	1	18	0.64
OSBPL7	1	19	0.65
CELSR2	1	32	0.83

7.2.4. Whole exome analysis

25 samples carrying a mutation in known FH genes and 29 with the LDL-C gene score above 1.16 were removed from further analysis.To interrogate the whole exome a burden test was performed between 71 cases and 1,926 controls. There were 4,326 genes with one or more *rare functional* variant in cases. In order to remove calls less likely to influence the FH phenotype and to increase the detection power, further analysis were only limited to genes where a maximum of 4 *rare functional* variants were seen in the controls, based on the expected prevalence of FH of 1 in 500 in the general population, and therefore any gene with > 4 *rare functional* variants in the controls were excluded. The next step involved a visual validation of the quality of calls performed using the Human Genome 19 on the Integrative Genomic Viewer (IGV) (Thorvaldsdottir et al. 2013). In order to avoid false negative calls, data produced by a less stringent variant calling pipeline (using a single-sample calling) were searched for rare variants in the significant genes. An additional loss-of-function variant, a premature stop codon at the position c.244C>T (p.Q81*) was found in the *CH25H* gene, in an FH patient sample with a low LDL-C SNP score After adjusting for the false positives and false negatives, *CH25H* remained the top gene ($p < 1.4x10^{-3}$) with three variants in the cases and two in the controls (Table 36).

Table 36

Summary of the genes and their variants, which shown a significant excess of *rare functional* variants in FH cases in comparison to the controls.

Gene		Ch	Number of variants in cases (n=71)	Number of variants in controls (n=1,926)		p value
CH25H		10	3	2		0.001
	Cases		ENST00000371852:exon1:c.G568A:p.V190I; exon1:c.A716C:p.H239P; exon1:c.C244T:p.Q82X			
	Controls		ENST00000371852:exon1:c.T742G:p.C248G; exon1:c.C590A:p.P197Q			
TSPAN4		11	3	3		0.002
	Cases		3X ENST00000397408:exon4:c.230_232del:p.77_78del			
	Controls		ENST00000346501:exon5:c.G389A:p.R130Q; 2X exon5:c.G439A:p.A147T			
HSPB7		1	2	0		0.003
	Cases		2X ENST00000311890:exon2:c.199+7G>A			
	Controls		None			
KLRC1		12	2	0		0.003
	Cases		ENST00000544822:exon5:c.G333C:p.Q111H; exon3:c.C178T:p.H60Y			
	Controls		None			
MOAP1		14	3	4		0.004
	Cases		ENST00000556883:exon2:c.C707T:p.A236V; exon2:c.G476C:p.C159S; exon2:c.A182G:p.N61S			
	Controls		ENST00000556883:exon2:c.C655G:p.R219G; exon2:c.C627A:p.S209R; exon2:c.C264G:p.I88M; exon2:c.A919G:p.I307V			
RBM25		14	3	4		0.004
	Cases		ENST00000261973:exon6:c.A454T:p.I152F; exon2:c.T50C:p.L17P; exon11:c.C1364A:p.A455D			

	Controls		ENST00000261973:exon7:c.C671T:p.A224V; exon11:c.A1273G:p.R425G; exon18:c.G2392A:p.V798I; exon2:c.T7C:p.F3L			
ANP32E		1	2	1		0.008
	Cases		ENST00000436748:exon3:c.G227C:p.S76T; ENST00000533654:exon4:c.A434G:p.K145R			
	Controls		ENST00000436748:exon6:c.G629T:p.R210L			
CABP5		19	2	1		0.008
	Cases		ENST00000293255:exon4:c.C281A:p.T94N; exon3:c.G201A:p.M67I			
	Controls		ENST00000293255:exon3:c.A169C:p.M57L			
CELA2B		1	2	1		0.008
	Cases		ENST00000375910:exon6:c.G576A:p.W192X; ENST00000422901:exon3:c.G271A:p.G91R			
	Controls		ENST00000375910:exon7:c.T739C:p.Y247H			
INSIG2		2	2	1		0.008
	Cases		ENST00000245787:exon2:c.T89C:p.I30T; exon2:c.C236T:p.T79M			
	Controls		ENST00000245787:exon4:c.G376A:p.D126N			
KCTD7		7	2	1		0.008
	Cases		ENST00000275532:exon4:c.G814A:p.V272M; exon4:c.C758T:p.S253L			
	Controls		ENST00000275532:exon4:c.G506A:p.R169Q			
MR1		1	2	1		0.008
	Cases		ENST00000367579:exon5:c.A747T:p.E249D; exon4:c.C740T:p.P247L			
	Controls		ENST00000367579:exon4:c.G654A:p.W218X			
MRO		18	2	1		0.008
	Cases		ENST00000436348:exon5:c.G578A:p.R193Q; exon5:c.G565A:p.V189I			
	Controls		ENST00000436348:exon3:c.A223G:p.S75G			
NR2E1		6	2	1		0.008
	Cases		ENST00000368983:exon1:c.G136A:p.G46S; exon5:c.A634G:p.M212V			
	Controls		ENST00000368983:exon7:c.G1000A:p.V334I			

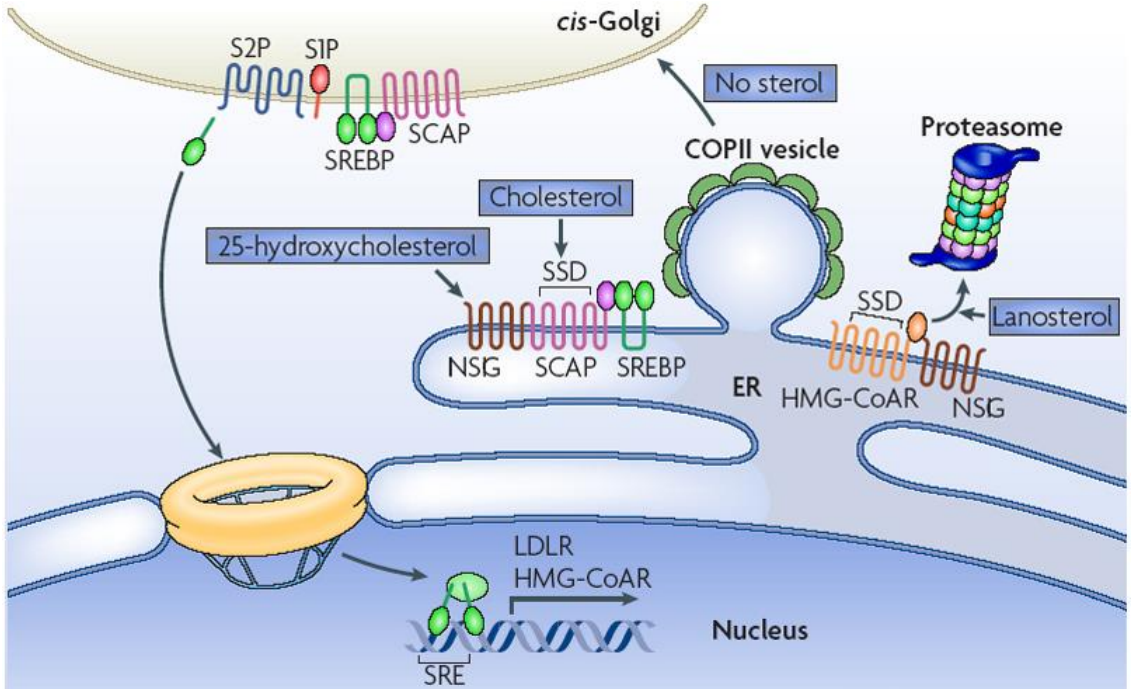
NXT2		X	2	1		0.008
	Cases		2X ENST00000218004:exon1:c.A82C:p.T28P			
	Controls		ENST00000218004:exon2:c.C155T:p.T52I			
PABPC1		8	2	1		0.008
	Cases		ENST00000318607:exon9:c.A1250C:p.Q417P;exon10:c.G1364A:p.R455H			
	Controls		ENST00000523555:exon3:c.226+3A>G			
PODXL		7	2	1		0.008
	Cases		ENST00000537928:exon3:c.G821A:p.R274K; exon5:c.A992G:p.H331R			
	Controls		ENST00000537928:exon8:c.C1246G:p.Q416E			
PUS3		11	2	1		0.008
	Cases		ENST00000530811:exon1:c.T74C:p.V25A; exon2:c.T824C:p.L275P			
	Controls		ENST00000530811:exon4:c.945-8T>C			
TXNDC15		5	2	1		0.008
	Cases		ENST00000511070:exon2:c.C130T:p.R44W; ENST00000507024:exon2:c.G91A:p.A31T			
	Controls		ENST00000358387:exon2:c.G534C:p.E178D			
WDR89		14	2	1		0.008
	Cases		ENST00000394942:exon2:c.T821C:p.L274S; exon2:c.A553G:p.M185V			
	Controls		ENST00000394942:exon2:c.A860G:p.D287G			

7.2.5. *CH25H* and *INSIG2* variants

CH25H codes for cholesterol 25-hydroxylase, known to catalyse the formation of the oxysterol - 25-hydroxycholesterol (25-HC) (Lund et al. 1998). The *INSIG2* gene, which also exhibited an excess of *rare functional* variants in the FH cohort in comparison to the controls ($p < 7.8x10^{-3}$) (Table 36), has been demonstrated to regulate the activity of SREBPs, a family of major lipid metabolism transcription factors, via direct biding of 25-HC (Radhakrishnan et al. 2007). Thus both genes, *CH25H* and *INSIG2*, are involved in the same pathway of cholesterol metabolism, as summarized in Figure 45.

Figure 45

CH25H and INSIG2 involvement in the regulation of SREBP activity (adapted from (Ikonen 2008)). CH25H catalyses the formation of 25-hydroxycholesterol from cholesterol. 25-hydroxycholesterol binds INSIG2 in the endoplasmic reticulum, which further prevents SCAP (SREBP cleavage activating protein) from transporting SREBP to the Golgi. Consequently, SREBP activity is inhibited and thus the expression of its target protein (including LDLR and HMG-CoAR) descreases.



There were three heterozygous variants found in *CH25H*, all confirmed by Sanger sequencing (Figure 46), of which one leads to a formation of a premature stop codon at residue 81, predicted to have a damaging effect on the protein, the second is affecting a well-conserved residue across species, c.568G>A (p.V190I), and third, c.716A>C (p.H239P), is altering one of the crucial residues of the His Box 3 domain, known to play a crucial role, together with His Boxes 1 and 2, in the catalytic activity of *CH25H* (Fox et al. 1994), as shown in Figure 47. Two rare functional variants were found in the control cohort, both being non-synonymous (p.P197Q and p.C248G). The p.P197Q is located in a conserved region of the protein however it is predicted as Tolerated/Benign/Neutral by SIFT/PolyPhen/Mutation Taster. The p.C248G variant affects a residue that is not conserved (Holmes et al. 2011).

Figure 46

Sanger sequencing confirmation of novel *CH25H* variants, A. c.716A>C (p.H239P) and c.568G>A (p.V190I); B. c.244C>T (p.Q81*). Primers used for the amplification of the region are highlighted in blue (A) and in purple (B).

A.

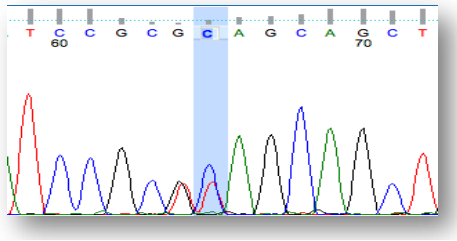
B.

CH25H sequencing (order #4108796)

```
90966991 TGCAGCCCCTCTGGGACCACCTGAGGAGCTGGGAGGCCCTCCTACAGTCGCCCTTCTTCC 90966932
90966931 CGGTCATCTTCTCCATCACCACATACRTRGGCTTTTGCCTGCCCTTCGTGGTCCTGGATA 90966872
90966871 TCCTGTGCTCCTGGGTGCCCGCCCTGCGGCGCTACAAGATCCAYCCTGACTTCTCGCCAT 90966812
90966811 CCGCGCAGCAGCTGCTACCTTGCCTGVGRCAGACCCTCTACCAGCAKGTGATGTTTGTGT 90966752
90966751 TCCCCGTRACGCTGMTGCAYTGGGCCYGCAGCCCGGCCCTCCTGCCCCACGAAGCTCCCG 90966692
90966691 AGCTGCTCCTGCTGCTGCACCACATCCTGTTCTGCCTGCYACTCTTCGACATGGAGTTCT 90966632
90966631 TYGTGTGGCACCTGCTGCACCACAAGGTGCCCTGGYTGTACCGCACCTTYCACAAGGTGC 90966572
90966571 ACCACCRGAACTCGTCCTCGTTCGCRCTGGCAACGCAGTATATGAGCGTCTGGGAACTGT 90966512
90966511 TTTCTTTGGGYTTCTTCGACATGATGAACGTCACACTGCTYGGGTGCCACCCGCTCACCA 90966452
90966451 CCCTGACCTTCCACGTGGTCAACATCTGKCTTTCCGTGGAGGWCCACTCCGGCTACAAYT 90966392
90966391 TCCCTTGGTCCACTCACAGACTGGTGCCCTTCGGGTGGTACGGGGGTGTGGTRCACCACG 90966332
90966331 ACCTGCATCACTCTCACTTTAACTGCAACTTCGCTCCRTACTTTACACACTGGGACAAAA 90966272
90966271 YACTGGGAACRCTGCGRACTGCATCTGTCCCAGCGCRRTGATGTGGCTGCGGTGGGTGCC 90966212
```

*Primers:
CH25H_03F and CH25H_04R

p.A80A (c.243G>T) and p.Q81* (c.244C>T) (UK10K_HYP5159267)



WT (UK10K_HYP5002218)

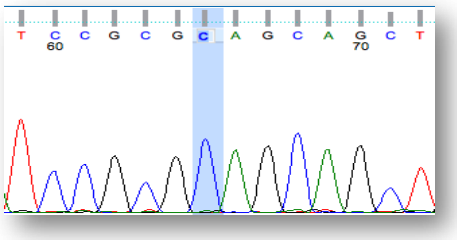
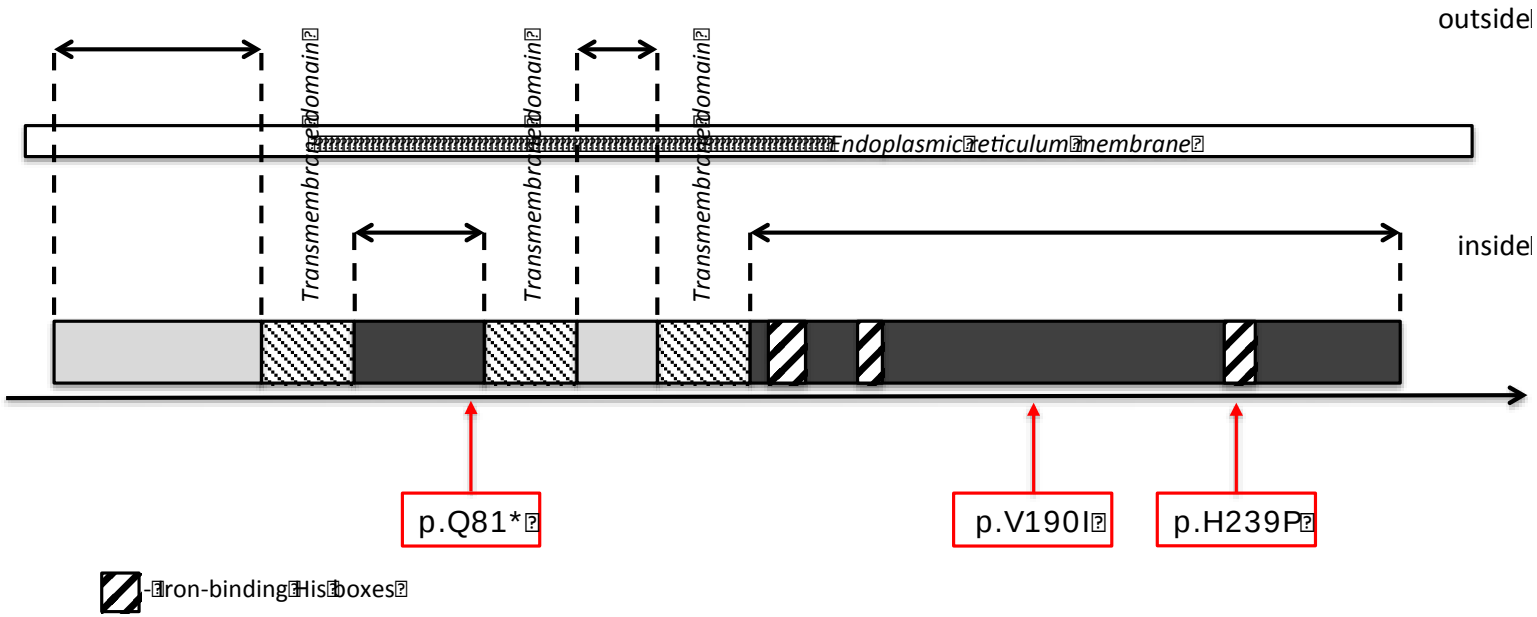


Figure 47

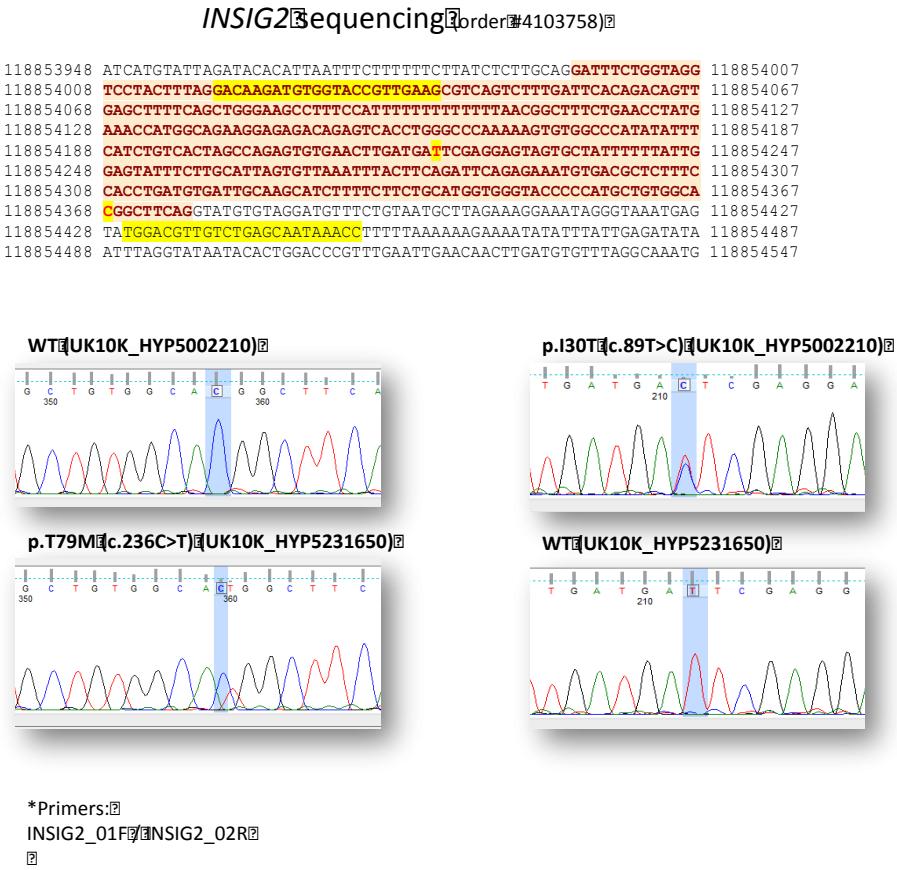
Schematic representation of the intronless *CH25H* gene and the localisation of novel variants identified in the FH cohort. *CH25H* encodes an enzyme, cholesterol 25-hydroxylase, known to be spanning the endoplasmic reticulum membrane, with two domains (including the N terminal) located outside of the membrane (in light grey), three 20 amino acid long transmembrane regions, and two domains positioned inside the membrane, which contain three His boxes, essential for the catalytic activity of the enzyme (Holmes et al. 2011)



Sanger sequencing also confirmed two *rare functional* variants in the *INSIG2* gene called in the cases, both non-synonymous changes (Figure 48). A mutation prediction report generated by Project HOPE (Venselaar et al. 2010) highlighted that the c.89T>C (p.I30T) variant will cause an empty space in the core of INSIG2 because of the size differences between the wild type Isoleucine and the smaller mutant - Threonine. The other variant, c.236C>T (p.T79M), located in the transmembrane domain of INSIG2 is predicted to have an effect on the hydrophobic interactions within the core of the protein or with the membrane lipids, because the mutant Methionine is more hydrophobic than the wild type Threonine. One rare missense variant was found in *INSIG2* in the controls (p.D126N), which was predicted as Tolerated/Probably Damaging/Disease Causing (by SIFT/PolyPhen/Mutation Taster).

Figure 48

Sanger sequencing confirmation of novel *INSIG2* variants, c.236C>T (p.T79M) and c.89T>V (p.I30T). Primers used for the amplification of the region are highlighted in yellow.



7.2.6. Re-sequencing of *CH25H*

Additional 160 unrelated FH patients, from independent cohorts studied in Chapter Three (Greece, Canada, Holland, South-Indian) who had LDL-C gene score below 1.16 cutoff for polygenic hypercholesterolaemia, were sequenced for changes within the *CH25H* gene. The gene was found to be highly conserved, with two common polymorphisms (MAF 0.19, and 0.20), which do not affect the amino acid composition of the protein.

7.3. Discussion

This study identified 25 mutations in known FH genes (23 in *LDLR* and 2 in *APOB*), which were missed by the current screening protocol. This finding confirmed that *LDLR* locus is highly heterogeneous and mutations within this gene account for the majority of FH causes. The need for an update of current screening methods has been previously discussed in Chapter Six. In addition to the known FH mutations, six novel *APOB* variants were identified, distributed across different exons, in five patients. The pathogenicity of these variants remains to be tested, however their very low frequency and absence from the control cohorts can indicate a damaging effect. Most of the current mutation screening strategies for FH are focused on a selected region of exon 26 of *APOB*, because of its established function (Boren et al. 1998). The whole exome sequencing enabled to extent the analysis to the entire coding sequence of the gene, by which novel variants, unique to the FH cohort, were found.

Rare functional variants in genes associated with LDL-C levels in the GWAS meta-analysis were not significantly overrepresented in the FH cohort, when compared to controls. This suggests that rare variants that have a major effect on function in these genes known to have common LDL-C variants of modest effect are unlikely to be a common cause of FH.

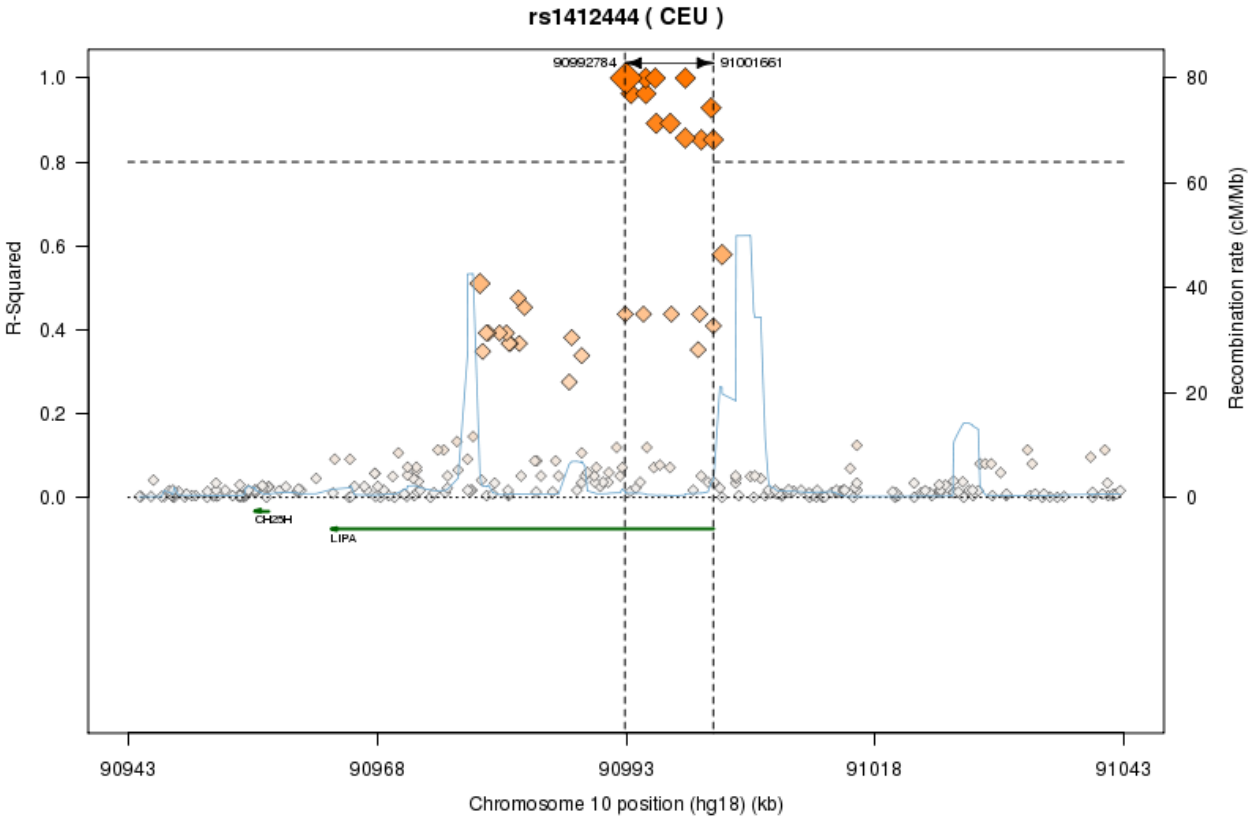
Although we found no evidence that rare variants in the LDL-C associated genes identified by GWAS were causing FH, the cumulative effect of common LDL-rising alleles in these genes was shown to be the likely cause of high LDL-C in a significant proportion (25%) of the examined patients. A gene score above the top decile for a healthy population cutoff (1.16) was also observed in two patients with considerably mild *LDLR* mutations, which demonstrates that common polymorphisms can contribute to the presentation of an individual carrying a mild effect FH mutation with LDL-C levels above the diagnostic threshold, as discussed in Chapter Five.

Genes *CH25H* and *INSIG2* are the strongest candidates for novel FH loci among the final 19 genes, showing the excess of *rare functional* variants, based on the available literature on functions of the proteins for which they code. *CH25H* encodes 25-cholesterol hydroxylase, which catalyses the formation of 25-HC from cholesterol. It has been demonstrated that both cholesterol and 25-HC can regulate the function of SREBP, a transcription factor known to regulate the expression of several key players in the lipid metabolism (Adams et al. 2004, Shao and Espenshade 2012). It is known that the regulation of SREBP activity depends on binding of 25-HC to INSIG2,

encoded by the *INSIG2* gene (Radhakrishnan et al. 2007). The recently updated GLGC GWAS study, with > 180,000 individuals had now identified an association at the genome wide level of LDL-C with an *INSIG2* gene variant (MAF = 0.08) (Willer et al. 2013b). A previous GWAS shown an association of *INSIG2* SNP with bone mineral density (Estrada et al. 2012). The *CH25H* variants identified in this study have not been observed in 1000 Genomes, 6500ESP and 69CG or the 1,926 control exomes. An additional novel variant has been found in the *CH25H* re-sequencing cohort of 160 FH patients. The possible effect of this promoter variant (c.-58C>T) on the expression of *CH25H* remains to be tested in the future work. *CH25H* is located in close proximity (<7 Kb) to the *LIPA* gene. A common (MAF= 0.38) intronic SNP (rs1412444) in *LIPA* has been shown to be associated with CHD in two independent GWAS (Wild et al. 2011, 'A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease' 2011), however the SNP is not in LD with the region surrounding *CH25H* (Figure 49). Homozygous mutations in *LIPA* are known to cause cholesterol ester storage disease due to deficiency of lysosomal acid lipase (LAL) (Patrick and Lake 1969). A recent case report described a patient with a homozygous splice site junction mutation in *LIPA* gene who was initially diagnosed with autosomal recessive hypercholesterolaemia (Stitziel et al. 2013). None of the *CH25H* rare variant carriers had an additional rare variant in *LIPA*, thus the *CH25H* variants could not mark a causative haplotype in this region, although variants in the intragenic region remain undiscovered.

Figure 49

Linkage disequilibrium (LD) region of *LIPA* SNP (rs1412444), which was shown to be associated with CHD. The variant does not share LD with any common variant in *CH25H* and its surrounding region. Plot was generated using SNAP tool (http://www.broadinstitute.org/mpg/snap/ldsearch.php).



In summary, in 125 DFH unrelated patients without an identified mutation by conventional screening methods, the analyses identified 25 disease-causing variants in already known FH loci, as well as six previously unreported *APOB* variants in five patients. LDL-C gene score analysis found that 31 (29 mutation negative) patients had SNP score in top decile of the general population and therefore had a definite polygenic aetiology, and an additional five had a potential functional variant in *CH25H* or *INSIG2*. This means that explanation for the FH phenotype is still lacking in 50% of the patients. This suggests that some causal variants could be missed at different stages of the data processing or analysis. The variant calling pipeline used for this study was carefully optimised for the majority of the exome regions, though some calls in poorly covered regions could be missed. There is a possibility that there are genetic causes located outside of the protein coding region, affecting protein expression, posttranscriptional stability, or altering gene splicing. Also, it is possible that the LDL-C gene score cutoff of 1.16 for polygenic hypercholesterolaemia is too stringent. Thus using the 9th decile cutoff of 1.08, in which a 41% of WHII individuals had LDL-C above the 4.9 mmol/l (mean LDL-C = 4.68 ±1.05mmol/l) FH diagnostic level, could be more appropriate. By doing so, the phenotype would be explained in an additional nine mutation-negative patients. Finally because the burden test results are dependent on the number of associated variants and variants diluting the signal, it is possible that novel FH mutations are located in a highly polymorphic gene, in which it is difficult to pick up the true mutation.

Thus overall, no major novel locus for FH was detected, with no gene having a functional variant in more than 3 patients. This suggests that the genetic

cause of FH in these unexplained cases is likely to be very heterogeneous, which complicates the novel gene discovery and diagnostic process.

VIII. CHAPTER EIGHT – General Discussion

8.1. Challenges of mutation prediction

The findings of this thesis show that, in the UK, the spectrum of FH mutations is markedly heterogeneous, and novel *LDLR* variants continue to be discovered. These results highlight the need for more advanced tools enabling a quick and decisive assessment of the functional effect of a variant on the protein. The already existing bioinformatic tools, such as PolyPhen, SIFT, Mutation Taster, and Project HOPE provide an evaluation of the variant's effect, based on multiple sequence alignments, and examining the degree of conservation of amino acid residues, Bayes classifiers, which predict the effect using the variant frequencies provided by large databases (including the 1000 Genomes Project and the Human Gene Mutation Database (HGMD)), or on the comparison of the physicochemical characteristics of the wild type and the substituted amino acids. The algorithms have proven to be accurate for prediction of the majority of human mutations and are optimized for high-throughput analysis of next generation sequencing data. However, the existing *in silico* tools are mainly designed to assess the effect of non-synonymous changes. Results of Chapter Three show that roughly 20% of the observed *LDLR* mutations are small insertions or deletions, 2% are variants located within or in close proximity to a splicing site, and 2% occur in the promoter region. The effect of such nucleotide changes cannot be examined by the tools mentioned above. Although some of the small insertions/deletions cause a frameshift and usually do not require further investigations, novel approaches examining the protein structure and/or interactions with other molecules are essential. A recently published

report on the UCL FH mutation database (Usifo et al. 2012) makes a use of the Single Amino Acid Polymorphism database (SAAPdb) (Hurst et al. 2009), which enables visualization of the structural effects of the studied variant. However, this method is limited to proteins that have an established crystal structure, therefore cannot be used for examining changes within, for example, the transmembrane domain of *LDLR*. Other advanced analytic tools of protein structure and conservation include ScoreCons (Lees et al. 2010), Jalview (Waterhouse et al. 2009) and PyMol (http://www.pymol.org). This structural analysis requires in-depth knowledge of the physiochemical interactions between amino acids, advanced bioinformatics skills and appropriate computing equipment, thus at the moment it is not widely available for clinical use.

The ability to predict the functional effect of a variant is especially important in the next generation sequencing era. The rapid development of high-throughput technologies over the past decade led to the discovery of thousands of variants, however identifying which are causal is not always successful. The Gene Ontology (GO) (http://www.geneontology.org) project can also provide an important tool for the analysis of next generation sequencing data. The GO project utilizes already published functional information to annotate genes and gene products. The information is incorporated in many different bioinformatics tools, which allows for the analysis of, for example, protein-protein interactions, or protein pathway networks. The GO project aims to completely annotate twelve reference genomes and biological systems, however currently only the Cardiovascular and Renal annotation projects are funded.

The analysis of variants that may potentially affect the correct splicing of pre-mRNA is still rather laborious. Currently available *in silico* tools, such as Human Splicing Finder (Desmet et al. 2009), compare the strength of the splice donor or acceptor sites and the branch points in the wild type and mutated state, by analyzing the length between the acceptor and the donor sites, the conservation of splicing factors motifs, and the frequency of the tested variant, however they cannot account for the complexity of a cellular environment, thus in some cases the results may be misleading. Although the preferred method for analyzing the effect of splice site variants is the direct analysis of mRNA from an appropriate tissue from a patient and a control, the genetic material may not always be available due to several different reasons, such as difficult tissue access. An *in vivo* approach, which employs a 'minigene' system has been developed to predict the potential consequence of splice site variants. The method applies minigene-based technologies used for alternative splicing analysis (Vibe-Pedersen et al. 1984). Any genomic region of interest- for example, an exon with part of its intronic flanking regions- can be amplified from normal or affected individuals and cloned into the minigene, forming a plasmid. The plasmid is then transfected in to an appropriate cell line where it is transcribed and the resulting mRNA can be analysed by Real-Time PCR (RT-PCR). This exon-trapping method was shown to be effective at correctly identifying the variant's effect *in vivo* and it can be used to analyse complex exonic/intronic regions (Baralle et al. 2003, Cooper 2005, Kishore et al. 2008). The minigene system is currently being tested for *LDLR* splice variants at the UCL Cardiovascular Genetics lab.

In a small proportion of FH patients the *LDLR* gene has no defect and the defective plasma clearance of LDL particles due to a defect in ApoB can be examined by analyzing the mutant:wild type ratios of ApoB in the plasma. Authors of a recent report ((Thomas et al.) published online on 13 Jun 2013) applied mass spectrometry technology to analyse a novel *APOB* variant, p.R50W, which was also identified in this project as a potential FH mutation (Chapter Seven). The authors demonstrated, that in the patients plasma, there was 2.5-3.5 fold more of the mutated ApoB peptide in comparison to the wild type ApoB peptide. The abundance of the two peptides was quantified using synthetic peptides. The finding was supported by a family co-segregation of the variant, which overall demonstrated that the ApoB p.W50 has a significant impact on the plasma LDL clearance. This method can be adapted for assessing other *APOB* variants, however it requires a multidisciplinary collaboration.

It is widely accepted that one of the main criteria to evaluate the pathogenicity of a variant in an autosomal dominant disorder is a family co-segregation analysis (Cotton and Scriver 1998). However, in some instances segregation analysis of the phenotype with a single variant may give misleading results, for example, the co-segregation of the novel *LDLR* variant (p.Asn316Lys) in the Indian Family, as demonstrated in Chapter Four. Further analysis of the polygenic effect of common variants (discussed in Chapter Five) or the effect of rare LDL-lowering variants (such as variants in *PCSK9*) should be considered before a final conclusion is made.

8.2. Application of LDL-C SNPs score

This thesis provides the first validation of the findings of the study (Talmud et al. 2013), which demonstrated that in a proportion of FH patients raised LDL-C may have a polygenic cause. This is an important step in the translational research of hypercholesterolaemia. The LDL-C SNPs score analysis can be used to identify hypercholesterolaemic patients who don't have a monogenic cause, which is vital for the efficacy of cascade testing. It is expected that less than 50% of relatives of patients with polygenic hypercholesterolaemia will be affected (i.e. have LDL-C levels above the diagnostic cut-off for relatives (Starr et al. 2008)), which would compromise cascade testing for FH. Conversely, in individuals in the lower two deciles of the LDL-C SNPs score, it is extremely unlikely that their elevated cholesterol levels can be explained by a polygenic mechanism, suggesting that they are more likely to have a novel monogenic cause, yet to be identified. Therefore using the gene score information may increase the probability of finding novel rare FH variants with a strong effect. Furthermore, combining the LDL-C SNPs score information with the FH mutation information for the mutation positive patients can help to explain the LDL-C level variability between patients, which consequently may help to decide on the appropriate LDL-C lowering therapy.

Finally, it is important to compare the impact of a polygenic vs. a monogenic cause of hypercholesterolaemia on the development of atherosclerosis. For example, future studies should use non-invasive methods such as computerised tomography (CT) angiography and carotid intima media thickness (cIMT) ultrasound measures to examine the extent of the

progression of atherosclerosis in patients with similar LDL-C levels but a polygenic vs. a monogenic cause.

8.3. Novel FH loci

The whole exome sequencing analysis described in Chapter Seven did not identify a single 4th FH locus causing high LDL-C levels in patients negative for mutations in *LDLR*, *APOB*, *PCSK9* and *LDLRAP1*. Nonetheless, several novel variants, which are likely to cause FH, were discovered. The strongest candidates for further functional study include variants in *CH25H* and *INSIG2* genes. Both genes encode proteins which function in the pathway of SREBP regulation. Although proving that mutations in *CH25H* and *INSIG2* can cause FH requires further investigation, this finding can be important for the discovery of novel drug targets for the LDL-lowering therapy. However, if the identified five variants (three in *CH25H* and two in *INSIG2*) are shown to be FH-causing, their frequency is very low, and consequently the overall FH mutation detection rate will not increase significantly. This finding also suggests the possibility that there are genetic causes located outside of the protein coding region, or that the cause of hypercholesterolaemia is multifactorial, i.e. due to interactions of several environmental factors or environment-gene interactions. Thus, altogether the analysis demonstrate that the FH aetiology is much more heterogeneous, than previously thought.

Bibliography

(WHO), W. H. O. (2011) *The impact of chronic disease in India*, WHO Global Database.

Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., Hurles, M. E., McVean, G. A. and Consortium, G. P. (2010) 'A map of human genome variation from population-scale sequencing', *Nature,* 467(7319), 1061-73.

Abifadel, M., Varret, M., Rabes, J. P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., Derre, A., Villeger, L., Farnier, M., Beucler, I., Bruckert, E., Chambaz, J., Chanu, B., Lecerf, J. M., Luc, G., Moulin, P., Weissenbach, J., Prat, A., Krempf, M., Junien, C., Seidah, N. G. and Boileau, C. (2003) 'Mutations in PCSK9 cause autosomal dominant hypercholesterolemia', *Nat Genet,* 34(2), 154-6.

Acton, S., Rigotti, A., Landschulz, K. T., Xu, S., Hobbs, H. H. and Krieger, M. (1996) 'Identification of scavenger receptor SR-BI as a high density lipoprotein receptor', *Science,* 271(5248), 518-20.

Adams, C. M., Reitz, J., De Brabander, J. K., Feramisco, J. D., Li, L., Brown, M. S. and Goldstein, J. L. (2004) 'Cholesterol and 25-hydroxycholesterol inhibit activation of SREBPs by different mechanisms, both involving SCAP and Insigs', *The Journal of biological chemistry,* 279(50), 52772-80.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. and Sunyaev, S. R. (2010) 'A method and server for predicting damaging missense mutations', *Nature methods,* 7(4), 248-9.

Alonso, R., Defesche, J. C., Tejedor, D., Castillo, S., Stef, M., Mata, N., Gomez-Enterria, P., Martinez-Faedo, C., Forga, L. and Mata, P. (2009) 'Genetic diagnosis of familial hypercholesterolemia using a DNA-array based platform', *Clin Biochem,* 42(9), 899-903.

Arambepola, C., Farmer, A. J., Perera, R. and Neil, H. A. (2007) 'Statin treatment for children and adolescents with heterozygous familial hypercholesterolaemia: a systematic review and meta-analysis', *Atherosclerosis,* 195(2), 339-47.

Ashavaid, T. F., Altaf, A. K. and Nair, K. G. (2000) 'Molecular basis of familial hypercholesterolemia: An Indian experience', *Indian J Clin Biochem,* 15(Suppl 1), 11-9.

Austin, M. A., Hutter, C. M., Zimmern, R. L. and Humphries, S. E. (2004) 'Genetic causes of monogenic heterozygous familial hypercholesterolemia: a HuGE prevalence review', *American journal of epidemiology,* 160(5), 407-20.

Avis, H. J., Vissers, M. N., Stein, E. A., Wijburg, F. A., Trip, M. D., Kastelein, J. J. and Hutten, B. A. (2007) 'A systematic review and meta-analysis of statin therapy in children with familial hypercholesterolemia', *Arterioscler Thromb Vasc Biol,* 27(8), 1803-10.

Baigent, C., Keech, A., Kearney, P. M., Blackwell, L., Buck, G., Pollicino, C., Kirby, A., Sourjina, T., Peto, R., Collins, R., Simes, R. and Collaborators, C. T. T. C. (2005) 'Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins', *Lancet,* 366(9493), 1267-78.

Baralle, M., Baralle, D., De Conti, L., Mattocks, C., Whittaker, J., Knezevich, A., Ffrench-Constant, C. and Baralle, F. E. (2003) 'Identification of a mutation that perturbs NF1 agene splicing using genomic DNA samples and a minigene assay', *J Med Genet,* 40(3), 220-2.

Barbagallo, C. M., Emmanuele, G., Cefalu, A. B., Fiore, B., Noto, D., Mazzarino, M. C., Pace, A., Brogna, A., Rizzo, M., Corsini, A., Notarbartolo, A., Travali, S. and Averna, M. R. (2003) 'Autosomal recessive hypercholesterolemia in a Sicilian kindred harboring the 432insA mutation of the ARH gene', *Atherosclerosis,* 166(2), 395-400.

Benn, M., Watts, G. F., Tybjaerg-Hansen, A. and Nordestgaard, B. G. (2012) 'Familial hypercholesterolemia in the danish general population: prevalence, coronary artery disease, and cholesterol-lowering medication', *J Clin Endocrinol Metab,* 97(11), 3956-64.

Bennet, A. M., Di Angelantonio, E., Ye, Z., Wensley, F., Dahlin, A., Ahlbom, A., Keavney, B., Collins, R., Wiman, B., de Faire, U. and Danesh, J. (2007) 'Association of apolipoprotein E genotypes with lipid levels and coronary risk', *JAMA,* 298(11), 1300-11.

Berge, K. E., Ose, L. and Leren, T. P. (2006) 'Missense mutations in the PCSK9 gene are associated with hypocholesterolemia and possibly increased response to statin therapy', *Arterioscler Thromb Vasc Biol,* 26(5), 1094-100.

Bertolini, S., Pisciotta, L., Rabacchi, C., Cefalù, A. B., Noto, D., Fasano, T., Signori, A., Fresa, R., Averna, M. and Calandra, S. (2013) 'Spectrum of mutations and phenotypic expression in patients with autosomal dominant hypercholesterolemia identified in Italy', *Atherosclerosis,* 227(2), 342-8.

Bhathena, S. J. (2006) 'Relationship between fatty acids and the endocrine and neuroendocrine system', *Nutr Neurosci,* 9(1-2), 1-10.

Boer, J. M., Feskens, E. J., Verschuren, W. M., Seidell, J. C. and Kromhout, D. (1999) 'The joint impact of family history of myocardial infarction and other risk factors on 12-year coronary heart disease mortality', *Epidemiology,* 10(6), 767-70.

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', *The Journal of clinical investigation,* 101(5), 1084-93.

Brown, M. S. and Goldstein, J. L. (1974a) 'Familial hypercholesterolemia: defective binding of lipoproteins to cultured fibroblasts associated with impaired regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity', *Proc Natl Acad Sci U S A,* 71(3), 788-92.

Brown, M. S. and Goldstein, J. L. (1974b) 'Suppression of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity and inhibition of growth of human fibroblasts by 7-ketocholesterol', *J Biol Chem,* 249(22), 7306-14.

Brown, M. S. and Goldstein, J. L. (1979) 'Receptor-mediated endocytosis: insights from the lipoprotein receptor system', *Proceedings of the National Academy of Sciences of the United States of America,* 76(7), 3330-7.

Campagna, F., Martino, F., Bifolco, M., Montali, A., Martino, E., Morrone, F., Antonini, R., Cantafora, A., Verna, R. and Arca, M. (2008) 'Detection of familial hypercholesterolemia in a cohort of children with hypercholesterolemia: results of a family and DNA-based screening', *Atherosclerosis,* 196(1), 356-64.

Catapano, A. L., Reiner, Z., De Backer, G., Graham, I., Taskinen, M. R., Wiklund, O., Agewall, S., Alegria, E., Chapman, M., Durrington, P., Erdine, S., Halcox, J., Hobbs, R., Kjekshus, J., Filardi, P. P., Riccardi, G., Storey, R. F. and Wood, D. (2011) 'ESC/EAS Guidelines for the management of dyslipidaemias The Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and the European Atherosclerosis Society (EAS)', *Atherosclerosis,* 217(1), 3-46.

Cenarro, A., Garcia-Otin, A. L., Tejedor, M. T., Solanas, M., Jarauta, E., Junquera, C., Ros, E., Mozas, P., Puzo, J., Pocovi, M. and Civeira, F. (2011) 'A presumptive new locus for autosomal dominant hypercholesterolemia mapping to 8q24.22', *Clinical genetics,* 79(5), 475-81.

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(6), 3116-23.

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., Nayir, A., Bakkaloglu, A., Ozen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S. and Lifton, R. P. (2009) 'Genetic diagnosis by whole exome capture and massively parallel DNA sequencing', *Proceedings of the National Academy of Sciences of the United States of America,* 106(45), 19096-101.

Choi, S. Y., Fong, L. G., Kirven, M. J. and Cooper, A. D. (1991) 'Use of an anti-low density lipoprotein receptor antibody to quantify the role of the LDL receptor in the removal of chylomicron remnants in the mouse in vivo', *J Clin Invest,* 88(4), 1173-81.

Coffey, A. J., Kokocinski, F., Calafato, M. S., Scott, C. E., Palta, P., Drury, E., Joyce, C. J., Leproust, E. M., Harrow, J., Hunt, S., Lehesjoki, A. E., Turner, D. J., Hubbard, T. J. and Palotie, A. (2011) 'The GENCODE exome: sequencing the complete human exome', *European journal of human genetics : EJHG,* 19(7), 827-31.

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(2), 161-5.

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C. H., Kristiansson, K., Macarthur, D. G., Macdonald, J. R., Onyiah, I., Pang, A. W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N. P., Lee, C., Scherer, S. W., Hurles, M. E. and Consortium, W. T. C. C. (2010) 'Origins and functional impact of copy number variation in the human genome', *Nature,* 464(7289), 704-12.

Cooper, G. M. and Shendure, J. (2011) 'Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data', *Nature reviews. Genetics,* 12(9), 628-40.

Cooper, T. A. (2005) 'Use of minigene systems to dissect alternative splicing elements', *Methods,* 37(4), 331-40.

Cotton, R. G. and Scriver, C. R. (1998) 'Proof of "disease causing" mutation', *Hum Mutat,* 12(1), 1-3.

Cuchel, M., Bloedon, L. T., Szapary, P. O., Kolansky, D. M., Wolfe, M. L., Sarkis, A., Millar, J. S., Ikewaki, K., Siegelman, E. S., Gregg, R. E. and

Rader, D. J. (2007) 'Inhibition of microsomal triglyceride transfer protein in familial hypercholesterolemia', *N Engl J Med,* 356(2), 148-56.

Damgaard, D., Larsen, M. L., Nissen, P. H., Jensen, J. M., Jensen, H. K., Soerensen, V. R., Jensen, L. G. and Faergeman, O. (2005) 'The relationship of molecular genetic to clinical diagnosis of familial hypercholesterolemia in a Danish population', *Atherosclerosis,* 180(1), 155-60.

Davis, C. G., Elhammer, A., Russell, D. W., Schneider, W. J., Kornfeld, S., Brown, M. S. and Goldstein, J. L. (1986) 'Deletion of clustered O-linked carbohydrates does not impair function of low density lipoprotein receptor in transfected fibroblasts', *J Biol Chem,* 261(6), 2828-38.

Davis, C. G., Goldstein, J. L., Sudhof, T. C., Anderson, R. G., Russell, D. W. and Brown, M. S. (1987) 'Acid-dependent ligand dissociation and recycling of LDL receptor mediated by growth factor homology region', *Nature,* 326(6115), 760-5.

Day, I. N., Bolla, M., Haddad, L., O'Dell, S. and Humphries, S. E. (1996) 'Microtiter Array Diagonal Gel Electrophoresis (MADGE) for Population Scale Genotype Analyses', *Methods in molecular medicine,* 5, 269-81.

Day, I. N., Haddad, L., O'Dell, S. D., Day, L. B., Whittall, R. A. and Humphries, S. E. (1997a) 'Identification of a common low density lipoprotein receptor mutation (R329X) in the south of England: complete linkage disequilibrium with an allele of microsatellite D19S394', *J Med Genet,* 34(2), 111-6.

Day, I. N., Whittall, R. A., O'Dell, S. D., Haddad, L., Bolla, M. K., Gudnason, V. and Humphries, S. E. (1997b) 'Spectrum of LDL receptor gene mutations in heterozygous familial hypercholesterolemia', *Hum Mutat,* 10(2), 116-27.

Dedoussis, G. V., Schmidt, H. and Genschel, J. (2004) 'LDL-receptor mutations in Europe', *Hum Mutat,* 24(6), 443-59.

DeLong, E. R., DeLong, D. M. and Clarke-Pearson, D. L. (1988) 'Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach', *Biometrics,* 44(3), 837-45.

Desai, A. and Jere, A. (2012) 'Next-generation sequencing: ready for the clinics?', *Clinical genetics,* 81(6), 503-10.

Desmet, F. O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M. and Beroud, C. (2009) 'Human Splicing Finder: an online bioinformatics tool to predict splicing signals', *Nucleic acids research,* 37(9), e67.

Di Angelantonio, E., Sarwar, N., Perry, P., Kaptoge, S., Ray, K. K., Thompson, A., Wood, A. M., Lewington, S., Sattar, N., Packard, C. J., Collins, R., Thompson, S. G., Danesh, J. and Collaboration, E. R. F. (2009) 'Major lipids, apolipoproteins, and risk of vascular disease', *JAMA,* 302(18), 1993-2000.

Dias, C. S., Shaywitz, A. J., Wasserman, S. M., Smith, B. P., Gao, B., Stolman, D. S., Crispino, C. P., Smirnakis, K. V., Emery, M. G., Colbert, A., Gibbs, J. P., Retter, M. W., Cooke, B. P., Uy, S. T., Matson, M. and Stein, E. A. (2012) 'Effects of AMG 145 on low-density lipoprotein cholesterol levels: results from 2 randomized, double-blind, placebo-controlled, ascending-dose phase 1 studies in healthy volunteers and hypercholesterolemic subjects on statins', *J Am Coll Cardiol,* 60(19), 1888-98.

Djurovic, S. and Berg, K. (1997) 'Epidemiology of Lp(a) lipoprotein: its role in atherosclerotic/thrombotic disease', *Clin Genet,* 52(5), 281-92.

Do, R., Xie, C., Zhang, X., Mannisto, S., Harald, K., Islam, S., Bailey, S. D., Rangarajan, S., McQueen, M. J., Diaz, R., Lisheng, L., Wang, X., Silander, K., Peltonen, L., Yusuf, S., Salomaa, V., Engert, J. C. and Anand, S. S. (2011) 'The effect of chromosome 9p21 variants on cardiovascular disease may be modified by dietary intake: evidence from a case/control and a prospective study', *PLoS Med,* 8(10), e1001106.

Dvir, H., Shah, M., Girardi, E., Guo, L., Farquhar, M. G. and Zajonc, D. M. (2012) 'Atomic structure of the autosomal recessive hypercholesterolemia phosphotyrosine-binding domain in complex with the LDL-receptor tail', *Proc Natl Acad Sci U S A,* 109(18), 6916-21.

Eiland, L. S. and Luttrell, P. K. (2010) 'Use of statins for dyslipidemia in the pediatric population', *J Pediatr Pharmacol Ther,* 15(3), 160-72.

Estrada, K., Styrkarsdottir, U., Evangelou, E., Hsu, Y. H., Duncan, E. L., Ntzani, E. E., Oei, L., Albagha, O. M., Amin, N., Kemp, J. P., Koller, D. L., Li, G., Liu, C. T., Minster, R. L., Moayyeri, A., Vandenput, L., Willner, D., Xiao, S. M., Yerges-Armstrong, L. M., Zheng, H. F., Alonso, N., Eriksson, J., Kammerer, C. M., Kaptoge, S. K., Leo, P. J., Thorleifsson, G., Wilson, S. G., Wilson, J. F., Aalto, V., Alen, M., Aragaki, A. K., Aspelund, T., Center, J. R., Dailiana, Z., Duggan, D. J., Garcia, M., Garcia-Giralt, N., Giroux, S., Hallmans, G., Hocking, L. J., Husted, L. B., Jameson, K. A., Khusainova, R., Kim, G. S., Kooperberg, C., Koromila, T., Kruk, M., Laaksonen, M., Lacroix, A. Z., Lee, S. H., Leung, P. C., Lewis, J. R., Masi, L., Mencej-Bedrac, S., Nguyen, T. V., Nogues, X., Patel, M. S., Prezelj, J., Rose, L. M., Scollen, S., Siggeirsdottir, K., Smith, A. V., Svensson, O., Trompet, S., Trummer, O., van Schoor, N. M., Woo, J., Zhu, K., Balcells, S., Brandi, M. L., Buckley, B. M., Cheng, S., Christiansen, C., Cooper, C., Dedoussis, G., Ford, I., Frost, M., Goltzman, D., Gonzalez-Macias, J.,

Kahonen, M., Karlsson, M., Khusnutdinova, E., Koh, J. M., Kollia, P., Langdahl, B. L., Leslie, W. D., Lips, P., Ljunggren, O., Lorenc, R. S., Marc, J., Mellstrom, D., Obermayer-Pietsch, B., Olmos, J. M., Pettersson-Kymmer, U., Reid, D. M., Riancho, J. A., Ridker, P. M., Rousseau, F., Slagboom, P. E., Tang, N. L., et al. (2012) 'Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture', *Nat Genet,* 44(5), 491-501.

Feher, M. and Richmond, W. (1997) *Lipids and Lipid Disorders,* Second edition ed., Mosby International Limited.

Forman-Kay, J. D. and Pawson, T. (1999) 'Diversity in protein recognition by PTB domains', *Curr Opin Struct Biol,* 9(6), 690-5.

Fox, B. G., Shanklin, J., Ai, J., Loehr, T. M. and Sanders-Loehr, J. (1994) 'Resonance Raman evidence for an Fe-O-Fe center in stearoyl-ACP desaturase. Primary sequence identity with other diiron-oxo proteins', *Biochemistry,* 33(43), 12776-86.

Frank-Kamenetsky, M., Grefhorst, A., Anderson, N. N., Racie, T. S., Bramlage, B., Akinc, A., Butler, D., Charisse, K., Dorkin, R., Fan, Y., Gamba-Vitalo, C., Hadwiger, P., Jayaraman, M., John, M., Jayaprakash, K. N., Maier, M., Nechev, L., Rajeev, K. G., Read, T., Röhl, I., Soutschek, J., Tan, P., Wong, J., Wang, G., Zimmermann, T., de Fougerolles, A., Vornlocher, H. P., Langer, R., Anderson, D. G., Manoharan, M., Koteliansky, V., Horton, J. D. and Fitzgerald, K. (2008) 'Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates', *Proc Natl Acad Sci U S A,* 105(33), 11915-20.

Futema, M., Plagnol, V., Whittall, R. A., Neil, H. A. and Humphries, S. E. (2012) 'Use of targeted exome sequencing as a diagnostic tool for Familial Hypercholesterolaemia', *J Med Genet,* 49(10), 644-649.

Futema, M., Whittall, R. A., Kiley, A., Steel, L. K., Cooper, J. A., Badmus, E., Leigh, S. E., Karpe, F. and Neil, H. A. (2013) 'Analysis of the frequency and spectrum of mutations recognised to cause familial hypercholesterolaemia in routine clinical practice in a UK specialist hospital lipid clinic', *Atherosclerosis*.

Garcia, C. K., Wilund, K., Arca, M., Zuliani, G., Fellin, R., Maioli, M., Calandra, S., Bertolini, S., Cossu, F., Grishin, N., Barnes, R., Cohen, J. C. and Hobbs, H. H. (2001) 'Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein', *Science,* 292(5520), 1394-8.

'A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease', (2011) *Nat Genet,* 43(4), 339-44.

Gibbons, G. F. (1990) 'Assembly and secretion of hepatic very-low-density lipoprotein', *Biochem J,* 268(1), 1-13.

Gill, P. J., Harnden, A. and Karpe, F. (2012) 'Familial hypercholesterolaemia', *BMJ,* 344, e3228.

Glomset, J. A. (1968) 'The plasma lecithins:cholesterol acyltransferase reaction', *J Lipid Res,* 9(2), 155-67.

Glynou, K., Laios, E., Drogari, E. and Tsaoussis, V. (2008) 'Development of a universal chemiluminometric genotyping method for high-throughput detection of 7 LDLR gene mutations in Greek population', *Clin Biochem,* 41(4-5), 335-42.

Goel, P. K., Bharti, B. B., Pandey, C. M., Singh, U., Tewari, S., Kapoor, A., Garg, N. and Sinha, N. (2003) 'A tertiary care hospital-based study of conventional risk factors including lipid profile in proven coronary artery disease', *Indian Heart J,* 55(3), 234-40.

Goldstein, J. L. and Brown, M. S. (1974) 'Binding and degradation of low density lipoproteins by cultured human fibroblasts. Comparison of cells from a normal subject and from a patient with homozygous familial hypercholesterolemia', *J Biol Chem,* 249(16), 5153-62.

Goldstein, J. L., DeBose-Boyd, R. A. and Brown, M. S. (2006) 'Protein sensors for membrane sterols', *Cell,* 124(1), 35-46.

Goldstein, J. L., Kita, T. and Brown, M. S. (1983) 'Defective lipoprotein receptors and atherosclerosis. Lessons from an animal counterpart of familial hypercholesterolemia', *N Engl J Med,* 309(5), 288-96.

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', *The Journal of clinical investigation,* 52(7), 1544-68.

Graham, C. A., McClean, E., Ward, A. J., Beattie, E. D., Martin, S., O'Kane, M., Young, I. S. and Nicholls, D. P. (1999) 'Mutation screening and genotype:phenotype correlation in familial hypercholesterolaemia', *Atherosclerosis,* 147(2), 309-16.

Graham, C. A., McIlhatton, B. P., Kirk, C. W., Beattie, E. D., Lyttle, K., Hart, P., Neely, R. D., Young, I. S. and Nicholls, D. P. (2005) 'Genetic screening protocol for familial hypercholesterolemia which includes splicing defects gives an improved mutation detection rate', *Atherosclerosis,* 182(2), 331-40.

Graham, M. J., Lemonidis, K. M., Whipple, C. P., Subramaniam, A., Monia, B. P., Crooke, S. T. and Crooke, R. M. (2007) 'Antisense inhibition of

proprotein convertase subtilisin/kexin type 9 reduces serum LDL in hyperlipidemic mice', *J Lipid Res,* 48(4), 763-7.

Green, R. C., Berg, J. S., Grody, W. W., Kalia, S. S., Korf, B. R., Martin, C. L., McGuire, A. L., Nussbaum, R. L., O'Daniel, J. M., Ormond, K. E., Rehm, H. L., Watson, M. S., Williams, M. S. and Biesecker, L. G. (2013) 'ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing', *Genet Med,* 15(7), 565-74.

Gudnason, V., Sigurdsson, G., Nissen, H. and Humphries, S. E. (1997) 'Common founder mutation in the LDL receptor gene causing familial hypercholesterolaemia in the Icelandic population', *Hum Mutat,* 10(1), 36-44.

Gupta, N., Fisker, N., Asselin, M. C., Lindholm, M., Rosenbohm, C., Ørum, H., Elmén, J., Seidah, N. G. and Straarup, E. M. (2010) 'A locked nucleic acid antisense oligonucleotide (LNA) silences PCSK9 and enhances LDLR expression in vitro and in vivo', *PLoS One,* 5(5), e10682.

Hadfield, S. G., Horara, S., Starr, B. J., Yazdgerdi, S., Bhatnagar, D., Cramb, R., Egan, S., Everdell, R., Ferns, G., Jones, A., Marenah, C. B., Marples, J., Prinsloo, P., Sneyd, A., Stewart, M. F., Sandle, L., Wang, T., Watson, M. S. and Humphries, S. E. (2008) 'Are patients with familial hypercholesterolaemia well managed in lipid clinics? An audit of eleven clinics from the Department of Health Familial Hypercholesterolaemia Cascade Testing project', *Ann Clin Biochem,* 45(Pt 2), 199-205.

Hattori, H., Hirayama, T., Nobe, Y., Nagano, M., Kujiraoka, T., Egashira, T., Ishii, J., Tsuji, M. and Emi, M. (2002) 'Eight novel mutations and functional impairments of the LDL receptor in familial hypercholesterolemia in the north of Japan', *J Hum Genet,* 47(2), 80-7.

Heath, K. E., Humphries, S. E., Middleton-Price, H. and Boxer, M. (2001) 'A molecular genetic service for diagnosing individuals with familial hypercholesterolaemia (FH) in the United Kingdom', *European journal of human genetics : EJHG,* 9(4), 244-52.

Hirayama, T., Yamaki, E., Hata, A., Tsuji, M., Hashimoto, K., Yamamoto, M. and Emi, M. (1998) 'Five familial hypercholesterolemic kindreds in Japan with novel mutations of the LDL receptor gene', *J Hum Genet,* 43(4), 250-4.

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', *Annual review of genetics,* 24, 133-70.

Holmes, R. S., Vandeberg, J. L. and Cox, L. A. (2011) 'Genomics and proteomics of vertebrate cholesterol ester lipase (LIPA) and cholesterol 25-hydroxylase (CH25H)', *3 Biotech,* 1(2), 99-109.

Hooper, A. J., Marais, A. D., Tanyanyiwa, D. M. and Burnett, J. R. (2007) 'The C679X mutation in PCSK9 is present and lowers blood cholesterol in a Southern African population', *Atherosclerosis,* 193(2), 445-8.

Hooper, A. J., Nguyen, L. T., Burnett, J. R., Bates, T. R., Bell, D. A., Redgrave, T. G., Watts, G. F. and van Bockxmeer, F. M. (2012) 'Genetic analysis of familial hypercholesterolaemia in Western Australia', *Atherosclerosis,* 224(2), 430-4.

Hostetler, H. A., Kier, A. B. and Schroeder, F. (2006) 'Very-long-chain and branched-chain fatty acyl-CoAs are high affinity ligands for the peroxisome proliferator-activated receptor alpha (PPARalpha)', *Biochemistry,* 45(24), 7669-81.

Hostetler, H. A., Petrescu, A. D., Kier, A. B. and Schroeder, F. (2005) 'Peroxisome proliferator-activated receptor alpha interacts with high affinity and is conformationally responsive to endogenous ligands', *J Biol Chem,* 280(19), 18667-82.

Huijgen, R., Sjouke, B., Vis, K., de Randamie, J. S., Defesche, J. C., Kastelein, J. J., Hovingh, G. K. and Fouchier, S. W. (2012) 'Genetic variation in APOB, PCSK9, and ANGPTL3 in carriers of pathogenic autosomal dominant hypercholesterolemic mutations with unexpected low LDL-Cl Levels', *Hum Mutat,* 33(2), 448-55.

Humphries, S. E., Cranston, T., Allen, M., Middleton-Price, H., Fernandez, M. C., Senior, V., Hawe, E., Iversen, A., Wray, R., Crook, M. A. and Wierzbicki, A. S. (2006a) 'Mutational analysis in UK patients with a clinical diagnosis of familial hypercholesterolaemia: relationship with plasma lipid traits, heart disease risk and utility in relative tracing', *Journal of molecular medicine,* 84(3), 203-14.

Humphries, S. E., Norbury, G., Leigh, S., Hadfield, S. G. and Nair, D. (2008) 'What is the clinical utility of DNA testing in patients with familial hypercholesterolaemia?', *Curr Opin Lipidol,* 19(4), 362-8.

Humphries, S. E., Whittall, R. A., Hubbart, C. S., Maplebeck, S., Cooper, J. A., Soutar, A. K., Naoumova, R., Thompson, G. R., Seed, M., Durrington, P. N., Miller, J. P., Betteridge, D. J. and Neil, H. A. (2006b) 'Genetic causes of familial hypercholesterolaemia in patients in the UK: relation to plasma lipid levels and coronary heart disease risk', *J Med Genet,* 43(12), 943-9.

Humphries, S. E., Whittall, R. A., Hubbart, C. S., Maplebeck, S., Cooper, J. A., Soutar, A. K., Naoumova, R., Thompson, G. R., Seed, M., Durrington, P. N., Miller, J. P., Betteridge, D. J. and Neil, H. A. (2006c)

'Genetic causes of familial hypercholesterolaemia in patients in the UK: relation to plasma lipid levels and coronary heart disease risk', *J Med Genet,* 43(12), 943-9.

Hurst, J. M., McMillan, L. E., Porter, C. T., Allen, J., Fakorede, A. and Martin, A. C. (2009) 'The SAAPdb web resource: a large-scale structural analysis of mutant proteins', *Hum Mutat,* 30(4), 616-24.

Ikonen, E. (2008) 'Cellular cholesterol trafficking and compartmentalization', *Nat Rev Mol Cell Biol,* 9(2), 125-38.

Innerarity, T. L., Weisgraber, K. H., Arnold, K. S., Mahley, R. W., Krauss, R. M., Vega, G. L. and Grundy, S. M. (1987) 'Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding', *Proceedings of the National Academy of Sciences of the United States of America,* 84(19), 6919-23.

Jelassi, A., Jguirim, I., Najah, M., Abid, A. M., Boughamoura, L., Maatouk, F., Rouis, M., Boileau, C., Rabes, J. P., Slimane, M. N. and Varret, M. (2009) 'Limited mutational heterogeneity in the LDLR gene in familial hypercholesterolemia in Tunisia', *Atherosclerosis,* 203(2), 449-53.

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., Voight, B. F., Elosua, R., Salomaa, V., O'Donnell, C. J., Dallinga-Thie, G. M., Anand, S. S., Yusuf, S., Huff, M. W., Kathiresan, S., Cao, H. and Hegele, R. A. (2011) 'An increased burden of common and rare lipid-associated risk alleles contributes to the phenotypic spectrum of hypertriglyceridemia', *Arteriosclerosis, thrombosis, and vascular biology,* 31(8), 1916-26.

Kandutsch, A. A., Chen, H. W. and Heiniger, H. J. (1978) 'Biological activity of some oxygenated sterols', *Science,* 201(4355), 498-501.

Kandutsch, A. A., Chen, H. W. and Shown, E. P. (1977) 'Binding of 25-hydroxycholesterol and cholesterol to different cytoplasmic proteins', *Proc Natl Acad Sci U S A,* 74(6), 2500-3.

Kannel, W. B. (2005) 'Overview of hemostatic factors involved in atherosclerotic cardiovascular disease', *Lipids,* 40(12), 1215-20.

Khachadurian, A. K. and Uthman, S. M. (1973) 'Experiences with the homozygous cases of familial hypercholesterolemia. A report of 52 patients', *Nutr Metab,* 15(1), 132-40.

Khamis A, P. J., Lench N, Taylor A, Leigh S, Humphries SE (2012) *Analysis of four LDLR 5'UTR and promoter variants in patients with familial hypercholesterolaemia*, unpublished.

Kishore, S., Khanna, A. and Stamm, S. (2008) 'Rapid generation of splicing reporters with pSpliceExpress', *Gene,* 427(1-2), 104-10.

Koren, M. J., Scott, R., Kim, J. B., Knusel, B., Liu, T., Lei, L., Bolognese, M. and Wasserman, S. M. (2012) 'Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 as monotherapy in patients with hypercholesterolaemia (MENDEL): a randomised, double-blind, placebo-controlled, phase 2 study', *Lancet,* 380(9858), 1995-2006.

Kotze, M. J., Loubser, O., Thiart, R., de Villiers, J. N., Langenhoven, E., Theart, L., Steyn, K., Marais, A. D. and Raal, F. J. (1997) 'CpG hotspot mutations at the LDL receptor locus are a frequent cause of familial hypercholesterolaemia among South African Indians', *Clin Genet,* 51(6), 394-8.

Kwon, H. J., Lagace, T. A., McNutt, M. C., Horton, J. D. and Deisenhofer, J. (2008) 'Molecular basis for LDL receptor recognition by PCSK9', *Proceedings of the National Academy of Sciences of the United States of America,* 105(6), 1820-5.

Lee, W. K., Haddad, L., Macleod, M. J., Dorrance, A. M., Wilson, D. J., Gaffney, D., Dominiczak, M. H., Packard, C. J., Day, I. N., Humphries, S. E. and Dominiczak, A. F. (1998) 'Identification of a common low density lipoprotein receptor mutation (C163Y) in the west of Scotland', *J Med Genet,* 35(7), 573-8.

Lees, J., Yeats, C., Redfern, O., Clegg, A. and Orengo, C. (2010) 'Gene3D: merging structure and function for a Thousand genomes', *Nucleic Acids Res,* 38(Database issue), D296-300.

Lehrman, M. A., Goldstein, J. L., Russell, D. W. and Brown, M. S. (1987a) 'Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia', *Cell,* 48(5), 827-35.

Lehrman, M. A., Russell, D. W., Goldstein, J. L. and Brown, M. S. (1987b) 'Alu-Alu recombination deletes splice acceptor sites and produces secreted low density lipoprotein receptor in a subject with familial hypercholesterolemia', *J Biol Chem,* 262(7), 3354-61.

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', *Annals of Human Genetics,* 72(Pt 4), 485-98.

Leigh, S. E., Leren, T. P. and Humphries, S. E. (2009) 'Commentary PCSK9 variants: A new database', *Atherosclerosis,* 203(1), 32-3.

Leitersdorf, E., Van der Westhuyzen, D. R., Coetzee, G. A. and Hobbs, H. H. (1989) 'Two common low density lipoprotein receptor gene mutations cause familial hypercholesterolemia in Afrikaners', *J Clin Invest,* 84(3), 954-61.

Leren, T. P., Manshaus, T., Skovholt, U., Skodje, T., Nossen, I. E., Teie, C., Sørensen, S. and Bakken, K. S. (2004) 'Application of molecular genetics for diagnosing familial hypercholesterolemia in Norway: results from a family-based screening program', *Semin Vasc Med,* 4(1), 75-85.

Libby, P., Ridker, P. M. and Hansson, G. K. (2011) 'Progress and challenges in translating the biology of atherosclerosis', *Nature,* 473(7347), 317-25.

Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E. and Wittwer, C. (2004) 'Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons', *Clin Chem,* 50(7), 1156-64.

Lin, Q., Ruuska, S. E., Shaw, N. S., Dong, D. and Noy, N. (1999) 'Ligand selectivity of the peroxisome proliferator-activated receptor alpha', *Biochemistry,* 38(1), 185-90.

Liyanage, K. E., Burnett, J. R., Hooper, A. J. and van Bockxmeer, F. M. (2011) 'Familial hypercholesterolemia: epidemiology, Neolithic origins and modern geographic distribution', *Critical reviews in clinical laboratory sciences,* 48(1), 1-18.

Ludwig, E. H. and McCarthy, B. J. (1990) 'Haplotype analysis of the human apolipoprotein B mutation associated with familial defective apolipoprotein B100', *Am J Hum Genet,* 47(4), 712-20.

Lund, E. G., Kerr, T. A., Sakai, J., Li, W. P. and Russell, D. W. (1998) 'cDNA cloning of mouse and human cholesterol 25-hydroxylases, polytopic membrane proteins that synthesize a potent oxysterol regulator of lipid metabolism', *J Biol Chem,* 273(51), 34316-27.

Mabuchi, H., Haba, T., Ueda, K., Ueda, R., Tatami, R., Ito, S., Kametani, T., Koizumi, J., Miyamoto, S., Ohta, M., Takeda, R., Takegoshi, T. and Takeshita, H. (1977) 'Serum lipids and coronary heart disease in heterozygous familial hypercholesterolemia in the Hokuriku District of Japan', *Atherosclerosis,* 28(4), 417-23.

Mamotte, C. D. and van Bockxmeer, F. M. (1993) 'A robust strategy for screening and confirmation of familial defective apolipoprotein B-100', *Clin Chem,* 39(1), 118-21.

'A map of human genome variation from population-scale sequencing', (2010) *Nature,* 467(7319), 1061-73.

Marks, D., Thorogood, M., Farrer, J. M. and Humphries, S. E. (2004) 'Census of clinics providing specialist lipid services in the United Kingdom', *J Public Health (Oxf),* 26(4), 353-4.

Marks, D., Thorogood, M., Neil, H. A. and Humphries, S. E. (2003) 'A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia', *Atherosclerosis,* 168(1), 1-14.

Marmot, M. G., Smith, G. D., Stansfeld, S., Patel, C., North, F., Head, J., White, I., Brunner, E. and Feeney, A. (1991) 'Health inequalities among British civil servants: the Whitehall II study', *Lancet,* 337(8754), 1387-93.

Marques-Pinheiro, A., Marduel, M., Rabes, J. P., Devillers, M., Villeger, L., Allard, D., Weissenbach, J., Guerin, M., Zair, Y., Erlich, D., Junien, C., Munnich, A., Krempf, M., Abifadel, M., Jais, J. P., Boileau, C. and Varret, M. (2010) 'A fourth locus for autosomal dominant hypercholesterolemia maps at 16q22.1', *European journal of human genetics : EJHG,* 18(11), 1236-42.

McEwan, I. J. (2009) 'Nuclear receptors: one big family', *Methods Mol Biol,* 505, 3-18.

Michaely, P., Li, W. P., Anderson, R. G., Cohen, J. C. and Hobbs, H. H. (2004) 'The modular adaptor protein ARH is required for low density lipoprotein (LDL) binding and internalization but not for LDL receptor clustering in coated pits', *J Biol Chem,* 279(32), 34023-31.

Miller, S. A., Dykes, D. D. and Polesky, H. F. (1988) 'A simple salting out procedure for extracting DNA from human nucleated cells', *Nucleic acids research,* 16(3), 1215.

Miltiadous, G., Elisaf, M., Bairaktari, H., Xenophontos, S. L., Manoli, P. and Cariolou, M. A. (2001) 'Characterization and geographic distribution of the low density lipoprotein receptor (LDLR) gene mutations in northwestern Greece', *Hum Mutat,* 17(5), 432-3.

Mohan, V., Deepa, R., Rani, S. S., Premalatha, G. and No.5), C. U. P. S. C. (2001) 'Prevalence of coronary artery disease and its relationship to lipids in a selected population in South India: The Chennai Urban Population Study (CUPS No. 5)', *J Am Coll Cardiol,* 38(3), 682-7.

Morrone, D., Weintraub, W. S., Toth, P. P., Hanson, M. E., Lowe, R. S., Lin, J., Shah, A. K. and Tershakovec, A. M. (2012) 'Lipid-altering efficacy of ezetimibe plus statin and statin monotherapy and identification of factors associated with treatment response: a pooled analysis of over 21,000 subjects from 27 clinical trials', *Atherosclerosis,* 223(2), 251-61.

Motazacker, M. M., Pirruccello, J., Huijgen, R., Do, R., Gabriel, S., Peter, J., Kuivenhoven, J. A., Defesche, J. C., Kastelein, J. J., Hovingh, G. K.,

Zelcer, N., Kathiresan, S. and Fouchier, S. W. (2012) 'Advances in genetics show the need for extending screening strategies for autosomal dominant hypercholesterolaemia', *European heart journal*.

Mozas, P., Cenarro, A., Civeira, F., Castillo, S., Ros, E. and Pocovi, M. (2000) 'Mutation analysis in 36 unrelated Spanish subjects with familial hypercholesterolemia: identification of 3 novel mutations in the LDL receptor gene', *Hum Mutat,* 15(5), 483-4.

Myant, N. B., Forbes, S. A., Day, I. N. and Gallagher, J. (1997) 'Estimation of the age of the ancestral arginine3500-->glutamine mutation in human apoB-100', *Genomics,* 45(1), 78-87.

Naoumova, R. P., Tosi, I., Patel, D., Neuwirth, C., Horswell, S. D., Marais, A. D., van Heyningen, C. and Soutar, A. K. (2005) 'Severe hypercholesterolemia in four British families with the D374Y mutation in the PCSK9 gene: long-term follow-up and treatment response', *Arterioscler Thromb Vasc Biol,* 25(12), 2654-60.

Neil, A., Cooper, J., Betteridge, J., Capps, N., McDowell, I., Durrington, P., Seed, M. and Humphries, S. E. (2008) 'Reductions in all-cause, cancer, and coronary mortality in statin-treated patients with heterozygous familial hypercholesterolaemia: a prospective registry study', *European heart journal,* 29(21), 2625-33.

Neil, H. A., Hammond, T., Huxley, R., Matthews, D. R. and Humphries, S. E. (2000) 'Extent of underdiagnosis of familial hypercholesterolaemia in routine practice: prospective registry study', *BMJ,* 321(7254), 148.

Neil, H. A., Seagroatt, V., Betteridge, D. J., Cooper, M. P., Durrington, P. N., Miller, J. P., Seed, M., Naoumova, R. P., Thompson, G. R., Huxley, R. and Humphries, S. E. (2004) 'Established and emerging coronary risk factors in patients with heterozygous familial hypercholesterolaemia', *Heart,* 90(12), 1431-7.

Ng, P. C. and Henikoff, S. (2003) 'SIFT: Predicting amino acid changes that affect protein function', *Nucleic acids research,* 31(13), 3812-4.

Nherera, L., Marks, D., Minhas, R., Thorogood, M. and Humphries, S. E. (2011) 'Probabilistic cost-effectiveness analysis of cascade screening for familial hypercholesterolaemia using alternative diagnostic and identification strategies', *Heart,* 97(14), 1175-81.

Nordestgaard, B. G., Chapman, M. J., Humphries, S. E., Ginsberg, H. N., Masana, L., Descamps, O. S., Wiklund, O., Hegele, R. A., Raal, F. J., Defesche, J. C., Wiegman, A., Santos, R. D., Watts, G. F., Parhofer, K. G., Kees Hovingh, G., Kovanen, P. T., Boileau, C., Averna, M., Borén, J., Bruckert, E., Catapano, A. L., Kuivenhoven, J. A., Pajukanta, P., Ray, K., Stalenhoef, A. F., Stroes, E., Taskinen, M. R., Tybjærg-Hansen, A. and Panel, f. t. E. A. S. C. (2013) 'Familial

hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: Consensus Statement of the European Atherosclerosis Society', *Eur Heart J*.

Oram, J. F. and Lawn, R. M. (2001) 'ABCA1. The gatekeeper for eliminating excess tissue cholesterol', *J Lipid Res,* 42(8), 1173-9.

Oram, J. F. and Yokoyama, S. (1996) 'Apolipoprotein-mediated removal of cellular cholesterol and phospholipids', *J Lipid Res,* 37(12), 2473-91.

Patel, A. Y., Pillarisetti, J., Marr, J. and Vacek, J. L. (2013) 'Ezetimibe in combination with a statin does not reduce all-cause mortality', *J Clin Med Res,* 5(4), 275-80.

Patrick, A. D. and Lake, B. D. (1969) 'Deficiency of an acid lipase in Wolman's disease', *Nature,* 222(5198), 1067-8.

Plagnol, V., Curtis, J., Epstein, M., Mok, K., Stebbings, E., Grigoriadou, S., Wood, N. W., Hambleton, S., Burn, S. O., Thrasher, A., Kumararatne, D., Doffinger, R. and Nejentsev, S. (2012) 'A robust model for read count data in exome sequencing experiments and implications for copy number variant calling', *Bioinformatics*.

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(3), 1225-34.

Qian, Y. W., Schmidt, R. J., Zhang, Y., Chu, S., Lin, A., Wang, H., Wang, X., Beyer, T. P., Bensch, W. R., Li, W., Ehsani, M. E., Lu, D., Konrad, R. J., Eacho, P. I., Moller, D. E., Karathanasis, S. K. and Cao, G. (2007) 'Secreted PCSK9 downregulates low density lipoprotein receptor through receptor-mediated endocytosis', *J Lipid Res,* 48(7), 1488-98.

Raal, F., Scott, R., Somaratne, R., Bridges, I., Li, G., Wasserman, S. M. and Stein, E. A. (2012) 'Low-Density Lipoprotein Cholesterol-Lowering Effects of AMG 145, a Monoclonal Antibody to Proprotein Convertase Subtilisin/Kexin Type 9 Serine Protease in Patients With Heterozygous Familial Hypercholesterolemia: The Reduction of LDL-C With PCSK9 Inhibition in Heterozygous Familial Hypercholesterolemia Disorder (RUTHERFORD) Randomized Trial', *Circulation,* 126(20), 2408-17.

Rabès, J. P., Varret, M., Devillers, M., Aegerter, P., Villéger, L., Krempf, M., Junien, C. and Boileau, C. (2000) 'R3531C mutation in the apolipoprotein B gene is not sufficient to cause hypercholesterolemia', *Arterioscler Thromb Vasc Biol,* 20(10), E76-82.

Rabès, J. P., Varret, M., Saint-Jore, B., Erlich, D., Jondeau, G., Krempf, M., Giraudet, P., Junien, C. and Boileau, C. (1997) 'Familial ligand-defective apolipoprotein B-100: simultaneous detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a French population', *Hum Mutat,* 10(2), 160-3.

Radhakrishnan, A., Ikeda, Y., Kwon, H. J., Brown, M. S. and Goldstein, J. L. (2007) 'Sterol-regulated transport of SREBPs from endoplasmic reticulum to Golgi: oxysterols block transport by binding to Insig', *Proceedings of the National Academy of Sciences of the United States of America,* 104(16), 6511-8.

Reese, M. G., Eeckman, F. H., Kulp, D. and Haussler, D. (1997) 'Improved splice site detection in Genie', *J Comput Biol,* 4(3), 311-23.

'Risk of fatal coronary heart disease in familial hypercholesterolaemia. Scientific Steering Committee on behalf of the Simon Broome Register Group', (1991) *BMJ,* 303(6807), 893-6.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J. P. (2011) 'Integrative genomics viewer', *Nature biotechnology,* 29(1), 24-6.

Rodgers, G. M. and Conn, M. T. (1990) 'Homocysteine, an atherogenic stimulus, reduces protein C activation by arterial and venous endothelial cells', *Blood,* 75(4), 895-901.

Roheim, P. S. (1986) 'Atherosclerosis and lipoprotein metabolism: role of reverse cholesterol transport', *Am J Cardiol,* 57(5), 3C-10C.

Rohlmann, A., Gotthardt, M., Hammer, R. E. and Herz, J. (1998) 'Inducible inactivation of hepatic LRP gene by cre-mediated recombination confirms role of LRP in clearance of chylomicron remnants', *J Clin Invest,* 101(3), 689-95.

Ross, R. (1999) 'Atherosclerosis--an inflammatory disease', *N Engl J Med,* 340(2), 115-26.

Rubinsztein, D. C., Jialal, I., Leitersdorf, E., Coetzee, G. A. and van der Westhuyzen, D. R. (1993) 'Identification of two new LDL-receptor mutations causing homozygous familial hypercholesterolemia in a South African of Indian origin', *Biochim Biophys Acta,* 1182(1), 75-82.

Schwarz, J. M., Rodelsperger, C., Schuelke, M. and Seelow, D. (2010) 'MutationTaster evaluates disease-causing potential of sequence alterations', *Nature methods,* 7(8), 575-6.

Seftel, H. C., Baker, S. G., Sandler, M. P., Forman, M. B., Joffe, B. I., Mendelsohn, D., Jenkins, T. and Mieny, C. J. (1980) 'A host of

hypercholesterolaemic homozygotes in South Africa', *British medical journal,* 281(6241), 633-6.

Setia, N., Verma, I. C., Khan, B. and Arora, A. (2012) 'Premature coronary artery disease and familial hypercholesterolemia: need for early diagnosis and cascade screening in the Indian population', *Cardiol Res Pract,* 2012, 658526.

Shan, L., Pang, L., Zhang, R., Murgolo, N. J., Lan, H. and Hedrick, J. A. (2008) 'PCSK9 binds to multiple receptors and can be functionally inhibited by an EGF-A peptide', *Biochem Biophys Res Commun,* 375(1), 69-73.

Shao, W. and Espenshade, P. J. (2012) 'Expanding roles for SREBP in metabolism', *Cell Metab,* 16(4), 414-9.

Sharma, P., Boyers, D., Boachie, C., Stewart, F., Miedzybrodzka, Z., Simpson, W., Kilonzo, M., McNamee, P. and Mowatt, G. (2012) 'Elucigene FH20 and LIPOchip for the diagnosis of familial hypercholesterolaemia: a systematic review and economic evaluation', *Health Technol Assess,* 16(17), 1-266.

Slack, J. (1969) 'Risks of ischaemic heart-disease in familial hyperlipoproteinaemic states', *Lancet,* 2(7635), 1380-2.

Soria, L. F., Ludwig, E. H., Clarke, H. R., Vega, G. L., Grundy, S. M. and McCarthy, B. J. (1989) 'Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100', *Proceedings of the National Academy of Sciences of the United States of America,* 86(2), 587-91.

Soutar, A. K., Knight, B. L. and Patel, D. D. (1989) 'Identification of a point mutation in growth factor repeat C of the low density lipoprotein-receptor gene in a patient with homozygous familial hypercholesterolemia that affects ligand binding and intracellular movement of receptors', *Proc Natl Acad Sci U S A,* 86(11), 4166-70.

Soutar, A. K. and Naoumova, R. P. (2007) 'Mechanisms of disease: genetic causes of familial hypercholesterolemia', *Nature clinical practice. Cardiovascular medicine,* 4(4), 214-25.

Soutar, A. K., Naoumova, R. P. and Traub, L. M. (2003) 'Genetics, clinical phenotype, and molecular cell biology of autosomal recessive hypercholesterolemia', *Arterioscler Thromb Vasc Biol,* 23(11), 1963-70.

Starr, B., Hadfield, S. G., Hutten, B. A., Lansberg, P. J., Leren, T. P., Damgaard, D., Neil, H. A. and Humphries, S. E. (2008) 'Development of sensitive and specific age- and gender-specific low-density lipoprotein cholesterol cutoffs for diagnosis of first-degree relatives with

familial hypercholesterolaemia in cascade testing', *Clinical chemistry and laboratory medicine : CCLM / FESCC,* 46(6), 791-803.

Stein, E. A., Dufour, R., Gagne, C., Gaudet, D., East, C., Donovan, J. M., Chin, W., Tribble, D. L. and McGowan, M. (2012a) 'Apolipoprotein B synthesis inhibition with mipomersen in heterozygous familial hypercholesterolemia: results of a randomized, double-blind, placebo-controlled trial to assess efficacy and safety as add-on therapy in patients with coronary artery disease', *Circulation,* 126(19), 2283-92.

Stein, E. A., Mellis, S., Yancopoulos, G. D., Stahl, N., Logan, D., Smith, W. B., Lisbon, E., Gutierrez, M., Webb, C., Wu, R., Du, Y., Kranz, T., Gasparino, E. and Swergold, G. D. (2012b) 'Effect of a monoclonal antibody to PCSK9 on LDL cholesterol', *N Engl J Med,* 366(12), 1108-18.

Sterling, R., Henderson, G. E. and Corbie-Smith, G. (2006) 'Public willingness to participate in and public opinions about genetic variation research: a review of the literature', *Am J Public Health,* 96(11), 1971-8.

Stitziel, N. O., Fouchier, S. W., Sjouke, B., Peloso, G. M., Moscoso, A. M., Auer, P. L., Goel, A., Gigante, B., Barnes, T. A., Melander, O., Orho-Melander, M., Duga, S., Sivapalaratnam, S., Nikpay, M., Martinelli, N., Girelli, D., Jackson, R. D., Kooperberg, C., Lange, L. A., Ardissino, D., McPherson, R., Farrall, M., Watkins, H., Reilly, M. P., Rader, D. J., de Faire, U., Schunkert, H., Erdmann, J., Samani, N. J., Charnas, L., Altshuler, D., Gabriel, S., Kastelein, J. J., Defesche, J. C., Nederveen, A. J., Kathiresan, S. and Hovingh, G. K. (2013) 'Exome sequencing and directed clinical phenotyping diagnose cholesterol ester storage disease presenting as autosomal recessive hypercholesterolemia', *Arterioscler Thromb Vasc Biol,* 33(12), 2909-14.

Sudhof, T. C., Goldstein, J. L., Brown, M. S. and Russell, D. W. (1985) 'The LDL receptor gene: a mosaic of exons shared with different proteins', *Science,* 228(4701), 815-22.

Sun, X. M., Eden, E. R., Tosi, I., Neuwirth, C. K., Wile, D., Naoumova, R. P. and Soutar, A. K. (2005) 'Evidence for effect of mutant PCSK9 on apolipoprotein B secretion as the cause of unusually severe dominant hypercholesterolaemia', *Human molecular genetics,* 14(9), 1161-9.

Talmud, P. J., Drenos, F., Shah, S., Shah, T., Palmen, J., Verzilli, C., Gaunt, T. R., Pallas, J., Lovering, R., Li, K., Casas, J. P., Sofat, R., Kumari, M., Rodriguez, S., Johnson, T., Newhouse, S. J., Dominiczak, A., Samani, N. J., Caulfield, M., Sever, P., Stanton, A., Shields, D. C., Padmanabhan, S., Melander, O., Hastie, C., Delles, C., Ebrahim, S., Marmot, M. G., Smith, G. D., Lawlor, D. A., Munroe, P. B., Day, I. N., Kivimaki, M., Whittaker, J., Humphries, S. E. and Hingorani, A. D. (2009) 'Gene-centric association signals for lipids and apolipoproteins

identified via the HumanCVD BeadChip', *Am J Hum Genet,* 85(5), 628-42.

Talmud, P. J., Shah, S., Whittall, R., Futema, M., Howard, P., Cooper, J. A., Harrison, S. C., Li, K., Drenos, F., Karpe, F., Neil, H. A., Descamps, O. S., Langenberg, C., Lench, N., Kivimaki, M., Whittaker, J., Hingorani, A. D., Kumari, M. and Humphries, S. E. (2013) 'Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study', *Lancet*.

Taylor, A., Martin, B., Wang, D., Patel, K., Humphries, S. E. and Norbury, G. (2009) 'Multiplex ligation-dependent probe amplification analysis to screen for deletions and duplications of the LDLR gene in patients with familial hypercholesterolaemia', *Clinical genetics,* 76(1), 69-75.

Taylor, A., Patel, K., Tsedeke, J., Humphries, S. E. and Norbury, G. (2010a) 'Mutation screening in patients for familial hypercholesterolaemia (ADH)', *Clinical genetics,* 77(1), 97-9.

Taylor, A., Wang, D., Patel, K., Whittall, R., Wood, G., Farrer, M., Neely, R. D., Fairgrieve, S., Nair, D., Barbir, M., Jones, J. L., Egan, S., Everdale, R., Lolin, Y., Hughes, E., Cooper, J. A., Hadfield, S. G., Norbury, G. and Humphries, S. E. (2010b) 'Mutation detection rate and spectrum in familial hypercholesterolaemia patients in the UK pilot cascade project', *Clinical genetics,* 77(6), 572-80.

Teer, J. K. and Mullikin, J. C. (2010) 'Exome sequencing: the sweet spot before whole genomes', *Human molecular genetics,* 19(R2), R145-51.

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., Johansen, C. T., Fouchier, S. W., Isaacs, A., Peloso, G. M., Barbalic, M., Ricketts, S. L., Bis, J. C., Aulchenko, Y. S., Thorleifsson, G., Feitosa, M. F., Chambers, J., Orho-Melander, M., Melander, O., Johnson, T., Li, X., Guo, X., Li, M., Shin Cho, Y., Jin Go, M., Jin Kim, Y., Lee, J. Y., Park, T., Kim, K., Sim, X., Twee-Hee Ong, R., Croteau-Chonka, D. C., Lange, L. A., Smith, J. D., Song, K., Hua Zhao, J., Yuan, X., Luan, J., Lamina, C., Ziegler, A., Zhang, W., Zee, R. Y., Wright, A. F., Witteman, J. C., Wilson, J. F., Willemsen, G., Wichmann, H. E., Whitfield, J. B., Waterworth, D. M., Wareham, N. J., Waeber, G., Vollenweider, P., Voight, B. F., Vitart, V., Uitterlinden, A. G., Uda, M., Tuomilehto, J., Thompson, J. R., Tanaka, T., Surakka, I., Stringham, H. M., Spector, T. D., Soranzo, N., Smit, J. H., Sinisalo, J., Silander, K., Sijbrands, E. J., Scuteri, A., Scott, J., Schlessinger, D., Sanna, S., Salomaa, V., Saharinen, J., Sabatti, C., Ruokonen, A., Rudan, I., Rose, L. M., Roberts, R., Rieder, M., Psaty, B. M., Pramstaller, P. P., Pichler, I., Perola, M., Penninx, B. W., Pedersen, N. L., Pattaro, C., Parker, A. N., Pare, G., Oostra, B. A., O'Donnell, C. J., Nieminen, M. S., Nickerson, D. A., Montgomery, G. W., Meitinger, T., McPherson, R.,

McCarthy, M. I., ., et al. (2010) 'Biological, clinical and population relevance of 95 loci for blood lipids', *Nature,* 466(7307), 707-13.

Thomas, E. R. A., Atanur, S. S., Norsworthy, P. J., Encheva, V., Snijders, A. P., Game, L., Vandrovcova, J., Siddiq, A., Seed, M., Soutar, A. K. and Aitman, T. J. 'Identification and biochemical analysis of a novel APOB mutation that causes autosomal dominant hypercholesterolemia'.

Thorvaldsdottir, H., Robinson, J. T. and Mesirov, J. P. (2013) 'Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration', *Brief Bioinform,* 14(2), 178-92.

Toth, P. P. and Davidson, M. H. (2005) 'Cholesterol absorption blockade with ezetimibe', *Curr Drug Targets Cardiovasc Haematol Disord,* 5(6), 455-62.

Tybjaerg-Hansen, A. and Humphries, S. E. (1992) 'Familial defective apolipoprotein B-100: a single mutation that causes hypercholesterolemia and premature coronary artery disease', *Atherosclerosis,* 96(2-3), 91-107.

Umans-Eckenhausen, M. A., Defesche, J. C., Sijbrands, E. J., Scheerder, R. L. and Kastelein, J. J. (2001) 'Review of first 5 years of screening for familial hypercholesterolaemia in the Netherlands', *Lancet,* 357(9251), 165-8.

Usifo, E., Leigh, S. E., Whittall, R. A., Lench, N., Taylor, A., Yeats, C., Orengo, C. A., Martin, A. C., Celli, J. and Humphries, S. E. (2012) 'Low-density lipoprotein receptor gene familial hypercholesterolemia variant database: update and pathological assessment', *Ann Hum Genet,* 76(5), 387-401.

Usifo, E. L., S.E.A.; Whittall, R.A.; Lench, N.; Taylor, A.; Yeates, C.; Orengo, C.A.; Martin, A.C.R.; and Humphries, S.E. (2012) 'Low Density Lipoprotein Receptor Gene Familial Hypercholesterolemia Variant Database: update and pathological assessment', *Annals of Human Genetics*.

Van Gaal, L. F., Peeters, A. V., De Block, C. E., de Leeuw, I. H., Thiart, R. and Kotze, M. J. (2001) 'Low-density lipoprotein receptor gene mutation analysis and clinical correlation in Belgian hypercholesterolaemics', *Mol Cell Probes,* 15(6), 329-36.

Varret, M., Abifadel, M., Rabès, J. P. and Boileau, C. (2008) 'Genetic heterogeneity of autosomal dominant hypercholesterolemia', *Clin Genet,* 73(1), 1-13.

Venselaar, H., Te Beek, T. A., Kuipers, R. K., Hekkelman, M. L. and Vriend, G. (2010) 'Protein structure analysis of mutations causing inheritable

diseases. An e-Science approach with life scientist friendly interfaces', *BMC bioinformatics,* 11, 548.

Versmissen, J., Oosterveer, D. M., Yazdanpanah, M., Defesche, J. C., Basart, D. C., Liem, A. H., Heeringa, J., Witteman, J. C., Lansberg, P. J., Kastelein, J. J. and Sijbrands, E. J. (2008) 'Efficacy of statins in familial hypercholesterolaemia: a long term cohort study', *BMJ,* 337, a2423.

Vibe-Pedersen, K., Kornblihtt, A. R. and Baralle, F. E. (1984) 'Expression of a human alpha-globin/fibronectin gene hybrid generates two mRNAs by alternative splicing', *EMBO J,* 3(11), 2511-6.

Visser, M. E., Wagener, G., Baker, B. F., Geary, R. S., Donovan, J. M., Beuers, U. H., Nederveen, A. J., Verheij, J., Trip, M. D., Basart, D. C., Kastelein, J. J. and Stroes, E. S. (2012) 'Mipomersen, an apolipoprotein B synthesis inhibitor, lowers low-density lipoprotein cholesterol in high-risk statin-intolerant patients: a randomized, double-blind, placebo-controlled trial', *Eur Heart J,* 33(9), 1142-9.

Voight, B. F., Peloso, G. M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen, M. K., Hindy, G., Holm, H., Ding, E. L., Johnson, T., Schunkert, H., Samani, N. J., Clarke, R., Hopewell, J. C., Thompson, J. F., Li, M., Thorleifsson, G., Newton-Cheh, C., Musunuru, K., Pirruccello, J. P., Saleheen, D., Chen, L., Stewart, A., Schillert, A., Thorsteinsdottir, U., Thorgeirsson, G., Anand, S., Engert, J. C., Morgan, T., Spertus, J., Stoll, M., Berger, K., Martinelli, N., Girelli, D., McKeown, P. P., Patterson, C. C., Epstein, S. E., Devaney, J., Burnett, M. S., Mooser, V., Ripatti, S., Surakka, I., Nieminen, M. S., Sinisalo, J., Lokki, M. L., Perola, M., Havulinna, A., de Faire, U., Gigante, B., Ingelsson, E., Zeller, T., Wild, P., de Bakker, P. I., Klungel, O. H., Maitland-van der Zee, A. H., Peters, B. J., de Boer, A., Grobbee, D. E., Kamphuisen, P. W., Deneer, V. H., Elbers, C. C., Onland-Moret, N. C., Hofker, M. H., Wijmenga, C., Verschuren, W. M., Boer, J. M., van der Schouw, Y. T., Rasheed, A., Frossard, P., Demissie, S., Willer, C., Do, R., Ordovas, J. M., Abecasis, G. R., Boehnke, M., Mohlke, K. L., Daly, M. J., Guiducci, C., Burtt, N. P., Surti, A., Gonzalez, E., Purcell, S., Gabriel, S., Marrugat, J., Peden, J., Erdmann, J., Diemert, P., Willenborg, C., Konig, I. R., Fischer, M., Hengstenberg, C., Ziegler, A., Buysschaert, I., Lambrechts, D., Van de Werf, F., Fox, K. A., El Mokhtari, N. E., Rubin, D., Schrezenmeir, J., Schreiber, S., et al. (2012) 'Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study', *Lancet,* 380(9841), 572-80.

Vuorio, A., Kuoppala, J., Kovanen, P. T., Humphries, S. E., Strandberg, T., Tonstad, S. and Gylling, H. (2010) 'Statins for children with familial hypercholesterolemia', *Cochrane Database Syst Rev,* (7), CD006401.

Vuorio, A. F., Turtola, H., Piilahti, K. M., Repo, P., Kanninen, T. and Kontula, K. (1997) 'Familial hypercholesterolemia in the Finnish north Karelia. A

molecular, clinical, and genealogical study', *Arterioscler Thromb Vasc Biol,* 17(11), 3127-38.

Wang, X., Li, X., Zhang, Y. B., Zhang, F., Sun, L., Lin, J., Wang, D. M. and Wang, L. Y. (2011) 'Genome-wide linkage scan of a pedigree with familial hypercholesterolemia suggests susceptibility loci on chromosomes 3q25-26 and 21q22', *PloS one,* 6(10), e24838.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. and Barton, G. J. (2009) 'Jalview Version 2--a multiple sequence alignment editor and analysis workbench', *Bioinformatics,* 25(9), 1189-91.

Webb, J. C., Sun, X. M., Patel, D. D., McCarthy, S. N., Knight, B. L. and Soutar, A. K. (1992) 'Characterization of two new point mutations in the low density lipoprotein receptor genes of an English patient with homozygous familial hypercholesterolemia', *J Lipid Res,* 33(5), 689-98.

Wenham, P. R., Henderson, B. G., Penney, M. D., Ashby, J. P., Rae, P. W. and Walker, S. W. (1997) 'Familial ligand-defective apolipoprotein B-100: detection, biochemical features and haplotype analysis of the R3531C mutation in the UK', *Atherosclerosis,* 129(2), 185-92.

Whittall, R. A., Scartezini, M., Li, K., Hubbart, C., Reiner, Z., Abraha, A., Neil, H. A., Dedoussis, G. and Humphries, S. E. (2010a) 'Development of a high-resolution melting method for mutation detection in familial hypercholesterolaemia patients', *Ann Clin Biochem,* 47(Pt 1), 44-55.

Whittall, R. A., Scartezini, M., Li, K., Hubbart, C., Reiner, Z., Abraha, A., Neil, H. A., Dedoussis, G. and Humphries, S. E. (2010b) 'Development of a high-resolution melting method for mutation detection in familial hypercholesterolaemia patients', *Ann Clin Biochem,* 47(Pt 1), 44-55.

Wierzbicki, A. S., Humphries, S. E. and Minhas, R. (2008) 'Familial hypercholesterolaemia: summary of NICE guidance', *BMJ,* 337, a1095.

Wild, P. S., Zeller, T., Schillert, A., Szymczak, S., Sinning, C. R., Deiseroth, A., Schnabel, R. B., Lubos, E., Keller, T., Eleftheriadis, M. S., Bickel, C., Rupprecht, H. J., Wilde, S., Rossmann, H., Diemert, P., Cupples, L. A., Perret, C., Erdmann, J., Stark, K., Kleber, M. E., Epstein, S. E., Voight, B. F., Kuulasmaa, K., Li, M., Schafer, A. S., Klopp, N., Braund, P. S., Sager, H. B., Demissie, S., Proust, C., Konig, I. R., Wichmann, H. E., Reinhard, W., Hoffmann, M. M., Virtamo, J., Burnett, M. S., Siscovick, D., Wiklund, P. G., Qu, L., El Mokthari, N. E., Thompson, J. R., Peters, A., Smith, A. V., Yon, E., Baumert, J., Hengstenberg, C., Marz, W., Amouyel, P., Devaney, J., Schwartz, S. M., Saarela, O., Mehta, N. N., Rubin, D., Silander, K., Hall, A. S., Ferrieres, J., Harris, T. B., Melander, O., Kee, F., Hakonarson, H., Schrezenmeir, J., Gudnason, V., Elosua, R., Arveiler, D., Evans, A., Rader, D. J., Illig, T., Schreiber, S., Bis, J. C., Altshuler, D., Kavousi, M., Witteman, J. C., Uitterlinden, A. G., Hofman, A., Folsom, A. R., Barbalic, M.,

Boerwinkle, E., Kathiresan, S., Reilly, M. P., O'Donnell, C. J., Samani, N. J., Schunkert, H., Cambien, F., Lackner, K. J., Tiret, L., Salomaa, V., Munzel, T., Ziegler, A. and Blankenberg, S. (2011) 'A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease', *Circ Cardiovasc Genet,* 4(4), 403-12.

Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M. L., Mora, S., Beckmann, J. S., Bragg-Gresham, J. L., Chang, H. Y., Demirkan, A., Den Hertog, H. M., Do, R., Donnelly, L. A., Ehret, G. B., Esko, T., Feitosa, M. F., Ferreira, T., Fischer, K., Fontanillas, P., Fraser, R. M., Freitag, D. F., Gurdasani, D., Heikkila, K., Hypponen, E., Isaacs, A., Jackson, A. U., Johansson, A., Johnson, T., Kaakinen, M., Kettunen, J., Kleber, M. E., Li, X., Luan, J., Lyytikainen, L. P., Magnusson, P. K., Mangino, M., Mihailov, E., Montasser, M. E., Muller-Nurasyid, M., Nolte, I. M., O'Connell, J. R., Palmer, C. D., Perola, M., Petersen, A. K., Sanna, S., Saxena, R., Service, S. K., Shah, S., Shungin, D., Sidore, C., Song, C., Strawbridge, R. J., Surakka, I., Tanaka, T., Teslovich, T. M., Thorleifsson, G., Van den Herik, E. G., Voight, B. F., Volcik, K. A., Waite, L. L., Wong, A., Wu, Y., Zhang, W., Absher, D., Asiki, G., Barroso, I., Been, L. F., Bolton, J. L., Bonnycastle, L. L., Brambilla, P., Burnett, M. S., Cesana, G., Dimitriou, M., Doney, A. S., Doring, A., Elliott, P., Epstein, S. E., Eyjolfsson, G. I., Gigante, B., Goodarzi, M. O., Grallert, H., Gravito, M. L., Groves, C. J., Hallmans, G., Hartikainen, A. L., Hayward, C., Hernandez, D., Hicks, A. A., Holm, H., Hung, Y. J., Illig, T., Jones, M. R., Kaleebu, P., Kastelein, J. J., Khaw, K. T., Kim, E., et al. (2013a) 'Discovery and refinement of loci associated with lipid levels', *Nat Genet,* 45(11), 1274-83.

Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M. L., Mora, S., Beckmann, J. S., Bragg-Gresham, J. L., Chang, H. Y., Demirkan, A., Den Hertog, H. M., Do, R., Donnelly, L. A., Ehret, G. B., Esko, T., Feitosa, M. F., Ferreira, T., Fischer, K., Fontanillas, P., Fraser, R. M., Freitag, D. F., Gurdasani, D., Heikkila, K., Hypponen, E., Isaacs, A., Jackson, A. U., Johansson, A., Johnson, T., Kaakinen, M., Kettunen, J., Kleber, M. E., Li, X., Luan, J., Lyytikainen, L. P., Magnusson, P. K., Mangino, M., Mihailov, E., Montasser, M. E., Muller-Nurasyid, M., Nolte, I. M., O'Connell, J. R., Palmer, C. D., Perola, M., Petersen, A. K., Sanna, S., Saxena, R., Service, S. K., Shah, S., Shungin, D., Sidore, C., Song, C., Strawbridge, R. J., Surakka, I., Tanaka, T., Teslovich, T. M., Thorleifsson, G., Van den Herik, E. G., Voight, B. F., Volcik, K. A., Waite, L. L., Wong, A., Wu, Y., Zhang, W., Absher, D., Asiki, G., Barroso, I., Been, L. F., Bolton, J. L., Bonnycastle, L. L., Brambilla, P., Burnett, M. S., Cesana, G., Dimitriou, M., Doney, A. S., Doring, A., Elliott, P., Epstein, S. E., Eyjolfsson, G. I., Gigante, B., Goodarzi, M. O., Grallert, H., Gravito, M. L., Groves, C. J., Hallmans, G., Hartikainen, A. L., Hayward, C., Hernandez, D., Hicks, A. A., Holm, H., Hung, Y. J., Illig, T., Jones, M. R., Kaleebu, P., Kastelein, J. J.,

Khaw, K. T., Kim, E., et al. (2013b) 'Discovery and refinement of loci associated with lipid levels', *Nat Genet*.

Williams, R. R., Hunt, S. C., Schumacher, M. C., Hegele, R. A., Leppert, M. F., Ludwig, E. H. and Hopkins, P. N. (1993) 'Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics', *Am J Cardiol,* 72(2), 171-6.

Yu, W., Nohara, A., Higashikata, T., Lu, H., Inazu, A. and Mabuchi, H. (2002) 'Molecular genetic analysis of familial hypercholesterolemia: spectrum and regional difference of LDL receptor gene mutations in Japanese population', *Atherosclerosis,* 165(2), 335-42.

Yue, P., Averna, M., Lin, X. and Schonfeld, G. (2006) 'The c.43_44insCTG variation in PCSK9 is associated with low plasma LDL-cholesterol in a Caucasian population', *Hum Mutat,* 27(5), 460-6.

Yusuf, S., Reddy, S., Ounpuu, S. and Anand, S. (2001) 'Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization', *Circulation,* 104(22), 2746-53.

Zakharova, F. M., Golubkov, V. I., Mandel'shtam, M. I. u., Lipovetskiĭ, B. M. and Gaĭtskhoki, V. S. (2001) '[Identification of novel missense mutation G571E, novel silent mutation H229H, nonsense mutation C74X, and four single nucleotide polymorphisms in the low-density lipoprotein receptor in patients with familial hypercholesterolemia from St. Petersburg]', *Bioorg Khim,* 27(5), 393-6.

Zuliani, G., Arca, M., Signore, A., Bader, G., Fazio, S., Chianelli, M., Bellosta, S., Campagna, F., Montali, A., Maioli, M., Pacifico, A., Ricci, G. and Fellin, R. (1999) 'Characterization of a new form of inherited hypercholesterolemia: familial recessive hypercholesterolemia', *Arterioscler Thromb Vasc Biol,* 19(3), 802-9.

Appendix

Table I

Summary of methods used for the initial FH mutation screening, before whole exome sequencing.

Sample ID	Original study cohort	Reference	*LDLR*	MLPA for *LDLR*	*APOB*	*PCSK9*
UK10K_HYP5231659	Australian FH	(Hooper et al. 2012)	all exons and promoter by Big Dye Terminator chemistry (Applied Biosystems) sequencing	yes	fragment of exon 26, and exon 29 by Big Dye Terminator chemistry (Applied Biosystems) sequencing	exon 7 by Big Dye Terminator chemistry (Applied Biosystems) sequencing
UK10K_HYP5231660	Australian FH					
UK10K_HYP5231661	Australian FH					
UK10K_HYP5231662	Australian FH					
UK10K_HYP5231663	Australian FH					
UK10K_HYP5231664	Australian FH					
UK10K_HYP5231665	Australian FH					
UK10K_HYP5231666	Australian FH					
UK10K_HYP5231667	Australian FH					
UK10K_HYP5231668	Australian FH					
UK10K_HYP5231669	Australian FH					
UK10K_HYP5231670	Australian FH					
UK10K_HYP5231671	Australian FH					
UK10K_HYP5231672	Australian FH					
UK10K_HYP5231673	Australian FH					
UK10K_HYP5231674	Australian FH					
UK10K_HYP5269604	Israeli FH	None	all exons and promoterby SSCP	no	APOB fragment of exon 26 by SSCP	no
UK10K_HYP5269605	Israeli FH					
UK10K_HYP5269606	Israeli FH					
UK10K_HYP5269607	Israeli FH					

UK10K_HYP5269608	Israeli FH					
UK10K_HYP5269609	Israeli FH					
UK10K_HYP5358903	Israeli FH					
UK10K_HYP5358904	Israeli FH					
UK10K_HYP5358905	Israeli FH					
UK10K_HYP5358906	Israeli FH					
UK10K_HYP5231679	Italy FH	(Bertolini et al. 2013)	all exons and promoter by Sanger sequencing	yes	c.9216 to c.11788 +152 nt of intron 26 by Sanger sequencing	all exons by Sanger sequencing
UK10K_HYP5231676	Italy FH					
UK10K_HYP5231677	Italy FH					
UK10K_HYP5231678	Italy FH					
UK10K_HYP5269570	Nothern Ireland FH	(Graham et al. 1999)	all exons and promoter by TTGE/DDGE	yes	RFLP for p.R3527Q	no
UK10K_HYP5269571	Nothern Ireland FH					
UK10K_HYP5269572	Nothern Ireland FH					
UK10K_HYP5269573	Nothern Ireland FH					
UK10K_HYP5269574	Nothern Ireland FH					
UK10K_HYP5269575	Nothern Ireland FH					
UK10K_HYP5269576	Nothern Ireland FH					
UK10K_HYP5269577	Nothern Ireland FH					
UK10K_HYP5269578	Nothern Ireland FH					
UK10K_HYP5269581	Nothern Ireland FH					
UK10K_HYP5159271	Oxford FH	(Futema et al. 2013)	all exons and promoter by HRM	yes	fragment of exon 26 by HRM and ARMS	ARMS for p.D374Y
UK10K_HYP5159272	Oxford FH					
UK10K_HYP5159273	Oxford FH					
UK10K_HYP5159274	Oxford FH					
UK10K_HYP5159275	Oxford FH					
UK10K_HYP5231650	Oxford FH					
UK10K_HYP5231651	Oxford FH					

UK10K_HYP5231652	Oxford FH					
UK10K_HYP5231653	Oxford FH					
UK10K_HYP5231654	Oxford FH					
UK10K_HYP5231655	Oxford FH					
UK10K_HYP5231656	Oxford FH					
UK10K_HYP5231657	Oxford FH					
UK10K_HYP5231658	Oxford FH					
UK10K_HYP5159266	RFH	(Taylor et al. 2010b)	all exons and promoter by SSCP or Sanger sequencing	yes	ARMS for p.R3527Q	ARMS for p.D374Y
UK10K_HYP5159267	RFH					
UK10K_HYP5159268	RFH					
UK10K_HYP5159269	RFH					
UK10K_HYP5159270	RFH					
UK10K_HYP5002209	SBBHF	(Whittall et al. 2010b, Mamotte and van Bockxmeer 1993, Humphries et al. 2006b, Taylor et al. 2009)	all exons and promoter by HRM	yes	RFLP for p.R3527Q	all exons by HRM
UK10K_HYP5002210	SBBHF					
UK10K_HYP5002211	SBBHF					
UK10K_HYP5002212	SBBHF					
UK10K_HYP5002213	SBBHF					
UK10K_HYP5002214	SBBHF					
UK10K_HYP5002215	SBBHF					
UK10K_HYP5002216	SBBHF					
UK10K_HYP5002217	SBBHF					
UK10K_HYP5002218	SBBHF					
UK10K_HYP5002219	SBBHF					
UK10K_HYP5002220	SBBHF					
UK10K_HYP5002221	SBBHF					
UK10K_HYP5002222	SBBHF					
UK10K_HYP5002223	SBBHF					

UK10K_HYP5002224	SBBHF
UK10K_HYP5002225	SBBHF
UK10K_HYP5002226	SBBHF
UK10K_HYP5002227	SBBHF
UK10K_HYP5002228	SBBHF
UK10K_HYP5002229	SBBHF
UK10K_HYP5002230	SBBHF
UK10K_HYP5002231	SBBHF
UK10K_HYP5002232	SBBHF
UK10K_HYP5062209	SBBHF
UK10K_HYP5062210	SBBHF
UK10K_HYP5062211	SBBHF
UK10K_HYP5062212	SBBHF
UK10K_HYP5062213	SBBHF
UK10K_HYP5062214	SBBHF
UK10K_HYP5062215	SBBHF
UK10K_HYP5062216	SBBHF
UK10K_HYP5062217	SBBHF
UK10K_HYP5062218	SBBHF
UK10K_HYP5062219	SBBHF
UK10K_HYP5062220	SBBHF
UK10K_HYP5062221	SBBHF
UK10K_HYP5062222	SBBHF
UK10K_HYP5062223	SBBHF
UK10K_HYP5062224	SBBHF
UK10K_HYP5062225	SBBHF
UK10K_HYP5062226	SBBHF

UK10K_HYP5062227	SBBHF
UK10K_HYP5062228	SBBHF
UK10K_HYP5062229	SBBHF
UK10K_HYP5062230	SBBHF
UK10K_HYP5062231	SBBHF
UK10K_HYP5062232	SBBHF
UK10K_HYP5269585	SBBHF
UK10K_HYP5269589	SBBHF
UK10K_HYP5269595	SBBHF
UK10K_HYP5269597	SBBHF
UK10K_HYP5269598	SBBHF
UK10K_HYP5269601	SBBHF
UK10K_HYP5269602	SBBHF
UK10K_HYP5315266	SBBHF
UK10K_HYP5315268	SBBHF
UK10K_HYP5315271	SBBHF
UK10K_HYP5315273	SBBHF
UK10K_HYP5315275	SBBHF
UK10K_HYP5358898	SBBHF
UK10K_HYP5358899	SBBHF
UK10K_HYP5358900	SBBHF
UK10K_HYP5358901	SBBHF
UK10K_HYP5358902	SBBHF

Table II
Sequences of primers used for HRM product amplification.

Gene	Fragment	Forward	Reverse	Forward primer 5' - 3'	Reverse primer 5' - 3'	Region Covered
LDLR	Prom	FH110	FH130	CAGCTCTTCACCGGAGACCC	ACCTGCTGTGTCCTAGCTGG	c.-298 - c.-62
LDLR	Exon 1	FH544f	FH545r	AATCACCCCACTGCAAACTC	GGGCTCCCTCTCAACCTATT	c.-139 - c.67+23
LDLR	Exon 2	FH112	FH132	TTGAGAGACCCTTTCTCCTTTTCC	GCATATCATGCCCAAAGGGG	c.68-10 - c.190+6
LDLR	Exon 3	E3HRMfor	E3HRMrev	TCAGTGGGTCTTTCCTTTGAG	CAGGACCCCGTAGAGACAAA	c.191-28 - c.313+58
LDLR	Exon 4(1)	FH504 F0	FH505 R0	TGGTGTTGGGAGACTTCACA	CACTCATCCGAGCCATCTTC	c.314-35 - c.519
LDLR	Exon 4(2)	FH506 F1	FH507 R1	AAGTGCATCTCTCGGCAGTT	CCCCTTGGAACACGTAAAGA	c.377 - c.557
LDLR	Exon 4(3)	FH508 F2	FH509 R2	AGCTTCCAGTGCAACAGCTC	CATACCGCAGTTTTCCTCGT	c.474 - c.679
LDLR	Exon 4(4)	FH510 F3	FH511 R3	TGTTCCAAGGGGACAGTAGC	AAATCACTGCATGTCCCACA	c.586 - c.694+60
LDLR	Exon 5	FH116	FH136	AGAAAATCAACACACTCTGTCCTG	GGAAAACCAGATGGCCAGCG	c.695-8 - c.817+5
LDLR	Exon 6	FH117	FH137	TCCTCCTTCCTCTCTCTGGC	TCTGCAAGCCGCCTGCACCG	c.818-8 - c.940+8
LDLR	Exon 7	FH118	FH138	GGCGAAGGGATGGGTAGGGG	GTTGCCATGTCAGGAAGCGC	c.941-38 - c.1060+36
LDLR	Exon 8	FH567	FH568	CTAGCCATTGGGGAAGAGCC	TGCCTGCAAGGGGTGAGGC	c.1061-31 - c.1186+30
LDLR	Exon 9	FH9	FH28	TCCATCGACGGGTCCCCTCTGACCC	AGCCCTCATCTCACCTGCGGGCCAA	c.1187-26 - c.1358+25
LDLR	Exon 10	FH10	FH29	AGATGAGGGCTCCTGGTGCGATGCC	GCCCTTGGTATCCGCAACAGAGACA	c.1359-26 - c.1490
LDLR	Exon 10	FH11	FH141	GATCCACAGCAACATCTACTGGACC	AGCCCTCAGCGTCGTGGATA	c.1475 - c.1586+5
LDLR	Exon 11	FH122	FH142	TCCTCCCCCGCCCTCCAGCC	GCTGGGACGGCTGTCCTGCG	c.1587-28 - c.1705+7
LDLR	Exon 12	FH13	FH32	GCACGTGACCTCTCCTTATCCACTT	CACCTAAGTGCTTCGATCTCGTACG	c.1706-10 - c.1845+10
LDLR	Exon 13	FH550f	FH551r	AGAGGGTGGCCTGTGTCTC	TCCACAAGGAGGTTTCAAGG	c.1846-47 - c.1987+29
LDLR	Exon 14	FH265	FH145	GAATCTTCTGGTATAGCTGAT	GCAGAGAGAGGCTCAGGAGG	c.1988-39 - c.2140+46
LDLR	Exon 15	FH16	FH146	GAAGGGCCTGCAGGCACGTGGCACT	GTGTGGTGGCGGGCCCAGTCTTT	c.2141-26 - c.2311+2
LDLR	Exon 16	FH127	FH147	CCTTCCTTTAGACCTGGGCC	CATAGCGGGAGGCTGTGACC	c.2312-23 - c.2389+32
LDLR	Exon 17	FH128	FH148	GGGTCTCTGGTCTCGGGCGC	GGCTCTGGCTTTCTAGAGAGGG	c.2390-33 - c.2547+10
LDLR	Exon 18	FH129	FH149	GCCTGTTTCCTGAGTGCTGG	TCTCAGGAAGGGTTCTGGGC	c.2548-35 - c.2607
PCSK9	Exon 7	FH433	FH434	CCCTCTCTTGGGCTCCTTTCT	AAAGGGGCTGTTAGCATCACG	c.997-27 - c.1180+29
APOB	Exon 26	FH521	FH522	TGTCAAGGGTTCGGTTCTTT	GGGTGGCTTTGCTTGTATGT	c.10516 - c.10745

Table III
Gene burden test results for genes in loci associated with FH in family linkage studies.

Chromosomal region	Gene name	Number of rare functional variants		p value
		cases (n=71)	controls (n=1926)	
21q22 (Wang et al. 2011)	*KRTAP10-11*	2	2	0.02
	PFKL	2	4	0.03
	DSCR8	1	0	0.05
	KRTAP11-1	1	0	0.05
	ERG	2	8	0.09
	KRTAP19-8	1	1	0.10
	LRRC3	2	10	0.13
	RCAN1	1	2	0.15
	SIM2	1	2	0.15
	SYNJ1	3	24	0.16
	CBR1	2	15	0.22
	COL18A1	3	30	0.24
	ZNF295	2	17	0.26
	C21orf59	1	5	0.27
	C21orf90	1	5	0.27
	ETS2	1	5	0.27
	C21orf56	1	6	0.31
	KRTAP12-4	1	6	0.31
	PKNOX1	1	6	0.31
	PCNT	5	68	0.33

	BACH1	1	7	0.35
	MX1	1	7	0.35
	BRWD1	2	24	0.39
	PRDM15	2	25	0.41
	HLCS	1	9	0.41
	PRMT2	1	9	0.41
	DOPEY2	3	43	0.43
	SUMO3	1	10	0.44
	MX2	1	11	0.47
	TTC3	2	29	0.48
	AIRE	1	13	0.53
	TRPM2	2	35	0.58
	ABCG1	1	16	0.60
	FTCD	1	17	0.62
	ITSN1	1	20	0.67
	LSS	1	20	0.67
	DSCAM	1	21	0.69
	COL6A2	1	23	0.72
	C21orf2	1	24	0.74
	TRAPPC10	1	24	0.74
	TSPEAR	1	24	0.74
	UMODL1	1	30	0.81
	ITGB2	1	33	0.84
	MCM3AP	1	36	0.86
	URB1	1	57	0.95
16q22 (Marques-	*CMTM2*	2	3	0.02

Pinheiro et al. 2010)				
	HSF4	2	6	0.06
	KCTD19	2	7	0.08
	CES4A	2	8	0.09
	CMTM4	1	2	0.15
	TMEM208	1	3	0.19
	CMTM3	1	5	0.27
	DPEP3	1	5	0.27
	TMCO7	1	5	0.27
	PLEKHG4	2	19	0.30
	C16orf48	1	6	0.31
	CDH16	2	20	0.32
	CES3	1	11	0.47
	CDH3	1	13	0.53
	COG4	1	14	0.55
	GFOD2	1	16	0.60
	TSNAXIP1	1	17	0.62
	FHOD1	1	19	0.65
	SLC12A4	1	19	0.65
	FUK	1	23	0.72
8q24 (Cenarro et al. 2011)	*WISP1*	3	25	0.17
	ST3GAL1	1	8	0.38
	ZFAT	2	28	0.46
3q25 (Wang et	*HPS3*	3	15	0.06

al. 2011)

ZBBX	2	8	0.09
NMD3	1	1	0.10
TRIM59	1	1	0.10
MLF1	2	11	0.14
ANKUB1	1	2	0.15
IL12A	1	2	0.15
OTOL1	2	13	0.18
C3orf80	1	3	0.19
WWTR1	1	4	0.23
SLITRK3	2	17	0.26
SMC4	1	6	0.31
B3GALNT1	1	9	0.41
MFSD1	1	9	0.41
SI	2	35	0.58
MED12L	2	36	0.59
MECOM	1	24	0.74
IGSF10	1	34	0.84