8-15-2013 12:00 AM

# Resolving the Genetic Etiology of Hypercholesterolemia in Familial Combined Hyperlipidemia

Mary A. Bamimore
*The University of Western Ontario*

Supervisor
Dr. Robert A. Hegele
*The University of Western Ontario*

Graduate Program in Biochemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

# Resolving The Genetic Etiology Of Hypercholesterolemia In Familial Combined Hyperlipidemia

(Thesis format: Monograph )

by

Mary Aderayo Bamimore

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

# Abstract

Hypercholesterolemia is a classical risk factor for cardiovascular disease development. The genetic etiology of hypercholesterolemia in familial combined hyperlipidemia (FCH), one of the most common genetic dyslipidemias, is poorly understood. We aimed at understanding the genetic etiology of hypercholesterolemia in FCH.

Sequencing, genotyping and computational analyses were performed in a case-control setting to better understand the 'nature' aspect of hypercholesterolemia in FCH. My findings suggest that FCH more likely has a polygenic basis.

All my findings have shown that the genetic definition of a disease, especially relatively common diseases like FCH that have been previously considered to be monogenic, may need to be reconsidered. Thus findings from my studies of FCH support a new direction in thinking about the genetic etiology of this common human hyperlipidemia.

**Key Words**: Hypercholesterolemia, Familial Combined Hyperlipidemia, Rare Variants, Genetic risk scores, Population Genetics, Genomics, Cardiovascular disease

# Co-Authorship

The contributions of co-authors in this thesis are acknowledged below and are outlined:

Mr. Adam McIntyre- DNA isolation for some familial combined hyperlipidemia (FCH) and familial hypertriglyceridemia (FHTG) patient samples that were studied in Chapters 2 and 3.

Dr. Henian Cao- Whole genome amplification of all FCH and FHTG patient samples studied in Chapter 3.

Drs. Henian Cao and Jian Wang- Sanger sequencing of some FCH patients reported in Chapter 3. Mr. Matthew Ban- Assistance with statistical analyses throughout all thesis subprojects.

Dr. Zuhier Awan (McGill University)- Collaboration for sequence analysis of familial hypercholesterolemia (FH) mutation negative patient samples in Chapter 4.

Professor Steve Humphries (University College, London UK)- Genotyping of all FH mutation negative patients and FH mutation positive controls in Chapter 4.

# Dedication

I dedicate this thesis to my angel, beloved mother and best friend, Alaba, who has always been a great source of motivation and inspiration.

I also dedicate this thesis to my esteemed graduate supervisor, Dr. Robert A. Hegele, who made everything a reality.

*Aderayo*

# Acknowledgements

Firstly, words cannot express my deepest gratitude to my esteemed graduate supervisor, Dr. Robert A. Hegele, for giving me the opportunity to pursue graduate work in his laboratory. I thank Dr. Robert A. Hegele for his excellence in mentorship, education and guidance in the entire course of my graduate training. The excellent support from Dr. Robert A. Hegele has greatly helped me at an academic level and at a personal level, for which I am eternally grateful.

I cannot but express my deepest appreciation to Dr. Henian Cao. All the projects in the course of my graduate training could not have been brought to fruition without the endless support and training from Dr. Henian Cao, for which I am forever appreciative.

Many thanks to Dr. Jian Wang and Mr. Matthew Ban, who also significantly contributed to my graduate training. Great thanks to Mr. John Robison, whose non-academic mentorship and positivity guided me through the entire journey of my academic program. I extend my gratitude to Dr. Robert A. Hegele's entire staff and graduate students, who are a family in the world of research and made my graduate training an unforgettable experience.

Special thanks to Dr. Kathleen Hill and Dr. Murray Huff, who were my Committee Advisors; my graduate training would be incomplete without them. I must also acknowledge the Canadian Institute of Health Research Strategic Training Program fellowship for funding my Master's program

Last, but by no means least, I am eternally grateful to my mother and only best friend, Alaba. I take this opportunity to express my gratitude to my mother for being

my unconditional support, my encouragement and my inspiration throughout the entire journey of my graduate program. Finally, my only sibling, Jay, deserves my wholehearted thanks for his ever loving support throughout my graduate program.

## Table of Contents

## List of Tables

**List of Figures**

## List of Abbreviations

| | |
|---|---|
| *ABCA1* | ATP-binding cassette |
| *ABCG5* | ATP-binding cassette |
| *ABCG8* | ATP-binding cassette |
| ADH | Autosomal Dominant Hypercholesterolemia |
| *ANGPTL3* | angiopoietin-like 3 |
| *APOA1* | apolipoprotein A-I |
| *APOA5* | apolipoprotein A-V |
| *APOB* | apolipoprotein B gene |
| *APOB* | apolipoprotein B |
| *APOC2* | apolipoprotein C-II |
| *APOE* | apolipoprotein E gene |
| *APOE* | apolipoprotein E |
| *ARH* | Autosomal Recessive Hypercholesterolemia gene |
| *ARH* | low density lipoprotein receptor adaptor protein 1 |
| BHF | British Heart Foundation |
| BMI | Body Mass Index |
| bp | base pair |
| CAD | Coronary Artery Disease |
| *CETP* | cholesteryl ester transfer protein |
| CF | Cystic Fibrosis |
| CF | Cystic Fibrosis |
| CHD | Coronary Heart Disease |
| CIP-ExoI | Calf intestinal Phosphatase-Exonuclease I |
| CM | Chylomicron |
| CVD | Cardiovascular Disease |
| *CYP7A1* | cholesterol 7-alpha-hydroxylase |
| *Diet1* | variability in response to cholesterol enriched atherogenic diet |
| DMSO | Dimethyl sulfoxide |
| DNA | deoxyribonucleic acid |
| FCH | Familial Combined Hyperlipidemia |
| FH | Familial Hypercholesterolemia |
| FH/M+ve | Familial Hypercholesterolemia Mutation positive patient |
| FH/M-ve | Familial Hypercholesterolemia Mutation negative patient |
| FHTG | Familial Hypertriglyceridemia |
| GLGC | Global Lipids Genetics Consortium |
| GRS | Genetic Risk Score |
| GWAS | Genome Wide Association Study |

| | |
|---|---|
| HDL | High Density Lipoprotein |
| HDL-C | High Density Lipoprotein- cholesterol |
| HMGCoA | 3-hydroxy-3-methyl-glutaryl-CoA |
| HTG | Hypertriglyceridemia |
| IDL | Intermediate Density Lipoprotein |
| *IDOL* | Inducible degrader of LDLR gene |
| IHD | Ischemic Heart Disease |
| *LCAT* | lecithin-cholesterol acyltransferase |
| LD | Linkage Disequilibrium |
| LDL | Low Density Lipoprotein |
| LDL-C | Low Density Lipoprotein-cholesterol |
| *LDLR* | Low Density Lipoprotein Receptor gene |
| *LDLR* | low density lipoprotein receptor |
| *LDLRAP1* | low density lipoprotein receptor adaptor protein 1 |
| *LIPA* | lipase A |
| *LIPC* | lipase |
| *LPL* | lipoprotein lipase |
| MAF | Minor allele frequency |
| MIM | Mendelian Inheritance in Man |
| MSA | Multiple Sequence Alignment |
| *MTP* | microsomal triglyceride transfer protein |
| *IDOL* | Myosin regulatory light chain interacting protein |
| NCBI | National Center for Biotechnology Information |
| OMIM | Online Mendelian Inheritance in Man |
| OR | Odds Ratio |
| PCR | Polymerase Chain Reaction |
| *PCSK9* | Proprotein Convertase Subtilisin/kexin Type 9 gene |
| *PCSK9* | proprotein convertase subtilisin/kexin type 9 |
| PKU | Phenylketonuria |
| PolyPhen-2 | Phenotyping version 2 |
| *PSMD9* | Proteosome Modulator 9 gene |
| PTB | Phosphotyrosine Binding |
| *SAR1B* | SAR1 homolog B |
| SIFT | Sorting Intolerant from Tolerant |
| *SLC10A2* | solute carrier family 10 (sodium/bile acid cotransporter family) |
| SNP | Single nucleotide polymorphism |
| TC | Total cholesterol |
| TC | Total Cholesterol |
| TG | Triglyceride |
| TG | Triglyceride |
| VLDL | Very Low Density Lipoprotein |
| WHO | World Health Organization |

# Chapter 1: Introduction

## 1.1 Cardiovascular Disease

### 1.1.1 Definition of cardiovascular disease (CVD) and current statistics on Cardiovascular Disease

Cardiovascular disease (CVD) is the generic term that describes any disease that affects the cardiovascular system, including the heart and blood vessels. Many diseases fall under the classification of CVD. Ischemic heart disease and stroke are common examples of CVD. In ischemic heart disease, there is reduced supply of blood to the heart muscle; in stroke there is reduced blood flow to the brain. Some of the various types of CVD are shown in **Table 1.1**.

CVD is the leading cause of mortality in North America. Even though the relative rates of mortality due to CVD have declined due to improvements in disease management and prevention, the absolute numbers of patients are rising due to the aging of the "baby boomer" generation and CVD in absolute terms still remains the leading cause of death and disability (*1, 2*). **Figure 1.1** summarizes the statistics of mortality rates of CVD in North America over time.

CVD is also becoming the leading cause of death worldwide because of the changes in diet and lifestyle of individuals in developing countries. Individuals from developing countries are adopting the stereotypical Western lifestyle of low physical activity and unhealthy diet, causing CVD to become the leading cause of death globally. **Table 1.2** summarizes global CVD statistics.

### 1.1.2 Risk Factors for CVD according to the Framingham Heart Study

Epidemiological studies, such as the Framingham Heart Study and its derived risk score, have shown that there are certain common factors that put an individual at risk for developing CVD (*3*). These Framingham risk factors are diabetes, hypertension, age, sex, obesity, cigarette smoking, elevated plasma low density lipoprotein cholesterol (LDL-C) levels and depressed high density lipoprotein cholesterol (HDL-C) levels (*3*, *4*). **Table 1.3** briefly describes how each risk factor contributes to increased CVD

**Table 1.1** Brief description of the some of the different types of cardiovascular diseases. Information in this table was taken from (*5*), (*6*)

| Type of CVD | Brief Description |
|---|---|
| Coronary Heart Disease | Coronary Heart Disease (CHD) is also known as coronary artery disease (CAD) and Ischemic Heart Disease (IHD). In CHD there is blockage in the coronary artery of the heart. This blockage deprives the heart muscle of oxygen and vital nutrients, which could result in myocardial infarction (*7*) |
| Stroke | Impairment of brain functions due to reduced blood flow to the brain. Lack of blood flow may be due to rupture of vessel wall (hemorrhagic) or due to block (ischemic) (*5*) |
| Hypertensive heart disease | The abnormal regulation of systemic blood pressure due to higher than normal arterial blood pressure (*6*) |
| Rheumatic heart disease | Heart disease where there is inflammation of the heart muscle due to rheumatic fever. Rheumatic fever is brought about by infection with the streptococcus bacteria (*6*) |
| Congenital heart disease | Heart disease in which there is malformation in the heart organ from birth. These malformation may be due to genetic defects or may be due to environmental exposure to teratogens during gestation (*6*) |

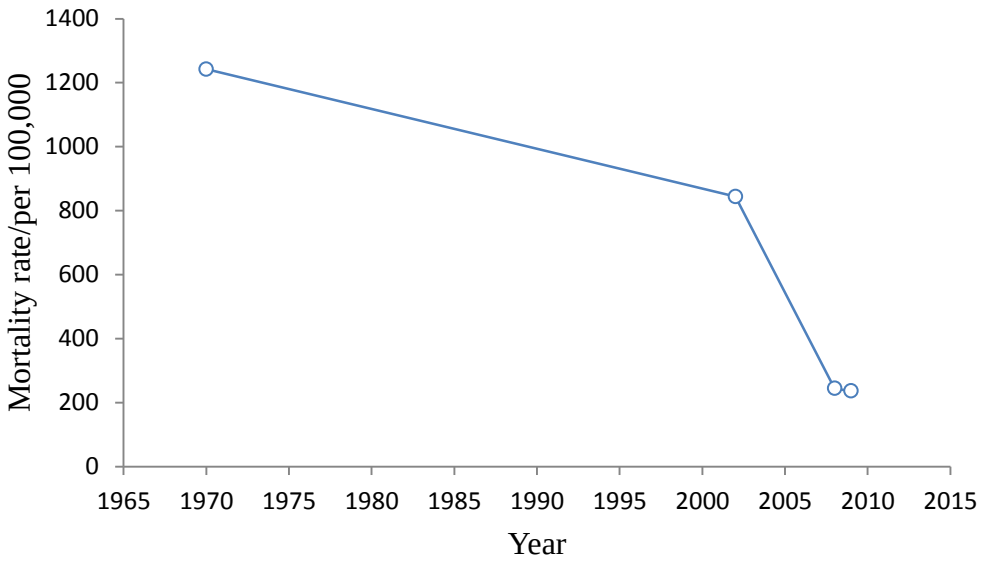



**Figure 1.1** North American Mortality rates of cardiovascular disease from 1970 to 2002. Information from this table was taken from (*2*, *8*, *9*)

**Table 1.2** Number of global deaths, in 2002, due to various types of cardiovascular diseases.

| Type of CVD | Global deaths due to type of CVD | Proportion of deaths |
|---|---|---|
| Coronary Heart Disease | 7.2 million | 43% |
| Stroke | 5.5 million | 33% |
| Other forms of Heart Disease | 2.4 million | 14% |
| Hypertensive Heart Disease | 0.9 million | 5% |
| Inflammatory Heart Disease | 0.4 million | 3% |
| Rheumatic Heart Disease | 0.3 million | 2% |
| Total number of deaths | 16.7 million | 100% |

Note: Proportion of death refers to proportion of global deaths due to CVD that result from a particular type of CVD. Information in this table was taken from World Health Organization (WHO) (http://www.who.int/cardiovascular_diseases/en/cvd_atlas_01_types.pdf Accessed July 1st 2013).

**Table 1.3.** Cardiovascular Disease risk factors according to the Framingham Heart Study with brief explanation of each risk factor(*3*).

| Risk Factor | Brief Description of association |
|---|---|
| Age | CVD risk increases with age (*3*) |
| Sex | Males and postmenopausal women are at greater risk than premenopausal women(*3*) |
| BMI | Overweight individuals are at a greater risk of CVD development(*3*) |
| Diabetes | Diabetes increases risk(*3*) |
| Smoking | Smoking increases risk of CVD(*3*) |
| Hypertension | High systolic blood pressure increases CVD risk(*3*) |
| Total Cholesterol | High cholesterol due to elevated LDL-C increases risk(*3*) |
| High Density Lipoprotein Cholesterol (HDL-C) | HDL-C is a negative risk factor – high levels of HDL-C reduce the risk of CHD(*3*) |

## 1.2 Hypercholesterolemia

Cholesterol belongs to the family of organic compounds called steroids (*10*). A steroid can be defined as a compound that has 17 carbon atoms arranged in 4 rings (*11*). Steroids fall under a large group of macromolecules called lipids; so cholesterol is a lipid. A lipid can be defined as an organic molecule that does not chemically interact with water (*12*). However this definition of a lipid is contextual as many lipids are amphipathic and the majority of the structure of amphipathic lipids is still hydrophobic. Cholesterol is a good example of an amphipathic lipid, in which a majority of its structure is hydrophobic, while a minority of its structure is hydrophilic (i.e. the hydroxyl group). The structure of cholesterol is shown in **Figure 1.2**

Cholesterol is essential to life. Cholesterol is required for the formation of cell membrane, bile acids, steroid hormones and formation of vitamin D. Humans have endogenous and exogenous sources of cholesterol. Cholesterol is synthesized by the liver in humans (endogenous source). Cholesterol is also obtained from the diet (exogenous source) (*10*).

Hypercholesterolemia is a condition where there is an aberrantly elevated concentration of cholesterol in the blood. As alluded to in Section 1.1.2, hypercholesterolemia increases an individual's risk for CVD. Section 1.2.4 will discuss how hypercholesterolemia increases CVD risk. An individual's cholesterol levels can be somewhat influenced by environmental factors such as dietary intake of cholesterol. However, genetic factors account for over 50%, and perhaps up to 80% of inter-individual variation in cholesterol levels in the human population (*13*).

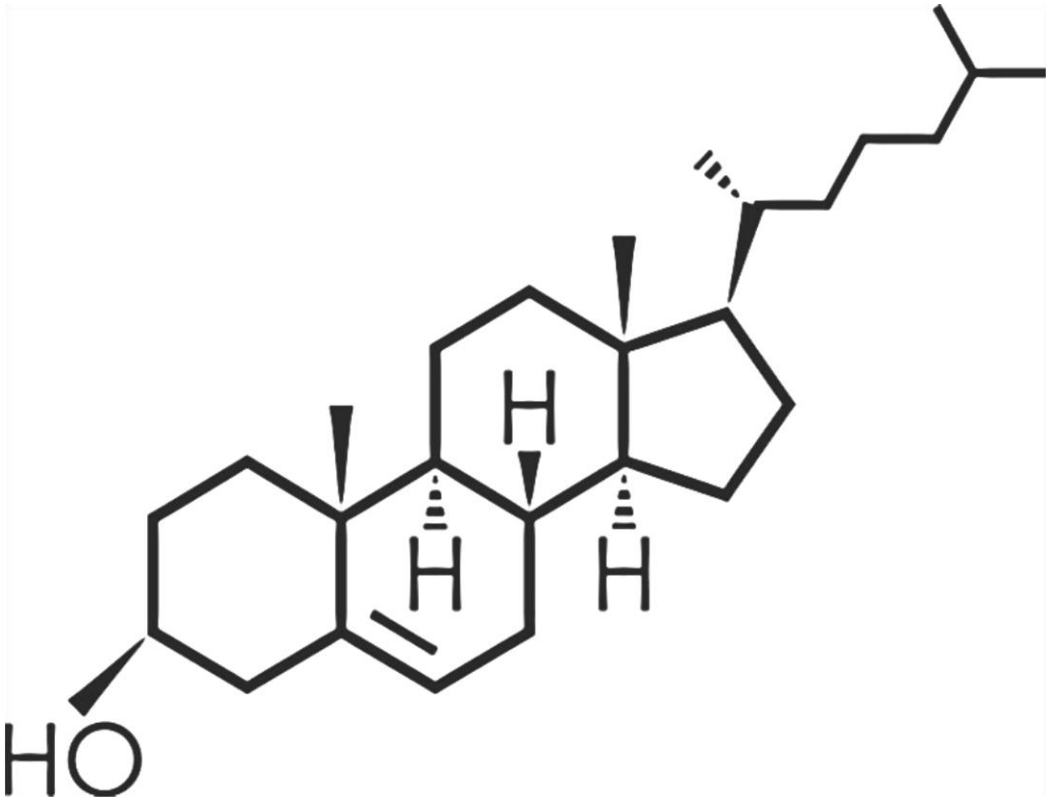



**Figure 1.2 Chemical Structure of Cholesterol.** It is a general steroid with some odd hydrocarbon side chain. Cholesterol is an organic molecule that is a steroid. Cholesterol has 17 carbon atoms arranged in 4 ring structures. Even though the definition of a lipid is a molecule that is hydrophobic, there are many lipids, like cholesterol, that are amphipathic with a majority of its structure being hydrophobic. The hydroxyl (OH) group of cholesterol is the only hydrophilic portion of the molecule and that hydrophilic portion of the molecule is still not enough to make it soluble in the aqueous blood. This figure shows the entire molecule soluble in aqueous blood. This is why cholesterol must be transported in the form of a lipoprotein. This figure is taken and modified from Food and Health Communications, Inc. (http://dev.foodandhealth.com/clipart.php?cat=9&img=Cholesterol_Structure.jpg )

### 1.2.1 Lipoprotein metabolism in general

Lipids such as cholesterol are insoluble in the aqueous blood. So, cholesterol needs to be transported within macromolecular complexes called lipoproteins that contain specific biological transporter proteins called apolipoproteins (or apoproteins). A lipoprotein can be defined as the assembly of a lipid and a protein (i.e. apolipoprotein) as a single unit. An apo-protein also refers to a protein in the state where it is unbound to its ligand (*14*). Within a lipoprotein, lipids such as cholesterol are bound to and held within the surface apolipoproteins and thus can be transported through the bloodstream (*14*). Cholesterol is transported as two main types of lipoproteins; the cholesterol within these particles is identical, but it is the biochemical behaviour of the particle which determines whether the cholesterol is "good" or "bad" in popular parlance. Cholesterol is transported from its site of synthesis (i.e. the liver) to other parts of the body within low density lipoprotein (LDL) particles. Excess or unused cholesterol is transported from body tissues back to the liver with high density lipoprotein (HDL) particles (*14*). There are various types of lipoproteins in human blood and they all have one main function, which is the transportation of lipids in the blood. Other types of lipids such as triglycerides are transported in the blood as very low density lipoprotein (VLDL) or chylomicrons.

Cholesterol carried within LDL particles can become embedded within the interior wall of the arteries, which with repeated deposition over time leads to atherosclerosis (which shall be explained in section 1.2.4). That is why LDL-C is colloquially referred to as 'bad cholesterol'; as mentioned, the cholesterol is identical to the substance found inside HDL particles, but by virtue of the fact that LDL particles (but not HDL particles) can become deposited within the arterial wall, the cholesterol within them gets its "bad" reputation. Hypercholesterolemia is usually the

result of abnormally elevated plasma LDL particles and excess LDL-C. In other words, aberrantly elevated LDL corresponds to excess cholesterol because cholesterol in the blood is mainly carried within LDL particles (*10*)

A minor proportion of cholesterol in the blood is carried within HDL particles: the cholesterol within HDL particles has been extracted from peripheral tissues, is transported back into the liver and metabolized for bile acid synthesis. Because cholesterol inside the HDL particle has been removed from peripheral cells and the arterial wall, and is ultimately cleared from the blood, the identical cholesterol inside the HDL particle is referred to by lay people as "good cholesterol" based not on the properties of the cholesterol molecules, but rather based on the biochemical behaviour of the HDL particles (*10*). The Framingham study showed that high levels of HDL-C are associated with reduced CVD risk in patients, which is consistent with what is understood about the action of HDL particles biochemically and physiologically.

### 1.2.2 Low Density Lipoprotein Cholesterol (LDL-C) metabolism

Two thirds of the body's cholesterol is contained within LDL (*15*) and this is mainly endogenous cholesterol made in the liver (*16*). Cholesterol is synthesized in a multi-step biochemical process from acetyl-CoA and hydroxymethylglutaryl-CoA (HMG-CoA) reductase, which is the enzyme that catalyzes the rate determining step in the process (*16*). LDL particle formation can be either the result of direct release into circulation from the liver, or through conversion of circulating VLDL to LDL in bloodstream. The proportion of LDL produced via VLDL varies from individual to individual. On average about two-thirds of circulating LDL is derived from conversion from VLDL and about one-third is derived from direct synthesis by the liver (*17*).

Catabolism of LDL is brought about by the LDL receptor pathway, whereby LDL is internalized in the cell through receptor mediated endocytosis (*15*). After receptor-mediated endocytosis, LDL is degraded within the lysosomes, and the cholesterol released suppresses HMG-CoA reductase; thus cholesterol regulates its own synthesis through negative feedback (*18*). Cholesterol within the cell is esterified and can be used for a variety of important functions depending on the cell type. The LDL receptor is then recycled back to the cell surface, and the process of receptor mediated endocytosis can resume. It is very important to appreciate that the LDL receptor is the main regulator of LDL-C levels in the blood. Exogenous cholesterol is absorbed into the blood and transported in chylomicrons, and through breakdown, this cholesterol finally reaches the liver and is repackaged within VLDL and LDL particles (*16*). **Figure 1.3** summarizes LDL-C metabolism.

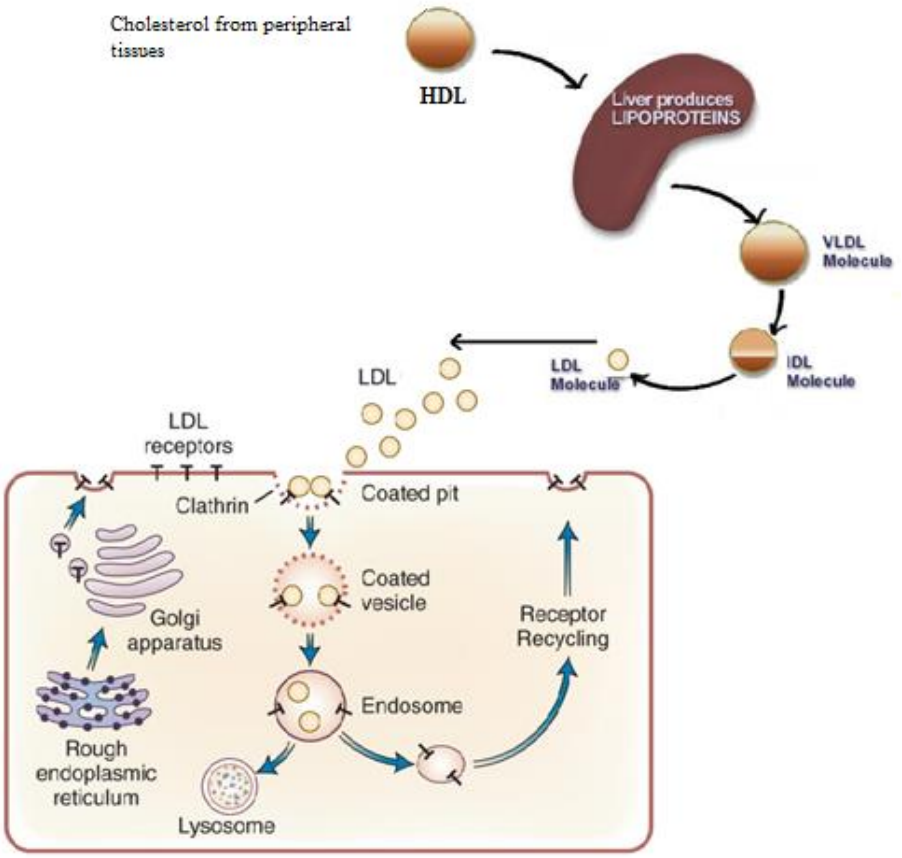



**Figure 1.3 LDL metabolism.** Most of the LDL in blood is derived from VLDL catabolism. When the triglycerides in VLDL are hydrolyzed, the remnants are referred to as IDL. Some IDL gets cleared from plasma by the liver. The remaining IDL that does not get cleared undergoes further triglyceride hydrolysis and becomes LDL. Some LDL found in plasma is also produced directly from the liver. LDL gets cleared from the blood through the LDL receptor pathway. In the LDL receptor pathway, LDL binds to its receptor, LDLR, which is expressed on the cell surface of most cells especially the liver. The complex of LDL-and LDLR enters the coated pit and is internalized. The coated vesicle loses its clathrin coat and becomes an endosome, which is the site of lipoprotein and receptor dissociation. The LDL receptor recycles to the cell surface, and the lipoproteins are degraded in the lysosomes. HDL transports cholesterol from peripheral tissues back into the liver. This figures was taken and modified from http://health-7.com/imgs/15/947.jpg

### 1.2.3 High Density Lipoprotein Cholesterol (HDL-C) metabolism

HDL transports cholesterol from peripheral tissues, including the arterial wall, back into liver (*19*), which is a process referred to as "reverse cholesterol transport". HDL-cholesterol is known as the 'good cholesterol' because it has been removed from potentially dangerous sites by the process of reverse cholesterol transport. HDL also has beneficial properties: for instance it blocks the proatherogenic oxidation of LDL in the vessel wall (*19*). HDL has been seen as cardio-protective. HDL-C levels had consistently shown an inverse relationship with CVD risk (*19*). Researchers have questioned whether or not there is direct causality between HDL and CVD risk; specifically, is HDL directly protective or is it merely a marker of some other entity or process that is directly protecting the heart and arteries (*19*)? Thus the precise role and mechanism of action of HDL are currently controversial in the field (*20*).

### 1.2.4 Mechanism of atherosclerotic plaque formation due to elevated plasma LDL-C levels

The arterial wall is made up of three layers, namely, (i) tunica intima, (ii) tunica media and (iii) tunica adventitia (*21*). A schematic figure of an artery is shown in **Figure 1.4.** Atherosclerosis can be said to start with a lesion that occurs in the endothelium of the arterial wall; this can be due to a toxin from cigarette smoke, or chemical or physical stress, such as high blood pressure. This initial injury or lesion then causes LDL to enter the arterial wall and accumulate in between the endothelium and tunica intima. When the LDL-C level is high in the plasma and remains in plasma for long,, it is more likely to become oxidized.

The properties of oxidized LDL are different from that of regular (i.e. native) LDL (*22*). Oxidized LDL is an immunogen, i.e. it triggers an immune response. As a

result of the immune response (*23*), monocytes from the bloodstream come to the site of oxidized LDL as the monocytes recognize oxidized LDL as a foreign substance. As the monocytes enter and take up residence within the arterial wall, they ingest the oxidized LDL and become macrophages. When the macrophages become filled up with oxidized lipids, they enlarge and take on a "foamy" appearance, which is why they are referred to as "foam cells". These foam cells eventually die, leaving the cholesterol permanently embedded within the arterial wall. As the process repeats over time, the cholesterol within the wall builds up into structures called "plaques" which can begin to cause narrowing and eventually occlusions of the artery. Larger and more mature plaques narrow the lumen of the artery; they are also prone to bursting or rupture and these unpredictable and dramatic events can suddenly completely block the artery, leading to a heart attack (myocardial infarction) or a stroke, depending on the anatomical location of the artery (*23*, *24*). **Figure 1.5** visually summarizes the process of atherosclerosis.

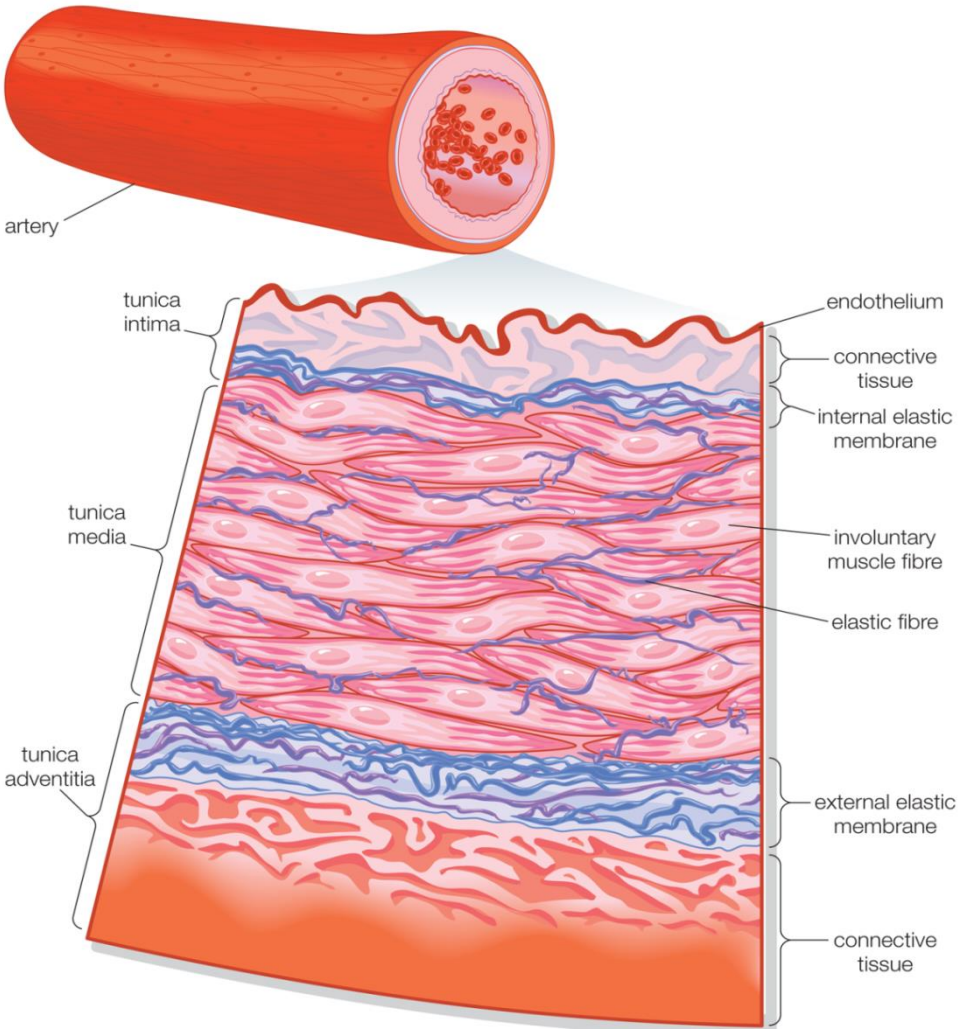



**Figure 1.4 Structure of the artery.** The artery is made up of a three layers and an innermost endothelium. From the endothelium, the three layers are: tunica intima, tunica media and tunica adventitia, respectively. Atherosclerosis occurs between the endothelium and tunica intima (or essentially the tunica intima). This figure was taken from Encyclopedia Britannica (*21*).

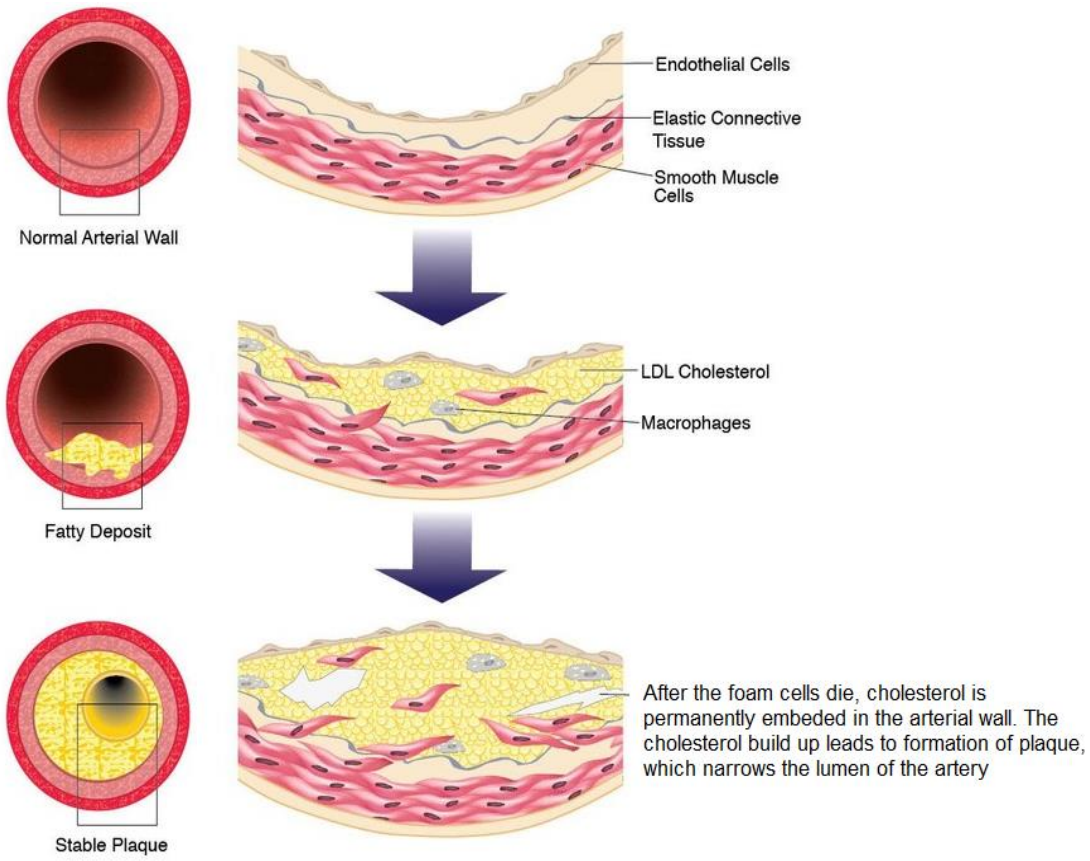


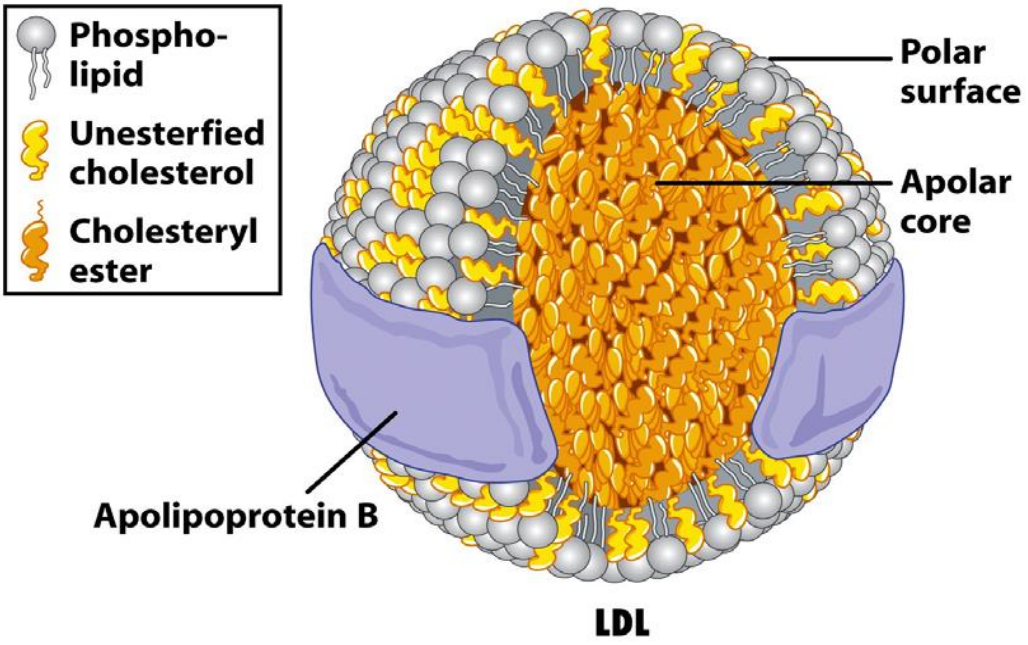

**Figure 1.5 Atherosclerotic plaque formation.** Native LDL becomes trapped between endothelium and tunica intima and undergoes oxidation. Then the resident monocytes transform into macrophages that take up oxidized LDL. After the macrophages consume oxidized LDL, they becoming foam cells. The foam cells enlarge and die, which leads to cholesterol being permanently embedded in these foam cells. This buildup of cholesterol leads to formation of structures called plaque. Plaque formation leads to narrowing of the lumen and thus reduces blood flow in the artery. This figure was taken and modified from Rochester Institute of Technology ( http://cias.rit.edu/faculty-staff/101/student/287 )

### 1.2.5 Low Density Lipoprotein (LDL): Structure and Function

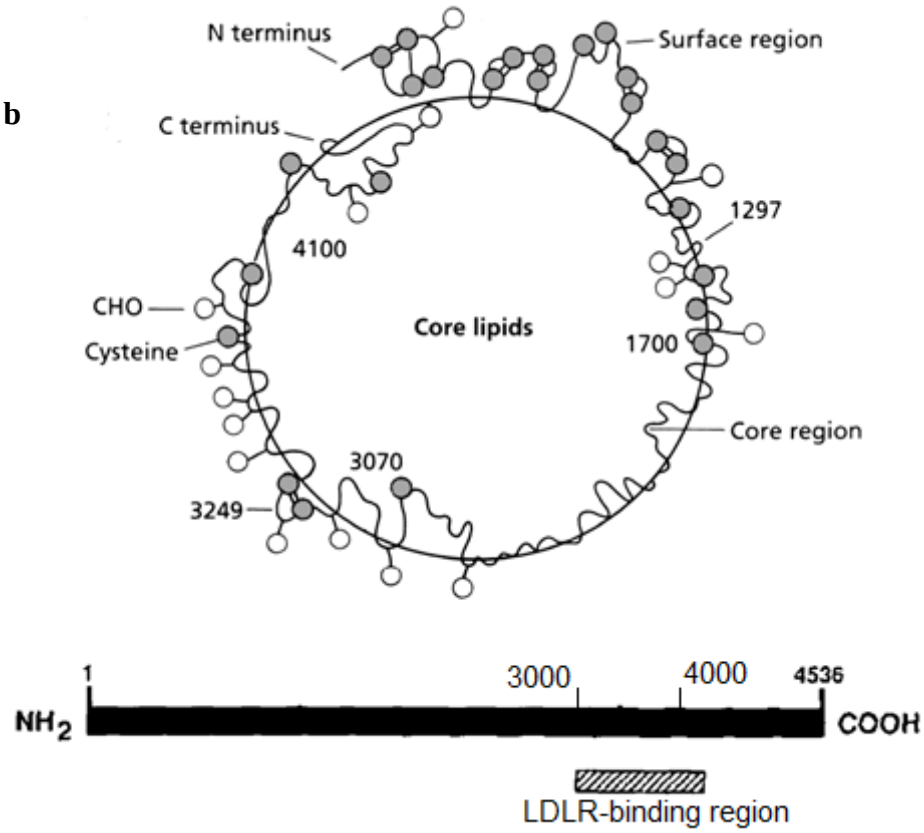
LDL is the lipoprotein through which cholesterol and esters of cholesterol are transported in blood (*25*). LDL is a spherical amphipathic assembly, where the hydrophilic portion is outwards facing and interacting with aqueous blood and the hydrophobic portion is inwards. **Figure 1.6** shows the structure of LDL (*25*). The hydrophobic core of LDL consists of esterified cholesterol (or cholesteryl ester) and some triglyceride. The phospholipid and apolipoprotein are found on the particle surface. The main apolipoprotein in LDL is apolipoprotein (apo) B-100 (*25*).

**Figure 1.6 (a) Structure of LDL** (*25*) Low Density Lipoprotein (LDL) is an assembly of proteins and lipids. LDL is the most abundant cholesterol carrying lipoprotein. The unesterified cholesterol and phospholipid faces outward that is surrounded by aqueous blood. The hydrophilic portion of the phospholipids faces the aqueous blood and thus makes up the polar surface of LDL. The esterified cholesterol (or cholesteryl ester) is the inner portion and makes up the hydrophobic (non-polar core). Esterified cholesterol is cholesterol with its hydroxyl group esterified. Esterification of cholesterol ensures efficient transportation of cholesterol because more cholesterol can be packed into the non-polar core when it is esterified **(b) (Top) Representation of ApoB-100 of LDL**(*26*) The large circle represents the lipids portion of LDL (i.e. the lipid core), 70% of which is cholesterol. ApoB is believed to wrap around the lipid core as shown. The dark circles represents the cysteine residues and the unshaded circles on ApoB represent the N-glycosylated carbohydrates. Part of ApoB is exposed to the surface and part of ApoB is buried in the lipid core. **(Below)** (*27*) Another representation of ApoB. The LDLR binding region is believed to be included in residues 3000 to 4000.

a


Phospho-
lipid
Unesterfied
cholesterol
Cholesteryl
ester
Polar
surface
Apolar
core
Apolipoprotein B
LDL

b


N terminus
Surface region
C terminus
1297
4100
CHO
Cysteine
Core lipids
1700
Core region
3070
3249
1
3000
4000
4536
NH2
COOH
LDLR-binding region

### 1.2.6 Proteins that normally control LDL-C levels: HMG CoA reductase, LDLR, ApoB, PSCK9, ARH, and IDOL

Cholesterol is a highly regulated molecule, especially within cells. Proteins that normally control plasma LDL Cholesterol (LDL-C) levels are: HMG CoA reductase, Low Density Lipoprotein Receptor (LDLR) (*18*), Apolipoprotein B-100 (ApoB) (*28*), PCSK9 proprotein convertase subtilisin/kexin type 9 (PSCK9) (*29*), LDLR adaptor protein (LDLRAP1) also called the Autosomal Recessive Hypercholesterolemia gene (ARH), and Inducible Degrader of LDLR (IDOL) (*30*). All of these proteins, except for HMG CoA reductase and ARH will be focused on in this thesis. The downstream effect of defective LDLR and ApoB is increased plasma LDL-C and total cholesterol levels (*18, 31*), while the downstream effect of defective PCSK9 function is decreased plasma LDL-C and total cholesterol levels (*29, 30*). Section 1.2.2 discussed HMG-CoA reductase, while section 1.4 will discuss the other proteins. **Figure 1.7** is a visual representation of how all six proteins interact to affect LDL metabolism.

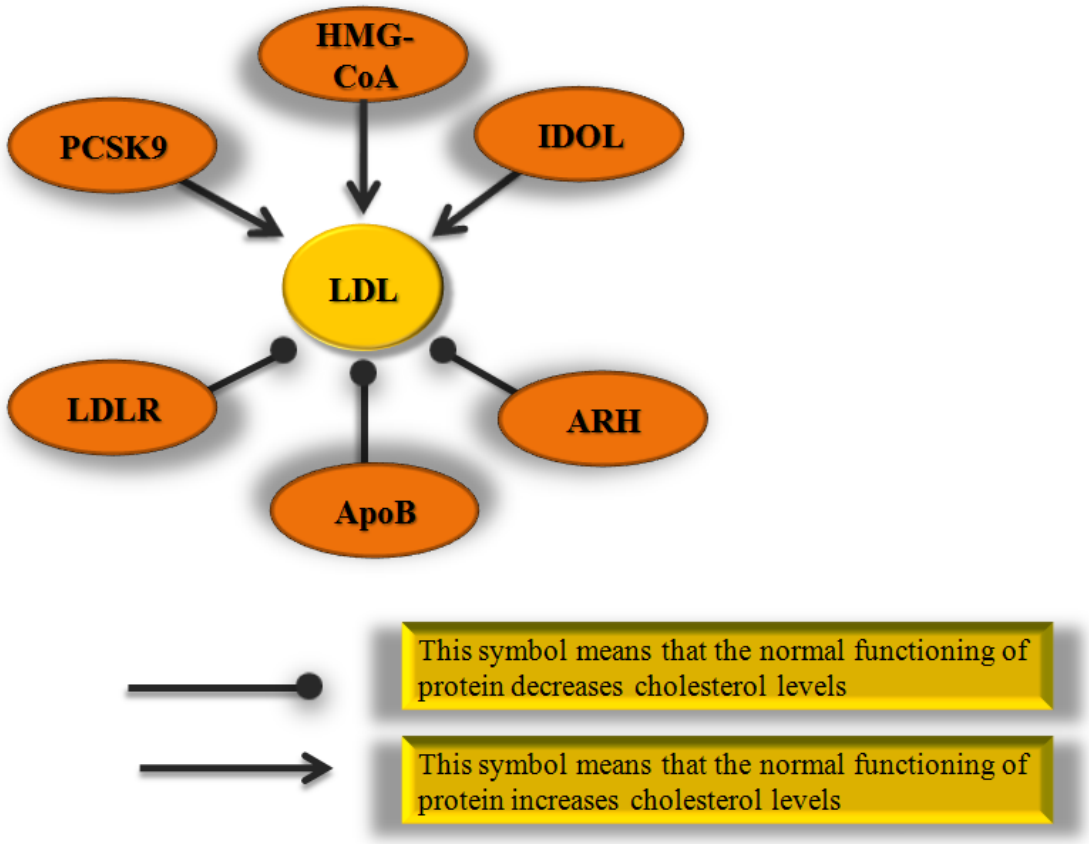



**Figure 1.7 Schematic of how LDLR, ApoB, PCSK9, ARH, HMG CoA and IDOL affect cholesterol levels.** Normal functioning of HMG-CoA reductase (written as HMG-CoA in the diagram), IDOL and PCSK9 increase LDL-cholesterol levels. Normal functioning of LDLR, ApoB and ARH decrease LDL-cholesterol levels.

## 1.3 Dyslipidemias

Dyslipidemia is the term used to describe abnormal levels of lipids in the blood. As alluded to in Section 1.2, lipids are macromolecules that are vital to many functions within the human body, and show a range of normal levels in the blood. So, just like water, too much or too little cholesterol can be pathogenic. Most pathogenic human dyslipidemias are hyperlipidemias, so literature focuses on hyperlipidemia (*32*).

### 1.3.1 Fredrickson's classification of hyperlipidemia

Lipids and lipoproteins have been described in detail in section 1.2. The Fredrickson classification system describes the various hyperlipidemias that can affect patients based on the lipoprotein that is increased in the plasma. The Fredrickson classification says nothing about the etiology of the phenotype; it simply describes the phenotype based on the pattern of lipoprotein elevation. So, a particular Fredrickson phenotype may result from multiple genetic defects. The Fredrickson scheme does not include a description of any human hypolipidemia. **Table 1.4** summarizes the Fredrickson classification (*33*). Because an individual's plasma lipid levels normally rise after a meal, hyperlipidemia is defined after a 12-16 hour fast (*33*) (*34*). The World Health Organization (WHO) uses the Fredrickson classification to describe hyperlipidemia (*34*). Most, but not all human hyperlipidemias are described using the Fredrickson system (*34*).

In **Table 1.4**, the word "Familial Hyperlipidemia" is used. However, usage of the word 'Familial' needs to be clarified here. In the field and colloquially, 'Familial' often implies a monogenic etiology (*35*). However, familial does not necessarily need to be monogenic because the familial hyperlipidemias in Table 1.4 are primary

hyperlipidemias in that they cluster in families but do not necessarily follow a specific Mendelian pattern of inheritance (*35*).

Another way of classifying hyperlipidemia is using "primary" or "secondary" nomenclature. Primary hyperlipidemias usually result from a genetic defect of some sort, while secondary hyperlipidemias result from other existing diseases in an individual. For instance, a poor diet, excessive alcohol intake, obesity, diabetes, thyroid disease, liver disease, kidney disease, autoimmune disease and certain medications such as corticosteroids or drugs that target the human immunodeficiency virus, can each cause secondary hyperlipidemia (*34*).

**Table 1.4** Fredrickson classification of hyperlipidemia.

| **Fredrickson classification** | **Familial Hyperlipidemia** | **Elevated Lipoproteins** | **Elevated Lipids** |
|---|---|---|---|
| Type 1 | Chylomicronemia | CM | TG |
| Type 2A | Hypercholesterolemia | LDL | TC |
| Type 2B | Combined hyperlipidemia | VLDL, LDL | TC, TG |
| Type 3 | Dysbetalipoproteinemia | IDL | TC, TG |
| Type 4 | Hypertriglyceridemia | VLDL | TC |
| Type 5 | Mixed Hyperlipidemia | VLDL, CM | TC, TG |

**Note:** Abbreviations: CM, Chylomicron; LDL, Low Density Lipoprotein; VLDL, Very Low Density Lipoprotein; IDL, Intermediate Density Lipoprotein; TC, Total Cholesterol, TG, Triglyceride. Information in this table was taken from Fredrickson *et al* (*33*)

#### 1.3.1.1 Focus on Fredrickson types 2A and 2B (FH and FCH)

Fredrickson Type 2A and Type 2B are the only phenotypes that include hypercholesterolemia in its phenotypic description. The diseases Familial Hypercholesterolemia (FH) and Familial Combined Hyperlipidemia (FCH) are designated as Type 2A (elevated LDL only) and Type 2B (elevated LDL and VLDL both) Fredrickson phenotype classes, respectively. The disease Familial Hypertriglyceridemia (FHTG) is identical with the Fredrickson Type 4 (elevated VLDL) phenotype. Previous work from our lab has shown that the FCH and FHTG have a common genetic etiology for elevated triglyceride (TG), which is a combination of many single nucleotide polymorphisms (or SNPs) that each contribute a small amount to raise levels of, but which cumulatively act to raise TG to a clinically relevant level. My hypothesis was that FCH may be a condition that is due to the co-existence of FH (a disease that is due to rare mutations in the LDLR that raise LDL-C levels) and FHTG (a disease that is due to common polymorphisms that raise TG levels) (**Figure 1.8**).

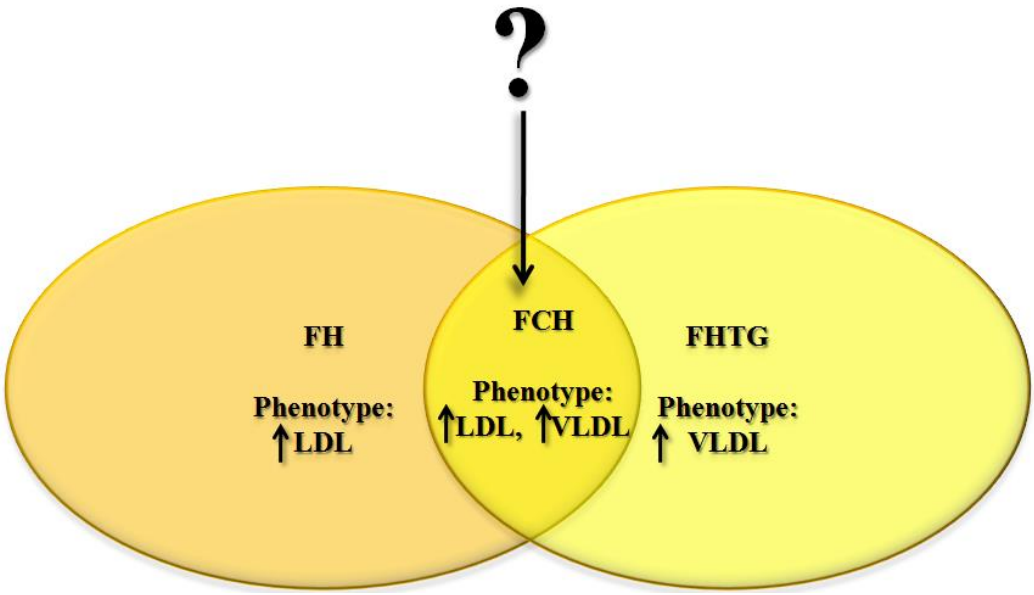



**Figure 1.8 Venn diagram suggesting a genetic model for Familial Combined Hyperlipidemia (FCH).** Elevated LDL occurs in FH and elevated VLDL occurs in FHTG. Elevated LDL and elevated VLDL occurs in FCH. Previous work from Hegele lab has shown that the TG (essentially VLDL) elevation in FCH and FHTG is due to similar genetic etiology, namely the accumulation of many common SNPs, each with a small effect on risk. So I hypothesized that FCH could reflect the coexistence of FH (rare variants that raise LDL-C) and FHTG (common variants that raise VLDL).

### 1.3.2 Role of genetics in dyslipidemia

#### 1.3.2.1 Genetic Variation and Disease

Most complex diseases have both an environmental etiology and genetic etiology. The ratio of environmental factors to genetic factors in the causation of a disease varies from disease to disease and from individual to individual. There are various types of genetic variation. I define genetic variation can be defined as any change in the DNA sequence in an individual's DNA. This change in sequence can be brought about, for instance, either by small substitutions of nucleotide or by larger gains or loss of DNA, such as insertions and deletions. Sometimes the word 'genetic variation' is considered to be synonymous with 'mutation'. The meaning of the word mutation depends on the context in which it is used. In a clinical setting, mutation usually refers to a rare genetic variation that leads to a dysfunctional gene product, which consequently leads to lack of wellbeing. So, in clinical terms, every mutation is a genetic variation but not every genetic variation is a mutation. The definition of mutation in this thesis is that of the clinical setting, in which it refers to a rare molecular event that likely affects the normal function of the gene product and could lead to an abnormal phenotype or disease.

Cystic Fibrosis (CF) and phenylketonuria (PKU) are examples of diseases that result mainly from genetic variation in a single gene that is sufficient to cause the disease. These are referred to as "monogenic diseases" and are individually rare in the population. For instance, the prevalence of CF in North America is about 1 out of 2500 live births, which makes CF the most common recessive disorder in individuals of European descent (*36*). Phenylketonuria is also another monogenic disorder (*37*) Cystic fibrosis (CF) is an autosomal recessive disease where individuals have a

homozygous (or compound heterozygous) mutation in the cystic fibrosis transmembrane conductance regulator gene (*CFTR*). In CF, individuals have pathological changes in tissues that express *CFTR* including secretory cells, sinuses, lungs, pancreas and reproductive tracts. CF is most pronounced in the airways. The deletion of the amino acid phenylalanine at amino acid position 508 occurs in more than half of Caucasian CF cases (*38*) (*36*).

Similarly, PKU is a rare disorder whose prevalence varies from population to population but generally affects ~1 in 10000 live births. Even so, PKU is the most common inborn error of amino acid metabolism (*37*). PKU is an autosomal recessive disease where individuals do not metabolize the amino acid phenylalanine, which leads to elevated levels in the blood and toxic levels in the brain (*37*). The genetic defect is in the phenylalanine hydroxylase gene (*PAH*) and missense mutations occur in majority of cases in PKU.

Identifying causal mutations in a monogenic disease helps unraveling pathways in various biochemical processes. For instance, identifying the casual LDL receptor gene (*LDLR*) mutation in FH (*39*) increased the understanding of the biochemistry of cholesterol metabolism through LDLR and knowledge of receptor mediated endocytosis, which is a mechanism used by many proteins and not just LDLR.

Diseases can be categorized, based on their prevalence in the population, as common diseases or rare diseases. According to the World Health Organization (WHO) a rare disease, also known as an orphan disease, is a disease defined as affecting individuals at a frequency of 65 to 100 in 100,000 (*40*). In contrast, a common disease can be defined as a disease that is found much more frequently in the

population, like cancer and CVD, which together affect one-third to one-half of all people.

There have been various models for explaining genetic etiology of common diseases. The Common Disease-Common Variant (CDCV) Hypothesis is a model whereby common diseases are said to result from accumulation of a moderate number of common variants, each of which contributes to a certain small percent of the disease risk (*41*). However, common variants identified by GWAS so far explain only a small portion of the genetic component of most common diseases, which has given rise to the 'missing heritability' problem.

In the field of genomics/genetics, the missing heritability problem essentially refers to the proportion of genetic susceptibility that is not explained by GWAS identified loci such as common variants (*41*). To solve the missing heritability problem, the missing genetic component has been attributed to either of three models, namely, (i) the infinitesimal model, (ii) the rare allele model and (iii) the broad sense heritability model.

In the infinitesimal model, the genetic component is explained by numerous amounts of common variants (each of small effect size) each contributing a small percent to disease risk. In the rare allele model, the genetic component is explained by many rare variants (each of large effect size) each contributing to disease risk. In the broad sense heritability model, the genetic component is explained by a combination of environmental, genotypic and epigenetic interactions. Also, some propose that other mechanism, such as gene X gene interactions, gene X environment interactions or epigenetic factors can help explain the "missing heritability" for many common diseases.

GWAS are not sufficiently powered to detect variants under any of these three models. (*41*). So, researchers have a choice of which model best fits their hypothesis/ experiments as none of the models has been shown to be better over the other (*41*).

#### 1.3.2.1.1 Common Variants

Common variants are defined as variants with a minor allele frequency (MAF) of greater than 5% in the general population (*42*). There are various types of common variants including SNPs. SNPs have been used for studying complex diseases/traits such as dyslipidemia (*43*), but also many other traits. SNPs have been used to identify loci that were later discovered to be involved in lipid metabolism (*44*). Common variants that predispose individuals to disease are normally non-disease causing in and of themselves.

The effect size is the statistic that refers to the magnitude of an effect such as magnitude of regression coefficient and mean difference (*45*). Effect size can be represented as an Odds Ratio. The Odds Ratio is the ratio the odds of an event occurring case cohort to odds of an event occurring in the control cohort (*45*). The effect size of a variant essentially refers to the increase in risk that is conferred by the variant. Common variants usually have small effect sizes. Common variants do not cause disease; they increase susceptibility to disease. However, there are exceptions; for instance, a common SNP is associated with a huge risk (odds ratio of almost 7-fold) for developing the eye disease macular degeneration (*46*). However, the effect sizes for most SNPs so far discovered in GWAS of common diseases and phenotypes ranges from 1.2- to 2-fold.

#### 1.2.3.1.2 Rare Variants

A rare variant is defined as a variant with a minor allele frequency (MAF) of less than 1% in the population; variants with a MAF of 1%-5% are called uncommon variants (*42*). Mutations (defined in section 1.3.2.1) are usually rare variants. Mutations can often have large effect sizes. However, there are exceptions, because next generation sequencing of the genomes of completely healthy people has now revealed hundreds of thousands of new rare variants, but because most of these people were essentially healthy, all these mutations cannot be assumed to have large effects on disease risk. Nevertheless, the general notion is that mutations are rare variants that can have large effects on disease risk.

Under the rare variant model (which was briefly described in Section 1.3.2.1), mutations are likely to be rare variants because of evolutionary theory. In evolutionary theory, variants that are deleterious to fitness such as disease causing variants (i.e. mutations) are selected against and therefore cannot be common in the population. This phenomenon is called purifying selection and it refers to selection against fitness-reducing variants (such as disease causing variants) such that their frequency is kept low in the population or even eliminated in the population. So, existence of rare variants in the population is a balance between purifying selection and high mutation rates that give rise to susceptibility variants, such that the balance leads to the frequency of such variants being 1% or slightly more if it has a moderate effect on fitness. Evolutionary theory is one of the strongest supports for the rare allele model (*41*). The hypothesis of our resequencing study, which was our first study, (explained in Section 1.8.1 of the thesis) was made in light of the rare allele model.

#### 1.3.2.2 Monogenic dyslipidemias

Monogenic dyslipidemias refer to dyslipidemias in which the genetic etiology can be narrowed down to one gene. Thus monogenic dyslipidemias typically follow a Mendelian pattern of inheritance. **Table 1.5** gives a list of dyslipidemias that are considered to follow a Mendelian pattern of inheritance(*47*) their OMIM numbers (i.e. Online Mendelian Inheritance in Man in The NCBI database) have also be given for reference. FH is a classic example of a monogenic dyslipidemia. However, recent research in the field has shown that an alternate etiology for a proportion of cases of FH – perhaps 20%, can actually be polygenic (*48*).

#### 1.3.2.3 Polygenic dyslipidemias

Polygenic dyslipidemia refers to dyslipidemia where multiple genes contribute to the disease phenotype. Polygenic dyslipidemia would fit the infinitesimal model (which was briefly described in Section 1.3.2.1 of the thesis). Diseases where etiology cannot be narrowed down to a single or few genes are by default categorized as polygenic. However, not finding a single particular gene (or few genes) involved in disease etiology may sometimes reflect the science or technology involved in unraveling disease etiology.

**Table 1.5** Brief descriptions of monogenic dyslipidemias and their OMIM number for reference

| Monogenic Dyslipidemia | Gene | OMIM number |
|---|---|---|
| 1-Autosomal dominant Familial Hypercholesterolemia due to defective | *LDLR* | 606945 |
| 2-Autosomal dominant Familial Hypercholesterolemia due to defective | *APOB* | 107730 |
| 3-Autosomal dominant Familial Hypercholesterolemia due to defective | *PCSK9* | 607786 |
| 4-Autosomal Recessive Familial Hypercholesterolemia due defective | *ARH* | 605747 |
| 5- Cholesteryl ester storage disease due to defective | *LIPA* | 613497 |
| 6-Hypobetalipoproteinemia due to defective | *APOB* | 107730 |
| 7-Primary Bile Acid malabsorption due to defective | *SLC10A2* | 601295 |
| 8-Analphalipoproteinemia due to defective | *APOA1* | 107680 |
| 9-Familial LCAT deficiency due to defective | *LCAT* | 606967 |
| 10-FamilialHypoalphalipoproteinemia due to defective | *ABCA1* | 600046 |
| 11-Hepatic Lipase deficiency due to defective | *LIPC* | 151670 |
| 12-Hyperchylomicronemia due to defective | *APOC2* | 608083 |
| 13-Hyperchylomicronemia due to defective | *LPL* | 609708 |
| 14-Dysbetalipoproteinemia due to defective | *APOE* | 107741 |
| 15- Hypobetalipoproteinemia due to defective | *PCSK9* | 607786 |
| 16-Sitosterolemia due defective | *ABCG5* | 605459 |
| 17-Sitosterolemia due to defective | *ABCG8* | 605460 |
| 18-CETP deficiency due to defective | *CETP* | 118470 |
| 19-Abetalipoproteinemia (ABL) due to defective | *MTP* | 157147 |
| 20-Familial Combined Hypolipidemia due to defective | *ANGPTL3* | 604774 |
| 21-Fredrickson's Type V Hypertriglyceridemia due to defective | *APOA5* | 606368 |
| 22- Cholesterol 7α-hydroxylase deficiency due to defective | *CYP7A1* | 118455 |
| 23- Chylomicron retention disease due to defective | *SAR1B* | 607690 |

OMIM, Online Mendelian Inheritance in Man **(http://www.ncbi.nlm.nih.gov/omim/)**
**Abbreviations:**
*LDLR*, low density lipoprotein receptor
*APOB*, Apolipoprotein B
*PCSK9*, proprotein convertase subtilisin/kexin type 9
*ARH*, low density lipoprotein receptor adaptor protein 1
*LIPA*, lipase A
*SLC10A2*, solute carrier family 10 (sodium/bile acid cotransporter family), member 2
*APOA1*, apolipoprotein A-I
*LCAT*, lecithin-cholesterol acyltransferase
*ABCA1*, ATP-binding cassette, sub-family A (ABC1), member 1
*LIPC*, lipase
*APOC2*, apolipoprotein C-II
*LPL*, lipoprotein lipase
*APOE*, apolipoprotein E
*ABCG5*, ATP-binding cassette, sub-family G (WHITE), member 5
*ABCG8*, ATP-binding cassette, sub-family G (WHITE), member 8
*CETP*, cholesteryl ester transfer protein, plasma
*MTP*, microsomal triglyceride transfer protein
*ANGPTL3*, angiopoietin-like 3
*APOA5*, apolipoprotein A-V
*CYP7A1*, cholesterol 7-alpha-hydroxylase
*SAR1B*, SAR1 homolog B

## 1.4 Familial Hypercholesterolemia

### 1.4.1 Characterization, clinical features, diagnosis and clinical genetics of Familial Hypercholesterolemia

In the disease Familial Hypercholesterolemia (FH), individuals have abnormally elevated LDL-C levels. FH typically shows a Mendelian pattern of inheritance, as it is usually a single gene disorder. FH can be inherited in an autosomal dominant manner (in which case is referred to as Autosomal Dominant Hypercholesterolemia) and an autosomal recessive manner (in which case it can be referred to as Autosomal Recessive Hypercholesterolemia). The prevalence of the heterozygous form of FH is 1in 500 and the prevalence of the homozygous form of FH is 1 in 1,000,000 (*49*).

In FH, there is reduced clearance of LDL-C from plasma because of defective activity of LDLR. FH heterozygotes have a 2- to 3-fold increase in LDL cholesterol levels. Approximately half of FH heterozygotes develop tendon xanthomas, xanthelasmas, premature corneal arcus and CHD by the $4^{th}$ or $5^{th}$ decade of life. FH homozygotes have a 5- to 8-fold increase in LDL-C levels and develop CHD in the $2^{nd}$ decade of life.

The way FH is diagnosed can vary from clinician to clinician, but there are standard clinical methods for diagnosis mainly that involve observation of elevated plasma LDL-C levels (that is unexplained by secondary causes), a personal history or family history of CHD or myocardial infarction, and the presence of xanthomas on physical examination (*50*).

### 1.4.2 Genetic Etiology of Familial Hypercholesterolemia

Autosomal Dominant Hypercholesterolemia (ADH) can be caused by a mutation in any one of three genes, namely *LDLR* (Low Density Lipoprotein Receptor gene), *APOB* (Apolipoprotein B-100 gene) and *PCSK9* (Proprotein Convertase Subtilisin/kexin Type 9 gene). Autosomal Recessive Hypercholesterolemia (ARH) is caused by a mutation in the two copies of the LDLR gene, or the Autosomal Recessive Hypercholesterolemia gene (*ARH*). The autosomal dominant form of FH is much more common than the autosomal recessive form of FH. The frequency of FH due to *LDLR*, *APOB*, *PCSK9* and *ARH* is 52%-76%, 2-10%, 2% and 2%, respectively. Thus, mutation in *LDLR* is the most common cause of heterozygous FH (*49*).

#### 1.4.2.1 Low Density Lipoprotein Receptor gene *LDLR*: Structure and function

*LDLR* codes for the Low Density Lipoprotein Receptor (LDLR) and this gene is mainly expressed in the in liver. LDLR is an ~860-amino acid cell surface glycoprotein (*51*). The most important physiological ligand for the LDLR is LDL, which carries ~70% of cholesterol in humans (*51*) (LDL has been described in Section 1.2.5). LDLR plays an important role in cholesterol homeostasis because its main function is to clear LDL from plasma.

LDLR has five domains, namely: (i) Ligand Binding Domain, (ii) EGF Precursor Homology, (iii) O-linked Sugar Domain, (iv) Membrane Spanning Domain and (v) Cytoplasmic Domain. The schematic figure of mature LDLR (i.e. LDLR with the 21 amino acid signal peptide sequence removed) is shown in **Figure 1.9** (*39*) shows the structure of cholesterol. Figure 1.3 shows how LDLR is involved in clearance of LDL from blood.

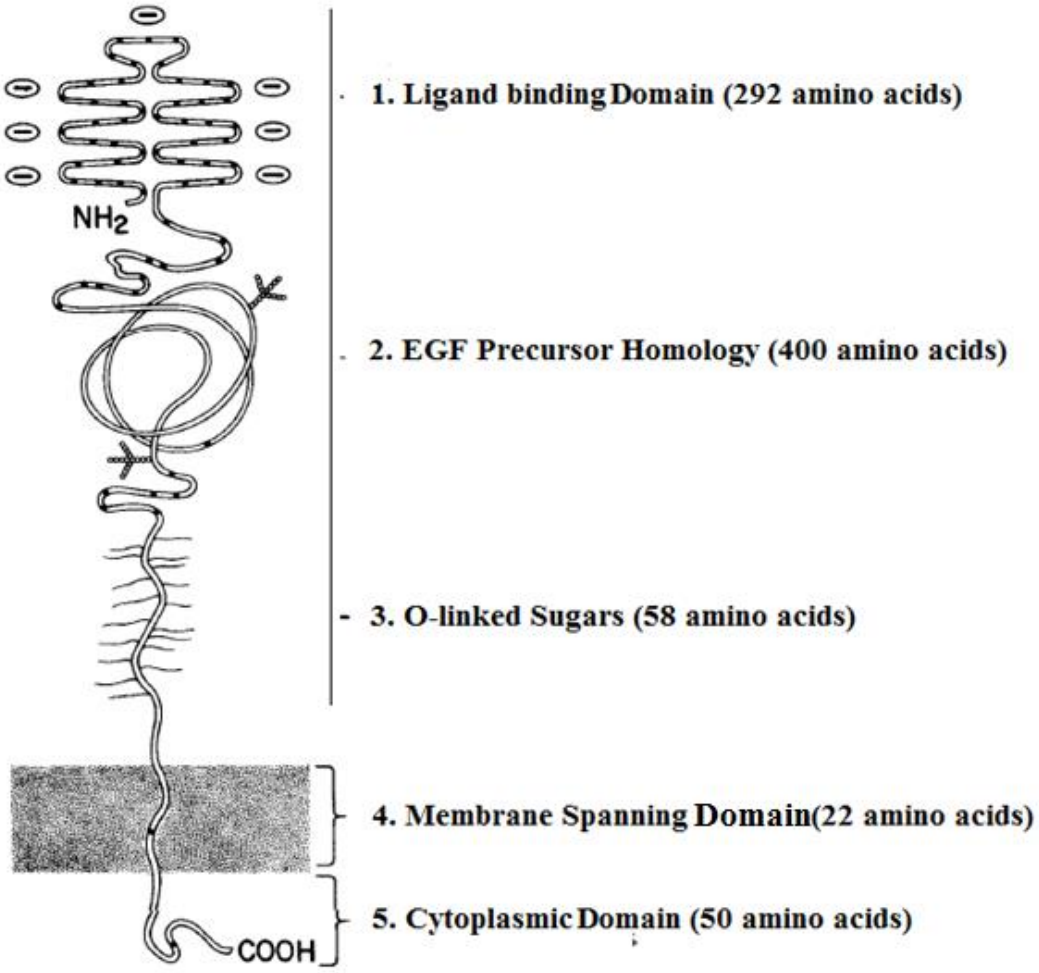
The first 21 amino acids of LDLR is the signal peptide sequence which gets cleaved soon after the protein is translated. The 21 amino acid signal directs location of translation on the membrane of the Endoplasmic Reticulum (ER) as the protein is getting translated (*39*).

The ligand binding domain as its name implies, is the domain that binds LDL. It is a cysteine rich domain and many disulphide bonds are present in this domain. This extensive disulphide structure gives this domain stability. This domain is negatively charged and this negative charge is complementary to the charge of regions of apo E (*39*). The second domain is the Epidermal Growth Factor (EGF) precursor homology domain and the name was given because this domain of LDLR resembles the part of the extracellular domain of EGF (*39*). The third domain is called O-linked sugar domain because of the clustering of O-linked sugar chains (*39*). The fourth domain, which is the membrane spanning domain, is rich in hydrophobic amino acid residues so that it can interact with the hydrophobic cell membrane (*39*). The fifth domain is the cytoplasmic domain and this domain is important for the clustering of LDLR in clathrin-coated pits that occurs in LDL clearance by LDLR. LDLR clusters and then internalizes LDL. This clathrin is important for this clustering of LDLR (*39*).

In clearance of LDL from blood (**Figure 1.3**), receptor-ligand complexes occurs in clathrin coated pits. LDLR clusters in clathrin coated pits. The receptor-ligand complex is internalized into the cell within the clathrin coated pits. This complex is delivered to the endosomes where the pH is low (i.e. acidic). At this low pH, the receptor dissociates from the ligand; the receptor gets recycled back to the cell surface and the ligand moves from the endosome to the lysosome. In the lysosome, LDL is hydrolysed and cholesterol is released into the cell (*51*). This entire process is referred to as receptor-mediated endocytosis.

A loss-of-function mutation in the *LDLR* gene leads to less or no clearance of LDL, thus resulting in hypercholesterolemia. For the LDLR, mutations have been classified into 4 classes (*39*). In Class I mutations, no receptor is synthesized due to a major deletion mutation that results in no protein product expressed. In Class II mutations, LDLR is synthesized but does not undergo its normal post translational modification, which occurs in the ER. So, LDLR does not get to the cell surface; LDLR remains in ER until it is degraded. In Class III mutations, LDLR is synthesized and reaches the cell surface but fails to bind its ligand. In Class IV mutations, LDLR is synthesized, reaches cell surface, binds to LDL but fails to cluster and clustering of LDLR (which occurs in clathrin coated pits), is vital for receptor mediated endocytosis. So, for LDLR, the mutations are classified based on the aspect of receptor mediated endocytosis it is affecting and not the type of mutation. The downstream effect of all the classes is reduced clearance of LDL. *LDLR* mutations are the most common cause of FH as alluded to earlier.

**Figure 1.9 Structure of LDLR (a)** The multidomain LDLR. LDLR has five domains. The Ligand binding domain is for binding of LDL and the cytoplasmic domain is important for internalization of LDLR –LDL complex into the cell. The membrane spanning domain is rich is hydrophobic residues so that it can interact with the hydrophobic layer of the cell membrane. The EGF (Epidermal Growth Factor) precursor homology domain is homologous to part of EGF. **Figure 1.9** was modified from (*39*)**. Figure 1.3** shows the mechanism of LDLR action. The main function of LDLR is to clear LDL from the blood. LDLR clears LDL from the blood through receptor mediated endocytosis. In receptor mediated endocytosis, LDLR binds LDL. The LDLR-LDL complex gets internalized into the cell. The LDLR-LDL complex enters the endosome. In the endosome the pH drops which dissociates the LDLR-LDL complex. LDLR gets recycles back to the cell surface, while LDL is transported to the lysosome and gets metabolized in the lysosome. In the lysosome, cholesterol is released from LDL.


NH2
COOH
1. Ligand binding Domain (292 amino acids)
2. EGF Precursor Homology (400 amino acids)
3. O-linked Sugars (58 amino acids)
4. Membrane Spanning Domain(22 amino acids)
5. Cytoplasmic Domain (50 amino acids)

### 1.4.2.2 Apolipoprotein B-100 gene *APOB*: Structure and function

*APOB* codes for both Apolipoprotein B-48 (ApoB-48) and Apolipoprotein B-100 (ApoB-100). The difference between the two isoforms of the protein is that the ApoB-48 isoform, which is 48% of the length of the full-length ApoB-100 isoform, is expressed in the small intestine. In contrast, the ApoB-100 isoform is expressed in the liver. ApoB is the ligand that LDLR recognizes in LDL (*52*) (Apolipoproteins and LDL were discussed in Section 1.2). ApoB, which is one of the largest monomeric proteins known (*53*) is the apolipoprotein of LDL (*52*). Because of the insoluble nature of ApoB, it has been difficult to completely study its tertiary structure (*26*) (*53*). So researchers have used experimental and *in silico* data to predict the three dimensional structure of ApoB. **Figure 1.6b** is also a diagram for ApoB structure. Some cases of FH result from a perfectly normal LDLR but a mutation in *APOB*; these mutations disrupt the ApoB structure such that it cannot bind with LDLR and thus clearance of LDL from plasma gets disrupted (*27*). *APOB* mutation is the second most common cause of FH as alluded to earlier.

**Figure 1.6b** (Top) was modified from (*26*) and **Figure 1.6 b** (Bottom) was modified from (*27.*) ApoB, which is made up of 4536 amino acids, is believed to wrap around the spherical lipid core The Receptor binding region of the ApoB protein is believed to span amino acid residues 3000 to 4000 (*27*) .

### 1.4.2.3 Proprotein Convertase Subtilisin/kexin Type 9 gene *PCSK9*: Structure and function

*PCSK9* codes for Proprotein Convertase Subtilisin/kexin Type 9 protein (PCSK9). PCSK9, which is a 692 amino acid protein that normally degrades LDLR.

A loss of function mutation in PCSK9 leads to less degradation of LDLR and thus increases clearance of LDL by LDLR. A gain of function mutation in PCSK9 leads to less clearance of LDL by LDLR. Thus, gain of function mutations are causative of Familial Hypercholesterolemia (*54*, *55*).

The mature PCSK9 is shown is **Figure 1.11a** PCSK9 has three domains, namely the prodomain, the catalytic domain and lastly the C-terminal domain. PCSK9 binds the cell surface LDLR by directly interacting with the EGF Precursor Homology Domain of LDLR and PCSK9 binds LDLR with its catalytic domain (*56*). This interaction of PCSK9 with LDLR targets LDLR for lysosomal degradation (*57*).

#### 1.4.2.4 Autosomal Recessive Hypercholesterolemia gene *ARH*: Structure and function

ARH (Autosomal Recessive Hypercholesterolemia protein), also called LDLR associated protein (LDLRAP1) interacts with the cytoplasmic tail of LDLR. The Phosphotyrosine binding (PTB) domain of ARH interacts with the NPXY consensus in LDLR cytoplasmic tail domain and the consensus of NPXY is required for internalization of LDLR-ligand complex (*58*). ARH, which is a 308 amino acid protein, is required for the internalization of LDL-LDLR complex and ARH is haploinsufficient; mutation in both copies of ARH causes FH (*59*). **Figure 1.11b** is a simple diagram of ARH and was modified from (*58*). As mentioned earlier, autosomal recessive mutation in ARH is fourth most common cause of FH.

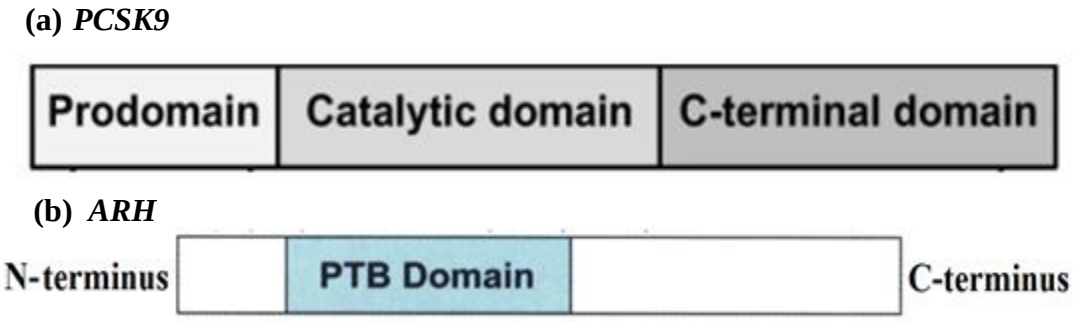



**Figure 1.10 Structure of PCSK9 and ARH (a)** Simple diagram of PCSK9. From N-terminus to C-terminus, PCSK9 has the prodomain, the catalytic domain and the C-terminal domain. PCSK9 degrades LDLR through its interaction with its catalytic domain. The catalytic domain of PCSK9 interacts with EGF Precursor homology of LDLR Figure 1.11a was taken and modified from Abifadel *et al* **(*55*) (b)** Simple Diagram of ARH. ARH has a PTB (phosphotyrosine binding) domain. This PTB domain interacts with the NPXY consensus sequence of the cytoplasmic tail domain if LDLR and this interaction is necessary for the internalization of LDLR-LDL complex into the cell. Figure 1.11b was modified from (*58*)

### 1.4.2.5 Role of polygenic susceptibility in FH - Humphries 2013 *Lancet* paper (*48*)

A minority of FH patients – perhaps 10 to 20% depending on the population - do not have mutations in any of the four known FH–causing genes (i.e. *LDLR*, *APOB*, *PCSK9* and *ARH*). These patients are known as Mutation Negative Familial Hypercholesterolemia (FH/M-ve) patients (*48*).

According to the Global Lipid Genetics Consortium (GLGC), there are 37 SNPs that have been shown to be associated with LDL-C levels (*43*). Researchers (*48*) used 12 of the 37 SNPs to test for accumulation of risk alleles in FH/M-ve patients to see if FH/M-ve could have an alternate polygenic etiology (*48*). Researchers tested this hypothesis in the British population and found that FH/M-ve patients have a greater accumulation of the 12 GLGC-identified risk alleles than in FH patient with mutation in any of the four known FH-causing genes. Thus FH can also be a polygenic disease, particularly when mutations in the genes that cause the monogenic form are absent (*48*).

### 1.4.2.6 Role of *APOE* in FH

Recent studies have shown that a mutation in the apo E gene (*APOE)* segregates with Familial Hypercholesterolemia (*60*). Marduel *et al* (*60*) was the first to report segregation of an *APOE* deletion mutation, the *APOE* Leu 167 del mutation. The *APOE* Leu 167 del mutation was also found in another family where the mutation segregated with FH (data yet to be published). In both cases, the *APOE* mutation segregated with Autosomal dominant form of FH.

### 1.4.3 Treatment of Familial Hypercholesterolemia

#### 1.4.3.1 Statin treatment: Brief explanation of cholesterol lowering by diet, statins and other drugs - ezetimibe, bile acid sequestrants

Hypercholesterolemia can be reduced by lowering the amount of cholesterol in the diet. Hypercholesterolemia can also be reduced by exercise (*61*). This change in lifestyle is often the first step in lipid lowering. As a result of knowledge of cholesterol metabolism, certain drugs have been developed to lower cholesterol. Acetyl CoA is a precursor in cholesterol biosynthesis and statins are structural analogues of acetyl CoA. So, statins lower cholesterol by preventing HMG CoA reductase from metabolizing its natural substrate (*61*). Statins also lower cholesterol because the effect of reduced cholesterol biosynthesis is upregulation of *LDLR* such that more LDLR are expressed on the cell surface (*61*).

There are other drugs that lower cholesterol by targeting various aspects of cholesterol biochemistry. For instance, ezetimibe blocks the absorption of cholesterol from the small intestine (*61*) and bile acid sequestrants increase the conversion of cholesterol to bile acids thus depleting liver cholesterol levels leading to upregulation of the LDLR (*62*).

#### 1.4.3.2 Kaplan-Meier survival curve for Familial Hypercholesterolemia

Kaplan-Meier curves are survival curves that can display differences due to various interventions carried out in the population. Survival curves show rates of

survival (i.e. percent of patients that are living) of individuals as a function of time. **Figure 1.12** shows a Kaplan-Meier survival curve for FH patients on statin and those not on statin treatment. The FH patients on statin treatment had much greater survival rates than FH patients who were not on statin treatment (*63*). Thus Kaplan-Meier survival curves emphasize the importance of early treatment with statins leading to better prognosis.

### 1.4.4 *IDOL, PSMD9* and cholesterol metabolism

Expression of LDLR is highly regulated. Myosin regulatory light chain interacting protein (IDOL) has been identified as an Inducible Degrader of LDR (IDOL) protein; thus IDOL is sometimes referred to as IDOL. IDOL degrades LDLR through a pathway independent of PCSK9. IDOL mediates ubiquitination and degradation of LDLR (*30, 64*). Thus a gain of function mutation in *IDOL* will lead to less LDL clearance and loss of function mutation will lead to more LDL clearance. This function makes IDOL a suitable gene to study in diseases where hypercholesterolemia occurs such as FH and FCH.

The gene for Proteasome Modulator 9 **(***PSMD9*) is localized 12q24 (*65*). Linkage of *PSMD9* locus to primary hypercholesterolemia while studying a rare family and suggested that this locus be tested in disease where primary hypercholesterolemia occurs. Thus PSMD9 is also a candidate gene for hypercholesterolemia based on chromosomal localization.

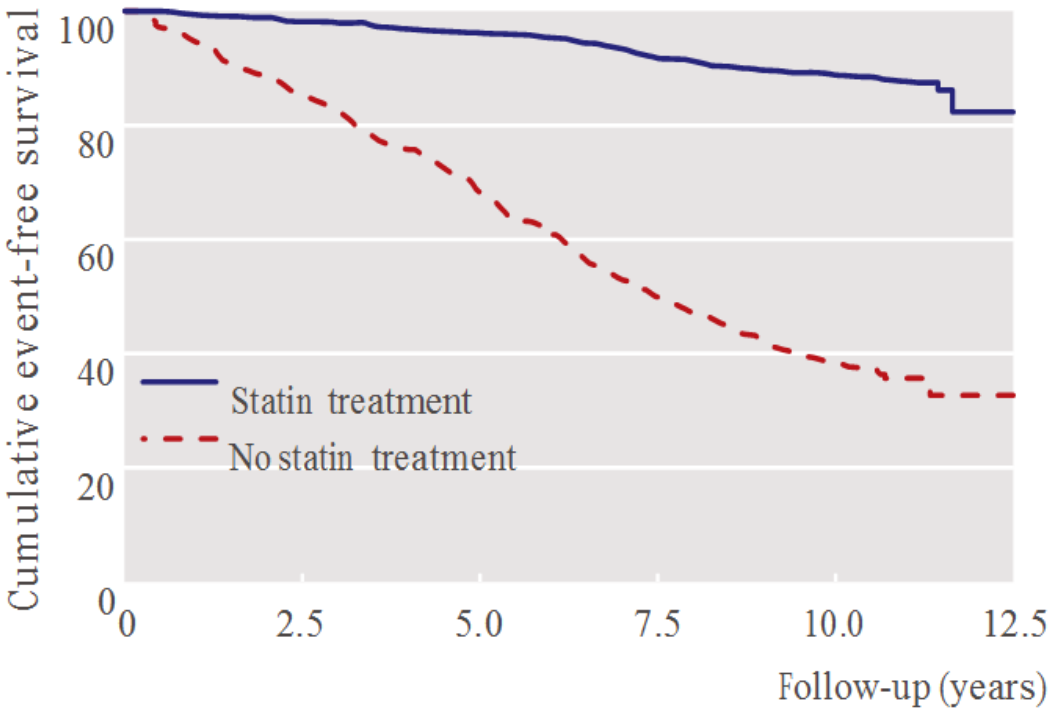



**Figure 1.11 Kaplan-Meier Survival Curve for FH patients on statin treatment and not on statin treatment** (*63*). The survival rate is represented on the vertical axis. The survival rate essentially shows the percentage of FH patients that are still living over the course of time (which is represented on the horizontal axis). FH patients on statin treatment have much better survival rates over time compared to FH patients who are not statin treatment. (*63*).

### 1.5 Familial Combined Hyperlipidemia

FCH has a prevalence of 1% in the Western population, making it the most common genetic dyslipidemia. Because of the hypercholesterolemia that is characteristic of FCH, having FCH puts an individual at risk for CVD; FCH has been estimated to occur in 20% of individuals with Coronary Heart Disease (CHD), which is a form of CVD (*66, 67*).

Familial Combined Hyperlipidemia (FCH) disease was characterized by an affected proband having both elevated plasma cholesterol and triglyceride levels due to elevated LDL and VLDL, respectively (*68*). Other affected family members can either have elevated cholesterol, or elevated TG or both, thus making FCH distinct from FH and Familial Hypertriglyceridemia (FHTG). FCH is distinct from FH and FHTG because (i) in FH, a clear vertical pattern of inheritance is observed and triglyceride levels are normal in affect family members and (ii) in FHTG, cholesterol levels are always normal in affected family members (*68*). Interestingly, it was suggested that the primary metabolic defect in FCH is in TG metabolism with secondary effects on cholesterol metabolism and that FCH could be a monogenic disorder.

The genetic etiology of FCH is currently considered to be polygenic (*69*). Previous work from our the Hegele lab has explained genetic basis for primary hypertriglyceridemia, including FCH and FHTG, with a combination of both common variants and rare variants (*70, 71*) explaining 42% of variation in HTG.

The genetic etiology of hypercholesterolemia in FCH is poorly understood. Many genes have been implicated through large linkage association studies (*66*), yet the genetic basis for hypercholesterolemia is still unknown. Part of this thesis focused

on unraveling the genetic etiology of hypercholesterolemia in FCH, more so hypercholesterolemia is a classical risk factor for CVD development.

### 1.6 *DIET1* and dyslipidemia

*DIET1* was first discovered in mice over a decade ago. The gene is located on chromosome 2 in the mouse genome (*72*). *DIET1* has not been fully annotated in the human genome. Recently, *DIET1* has been shown to affect lipid metabolism in mice and cultured human cells (*73*). Thus the gene is now becoming a subject of interest in the field (*74*).

*DIET1* was somewhat serendipitously discovered. A particular strain of mice was used as a model to study genes involved in lipid metabolism. *DIET1* was discovered when a *de novo* (i.e. spontaneous) mutation occurred in this particular strain of mice and gave rise to a new strain of mice (*72*). So, there was now the old strain (i.e. the original strain) and a new strain that was phenotypically different. When both the old strain and new strain were given an atherogenic diet (i.e. a diet high in cholesterol), the old strain of mice showed increased cholesterol levels and formation of atherosclerosis but the new strain did not. The new strain did not show any increased cholesterol levels and did not develop atherosclerosis. So the phenotype of this new strain was described as ‘resistant to diet-induced hypercholesterolemia and atherosclerosis’ (*72*).

Because both the old and new strains of mice had extensive genetic identity, the new phenotype had to be a result of a gene or a few genes. Using various genetic approaches, such as genetic crosses and linkage analyses, a mutation in DIET1 in mice was associated with the phenotype (*72*)

Further studies were aimed at understanding the metabolic role and gene expression of *DIET1* (*13*). Phan *et al* used the genetic approach of gene expression profiling to unravel the metabolism of *DIET1*. *DIET1* is involved with increased bile acid synthesis and excretion. Thus the mutation at the *DIET1* locus caused the new strain of mice to be resistant to diet-induced hypercholesterolemia and atherosclerosis because plasma cholesterol concentration decreases when there is increased bile acid synthesis and bile acid excretion (*13*)

There have been some human studies on *DIET1*. The chromosomal location of *DIET1* in humans is 10p12 and *DIET1* encodes a predicted protein of 2156 amino acids (*73*). Human and mice *DIET1* share 70% similarity (*73*). *DIET1* is expressed in the small intestine in humans and mice (*73*). The role of *DIET1* at the metabolic level in both human and mice has been reported (*73*).

Cultured human and mouse intestinal cells were examined to further explain *DIET1* metabolism (*73*). Bile acid synthesis is controlled by negative feedback regulation. Bile acid synthesis upregulates a protein called fibroblast growth factor (FGF) 15 in mice (or FGF-19 in humans). The downstream effect of FGF-15/19 is inhibited bile acid synthesis (*73*). *DIET1* upregulates FGF-15/19. So mice and cultured human intestinal cells deficient in *DIET1* do not express FGF-15/19 do not have repression of bile acid synthesis. Increased bile acid synthesis and increased bile acid excretion are ways of lowering cholesterol in body. Thus, *DIET1* deficiency leads to hypocholesterolemia (*73, 74*).

*DIET1* was shown to recently be associated with TG levels in mice (data yet to be published, communicated from Professor Karen Reue, Department of Genetics, University of California, Los Angeles). Since the Hegele lab has examined genetic factors accounting for variation hypertriglyceridemia more than any other group (*70,*

*71*, *75*) I wanted to test for associations between common variants in *DIET1* region and hypertriglyceridemia.

### 1.7 Genetic approaches to identifying disease etiology

There are various approaches —in the terms of study design and techniques — to identifying disease etiology in the field of genetics/genomics and the repertoire of approaches increases with advancing technologies. Various designs employed in identifying disease etiology include family studies, case-control studies and population studies. Various techniques used in identifying disease etiology include resequencing and genotyping.

In a family study design, genetic variation is evaluated in family members affected with a particular disease. Various analyses, essentially referred to as pedigree analyses, are performed to identify whether the genetic variation segregates with the disease in the disease affected family. Linkage analyses and autozygosity mapping are ways of analyzing pedigree information. In linkage analyses the LOD Score is essentially measure of likelihood of linkage divided b likelihood of no linkage; LOD score is calculated from pedigree information (*76*). Autozygosity mapping, which is another form of linkage analyses, can be used in identifying genes involved in autosomal recessive disorders (*77*) . These approaches are also used for better understanding genetic etiology of traits that are non-disease related (*78*) such as height and hair colour.

The advent of advanced technologies, such as next generation sequencing, has birthed various genetic/genomic approaches including Genome Wide Association Studies (GWAS) in large epidemiological settings, and resequencing of candidate (suspect) genes in smaller well –defined phenotypic extremes or case-control settings.

So, researchers in the field have a variety of approaches, in terms of study design and technique, to choose from to understand genetic etiologies of diseases. For GWASs, chip-based microarray technology is used for assaying millions of SNPs. Illumina and Affymetrix are two common platforms used for most GWASs (*79*).

Genotyping can be defined as any technique that enables identification of genetic variation in an individual. The difference between genotyping and sequencing is that genotyping requires prior knowledge of the genetic variation so novel genetic variations cannot be discovered. However sequencing does not require any knowledge of prior genetic variation, so novel genetic variation can be discovered. Genotyping can be accomplished by various techniques such as restriction enzyme length polymorphism and TaqMan assays. Sanger sequencing can also be employed for genotyping. When using the resequencing technique, the DNA sequence of the gene is sequenced.

In Genome Wide Association Studies (GWASs), common genetic variants, such as SNPs, are genotyped all across the genome to see if there is an association between genetic variation and a trait (including disease) (*80*). GWASs are employed for identifying susceptibility loci for complex traits (*80*). In GWASs, the difference in frequency of genetic variation between cases and controls is tested to see if the gene variation is associated with susceptibility to the complex disease.

### 1.7.1 Resequencing candidate genes to test for accumulation of rare variants in case-control cohorts

Our context of a mutation is a genetic variation that leads to protein dysfunction which eventually leads to lack of individuals' well-being. So, missense rare variants are likely to be mutations. If rare missense variants occur significantly

more frequently in a gene of a diseased individual relative to the non-disease individual, it indicates that the gene is associated with disease etiology directly or indirectly. Given that functional verification of a genetic variant consumes a lot of time and resources, it is reasonable to verify rare missense variants that significantly accumulate in disease individuals.

Previous work from the Hegele lab has used the approach of resequencing candidate genes in better understanding disease etiology. The Hegele lab used the resequencing approach to test for significant accumulation of rare missense variants in candidate genes for hypertriglyceridemia (*71, 75*).

### 1.7.2 Using genotyping to compare Genetic risk score in case-control cohorts

Sometimes selected common variants such as SNPs are genotyped in cases and controls to create a genetic risk score. An individual's genetic risk score is a measure of the sum of risk alleles present in an individual. The SNPs are genotyped to see if cases have more of the risk allele relative to controls. If the mean genetic risk score is higher in cases relative to controls, it shows that those sets of SNPs genotyped or the set of risk alleles increase disease susceptibility

## 1.8 *In silico* analyses

The effect of all rare missense variants on function of gene products of the four candidate gene were predicted using the *in silico* tools Polymorphism Phenotyping version 2 (PolyPhen-2) and Sorting Intolerant from Tolerant (SIFT). SIFT bases its prediction on multiple sequence alignments and amino acid conservation to determine whether a missense mutation is deleterious or not (*81, 82*).

PolyPhen-2 bases its prediction on amino acid sequence alignment and structural alignment of the protein (*83*).

PolyPhen-2 is a free online computational tool that is very commonly used to predict the effect(s) of missense variants on protein function. PolyPhen-2 uses 8 sequence-based predictive features and 3-structure based predictive features in its prediction model, where these predictive features essentially compare the wild type protein with the mutant protein (i.e. protein with the missense rare variant). PolyPhen-2 also reports estimates of false positive rate and true positive rate (*84*) . PolyPhen-2 qualitatively appraises a mutation to be either 'benign', 'probably damaging' or 'possibly damaging' depending on the false positive rate of the prediction model. The 'Possibly damaging' appraisal is a less confident prediction than 'probably damaging' because the false positive rate is higher in 'possibly damaging'(*84*).

SIFT is another free online computational tool that is commonly used for predicting the effect of missense variants. Using query sequences, SIFT compiles sequences of functionally related protein and calculates the probability of finding all 20 amino acid at each amino acid position, where the probabilities are recorded in a scaled probability matrix. A mutation is said to affect the protein if the scaled probability, or 'SIFT score', is below a certain threshold. An amino acid that is conserved throughout evolution is more likely to be intolerant to substitution than an amino acid that is not conserved. SIFT also gives a conservation value for each position, where 0 is the conservation value if all amino acids are seen and 4.25 is the conservation value if only one amino acid is seen throughout the homologous family of proteins at that position. SIFT ensures that the final set of aligned sequences has a median conservation value of approximately 3.00, because a median conservation value of 3.25 produces predictions with low confidence (*85*). The only limitation of

SIFT is that it does not use structural predictions — although structural predictions are said to marginally improve predictions (*85*).

Multiple sequence alignment was also performed for each of the 4 candidate genes to visually observe conservation of amino acid position of the detected missense rare variants. In a multiple sequence alignment, amino acid sequences of a particular protein from different species are aligned to derive potential evolutionary or functional significance of each amino acid residue (*86*). So in a multiple sequence alignment, each single row represents the amino acid sequence of a protein from one species, with gaps inserted so that homologous residues appear in the same positions across the species used for the alignment, Here, homologous is context dependent (*86*). In the evolutionary context, homology refers to the amino acid residues having common evolutionary ancestry. In the context of structural biology, equivalence refers to the analogous amino acid residues belonging to the homologous fold in the set of proteins. In the context of molecular biology, equivalence refers to the amino acid residues having similar functional roles in the set of proteins (*86*). Clustal Omega was the multiple sequence alignment tool used for multiple sequence alignments for the 4 candidate genes of the first project (*87*). Jalview 2.8 was used to analyze the multiple sequence alignments (*88*).

### 1.9 General Thesis Project Aims

The overall theme of this thesis was to better understand the genetic etiology of hypercholesterolemia in Familial Combined Hyperlipidemia (FCH). There were three main studies in this thesis. Each study was composed of various projects. Even though the second study and part of the last study focused on Familial

Hypercholesterolemia and hypertriglyceridemia, respectively, the studies still converge towards understanding the two phenotypes characteristic of FCH, namely hypercholesterolemia and hypertriglyceridemia.

In all the thesis projects, whenever FCH patients were used as cases, their controls were individuals with FHTG. This is because previous work from the Hegele lab has shown that individuals with FCH and FHTG share a common genetic architecture for hypertriglyceridemia. To understand hypercholesterolemia in FCH, we needed to control for the hypertriglyceridemia phenotype of FCH, by using individuals with FHTG as controls.

### 1.9.1 Study I- Resequencing candidate genes in FCH

I hypothesized that there is an accumulation of rare missense variants in the *LDLR*, *APOB*, *PCSK9* and *IDOL* in FCH cases relative to controls. I tested this hypothesis using Sanger Sequencing to see if hypercholesterolemia in FCH is a result of FH-causing mutations. The gene responsible for the autosomal recessive form of Familial Hypercholesterolemia, *ARH,* was not sequenced because FCH does not show an autosomal recessive pattern and the frequency of FH dues to ARH mutations is extremely low, meaning that I was less likely to detect any *ARH* mutation.

### 1.9.2 Study II- Understanding hypercholesterolemia in Mutation Negative Familial Hypercholesterolemia

There were two projects for this study and both studies were collaborations.

**Project I**

In the first collaboration project, I tested the hypothesis that Mutation Negative Familial Hypercholesterolemia patients (FH/M-ve) have a higher mean

LDL-C genetic risk score than Mutation Positive Familial Hypercholesterolemia (FH/M+ve) controls. The genetic risk score was determined by calculating how many risk alleles of 12 SNPs were present in FH/M-ve patients. These 12 SNPs (See Section 1.4.2.5) were identified in Global Lipids Genetics Consortium (GLGC)(*43*). This has already been studied in the British population (*48*). So my project was a replication study in the Canadian Population.

**Project II**

In the second collaboration project, *APOE* was sequenced in all our FH/M-ve patients because our collaborators found an *APOE* FH-causing mutation, namely the *APOE* Leu 167 del (data yet to be published). Thus, all the FH/M-ve patients were sequenced for the *APOE* Leu 167 del, as well as any novel mutations.

### 1.9.3. Study III – Use of laboratory GWAS data for further understanding hypercholesterolemia and hypertriglyceridemia

Three projects composed the third study. In all three projects the same GWAS data from the Hegele lab (*70*) were used. The GWAS data were analyzed using a Unix-based program called PLINK (*89*). PLINK can be defined as a program that allows analyses of various types of genomic and genetic data. The GWAS data was genotyped SNPs across the entire human genome; more information on the GWAS data can be found in Johansen *et al* (*70*).In general, GWAS data are a wealth of genetic information and thus makes GWAS data a useful resource for testing new hypotheses, performing new analyses and consequently generating new findings.

**Project I**

The Hegele lab identified many, but not all, of the genetic variation in HTG. In the first project of my third study, I tested the hypothesis that *DIET1* is associated with hypertriglyceridemia. As mentioned in Section 1.6 of this thesis, *DIET1* has shown association with HTG in mice. So, I questioned whether genetic variation in *DIET1* could further explain genetic variation in HTG, which is a component phenotype characteristic of FCH. The key experiment was logistic regression to see if there was an accumulation of risk alleles in *DIET1* region in HTG patients.

**Project II**

In the second study of my third project, I wanted to see if there is an accumulation of SNPs in the *PSMD9* region in FCH cases relative to controls since because *PSMD9* locus has been associated with hypercholesterolemia. The key experiment was performing logistic regression for the *PSMD9* region in FCH cases and FHTG controls.

**Project III**

The Global Lipids Genetic Consortium (GLGC) identified 37 SNPs to affect LDL cholesterol level in the general population (*43*). In the third project of my third study, I studied the 37 GLGC identified LDL-C SNPs for association with hypercholesterolemia in FCH. Then I tested whether FCH cases have a greater LDL-genetic risk score, where the genetic risk score was a measure of risk alleles of the 37 SNPs.

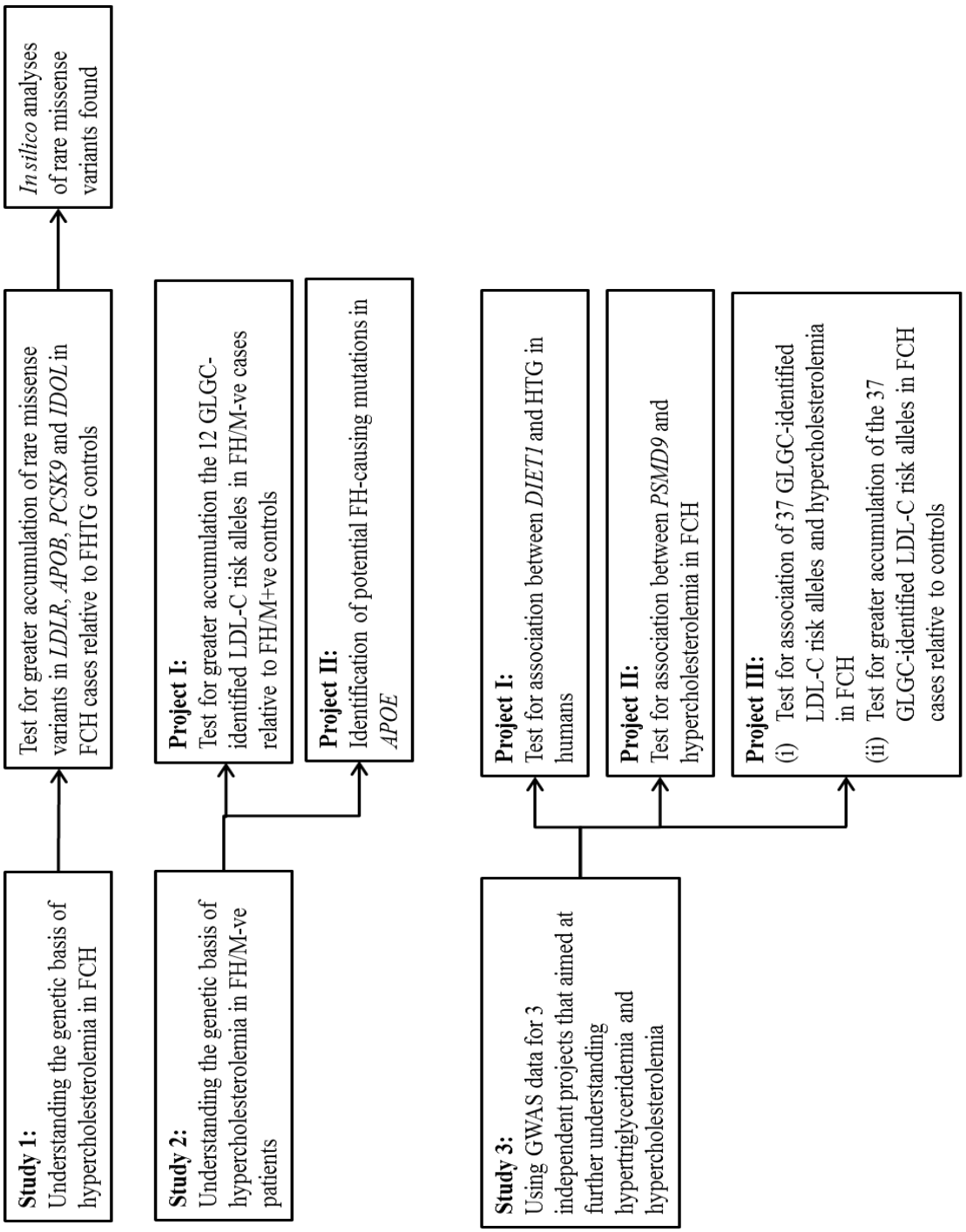



**Figure 1.12** Summary of Thesis projects

# Chapter 2: Materials and Methods

## 2.1 Study subjects

All study subjects provided informed consent for use of their DNA for research purposes, including DNA extraction, sequencing and analyses. This study was approved by the University of Western Ontario Institutional Review Board (protocol number 07920E) (Ethics Approval notice attached in the Appendix). The study subjects involved in all the three projects were of self-declared European ancestry.

### 2.1.1 Familial Combined Hyperlipidemia patients (Fredrickson Type 2B Phenotype)

For the first project, a total of 138 cases were Familial Combined Hyperlipidemia (FCH) patients (MIM 144250). FCH patients were unrelated and of self-declared European ethnicity. FCH patients were diagnosed as the affected individual having plasma total cholesterol concentration above the 90th percentile (>7.7 mmol/L) in addition to having plasma triglyceride concentration above the 90th percentile (>3.4 mmol/L), controlled for age and sex, according to reference levels for the North American population. All the FCH cases were patients from the Lipid Genetics Clinic in London, Ontario.

### 2.1.2 Familial Hypercholesterolemia patients (Fredrickson Type 2A Phenotype)

For the second project, a total of 44 cases were Mutation Negative Familial Hypercholesterolemia (FH/M-ve). 44 controls were Mutation Positive Familial

Hypercholesterolemia (FH/M+ve) patients. All of these 88 Familial Hypercholesterolemia patients (MIM 143890) were from the Lipid Genetics Clinic in London, Ontario. Familial Hypercholesterolemia was diagnosed as having Low Density Lipoprotein cholesterol (LDL-C) above the 95$^{th}$ percentile (>5.2 mmol/L), controlled for age and sex, according to reference levels for the North American population. Cases and control were matched for age and sex.

### 2.1.3 Hypertriglyceridemia patients

#### 2.1.3.1 Familial Hypertriglyceridemia patients (Fredrickson Type 4 phenotype)

For the first project, there were 94 Hypertriglyceridemia patients of the Fredrickson Type 4 phenotype classification (Familial Hypertriglyceridemia) used as controls. Familial Hypertriglyceridemia (FHTG) was diagnosed as having triglycerides above the 90$^{th}$ percentile (>3.7 mmol/L), controlled for age and sex, in the North American population. All the Type 4 controls were patients from the Lipid Genetics Clinic in London, Ontario. FHTG patients were used as for the hypertriglyceridemia component of FCH.

#### 2.1.3.2 Polygenic Hypertriglyceridemia patients (Fredrickson Types 2B, 3, 4 and 5)

For the third project, hypertriglyceridemia (HTG) patients were cases who constituted individuals clinically diagnosed with of all the 4 Fredrickson polygenic hypertriglyceridemia phenotypic classification, namely: Type 2B (MIM 144250), Type 3 (MIM 107741), Type 4 (MIM 144600) and Type 5 (MIM 144650). HTG was diagnosed as having an untreated 12 hour fasting plasma triglyceride concentration

above the 90$^{th}$ percentile (>3.4 mmol/L) on at least two occasions. These polygenic HTG cases were patients from the Lipid Genetics Clinic in London, Ontario. In total there were 504 HTG patients (cases) and 1254 mostly normolipidemic controls.

Familial Hypercholesterolemia (Fredrickson Type 2A phenotype) patients constituted 4% of controls; healthy Canadian individuals of European descent ascertained through the Study of Health Assessment and Risk in Ethnic Groups (*90*) constituted 18% of controls; healthy individuals from the Myocardial Infarction Genetics Consortium (*91*) constituted 78% of controls. Familial Hypercholesterolemia patients were used as controls to counterbalance the increased cholesterol phenotype that is seen in HTG patients.

## 2.2 DNA samples of study subjects

### 2.2.1 DNA extraction

Genomic DNA (gDNA) samples were collected from whole blood that was drawn from study subjects. gDNA was isolated from whole blood of study subjects using the Puregene DNA isolation kit (Gentra Systems, QIAGEN Inc, Mississauga, ON, Canada) according to manufacturer's instructions.

### 2.2.2 Whole genome amplification

Whole genome amplification was performed, according to manufacturer's instructions, on extracted gDNA using the Illustra GenomiPhi HY DNA Amplification Kit (GE Healthcare Life Sciences, Mississauga, ON, Canada). WGA was performed for FCH and FHTG gDNA samples because gDNA samples were limited. The number of gDNA samples was limited as those samples that had been

collected over the course of 15 years and had been used to varying degrees in earlier studies.

### 2.2.3 DNA quantification

For the second project, gDNA was quantified to a final concentration of approximately 50ng/µL for genotyping. DNA quantification was performed by measuring the concentration of 1µL gDNA on the ND-1000 Spectrophotometer (Nanodrop, Thermo Scientific, Mississauga, ON, Canada) for each gDNA FH sample. 88 FH gDNA samples were diluted to a concentration of y 50ng/µL, which is considered optimal for subsequent procedures, including genotyping.

## 2.3 DNA amplification by Polymerase Chain Reaction

The polymerase chain reaction (PCR) was performed to amplify target regions of WGA DNA of the 4 candidate genes in cases and controls of the first project. Target regions in *LDLR* were the 18 coding regions, 100bp outside each exon and the promoter. Target regions in *APOB* were exons 26 and 29 since 90% of reported hyperlipidemia-associated mutations reside within these regions. The target region in *PCSK9* for sequencing was exon 7, since more than 60% of disease-associated variants have been shown to reside in this exon (*92*). The target regions in *IDOL* were 7 exons, since no disease-causing mutations have yet been reported in this gene.

### 2.3.1 Primer list for candidate genes

**Tables 2.1** to **2. 4** represent the forward (F) and reverse (R) primers of the target regions for the first project and annealing temperature for each target region. Primers

were designed using the free online software, Primer3 The length of both forward and reverse primers had to be between 18 to 22 bp; the difference in length of the primers was not more than 1 nucleotide; the GC content of each primer must be between 40%-60%; the forward and reverse primer must have similar melting temperatures ($T_m$ ) (i.e. not more than 1˚C difference )and repeat sequences in primers were avoided (*93*). Factors that affect the melting temperature include GC content, concentration of ions and DNA length (http://www.entelechon.com/2008/08/dna-melting-temperature/ ).

**Figures 2.1** to **2.4** represents the gene structure of each candidate gene and the regions of the genes that were sequenced. Figures were generated by mainly using *GenomeGraphs* software package (*94*) in R statistical programming environment (*95*). Ensemble ID (ENSG ID) of all genes were required for generation of figures:

*LDLR*: ENSG00000130164,

*APOB*: ENSG00000084674,

*PCSK9*: ENSG00000169174

*IDOL*: ENSG00000007944

**Table 2.1. Primer sequence and annealing temperature of the 18 exons and promoter of *LDLR***

| Exon | Annealing temperature (˚C) | Amplicon size (bp) | Primer sequence (5' to 3') |
|---|---|---|---|
| Promoter | 58 | 410 | F: CAGGAGGATCTTTCAGAAGATGCG<br>R: AGGAGCAAGGCGACGGTCCAG |
| 1 | 58 | 438 | F: GGACTGGAGTGGGAATCAGA<br>R: TTACCCCACAAGTCTCCCAG |
| 2 | 58 | 486 | F: GTGCTTGCTTAATTCCCTGG<br>R: TCAAAATCCACTGGCCAC |
| 3 | 58 | 501 | F: GAGACAGGGTTTCACTATATTGGC<br>R: ACAAACCCGAAGAGGTAGCA |
| 4 | 58 | 609 | F: GCAGTGGTTCAGAGTCCATGG<br>R: TCCCAATAAGCTAACAGCAACCATCGG |
| 5 | 58 | 520 | F: CTCAAGCAGTTGGAACCACA<br>R: GCGAGACTCCGTCTCAAAAC |
| 6 | 58 | 357 | F: GTGCTGGGATTACAGGCACAAAC<br>R: CCTACAGCACTCATGTCTCAGTC |
| 7 | 60 | 470 | F: ACATGCCTGTAGTCCCAGCTACTT<br>R: CAAGAAACTCTGGCCAGCCAATGA |
| 8 | 58 | 362 | F: TTACATCTCCCGAGAGGCTGG<br>R: GGTCAGGGGATATGAGTCTGTG |
| 9 | 58 | 400 | F: GGAGGTCTTTTCCACCCTCT<br>R: CTGAGGCAGGAGGAGAGAAG |
| 10 | 58 | 498 | F: AGCGAGTACACCAGCCTCATC<br>R: GCCCACTAACCAGTTCCTGA |
| 11 | 58 | 467 | F: CCCAAACAAGCCACATTTGGAGTTTGGGGTTC<br>R: AAAGAGGGAAACCTTCAGGGAGCAGCTTGG |
| 12 | 58 | 467 | F: TGTGACCTGCAACTCCCCTAC<br>R: CTCAGGTCTAAGACCTCCTCC |
| 13 | 61 | 526 | F: AGGCTGAAGCAAGAGAATCG<br>R: GGTGGTCCTCTCACACCAG |
| 14 | 58 | 502 | F: TCTCTTCCACAACCTCACCC<br>R: CATCAAAGGGGAACTGGGTA |
| 15 | 61 | 500 | F: AGAGATGGTATTTTGCCATGTTG<br>R: GATAGGGAAACTGAGGGCCCAGAG |
| 16 | 58 | 508 | F: CCGGAATTGAGTCCTACAACC<br>R: TCTCGGTGAGGCTATTCCAC |
| 17 | 58 | 440 | F: GTCAAGGTTATGGTACGATGC<br>R: TTCCTCTACACCACCAAGGC |
| 18 | 58 | 574 | F: ACTGAATCCGGTACTCACCG<br>R: GTGCCATCTGCTGTTGTGTG |

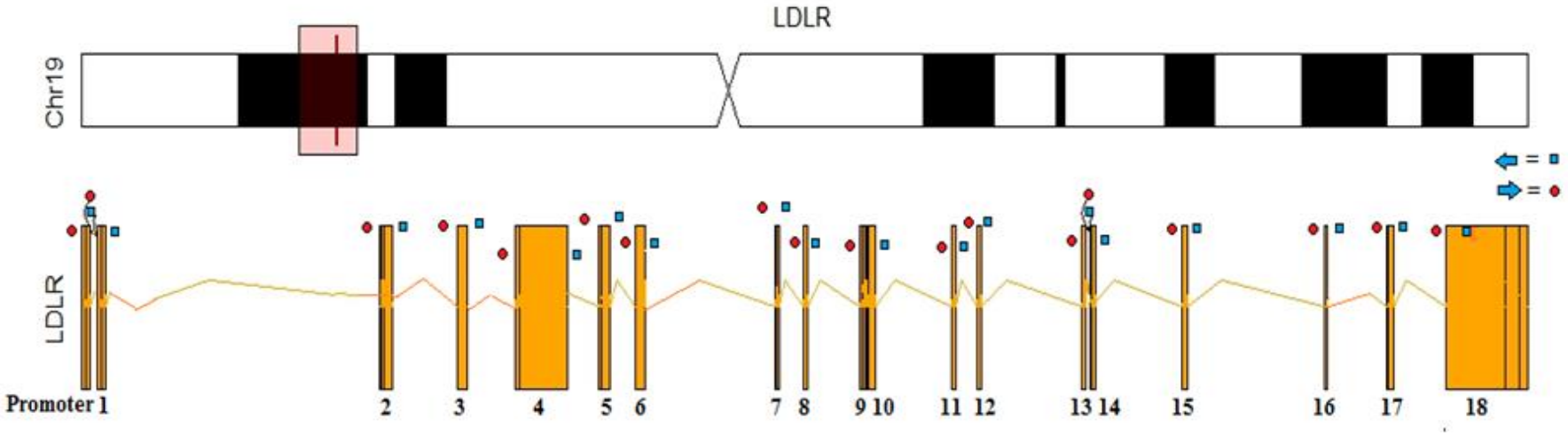



**Figure 2.1.** Gene structure of *LDLR* (ENSG00000130164): (a) the chromosomal location of *LDLR* on the p (short) arm of chromosome 19. The position of *LDLR* is represented by the vertical line and the region surrounding *LDLR* is shaded in red; (b) the rectangles are symbolic of the exons in *LDLR* and the lines joining the rectangles are symbolic of the introns. Circles represent forward primer and squares represent reverse primer; primers are placed approximately 100 bp outside the intron-exon boundary.

**Table 2.2 Primer sequence and annealing temperature of the 19 amplicons of exon 26 and 3 amplicons of exon 29 of *APOB***

| Exon | Annealing temperature (˚C) | Amplicon size (bp) | Primer Sequence (5' to 3') |
|---|---|---|---|
| 26-1 | 60 | 580 | F:TGCATTACAGATGGAGGAGTC |
| | | | R:TTTGCAGATCAGAGGTGGAGG |
| 26-2 | 60 | 590 | F:TTCGTTCTATGCTAAAGGCACA |
| | | | R:AGCGGCCATTTGTTGTTAAT |
| 26-3 | 60 | 581 | F:AAAATTAATAGTGGTGCTCACAAGG |
| | | | R:TTTAGGTTACCAGCCACATGC |
| 26-4 | 60 | 590 | F:CAGCTCTGACAAGTTTTATAAGCAA |
| | | | R:GGGCACTGACTTTGTGTTCA |
| 26-5 | 60 | 557 | F:CGCTCTCTGGGGAGAACATA |
| | | | R:TGATGTGCTTCAGGTTTCTCTG |
| 26-6 | 60 | 556 | F:ACCAAGATGTTCACTCCATTAACC |
| | | | R:CAGGATGCAGTACTACTTCCAC |
| 26-7 | 60 | 512 | F:TTGATGAGCACTATCATATCCGTG |
| | | | R:TTGTAGGACATTGCTTAGCTTCTG |
| 26-8 | 60 | 589 | F:ATCCTTCAGAGCCAAAGTC |
| | | | R:TCCTGCTGAATGTCCATTTG |
| 26-9 | 60 | 580 | F:AAGGCCACAGTTGCAGTGTA |
| | | | R:CTACAAAGTCAATTGTAAAGGAAGGA |
| 26-10 | 60 | 510 | F:CAGATTTGAGGATTCCATCAGTTCAG |
| | | | R:AAGCTGCGATACCTGCTTCGTTTG |
| 26-11 | 60 | 580 | F:TACCTACTTTTGGCAAGCTATACA |
| | | | R:TGTGATTCATGTGTTCCCTCA |
| 26-12 | 60 | 593 | F:ATTGAACATCCCCAAACTGG |
| | | | R:TTACTTGCCAACTTGCTTGC |
| 26-13 | 60 | 487 | F:TTTGAGATCACGGCATCCACAAAC |
| | | | R:TGTCAAAGGATTTGATGCTCTGAC |
| 26-14 | 60 | 580 | F:AAGAAAAACAAACACAGGCATTC |
| | | | R:AAGATGAAGAAAGGAGATGAGCA |
| 26-15 | 60 | 557 | F:CTCTTCCAGATTTCAAGGAATTGTG |
| | | | R:CTTGACATCTCCTTTGGTAGATG |
| 26-16 | 60 | 544 | F:CAATTCTTCAATGCTGTACTCTACC |
| | | | R:GACCTGGCTCTGGAAAGACC |
| 26-17 | 60 | 582 | F:TCAGCTCTTGTTCAGGTCCA |
| | | | R:TTTTACCTCGGGGAGTGTTG |
| 26-18 | 60 | 549 | F:TCAGTTCTTGTCATGCCTACG |
| | | | R:TAGGAACTGTACGGTTGAGCTG |
| 26-19 | 60 | 580 | F:TCCTTCCTTTCAAGCACTGAC |
| | | | R:TTTTGTGTGTTCCCAAAACTG |
| 29-1A | 60 | 580 | F:TTGTGTAATTGGAGTAATTGAAAACA |
| | | | R:GAAACTGGAATCTGGGGAAG |
| 29-2 | 60 | 613 | F:CCATATGAAAGTCAAGCATCTGA |

| | | | |
|---|---|---|---|
| | | | R:TTCACGAAGGGCCATAATGT |
| 29-3A | 60 | 657 | F:ATTCAAAACGAGCTTCAGGA |
| | | | R:TGTGAAAGTTCAATTGGAAAAGA |

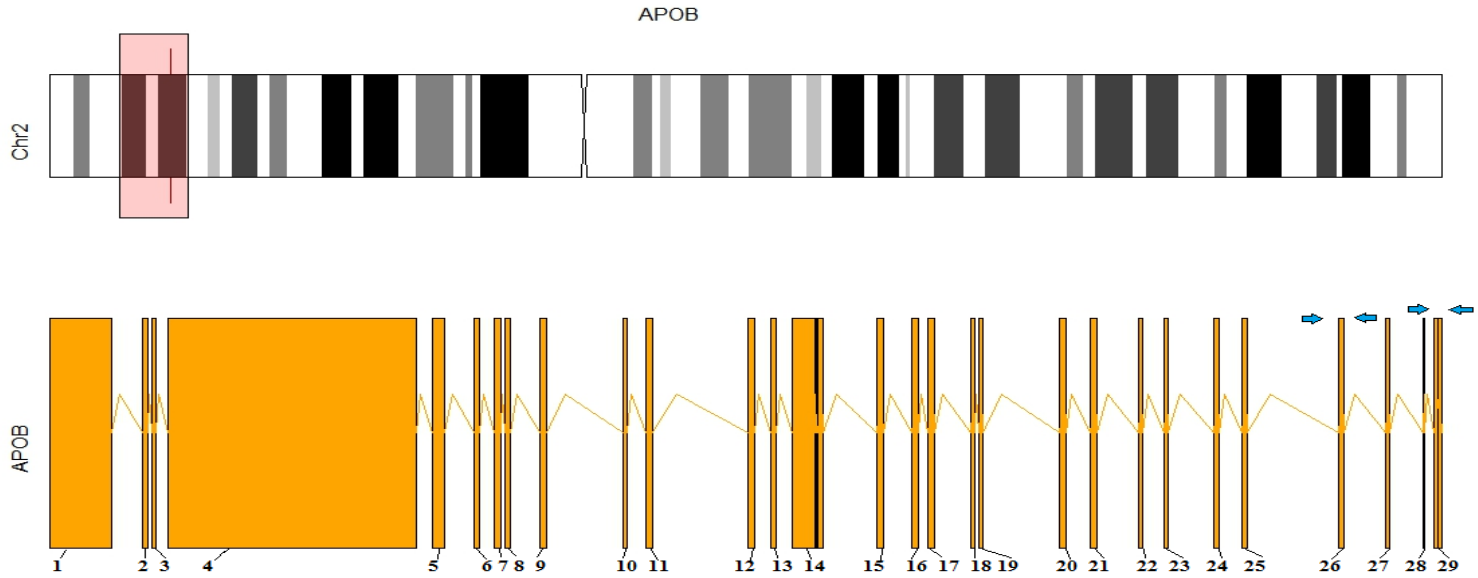



**Figure 2.2** . Gene structure of *APOB* (ENSG00000084674): (a) the chromosomal location of *APOB* on the p arm of chromosome 2. The position of *APOB* is represented by the vertical line and the region surrounding *APOB* is shaded in red; (b) the rectangles are symbolic of the exons in *APOB* and the lines joining the rectangles are symbolic of the introns. The arrows represent the primers; primers are placed approximately 100 bp outside the intron-exon boundary. Only exons 26 and 29 were sequenced in 138 FCH cases and 94 controls

**Table 2.3. Primer sequence and annealing temperature of exon 7 of *PCSK9***

| Exon | Annealing temperature (℃) | Amplicon size (bp) | Primer Sequence |
|---|---|---|---|
| 7 | 59 | 451 | F: CAGAGTTCTGCCTGGGCAGTC<br>R: GAGTGTCCTTGAAGGCACCATC |

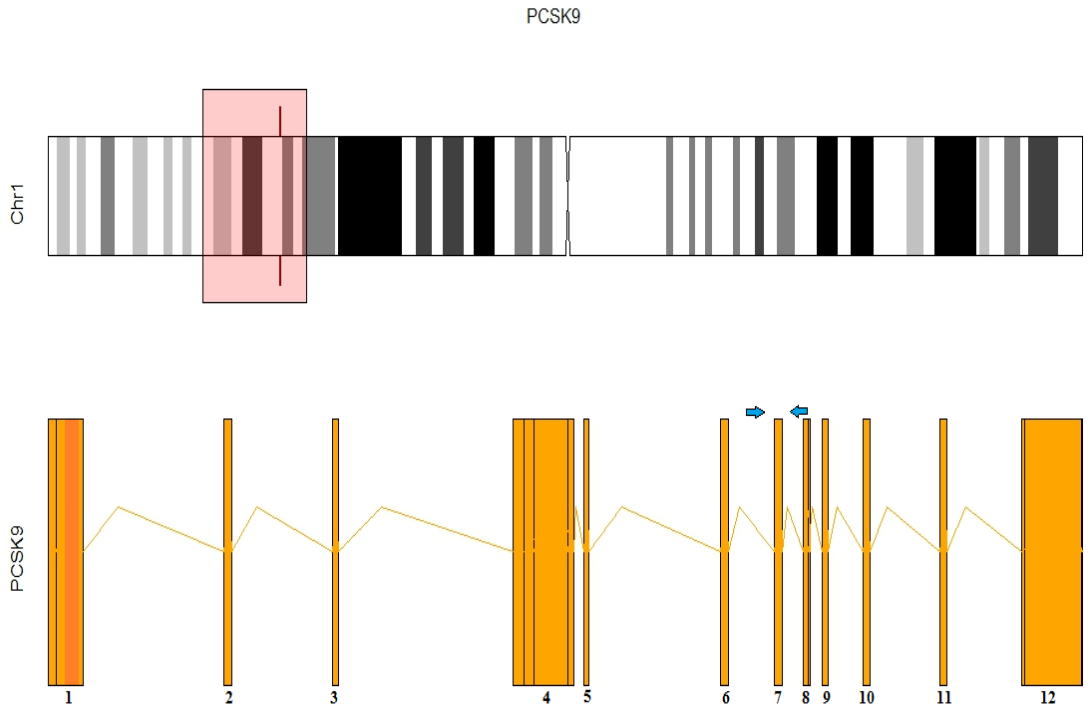



**Figure 2.3**. Gene structure of *PCSK9* (ENSG00000169174): (a) the chromosomal location of *PCSK9* on the p arm of chromosome 1. The position of *PCSK9* is represented by the vertical line and the region surrounding *PCSK9* is shaded in red; (b) the rectangles are symbolic of the exons in *PCSK9* and the lines joining the rectangles are symbolic of the introns. The arrows represent the primers; primers are placed approximately 100bp outside the intron-exon boundary.

**Table 2.4 . Primer sequence and annealing temperature of the 7 exons of *IDOL***

| Exon | Annealing temperature (˚C) | Amplicon Size (bp) | Primer Sequence (5' to 3') |
|---|---|---|---|
| 1 | 62 | 286 | F:GAGGGCCAGCCCTCTCCGAGTCCG |
| | | | R:TAGTAGGGGGCGCGCCAGAGTGCC |
| 2 | 60 | 413 | F:TGGTATCATTGGAGCCGTGGAACT |
| | | | R:CACTCAGACCAAGTAGGTAGCTCC |
| 3 | 60 | 346 | F:GCTGAGATTGATGTCAGGTTATCC |
| | | | R:TGCCTCGAACATCAGAGAGCTCAA |
| 4 | 58 | 324 | F:TGAGATCCCAGTGTCTTAGACGTT |
| | | | R:GAGCTGACTGTCGAGTAAATCCCT |
| 5 | 60 | 311 | F:CCACAAAGGCACACACATGGTGAA |
| | | | R:ACCGTAGAAACCTGGTTGTCACCT |
| 6 | 60 | 602 | F:GGAGATGTTAGAGAAACAGAGGTG |
| | | | R:ACAAAGACCCTTTCCGGGTGAAGA |
| 7 | 60 | 340 | F:TGTGAGACGGCAAAGATCTCTACC |
| | | | R:TGGTCCCATGACTGGAGTTGTTGA |

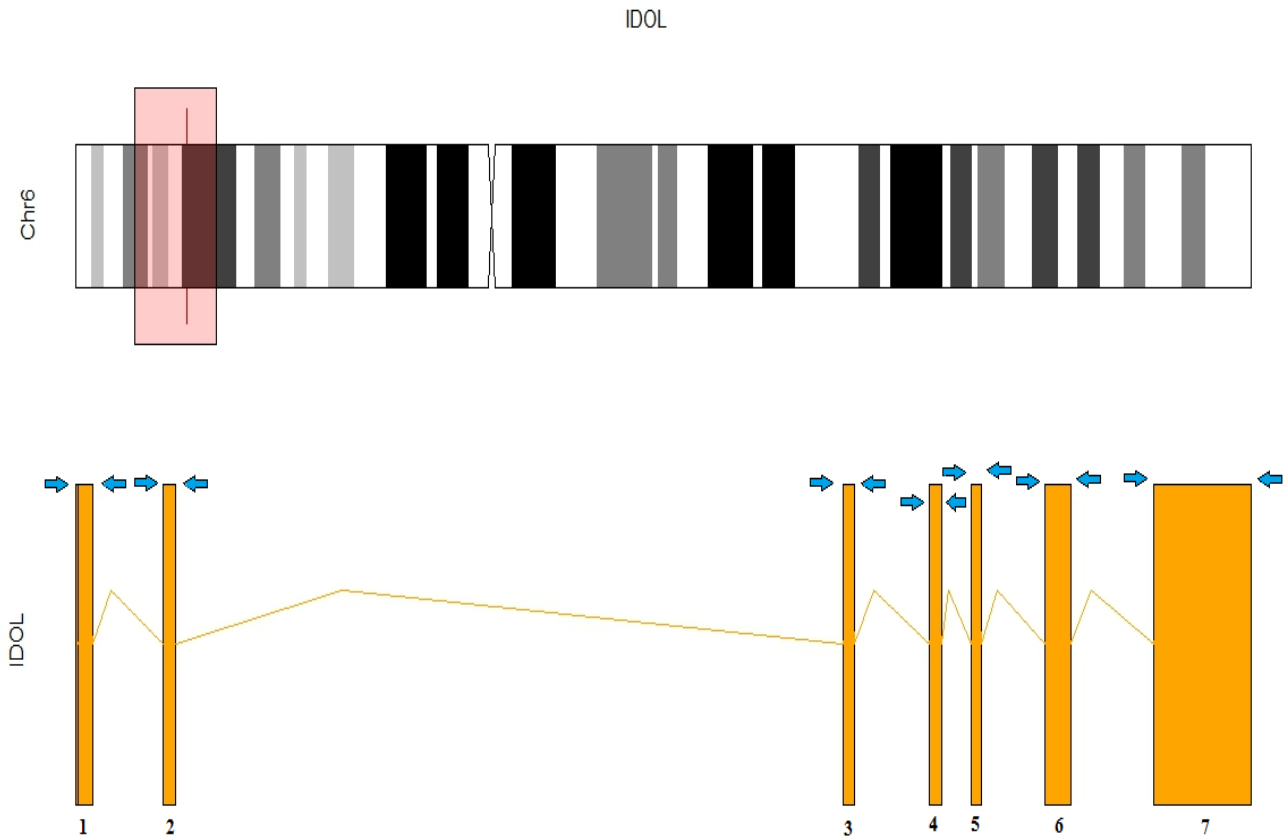



**Figure 2.4**. Gene structure of *IDOL* (ENSG00000007944): (a) the chromosomal location of *IDOL* on the p arm of chromosome 6. The position of *IDOL* is represented by the vertical line and the region surrounding *IDOL* is shaded in red; (b) the rectangles are symbolic of the exons in *IDOL* and the lines joining the rectangles are symbolic of the introns. The arrows represent the primers; primers are placed approximately 100bp outside the intron-exon boundary.

### 2.3.2 Polymerase Chain Reaction

WGA DNA samples of FCH and FHTG patients were used for Polymerase Chain Reaction (PCR). The PCR kit used was Life Technologies Platinum® Taq (Life Technologies Inc., Burlington, ON, Canada). WGA DNA Samples and PCR reaction mixtures were placed in 96-well plates in a DNA Thermocycler (Life Technologies Inc. Burlington, ON, Canada). Every reaction well contained 1µL of WGA DNA, 2µL 10x $MgCl_2$ PCR buffer, 3.2µL of 10mM of each of the 4 dNTPs , 0.33µL of 0.6pmol/µL of the forward primer, 0.33µL of 0.6pmol/µL of the reverse primer, 0.6µL of 50mM $MgCl_2$ and 0.1µL of 5U/µL Taq Polymerase, yielding a final volume of 20µL PCR reaction mixture. The thermocycler conditions were comprised of 3 stages. The first stage was the initial denaturing stage for 5 minutes at 95℃ to separate DNA strands, followed by second stage that was composed of 30 cycles of 3 steps. The first step was denaturation for 30 seconds at 95℃, the second step was annealing for 30 seconds at 60℃ and the third step was elongation for 30 seconds at 72℃. The final stage was the final elongation stage, carried out at 72℃ for 10 minutes to ensure that every amplified DNA strand was fully extended.

For some target regions, the PCR reaction mixture and Thermocycler conditions were slightly different as those target regions required addition of 99.9% dimethyl sulfoxide (DMSO) (Sigma-Aldrich®, Oakville, ON, Canada) to ease double strand separation. **Table 2.5** contains details of those PCR reaction conditions. Only differences in PCR conditions are displayed in the table — otherwise PCR reaction conditions are as stated in this section.

**Table 2.5.** PCR reaction mixture for target regions that required DMSO

| Gene | Exon | Reaction Mixture |
|---|---|---|
| *LDLR* | 4 | 1.5uL DMSO, 10.8uL distilled water, extension time 45 seconds |
| | 7,13,15 | 1.5uL DMSO, 10.8uL distilled water, 0.2uL Taq Polymerase, 2.0uL WGA DNA |
| *IDOL* | 1 | 1.5uL DMSO, 10.8uL distilled water, 0.2uL Taq Polymerase |

### 2.3.3 Gel electrophoresis

A 1% agarose gel was used to visualize PCR products under UV light to confirm satisfactory amplification of PCR products before sending samples to sequencing. The agarose gel was made by dissolving 1g of agarose powder (Bioshop® Canada Inc., Burlington, ON, Canada) in 100mL of distilled water and boiling for 5 minutes. PCR products containing loading dye were loaded on to the agarose gel and Loading dye contained glycerol. In all cases, negative control reactions showed no bands which confirmed no contamination; if negative controls had been positive for a PCR product (i.e. showed band on agarose gel), it would have meant that our experimental WGA DNA samples had been contaminated with other DNA samples from other sources, such as bacteria and PCR for contaminated samples was repeated.

### 2.3.4 Calf intestinal phosphatase (CIP)- Exonuclease (Exo) I treatment and preparation of sample for Sanger sequencing

Contaminants such as single strands, primers and free dNTPs were removed from PCR product using calf intestinal phosphatase – exonuclease I (CIP-ExoI) treatment on PCR product in a thermocycler. CIP-ExoI treatment was performed to ensure that only PCR products, and no contaminants, had been sequenced. In each case, 6µL of CIP-ExoI treatment was added to 6µL of PCR product in a reaction well. CIP-ExoI treatment was made up of 5.7µL of distilled water, 2 Units of Exonuclease I (New England BioLabs ® Inc., Ipswich, Massachusetts, United States ) and 2 Units Calf Intestinal Phosphatase (New England BioLabs ® Inc., Ipswich, Massachusetts,

United States ). The thermocycler conditions for CIP-ExoI treatment were 37˚C for 1 hour, to activate the CIP-ExoI enzyme, followed by 72˚C for 15 minutes to stop all enzyme activity.

Reactions containing 2.5µL of CIP- ExoI treated PCR product, 2.5µL of distilled water and 2.5 uL of the sequencing primer, yielding a total of 7.5uL, were sent to the London Regional Genomic Centre (LRGC) (London, Ontario, Canada) for sequencing. The sequencing primer was a ¼ dilution of either of the forward or reverse PCR primer from stock solution, so that two 7.5uL PCR products were sent to sequencing for each DNA sample, where one 7.5uL PCR product contained the forward primer and the second 7.5uL PCR product contained the reverse primer. Automated chain termination Sanger sequencing was performed on PCR products at LRGC using the ABI 3730 (Life Technologies, Burlington, Ontario, Canada). For each PCR product sent to sequencing, electropherograms, which were the hard copies of the DNA sequence of target region of each sample, were produced.

### 2.3.5 Sanger Sequencing

In the London Regional Genomic Centre (LRGC), the 7.5µL CIP-ExoI treated PCR products of all target regions were Sanger Sequenced by the involvement of four steps, namely, denaturation, chain termination of PCR amplification, purification and sequencing, respectively. In total, there were 11,368 target regions that were sequenced for the first project.

The samples of PCR amplicons were centrifuged at 1000g for one minute on the Beckman Coulter TJ-25 Centrifuge (Biotech Equipment Sales Inc., San Francisco, California, United States). After centrifugation, the samples are placed on an Applied Biosystems GeneAMP 9700 Thermocycler (Life Technologies Inc., Burlington,

Ontario, Canada) at 98°C for 5 minutes for denaturing of samples. The samples were held at 4°C in the thermocycler to keep the double stranded PCR product unwound. After the denaturing, the samples were again centrifuged at 1000g for one minute. After centrifugation, the samples were placed on an ice pack to maintain the unwound structure.

4µL of Applied Biosystems BigDye® Terminator Master Mix (Life Technologies Inc., Burlington, Ontario, Canada) was added to 7.5µL denatured PCR product. The Master Mix was made up of 12.5% BigDye Terminator, 25% 5 x Buffer AB and 62.5% distilled water. The positive control had the 4µL BDT Master Mix, 2µL pGEM (Promega Corporation, Madison, Wisconsin, United States), 2µL Applied Biosystems® Control Primer (M13F) (Life Technologies, Burlington, Ontario, Canada), and 2µL distilled water. The PCR product and positive control containing Master Mix, making a total volume of 11.5µL in each well, were each centrifuged at 1000g for one minute. After centrifugation, the samples were placed on the Applied Biosystems GeneAMP 9700 Thermocycler for chain termination PCR. A hot start PCR was performed at 80℃. The thermocycler conditions comprised 2 stages. The first stage was an initial denaturing stage for 5 minutes at 96℃ to separate DNA strands, followed by the second stage that was composed of 30 cycles of 3 steps. The first step of these was denaturation for 20 seconds at 96℃, the second step was annealing for 15 seconds at 50℃ and third step was elongation for 4 minutes at 60℃. After the second stage, the samples were stored at 4℃ in the thermocycler.

The samples were then purified to remove contaminants such as free dNTPs, proteins, salts and unincorporated dye. Purification was carried out using an Edge Plate (Edge BioSystems, Gaithersburg, Maryland, United States).

The purified samples were then placed on Applied Biosystems 3730 Analyzer (Life technologies Inc., Burlington, Ontario, Canada) for sequencing.

The raw data from the Applied Biosystems 3730 Analyzer was processed using the Applied Biosystems DNA Sequencing Analysis Software version 5.3.1 (Life Technologies Inc., Burlington, Ontario, Canada). After processing of raw data, electropherograms, which were the hard copies of the DNA sequence of target region, were generated. Electronic versions of the electropherograms were also available.

The samples were denatured to unwind the double-stranded PCR products so that there would be ease of access of sequencing reagents.

Hot Start PCR minimizes amplification of non-specific sequences because it prevents non-specific primer annealing, which typically occurs at lower temperatures (i.e. below 65℃) (*96*).

In chain termination PCR, fluorescently labeled dideoxy nucleotides are used. Each of the 4 dideoxy nucleotides fluoresces at a different colour. The rationale for using dideoxy nucleotide was to generate PCR products that end with the fluorescently labeled dideoxy nucleotide and generate chain terminated PCR products that differ from each other by one nucleotide. The fluorescently labeled dideoxy nucleotides is able to terminate PCR reaction because the Dideoxy NTPs lack the 3' hydroxyl group, which prevents incorporation of another nucleotide as no phosphodiester bond can be created (*97*, *98*). **Figure 2.5** is a simplified visual representation of the reactions that occur in Sanger Sequencing.

Within the 3730 sequencer, the fluorescently labeled PCR products pass through a capillary tube that separates chain terminated PCR products on the basis of size. The smallest PCR product is read first, the next PCR product that is 1bp longer is read next and so on. The colour of each fluorescent dye corresponds to a particular

nucleotide, which essentially leads to each nucleotide position of the PCR product being read as a sequence.

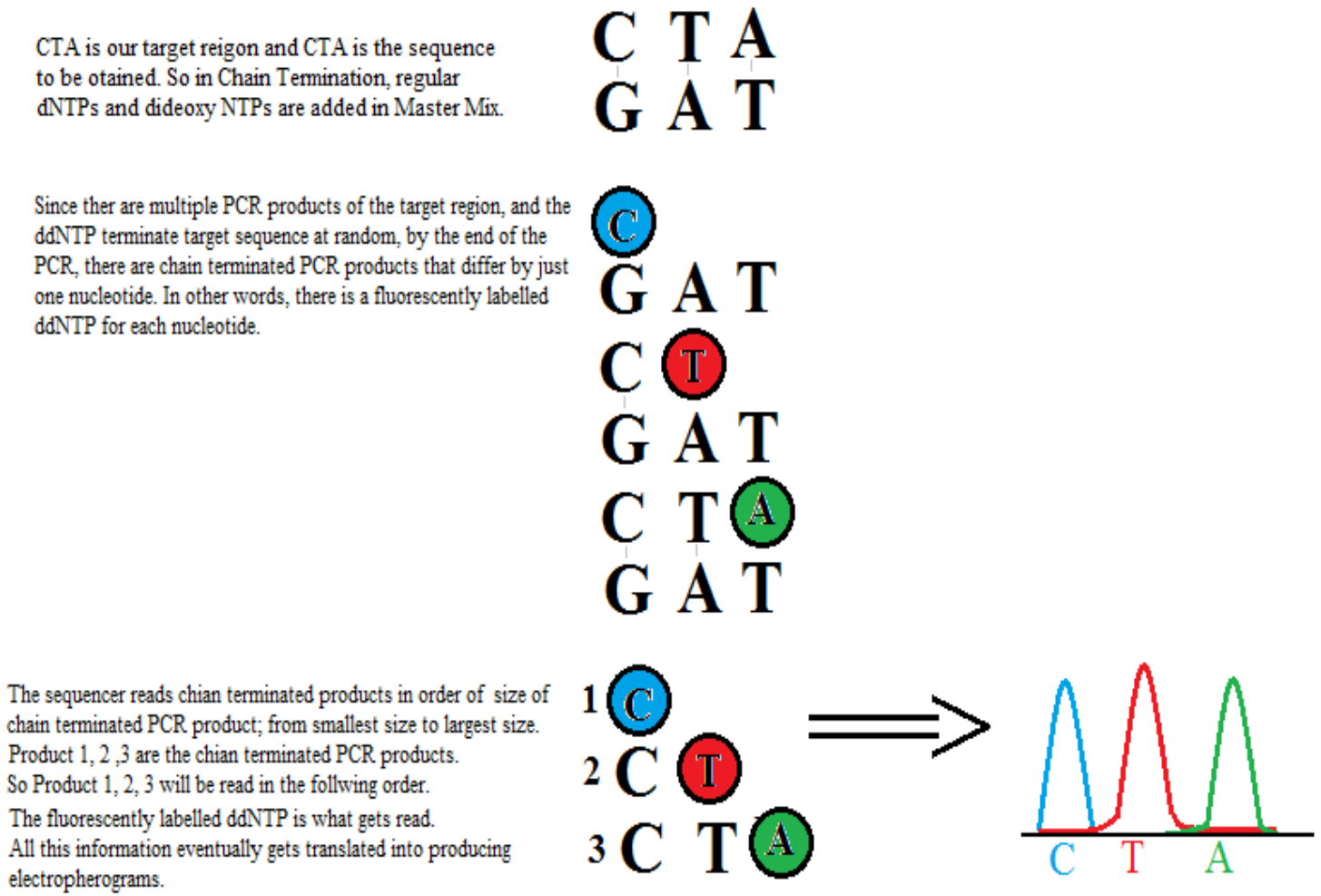



**Figure 2.5** Simplified visual representation of essentially what occurs in Sanger Sequencing.

### 2.3.6 DNA sequence analyses

Electropherograms of all target regions were analyzed using SeqScape® software version 2.6 (Applied Biosystems), which is a standard software for mutation detection and analysis. Reference genomes (hg19) were obtained from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/ ) and target regions were compared with reference genomes. All genotype information was entered into the lab database and genotype information was used for statistical analyses. The NCBI reference genomic DNA (the NC number) and cDNA (the NM number) for the candidate genes were as follows:

*LDLR*: NC_000019.9 and NM_000527.4;

*APOB*: NC_000002.11 and NM_000384.2;

*PCSK9*: NC_000001.10 and NM_174936.3

*IDOL* NC_000006.11 and NM_013262.3

### 2.3.7 Power Calculations

Power calculations were performed using the free software PS-Power and Sample Size Calculation (*99*). Power was determined after calculating rare variant accumulation and obtaining Odds Ratios for the risk alleles of all the candidate genes. Significance level used was 5% in power calculation for all candidate genes.

**2.4 SNP genotyping for Mutation Negative Familial Hypercholesterolemia patients and Mutation Positive Familial Hypercholesterolemia patients (Fredrickson Type 2A phenotype)**

As part of a collaboration to study patients with clinical FH who were negative for mutations in known genes, 44 Mutation Negative Familial Hypercholesterolemia (FH/M-) patients (cases) and 44 Mutation Positive Familial Hypercholesterolemia (FH/M+) patients (age- and sex-matched controls) were selected. Samples were sent to the British Heart Foundation (BHF) Laboratories, Institute Cardiovascular Science, University College London, England United Kingdom to obtain an LDL-C genetic risk score using the top 12 Global Lipids Genetic Consortium (GLGC) identified SNPs (reference). We hypothesized that FH/M-ve patients would have a significantly greater accumulation of LDL-C raising SNPs than FH/M+ve patients. This hypothesis was first evaluated and proven in a British study of patients predominantly resident in the United Kingdom (*48*). This study was the first study to test this hypothesis in a Canadian population; our study also served as an independent replication study for the work spear-headed by our collaborators.

The GLGC was an international project that identified 95 loci associated with lipid traits at the genome wide level (*43*). 37 loci were associated with LDL-C and 12 of those 37 were selected for SNP genotyping on the 88 Canadian FH/M+ve and FH/M-ve samples. These 12 SNPs were chosen for genotyping because a previous study, preformed in a British population, showed that that FH/M- patients have a significantly higher LDL-C genetic risk score, where those 12 SNPs were used for constructing LDL-C genetic risk score (*48*).

SNP genotyping was the key experiment of the second project of the thesis.

Genotyping was performed on gDNA samples that were diluted to 50ng/uL as this was the optimal concentration for subsequent procedures for genotyping.

### 2.4.1 List of GLGC identified SNPs

**Table 2.6** shows the 12 GLGC identified SNPs that were genotyped to calculate the LDL-C genetic risk score for all 88 FH patients. The risk alleles (bolded) are the LDL-C raising alleles.

### 2.4.2 Calculation of LDL-C genetic risk score

12 common LDL-C raising alleles identified by GLGC were used to construct a weighted LDL-C-raising genetic risk score. For each individual, LDL-C genetic risk score was calculated by calculating the weighted sum of the risk alleles. The weights used corresponded to the weight of the allele effect size as determined in GLGC (*43*).

### 2.4.3 *APOE* gene resequencing in Mutation Negative Familial Hypercholesterolemia patients

All 3 coding regions of *APOE* were sequenced in 95 FH/M-ve patients. *APOE* was sequenced to identify any potential FH-causing mutations as a recent collaborative study has showed an *APOE* amino acid deletion variant namely, *APOE* Leu 167 del to be FH-causing.

*APOE* was sequenced to screen for this particular amino acid deletion variant and for possible novel FH-causing or FH-associated variants. *APOE* was sequenced by following identical procedures of sections 2.3.2 to 2.3.5 of the Materials and Methods section. For the 3 coding regions of *APOE*, the PCR reaction mixture and

thermocyler conditions were slightly different, as the 3 coding regions required addition of 99.9% dimethyl sulfoxide (DMSO) (Sigma-Aldrich®, Oakville, ON, Canada) to ease double strand separation. PCR reaction conditions for all three coding regions of *APOE* were identical to those for the first exon of *IDOL*, shown in **Table 2.5**. Only differences in PCR conditions for the three coding regions are displayed in **Table 2.7**— otherwise PCR reaction conditions for all three coding regions are as stated in section 2.3.2.

**Table 2.6. List of GLGC identified SNPs that were genotypes in the 88 Familial Hypercholesterolemia cohort and their effect sizes**

| CHR | SNP | Gene | Minor* | Common* | GLGC Weight for Score Calculation |
|---|---|---|---|---|---|
| 1 | rs2479409 | *PCSK9* | **G** | A | 0.052 |
| 1 | rs629301 | *CELSR2* | G | **T** | 0.15 |
| 2 | rs1367117 | *APOB* | **A** | G | 0.1 |
| 2 | rs4299376 | *ABCG8* | **G** | T | 0.071 |
| 6 | rs1564348 | *SLC22A1* | C | **T** | 0.014 |
| 6 | rs1800562 | *HFE* | A | **G** | 0.057 |
| 6 | rs3757354 | *IDOL* | T | **C** | 0.037 |
| 11 | rs11220462 | *ST3GAL4* | **A** | G | 0.05 |
| 14 | rs8017377 | *KIAA1305* | **A** | G | 0.029 |
| 19 | rs6511720 | *LDL-R* | T | **G** | 0.18 |
| 19 | rs429358 | *APOEψ* | C | T | |
| 19 | rs7412 | *APOEψ* | T | C | |
| 19 | ε2ε2 | *APOE* | | | -0.9 |
| 19 | ε2ε3 | *APOE* | | | -0.4 |
| 19 | ε2ε4 | *APOE* | | | 0.2 |
| 19 | ε3ε3 | *APOE* | | | 0 |
| 19 | ε3ε4 | *APOE* | | | 0.1 |
| 19 | ε4ε4 | *APOE* | | | 0.2 |

*LDL-C raising alleles are indicated in bold. Effect sizes taken from (*48*)

Note: Abbreviations: CHR: Chromosome, SNP: single nucleotide polymorphism

[ψ] APOE weights were based on haplotypic effects taken from (*100*) as described in Methods.

For calculating the effect size of *APOE*, the *APOE* genotype was first determined and then given an *APOE* risk score calculation.

**2.5 SNP genotyping for patients with hypertriglyceridemia**

Genome wide association study (GWAS) data were obtained for the third project (*70*) and were used in genomic regions that were significantly associated with plasma triglyceride levels. To accomplish this, 463 HTG patients and 1197 healthy controls were genotyped for SNPs across the entire genome using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA)(*70*). HTG patients studied had Fredrickson types 2B, 3, 4 and 5. Controls were healthy normolipidemic individuals, as discussed in Section 2.1.3.2 of the Materials and Methods section. This data were used to evaluate the association between the *DIET1* region and hypertriglyceridemia in humans.

**Table 2.7. Primer sequence and annealing temperature of the 3 coding regions of *APOE***

| Exon | Annealing temperature (˚C) | Amplicon Size (bp) | Primer Sequence (5' to 3') |
|---|---|---|---|
| 2 | 60 | 237 | F: GGGAGGAGTCCTCACTGGCGGTTG |
| | | | R: GCCAGGAGCAGCACAGAAGCCTC |
| 3 | 60 | 303 | F: TGCCTGGACGGGGTCAGAAGGAC |
| | | | R: CTGGGGAGGTATAGCCGCCCACCAG |
| 4 | 62 | 830 | F: ATCAAGCTTTCGCCCGCCCCATCCCAGCCCTTC |
| | | | R: CGTGAATTCGCATGGCTGCAGGCTTCGGCGTTC |

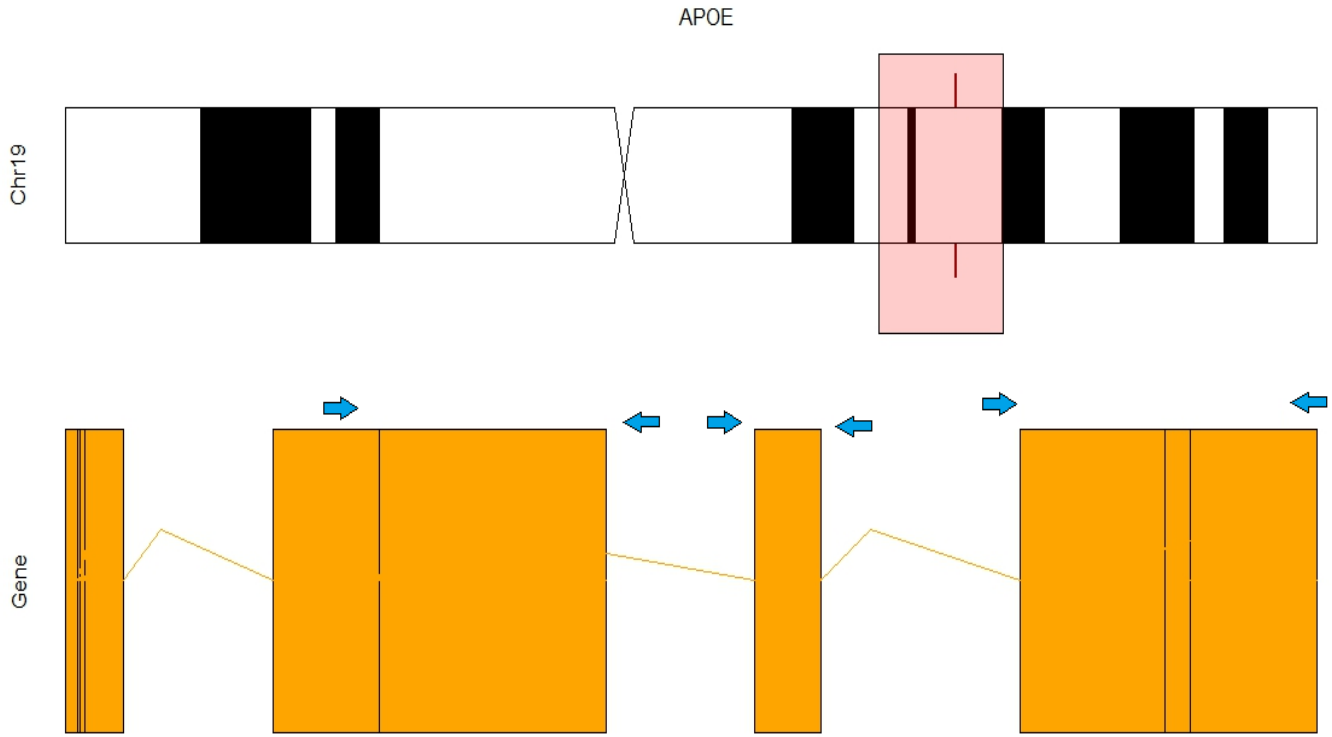



**Figure 2.6** . Gene structure of *APOE* (ENSG00000130203) (a) the chromosomal location of *APOE* on the q arm of chromosome 19. The position of *APOE* is represented by the vertical line and the region surrounding *APOE* is shaded in red (b) the rectangles are symbolic of the exons in *APOE* and the lines joining the rectangles are symbolic of the introns. The arrows represent the primers.

### 2.6 *DIET1* association analyses tested with PLINK

Single nucleotide polymorphisms (SNPs) in the *DIET 1* region, which spans chromosome region 19 10:19,377,700-20,063,500 in the hg19 genome build, were compared in cases and controls to test for association between the *DIET1* region and hypertriglyceridemia. The *DIET1* region maps to the 10p12.31 chromosomal region. All analyses was done using the free whole genome association analysis toolset, PLINK (*89*). PLINK can perform large-scale analysis of phenotype/genotype data in a computationally efficient manner. The logistic regression command was used in all analyses; SNP association was performed using logistic regression adjusted for covariates such as sex and Body Mass Index (BMI). In logistic regression, a correlation between a dependent categorical variable and a continuous or categorical independent variable is tested for (*101*). In logistic regression, no assumption is made about the distribution of the independent variable (*101*).

### 2.7 Statistical analyses

For all three projects, statistical analyses were performed using various tools on data obtained. For the first project, analyses were performed on genotype data of cases and controls using subroutines within SAS version 9.3 (SAS Institute, Cary NC). For the second project, analyses were done on the LDL-C genetic risk scores of all 88 FH patients using subroutines within SAS version 9.3 (SAS Institute, Cary NC). For the third project, analyses were done on the GWAS genotype data from HTG patients and healthy controls subroutines within PLINK. Significance in each case was set at a nominal P-value $< 0.05$, with adjustment for multiple comparisons where appropriate.

### 2.7.1 Testing for rare variant accumulation using the chi-square test

The general strategy for testing the first hypothesis was as follows: (i) record genotypes of all 138 cases and 94 controls, from Sanger sequencing into a database; and (ii) rare variants were defined as variants with a minor allele frequency of <1% in the combined cohort of FCH cases and FHTG controls. Then we tested for accumulation of only missense rare variants. Chi- square analysis was used to test for accumulation across each gene and across the four candidate genes together. Chi square tests were performed using SAS 9.3 statistical software.

### 2.7.2 Comparing mean LDL-C genetic risk score using Wilcoxon's signed rank test

Normality tests showed that the LDL-C genetic risk scores of cases and controls were not normally distributed. So, non-parametric statistical analyses were performed. Since the samples were matched, Wilcoxon's signed rank test was performed to see if there is any difference in the between means of LDL-C genetic risk scores in cases and controls. SAS Enterprise Guide 4.3 was used for Wilcoxon's signed rank tests.

### 2.7.3 Testing for SNP association between *DIET1* locus and hypertriglyceridemia using logistic regression

Logistic regression was performed to observe if there was any association between *DIET 1* region and the dichotomous trait of disease status. All analyses were done with PLINK. First, SNPs in the region of interested were selected from the GWAS

data. There were a total of **4,808 SNPs** were genotyped in the 10p12.31 region. Logistic regression was adjusted for covariates such as sex, diabetes status, population substructure and body mass index (BMI). (*101*)

# Chapter 3: Results I— Resequencing candidate genes in Familial Combined Hyperlipidemia cases and Familial Hypertriglyceridemia controls

## 3.1 Study subjects

Cases were FCH patients whose biochemical phenotype

### 3.1.1 Demographics of patients with Familial Combined Hyperlipidemia and Familial Hypertriglyceridemia

**Table 3.1** shows all the baseline clinical characteristics of 138 Familial Combined Hyperlipidemia (FCH) cases and 94 Familial Hypertriglyceridemia (FHTG) controls. The mean and standard error of body mass index (BMI), total cholesterol (TC), triglyceride (TG), low density lipoprotein cholesterol (LDL-C), high density lipoprotein cholesterol (HDL-C) are given in **Table 3.1**. Where possible, cases and controls were matched for age, sex and BMI.

## 3.2 Low Density Lipoprotein Receptor gene *LDLR*

The DNA sequence of the 18 exons, 100 bp regions flanking each exon and about 1000 bp of the promoter region of *LDLR* were sequenced and all detected variants were entered into the laboratory database. Silent variants, missense variants, intronic variants and splice site variants were found. Only rare missense variants were included for statistical analyses because of all types of variants, these are most likely to be disease causing. A list of all variants found in the *LDLR* gene in the case-control cohort is listed in the Appendix.

### 3.2.1 List of all *LDLR* missense rare variants found

I hypothesized that there is an accumulation of rare missense variants in the *LDLR* gene in cases relative to controls. **Table 3.2** shows the list of missense rare variants that were found in the *LDLR*. In total there were 7 rare missense variants in cases and 4 rare missense variants in controls. All the rare missense variants in *LDLR*, except for one ,were exclusive to either cases or controls (**Table 3.2**).

### 3.2.2 *In silico* analyses

The effect of the amino acid substitutions on LDLR function was predicted using the *in silico* tools PolyPhen-2 and SIFT. **Table 3.2** shows the *in silico* predicted effect of amino acid change on LDLR function. Rare missense variants that were predicted to be deleterious by both PolyPhen-2 and SIFT were found only in FCH cases for *LDLR.*

### 3.2.3 Test for rare missense variant accumulation

Chi squared analysis was performed to test for differences in the frequency of missense rare variants in *LDLR* in FCH cases compared to FHTG controls. **Table 3.3** shows results of the Chi squared test. There was a greater accumulation of missense rare variants in cases, as shown by an odds ratio (OR) of 1.2 (**Table 3.3**). However, this greater accumulation of missense rare variants was not significant because the confidence interval of the extended odds ratio included 1.0. In retrospect, the absence of statistical significance is not surprising given the sample size and statistical power. The power to detect a difference of this magnitude, given the sample size of the *LDLR* resequencing project was 6%.

### 3.3 Apolipoprotein B-100 gene *APOB*

The DNA sequence of the 19 amplicons of *APOB* exon 26, all 3 amplicons of *APOB* exon 29 and the 100 bp intronic regions flanking exons 26 and 29 were analyzed and variants were entered into the laboratory database. *APOB* exons 26 and 29 were chosen as target regions because these exons have each been shown to be a hotspot for FH-causing mutations. Silent variants, missense variants, deletion variants, intronic variants and splice site variants were found. As with the *LDLR* gene, only missense rare variants were included for statistical analyses. Only missense rare variants were included for statistical analyses because these variants, of all types of variants are most likely to be disease causing. All the variants found in exons 26 and 29 of *APOB* in the case-control cohort are listed in the Appendix.

#### 3.3.1 List of all *APOB* gene rare missense variants

I hypothesized that there is an accumulation of rare missense variants in *APOB* exons 26 and 29 in cases relative to controls. *APOB* exons 26 and 29 were sequenced in 138 cases and 94 controls. **Table 3.4** shows the list of rare missense variants that were found in *APOB*. In total there were 10 rare missense variants in cases and 9 rare missense variants in controls. All rare missense variants in *APOB* were exclusive to either cases or controls (**Table 3.4**).

### 3.3.2 *In silico* analyses

Effect of amino acid substitution on ApoB function was predicted using the *in silico* tools PolyPhen-2 and SIFT. **Table 3.4** shows the *in silico* predicted effect of amino acid change on ApoB function. More missense rare variants were predicted to be deleterious by both PolyPhen-2 and SIFT in cases (**Table 3.4**).

### 3.3.3 Test for rare missense variant accumulation

Chi square analysis was performed to determine whether the frequency of rare missense variants in *APOB* differed between FCH cases and FHTG controls. The accumulation of rare variants in cases was not greater as shown by an OR of 0.74 and by an extended confidence interval that included 1.00 ($P$=0.63) (**Table 3.5**). . In retrospect, the absence of statistical significance is not surprising given the sample size and statistical power. The power to detect a difference of this magnitude in *APOB* rare missense variants, given the sample size was 10%

## 3.4 Proprotein Convertase Subtilisin/kexin Type 9 gene (*PCSK9*)

The DNA sequence of exon 7 of *PCSK9* was analyzed in 138 FCH cases and 94 FHTG controls. Exon 7 of *PCSK9* was chosen as target region as this exon has recently been considered a hotspot for FH-causing mutations. No variants were found in *PCSK9* in 138 FCH cases and 94 FHTG controls, so no further confidence or statistical analyses were performed.

### 3.5 Inducible degrader of Low Density Lipoprotein receptor gene (*IDOL*)

The DNA sequence of all coding regions of *IDOL* was analyzed and patients' genotypes were entered into the laboratory database. *IDOL* has recently been implicated in cholesterol metabolism as it is a degrader of LDLR. Only missense rare variants were included for statistical analyses. List of all variants found in the 7 exons of IDOL are found in the Appendix.

#### 3.5.1 List of all missense rare variants found

We hypothesized that there is a greater accumulation of rare missense variants in coding regions of *IDOL* in FCH cases relative to FHTG controls. The 7 exons of *IDOL* were sequenced in 138 cases and 94 controls. **Table 3.6** shows the list of rare missense variants that were found in *IDOL*. In total there were 2 rare missense variants in cases and 2 rare missense variants in controls. All the rare missense variants except one were exclusive to either cases or controls in *IDOL* (**Table 3.6**).

#### 3.5.2 *In silico* analyses

The effect of amino acid substitution on IDOL function was predicted using the *in silico* tools PolyPhen-2 and SIFT. **Table 3.6** shows the *in silico* predicted effect of amino acid change on IDOL function. The rare missense variants that were exclusive to cases and controls were predicted to be deleterious by both PolyPhen-2 and SIFT (**Table 3.6**).

### 3.5.3 Test for rare missense variant accumulation

Chi square analysis was performed to determine whether the frequency of missense rare variants in *IDOL* was different in FCH cases compared to FHTG controls. **Table 3.7** shows results of the chi square analysis. There was no greater accumulation of rare missense variants in cases as shown by an OR of 0.69 and by a confidence interval that included 1.00 (*P*=1.00) (**Table 3.7**). . In retrospect, the absence of statistical significance is not surprising given the sample size and statistical power. The power to detect a difference of this magnitude, given the sample size was 7.3%.

## 3.6 Evaluating accumulation of functionally verified variants

### 3.6.1 Test for rare missense variant accumulation across candidate genes (grand total)

There was no statistically significant accumulation of rare missense variants in the individual candidate genes. Therefore, statistical accumulation of missense rare variants across all 3 main candidate genes as a grand total was determined.

Chi squared analysis was performed to determine whether the frequency of missense rare variants across *LDLR*, *APOB* and *IDOL* genes was different in FCH cases compared to FHTG controls. **Table 3.8** shows results of the chi square analysis. There was no greater accumulation of rare missense variants in cases as shown by an OR of 0.85 (**Table 3.8**). This indicates no enrichment of rare missense variants in FCH cases vs FHTG controls and that if anything, there is a non-significant trend that the frequency being higher in HTG patients.

### 3.6.2 Test for accumulation of functionally verified missense rare variants across candidate genes (grand total)

The literature was searched to date (June 1st, 2013) to determine whether any of the rare missense rare variants found in cases and controls have been functionally verified. Interestingly, 4 functionally verified rare missense variants were found only in cases, namely: *LDLR*: p.G314S, p.D333V, p.V806I and *APOB*: p.R3500W variants, at a nominal level approaching statistical significance ($P$=0.09) (**Table 3.9**). Rare missense variants that were functionally verified in literature and/or predicted deleterious by both PolyPhen-2 and SIFT were also tested for significant accumulation using Fisher's Exact test. There was a significant accumulation of rare variants as shown by an OR of 2.4, but the accumulation was non-significant ($P=0.25$) and included an OR of 1.00 (**Table 3.10**).

## 3.7 Multiple sequence alignment analyses

Multiple sequence alignment was done using Clustal Omega (*87*) and Jalview 2.8 (*88*) was used to analyze multiple sequence alignments.

### 3.7.1 Multiple Sequence Alignments for *LDLR*

According to NCBI, *LDLR* is conserved in Humans (*Homo Sapiens*), Rhesus Monkey (*Macca Mulata*), House Mouse (*Mus Musculus*), Cattle (*Bos Tarus*), Chimpanzee (*Pan troglodytes*), Dog (*Canis lupus familiaris*), Rat (*Rattus norvegicus*) and Zebrafish (*Danio rerio*). Multiple sequence alignments (MSA) *LDLR* amino acid sequence was performed to observe whether the rare missense variants found in *LDLR* were conserved (**Figure 3.1**).

The rare missense variants found in *LDLR* are marked by the red rectangles with the variant name on top. For the p.G-2R variant, the residues marked by green rectangles represent electrostatic/ionic bonds that could be formed by the p.G-2R mutation. For the p.D333V variant, the residues marked by blue rectangles represent possible electrostatic/ionic bond that may have been disrupted by the p.D333V mutation. For the p.C677G mutation, the residues marked by blue represent possible di-sulphide bonds that may be disrupted with p.C677G (**Figure 3.1**).

### 3.7.2 Multiple Sequence Alignments for *APOB*

According to NCBI, *APOB* is conserved in Humans (*Homo Sapiens*), Rhesus Monkey (*Macca Mulata*), House Mouse (*Mus Musculus*), Cattle (*Bos Tarus*), Chimpanzee (*Pan troglodytes*), Dog (*Canis lupus familiaris*), Rat (*Rattus norvegicus* ) and Zebrafish (*Danio rerio*) and Chicken (*Gallus gallus*) . Multiple sequence alignment of the amino acid sequence of *APOB* was performed to observe whether or not the rare missense variants found in *APOB* were in conserved regions of the protein (**Figure 3.2**).

The rare missense variants found in *APOB* are marked by the red rectangles with the variant name on top. For the p.K1615R mutation, the residues marked by blue rectangles represent the electrostatic bonds may be strengthened and/ or weakened by the p.K1615R mutation. For the p.E2539K, p.R1662H, p.R2192C and p.R3500W, the residues marked by the blue rectangles represent electrostatic bonds that may be disrupted with the respective mutations. For the p.E2539K and p.T3020R mutations, the residues marked by green rectangles represent electrostatic bonds that may have formed with respective residues (**Figure 3.2**).

### 3.7.3 Multiple sequence alignments for *IDOL*

According to NCBI, *IDOL* is conserved in Humans (*Homo Sapiens*), Rhesus Monkey (*Macca Mulata*), House Mouse (*Mus Musculus*), Cattle (*Bos Tarus*), Chimpanzee (*Pan troglodytes*), Dog (*Canis lupus familiaris*), Fruit fly (*Drosophila melanogaster*), Mosquito (Culicidae) and Zebrafish (*Danio rerio*) and Chicken (*Gallus gallus*) . Multiple sequence alignment of the amino acid sequence of *IDOL* was performed to observe whether the missense rare variants found in *IDOL* were in conserved regions of the protein. High conservation was an indication of importance of the residue (**Figure 3.3**).

The rare missense variants found in *IDOL* are marked by the red rectangles with the variant name on top. For the p.C31Y, the residues marked by blue rectangles represent residues that may disrupt disulphide bonds with the p.C31Y mutation. For the p.R372W mutation, the residues marked by blue rectangles represent electrostatic bonds that may be disrupted with an p.R372W mutation (**Figure 3.3**).

**Table 3.1 Mean Values (± Standard Error) of Clinical traits of FCH cases and FHTG controls**

| Population | N | Female,% | Age, years | BMI, $kg/m^2$ | TC, mmol/L | HDL-C, mmol/L | LDL-C, mmol/L | TG, mmol/L |
|---|---|---|---|---|---|---|---|---|
| FCH | 138 | 39.9 | 51.7±1.1 | 29.2±0.35 | 7.9±0.13 | 1.1±0.026 | 4.8±0.17 | 5.1±0.17 |
| FHTG | 94 | 23.4 | 52.0±1.3 | 30.4±0.47 | 5.1±0.1 | 0.81±0.024 | 2.7±0.14 | 4.9±0.16 |

**Table 3.2 Missense rare variants found in the18 exons and promoter of *LDLR* in 138 FCH cases and 94 FHTG controls. The predicted effect on protein function using *in silico* predictive tools PolyPhen-2 and SIFT are listed.**

| Variant | Population | Cases | Controls | SIFT | SIFT score | PolyPhen-2 | PolyPhen-2 score |
|---|---|---|---|---|---|---|---|
| p.T705I | Both | 1 | 2 | Tolerated | 0.51 | Benign | 0.021 |
| p.G-2R | Cases | 1 | 0 | Tolerated | 0.16 | Benign | 0.12 |
| p.G314S | Cases | 1 | 0 | Damaging | 0.04 | Probably Damaging | 1.00 |
| p.D333V | Cases | 1 | 0 | Damaging | 0 | Probably Damaging | 1.00 |
| p.L561P | Cases | 1 | 0 | Damaging | 0 | Probably Damaging | 1.00 |
| p.C677G | Cases | 1 | 0 | Damaging | 0 | Probably Damaging | 0.997 |
| p.V806I | Cases | 1 | 0 | Tolerated | 0.06 | Probably Damaging | 1.00 |
| p.T41M | Controls | 0 | 1 | Tolerated | 0.14 | Probably Damaging | 1.00 |
| p.A585S | Controls | 0 | 1 | Tolerated | 0.32 | Benign | 0.141 |

**Table 3.3 Rare missense Variant accumulation of non-synonymous rare variants, in *LDLR*, in Cases and Controls**

| Variant Selection | FCH | Controls | OR(95% CI) | *P* Value |
|---|---|---|---|---|
| Missense <1% | 7 | 4 | 1.20 (0.34-4.23) | 1.00 |

**Table 3.4 Missense rare variants found in the exons 26 and 29 of *APOB* in 138 FCH cases and 94 FHTG controls and the predicted effect on protein function using *in silico* predictive tools PolyPhen-2 and SIFT.**

| Variant | Population | Cases | Controls | SIFT | SIFT Score | PolyPhen-2 | PolyPhen-2 Score |
|---|---|---|---|---|---|---|---|
| p.C1395Y | Cases | 1 | 0 | Tolerated | 0.46 | Benign | 0.002 |
| p.E2539K | Cases | 1 | 0 | Tolerated | 0.14 | Benign | 0.112 |
| p.M2331I | Cases | 1 | 0 | Tolerated | 0.18 | Benign | 0.002 |
| p.M4293V | Cases | 1 | 0 | Tolerated | 0.09 | Benign | 0 |
| p.R1662H | Cases | 1 | 0 | Damaging | 0.01 | Probably Damaging | 1 |
| p.R3500W | Cases | 1 | 0 | Damaging | 0 | Probably Damaging | 1 |
| p.S3252G | Cases | 2 | 0 | Damaging | 0.03 | Possibly Damaging | 0.532 |
| p.T3020R | Cases | 1 | 0 | Tolerated | 0.12 | Benign | 0.206 |
| p.T3799M | Cases | 1 | 0 | Damaging | 0 | Probably Damaging | 1 |
| p.T4457M | Cases | 3 | 0 | Tolerated | 0.3 | Benign | 0.001 |
| p.I2286V | Cases | 1 | 0 | Tolerated | 0.57 | Benign | 0 |
| p.E2539D | Controls | 0 | 1 | Tolerated | 0.06 | Possibly Damaging | 0.763 |
| p.K1615R | Controls | 0 | 1 | Damaging | 0.08 | Probably Damaging | 1 |
| p.Q3405E | Controls | 0 | 2 | Damaging | 0.22 | Benign | 0.292 |
| p.R2192C | Controls | 0 | 1 | Damaging | 0 | Benign | 0.116 |
| p.S1586T | Controls | 0 | 1 | Tolerated | 0.8 | Benign | 0.006 |
| p.S3267P | Controls | 0 | 1 | Damaging | 0 | Probably Damaging | 1 |
| p.S4403T | Controls | 0 | 1 | Tolerated | 0.32 | Benign | 0.008 |
| p.V4101M | Controls | 0 | 1 | Tolerated | 0.45 | Benign | 0.007 |

**Table 3.5 Rare Missense Variant accumulation in exons26 and 29 of *APOB*, in Cases and Controls**

| Variant Selection | FCH | Controls | OR(95% CI) | *P* - Value |
|---|---|---|---|---|
| Missense <1% | 10 | 9 | 0.74(0.29-1.90) | 0.63 |

**Table 3.6 Rare missense variants found in the exons of *IDOL* in 138 FCH cases and 94 FHTG controls. The predicted effect on protein function using *in silico* predictive tools PolyPhen-2 and SIFT is listed**

| Variant | Population | Cases | Controls | SIFT | SIFT score | PolyPhen-2 | PolyPhen-2 score |
|---|---|---|---|---|---|---|---|
| p.R372W | Cases | 1 | 0 | Damaging | 0.01 | Probably Damaging | 0.990 |
| p.V339I | Both | 1 | 1 | Tolerated | 0.35 | Benign | 0.002 |
| p.C31Y | Controls | 0 | 1 | Damaging | 0 | Probably Damaging | 0.999 |

**Table 3.7. Rare missense variant accumulation of 7 exons of *IDOL*, in Cases and Controls**

| Variant Selection | FCH | Controls | OR (95% CI) | *P* -Value |
|---|---|---|---|---|
| Missense<1% | 2 | 2 | 0.69 (0.095-4.96) | 1.00 |

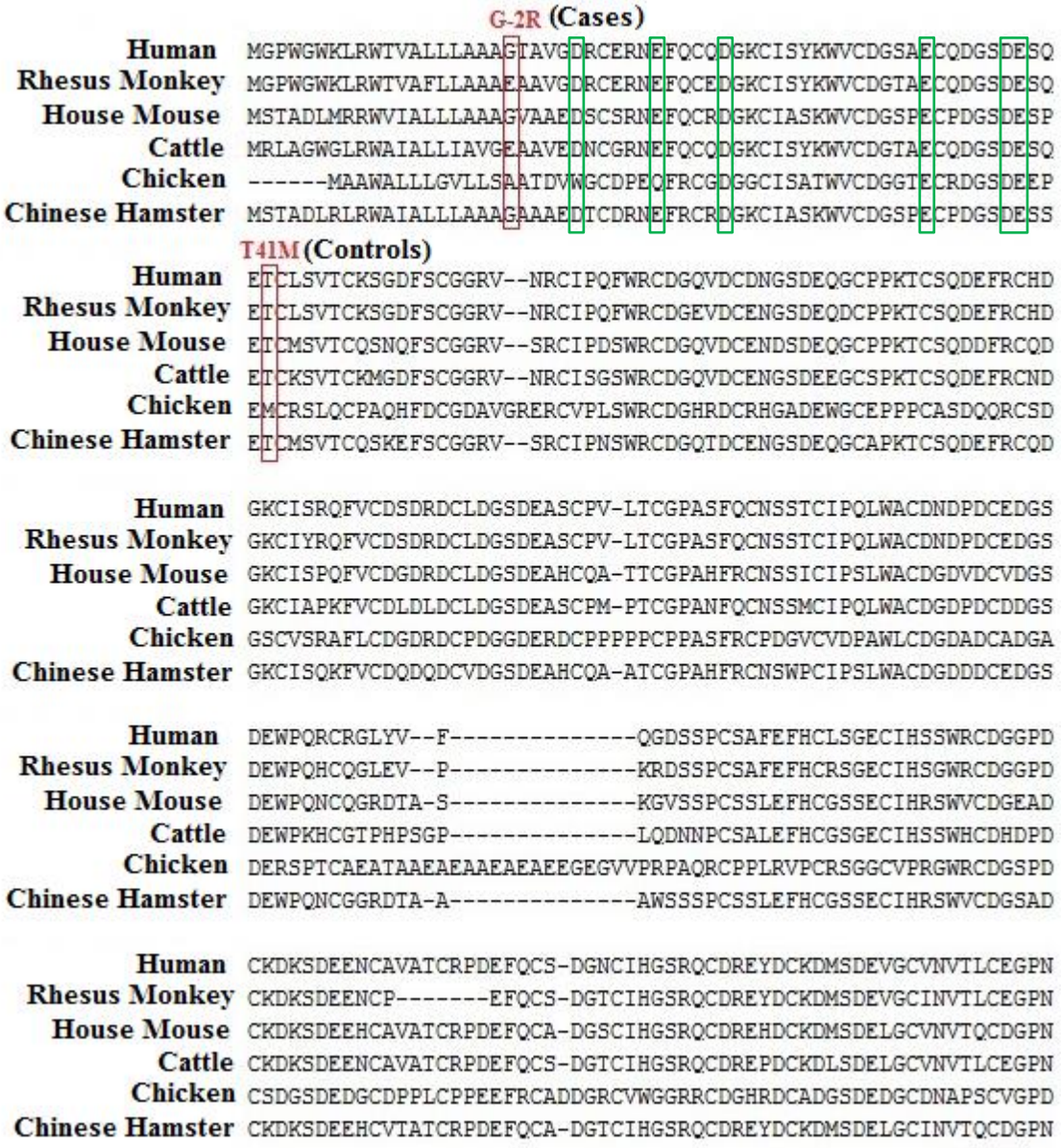
**Table 3.8 Rare missense variant accumulation in *LDLR*, *APOB* and *IDOL* in Cases and Controls**

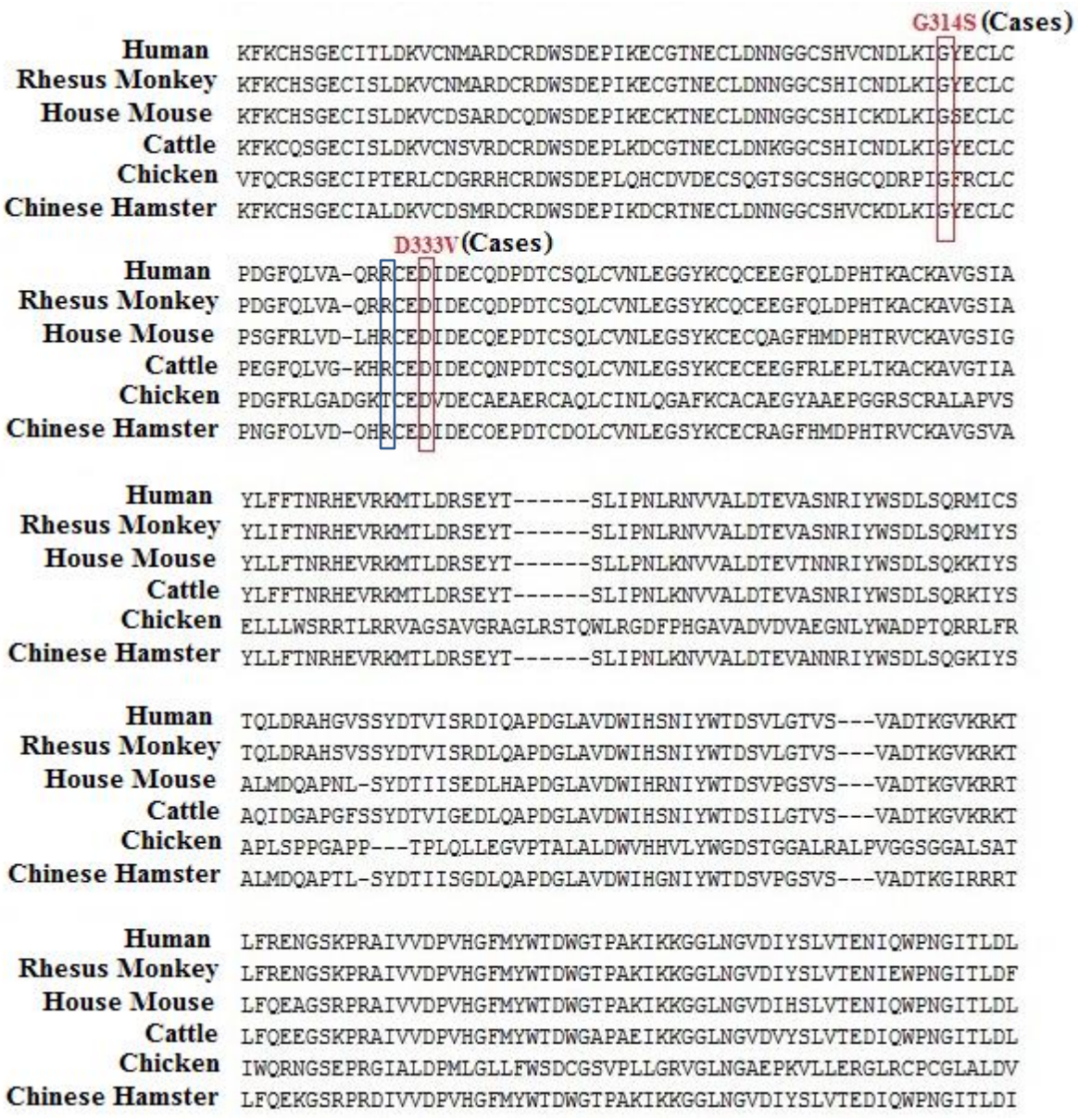
| Variant Selection | FCH | Controls | OR(95% CI) | *P* - Value |
|---|---|---|---|---|
| Missense <1% | 19 | 15 | 0.85 (0.39-1.90) | 0.68 |

**Table 3.9 Rare missense variant accumulation functionally verified and/or predicted to be deleterious by both PolyPhen-2 and SIFT, in *LDLR*, *APOB* and *IDOL*, in cases and controls**

| Variant Selection | FCH | Controls | OR (95% CI) | *P* Value |
|---|---|---|---|---|
| Missense <1% | 10 | 3 | 2.4 (0.59-11.4) | 0.25 |

**Figure 3.1 Multiple Sequence Alignment of *LDLR* . Multiple Sequence Alignment of *LDLR* amino acid sequence from 6 species, namely: Human, Rhesus Monkey, House Mouse, Cattle, Chicken and Chinese Hamster.**




G314S (Cases)
Human KFKCHSGECITLDKVCNMARDCRDWSDEPIKECGTNECLDNNGGCSHVCNDLKIGYECLC
Rhesus Monkey KFKCHSGECISLDKVCNMARDCRDWSDEPIKECGTNECLDNNGGCSHICNDLKIGYECLC
House Mouse KFKCHSGECISLDKVCDSARDCQDWSDEPIKECKTNECLDNNGGCSHICKDLKIGSECLC
Cattle KFKCQSGECISLDKVCNSVRDCRDWSDEPLKDCGTNECLDNKGGCSHICNDLKIGYECLC
Chicken VFQCRSGECIPTERLCDGRRHCRDWSDEPLQHCDVDECSQGTSGCSHGCQDRPIGFRCLC
Chinese Hamster KFKCHSGECIALDKVCDSMRDCRDWSDEPIKDCRTNECLDNNGGCSHVCKDLKIGYECLC
D333V (Cases)
Human PDGFQLVA-QRRCEDIDECQDPDTCSQLCVNLEGGYKCQCEEGFQLDPHTKACKAVGSIA
Rhesus Monkey PDGFQLVA-QRRCEDIDECQDPDTCSQLCVNLEGSYKCQCEEGFQLDPHTKACKAVGSIA
House Mouse PSGFRLVD-LHRCEDIDECQEPDTCSQLCVNLEGSYKCECQAGFHMDPHTRVCKAVGSIG
Cattle PEGFQLVG-KHRCEDIDECQNPDTCSQLCVNLEGSYKCECEEGFRLEPLTKACKAVGTIA
Chicken PDGFRLGADGKTCEDVDECAEAERCAQLCINLQGAFKCACAEGYAAEPGGRSCRALAPVS
Chinese Hamster PNGFQLVD-QHRCEDIDECQEPDTCDQLCVNLEGSYKCECRAGFHMDPHTRVCKAVGSVA
Human YLFFTNRHEVRKMTLDRSEYT------SLIPNLRNVVALDTEVASNRIYWSDLSQRMICS
Rhesus Monkey YLIFTNRHEVRKMTLDRSEYT------SLIPNLRNVVALDTEVASNRIYWSDLSQRMIYS
House Mouse YLLFTNRHEVRKMTLDRSEYT------SLLPNLKNVVALDTEVTNNRIYWSDLSQKKIYS
Cattle YLFFTNRHEVRKMTLDRSEYT------SLIPNLKNVVALDTEVASNRIYWSDLSQRKIYS
Chicken ELLLWSRRTLRRVAGSAVGRAGLRSTQWLRGDFPHGAVADVDVAEGNLYWADPTQRRLFR
Chinese Hamster YLLFTNRHEVRKMTLDRSEYT------SLIPNLKNVVALDTEVANNRIYWSDLSQGKIYS
Human TQLDRAHGVSSYDTVISRDIQAPDGLAVDWIHSNIYWTDSVLGTVS---VADTKGVKRKT
Rhesus Monkey TQLDRAHSVSSYDTVISRDLQAPDGLAVDWIHSNIYWTDSVLGTVS---VADTKGVKRKT
House Mouse ALMDQAPNL-SYDTIISEDLHAPDGLAVDWIHRNIYWTDSVPGSVS---VADTKGVKRRT
Cattle AQIDGAPGFSSYDTVIGEDLQAPDGLAVDWIHSNIYWTDSILGTVS---VADTKGVKRKT
Chicken APLSPPGAPP---TPLQLLEGVPTALALDWVHHVLYWGDSTGGALRALPVGGSGGALSAT
Chinese Hamster ALMDQAPTL-SYDTIISGDLQAPDGLAVDWIHGNIYWTDSVPGSVS---VADTKGIRRRT
Human LFRENGSKPRAIVVDPVHGFMYWTDWGTPAKIKKGGLNGVDIYSLVTENIQWPNGITLDL
Rhesus Monkey LFRENGSKPRAIVVDPVHGFMYWTDWGTPAKIKKGGLNGVDIYSLVTENIEWPNGITLDF
House Mouse LFQEAGSRPRAIVVDPVHGFMYWTDWGTPAKIKKGGLNGVDIHSLVTENIQWPNGITLDL
Cattle LFQEEGSKPRAIVVDPVHGFMYWTDWGAPAEIKKGGLNGVDVYSLVTEDIQWPNGITLDL
Chicken IWQRNGSEPRGIALDPMLGLLFWSDCGSVPLLGRVGLNGAEPKVLLERGLRCPCGLALDV
Chinese Hamster LFQEKGSRPRDIVVDPVHGFMYWTDWGTPAKIKKGGLNGVDIYSLVTEDIQWPNGITLDI

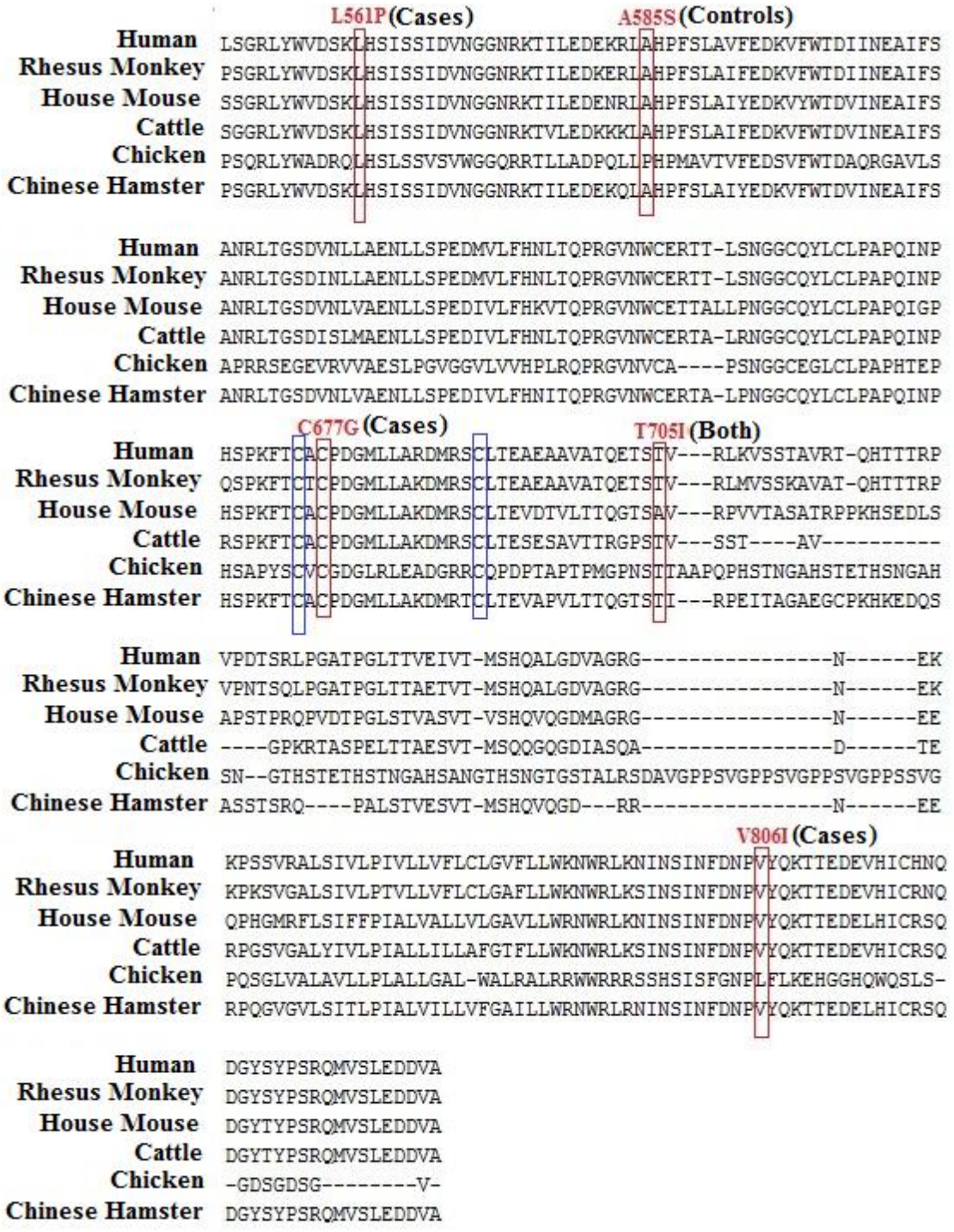
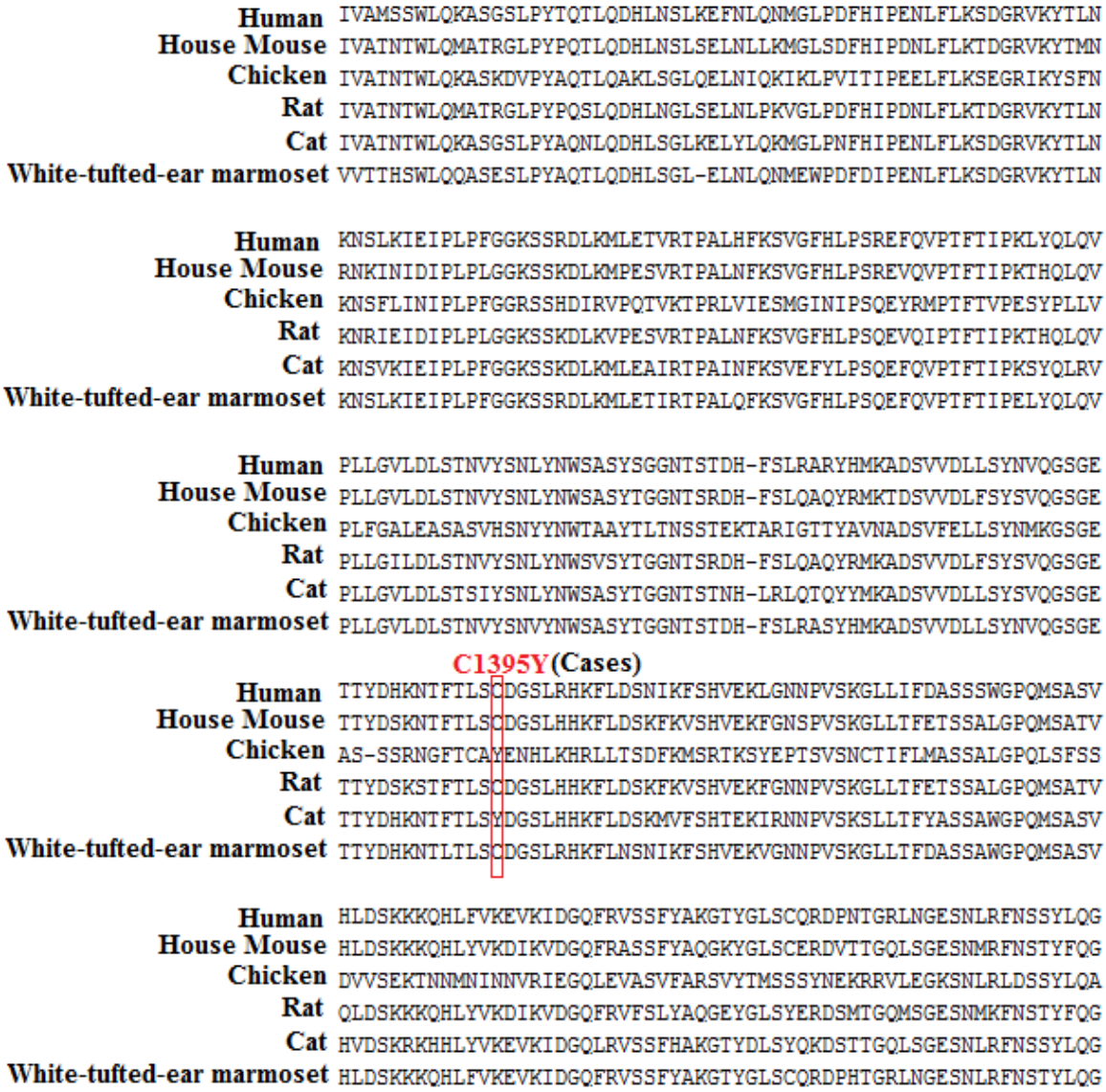
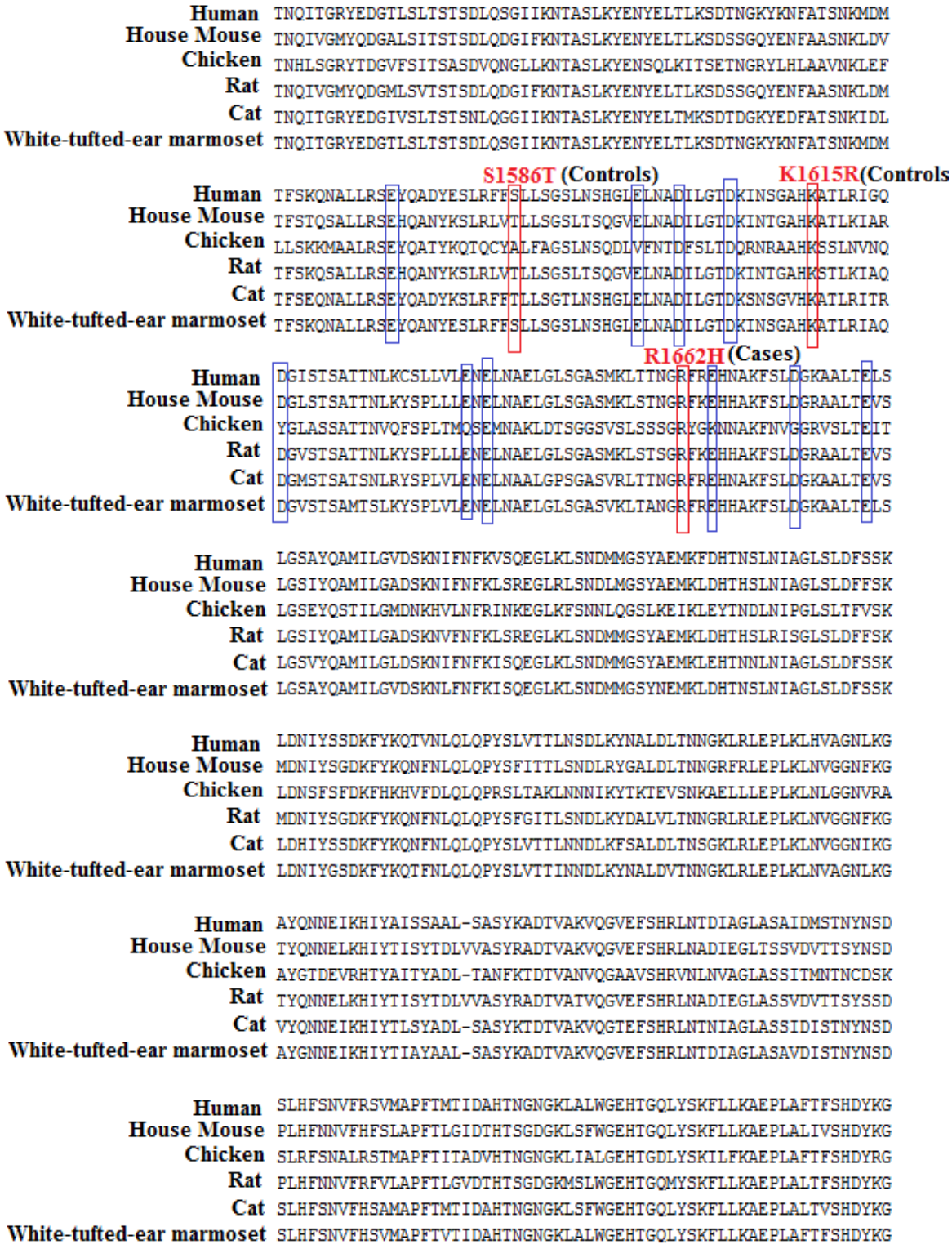
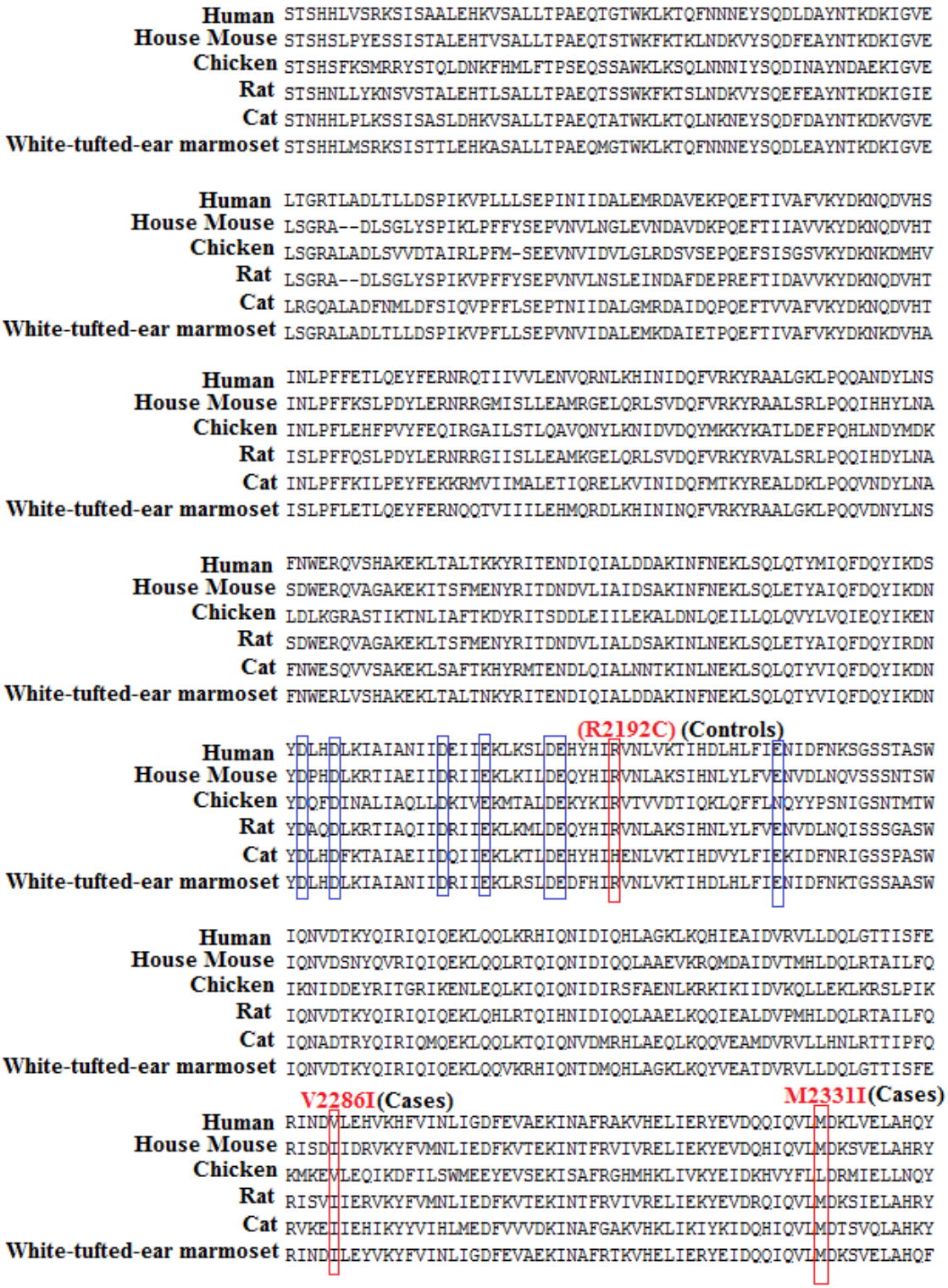
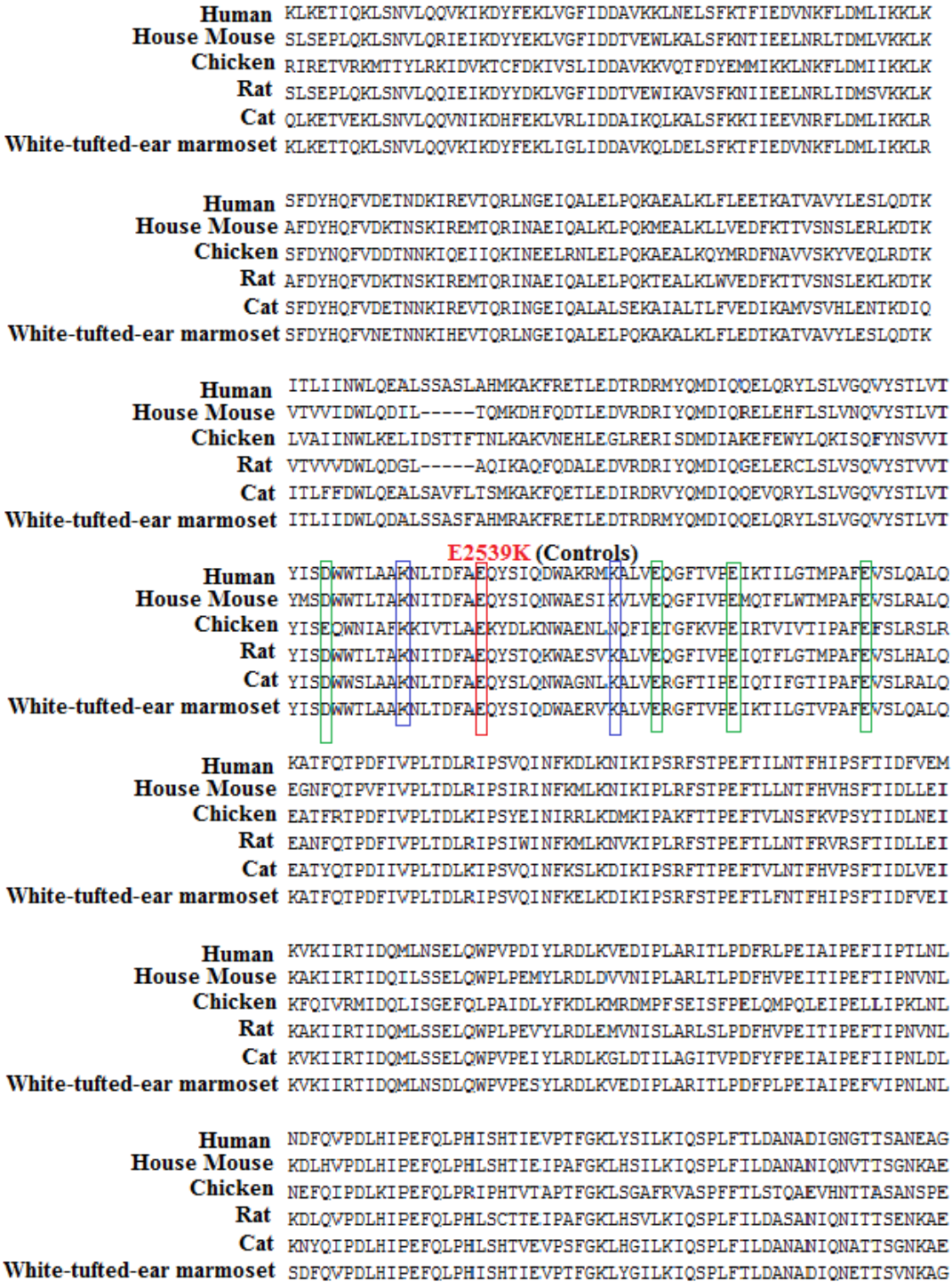

L561P (Cases)
A585S (Controls)
Human LSGRLYWVDSKLHSISSIDVNGGNRKTILEDEKRLAHPFSLAVFEDKVFWTDIINEAIFS
Rhesus Monkey PSGRLYWVDSKLHSISSIDVNGGNRKTILEDKERLAHPFSLAIFEDKVFWTDIINEAIFS
House Mouse SSGRLYWVDSKLHSISSIDVNGGNRKTILEDENRLAHPFSLAIYEDKVYWTDVINEAIFS
Cattle SGGRLYWVDSKLHSISSIDVNGGNRKTVLEDKKKLAHPFSLAIFEDKVFWTDVINEAIFS
Chicken PSQRLYWADRQLHSLSSVSVWGGQRRTLLADPQLLPHPMAVTVFEDSVFWTDAQRGAVLS
Chinese Hamster PSGRLYWVDSKLHSISSIDVNGGNRKTILEDEKQLAHPFSLAIYEDKVFWTDVINEAIFS
Human ANRLTGSDVNLLAENLLSPEDMVLFHNLTQPRGVNWCERTT-LSNGGCQYLCLPAPQINP
Rhesus Monkey ANRLTGSDINLLAENLLSPEDMVLFHNLTQPRGVNWCERTT-LSNGGCQYLCLPAPQINP
House Mouse ANRLTGSDVNLVAENLLSPEDIVLFHKVTQPRGVNWCETTALLPNGGCQYLCLPAPQIGP
Cattle ANRLTGSDISLMAENLLSPEDIVLFHNLTQPRGVNWCERTA-LRNGGCQYLCLPAPQINP
Chicken APRRSEGEVRVVAESLPGVGGVLVVHPLRQPRGVNVCA----PSNGGCEGLCLPAPHTEP
Chinese Hamster ANRLTGSDVNLVAENLLSPEDIVLFHNITQPRGVNWCERTA-LPNGGCQYLCLPAPQINP
C677G (Cases)
T705I (Both)
Human HSPKFTCACPDGMLLARDMRSCLTEAEAAVATQETSTV---RLKVSSTAVRT-QHTTTRP
Rhesus Monkey QSPKFTCTCPDGMLLAKDMRSCLTEAEAAVATQETSTV---RLMVSSKAVAT-QHTTTRP
House Mouse HSPKFTCACPDGMLLAKDMRSCLTEVDTVLTTQGTSAV---RPVVTASATRPPKHSEDLS
Cattle RSPKFTCACPDGMLLAKDMRSCLTESESAVTTRGPSTV---SST----AV----------
Chicken HSAPYSCVCGDGLRLEADGRRCQPDPTAPTPMGPNSTTAAPQPHSTNGAHSTETHSNGAH
Chinese Hamster HSPKFTCACPDGMLLAKDMRTCLTEVAPVLTTQGTSTI---RPEITAGAEGCPKHKEDQS
Human VPDTSRLPGATPGLTTVEIVT-MSHQALGDVAGRG----------N------EK
Rhesus Monkey VPNTSQLPGATPGLTTAETVT-MSHQALGDVAGRG----------N------EK
House Mouse APSTPRQPVDTPGLSTVASVT-VSHQVQGDMAGRG----------N------EE
Cattle ----GPKRTASPELTTAESVT-MSQQGQGDIASQA----------D------TE
Chicken SN--GTHSTETHSTNGAHSANGTHSNGTGSTALRSDAVGPPSVGPPSVGPPSVGPPSSVG
Chinese Hamster ASSTSRQ----PALSTVESVT-MSHQVQGD---RR----------N------EE
V806I (Cases)
Human KPSSVRALSIVLPIVLLVFLCLGVFLLWKNWRLKNINSINFDNPVYQKTTEDEVHICHNQ
Rhesus Monkey KPKSVGALSIVLPTVLLVFLCLGAFLLWKNWRLKSINSINFDNPVYQKTTEDEVHICRNQ
House Mouse QPHGMRFLSIFFPIALVALLVLGAVLLWRNWRLKNINSINFDNPVYQKTTEDELHICRSQ
Cattle RPGSVGALYIVLPIALLILLAFGTFLLWKNWRLKSINSINFDNPVYQKTTEDEVHICRSQ
Chicken PQSGLVALAVLLPLALLGAL-WALRALRRWWRRRSSHSISFGNPLFLKEHGGHQWQSLS-
Chinese Hamster RPQGVGVLSITLPIALVILLVFGAILLWRNWRLRNINSINFDNPVYQKTTEDELHICRSQ
Human DGYSYPSRQMVSLEDDVA
Rhesus Monkey DGYSYPSRQMVSLEDDVA
House Mouse DGYTYPSRQMVSLEDDVA
Cattle DGYTYPSRQMVSLEDDVA
Chicken -GDSGDSG--------V-
Chinese Hamster DGYSYPSRQMVSLEDDVA

**Figure 3.2 Multiple Sequence Alignment of *APOB*. Multiple Sequence Alignment of *APOB* amino acid sequence from 6 species, namely: Human, House Mouse, Chicken Rat, Cat and White-tufted-ear marmoset. APOB sequence is from amino acid 1231 to 4563.**

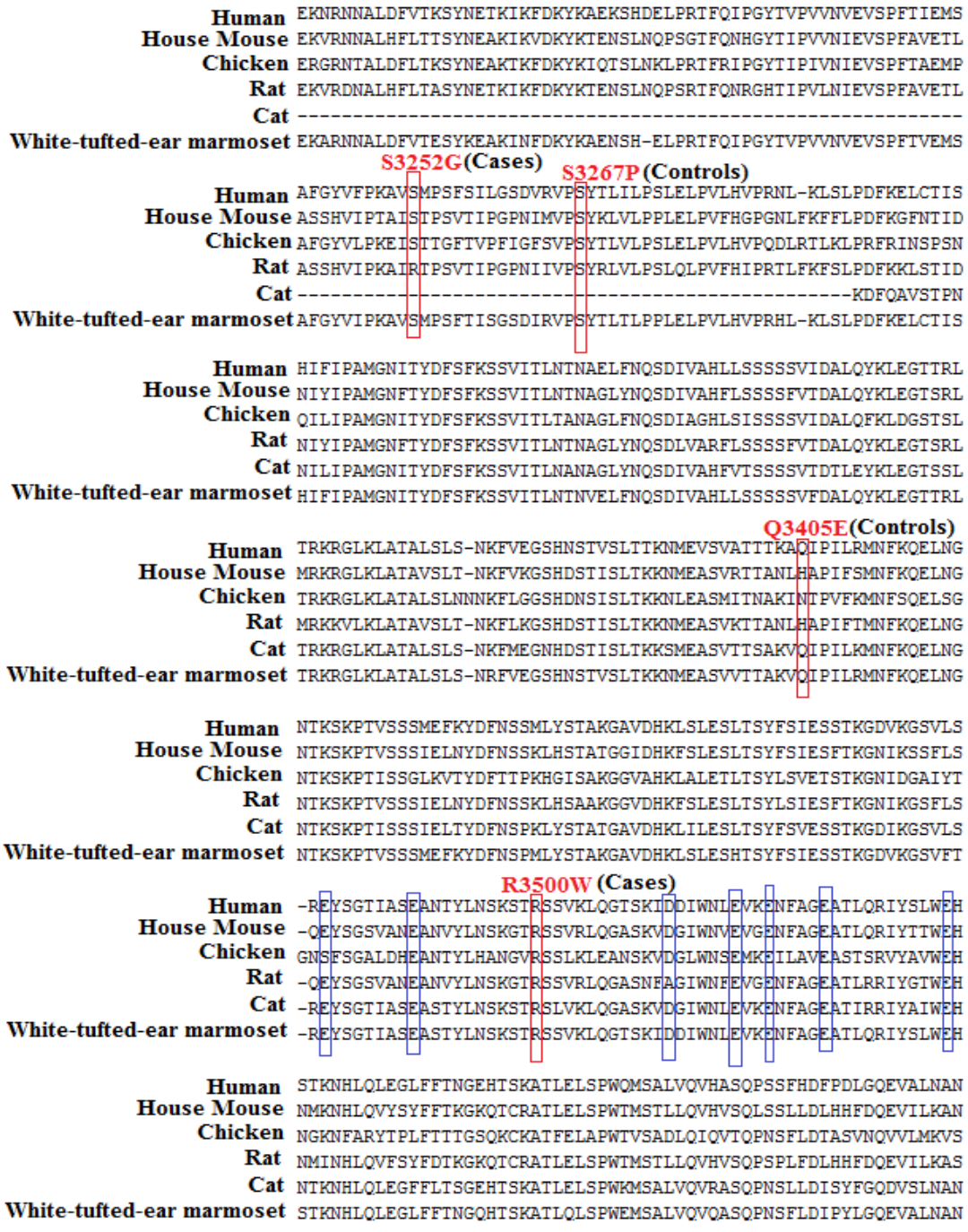



S1586T (Controls)
K1615R(Controls
R1662H (Cases)
Human TNQITGRYEDGTLSLTSTSDLQSGIIKNTASLKYENYELTLKSDTNGKYKNFATSNKMDM
House Mouse TNQIVGMYQDGALSITSTSDLQDGIFKNTASLKYENYELTLKSDSSGQYENFAASNKLDV
Chicken TNHLSGRYTDGVFSITSASDVQNGLLKNTASLKYENSQLKITSETNGRYLHLAAVNKLEF
Rat TNQIVGMYQDGMLSVTSTSDLQDGIFKNTASLKYENYELTLKSDSSGQYENFAASNKLDM
Cat TNQITGRYEDGIVSLTSTSNLQGGIIKNTASLKYENYELTMKSDTDGKYEDFATSNKIDL
White-tufted-ear marmoset TNQITGRYEDGTLSLTSTSDLQSGIIKNTASLKYENYELTLKSDTNGKYKNFATSNKMDM
Human TFSKQNALLRSEYQADYESLRFFSLLSGSLNSHGLELNADILGTDKINSGAHKATLRIGQ
House Mouse TFSTQSALLRSEHQANYKSLRLVTLLSGSLTSQGVELNADILGTDKINTGAHKATLKIAR
Chicken LLSKKMAALRSEYQATYKQTQCYALFAGSLNSQDLVFNTDFSLTDQRNRAAHKSSLNVNQ
Rat TFSKQSALLRSEHQANYKSLRLVTLLSGSLTSQGVELNADILGTDKINTGAHKSTLKIAQ
Cat TFSEQNALLRSEYQADYKSLRFFTLLSGTLNSHGLELNADILGTDKSNSGVHKATLRITR
White-tufted-ear marmoset TFSKQNALLRSEYQANYESLRFFSLLSGSLNSHGLELNADILGTDKINSGAHKATLRIAQ
Human DGISTSATTNLKCSLLVLENELNAELGLSGASMKLTTNGRFREHNAKFSLDGKAALTELS
House Mouse DGLSTSATTNLKYSPLLLENELNAELGLSGASMKLSTNGRFKEHHAKFSLDGRAALTEVS
Chicken YGLASSATTNVQFSPLTMQSEMNAKLDTSGGSVSLSSSGRYGKNNAKFNVGGRVSLTEIT
Rat DGVSTSATTNLKYSPLLLENELNAELGLSGASMKLSTSGRFKEHHAKFSLDGRAALTEVS
Cat DGMSTSATSNLRYSPLVLENELNAALGPSGASVRLTTNGRFREHNAKFSLDGKAALTEVS
White-tufted-ear marmoset DGVSTSAMTSLKYSPLVLENELNAELGLSGASVKLTANGRFREHHAKFSLDGKAALTELS
Human LGSAYQAMILGVDSKNIFNFKVSQEGLKLSNDMMGSYAEMKFDHTNSLNIAGLSLDFSSK
House Mouse LGSIYQAMILGADSKNIFNFKLSREGLRLSNDLMGSYAEMKLDHTHSLNIAGLSLDFFSK
Chicken LGSEYQSTILGMDNKHVLNFRINKEGLKFSNNLQGSLKEIKLEYTNDLNIPGLSLTFVSK
Rat LGSIYQAMILGADSKNVFNFKLSREGLKLSNDMMGSYAEMKLDHTHSLRISGLSLDFFSK
Cat LGSVYQAMILGLDSKNIFNFKISQEGLKLSNDMMGSYAEMKLEHTNNLNIAGLSLDFSSK
White-tufted-ear marmoset LGSAYQAMILGVDSKNLFNFKISQEGLKLSNDMMGSYNEMKLDHTNSLNIAGLSLDFSSK
Human LDNIYSSDKFYKQTVNLQLQPYSLVTTLNSDLKYNALDLTNNGKLRLEPLKLHVAGNLKG
House Mouse MDNIYSGDKFYKQNFNLQLQPYSFITTLSNDLRYGALDLTNNGRFRLEPLKLNVGGNFKG
Chicken LDNSFSFDKFHKHVFDLQLQPRSLTAKLNNNIKYTKTEVSNKAELLLEPLKLNLGGNVRA
Rat MDNIYSGDKFYKQNFNLQLQPYSFGITLSNDLKYDALVLTNNGRLRLEPLKLNVGGNFKG
Cat LDHIYSSDKFYKQNFNLQLQPYSLVTTLNNDLKFSALDLTNSGKLRLEPLKLNVGGNIKG
White-tufted-ear marmoset LDNIYGSDKFYKQTFNLQLQPYSLVTTINNDLKYNALDVTNNGKLRLEPLKLNVAGNLKG
Human AYQNNEIKHIYAISSAAL-SASYKADTVAKVQGVEFSHRLNTDIAGLASAIDMSTNYNSD
House Mouse TYQNNELKHIYTISYTDLVVASYRADTVAKVQGVEFSHRLNADIEGLTSSVDVTTSYNSD
Chicken AYGTDEVRHTYAITYADL-TANFKTDTVANVQGAAVSHRVNLNVAGLASSITMNTNCDSK
Rat TYQNNELKHIYTISYTDLVVASYRADTVATVQGVEFSHRLNADIEGLASSVDVTTSYSSD
Cat VYQNNEIKHIYTLSYADL-SASYKTDTVAKVQGTEFSHRLNTNIAGLASSIDISTNYNSD
White-tufted-ear marmoset AYGNNEIKHIYTIAYAAL-SASYKADTVAKVQGVEFSHRLNTDIAGLASAVDISTNYNSD
Human SLHFSNVFRSVMAPFTMTIDAHTNGNGKLALWGEHTGQLYSKFLLKAEPLAFTFSHDYKG
House Mouse PLHFNNVFHFSLAPFTLGIDTHTSGDGKLSFWGEHTGQLYSKFLLKAEPLALIVSHDYKG
Chicken SLRFSNALRSTMAPFTITADVHTNGNGKLIALGEHTGDLYSKILFKAEPLAFTFSHDYRG
Rat PLHFNNVFRFVLAPFTLGVDTHTSGDGKMSLWGEHTGQMYSKFLLKAEPLALTFSHDYKG
Cat SLHFSNVFHSAMAPFTMTIDAHTNGNGKLSFWGEHTGQLYSKFLLKAEPLALTVSHDYKG
White-tufted-ear marmoset SLHFSNVFHSVMAPFTVTIDAHTNGNGKLALWGEHTGQLYSKFLLKAEPLAFTFSHDYKG


Human STSHHLVSRKSISAALEHKVSALLTPAEQTGTWKLKTQFNNNEYSQDLDAYNTKDKIGVE
House Mouse STSHSLPYESSISTALEHTVSALLTPAEQTSTWKFKTKLNDKVYSQDFEAYNTKDKIGVE
Chicken STSHSFKSMRRYSTQLDNKFHMLFTPSEQSSAWKLKSQLNNNIYSQDINAYNDAEKIGVE
Rat STSHNLLYKNSVSTALEHTLSALLTPAEQTSSWKFKTSLNDKVYSQEFEAYNTKDKIGIE
Cat STNHHLPLKSSISASLDHKVSALLTPAEQTATWKLKTQLNKNEYSQDFDAYNTKDKVGVE
White-tufted-ear marmoset STSHHLMSRKSISTTLEHKASALLTPAEQMGTWKLKTQFNNNEYSQDLEAYNTKDKIGVE
Human LTGRTLADLTLLDSPIKVPLLLSEPINIIDALEMRDAVEKPQEFTIVAFVKYDKNQDVHS
House Mouse LSGRA--DLSGLYSPIKLPFFYSEPVNVLNGLEVNDAVDKPQEFTIIAVVKYDKNQDVHT
Chicken LSGRALADLSVVDTAIRLPFM-SEEVNVIDVLGLRDSVSEPQEFSISGSVKYDKNKDMHV
Rat LSGRA--DLSGLYSPIKVPFFYSEPVNVLNSLEINDAFDEPREFTIDAVVKYDKNQDVHT
Cat LRGQALADFNMLDFSIQVPFFLSEPTNIIDALGMRDAIDQPQEFTVVAFVKYDKNQDVHT
White-tufted-ear marmoset LSGRALADLTLLDSPIKVPFLLSEPVNVIDALEMKDAIETPQEFTIVAFVKYDKNKDVHA
Human INLPFFETLQEYFERNRQTIIVVLENVQRNLKHINIDQFVRKYRAALGKLPQQANDYLNS
House Mouse INLPFFKSLPDYLERNRRGMISLLEAMRGELQRLSVDQFVRKYRAALSRLPQQIHHYLNA
Chicken INLPFLEHFPVYFEQIRGAILSTLQAVQNYLKNIDVDQYMKKYKATLDEFPQHLNDYMDK
Rat ISLPFFQSLPDYLERNRRGIISLLEAMKGELQRLSVDQFVRKYRVALSRLPQQIHDYLNA
Cat INLPFFKILPEYFEKKRMVIIMALETIQRELKVINIDQFMTKYREALDKLPQQVNDYLNA
White-tufted-ear marmoset ISLPFLETLQEYFERNQQTVIIILEHMQRDLKHININQFVRKYRAALGKLPQQVDNYLNS
Human FNWERQVSHAKEKLTALTKKYRITENDIQIALDDAKINFNEKLSQLQTYMIQFDQYIKDS
House Mouse SDWERQVAGAKEKITSFMENYRITDNDVLIAIDSAKINFNEKLSQLETYAIQFDQYIKDN
Chicken LDLKGRASTIKTNLIAFTKDYRITSDDLEIILEKALDNLQEILLQLQVYLVQIEQYIKEN
Rat SDWERQVAGAKEKLTSFMENYRITDNDVLIALDSAKINLNEKLSQLETYAIQFDQYIRDN
Cat FNWESQVVSAKEKLSAFTKHYRMTENDLQIALNNTKINLNEKLSQLQTYVIQFDQYIKDN
White-tufted-ear marmoset FNWERLVSHAKEKLTALTNKYRITENDIQIALDDAKINFNEKLSQLQTYVIQFDQYIKDN
(R2192C) (Controls)
Human YDLHDLKIAIANIIDEIIEKLKSLDEHYHIRVNLVKTIHDLHLFIENIDFNKSGSSTASW
House Mouse YDPHDLKRTIAEIIDRIIEKLKILDEQYHIRVNLAKSIHNLYLFVENVDLNQVSSSNTSW
Chicken YDQFDINALIAQLLDKIVEKMTALDEKYKIRVTVVDTIQKLQFFLNQYYPSNIGSNTMTW
Rat YDAQDLKRTIAQIIDRIIEKLKMLDEQYHIRVNLAKSIHNLYLFVENVDLNQISSSGASW
Cat YDLHDFKTAIAEIIDQIIEKLKTLDEHYHIHENLVKTIHDVYLFIEKIDFNRIGSSPASW
White-tufted-ear marmoset YDLHDLKIAIANIIDRIIEKLRSLDEDFHIRVNLVKTIHDLHLFIENIDFNKTGSSAASW
Human IQNVDTKYQIRIQIQEKLQQLKRHIQNIDIQHLAGKLKQHIEAIDVRVLLDQLGTTISFE
House Mouse IQNVDSNYQVRIQIQEKLQQLRTQIQNIDIQQLAAEVKRQMDAIDVTMHLDQLRTAILFQ
Chicken IKNIDDEYRITGRIKENLEQLKIQIQNIDIRSFAENLKRKIKIIDVKQLLEKLKRSLPIK
Rat IQNVDTKYQIRIQIQEKLQHLRTQIHNIDIQQLAAELKQQIEALDVPMHLDQLRTAILFQ
Cat IQNADTRYQIRIQMQEKLQQLKTQIQNVDMRHLAEQLKQQVEAMDVRVLLHNLRTTIPFQ
White-tufted-ear marmoset IQNVDTKYQIRIQIQEKLQQVKRHIQNTDMQHLAGKLKQYVEATDVRVLLDQLGTTISFE
V2286I(Cases)
M2331I(Cases)
Human RINDVLEHVKHFVINLIGDFEVAEKINAFRAKVHELIERYEVDQQIQVLMDKLVELAHQY
House Mouse RISDIIDRVKYFVMNLIEDFKVTEKINTFRVIVRELIEKYEVDQHIQVLMDKSVELAHRY
Chicken KMKEVLEQIKDFILSWMEEYEVSEKISAFRGHMHKLIVKYEIDKHVYFLLDRMIELLNQY
Rat RISVIIERVKYFVMNLIEDFKVTEKINTFRVIVRELIEKYEVDRQIQVLMDKSIELAHRY
Cat RVKEIIEHIKYYVIHLMEDFVVVDKINAFGAKVHKLIKIYKIDQHIQVLMDTSVQLAHKY
White-tufted-ear marmoset RINDILEYVKYFVINLIGDFEVAEKINAFRTKVHELIERYEIDQQIQVLMDKSVELAHQF


Human KLKETIQKLSNVLQQVKIKDYFEKLVGFIDDAVKKLNELSFKTFIEDVNKFLDMLIKKLK
House Mouse SLSEPLQKLSNVLQRIEIKDYYEKLVGFIDDTVEWLKALSFKNTIEELNRLTDMLVKKLK
Chicken RIRETVRKMTTYLRKIDVKTCFDKIVSLIDDAVKKVQTFDYEMMIKKLNKFLDMIIKKLK
Rat SLSEPLQKLSNVLQQIEIKDYYDKLVGFIDDTVEWIKAVSFKNIIEELNRLIDMSVKKLK
Cat QLKETVEKLSNVLQQVNIKDHFEKLVRLIDDAIKQLKALSFKKIIEEVNRFLDMLIKKLR
White-tufted-ear marmoset KLKETTQKLSNVLQQVKIKDYFEKLIGLIDDAVKQLDELSFKTFIEDVNKFLDMLIKKLR
Human SFDYHQFVDETNDKIREVTQRLNGEIQALELPQKAEALKLFLEETKATVAVYLESLQDTK
House Mouse AFDYHQFVDKTNSKIREMTQRINAEIQALKLPQKMEALKLLVEDFKTTVSNSLERLKDTK
Chicken SFDYNQFVDDTNNKIQEIIQKINEELRNLELPQKAEALKQYMRDFNAVVSKYVEQLRDTK
Rat AFDYHQFVDKTNSKIREMTQRINAEIQALELPQKTEALKLWVEDFKTTVSNSLEKLKDTK
Cat SFDYHQFVDETNNKIREVTQRINGEIQALALSEKAIALTLFVEDIKAMVSVHLENTKDIQ
White-tufted-ear marmoset SFDYHQFVNETNNKIHEVTQRLNGEIQALELPQKAKALKLFLEDTKATVAVYLESLQDTK
Human ITLIINWLQEALSSASLAHMKAKFRETLEDTRDRMYQMDIQQELQRYLSLVGQVYSTLVT
House Mouse VTVVIDWLQDIL-----TQMKDHFQDTLEDVRDRIYQMDIQRELEHFLSLVNQVYSTLVT
Chicken LVAIINWLKELIDSTTFTNLKAKVNEHLEGLRERISDMDIAKEFEWYLQKISQFYNSVVI
Rat VTVVVDWLQDGL-----AQIKAQFQDALEDVRDRIYQMDIQGELERCLSLVSQVYSTVVT
Cat ITLFFDWLQEALSAVFLTSMKAKFQETLEDIRDRVYQMDIQQEVQRYLSLVGQVYSTLVT
White-tufted-ear marmoset ITLIIDWLQDALSSASFAHMRAKFRETLEDTRDRMYQMDIQQELQRYLSLVGQVYSTLVT
E2539K (Controls)
Human YISDWWTLAAKNLTDFAEQYSIQDWAKRMKALVEQGFTVPEIKTILGTMPAFEVSLQALQ
House Mouse YMSDWWTLTAKNITDFAEQYSIQNWAESIKVLVEQGFIVPEMQTFLWTMPAFEVSLRALQ
Chicken YISEQWNIAFKKIVTLAEKYDLKNWAENLNQFIETGFKVPEIRTVIVTIPAFEFSLRSLR
Rat YISDWWTLTAKNITDFAEQYSTQKWAESVKALVEQGFIVPEIQTFLGTMPAFEVSLHALQ
Cat YISDWWSLAAKNLTDFAEQYSLQNWAGNLKALVERGFTIPEIQTIFGTIPAFEVSLRALQ
White-tufted-ear marmoset YISDWWTLAAKNLTDFAEQYSIQDWAERVKALVERGFTVPEIKTILGTVPAFEVSLQALQ
Human KATFQTPDFIVPLTDLRIPSVQINFKDLKNIKIPSRFSTPEFTILNTFHIPSFTIDFVEM
House Mouse EGNFQTPVFIVPLTDLRIPSIRINFKMLKNIKIPLRFSTPEFTLLNTFHVHSFTIDLLEI
Chicken EATFRTPDFIVPLTDLKIPSYEINIRRLKDMKIPAKFTTPEFTVLNSFKVPSYTIDLNEI
Rat EANFQTPDFIVPLTDLRIPSIWINFKMLKNVKIPLRFSTPEFTLLNTFRVRSFTIDLLEI
Cat EATYQTPDIIVPLTDLKIPSVQINFKSLKDIKIPSRFTTPEFTVLNTFHVPSFTIDLVEI
White-tufted-ear marmoset KATFQTPDFIVPLTDLRIPSVQINFKELKDIKIPSRFSTPEFTLFNTFHIPSFTIDFVEI
Human KVKIIRTIDQMLNSELQWPVPDIYLRDLKVEDIPLARITLPDFRLPEIAIPEFIIPTLNL
House Mouse KAKIIRTIDQILSSELQWPLPEMYLRDLDVVNIPLARLTLPDFHVPEITIPEFTIPNVNL
Chicken KFQIVRMIDQLISGEFQLPAIDLYFKDLKMRDMPFSEISFPELQMPQLEIPELLIPKLNL
Rat KAKIIRTIDQMLSSELQWPLPEVYLRDLEMVNISLARLSLPDFHVPEITIPEFTIPNVNL
Cat KVKIIRTIDQMLSSELQWPVPEIYLRDLKGLDTILAGITVPDFYFPEIAIPEFIIPNLDL
White-tufted-ear marmoset KVKIIRTIDQMLNSDLQWPVPESYLRDLKVEDIPLARITLPDFPLPEIAIPEFVIPNLNL
Human NDFQVPDLHIPEFQLPHISHTIEVPTFGKLYSILKIQSPLFTLDANADIGNGTTSANEAG
House Mouse KDLHVPDLHIPEFQLPHLSHTIEIPAFGKLHSILKIQSPLFILDANANIQNVTTSGNKAE
Chicken NEFQIPDLKIPEFQLPRIPHTVTAPTFGKLSGAFRVASPFFTLSTQAEVHNTTASANSPE
Rat KDLQVPDLHIPEFQLPHLSCTTEIPAFGKLHSVLKIQSPLFILDASANIQNITTSENKAE
Cat KNYQIPDLHIPEFQLPHLSHTVEVPSFGKLHGILKIQSPLFILDANANIQNATTSGNKAE
White-tufted-ear marmoset SDFQVPDLHIPEFQLPHISHTIEVPTFGKLYGILKIQSPLFTLDANADIQNETTSVNKAG

```
Human                          IAASITAKGESKLEVLNFDFQANAQLSNPKINPLALKESVKFSSKYLRTEHGSEMLFFGN
House Mouse                    IVASVTAKGESQFEALNFDFQAQAQFLELNPHPPVLKESMNFSSKHVRMEHEGEIVFDGK
Chicken                        FVTSLSAQATSKLDFLVFSVIADSHILAPEMKQLKLKNSMKVSHKFLKIDHTNEVVFLGT
Rat                            IVASVTARGESKFEALNFDFQAQAQFLELNANPLVLKESVNFSSKHVRMEHEGKILVSGK
Cat                            IAASITAKGESKLEVLNFDFQANGQLSDSNINPLVMKQSMRFSSKYIKTEHESEVLFFGN
White-tufted-ear marmoset      IAASITAKGESRLEVLSFDFQAKAQLSTPKINPLALKESVKFSSKYLRTEHDSEMLFFGN

Human                          AIEGKSNTVASLHTEKNTLELSNGVIVKINNQLTLDSNTKYFHKLNIPKLDFSSQADLRN
House Mouse                    AIEGKSDTVASLHTEKNEVEFNNGMTVKVNNQLTLDSHTKYFHKLSVPRLDFSSKASLNN
Chicken                        SVSGEAETRAKFSAIKNSIELQNNLMVNLQRKIQMQSGTAYSHRLNIPEADFSSQADLVN
Rat                            ALEGKSDTVARLHTEKNTVEFNNGIVVKINNQFTLDSQTKYFHKLSVPRLDFSSKASLNN
Cat                            AVEGKSNTVAGLHTEKNMLEFSNGVFVRINNQLTLDSNTKYFHKLNIPKLDFSSQAELRN
White-tufted-ear marmoset      AIEGKSNTVASLHTEKNTLELSNGAIVKINNQLTLDSNTKYFHNLNIPKLDFSSQADLRN

Human                          EIKTLLKAGHIAWTSSGKGSWKWACPRFSDEGTHESQISFTIEGPLTSFGLSNKINSKHL
House Mouse                    EIKTLLEAGHVALTSSGTGSWNWACPNFSDEGIHSSQISFTVDGPIAFVGLSNNINGKHL
Chicken                        NMTTEVEAGRISFTSNGKGNWKWTSPNFSDEGTHNSHATFRVDGPILIFFADYRINDRYL
Rat                            EIKTLLEAGHMAWTSSGTGSWNWACPNFSDEGIHSSKISFIVDGPIASFGLSNNINGKHL
Cat                            EIKTLLEAGHVAWTSSGIGSWKWAHQTFSDEGAHESQVSFTVEGPITSFGLSNKISSKHL
White-tufted-ear marmoset      EVKTLLEAGHVAWTSSGKGSWKWACPHFSDEGTHESQISFTMEGPLTSFELSNKINSKHL

Human                          RVNQNLVYESGSLNFSKLEIQSQVDSQHVGHSVLTAKGMALFGEGKAEFTGRHDAHLNGK
House Mouse                    RVIQKLTYESGFLNYSKFEVESKVESQHVGSSILTANGRALLKDAKAEMTGEHNANLNGK
Chicken                        KVSQSMRYECGFLSYATLQVQSEIESQRVGRSILNVKGTGQLGGMKVELTGSHNARLNGR
Rat                            RVVQKLTSESGFLNYSRFEVESKVESQHVGSSILTAEGRALLGDAKAEMTGEHNANLNGK
Cat                            RLSQNLVYESGFLDFSKFEIQSQVESQYVGRSVLAAKGTALLGERKAEITGNHDAHLNGK
White-tufted-ear marmoset      TVNQNLVYESGFLNFSKLEIQSRAESQHVGRSVLTAKGTALFGEGKAELTGRHDAHLNGK

                                                T3020R (Cases)
Human                          VIGTLKNSLFFSAQPFEITASTNNEGNLKVRFPLRLTGKIDFLNNYALFLSPSAQQASWQ
House Mouse                    VIGTLKNSLFFSAQPFEITASTNNEGNLKVGFPLKLTGKIDFLNNYALFLSPRAQQASWQ
Chicken                        ITGTVNNDFFFLVQPFEIRLLTNNDGNVKISFPMKLTGKIDFLNNYGLSLSSSVQQVSWQ
Rat                            VIGTLKNSLFFSAQPFEITASTNNEGNLKVSFPLKLTGKIDFLNNYALFLSPHAQQASWQ
Cat                            VIGTLKNSLFFTAQPFEITASTNNEGNLKVSFPLKLIGKVDFLNNYALFLSPSAQQASWQ
White-tufted-ear marmoset      VIGTLKNSLFFSAQPFEITVSTNNEGNLKVSFPIRVTGKIDFLNNYALFLSPSAQQAGWQ

Human                          VSARFNQYKYNQNFSAGNNENIMEAHVGINGEANLDFLNIPLTIPEMRLPYTIITTPPLK
House Mouse                    ASTRFNQYKYNQNFSAINNEHNIEASIGMNGDANLDFLNIPLTIPEINLPYTEFKTPLLK
Chicken                        ATGRFNQYRYSHNMSAGNNDDRIEAHVEMSGDANLDFLNIPLTIPQLHIPYTGIQTPQLK
Rat                            LSTRFNQYKYNQNFSAINNEHNMEASIVMNGDANLDFLNIPLTIPEINLPYTRFTTPLLK
Cat                            ASARFNQYKYHQNFSAGNNENSIEAHVGINGEANLDFLNIPLTIPEMTLPYTGLTTPQVK
White-tufted-ear marmoset      VSARFNQYKYSQNFSAENNENNMEAHVGINGEANLDFLNFPLTIPEMLLPYTAITTSPQK

Human                          DFSLWEKTGLKEFLKTTKQSFDLSVKAQYKKNKHRHSITNPLAVLCEFISQSIKSFDRHF
House Mouse                    DFSIWEETGLKEFLKTTKQSFDLSVKAQYKKNSDKHSIVVPLGMFYEFILNNVNSWDRKF
Chicken                        DYSLWEQAGLKNLLKTTRQSFDLNLNAQYEKNKDMHVIPLPLATVHEALNKYIIFFNKYF
Rat                            DFSIWEETGLKEFLKTTKQSFDLSIKAQYKKNRDKHSVVIPLKMFYEFMLNNVNSWDRKF
Cat                            DFSLWEKTGLKEFLKTTKQSFDLSVKAQYKKNKDKHSIPIPFYV----------------
White-tufted-ear marmoset      NFSLWEKTGLKELLKTTKQSFDLSIKAQYKKNKHKHSIANPLAVLCEFISQNIKSFDRHL
```


Human EKNRNNALDFVTKSYNETKIKFDKYKAEKSHDELPRTFQIPGYTVPVVNVEVSPFTIEMS
House Mouse EKVRNNALHFLTTSYNEAKIKVDKYKTENSLNQPSGTFQNHGYTIPVVNIEVSPFAVETL
Chicken ERGRNTALDFLTKSYNEAKTKFDKYKIQTSLNKLPRTFRIPGYTIPIVNIEVSPFTAEMP
Rat EKVRDNALHFLTASYNETKIKFDKYKTENSLNQPSRTFQNRGHTIPVLNIEVSPFAVETL
Cat ----------
White-tufted-ear marmoset EKARNNALDFVTESYKEAKINFDKYKAENSH-ELPRTFQIPGYTVPVVNVEVSPFTVEMS
S3252G(Cases)
S3267P(Controls)
Human AFGYVFPKAVSMPSFSILGSDVRVPSYTLILPSLELPVLHVPRNL-KLSLPDFKELCTIS
House Mouse ASSHVIPTAISTPSVTIPGPNIMVPSYKLVLPPLELPVFHGPGNLFKFFLPDFKGFNTID
Chicken AFGYVLPKEISTTGFTVPFIGFSVPSYTLVLPSLELPVLHVPQDLRTLKLPRFRINSPSN
Rat ASSHVIPKAIRTPSVTIPGPNIIVPSYRLVLPSLQLPVFHIPRTLFKFSLPDFKKLSTID
Cat ----------KDFQAVSTPN
White-tufted-ear marmoset AFGYVIPKAVSMPSFTISGSDIRVPSYTLTLPPLELPVLHVPRHL-KLSLPDFKELCTIS
Human HIFIPAMGNITYDFSFKSSVITLNTNAELFNQSDIVAHLLSSSSSVIDALQYKLEGTTRL
House Mouse NIYIPAMGNFTYDFSFKSSVITLNTNAGLYNQSDIVAHFLSSSSFVTDALQYKLEGTSRL
Chicken QILIPAMGNITYDFSFKSSVITLTANAGLFNQSDIAGHLSISSSSVIDALQFKLDGSTSL
Rat NIYIPAMGNFTYDFSFKSSVITLNTNAGLYNQSDLVARFLSSSSFVTDALQYKLEGTSRL
Cat NILIPAMGNITYDFSFKSSVITLNANAGLYNQSDIVAHFVTSSSSVIDTLEYKLEGTSSL
White-tufted-ear marmoset HIFIPAMGNITYDFSFKSSVITLNTNVELFNQSDIVAHLLSSSSSVFDALQYKLEGTTRL
Q3405E(Controls)
Human TRKRGLKLATALSLS-NKFVEGSHNSTVSLTTKNMEVSVATTTKAQIPILRMNFKQELNG
House Mouse MRKRGLKLATAVSLT-NKFVKGSHDSTISLTKKNMEASVRTTANLHAPIFSMNFKQELNG
Chicken TRKRGLKLATALSLNNNKFLGGSHDNSISLTKKNLEASMITNAKINTPVFKMNFSQELSG
Rat MRKKVLKLATAVSLT-NKFLKGSHDSTISLTKKNMEASVKTTANLHAPIFTMNFKQELNG
Cat TRKRGLKLATALSLS-NKFMEGNHDSTISLTKKSMEASVTTSAKVQIPILKMNFKQELNG
White-tufted-ear marmoset TRKRGLKLATALSLS-NRFVEGSHNSTVSLTKKNMEASVVTTAKVQIPILRMNFKQELNG
Human NTKSKPTVSSSMEFKYDFNSSMLYSTAKGAVDHKLSLESLTSYFSIESSTKGDVKGSVLS
House Mouse NTKSKPTVSSSIELNYDFNSSKLHSTATGGIDHKFSLESLTSYFSIESFTKGNIKSSFLS
Chicken NTKSKPTISSGLKVTYDFTTPKHGISAKGGVAHKLALETLTSYLSVETSTKGNIDGAIYT
Rat NTKSKPTVSSSIELNYDFNSSKLHSAAKGGVDHKFSLESLTSYLSIESFTKGNIKGSFLS
Cat NTKSKPTISSSIELTYDFNSPKLYSTATGAVDHKLILESLTSYFSVESSTKGDIKGSVLS
White-tufted-ear marmoset NTKSKPTVSSSMEFKYDFNSPMLYSTAKGAVDHKLSLESHTSYFSIESSTKGDVKGSVFT
R3500W (Cases)
Human -REYSGTIASEANTYLNSKSTRSSVKLQGTSKIDDIWNLEVKENFAGEATLQRIYSLWEH
House Mouse -QEYSGSVANEANVYLNSKGTRSSVRLQGASKVDGIWNVEVGENFAGEATLQRIYTTWEH
Chicken GNSFSGALDHEANTYLHANGVRSSLKLEANSKVDGLWNSEMKEILAVEASTSRVYAVWEH
Rat -QEYSGSVANEANVYLNSKGTRSSVRLQGASNFAGIWNFEVGENFAGEATLRRIYGTWEH
Cat -REYSGTIASEASTYLNSKSTRSLVKLQGASKVDGIWNLEVKENFAGEATIRRIYAIWEH
White-tufted-ear marmoset -REYSGTIASEASTYLNSKSTRSSVKLQGTSKIDDIWNLEVKENFAGEATLQRIYSLWEH
Human STKNHLQLEGLFFTNGEHTSKATLELSPWQMSALVQVHASQPSSFHDFPDLGQEVALNAN
House Mouse NMKNHLQVYSYFFTKGKQTCRATLELSPWTMSTLLQVHVSQLSSLLDLHHFDQEVILKAN
Chicken NGKNFARYTPLFTTTGSQKCKATFELAPWTVSADLQIQVTQPNSFLDTASVNQVVLMKVS
Rat NMINHLQVFSYFDTKGKQTCRATLELSPWTMSTLLQVHVSQPSPLFDLHHFDQEVILKAS
Cat NTKNHLQLEGFFLTSGEHTSKATLELSPWKMSALVQVRASQPNSLLDISYFGQDVSLNAN
White-tufted-ear marmoset STKNHLQLEGLFFTNGQHTSKATLQLSPWEMSALVQVQASQPNSFLDIPYLGQEVALNAN


Human TKNQKIRWKNEVRIHSGSFQSQVELSNDQEKAHLDIAGSLEGHLRFLKNIILPVYDKSLW
House Mouse TKNQKISWKGGVQVESRVLQHNAQFSNDQEEIRLDLAGSLDGQLWDLEAIFLPVYGKSLQ
Chicken PTDQKVGWKGEGQIQSLSLRHDMQLSNEKSNAKFDISGSLEGYMDFLKRINCAISKKSLW
Rat TKNQKVSWKSEVQVESQVLQHNAHFSNDQEEVRLDIAGSLEGQLWDLENFFLPAFGKSLR
Cat SEHQKVSWKSEVQVHSGFFQNNVHLSNDQEEARLDIASTLKVSLWFLKDIALPVYDKSLW
White-tufted-ear marmoset TKNQKIRWKSEARVHSGSLHSHVELSNDQEKAHLDIAGSLEGHLRFLKNIILPVYDKSLW
Human DFLKLDVTTSIGRRQHLRVSTAFVYTKNPNGYSFSIPVKVLADKF--IIPGLKLNDLNSV
House Mouse ELLQMD-----GKRQYLQASTSLLYTKNPNGYLLSLPVQELADRF--IIPGIKLND----
Chicken DILKLDVTTVADRKHYLNASASFIYRKSDDGYFFPMPVIRLSDGFTFSIPELHLKAPSPV
Rat ELLQID-----GKRQYLQASTSLHYTKNPNGYLLSLPVQELTDRF--IIPGLKLND----
Cat DLLKLDVTTSIDRKQYLHASTSLVYTKNPNGYHFSVPVQELADKF--IVPRLKLNGLSSG
White-tufted-ear marmoset DFLKLDVTTSMGRRQHLRLSTALVYTKNPNGYSFSIPVKVLADEF--IIPGLKLNDLNLV
Human LVMPTFHVPFTDLQVPSCKLDFREIQIYKKLRTSSFALNLPTLPEVKFPEVDVLTKYSQP
House Mouse ----------FSGVKIYKKLSTSPFALNLTMLPKVKFPGIDLLTQYSTP
Chicken LSTPEFRVPFSTLQVPAYTIDLRNIKIPQTLNIMPFDVNLPTLPKLRFPRVDVGANYITL
Rat ----------FSGIKIYKKLSTSPFALNLTMLPKVKFPGVDLLTQYSKP
Cat LVIPAFQVPFTDLQVPSYTFDFSEIKIYKKLSSLPFALSIPTLPKVKFPKVDVLTKYSEP
White-tufted-ear marmoset LVMPAFYVPFTDLQVPSYKLDFREIKMYKKLRTSSFALSLPTLPKVKFPEVDVLTKYSQP
T3799M (Cases)
Human EDSLIPFFEITVPESQLTVSQFTLPKSVSDGIAALDLNAVANKIADFELPTIIVPEQTIE
House Mouse EGSSVPIFEATIPEIHLTVSQFTLPKSLPVGNTVFDLNKLANMIADVDLPSVTLPEQTIV
Chicken EEYKIPYFEVTVPEYQITVSQFTLPKSISLGSFHVDLDEVANKIADFDLPTITIPEQKIE
Rat EGSSVPTFETTIPEIQLTVSQFTLPKSFPVGNTVFDLNKLTNLIADVDLPSITLPEQTIE
Cat EDASVPFFEITIPASQLTVSQFTLPKSISVGSAVEYLNEVASKIADFELPAITMPEQTIE
White-tufted-ear marmoset EDSLVPFFEITVPESQLTVSQFTLPKSVSVGGAVLDLNKVAKKIADFELPTITMPEQTIE
Human IPSIKFSVPAGIVIPSFQALTARFEVDSPVYNATWSASLKNKADYVETVLDSTCSSTVQF
House Mouse IPPLEFSVPAGIFIPFFGELTARAGMASPLYNVTWSAGWKTKADHVETFLDSMCTSTLQF
Chicken IPPLKVSLPAGIYIPSFGALTGSLKVASPLYNVTWRTDLTNKKESFEVSIDSTCSSTLQF
Rat IPSLEFSVPAGIFIPFFGELTAHVGMASPLYNVTWSTGWKNKADHVETFLDSTCSSTLQF
Cat IPSIKFSVPAGIFIPSFGALTAHFGVASPLYNATWSTGLKNKVEYVETFLNSTCSSTIQF
White-tufted-ear marmoset IPSVKFSVPAGVFIPSFGALTARFGVASPLYNATWSASLKNKADHLETALDSTCSSTIQF
Human LEYELNVLGTHKIEDGTLASKTKGTFAHRDFSAEYEEDGKYEGLQEWEGKAHLNIKSPAF
House Mouse LEYALKVVETHKIEEDLLTYNIKGTLQHCDFNVEYNEDGLFKGLWDWQGEAHLDITSPAL
Chicken LEYDLNVVSNYKYEEGKFVKKTVGSFAHRDLSANYIEDLALQGFGTVENTASLDVISPTF
Rat LEYALKVVGTHRIENDKFIYKIKGTLQHCDFNVKYNEDGIFEGLWDLEGEAHLDITSPAL
Cat LEYDLNVVGTHKIEDGVLFYRTKGTFAHHDISAEYKEDNKYRGLWDWKEDIRLDITSPTF
White-tufted-ear marmoset LEYDLNVVETLEIKNGSLTCKTKGTFAHRDFSAEYEEDGKYEGLKDWEGKARLDIKSPAF
Human TDLHLRYQKDKKGISTSAASPAVGTVGMDMDEDDDFSKWNFYYSPQSSPDKKLTIFKTEL
House Mouse TDFHLYYKEDKTSLSASAASSTIGTVGLDSSTDDQSVELNVYFHPQSPPEKKLSIFKTEW
Chicken ADVHVRYQMTDNMISSTVSSPSAGILGFLVQIETDTLKEKFYYRTLSAPQKDIDILKSEI
Rat TDFHLHYKEDKTSVSASAASPAIGTVSLDASTDDQSVRLNVYFRPQSPPDNKLSIFKMEW
Cat TDVHLHYQLNKNSLSSSVSSPAIGTVALDLEGNNNTLKWNLYYRPQSSLDKKLNIFKTEW
White-tufted-ear marmoset TDLHLHYQEDEKGLSISAASPAIGTMGMDIDEDDDFFKWNFYYSPQSSPDKKLTIFKTEL


Human RVRESDEETQIKVNWEEEAASGLLTSLKDNVPKATGVLYDYVNKYHWEHTGLTLREVSSK
House Mouse RYKESDGERYIKINWEEEAASRLLGSLKSNVPKASKAIYDYANKYHLEYVS-------SE
Chicken SFRNSD-IIQIKLNWKEDAAKDLLLGLKEKVPKMTSAVYKCVNRYYKEHMGLDISDATVI
Rat RDKESDGETYIKINWEEEAAFRLLDSLKSNVPKASEAVYDYVKKYHLGHAS-------SE
Cat RYQESDDKSQIKVSWEEEAVSELLSSLKDNVPKATGVLYDYANKYHQEYTGLNLRDASLK
White-tufted-ear marmoset RYQESDEELQIKVNWEEEAASGLLTSLKDNVPKATQALYDYVNKYHWEHTGLTLRGASSK
(V4101M) (Controls)
Human LRRNLQNNAEWVYQGAIRQIDDIDVRFQKAASGTTGTYQEWKDKAQNLYQELLTQEGQAS
House Mouse LRKSLQVNAE----HARRMVDEMNMSFQRVARD----------TYQNLYEEMLAQKSLSI
Chicken MKNILQNNADKAYMFAARQVDQMDVQLRTAANEASEKYQEMKVKARQLYKRAAEQAEQID
Rat LRKSLQNDAE----HAIRMVDEMNVNAQRVTRD----------TYQSLYKKMLAQESQSI
Cat LRRSLQNSAEWAYESTMRQTDEVDMGLQRVARATTRTYQQWKDKAQDLYHELLAQEGQAD
White-tufted-ear marmoset MRRNLRNKAEWVYQGAIRQIDDMDMWFQKAATGTTGNYQKWMDKAQNLYQELLTPEAQAS
Human FQGLKDNVFDGLVRVTQEFHMKVKHLIDSLIDFLNFPRFQFPGKPGIYTREELCTMFIRE
House Mouse PENLKKRVLDSIVHVTQKYHMAVMWLMDSFIHFLKFNRVQFPGYAGTYTVDELYTIVMKE
Chicken YQRIKARLLDATVDLMEEYHSKIRHLIDSVIEFLKTTKFQIPGLSEKYTGEELYLMTTEK
Rat PEKLKKMVLGSLVRITQKYHMAVTWLMDSVIHFLKFNRVQFPGNAGTYTVDELYTIAMRE
Cat FQGLQNKVFDRLMGVTQEYKDKVSHQVASLIDFLQFTRFQLPGKAGTYSRDELFTMVMRE
White-tufted-ear marmoset LQGFKDKVFDSLIRVTQEFRMKIKHLIDSLIDFLTFPRFQFPGKPGTYTRDELCTMVIRE
Human VGTVLSQVYSKVHNGSEILFSYFQDLV----------ITLPFELRKHKLIDVISMYRELL
House Mouse TKKSLSQLF----NGLGNLLSYVQNQVEKSRLINDITFKCPFFSKPCKLKDLILIFREEL
Chicken AAKTADICLSKLQ-------EYFDALI---AAISELEVRVPASETILRGRNVLDQIKEML
Rat TKKLLSQLF----NGLGHLFSYVQDQVEKSRVINDITFKCPFSPTPCKLKDVLLIFREDL
Cat VGQVLSQVYSKIHSGLEILFSYFQDLMEKSELIKDVEIKFPFNSESYKLRDVVLEFGNLL
White-tufted-ear marmoset VGMVLSQVYAKVHNGSEILFSYFQDLV----------IKLPFELRSQKLIDIILRCRELL
M4293V (Cases)
Human KDLSKEAQEVFKAIQSLKTTEVLRNLQDLLQFIFQLIEDNIKQLKEMKFTYLINYIQDEI
House Mouse NILSNIGQ------QDIKFTTILSSLQGFLERVLDIIEEQIKCLKDNESTC----VADHI
Chicken KHLQEKIRQTFVTLQEADFAGKLNQLKQVVQKTFQKAGNMVRSLQSKNFEDIKVQMQQLY
Rat NILSNLGQ------QDINFTTILSDFQSFLERLLDIIEEKIECLKNNESTC----VPDHI
Cat KSLSQNIQDALNNLQSIKTTEMLSHLQRYLERAFQEIEEEIQRLKGKKFTYLINDIQHVI
White-tufted-ear marmoset NDLSEEAQTVFKNMQSLKTTEVLHDLQQLLQFIFQI-EEDIKRLKEIKLTQLTNHIRDDI
Human NTIFSDYIPYVFKLLKENLCLNLHKFNEFIQNELQEASQELQQIHQYIMALREEYFDPSI
House Mouse NMVFKIQVPYAFKSLREDIYFVLGEFNDFLQSILQEGSYKLQQVHQYMKALREEYFDPSM
Chicken KDAMASDYAHKLRSLAENVKKYISQMKNFSQKTLQKVSENLQQLVLYIKALREEYFDPTT
Rat NMFFKTHIPFAFKSLRENIYSVFSEFNDFVQSILQEGSYKLQQVHQYMKAFREEYFDPSV
Cat NRIFKKYIPTAFTFLK----INFDKFNELVQNQLQEASQQLQQLHQLIKDLHKEYVDPTV
White-tufted-ear marmoset NTIFNDYIPFFFKILKEILCHNLHTFNEFIQNNLQEASQELQQIHQYIMALREEYFDPSV
(S4403T) (Controls)
Human VGWIVKYYELEEKIVSLIKNLLVALKDFHSEYIVSASNFTSQLSSQVEQFLHRNIQEYLS
House Mouse VGWIVKYYEIEENMVELIKTLLVSFRDVYSEYSVTAADFASKMSTQVEQFVSRDIREYLS
Chicken LGWSVKYYEVEDKVLGLLKNLMDTLVIWYNEYAKDLSDLVTRLTDQVRELVENYRQEYYD
Rat VGWTVKYYEIEEKMVDLIKTLLAPLRDFYSEYSVTAADFASKMSTQVEQFVSRDIREYLS
Cat VSWTSRYYELEEKIISLIKNLVDVLMDFHSKYTISTAGLISQLSNQVEQFVQKDIQEYLS
White-tufted-ear marmoset VGWTVKYYELEEKIVSLIKNLLVALKDFHSEYTASASGFASQLSTQVEQFLHGNIQEYLS

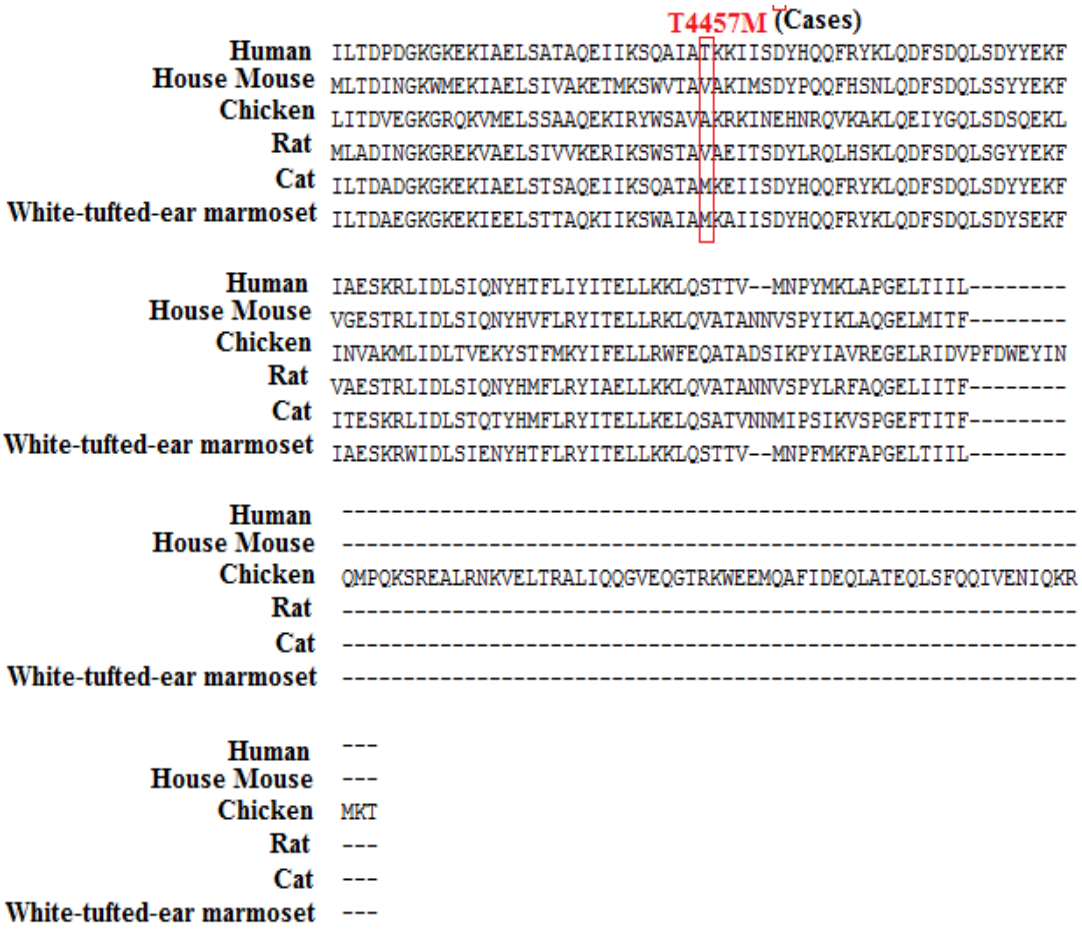
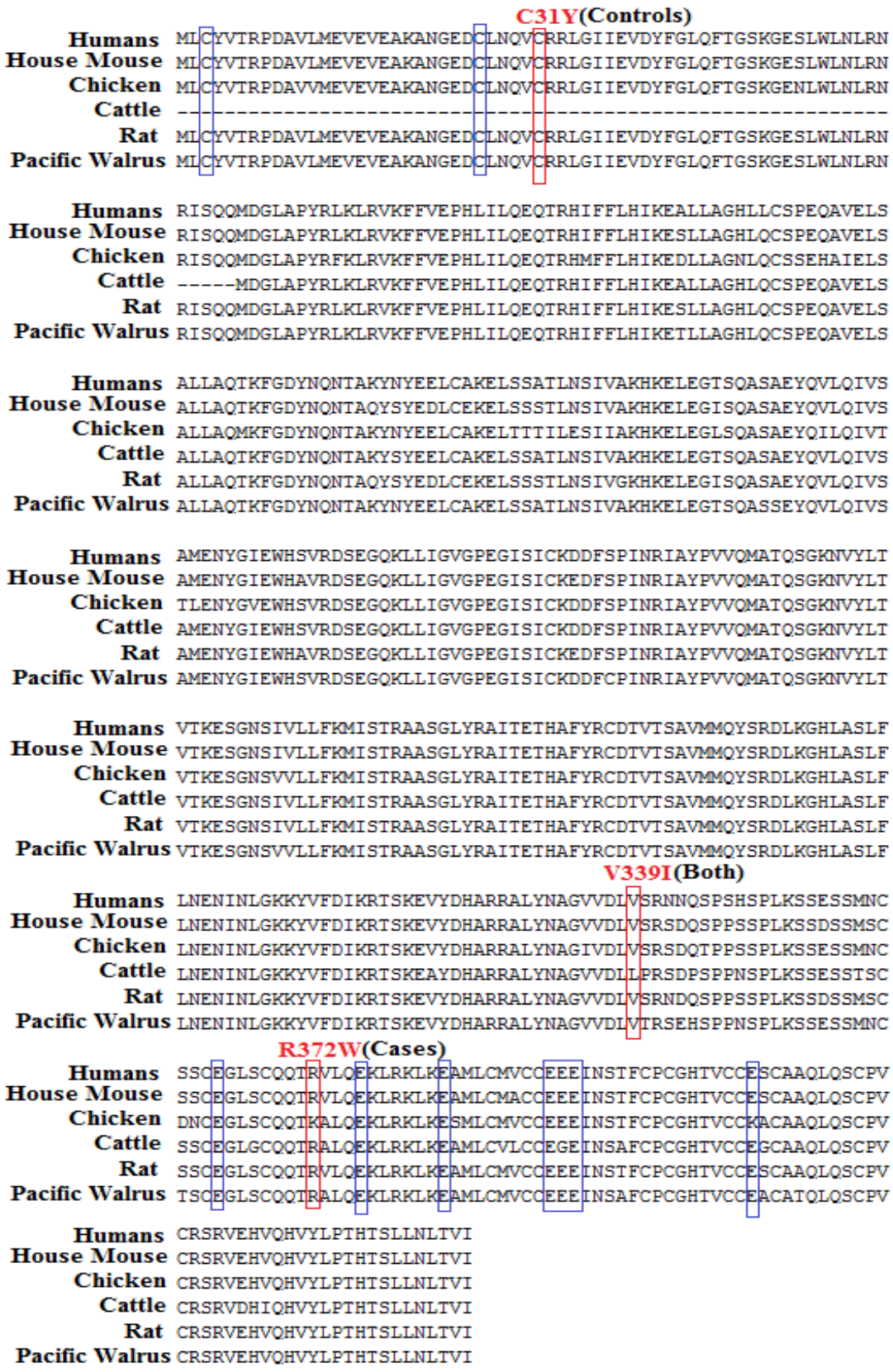

T4457M (Cases)
Human ILTDPDGKGKEKIAELSATAQEIIKSQAIATKKIISDYHQQFRYKLQDFSDQLSDYYEKF
House Mouse MLTDINGKWMEKIAELSIVAKETMKSWVTAVAKIMSDYPQQFHSNLQDFSDQLSSYYEKF
Chicken LITDVEGKGRQKVMELSSAAQEKIRYWSAVAKRKINEHNRQVKAKLQEIYGQLSDSQEKL
Rat MLADINGKGREKVAELSIVVKERIKSWSTAVAEITSDYLRQLHSKLQDFSDQLSGYYEKF
Cat ILTDADGKGKEKIAELSTSAQEIIKSQATAMKEIISDYHQQFRYKLQDFSDQLSDYYEKF
White-tufted-ear marmoset ILTDAEGKGKEKIEELSTTAQKIIKSWAIAMKAIISDYHQQFRYKLQDFSDQLSDYSEKF
Human IAESKRLIDLSIQNYHTFLIYITELLKKLQSTTV--MNPYMKLAPGELTIIL--------
House Mouse VGESTRLIDLSIQNYHVFLRYITELLRKLQVATANNVSPYIKLAQGELMITF--------
Chicken INVAKMLIDLTVEKYSTFMKYIFELLRWFEQATADSIKPYIAVREGELRIDVPFDWEYIN
Rat VAESTRLIDLSIQNYHMFLRYIAELLKKLQVATANNVSPYLRFAQGELIITF--------
Cat ITESKRLIDLSTQTYHMFLRYITELLKELQSATVNNMIPSIKVSPGEFTITF--------
White-tufted-ear marmoset IAESKRWIDLSIENYHTFLRYITELLKKLQSTTV--MNPFMKFAPGELTIIL--------
Human ----------
House Mouse ----------
Chicken QMPQKSREALRNKVELTRALIQQGVEQGTRKWEEMQAFIDEQLATEQLSFQQIVENIQKR
Rat ----------
Cat ----------
White-tufted-ear marmoset ----------
Human ---
House Mouse ---
Chicken MKT
Rat ---
Cat ---
White-tufted-ear marmoset ---

**Figure 3.3. Multiple Sequence Alignment of the *IDOL* protein. Multiple Sequence Alignment of *IDOL* amino acid sequence from 6 species, namely: Human, House Mouse, Chicken, Rat, Cattle, Rat and Pacific Walrus.**


C31Y(Controls)
Humans MLCYVTRPDAVLMEVEVEAKANGEDCLNQVCRRLGIIEVDYFGLQFTGSKGESLWLNLRN
House Mouse MLCYVTRPDAVLMEVEVEAKANGEDCLNQVCRRLGIIEVDYFGLQFTGSKGESLWLNLRN
Chicken MLCYVTRPDAVVMEVEVEAKANGEDCLNQVCRRLGIIEVDYFGLQFTGSKGENLWLNLRN
Cattle ----------
Rat MLCYVTRPDAVLMEVEVEAKANGEDCLNQVCRRLGIIEVDYFGLQFTGSKGESLWLNLRN
Pacific Walrus MLCYVTRPDAVLMEVEVEAKANGEDCLNQVCRRLGIIEVDYFGLQFTGSKGESLWLNLRN
Humans RISQQMDGLAPYRLKLRVKFFVEPHLILQEQTRHIFFLHIKEALLAGHLLCSPEQAVELS
House Mouse RISQQMDGLAPYRLKLRVKFFVEPHLILQEQTRHIFFLHIKESLLAGHLQCSPEQAVELS
Chicken RISQQMDGLAPYRFKLRVKFFVEPHLILQEQTRHMFFLHIKEDLLAGNLQCSSEHAIELS
Cattle -----MDGLAPYRLKLRVKFFVEPHLILQEQTRHIFFLHIKEALLAGHLQCSPEQAVELS
Rat RISQQMDGLAPYRLKLRVKFFVEPHLILQEQTRHIFFLHIKESLLAGHLQCSPEQAVELS
Pacific Walrus RISQQMDGLAPYRLKLRVKFFVEPHLILQEQTRHIFFLHIKETLLAGHLQCSPEQAVELS
Humans ALLAQTKFGDYNQNTAKYNYEELCAKELSSATLNSIVAKHKELEGTSQASAEYQVLQIVS
House Mouse ALLAQTKFGDYNQNTAQYSYEDLCEKELSSSTLNSIVAKHKELEGISQASAEYQVLQIVS
Chicken ALLAQMKFGDYNQNTAKYNYEELCAKELTTTILESIIAKHKELEGLSQASAEYQILQIVT
Cattle ALLAQTKFGDYNQNTAKYSYEELCAKELSSATLNSIVAKHKELEGTSQASAEYQVLQIVS
Rat ALLAQTKFGDYNQNTAQYSYEDLCEKELSSSTLNSIVGKHKELEGISQASAEYQVLQIVS
Pacific Walrus ALLAQTKFGDYNQNTAKYNYEELCAKELSSATLNSIVAKHKELEGTSQASSEYQVLQIVS
Humans AMENYGIEWHSVRDSEGQKLLIGVGPEGISICKDDFSPINRIAYPVVQMATQSGKNVYLT
House Mouse AMENYGIEWHAVRDSEGQKLLIGVGPEGISICKEDFSPINRIAYPVVQMATQSGKNVYLT
Chicken TLENYGVEWHSVRDSEGQKLLIGVGPEGISICKDDFSPINRIAYPVVQMATQSGKNVYLT
Cattle AMENYGIEWHSVRDSEGQKLLIGVGPEGISICKDDFSPINRIAYPVVQMATQSGKNVYLT
Rat AMENYGIEWHAVRDSEGQKLLIGVGPEGISICKEDFSPINRIAYPVVQMATQSGKNVYLT
Pacific Walrus AMENYGIEWHSVRDSEGQKLLIGVGPEGISICKDDFCPINRIAYPVVQMATQSGKNVYLT
Humans VTKESGNSIVLLFKMISTRAASGLYRAITETHAFYRCDTVTSAVMMQYSRDLKGHLASLF
House Mouse VTKESGNSIVLLFKMISTRAASGLYRAITETHAFYRCDTVTSAVMMQYSRDLKGHLASLF
Chicken VTKESGNSVVLLFKMISTRAASGLYRAITETHAFYRCDTVTSAVMMQYSRDLKGHLASLF
Cattle VTKESGNSIVLLFKMISTRAASGLYRAITETHAFYRCDTVTSAVMMQYSRDLKGHLASLF
Rat VTKESGNSIVLLFKMISTRAASGLYRAITETHAFYRCDTVTSAVMMQYSRDLKGHLASLF
Pacific Walrus VTKESGNSVVLLFKMISTRAASGLYRAITETHAFYRCDTVTSAVMMQYSRDLKGHLASLF
V339I(Both)
Humans LNENINLGKKYVFDIKRTSKEVYDHARRALYNAGVVDLVSRNNQSPSHSPLKSSESSMNC
House Mouse LNENINLGKKYVFDIKRTSKEVYDHARRALYNAGVVDLVSRSDQSPPSSPLKSSDSSMSC
Chicken LNENINLGKKYVFDIKRTSKEVYDHARRALYNAGIVDLVSRSDQTPPSSPLKSSESSMNC
Cattle LNENINLGKKYVFDIKRTSKEAYDHARRALYNAGVVDLLPRSDPSPPNSPLKSSESSTSC
Rat LNENINLGKKYVFDIKRTSKEVYDHARRALYNAGVVDLVSRNDQSPPSSPLKSSDSSMSC
Pacific Walrus LNENINLGKKYVFDIKRTSKEVYDHARRALYNAGVVDLVTRSEHSPPNSPLKSSESSMNC
R372W(Cases)
Humans SSCEGLSCQQTRVLQEKLRKLKEAMLCMVCCEEEINSTFCPCGHTVCCESCAAQLQSCPV
House Mouse SSCEGLSCQQTRVLQEKLRKLKEAMLCMACCEEEINSTFCPCGHTVCCESCAAQLQSCPV
Chicken DNCEGLSCQQTKALQEKLRKLKESMLCMVCCEEEINSTFCPCGHTVCCKACAAQLQSCPV
Cattle SSCEGLGCQQTRALQEKLRKLKEAMLCVLCCEGEINSAFCPCGHTVCCEGCAAQLQSCPV
Rat SSCEGLSCQQTRVLQEKLRKLKEAMLCMVCCEEEINSTFCPCGHTVCCESCAAQLQSCPV
Pacific Walrus TSCEGLSCQQTRALQEKLRKLKEAMLCMVCCEEEINSAFCPCGHTVCCEACATQLQSCPV
Humans CRSRVEHVQHVYLPTHTSLLNLTVI
House Mouse CRSRVEHVQHVYLPTHTSLLNLTVI
Chicken CRSRVEHVQHVYLPTHTSLLNLTVI
Cattle CRSRVDHIQHVYLPTHTSLLNLTVI
Rat CRSRVEHVQHVYLPTHTSLLNLTVI
Pacific Walrus CRSRVEHVQHVYLPTHTSLLNLTVI

# Chapter 4: Results II — Resolving genetic etiology of hypercholesterolemia in Mutation Negative Familial Hypercholesterolemia patients

## 4.1 Study subjects

Individuals with Familial Hypercholesterolemia have abnormally elevated cholesterol levels. I performed a case control study in which 44 Familial Hypercholesterolemia Mutation Negative (FH/M-ve) patients were cases and 44 Familial Hypercholesterolemia Mutation Positive (FH/M+ve) patients were controls. The cases and controls were matched for age and sex. I hypothesized that FH/M-ve patients have a significantly greater accumulation of LDL-C raising SNPs, as assessed by the genetic risk score, than FH/M+ve patients. This hypothesis was first tested in a British study (*48*). However, my study was the first to test this hypothesis in a Canadian population and also served as an independent replication study.

I also performed an independent study where all the coding regions of *APOE* were in 95 *APOE* FH/M-ve patients. For the FH/M-ve resequencing study, all 44 of the FH/M-ve patients from the FH case control study were included in the FH/M-ve *APOE* resequencing study. *APOE* was sequenced to identify any potential FH-causing mutations as a recent collaborative study has showed an *APOE* variant to FH-causing (data not yet published).

### 4.1.1 Patient demographics of Mutation negative Familial Hypercholesterolemia patients and Mutation positive Familial Hypercholesterolemia patients

In **Tables 4.1** and **4.2**, the clinical attributes of patients from the FH case-control study and *APOE* re-sequencing study are given, respectively. Paired t-tests were performed to compare clinical characteristics between cases and controls; paired t-test was used because the cases and controls were matched. Paired t-tests were conducted using SAS Enterprise guide version 4.3 software, with a nominal level of significance of $P < 0.05$.

## 4.2 LDL-C genetic risk score

12 of the 37 LDL-C raising SNPs identified by GLGC were genotyped in 44 FH/M-ve cases and 44 FH/M+ve controls. For each FH patient, the LDL-C genetic risk score was calculated as follows: The LDL score was the sum of the product of the risk allele and beta coefficient (or standardized regression coefficient). The beta coefficient essentially represents the standardized regression coefficient for the risk allele. The beta coefficient is representative of the effect size (i.e. estimate of how much the risk allele is raising LDL-C). **Table 4.3** shows the beta coefficients from GLGC (*43*) recalculated to mmol/L and the beta coefficients for *APOE* (*100*). **Table 4.3** shows the standardized regression coefficient for all the risk alleles used to calculate the LDL-C genetic risk score (*48*).

For example if an FH patient had two copies of the rs2479409 risk allele (*PCSK9*), one copy of the rs629301 risk allele (*CELSR2*) and ε4ε4, the LDL-C genetic risk score would be calculated as follows:

$(0.052*2)+(0.15*1)+ 0.2= 0.45$

### 4.2.1 List of LDL-C genetic risk score in Mutation Negative Familial Hypercholesterolemia patients and Mutation Positive Familial Hypercholesterolemia patients

The GLGC weighted LDL-C genetic risk scores for the 44 FH cases and their matched controls are all given in **Table 4.4**. The information in **Table 4.4** was obtained from our UK collaborators. All the 88 FH patients were genotyped for the 12 SNPs and the LDL-C genetic risk score was calculated for each patient by our UK collaborators (Professor Steve Humphries, British Heart Foundation Laboratories, Institute for Cardiovascular Science, University College London, United Kingdom).

**Table 4.2** gives partial information on patient demographics for *APOE* resequencing. All the FH/M-ve patients were collected over the years in the Hegele lab; some of the patients were from other sources. So, information clinical information on all 95 patients could not be obtained, however, I found clinical information on 59 of the 95 patients (62% of patients) (**Table 4.2**)

## 4.3 Test for accumulation of LDL-C SNPs

Normality tests showed that the LDL-C genetic risk scores for FH/M-ve cases and FH/M+ve controls were not normally distributed; they were skewed (normality test results are shown in Table 5 in the Appendix). Because cases and controls were paired, the LDL-C genetic risk scores were also paired. Therefore, the Wilcoxon Signed-Rank test was used to statistically determine whether the mean LDL- C genetic risk score was higher in FH/M-ve cases than in FH/M+ve controls, because it is the most appropriate

test to use to compare the mean of paired data that are not normally distributed. **Table 4.5** shows results from Wilcoxon's signed rank test. Even though the mean LDL-C genetic risk score was higher in FH/M-ve cases than in FH/M+ve controls, this difference in mean LDL-C genetic risk score was not statistically significant ($P$=0.43) (**Table 4.5**). **Table 4.6** shows all the descriptive statistics for the mean LDL-C genetic risk scores in FH/M-ve cases and FH/M+ve controls

**4.4 Comparison of LDL-C genetic risk scores to other ethnically different cohorts**

The LDL-C genetic risk scores from my Canadian case control cohort was then compared with two other different populations, namely, British and Belgian subpopulations (*48*). Information on the British and Belgium cohort is from Talmud et al (*48*). **Table 4.7** shows the mean LDL-C genetic risk scores, along with the p-values for the mean differences between cases and controls in the three different populations. **Table 4.7** also shows the sample size of the case-control cohort in the three populations. In all three cohorts, the mean LDL-C genetic risk score was greater in FH/M-ve cases than in FH/M+ve controls.

The mean LDL-C genetic risk scores of the cases and controls of the Canadian cohort are unequivocally comparable to the mean LDL-C genetic risk scores of cases and controls of the UK and Belgium cohort. Unlike the Canadian cohort, the sample size of the UK and Belgian cohorts is considerably larger (**Table 4.7**) and the differences in mean LDL-C genetic risk scores between FH/M-ve cases and FH/M+ve controls were significant in both the UK and Belgian cohort (**Table 4.7**). Post-hoc power calculations showed that the British and Belgium cohorts were sufficiently powered and the Canadian

cohort was underpowered. **Figure 4.1** shows how statistical power is a function of sample size.

### 4.5 *APOE* resequencing in FH/M-ve patients

Another recent collaborative study involving the Hegele lab and the Genest lab at McGill University identified an amino acid deletion mutation in *APOE* that segregated with the hypercholesterolemia phenotype in a family with Familial Hypercholesterolemia (FH). This mutation was designated as *APOE* Leu167del and was considered to be FH causing as it co-segregated with the FH phenotype in *LDLR* mutation negative individuals across multiple generations in the affected family. Because this example showed that mutations in *APOE* could cause an FH-like phenotype, the three coding regions of *APOE* were sequenced in 95 Canadian FH/M-ve patients to identify any possible FH-causing mutations. The APOE Leu 167 del was not found in this Canadian FH/M-ve cohort. **Table 4.8** shows all the variants that were found by Sanger sequencing. Intronic, silent and missense variants were found. I performed *in silico* prediction analysis of the rare missense variants found in the Canadian FH/M-ve cohort: these were the *APOE* : p.L46P and p.A91 T variants. These were absent from other databases. Only the p.L46P variant was predicted to be deleterious by PolyPhen-2 (**Table 4.9**).

Therefore, in Canadian FH patients who are negative for mutations in genes such as LDLR, APOB and PCSK9, there appears to be other genetic factors involved, including: 1) a trend towards a higher polygenic LDL-C genetic risk score; and 2) evidence for rare variants in APOE. The relatively small sample size here was consistent

with a pilot project to test these hypotheses; larger sample sizes will be needed to determine whether these findings can attain statistical significance.

**Table 4.1** Baseline characteristics (mean±SEM or percentage) of FH patients with and without a mutation in the FH case-control project

| Population | FH M+ | FH M- | p-value |
|---|---|---|---|
| **n** | 44 | 44 | N/A |
| **Female,%** | 56.8 | 56.8 | N/A |
| **Age, years** | 47.5±1.8 | 48.3±1.8 | 0.03 |
| **BMI, kg/m$^2$** | 29.1±1.0 | 28.8±1.0 | 0.67 |
| **TC, mmol/L** | 9.35±0.37 | 7.40±0.32 | <0.0001 |
| **LDL-C, mmol/L** | 7.36±0.3 | 5.24±0.3 | <0.0001 |
| **HDL-C, mmol/L** | 1.25±0.05 | 1.24±0.05 | 0.93 |
| **TG, mmol/L** | 1.73±0.15 | 2.05±0.14 | 0.1 |
| **Tendon xanthoma present, %** | 20.5 | 2.27 | 0.0097 |

Note: BMI, Body Mass Index; TC, Total Cholesterol; LDL-C, Low Density Lipoprotein Cholesterol; HDL-C; High Density Lipoprotein Cholesterol; TG, Triglyceride. Mean values of clinical attributes are represented as mean ± standard deviation. LDL-C level is not accurately calculated using the Friedewald equation for HTG patients when plasma TG concentration exceeds 4.5 mmol/L

**Table 4.2** Baseline characteristics (mean±SEM and percentages) of 59 of the 95 FH patients without a mutation for *APOE* resequencing project.

| Characteristic | value |
|---|---|
| **n** | 59 |
| **Female,%** | 50.8 |
| **Age, years** | 44.2 ±2.01 |
| **BMI, kg/m$^2$** | 27.3±0.96 |
| **TC, mmol/L** | 7.70±0.52 |
| **LDL-C, mmol/L** | 5.04±0.3 |
| **HDL-C, mmol/L** | 1.27±0.26 |
| **TG, mmol/L** | 1.93±0.12 |

Note: BMI, Body Mass Index; TC, Total Cholesterol; LDL-C, Low Density Lipoprotein Cholesterol; HDL-C; High Density Lipoprotein Cholesterol; TG, Triglyceride. Mean values of clinical attributes are represented as mean ± standard deviation. LDL-C level is not accurately calculated using the Friedewald equation for HTG patients when plasma TG concentration exceeds 4.5 mmol/L

**Table 4.3.** The standardized regression coefficient of the risk alleles for various SNPs (*43*).

| CHR | SNP | Gene | Risk allele | Standardized regression coefficient |
|---|---|---|---|---|
| 1 | rs2479409 | *PCSK9* | G | 0.052 |
| 1 | rs629301 | *CELSR2* | T | 0.15 |
| 2 | rs1367117 | *APOB* | A | 0.1 |
| 2 | rs4299376 | *ABCG8* | G | 0.071 |
| 6 | rs1564348 | *SLC22A1* | T | 0.014 |
| 6 | rs1800562 | *HFE* | G | 0.057 |
| 6 | rs3757354 | *IDOL* | C | 0.037 |
| 11 | rs11220462 | *ST3GAL4* | A | 0.05 |
| 14 | rs8017377 | *KIAA1305* | A | 0.029 |
| 19 | rs6511720 | *LDL-R* | G | 0.18 |
| 19 | rs429358 | *APOEψ* | | |
| 19 | rs7412 | *APOEψ* | | |
| 19 | | *APOE* | ε2ε2 | -0.9 |
| 19 | | *APOE* | ε2ε3 | -0.4 |
| 19 | | *APOE* | ε2ε4 | 0.2 |
| 19 | | *APOE* | ε3ε3 | 0 |
| 19 | | *APOE* | ε3ε4 | 0.1 |
| 19 | | *APOE* | ε4ε4 | 0.2 |

*Ψ APOE* weights (*100*)

The standardized regression coefficients were representative of effect size of risk alleles. The LDL-C genetic risk score was calculated for each FH patient by calculating the sum of the products of standardized regression coefficients and the count of risk alleles.

**Table 4.4 GLGC Weighted LDL-C genetic risk scores for all the FH patients in the FH case-control study**

| Sample.ID of 44 FH/M- patients | GLGC Weighted LDL-C genetic risk score | Sample.ID of 44 FH/M+ patients | GLGC Weighted LDL-C genetic risk score |
|---|---|---|---|
| 726 | -0.045590899 | 90 | -0.00188777 |
| 9035 | 0.307008016 | 573 | 0.420015515 |
| 4242 | 0.385621928 | 1123 | 0.5805534 |
| 366 | 0.640031031 | 2457 | 0.603568657 |
| 9075 | 0.72723041 | 4177 | 0.648461339 |
| 250 | 0.773467804 | 796 | 0.655340056 |
| 8784 | 0.783811739 | 2062 | 0.674993534 |
| 1746 | 0.814067752 | 9979 | 0.730799069 |
| 367 | 0.871476596 | 1999 | 0.747349366 |
| 3895 | 0.885880527 | 3420 | 0.840186191 |
| 8555 | 0.892086888 | 3606 | 0.846909748 |
| 8836 | 0.892940263 | 1905 | 0.865451254 |
| 9884 | 0.923454874 | 175 | 0.873028187 |
| 8983 | 0.926299456 | 787 | 0.873648823 |
| 3477 | 0.928032066 | 8832 | 0.877424358 |
| 6455 | 0.972071372 | 55 | 0.877941556 |
| 5965 | 0.972329971 | 846 | 0.907680372 |
| 683 | 1.006206361 | 8783 | 0.910886992 |
| 8728 | 1.021903283 | 877 | 0.9256271 |
| 8831 | 1.02560124 | 8712 | 0.929919834 |
| 8534 | 1.039824152 | 2453 | 0.933540212 |
| 3939 | 1.046547711 | 363 | 0.952935091 |
| 4789 | 1.057667442 | 393 | 0.966382207 |
| 1095 | 1.076027928 | 5614 | 0.971476596 |
| 1339 | 1.093612619 | 1844 | 0.976208947 |
| 169 | 1.098448409 | 3635 | 0.979570725 |
| 3646 | 1.107240755 | 5791 | 0.980087923 |
| 9885 | 1.124644426 | 11 | 0.987070079 |
| 8797 | 1.127747607 | 6310 | 1.024566847 |
| 1460 | 1.133436772 | 124 | 1.034134988 |
| 8954 | 1.134652185 | 825 | 1.038712179 |
| 1813 | 1.137574346 | 8557 | 1.065606413 |

| | | | |
|---|---|---|---|
| 8704 | 1.138091543 | 8967 | 1.072071372 |
| 3457 | 1.147323506 | 140 | 1.08215671 |
| 1805 | 1.150426687 | 154 | 1.089475044 |
| 8625 | 1.166976983 | 5665 | 1.105766744 |
| 8852 | 1.228782001 | 592 | 1.127928626 |
| 5732 | 1.231186966 | 8504 | 1.179131108 |
| 521 | 1.235505559 | 554 | 1.210343935 |
| 6302 | 1.242229117 | 495 | 1.226014998 |
| 8691 | 1.292655804 | 4625 | 1.227825187 |
| 8982 | NA | 810 | 1.28541505 |
| 1761 | NA | 650 | NA |
| 628 | NA | 560 | NA |

**Table 4.5. Mean (±SD) for LDL-C genetic risk score for Canadian case-control cohort**

| Statistic | FH/M-ve | FH/M+ve | *P*-value |
|---|---|---|---|
| Mean | 0.96±0.27 | 0.92±0.24 | 0.43 |

**Table 4.6. Descriptive statistics for LDL-C genetic risk score for the Canadian case-control cohort**

| Statistic | FH/M-ve | FH/M+ve |
|---|---|---|
| Mean | 0.96 | 0.92 |
| Standard Deviation | 0.3 | 0.2 |
| Median | 1.03 | 0.95 |
| Standard Error | 0.044 | 0.038 |

**Table 4.7. Mean weighted LDL-C genetic risk score (±SD) for UK, Belgium and Canadian Case-Control Cohort**

| Cohort | FH/M- | FH/M+ | n | *P*-value |
|---|---|---|---|---|
| UK | 1.0±0.2 | 0.95±0.2 | 640 | 0.0014 |
| Belgium | 0.99±0.19 | 0.92± 0.2 | 736 | 4.0xl0-6 |
| Canada | 0.96±0.27 | 0.92±0.23 | 88 | 0.45 |

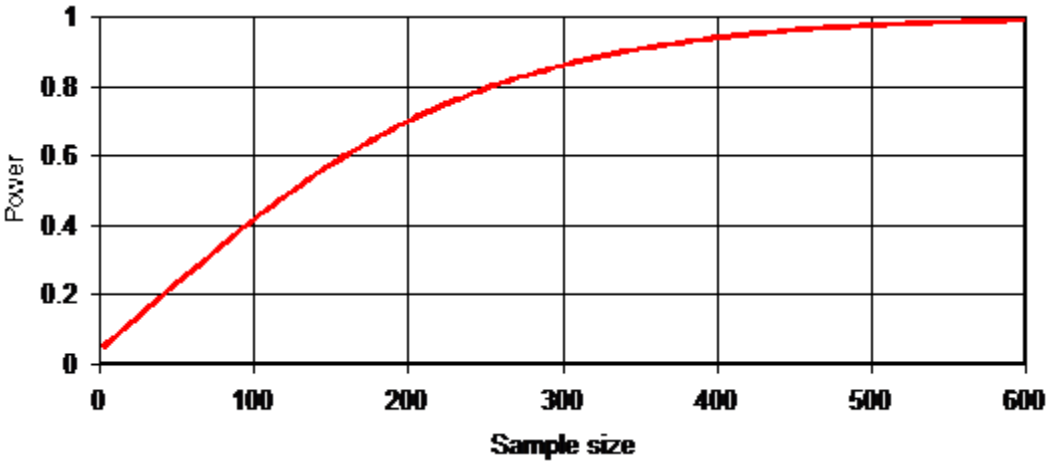



**Figure 4.1** Representation of level power as a function of sample size. As sample size increases, the power to detect statistical significance increases. Post hoc power calculations showed that British cohort sufficiently powered, while Canadian cohort was underpowered.

**Table 4.8. List of all variants found in 95 Mutation Negative Familial Hypercholesterolemia patients that were sequenced for *APOE***

| Variant | SNP identifier | MAF(%) | Minor allele carriers |
|---|---|---|---|
| c.43 +64C>T | rs.143063029 | 0.53 | 1 |
| p.S40S | - | 0.53 | 1 |
| p.L46P | rs769452 | 0.53 | 1 |
| p.A91T | - | 0.53 | 1 |
| p.A217A | rs.72654468 | 0.53 | 1 |
| c.43+78G>A | rs.769449 | 13.7 | 25 |
| p.C130R | rs.429358 | 16.8 | 32 |
| p.R176C | rs.7412 | 3.16 | 5 |

**Table 4.9 Rare missense variants found in *APOE* in 95 FH/M-ve patients that were resequenced**

| Variant | SIFT prediction | SIFT score | PolyPhen-2 prediction | Polyphen-2 score |
|---|---|---|---|---|
| p.L46P | Tolerated | 0.11 | Possibly damaging | 0.949 |
| p.A91T | Tolerated | 0.36 | Benign | 0.092 |

# Chapter 5: Results III—Association of *DIET1* SNPs with Hypertriglyceridemia

## 5.1 Study subjects

Study subjects were individuals clinically diagnosed with one of the 4 Fredrickson polygenic hypertriglyceridemia (HTG) phenotypes, namely, Type 2B (MIM 144250), Type 3 (107741), Type 4 (144600) or Type 5 (144650). The majority of control subjects were healthy normolipidemic individuals (96%) and the remaining controls were FH patients without HTG (4%). FH patients were included as controls to partially correct for the increased total cholesterol phenotype that is seen in some HTG patients. **Table 5.1** shows the clinical attributes of the case-control cohort for the *DIET1* analyses. There was a total of 463 HTG cases.

Out of all these 463 hypertriglyceridemia patients, all Familial Combined Hyperlipidemia (FCH) (Frederickson Type 2B) patients (n=159) and all Familial Hypertriglyceridemia (FHTG) (Fredrickson Type 4) patients (n=128) were used for the independent case-control study that tested for association of the *PSMD9* gene region with hypercholesterolemia. Testing this association aimed at identifying any common variants associated with the cholesterol component of FCH because of a recent publication in which the *PSMD9* locus had been associated with hypercholesterolemia (*65*)

## 5.2 Genotyping Results

HTG cases and controls had been genotyped for SNPs across the genome using Affymetrix version 6.0 microarrays, as stated in section 2.5 of the Materials and Methods

section. The specific subset of these genotypes within the *DIET1* region was used for the association analysis with HTG. The *DIET1* region in mice corresponds to the genomic coordinates chr 10:19,377,700 to 20,063,500 in the hg19 human genome build. This corresponds to chr 10: 19,417,706 to 20,103,506 genomic coordinates in the hg18 human genome build. The hg18 version of genome build in the human genome was used throughout *DIET1* analyses because the GWAS data *(70)* were based on hg18. I hypothesized that the *DIET1* region is associated with triglyceride levels in humans, because the mouse *DIET1* gene was associated with triglyceride levels in mice. If correct, the frequency of SNPs in *DIET1* region should differ between HTG cases and normotriglyceridemic controls.

The *DIET1* region corresponds to the 19,417,706 to 20,103,506 hg18 human genome coordinates. Since I was only interested in the *DIET1* region, statistical analyses was only performed on SNPs in the *DIET1* region from the GWAS data. Logistic regression was performed to test for association of SNPS in *DIET1* region and triglyceride (TG) level; logistic regression essentially tests whether any SNP allele or genotype has a different frequency in cases relative to controls.

Because of linkage disequilibrium, SNPs outside the 19,417,706 to 20,103,506 *DIET1* regions were selected; SNPs from hg18 19,317,706 to 24,003,506 human genomic coordinates on chromosome 10 were selected for logistic regression. In total, 4,808 SNPs were present within this expanded range. Logistic regression was adjusted for the following covariates: sex, body mass index (BMI), diabetes status and population substructure.

**5.3 List of most highly associated SNPs in *DIET1* locus with HTG**

Since there were 4,808 SNPs genotyped in the 19,317,706 to 24,003,506 hg18 *DIET1* region, the Bonferroni corrected p-value, below which an association is considered statistically significant, is (0.05/4,808) 0.00001. Bonferroni correction is an over-conservative form of adjusting p-values for multiple testing, where each association analysis of each SNP is essentially a 'test'. **Table 5.2** shows the top 5 five SNPs most highly associated with TG levels, in the *DIET1* region.

The most highly associated SNP with TG (i.e. smallest p-value) was rs2499065 (*P*=0.0008). Since the *DIET1* region in mice has currently not been annotated in the human genome, not much information on the rs2499065 variant presently exists; so it is not known whether the variant is intronic, silent, nonsense or amino acid changing.

Even though the p-value of the rs2499065 variant does not reach the overly-conservative p-value of 0.00001, it is still worth reporting the variant (as well the remaining top 4) because a substantial proportion of the 4,808 SNPs are in Linkage Disequilibrium (LD). So, in reality the SNPs that are actually associated due to some biological relationship with TG (and not due to LD) are likely much fewer than 4,808, although the precise number cannot be estimated. So this only makes the Bonferroni corrected p-value over-conservative as there are in actuality fewer 'tests'.

For the adjusted logistic regression, the rs2499065 variant had an OR of 1.4 with a 95% confidence interval that did not include 1.0 (1.1-1.6) (**Table 5.2**). The remaining SNPs, from the adjusted logistic regression, also had ORs above 1.0 with 95% Confidence Interval (CI) that did not include 1.0 (**Table 5.2**). **Figure 5.1** summarizes the

findings from the *DIET1* analyses. The data suggest that there might indeed be an association between SNPs in the *DIET1* region and TG; follow up experiments would be very reasonable for a future graduate student project.

### 5.4 *PSMD9* and cholesterol

Testing whether *PSMD9* is associated with hypercholesterolemia in FCH was conducted in an identical manner for *DIET1* analyses. The cases were FCH patients and controls were FHTG patients; so the same data from Johansen *et al* (*70*) were used. The hg19 human genomic coordinates for *PSMD9* is on chromosome 12: 122,326,646-122,355,771; the hg18 human genomic coordinates of *PSMD9* is chromosome 12: 120,811,029 -120,840,154 and the hg18 human genomic coordinates of *PSMD9* were used because the GWAS data from Johansen *et al (70)* were obtained when hg18 had been the current version of the human genome build.

Since we were only interested in the *PSMD9* region, statistical analyses were only performed on SNPs in the *PSMD9* region from the GWAS data. Logistic regression was performed to test for association of SNPS in *PSMD9* region and hypercholesterolemia; logistic regression essentially tests whether any SNP allele or genotype has a different frequency in cases relative to controls.

Because of LD, SNPs outside the 120,811,029 -120,840,154 region were selected; SNPs from the hg18 chr 12: 120,711,029 -120,940,154 human genomic coordinates were selected for logistic regression. In total, 134 SNPs were present within the chromosome 12 120,711,029 -120,940,154 region in hg18. Logistic regression was adjusted for the

following covariates: sex, body mass index (BMI), diabetes status and population substructure.

The most highly associated SNP from this region with FCH was rs1795964 (*P*=0.08). The Bonferroni-corrected p-value is (0.05/134) 0.0004. However, it is still worth reporting the variant as a good proportion of the SNPs are in Linkage Disequilibrium. So the SNPs that are actually associated due to some relationship with hypercholesterolemia (and not due to LD) are fewer than 134. So this only makes the Bonferroni corrected p-value over-conservative as there are in actuality fewer 'tests'. Nevertheless, it is unlikely that rs1795964 is playing a role in dyslipidemia susceptibility because the 95% Confidence interval of its OR includes 1.0 (**Table 5.3**)

According to NCBI, the rs1795964 SNP is a SNP that is found in the intronic region of a gene named SET domain containing 1B (*SETD1B*); *SETD1B* gene product is a component of a histone methyltransferase complex (*102*), so an obvious mechanistic connection to dyslipidemia is unclear. **Figure 5.2** summaries the *PSMD9* findings.

### 5.5 GLGC-identified LDL-C raising SNPs

The Global Lipids Genetic Consortium (GLGC) identified 37 SNPs that affects LDL-C levels (*43*). The list of these 37 SNPs is shown in **Table 5.4;** *(43).* The SNPs listed in the table are the SNPs most highly associated with LDL-C. For instance rs2131925, which is in the *ANGPTL3* gene, is a SNP primarily associated with TG levels but is also associated with other lipid variables. So, the "best SNP" is taken to be the SNP most strongly associated with a particular lipid, in cases where there are multiple associations. For instance, the "best SNP" in the *ANGPTL3* example is rs3850634

because it is most strongly associated with LDL-C. Because **Table 5.4** mainly shows SNPs that affect specifically LDL-C levels, it is slightly modified from Table 1 in (*43*).

The same GWAS data for FCH cases and HTG controls from (*70*) was used for two analyses: (i) to test for association between each of the 37 LDL-C SNPs and LDL-C levels in FCH cases and FHTG controls; and (ii) to determine whether more of risk alleles of the 37 SNPs accumulate in cases relative to controls. I tested for association of each of the 37 SNPs to see if any of these SNPs that affect LDL-C levels in the population also affects LDL-C levels in FCH cases. I performed the risk score analyses to see if a larger number of risk alleles from the 37 LDL-C associated SNPs accumulate in cases relative to controls.

### 5.5.1 Test for association of the 37 SNPs

Logistic regression was performed as in Section 5.4. The 37 SNPs shown in **Table 5.4** were tested in 159 FCH cases and 128 HTG controls. The Bonferroni corrected p-value was (0.05/37) 0.001. **Table 5.5** shows the results from logistic regression. The two SNPs with the smallest (i.e. most significant) p-values were rs629301 ($P$=0.04) and rs3757354 ($P$=0.05). The SNP with the smallest p-value but highest OR was rs1367117 (OR=1.34) ($P$=0.12).

### 5.5.2 Test for accumulation of all SNPs for LDL-C genetic risk score

I hypothesized that FCH cases would have a greater accumulation of LDL-C raising alleles relative to the HTG controls. Therefore, all 37 LDL-C raising SNPs were tested to see if FCH patients have a more accumulation of LDL-C associated risk alleles

and thus a higher genetic risk score. The unweighted LDL-C genetic risk score was used for comparison. Since the unweighted LDL-C genetic risk score was not normally distributed, the Wilcoxon Rank Sum test was used. The mean LDL-C genetic risk score was greater in FCH cases than in HTG controls at a rate almost approaching statistical significance ($P$=0.054).

**Summary**

The results of the association analysis of *DIET1* with TG, *PSMD9* and cholesterol and LDL-C genetic risk score (GRS) and hypercholesterolemia show suggestive positive trends that are close to, or of borderline statistical significance. The relatively small sample size here was consistent with a pilot project to test these hypotheses; larger sample sizes will be needed to determine whether these trends can attain statistical significance.

**Table 5.1.** Baseline characteristics (mean±SEM) of 463 HTG cases and 1197controls.

| Clinical attribute | HTG cases | Controls |
|---|---|---|
| **n** | 463 | 1197 |
| **Female,%** | 30.7 | 40.4 |
| **Age, years** | 50.9±13.0 | 47.8±11.1 |
| **BMI, kg/m$^2$** | 29.9±4.9 | 26.4±4.6 |
| **TC, mmol/L** | 8.2±3.9 | 5.3±1.3 |
| **LDL-C, mmol/L** | - | 3.4±1.2 |
| **HDL-C, mmol/L** | 0.9±0.3 | 1.4±0.4 |
| **TG, mmol/L** | 14.3±1.8 | 1.1±0.7 |

Note: BMI, Body Mass Index; TC, Total Cholesterol; LDL-C, Low Density Lipoprotein Cholesterol; HDL-C; High Density Lipoprotein Cholesterol; TG, Triglyceride. Mean values of clinical attributes are represented as mean ± standard deviation. LDL-C level is not accurately calculated using the Friedewald equation for HTG patients when plasma TG concentration exceeds 4.5 mmol/L

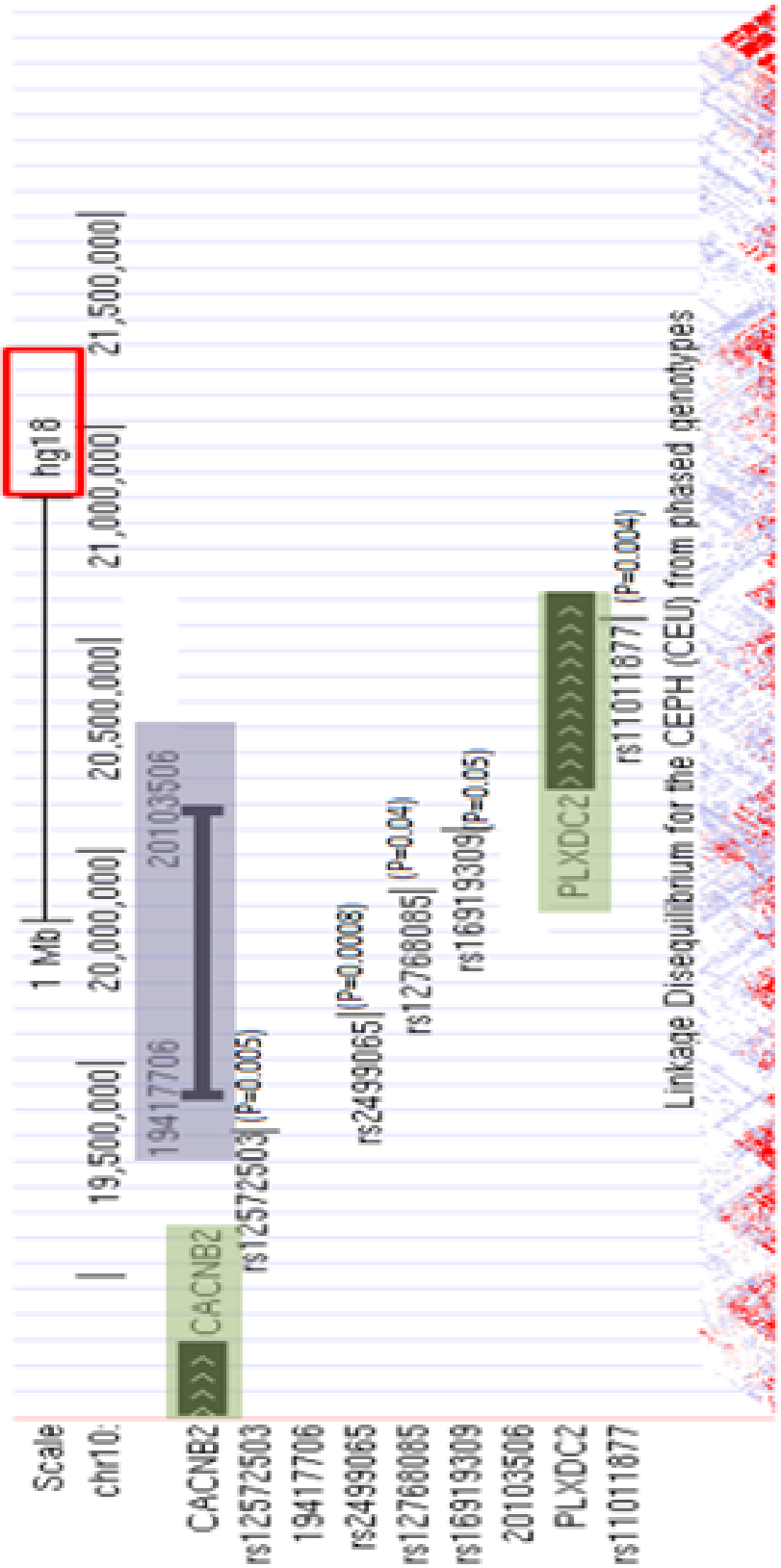
**Table 5.2. Top five associated SNPs, from *DIET1* gene region (hg19 19377700-20063500) in humans**

| SNP ID | Nucleotide position in Chromosome 10 in hg19 coordinates | Minor allele frequency | Nucleotide change | Odds Ratio with 95% CI (unadjusted*) | P-value (unadjusted*) | Odds Ratio with 95%CI (adjusted*) | P-value (adjusted*) |
|---|---|---|---|---|---|---|---|
| rs2499065 | 19572729 | 0.4 | T>C | 1.2 (1.1-1.4) | 0.007 | 1.4 (1.1-1.6) | 0.0008 |
| rs11011877 | 20505183 | 0.2 | A>G | 1.4 (1.2-1.6) | 0.0002 | 1.3 (1.1-1.7) | 0.004 |
| rs12572503 | 19297179 | 0.06 | C>G | 1.4 (1.1-1.9) | 0.01 | 1.6 (1.1-2.2) | 0.005 |
| rs12768085 | 19862640 | 0.005 | G>A | 2.2 (0.9-5.8) | 0.1 | 3.2 (1.1-9.5) | 0.04 |
| rs16919309 | 20010320 | 0.004 | T>C | 3.4 (1.2-9.7) | 0.03 | 3.5 (1.0-12.5) | 0.05 |

Note: CI, Confidence interval. Adjusted values were obtained from a logistic regression model that was adjusted for Sex, Body Mass Index (BMI), Diabetes Status and Population Substructure Unadjusted values were obtained from a logistic regression that was not adjusted for Sex, BMI, Diabetes Status and Population Substructure

**Figure 5.1. Representation of the top 5 SNPs in the 10p12.31 chromosome region using the UCSC genome browser template.** (a) Because UCSC has not released its Linkage disequilibrium (LD) map for the hg19 genome build, the hg18 genome build was used to show the LD pattern of SNPs in the region corresponding to the *DIET1* gene in humans in the European population. LD map is a representation of SNPs that are in LD with each other across the genome. Since SNPs are in LD with SPNs physically close, LD occurs in LD 'blocks'. The "CEPH (CEU) from phased genotypes" refers to the LD pattern for the European population. The 19,417,706-20,103,506 hg18 coordinates correspond to the hg19 19,377,700-20,063,500 coordinates. P-values from logistic regression adjusted for sex, BMI, diabetes status and population substructure are in the parentheses next to SNP ID. For the LD blocks in the European population, the intensity of the colour is proportional to the LD measure; the red colour represents a stronger LD measure than the less intense lavender purple colour. (b) is a zoomed in image of (a) that better shows which LD block the top SNPs are in. The green dotted lines show which LD block the top SNPs belong to.

a


Scale
chr10:
1 Mb
19,500,000
20,000,000
20,500,000
21,000,000
21,500,000
hg18
CACNB2
rs12572503
19417706
rs2499065
rs12768085
rs16919309
20103506
PLXDC2
rs11011877
(P=0.005)
(P=0.0008)
(P=0.04)
(P=0.05)
(P=0.004)
Linkage Disequilibrium for the CEPH (CEU) from phased genotypes

**b**

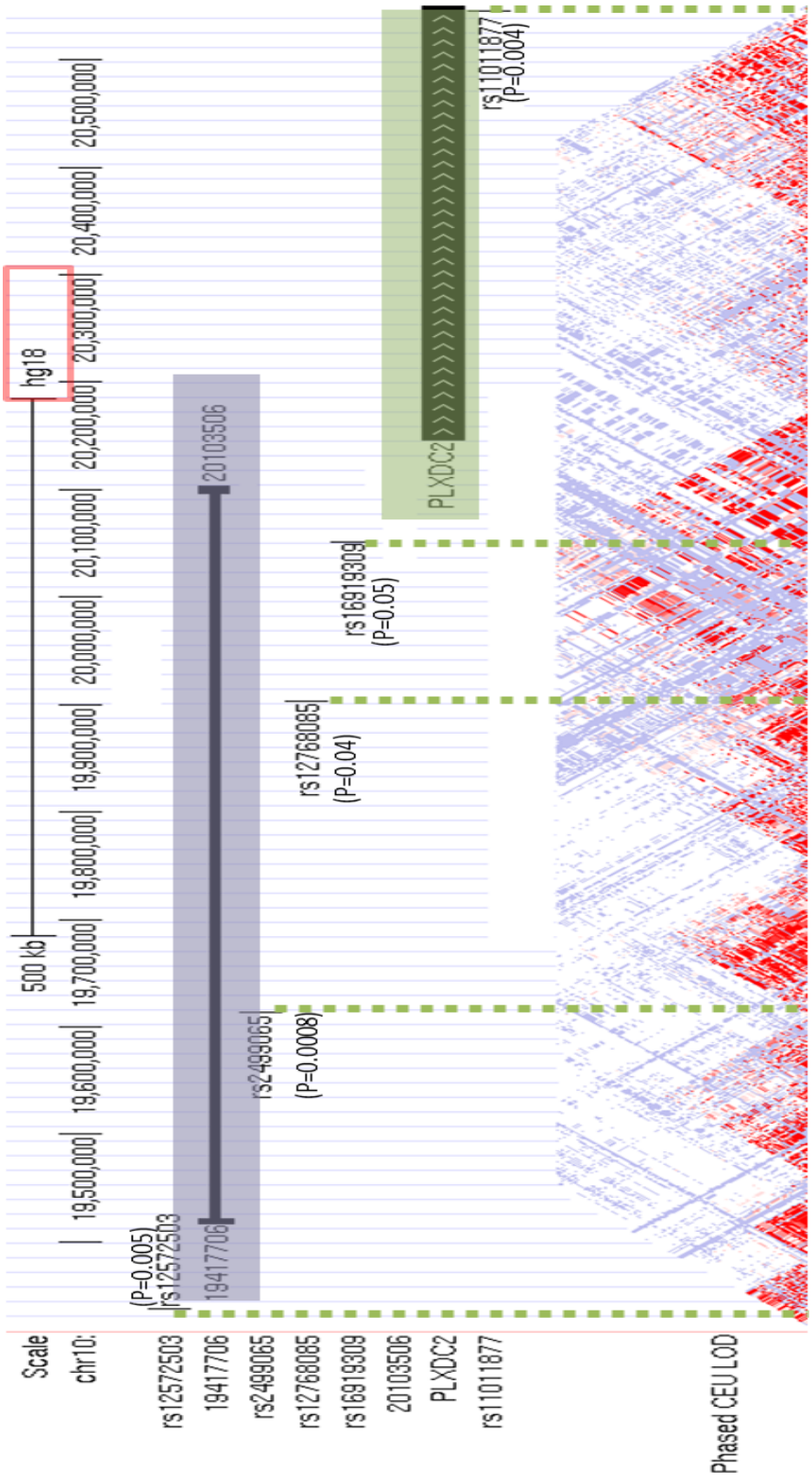

Scale
chr10:
19,500,000
19,600,000
19,700,000
19,800,000
19,900,000
20,000,000
20,100,000
20,200,000
20,300,000
20,400,000
20,500,000
500 kb
hg18
rs12572503
19417706
rs2499065
rs12768085
rs16919309
20103506
PLXDC2
rs11011877
Phased CEU LOD
(P=0.005)
(P=0.0008)
(P=0.04)
(P=0.05)
(P=0.004)

**Table 5.3 Most associated SNP in *PSMD9* region between FCH cases and FHTG controls**

| SNP ID | Nucleotide position in chromosome 12 in hg19 coordinates | Minor allele frequency | Nucleotide change | Odds Ratio with 95%CI (unadjusted) | P-value (unadjusted*) | Odds Ratio with 95%CI (adjusted*) | P-value (adjusted*) |
|---|---|---|---|---|---|---|---|
| rs1795964 | 122267117 | 0.41 | c.5599-855A>G | 1.3(0.9-1.8) | 0.17 | 1.4(0.95-1.7) | 0.085 |

Note:

CI, Confidence Interval

Adjusted values were obtained from a logistic regression model that was adjusted for Sex, Body Mass Index (BMI), Diabetes Status and Population Substructure Unadjusted values were obtained from a logistic regression that was not adjusted for Sex, BMI, Diabetes Status and Population Substructure

**Figure 5.2**. **Representation of the rs1795964 variant in the 12q24 chromosomal region using the UCSC genome browser template.** Because UCSC has not yet released its linkage disequilibrium (LD) map for the hg19 genome build, the hg18 genome build was used to show the LD pattern of rs1795964 non-coding SNP in the European population. The "CEPH (CEU) from phased genotypes" refers to the LD pattern for the European population. For the LD blocks in the European population, the intensity of the colour is proportional to the LD measure; the red colour represents stronger LD measure than the less intense lavender purple colour. The blue dotted line shows which LD block the rs1795964 belongs to.

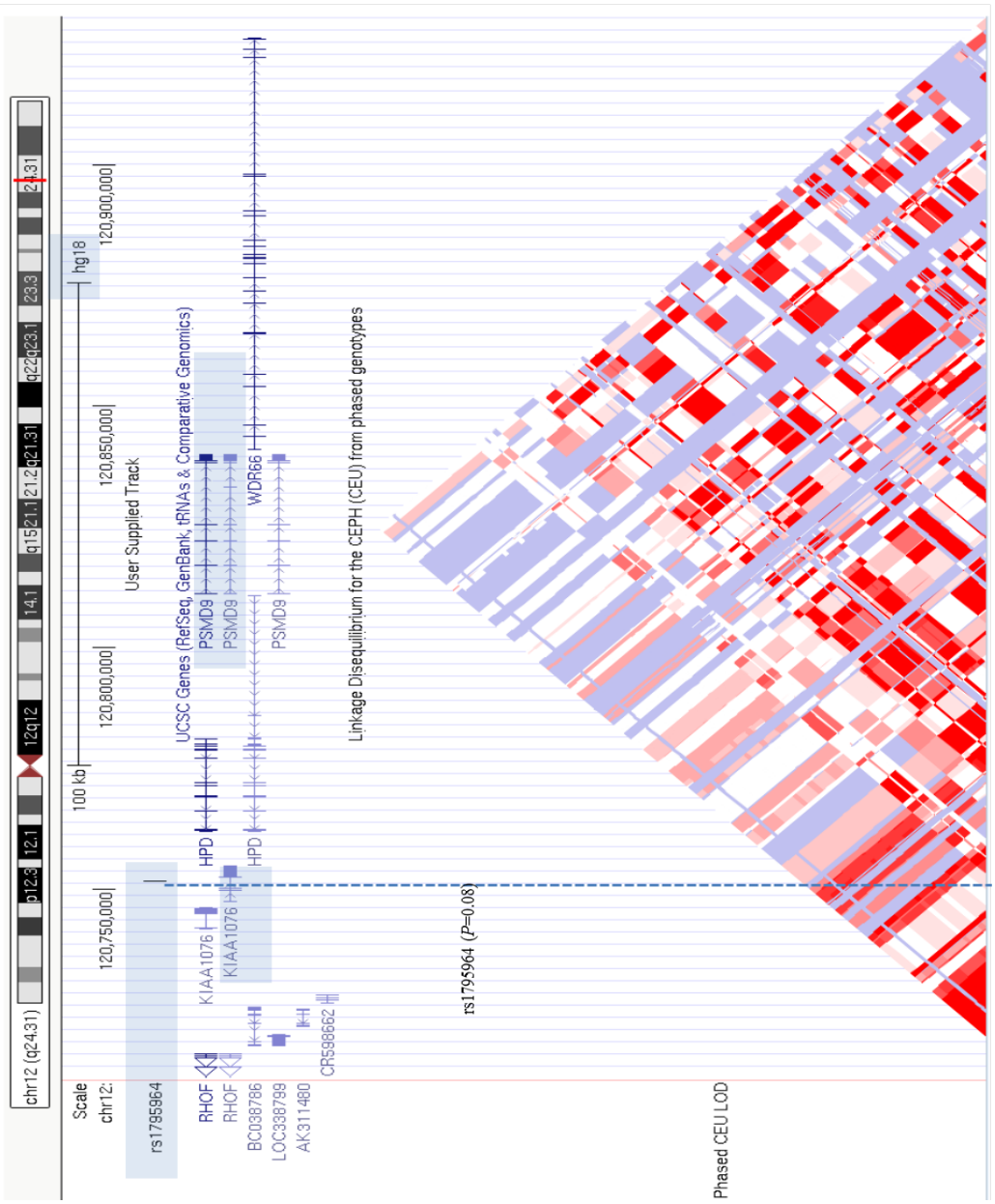

chr12 (q24.31)
p12.3
12.1
12q12
14.1
q15
21.1
21.2
q21.31
q22
q23.1
23.3
24.31
Scale
chr12:
100 kb
120,750,000
120,800,000
120,850,000
120,900,000
hg18
User Supplied Track
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
Linkage Disequilibrium for the CEPH (CEU) from phased genotypes
rs1795964
rs1795964 (P=0.08)
RHOF
KIAA1076
HPD
PSMD9
WDR66
BC038786
LOC338799
AK311480
CR598662
Phased CEU LOD

**Table 5.4** List of 37 GLGC identified SNPs that affect LDL-C levels in the general population

| SNP | Effect Size | Risk allele | Nearby genes | Chromosome |
|---|---|---|---|---|
| rs12027135 | -1.1 | T | *LDLRAP1* | 1 |
| rs2479409 | 2.01 | G | *PCSK9* | 1 |
| rs3850634 | -1.59 | T | *ANGPTL3* | 1 |
| rs629301 | -5.65 | T | *SORT1* | 1 |
| rs2807834 | -1.09 | G | *MOSC1* | 1 |
| rs514230 | -1.13 | T | *IRF2BP2* | 1 |
| rs1367117 | 4.05 | A | *APOB* | 2 |
| rs4299376 | 2.75 | G | *ABCG5/8* | 2 |
| rs12916 | 2.45 | C | *HMGCR* | 5 |
| rs6882076 | -1.67 | C | *TIMD4* | 5 |
| rs3757354 | -1.43 | C | *IDOL* | 6 |
| rs1800562 | -2.22 | G | *HFE* | 6 |
| rs3177928 | 1.83 | A | *HLA* | 6 |
| rs11153594 | -0.89 | C | *FRK* | 6 |
| rs1564348 | 1.95 | C | *LPA* | 6 |
| rs12670798 | 1.26 | C | *DNAH11* | 7 |
| rs217386 | -1.17 | G | *NPC1L1* | 7 |
| rs2126259 | -2.22 | C | *PPP1R3B* | 8 |
| rs1030431 | 0.95 | A | *CYP7A1* | 8 |
| rs2954022 | -1.84 | C | *TRIB1* | 8 |
| rs11136341 | 1.4 | G | *PLEC1* | 8 |
| rs649129 | 2.05 | T | *ABO* | 9 |
| rs1129555 | 1.08 | A | *GPAM* | 10 |
| rs174583 | -1.71 | C | *FADS1-2-3* | 11 |
| rs964184 | 2.85 | G | *APOA1-C3-A4-A5* | 11 |
| rs11220462 | 1.95 | A | *ST3GAL4* | 11 |
| rs11065987 | -0.97 | A | *BRAP* | 12 |
| rs1169288 | 1.42 | C | *HNF1* | 12 |
| rs2332328 | 1.17 | T | *NYNRIN* | 14 |
| rs247616 | -1.45 | C | *CETP* | 16 |
| rs2000999 | 2 | A | *HPR* | 16 |
| rs7225700 | -0.87 | C | *OSBPL7* | 17 |
| rs6511720 | -6.99 | G | *LDLR* | 19 |
| rs10401969 | -3.11 | T | *CILP2* | 19 |
| rs4420638 | 7.14 | G | *APOE-C1-C2* | 20 |
| rs2902941 | -0.98 | A | *MAFB* | 20 |

| | | | | |
|---|---|---|---|---|
| rs909802 | 1.41 | T | *TOP1* | 20 |

**Table 5.5 Results from testing for association, using logistic regression, of the 37 GLGC-identified LDL-C SNPS**

| SNP | Gene | OR(95% CI) | p-value |
|---|---|---|---|
| rs629301 | *SORT1* | 0.63 (0.41-0.98) | 0.04 |
| rs3757354 | *IDOL* | 0.65 (0.42-0.99) | 0.05 |
| rs1367117 | *APOB* | 1.34 (0.90-1.90) | 0.12 |

# Chapter 6: Discussion

Hypercholesterolemia is a classical major risk factor for cardiovascular disease (CVD), the most common cause of mortality in North America. Therefore, having Familial Hypercholesterolemia (FH) (Fredrickson Type 2A) or Familial Combined Hyperlipidemia (FCH) (Fredrickson Type 2B) puts an individual at risk for CVD development because of the hypercholesterolemia that is characteristic to both dyslipidemias. The etiology of hypercholesterolemia can be either environmental and/ or genetic. The underlying theme of all three projects comprising my thesis was to better understand the genetic etiology of hypercholesterolemia, with the heaviest focus on better understanding the genetic basis of hypercholesterolemia in FCH. In the course of this work, I therefore investigated patients with FCH, but also those with FH and with Familial Hypertriglyceridemia (HTG) (Fredrickson Type 4).

The overall hypothesis of my first study was that FCH could actually represent the simultaneous co-existence of FH and FHTG. My second study focused on FH and in particular non-classical genes that could be linked in some cases, such as *APOE* and also the polygenic LDL-C genetic risk score (comprised of SNPs). My third study focused on non-classical genetic determinants of HTG and hypercholesterolemia using GWAS data. So, the work done independently on FH and HTG, in essence converge towards further understanding the two phenotypes that are characteristic of FCH, namely hypercholesterolemia and hypertriglyceridemia.

Knowing the genetic basis of diseases is a major upstream step in understanding the biochemistry underlying the disease, which can consequently lead to implementation

of better diagnoses and better treatment. For instance, the Nobel-Prize winning discovery of the genetic basis of FH has led to the understanding of the disease at the biochemical level (*39*) and has also led to the development and implementation of LDL-C lowering drugs(*18*). Better diagnosis leads to better prognosis, as shown by Kaplan-Meier survival curves that showed better survival rate for FH patients on statin treatment (*63*).

The main goal of my first study was to better understand the genetic basis of hypercholesterolemia in FCH, which is the most common genetic dyslipidemia in the North American population. FCH is also the principal dyslipidemia in 20% of individuals with coronary heart disease (CHD) (*103*). The genetic etiology of HTG is better understood than the hypercholesterolemia component in FCH, due in large part of previous work from the Hegele lab. Previous work has explained 42% of genetic variation of HTG for the four Fredrickson polygenic HTG-associated phenotypes (Fredrickson Type 2B, Type III, Type 4 and Type V) (*71*). Low plasma levels of high density lipoprotein cholesterol (HDL-C), which are strongly associated with CVD risk, is also often characteristic of FCH (*104*). None of my projects focused on low HDL-C, perhaps fortunately since the direct role of HDL-C in causing CVD and CHD is recently facing questions (*20*).

My second study, which was composed of two sub-projects, was aimed at resolving the genetic etiology of hypercholesterolemia in FH patients, in whom the hyperlipidemia was not due to the any of the four known FH causing genes. The first sub-project of my second study tested whether the LDL-C genetic risk score was higher in Mutation Negative Familial Hypercholesterolemia (FH/M-ve) patients relative to Mutation Positive Familial Hypercholesterolemia (FH/M+ve) patients; in other words,

the first sub-project tested whether the genetic etiology of FH could be polygenic. The second sub-project of the second study aimed at looking for other rare monogenic causes of FH among patients in whom FH was not due to the any of the four known FH causing genes. Towards this goal, *APOE* was sequenced in the second sub-project.

For the first sub-project of the second study, I hypothesized an alternate genetic etiology for FH could be the accumulation of LDL-C raising SNPs. In the British study from our collaborators at University College, London (*48*), they were the first to test this hypothesis in a UK and Belgian cohort. So, my study was both the first to test this hypothesis in the Canadian population and also served as a replication study of theirs. I was also interested in the results of the second sub-project, because I reasoned that whatever caused hypercholesterolemia in FH/M-ve patients could also be cause hypercholesterolemia in FCH. So, my FH findings could converge to specify future directions on further studies on FCH.

For the third study of my thesis, a Genome Wide Association Study (GWAS) in hyperlipidemia patients and healthy controls was performed using data previously obtained (*70*). This data represents a wealth of data for various follow-up genetic analyses to test new hypotheses regarding hyperlipidemia. In particular I used this data for all three sub-projects of my third study. Note, this data studied 1197 healthy controls and 463 HTG cases (*70*). These HTG cases were patients with polygenic HTG, namely, FCH (Fredrickson Type 2B), Familial HTG (Fredrickson Type 4), Familial dysbetalipoproteinemia (Fredrickson Type III) and Mixed Dyslipidemia (Fredrickson Type V).

The *DIET1* gene in mice (also called *Diet1*) has been associated with HTG. . Synteny refers to similar genes present in similar chromosomal locations between species and the syntenic *DIET1* region has not been well-annotated in humans. Association of the syntenic *DIET1* region with HTG has not previously been evaluated. So, for the first sub-project of the third study, I was interested in finding whether or not *DIET1* plays a significant role in HTG in humans, by searching for any highly associated SNPs in the *DIET1* region. All the polygenic HTG patients from (*70*) were studied as my cases and all the healthy normolipidemic patients from (*70*) were my controls. As mentioned earlier, the Hegele lab has identified a large proportion of the genetic basis of HTG; 42% of susceptibility to HTG, encompassing Type 2B, Type III, Type 4 and Type V Fredrickson dyslipidemias is due to a combination of common SNPs and heterozygous rare variants. So, I was also interested in the first sub-project of the third study to see if I could further add to the explained proportion of the genetic basis of HTG (which is also the other characteristic biochemical phenotype of FCH).

The second sub-project of the second study aimed at identifying whether the genomic region harboring the *PSMD9* gene is associated with hypercholesterolemia in FCH, given that this region has recently been associated with hypercholesterolemia (*65*).

For the third sub-project of the third study, the same GWAS data from FCH patients and HTG patients was used in a case-control study, where FCH patients were defined as cases and HTG patients were defined as controls. I hypothesized that the 37 SNPs that affect LDL-C levels in the general population would also affect LDL-C levels in FCH, especially since the genetic architecture of FCH is increasingly appreciated as being polygenic. Out of the total of 463 HTG cases, FCH patients (n=159) were our cases

and HTG patients (n=128) were our controls. So, I tested for association of the 37 SNPs with LDL-C levels in FCH. Then we tested for whether FCH patients have a higher genetic LDL-C genetic risk score, where the LDL-C genetic risk score was calculated using the effect sizes of the 37 GLGC identified SNPs (*43*).

Even though my second and third studies focused on FH and HTG, respectively, findings from the latter two projects can lead to future directions for studying the two phenotypes (i.e. hypercholesterolemia and hypertriglyceridemia) that are the defining biochemical phenotypes in FCH.

### 6.1 Findings from resequencing of candidate genes in FCH

Before I undertook my first project, the genetic basis of hypercholesterolemia in FCH was poorly understood in the field. The first researchers to describe and characterize Familial Combined Hyperlipidemia (FCH) (*103*) described the disorder to be an autosomal dominant disorder that was the most common dyslipidemia in humans (*68*). Since FCH was proposed in the 1970's by Goldstein to be an autosomal dominant disease, researchers in the field have believed FCH was due to a single gene. Despite decades of heroic efforts, a single gene could not be found for most cases of FCH. Thus, by default, FCH was then considered to be polygenic, although there was no direct evidence for this idea until recently (*69*).

The hypothesis of my first project was that individuals with FCH have a greater accumulation of heterozygous rare and potentially deleterious mutations in the three Autosomal Dominant Hypercholesterolemia (ADH) causing genes, namely *LDLR*, *APOB* and *PCSK9*. We did not focus on *ARH* because (i) FCH does not show an autosomal recessive pattern and (ii) only a minute percent of the total population of FH is caused by

recessive mutations in *ARH.* Since work from the Hegele lab has previously explained the genetic etiology of hypertriglyceridemia (HTG) in FCH, I hypothesized that FCH could be a condition where individuals have polygenic susceptibility to HTG in combination with FH-causing mutations. My study was the first to sequence the three known ADH causing genes in a case-control setting. My study was also the first to utilize individuals with HTG as controls for studying genetic basis of hypercholesterolemia in FCH. Sanger sequencing, which is a Nobel Prize winning method (*105*), is still the gold standard for identifying novel and known mutations. Our study was the first to Sanger sequence *LDLR*, *APOB* and *PCSK9* genes in both FCH cases and age- and sex-matched HTG controls.

The ratio of rare missense variants for *LDLR* was 7 to 4, giving rise to an OR of 1.2 with a confidence interval that included 1.0. The comparable ORs were 0.74 and 0.69 for *APOB* and *IDOL,* respectively and the 95% confidence interval included 1.0 for both genes. No analyses could be done for *PCSK9* as no variants were found. Post hoc power calculations showed that the studies had a statistical power to detect small effects that was well below 80%. Therefore lack of significant results in my study can be explained by low power. Not having a definite idea of the power of my type of study is a limitation as effect sizes (as represented through ORs) of rare variants accumulating in candidate genes cannot be determined a priori, especially since my study was the first to perform this type of analysis. Realistically, finding appropriate sized samples from a single centre that would provide satisfactory statistical power is impractical, as samples of such patients are not easy to assemble. Nevertheless it is still worthwhile attempting such

studies, on the chance that there is a large biological effect that might be detectable with a relatively small sample. Again, the effect size was unknown a priori in my study.

Since I did not perform any functional work on any of the variants found, a literature search was performed on all 31 rare missense variants that were observed in the case-control studies to determine if any of the variants had already been functionally evaluated. I performed a literature search to date (June 1, 2013) by searching the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/) and Human Gene Mutation Database (HGMD) (http://www.hgmd.cf.ac.uk/ac/index.php).

### 6.1.1 Findings from literature search of all 31 missense variants found in my study

***LDLR:* p.T705I**

Loux *et al*, were the first to report this variant in a French family: a proband, and father and sister had the *LDLR*: *p.*T705I variants and were all FH patients (*106*). *LDLR*: *p.*T705I is a mutation occurring in the O-linked sugar domain of *LDLR (106).* Despite the fact that the *LDLR*: p.T705I variant tends to segregate in families(*107*) (*106*), Brussgaard *et al* mentioned that the role of *LDLR*: p.T705I in FH has been controversial (*107*). So Graham *et al* investigated the *LDLR:* p.T705I variant by screening the *LDLR:* p.T705I in 207 normolipidemic controls (*108*). Throughout the literature, the *LDLR:* p.T705I variant has been said to be in linkage disequilibrium (LD) with an intronic variant in of exon 7 (rs72658861) (*109*) (*108*). The rs72658861 variant and the *LDLR:* p.T705I variant were also in LD in our FCH cases. Likewise, the *LDLR:* p.T705I variant could also be in LD with an actual causative FH-causing mutation, which could explain why the p.T705I variant has been shown to segregate in FH families (*106*). Graham *et al* concluded that

the *LDLR:* p.T705I does not cause FH because it was found in normolipidemic controls at a population frequency of greater than 1% , which is above the cut-off frequency for defining a variant to be a mutation. In 2008, Leigh *et al* updated the University College London, UK database for Low Density Lipoprotein Receptor Familial Hypercholesterolemia database and agreed that the *LDLR:* p.T705I variant was considered a non-FH causing (*110*). *LDLR:* p.T705I was found in 2 of our normocholesterolemic HTG controls. So, with our data and literature search on the *LDLR:* p.T705I variant, the bulk of the data favour *LDLR:* p.T705I being non-causative of FH.

***LDLR:* p.G-2R**

Amsellems *et al* (*111*) were the first to discover the *LDLR* G-2R variant in a male FH patient. This patient inherited the *LDLR*:p.G-2R variant from his father, who did not have FH and also inherited an *LDLR:* p.V502M variant (rs28942080) from his mother, who had FH. *LDLR:* p.V502M was not seen in my study. The *LDLR:* p.V502M was first reported in Hobbs *et al* in an FH patient, but with no other information (*112*). Romano *et al* (*113*) showed that the *LDLR:* p.V502M variant lowers LDLR activity through functional studies. So, it is very likely that the FH male patient with the *LDLR:* p.G-2R, in Amsellems *et al*, had FH because of the *LDLR*: p.V502M variant he inherited from his FH mother and not the *LDLR:* p.G-2R variant he inherited from his non-FH father.

Amsellems *et al* did not report any functional work for the *LDLR:* p.G-2R variant, but they suggested that based on the structural and functional knowledge of *LDLR* gene, it is very likely the *LDLR:* p.G-2R variant is disease causing (*111*). For instance , the -2

position is conserved in mammals and is found in the signal peptide domain of *LDLR (111)* thus making it likely to be a disease causing variant. Fouchier *et al* (*114*) were the first to report the *LDLR:* p.G-2R variant in a Dutch Population. Fouchier *et al* simply reported the occurrence of the variant and did not provide any information on whether the variant segregated with FH phenotype (*114*). Amsellems *et al* reported that all the novel mutations they found segregated with the FH phenotype and were not found in 150 chromosomes of normolipidemic individuals (i.e. 75 normolipidemic individuals) (*111*). I also did not find the *LDLR:* p.G-2R in our normocholesterolemic controls. Nevertheless, functional work for *LDLR:* p.G-2R, which is currently non-existent in the literature, would be needed in order to confidently characterize it as an FH-causing variant. So I do not consider *LDLR:* p.G-2R to be an FH-causing mutation.

***LDLR:* p.G314S**

Hobbs *et al* were the first and so far only to report the *LDLR:* p.G314S variant in literature (*115*). Their functional studies showed that the *LDLR:* p.G314S variant lowers LDLR activity. Since, I found only one FCH patient with the *LDLR:* p.G314S variant with no occurrences in the FHTG controls, I consider the *LDLR:* p.G314S variant to be probably causative of the hypercholesterolemia seen in this single FCH patient.

***LDLR:* p.D333V**

Hobbs *et al* were the first to report the *LDLR:* p.D333V variant (*115*) and showed that the variant lowers LDLR activity through their functional studies. The *LDLR:* p.D333V variant was initially found in an FH individual. The *LDLR:* p.D333V variant

has also been mentioned in non-hyperlipidemia studies. In Liljedahl *et al* (*116*), the *LDLR:* p.D333V variant, along with other variants, was used in evaluation and comparison of microarray technologies. In Andreotti *et al* (*117*) the *LDLR:* p.D333V variant, along with other known LDL-C raising variants was shown to be associated with biliary tract cancers (*117*). So The *LDLR:* p.D333V variant was used because it was considered to be pathogenic, although not necessarily for hypercholesterolemia. Based on literature findings and because I found the *LDLR:* p.D333V variant in only FCH cases, I consider this variant to be possibly causative of the hypercholesterolemia seen in the FCH patients.

***LDLR:* p.L561P**

The Hegele lab was the first to report the *LDLR:* p.L561P variant and this variant was found in an FH individual (*118*). Wang *et al* did not perform any functional work on the *LDLR:* p.L561P variant (*118*). So I cannot definitively conclude that this variant is causative of hypercholesterolemia, even though I found this variant only in FCH cases.

***LDLR:* p.C677G**

This variant to date (June 2013) could not be found in the HGMD (http://www.hgmd.cf.ac.uk/ac/index.php ), NCBI(http://www-ncbi-nlm-nih-gov.proxy2.lib.uwo.ca/snp/?term=%28%28ldlr[Gene+Name]%29+AND+11231152[Base+Position]%29+AND+19[Chromosome] ), Ensembl (http://uswest.ensembl.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g=ENSG00000130164;r=19:11200038-

11244492;v=rs147509697;vdb=variation;vf=38263839#missense_variant_tablePanel ) and 1000 Genomes (http://browser.1000genomes.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g=ENSG00000130164;r=19:11200038-11244492;t=ENST00000558518;v=rs5931;vdb=variation;vf=1894#missense_variant_tablePanel) databases. So, while this variant is novel, no functional work has been performed and thus I cannot conclude that it is causative of hypercholesterolemia in our FCH cases.

***LDLR:* p.V806I**

Hobbs *et al* (*115*) found the *LDLR:* p.V806I variant in an FH individual; this variant showed lowered LDLR activity through their functional work. Lombardi *et al* (*119*) were the first to report the *LDLR:* p.V806I variant in an FH patients from the Dutch population. Zakharova *et al* (*120*) was the first to report the *LDLR: p.*V806I variant in a Russian population. Zakharova *et al* mentioned that the *LDLR:* p.V806I position is found in the internalization signal whose sequence is NPVY, where V is not conserved. So Zakharova *et al* mentioned that the substitution of Valine for Isoleucine may not affect LDLR function as the V in the NPVY internalization signal is not conserved (*120*). However, Zakharova *et al* did not perform any functional work to disfavour the *LDLR:* p.V806I variant being causative of FH. Also, the range of species used for their multiple sequence alignment may differ from the species I used for multiple sequence alignments. My multiple sequence alignment showed that the *LDLR:* p.V806I position was conserved across species in which *LDLR* is said to be conserved according to NCBI (Figure 3.1).

Laurie *et al* reported the pathogenicity of the variant to be uncertain (*121*). In 2010, Huijgen *et al* (*122*) classified the *LDLR:* p.V806I variant as non-pathogenic because they cited Huijgen *et al* cited Defesche *et al* (*123*), Fouchier *et al* (*114, 124*) and Lombardi *et al* (*125*).

Interestingly, Defesche *et al* (*123*) reported the *LDLR:* p.V806I variant in an FH patient but did not perform any functional work that would disfavour the variant being causative of FH and cited sources that did not even report the *LDLR:* p.V806I variant. These sources were Lombardi *et al* (*125*), Graham et al(*108*), Jensen *et al* (*126*) and Weiss *et al* (*127*). Fouchier *et al* (*114, 124*) reported the *LDLR:* p.V806I variant in FH patients but did not perform any functional work to disfavor the *LDLR:* p.V806I variant being causative of FH. Lomabrdi *et al* did not even mention the *LDLR:* p.V806I variant (*125*). In 2012, Huijgen *et al* (*128*) reported the *LDLR:* p.V806I variant to be non-pathogenic based on a criteria that was non-functional based.

Interestingly, from my literature search, the *LDLR:* p.V806I variant was essentially considered being non-causative of hypercholesterolemia, although most authors did not perform any functional work to support this. The only source that supported the *LDLR:* p.V806I variant being causative of FH performed functional work where the variant lowered LDLR activity (*115*). I also found the *LDLR:* p.V806I variant only in cases. Since the weight of evidence favours the *LDLR:* p.V806I variant as being causative of hypercholesterolemia, I consider the *LDLR:* p.V806I variant to be likely causative of hypercholesterolemia in the FCH cases.

***LDLR:* p.T41M**

Fouchier *et al* (*114*) were the first to report the *LDLR:* p.T41M mutation in a Dutch FH patient. However, I found this variant in only one HTG control subject. No functional work has been reported on this variant. So I cannot conclude that this variant is causative of FH more, especially since it was found only in normocholesterolemic HTG controls.

***LDLR:* p.A585S**

Sun *et al* (*129*) were the first to report the *LDLR:* p.A585S in an FH patient. Through some functional work, Sun *et al* predicted the effect of various variants they found in FH patients to be either mild or severe. However, Sun *et al* did not classify the functional effect of *LDLR:* p.A585S variant (*130*) (*129*). Since the *LDLR:* p. A585S variant only occurred in our normocholesterolemic controls, I consider it to not be causative of hypercholesterolemia in FCH.

The *APOB* gene product is the apolipoprotein that carries the hydrophobic lipid (cholesterol) contents of the LDL particle through the bloodstream. The interaction of ApoB with LDLR is vital for the internalization of LDL-C into the cells. Therefore, mutations in *APOB* that disrupt APOB-LDLR binding will prevent LDL-C from being internalized, which results in hypercholesterolemia. In my first project, I wanted to see if there was an accumulation of FH-causing variants in *APOB* in the FCH cases.

***APOB:* p.C1395Y, p.T3799M, p.I2286V, p.S1586T, p.V4101M**

These variants have each been reported in the dbSNP database, which is part of the NCBI database (http://www.ncbi.nlm.nih.gov/snp/ ). To date (June 2013), no papers could be found citing these variants in the dbSNP database.

***APOB:* p.E2539K**, **p.M2331I**, **p.M4293V**, **p.R1662W**, **p.S3252G**, **p.T3020R**, **p.T4457M**, **p.E2539D**, **p.R2192C**, **p.S3267P** and **p.S4403T**

Our lab was the first to report these variants, that were discovered by Johansen *et al* (*71*) in patients with polygenic Fredrickson HTG phenotypes (i.e. Fredrickson Type 2B, Type III, Type 4 and Type V). Sequencing of the target regions of *APOB* in all of the FCH patients and most HTG controls was already performed in Johansen *et al* (*71*). Sequencing of the target regions of *APOB* in the remaining of the 94 FHTG patients was performed in my first study. So the patients' sequence data from Johansen *et al* (*71*) and my first study were combined for subsequent analyses. No papers cited these variants in the dbSNP database (http://www.ncbi.nlm.nih.gov/snp/) and so no functional work was reported in literature.

***APOB:* p.R3500W**

Gaffney *et al* (*31*) were the first to report the *APOB:* p.R3500W variant. Gaffney *et al* screened hypercholesterolemic patients for mutations at position 3500 in *APOB*. *APOB:* p.R3500Q was the first *APOB* mutation established to cause FH (*31*). Gaffney *et al* compared the effect of *APOB:* p.R3500Q variants and the *APOB:* p.R3500W on the

APOB-LDLR interaction. Gaffney *et al* showed both variants bound defectively, found no difference in effect of the two mutations and concluded the *APOB:* p.R3500W variant is causative of hypercholesterolemia. Since I only found the *APOB:* p.R3500W variant in FCH cases, the *APOB:* p.R3500W variant very likely explained hypercholesterolemia in our FCH cases.

***APOB:* p.K1615R**

This variant to date (June 2013) could not be found in the HGMD (http://www.hgmd.cf.ac.uk/ac/index.php ), NCBI (http://www.ncbi.nlm.nih.gov/snp/), Ensemble (http://www.ensembl.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g=ENSG00000084674;r=2:21224301-21266945#missense_variant_tablePanel ) and 1000 genomes (http://browser.1000genomes.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g=ENSG00000084674;r=2:21224301-21266945#missense_variant_tablePanel ) databases. So, this variant is novel and since no functional work was done on this variant, I cannot definitively conclude that it is causative of hypercholesterolemia.

***APOB:* p.Q3404E**

The functional effect of the *APOB* : p.Q3405E variant has been contradictory throughout literature. Findings from Pullinger *et al* support that *APOB:* p.Q3405E does not cause hypercholesterolemia (*131*). In contrast, findings from Gaffney *et al (132)* support the idea that the *APOB:* p.Q3405E variant causes hypercholesterolemia.

Pullinger *et al* used fibroblast binding assays from 11 normolipidemic controls, 7 heterozygous *APOB:* p.Q3405E individuals and 1 homozygous *APOB:* p.Q3405E individual. There was no statistical difference in binding affinity in the 3 groups (*131*)

Gaffney *et al* studied growth of U937 cells from 23 normolipidemic individuals and 13 *APOB:* p.Q3405E individuals as a measure of LDL binding affinity. Gaffney *et al* found 13 heterozygous *APOB:* p.Q3405E individuals at a frequency of 1.4% and 1 homozygous *APOB:* p.Q3405E variant at a frequency of 0.1%. They compared *APOB:* p.Q3405E mutant cells with cells cultured from normolipidemic individuals. They showed that the mean (± SD) cell growth from normolipidemic individuals was 1.1 ± 0.32 units. The mean (± SD) cell growth from *APOB:* p.Q3405E individuals was 0.77± 0.24 units and they reported the difference in mean growth rates to be statistically significant ($P$=0.004).

High triglyceride levels lower the affinity of LDL for its receptor (*132*) and in Gaffney *et al*, a triglyceride level of 2.3mmol/L was defined as the upper limit of the normal range (*132*). After controlling for HTG, by removing HTG individuals, Gaffney *et al* still reported significantly lower LDL binding affinity ($P$=0.009) attributable to the *APOB:* p.Q3405E variant(*132*). However, Figure 5 from Gaffney *et al* showed a considerable discrepancy in cell growth for two separate assays done for the same severely HTG individual, who was removed to control for HTG. This huge variability from the same patient may lead to questioning of the reliability of cell growth values, even when controlled for HTG, for *APOB:* p.Q3405E individuals who had only a single assay performed. Table 1 in Gaffney *et al* reports the mean triglyceride level for the 13 *APOB:* p.Q3405E individuals to be 2.5 mmol/L, which is 0.2 mmol/L above their upper

limit for normal TG level range. So, it can be argued that this mean TG level could bias the significantly lower cell growth in cells derived from *APOB:* p.Q3405E individuals. Also, in the selection of study subjects, Gaffney *et al* stated that patients were referred to them and were not pre-screened. So, it is possible that some of these 13 *APOB:* p.Q3405E individuals actually had FCH, in which HTG is part of the definition.

Gaffney *et al* could not rule out the possibility of other (now) known FH causing mutations that were not detected. For instance, association of gain of function mutations in *PCSK9* with clinical FH was discovered in 2003 (*133*), which was well after Gaffney *et al,* who published their data in 1998. So, Gaffney *et al* could not have screened for FH causing gain-of-function mutations in *PCSK9*. So, it remains possible that the significantly lower cell growth found in *APOB:* p.Q3405E individuals was actually due to other unmeasured or undetected variants and not to *APOB:* p.Q3405E.

Pullinger *et al* proposed that the *APOB:* p.Q3405E variant does not cause hypercholesterolemia because the *APOB:* p.Q3405E variant segregated independently of hypercholesterolemia in a family diagnosed with FH (*131*). Gaffney *et al* also reported no statistical difference in U937 cell growth from three family members, in which the proband and proband's mother had the *APOB:* p.Q3405E genotype but the proband's brother did not have the *APOB:* p.Q3405E genotype. The U937 cell growth of all three family members was all comparable to U937 cell growth from healthy individuals (*132*). Gaffney *et al* screened 200 normolipidemic individuals for the *APOB:* p.Q3405E variant and found it in 4 of the 200 normolipidemic individuals. Despite reporting information that disfavours *APOB:* p.Q3405E being causative of hypercholesterolemia, Gaffney *et al* stated that the extra negative charge brought about the *APOB:* p.Q3405E variant may

alter the structural biology of APOB in such a way that hypercholesterolemia may result downstream (*132*). However, that statement was merely a suggestion or speculation. The balance of all the experimental findings do not support the idea that *APOB:* p.Q3405E is FH-causing,.

Gaffney *et al* and Pullinger *et al* are two papers in literature that discuss the *APOB:* p.Q3405E variant in detail. In my study, the *APOB:* p.Q3405E variant was only found in one normocholesterolemic HTG control and was not found at all in FCH cases. After evaluating all the findings from Gaffney *et al* and Pullinger *et al* along with only finding the *APOB:* p.Q3405E variant in HTG controls, I would argue that the balance of evidences favours the idea that *APOB:* p.Q3405E does not raise LDL-C levels.

***IDOL:* p.R372W** and **p.V339I**

These variants were not reported in HGMD (http://www.hgmd.cf.ac.uk/ac/index.php). These variants were reported in the dbSNP database (http://www.ncbi.nlm.nih.gov/snp/). No papers cited these variants in the dbSNP database and so no functional work was reported in literature. I cannot definitively conclude whether either one is associated with hypercholesterolemia.

***IDOL:* p.C31Y**

This variant to date (June 2013) could not be found in the HGMD (http://www.hgmd.cf.ac.uk/ac/index.php ), NCBI (http://www.ncbi.nlm.nih.gov/snp/), 1000 Genomes (http://browser.1000genomes.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g

=ENSG00000007944;r=6:16129356-16148479#missense_variant_tablePanel ) and Ensembl (http://www.ensembl.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g=ENSG00000007944;r=6:16129356-16148479#missense_variant_tablePanel ) databases. So, because this variant is novel and since no functional work was done on this variant, I cannot definitively conclude that it is causative of hypercholesterolemia.

***PCSK9***

Finally, with respect to PCSK9, given the rarity of mutations in this gene across the entire population of FH patients, it is perhaps not surprising that I did not find any variant in *PCSK9* in my study.

**All rare variants in FCH considered cumulatively**

My literature search of the variants that I found suggests that there was an accumulation of known functionally verified variants in FCH cases at a p-value almost approaching statistical significance ($P$=0.09). No known functionally verified variant was found in controls and 4 known functionally verified variants was found in cases namely *LDLR:* p.G314S, p.D333V,p.V806I and *APOB:* p.R3500W variants. These variants likely explain hypercholesterolemia in those particular FCH patients, which make up about 3%

The results of the re-sequencing of FH candidate genes in FCH cases and FHTG controls indicates non-significant trends that suggest that dysfunctional rare variants accumulate in cases relative to controls. The relatively small sample size here was

consistent with a pilot project to test these hypotheses; larger sample sizes will be needed to determine whether these trends attain statistical significance.

**Limitations of sequencing strategy to identify rare variants**

I only looked at the coding regions of *LDLR*, coding regions of *IDOL*, only exons 26 to 29 in *APOB* and only exon 7 in *PCSK9*. So I cannot rule out potential disease-causing variants outside these regions. Previous studies have shown that disease causing variants can be non-exonic (*134*) and SNPs from GWAS studies have shown that it is SNPs in non-coding regions that are associated with disease traits (*43*) — all of which are in line with findings from the ENCODE project that showed that most variants that control protein biochemistry are non-coding (*135*) .

However, I am still confident in the approach taken to test my hypothesis because it has been long established that non-synonymous rare variants are most likely to be disease causing (*136*). There have been 'success stories', where the discovery of missense rare variants being causative of a disease resulted from the approach where only coding regions were analysed (*137*) (*138*). One such success story that involved analyzing exonic regions is the discovery the *ANGPTL3* gene being causative of Familial Combined Hypolipidemia (*137*).

Most of the missense rare variants I found were either exclusive in FCH cases or FHTG controls; only the *LDLR:* p.T705I and *IDOL*: p.V339I were found in both FCH cases and controls. So, it is possible that the effect sizes of the other 29 missense rare variants vary, and the standard chi Square test is insensitive, since it assumes all missense rare variants to have an equal effect size.

**Table 6.1 Complete list of 31 missense rare variants observed in FCH-FHTG cohort and information from published research on these variants**

| Variant | Variant identifier | MAF | Functionally verified in literature as causative of hypercholesterolemia | References | Population |
|---|---|---|---|---|---|
| *LDLR:* p.T705I | rs45508991 | 0.0065 | No | Loux *et al* (*106*), **1992**<br>Brussgaard *et al(107)*, **2006**<br>Graham *et al* (*108*), **2006**<br>Leigh *et al* (*110*), **2008** | Cases: 1<br>Controls: 2 |
| *LDLR:* p.G-2R | rs147509697 | 0.0022 | No | Amsellems *et al* (*111*), **2002**<br>Fouchier *et al* (*114*), **2005** | Cases:1<br>Controls:0 |
| *LDLR:* p.G314S | CM920439 | 0.0022 | Yes | Hobbs *et al* (*115*), 1992 | Cases:1<br>Controls:0 |
| *LDLR:* p.D333V | rs5930 | 0.0022 | Yes | Hobbs *et al* (*115*), **1992** | Cases:1<br>Controls:0 |
| *LDLR:* p.L561P | CM014578 | 0.0022 | No | Wang *et al* (*118*), **2001** | Cases:1<br>Controls:0 |
| *LDLR:* p.C677G | N/A | 0.0022 | No | NOVEL | Cases:1<br>Controls:0 |
| *LDLR:* p.V806I | rs137853964 | 0.0022 | Yes | Hobbs *et al* (*115*), **1992**<br>Lombardi *et al (119)*, **2000**<br>Zakharova *et al* (*120*), **2005**<br>Huijgen *et al* (*122*), **2010**<br>Huijgen *et al* (*128*), **2012** | Cases:1<br>Controls:0 |
| *LDLR:* p.T41M | CM055350 | 0.0022 | No | Fouchier *et al* (*114*), **2005** | Cases:0<br>Controls:1 |
| *LDLR:* p.A585S | rs72658868 | 0.0022 | No | Sun *et al* (*129*) | Cases:0<br>Controls:1 |
| *APOB:* p.C1395Y | rs568413 | 0.0022 | No | NOVEL | Cases:1<br>Controls:0 |
| *APOB:* p.E2539K | rs1801696 | 0.0022 | No | Johansen *et al* (*71*), **2010** | Cases:1<br>Controls:0 |
| *APOB:* p.M2331I | CM105023 | 0.0022 | No | Johansen *et al* (*71*), **2010** | Cases:1<br>Controls:0 |
| *APOB:* p.M4293V | CM104782 | 0.0022 | No | Johansen *et al* (*71*), **2010** | Cases:1<br>Controls:0 |
| *APOB:* p.R1662H | rs151009667 | 0.0022 | | Johansen *et al* (*71*), **2010** | Cases:1<br>Controls:0 |

| | | | | | |
|---|---|---|---|---|---|
| *APOB:* p.R3500W | rs144467873 | 0.0022 | Yes | Gaffney *et al* (*31*), **1995** | Cases:1 Controls:0 |
| *APOB:* p.S3252G | rs12720854 | 0.0043 | No | Johansen *et al* (*71*), **2010** | Cases:2 Controls:0 |
| *APOB:* p.T3020R | rs61742323 | 0.0022 | No | Johansen *et al* (*71*), **2010** | Cases:1 Controls:0 |
| *APOB:* p.T3799M | rs61744153 | 0.0022 | No | NOVEL | Cases:1 Controls:0 |
| *APOB:* p.T4457M | rs12713450 | 0.0065 | No | Johansen *et al* (*71*), **2010** | Cases:3 Controls:0 |
| *APOB:* p.I2286V | rs584542 | 0.0022 | No | NOVEL | Cases:1 Controls:0 |
| *APOB:* p.E2539D | rs149306841 | 0.0022 | No | Johansen *et al* (*71*), **2010** | Cases:0 Controls:1 |
| *APOB:* p.K1615R | N/A | 0.0022 | No | NOVEL | Cases:0 Controls:1 |
| *APOB:* p.Q3405E | rs1042023 | 0.0043 | No | Pullinger *et al* (*131*), **1996** Gaffney *et al* (*132*), **1998** | Cases:0 Controls:2 |
| *APOB:* p.R2192C | rs141641980 | 0.0022 | No | Johansen *et al* (*71*), **2010** | Cases:0 Controls:1 |
| *APOB:* p.S1586T | rs61742247 | 0.0022 | No | NOVEL | Cases:0 Controls:1 |
| *APOB:* p.S3267P | rs12720855 | 0.0022 | No | Johansen *et al* (*71*), **2010** | Cases:0 Controls:1 |
| *APOB:* p.S4403T | rs72654426 | 0.0022 | No | Johansen *et al* (*71*), **2010** | Cases:0 Controls:1 |
| *APOB:* p.V4101M | rs1801703 | 0.0022 | No | NOVEL | Cases:0 Controls:1 |
| *IDOL:* p.R372W | rs141183183 | 0.0022 | No | NOVEL | Cases:1 Controls:0 |
| *IDOL:* p.V339I | rs142124143 | 0.0043 | No | NOVEL | Cases:1 Controls:1 |
| *IDOL:* p.C31Y | N/A | 0.0022 | No | NOVEL | Cases:0 Controls:1 |

Note: MAF, Minor allele frequency (defined as the frequency of the minor allele ≥ 0.01 in our combined FCH-FHTG cohort.

*LDLR*, Low Density Lipoprotein Receptor (gene);*APOB*, ApolipoproteinB-100 (gene); *MYLIP*, Myosin Regulatory Light Chain Interacting Protein (gene).

References refer to papers that cite the variants.

Population refers to how many of our cases and controls the variant was found in.

N/A, Not Applicable; N/A was reported under variant identifier when variant was novel. N/A was reported under References when there was no research article citing the variant.

Where variants did not have a dbSNP ID (i.e. rs ID), the HGMD accession ID was given as a variant identifier (which are the numbers prefixed with the letters CM)

### 6.1.2 *In silico* analyses of missense variants

The effects of all 31 missense rare variants from the candidate genes studied were predicted *in silico* using PolyPhen-2 and SIFT; results are shown in Chapter 3. Missense rare variants predicted to be deleterious both PolyPhen-2 and SIFT were only found in cases for *LDLR*, namely *LDLR:* p.G314S, p.D333V, p.L561P and p.C677G variants, of which the *LDLR:* p.G314S and *LDLR:* p.D333V variants were accurately predicted as being likely dysfunctional (Table 3.8). This is because these two variants were predicted to be deleterious when functional work showed reduced LDLR activity. In contrast, the *LDLR:* p.V806I has been shown to reduce LDLR activity but was predicted to be tolerated by SIFT (Table 3.8) thus indicating an inaccurate *in silico* prediction. However, only PolyPhen-2 predicted the *LDLR:* p.V806I variant to be deleterious— even though p.V806I (SIFT score 0.06) marginally escaped being predicted as deleterious, since a SIFT score of 0.05 or lower is predicted as being deleterious (Table 3.8).

For *APOB*, 4 missense rare variants were predicted to deleterious by both PolyPhen-2 and SIFT in FCH cases. These variants were the *APOB:* p.R1662H, p.R3500W, p.S3252G and p.T3799M variants of which the *APOB:* p.R3500W variant was predicted accurately. In HTG controls, 2 missense rare variants were predicted to deleterious by both PolyPhen-2 and SIFT, which namely were the *APOB:* p.K1615R and p.S3267P variants. With respect to being causative of hypercholesterolemia, these predictions are very likely to be inaccurate since the variants were found in only normocholesterolemic controls.

For *IDOL*, the p.R373W and p.C31Y variants were predicted to be deleterious by both PolyPhen-2 and SIFT, and were found in cases and controls, respectively. Since the

literature did not report any functional work on these variants, I cannot comment on the accuracy of *in silico* predictions. With respect to being causative of hypercholesterolemia, the *in silico* prediction for the *IDOL* p.C31Y variant is likely to be inaccurate because controls are normocholesterolemic.

The discrepancies in predictions of PolyPhen-2 and SIFT, in variants across candidate genes, could result from differences in algorithms and weighting priorities given to certain features by the two softwares.

For each of the candidate genes, Multiple sequence alignment (MSA) was done in order to visually analyze amino acid conservation in the region of each rare missense variant (Figures 3.1-3.3). For MSA of each candidate gene, 6 species were used for MSA. In the first 4 species, the gene, according to NCBI, was conserved; in the last two species, the candidate gene was not considered to be conserved. In general, some of the amino acid positions of the rare missense variants that I studied were completely conserved and others were not conserved.

After looking at MSAs of *LDLR*, the change of amino acid to the basic arginine at position -2 can create electrostatic bonds with nearby acidic amino acids (shown in green squares) that may negatively affect protein activity (Figure 3.1). These electrostatic bonds may be unfavourable in the sense that it could negatively affect protein activity. Protein activity could also be negatively affected due to the absence of Glycine and not necessarily the presence of Arginine, especially since Glycine is the amino acid that allows most three dimensional freedom.

The *LDLR:* p.T41 position is conserved in the 4 species that show the strongest *LDLR* conservation. The conservative mutation to a Methionine may not affect LDLR

activity and thus may explain why the *LDLR* T41M variant was found in a normocholesterolemic control (Figure 3.1).

The *LDLR:* p.G314 position is fully conserved across species. This mutation was shown to lower LDLR activity. This could be due to the lack of a Glycine at the *LDLR* G314 position, especially since Glycine is the amino acid that allows three dimensional freedom the most. This lowered activity could also be due to a non-hydrophobic amino acid such as Serine at the (Figure 3.1).

The *LDLR:* p.D333 position is also fully conserved across species and mutation to the hydrophobic Valine has been shown to lower LDLR activity functionally. This lowered activity may be due to disruption of electrostatic bonds with nearby basic amino acids (shown in blue squares) (Figure 3.1).

The *LDLR:* p.C677 position is fully conserved and mutation to a Glycine may disrupt disulphide bonds with nearby Cysteines (shown in blue squares) (Figure 3.1).

For all the missense rare variants in *LDLR*, *APOB* and *IDOL* the amino acids in blue squares represent amino acids that could have disrupted electrostatic interactions with amino acids at mutant positions. Amino acids at mutant positions are in red squares. Amino acids in green squares represent amino acids that could have formed electrostatic interactions with amino acids at mutant positions. The possible formation and/or disruption of electrostatic bonds could be unfavourable in that it may lower protein activity. Amino acids in blue squares also represent amino acids that could have disulphide bonds formed or broken—which could negatively affect protein activity.

**Limitations of bioinformatic analyses**

*In silico* predictions in and of themselves are not enough to determine whether an amino acid changing variant is causative of a disease. Literature has shown that predictions by PolyPhen-2 and SIFT are inaccurate; the concordance with functional studies, when performed, is only 50-60% (*139*). This is because variants that have been shown to be deleterious functionally were predicted to be non-deleterious and conversely, variants that have been shown to be non-deleterious functionally were predicted as deleterious (*139*). This discrepancies could be because the *in silico* programs do not take into consideration other aspects of protein biochemistry such as post-translational modification, protein-protein interactions, etc (*139*). Therefore, *in silico* predictions can be considered 'just slightly better than chance' and cannot not be used in clinical decisions such as diagnosis, therapeutics and prognosis. Nevertheless, *in silico* analyses of missense variants are a useful tool in the investigation of variants along with other information such as structural information, biological information, etc. *In silico* analyses, along with other information, can also be useful in the prioritizing of those variants to test functionally. So the information from *in silico* predictions of the 31 missense rare variants, as well as the MSA of the 3 candidate genes, would be useful for any future projects involving functional validation of the variants.

## 6.2 Findings from Familial Hypercholesterolemia Mutation Negative (FH/M-ve) subjects and Familial Hypercholesterolemia Mutation Positive (FH/M+ve)

My second study was composed of two collaborative projects, namely (i) the resequencing of *APOE* in FH/M-ve patients to look for possible undetected deleterious

mutations and (ii) testing whether FH/M-ve patients (cases) have a higher LDL-C genetic risk score than FH/M+ve subjects (controls). In the second collaborative study, the hypothesis had been first tested by our collaborators in the UK. So, my second collaborative study not only served as a replication study, but was the first study to test the hypothesis in the Canadian population.

In the first collaborative project of my second study, not a single potentially deleterious *APOE* mutation was found in the 95 FH/M-ve patients. However, our collaborators for this first study found *APOE* c.L167 del variant that segregated in an FH family and this *APOE* Leu 167 del variant was found to be causative of FH (data yet to be published). Their finding has now made *APOE* the fifth FH-causing gene. Our collaborator's findings is an example of a serendipitous discovery of another FH–causing gene as it is very likely that our collaborator's FH family is one of only few families in the world to have FH due to the *APOE* Leu 167 del mutation. To date (June 2013) an *APOE* Leu 167 del mutation has only been reported in one other large family as being causative of FH (*60*). Therefore, it is not surprising that our 95 FH/M-ve patients did not show any deleterious mutation. The second collaborative project of my second study showed that FH/M-ve patients had a higher mean LDL-C genetic risk score than FH/M+ve controls, meaning that the FH/M-ve patients have a greater accumulation of risk alleles. However, in my sample this greater accumulation was not statistically significant. The effect sizes and absolute values of the mean LDL-C genetic risk scores that I observed for FH/M-ve cases and FH/M+ve controls were comparable to those reported in our collaborators' FH/-ve cases and FH/M+ve controls, respectively (Table 4.7). However, post hoc power calculations showed that our UK collaborators had

sufficient statistical power, while the sample size in my study did not afford sufficient statistical power to detect a difference of this magnitude in the mean values (Table 4.7). So, our lack of statistical significance is likely due to restricted sample size and not due to lack of biological effect. In essence, my studies in the Canadian FH population replicated our UK collaborators' finding that an alternate polygenic etiology (as in the accumulation of risk alleles) can cause FH, thus making it both a polygenic and a monogenic disease.

### 6.3 Findings from analyses of GWAS data

Finally, I applied some of the GWAS data from (*70*) for all 3 projects that made up my third study. The aim of the first project of my third study was to identify if the *DIET1* locus is associated with polygenic HTG in humans, because *DIET1* is associated with HTG in mice. The second and third projects of my third study used only the FCH patients and HTG patients in (*70*) as cases and controls, respectively. The aim of my second project was to test whether the *PSMD9* locus, which was recently reported to be associated with hypercholesterolemia, is associated with hypercholesterolemia in FCH. The third project had two aims: (i) to test for association of the 37 GLGC-identified SNPs with hypercholesterolemia in FCH; and (ii) to test for accumulation of the 37 GLGC-identified LDL-C risk alleles in FCH cases relative to controls.

Association analysis of all the SNPs in the *DIET1* region with HTG did not reach the overly conservative Bonferroni-corrected p-value, thus I could not say that the *DIET1* region was associated with HTG in humans. The true p-value would likely be much less strict than the Bonferroni corrected p-value, since many SNPs at this locus were in linkage disequilibrium; however the LD was largely uncharacterized in this region, I

could not estimate what the appropriate correction should be, and so used the Bonferroni correction as my default approach. Limitations of the first project was sample size, as typical GWAS studies have very large sample sizes (*79*) (*70*). Nevertheless, I am still confident in testing whether the *DIET1* locus is associated with HTG in humans, since nothing was known about this unannotated region in humans previously, and also because my findings will be a useful starting point in future analysis of this region. If and when more is learned about the *DIET1* region in humans, it is likely that fewer SNPs would be tested for association, which would increase the statistical power since the adjustment of the nominal p-value would be less strict. Because my results might be useful for future analyses, I still reported 5 SNPs with the strongest associations — albeit all non-statistically significant (Table 5.2).

In my second project, I found that the most highly associated SNP —albeit not statistically significant— in the *PSMD9* region with FCH was the rs1795964 SNP ($P$=0.08). So it is possible that this region plays a role in hypercholesterolemia in FCH. However, the second project was similarly limited with respect to statistical power, because only subset of subjects from the database in (*70*) was used. However, as with the first project, the information could be useful for future studies.

Finally, I tested for association of each of the 37 GLGC-identified SNPs in the first part of my third project. The top three SNPs with strongest associations were all well below the Bonferroni the corrected p-value (Table 5.6). Since these 37 SNPs were not in LD, the Bonferroni corrected p-value is not overly conservative. Here I may not have detected any significance because of limited statistical power due to sample size.

Nevertheless, my findings would be valuable for future larger-sized meta-analyses studies that could include our samples.

My findings for the second part of the third project showed that there is a greater accumulation of risk alleles in FCH cases relative to controls at a rate borderline of statistical significance (*P*=0.054). Given the sample size, the inclusion of the 37 SNPs increased statistical power compared to testing each of the 37 SNPs individually. So, if I had the opportunity to study a few more patients, the p-value could have been well below 0.05. In essence, the 37 SNPs that affect LDL-C levels in the general population also appear to affect LDL-C levels in FCH. This is in line with the concept that the genetic etiology of FCH is polygenic (*66*).

**6.4 Conclusions**

Findings from my first study have shown that the presence of HTG and FH-causing mutations can interact to produce an FCH phenotype. In other words, my findings suggest that FCH can sometimes indeed result from the co-existence of FH and HTG genetic susceptibility. My finding of the *LDLR:* p.G314S, p.D333V, p.V806I and *APOB:* p.R3500W variants in the FCH cases supports the idea that rare FH-causing mutations are over-represented in FCH. My first study was the first of its kind. However, a similar—yet far from identical— study was performed by Civeira *et al (140),* where *LDLR* and *APOB* variants (203 *LDLR* variants and 4 *APOB* variants) were genotyped in only 143 unrelated FCH patients (*140*). Civeira *et al* found *LDLR* mutations in his FCH population. However, their study, unlike mine, was uncontrolled and so variants they found could also have been present in normocholesterolemic controls, including healthy

individuals and individuals with HTG. For instance, Civeira *et al* stated that none of the mutations they found were reported in normolipidemic individuals and one of the LDLR variants the genotyped was *LDLR:* p.T705I (*140*). Civeira *et al* also stated that the variant they genotyped were reported to be causative of FH (*140*). From our extensive literature search on each of our 31 *LDLR* missense rare variants, we discovered that there were variants reported to cause hypercholesterolemia in the HGMD database, when extensive literature search did not confirm those variants being causative of hypercholesterolemia such as the *APOB:* p.Q3405E, *LDLR:* p.T705I and *LDLR:* p.G-2R variants. Thus having sequence information from controls, which Civeira *et al* did not have, helped in our assessment of whether a variant was possibly causative of hypercholesterolemia. For instance, I found the same *LDLR:* p.T705I that Civeria *et al* found and interpreted as being FH causing, in our normocholesterolemic HTG controls, which helped me discard *LDLR:* p.T705I as being causative of hypercholesterolemia in FCH.

Even though the understanding of the genetic etiology of FCH in the field shifted from being autosomal dominant to polygenic, no one had confidently disprove that FCH is caused by FH causing mutations because no one ever sequenced the three well known FH causing genes. 'Absence of evidence (not knowing for sure that hypercholesterolemia in FCH is due to FH-causing mutations) is not evidence of absence (FCH is not due to FH-causing mutations)'. But the findings from my first study indicate that FCH is for the most part not due to FH causing mutations. Despite the fact that post hoc power calculations showed low statistical power for my first study, if the hypercholesterolemia in FCH had indeed been due to FH-causing mutations, my sample size would have been sufficiently powered. This is because, hypothetically speaking, had

I tested the same hypothesis in 138 FH cases and 94 controls, I would have had a greater accumulation of missense rare variants in our cases.

Findings from my second study support the idea that while FH is mainly monogenic, there are some cases in which there is a polygenic cause. For instance, findings from my study suggest that it might be warranted to test for greater accumulation of the 37 GLGC- identified LDL-C risk alleles in FH cases relative to healthy controls, especially when a mutation in any of the known genes is absent, since the genetic etiology of FH can sometimes be polygenic.

The biological macromolecules such as carbohydrates, proteins, lipids and nucleic acids are vital to human existence. So, it is not unreasonable that many genes are involved in the metabolism of these macromolecules, including lipids. Many genes are involved in cholesterol metabolism, thus mutation in such genes can lead to a clinical presentation of hypercholesterolemia. Therefore, findings from the *APOE* collaborative project would support searching for other possible monogenic causes of FH using next generation sequencing technologies, such as whole exome sequencing, in FH/M-ve cohorts.

Findings from testing for accumulation of the 37 LDL-C risk SNPs in FCH cases as well as findings from my first study show that genetic etiology of hypercholesterolemia in FCH, like in FH, can be monogenic and polygenic. Findings from testing for accumulation of the 37 LDL-C risk SNPs in FCH cases and findings from testing for accumulation of 12 of the 37 LDL-C SNPs in FH/M-ve cases show that the polygenic etiology of hypercholesterolemia in FCH and FH are similar in some patients.

All of these findings have shown that genetic definition of a disease, including monogenic diseases, cannot be too rigid, which might help clinicians make a better diagnosis, especially if it can be shown that these different etiologies predict a different prognosis or different response to treatment. My findings also support the idea that personalized medicine might one day be the standard of care for patients with hyperlipidemia, especially since the genetic etiology of a these diseases cannot be 'generalized' into a single cause. From an economic aspect, this would also encourage researching ways of making personalized medicine cost-effective.

All of my findings strongly support extending such investigative approaches to other monogenic diseases, where the causative gene does not explain the disease in some minority of cases. Finally, the findings from my studies support a new way of thinking and a different approach in unraveling the genetic etiology of monogenic and polygenic lipid disorder and perhaps other related diseases in the field.

## References


1. Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Benjamin, E. J., Berry, J. D., Borden, W. B., Bravata, D. M., Dai, S., Ford, E. S., Fox, C. S., Fullerton, H. J., Gillespie, C., Hailpern, S. M., Heit, J. A., Howard, V. J., Kissela, B. M., Kittner, S. J., Lackland, D. T., Lichtman, J. H., Lisabeth, L. D., Makuc, D. M., Marcus, G. M., Marelli, A., Matchar, D. B., Moy, C. S., Mozaffarian, D., Mussolino, M. E., Nichol, G., Paynter, N. P., Soliman, E. Z., Sorlie, P. D., Sotoodehnia, N., Turan, T. N., Virani, S. S., Wong, N. D., Woo, D., and Turner, M. B. (2012) Executive summary: heart disease and stroke statistics--2012 update: a report from the American Heart Association, *Circulation 125*, 188-197.
2. Jemal, A., Ward, E., Hao, Y., and Thun, M. (2005) Trends in the leading causes of death in the United States, 1970-2002, *JAMA 294*, 1255-1259.
3. Wilson, P. W., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., and Kannel, W. B. (1998) Prediction of coronary heart disease using risk factor categories, *Circulation 97*, 1837-1847.
4. Pencina, M. J., D'Agostino, R. B., Sr., Larson, M. G., Massaro, J. M., and Vasan, R. S. (2009) Predicting the 30-year risk of cardiovascular disease: the framingham heart study, *Circulation 119*, 3078-3084.
5. http://www.britannica.com.proxy2.lib.uwo.ca/EBchecked/topic/569347/stroke. (2013) stroke.
6. http://www.britannica.com.proxy2.lib.uwo.ca/EBchecked/topic/720793/cardiovascular-disease/33612/Hypertensive-heart-disease. (2013).
7. http://www.britannica.com.proxy2.lib.uwo.ca/EBchecked/topic/138261/coronary-heart-disease. (2013) coronary heart disease.
8. Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Benjamin, E. J., Berry, J. D., Borden, W. B., Bravata, D. M., Dai, S., Ford, E. S., Fox, C. S., Fullerton, H. J., Gillespie, C., Hailpern, S. M., Heit, J. A., Howard, V. J., Kissela, B. M., Kittner, S. J., Lackland, D. T., Lichtman, J. H., Lisabeth, L. D., Makuc, D. M., Marcus, G. M., Marelli, A., Matchar, D. B., Moy, C. S., Mozaffarian, D., Mussolino, M. E., Nichol, G., Paynter, N. P., Soliman, E. Z., Sorlie, P. D., Sotoodehnia, N., Turan, T. N., Virani, S. S., Wong, N. D., Woo, D., Turner, M. B., American Heart Association Statistics, C., and Stroke Statistics, S. (2012) Heart disease and stroke statistics--2012 update: a report from the American Heart Association, *Circulation 125*, e2-e220.
9. Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Borden, W. B., Bravata, D. M., Dai, S., Ford, E. S., Fox, C. S., Franco, S., Fullerton, H. J., Gillespie, C., Hailpern, S. M., Heit, J. A., Howard, V. J., Huffman, M. D., Kissela, B. M., Kittner, S. J., Lackland, D. T., Lichtman, J. H., Lisabeth, L. D., Magid, D., Marcus, G. M., Marelli, A., Matchar, D. B., McGuire, D. K., Mohler, E. R., Moy, C. S., Mussolino, M. E., Nichol, G., Paynter, N. P., Schreiner, P. J., Sorlie, P. D., Stein, J., Turan, T. N., Virani, S. S., Wong, N. D., Woo, D., and Turner, M. B. (2013) Executive summary: heart disease and stroke statistics--

2013 update: a report from the American Heart Association, *Circulation 127*, 143-152.

10. http://www.britannica.com.proxy2.lib.uwo.ca/EBchecked/topic/114106/cholesterol. (2013) cholesterol.
11. http://www.britannica.com.proxy2.lib.uwo.ca/EBchecked/topic/565825/steroid. (2013) steroid.
12. http://www.britannica.com.proxy2.lib.uwo.ca/EBchecked/topic/342808/lipid. (2013) lipid.
13. Phan, J., Pesaran, T., Davis, R. C., and Reue, K. (2002) The Diet1 locus confers protection against hypercholesterolemia through enhanced bile acid metabolism, *J Biol Chem 277*, 469-477.
14. http://www.britannica.com.proxy2.lib.uwo.ca/EBchecked/topic/342894/lipoprotein. (2013) lipoprotein.
15. Brown, M. S., and Goldstein, J. L. (1979) Receptor-mediated endocytosis: insights from the lipoprotein receptor system, *Proc Natl Acad Sci U S A 76*, 3330-3337.
16. Myant, N. B. (1973) Cholesterol metabolism, *J Clin Pathol Suppl (Assoc Clin Pathol) 5*, 1-4.
17. Rudel, L. L., Parks, J. S., Johnson, F. L., and Babiak, J. (1986) Low density lipoproteins in atherosclerosis, *J Lipid Res 27*, 465-474.
18. Goldstein, J. L., and Brown, M. S. (2009) The LDL receptor, *Arterioscler Thromb Vasc Biol 29*, 431-438.
19. Asztalos, B. F. (2004) High-density lipoprotein metabolism and progression of atherosclerosis: new insights from the HDL Atherosclerosis Treatment Study, *Curr Opin Cardiol 19*, 385-391.
20. Ng, D. S., Wong, N. C., and Hegele, R. A. (2013) HDL--is it too big to fail?, *Nat Rev Endocrinol 9*, 308-312.
21. http://www.britannica.com.proxy2.lib.uwo.ca/EBchecked/topic/36874/artery. (2013) artery.
22. Witztum, J. L., and Steinberg, D. (1991) Role of oxidized low density lipoprotein in atherogenesis, *J Clin Invest 88*, 1785-1792.
23. Bui, Q. T., Prempeh, M., and Wilensky, R. L. (2009) Atherosclerotic plaque development, *Int J Biochem Cell Biol 41*, 2109-2113.
24. http://www.britannica.com.proxy2.lib.uwo.ca/EBchecked/topic/40908/atherosclerosis. (2013) atherosclerosis.
25. Murtola, T., Vuorela, T. A., Hyvonen, M. T., Marrink, S.-J., Karttunen, M., and Vattulainen, I. (2011) Low density lipoprotein: structure, dynamics, and interactions of apoB-100 with lipids, *Soft Matter 7*, 8135-8141.
26. Yang, C. Y., Gu, Z. W., Weng, S. A., Kim, T. W., Chen, S. H., Pownall, H. J., Sharp, P. M., Liu, S. W., Li, W. H., Gotto, A. M., Jr., and et al. (1989) Structure of apolipoprotein B-100 of human low density lipoproteins, *Arteriosclerosis 9*, 96-108.

27. Innerarity, T. L., Mahley, R. W., Weisgraber, K. H., Bersot, T. P., Krauss, R. M., Vega, G. L., Grundy, S. M., Friedl, W., Davignon, J., and McCarthy, B. J. (1990) Familial defective apolipoprotein B-100: a mutation of apolipoprotein B that causes hypercholesterolemia, *J Lipid Res 31*, 1337-1349.
28. Gaffney, D., Reid, J. M., Cameron, I. M., Vass, K., Caslake, M. J., Shepherd, J., and Packard, C. J. (1995) INDEPENDENT MUTATIONS AT CODON-3500 OF THE APOLIPOPROTEIN-B GENE ARE ASSOCIATED WITH HYPERLIPIDEMIA, *Arteriosclerosis Thrombosis and Vascular Biology 15*, 1025-1029.
29. Lambert, G. (2007) Unravelling the functional significance of PCSK9, *Current opinion in lipidology 18*, 304-309.
30. Zelcer, N., Hong, C., Boyadjian, R., and Tontonoz, P. (2009) LXR regulates cholesterol uptake through Idol-dependent ubiquitination of the LDL receptor, *Science 325*, 100-104.
31. Gaffney, D., Reid, J. M., Cameron, I. M., Vass, K., Caslake, M. J., Shepherd, J., and Packard, C. J. (1995) Independent mutations at codon 3500 of the apolipoprotein B gene are associated with hyperlipidemia, *Arterioscler Thromb Vasc Biol 15*, 1025-1029.
32. Elmehdawi, R. (2008) Hypolipidemia: a word of caution, *Libyan J Med 3*, 84-90.
33. Fredrickson, D. S., and Lees, R. S. (1965) A System for Phenotyping Hyperlipoproteinemia, *Circulation 31*, 321-327.
34. Beaumont, J. L., Carlson, L. A., Cooper, G. R., Fejfar, Z., Fredrickson, D. S., and Strasser, T. (1970) Classification of hyperlipidaemias and hyperlipoproteinaemias, *Bull World Health Organ 43*, 891-915.
35. Brahm, A., and Hegele, R. A. (2013) Hypertriglyceridemia, *Nutrients 5*, 981-1001.
36. O'Sullivan, B. P., and Freedman, S. D. (2009) Cystic fibrosis, *Lancet 373*, 1891-1904.
37. Blau, N., van Spronsen, F. J., and Levy, H. L. (2010) Phenylketonuria, *Lancet 376*, 1417-1427.
38. Ratjen, F., and Doring, G. (2003) Cystic fibrosis, *Lancet 361*, 681-689.
39. Brown, M. S., and Goldstein, J. L. (1986) A receptor-mediated pathway for cholesterol homeostasis, *Science 232*, 34-47.
40. Nahler, G. (2009) Dictionary of Pharmaceutical Medicine, In *Dictionary of Pharmaceutical Medicine*, Springer-Verlag Vienna, Vienna.
41. Gibson, G. (2011) Rare and common variants: twenty arguments, *Nat Rev Genet 13*, 135-145.
42. Cirulli, E. T., and Goldstein, D. B. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing, *Nat Rev Genet 11*, 415-425.
43. 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., Cho, Y. S., Go, M. J., Kim, Y. J., Lee, J.-Y., Park, T., Kim, K., Sim, X., Ong, R. T.-H., Croteau-Chonka, D. C., Lange, L. A., Smith, J. D., Song, K., Zhao, J. H., Yuan,

X., Luan, J. a., Lamina, C., Ziegler, A., Zhang, W., Zee, R. Y. L., Wright, A. F., Witteman, J. C. M., 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. G., 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. J. H., 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., McArdle, W., Masson, D., Martin, N. G., Marroni, F., Mangino, M., Magnusson, P. K. E., Lucas, G., Luben, R., Loos, R. J. F., Lokki, M.-L., Lettre, G., Langenberg, C., Launer, L. J., Lakatta, E. G., Laaksonen, R., Kyvik, K. O., Kronenberg, F., Koenig, I. R., Khaw, K.-T., Kaprio, J., Kaplan, L. M., Johansson, A., Jarvelin, M.-R., Janssens, A. C. J. W., Ingelsson, E., Igi, W., Hovingh, G. K., Hottenga, J.-J., Hofman, A., Hicks, A. A., Hengstenberg, C., Heid, I. M., Hayward, C., Havulinna, A. S., Hastie, N. D., Harris, T. B., Haritunians, T., Hall, A. S., Gyllensten, U., Guiducci, C., Groop, L. C., Gonzalez, E., Gieger, C., Freimer, N. B., Ferrucci, L., Erdmann, J., Elliott, P., Ejebe, K. G., Doering, A., Dominiczak, A. F., Demissie, S., Deloukas, P., de Geus, E. J. C., de Faire, U., Crawford, G., Collins, F. S., Chen, Y.-d. I., Caulfield, M. J., Campbell, H., Burtt, N. P., Bonnycastle, L. L., Boomsma, D. I., Boekholdt, S. M., Bergman, R. N., Barroso, I., Bandinelli, S., Ballantyne, C. M., Assimes, T. L., Quertermous, T., Altshuler, D., Seielstad, M., Wong, T. Y., Tai, E. S., Feranil, A. B., Kuzawa, C. W., Adair, L. S., Taylor, H. A., Jr., Borecki, I. B., Gabriel, S. B., Wilson, J. G., Holm, H., Thorsteinsdottir, U., Gudnason, V., Krauss, R. M., Mohlke, K. L., Ordovas, J. M., Munroe, P. B., Kooner, J. S., Tall, A. R., Hegele, R. A., Kastelein, J. J. P., Schadt, E. E., Rotter, J. I., Boerwinkle, E., Strachan, D. P., Mooser, V., Stefansson, K., Reilly, M. P., Samani, N. J., Schunkert, H., Cupples, L. A., Sandhu, M. S., Ridker, P. M., Rader, D. J., van Duijn, C. M., Peltonen, L., Abecasis, G. R., Boehnke, M., and Kathiresan, S. (2010) Biological, clinical and population relevance of 95 loci for blood lipids, *Nature 466*, 707-713.

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

45. Nakagawa, S., and Cuthill, I. C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists, *Biol Rev Camb Philos Soc 82*, 591-605.

46. Zareparsi, S., Branham, K. E., Li, M., Shah, S., Klein, R. J., Ott, J., Hoh, J., Abecasis, G. R., and Swaroop, A. (2005) Strong association of the Y402H variant

in complement factor H at 1q32 with susceptibility to age-related macular degeneration, *Am J Hum Genet 77*, 149-153.
47. Rahalkar, A. R., and Hegele, R. A. (2008) Monogenic pediatric dyslipidemias: classification, genetics and clinical spectrum, *Mol Genet Metab 93*, 282-294.
48. 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 381*, 1293-1301.
49. Garg, A., and Simha, V. (2007) Update on dyslipidemia, *J Clin Endocrinol Metab 92*, 1581-1589.
50. Sullivan, D. (2007) Guidelines for the diagnosis and management of familial hypercholesterolaemia, *Heart Lung Circ 16*, 25-27.
51. Jeon, H., and Blacklow, S. C. (2005) Structure and physiologic function of the low-density lipoprotein receptor, *Annu Rev Biochem 74*, 535-562.
52. Cardin, A. D., Price, C. A., Hirose, N., Krivanek, M. A., Blankenship, D. T., Chao, J., and Mao, S. J. (1986) Structural organization of apolipoprotein B-100 of human plasma low density lipoproteins. Comparison to B-48 of chylomicrons and very low density lipoproteins, *J Biol Chem 261*, 16744-16748.
53. Segrest, J. P., Jones, M. K., De Loof, H., and Dashti, N. (2001) Structure of apolipoprotein B-100 in low density lipoproteins, *J Lipid Res 42*, 1346-1367.
54. Seidah, N. G., Benjannet, S., Wickham, L., Marcinkiewicz, J., Jasmin, S. B., Stifani, S., Basak, A., Prat, A., and Chretien, M. (2003) The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): liver regeneration and neuronal differentiation, *Proc Natl Acad Sci U S A 100*, 928-933.
55. 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*, 154-156.
56. Kwon, H. J., Lagace, T. A., McNutt, M. C., Horton, J. D., and Deisenhofer, J. (2008) Molecular basis for LDL receptor recognition by PCSK9, *Proc Natl Acad Sci U S A 105*, 1820-1825.
57. Konrad, R. J., Troutt, J. S., and Cao, G. (2011) Effects of currently prescribed LDL-C-lowering drugs on PCSK9 and implications for the next generation of LDL-C-lowering agents, *Lipids Health Dis 10*, 38.
58. 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*, 1394-1398.
59. 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*, 34023-34031.

60. Marduel, M., Ouguerram, K., Serre, V., Bonnefont-Rousselot, D., Marques-Pinheiro, A., Erik Berge, K., Devillers, M., Luc, G., Lecerf, J. M., Tosolini, L., Erlich, D., Peloso, G. M., Stitziel, N., Nitchke, P., Jais, J. P., Abifadel, M., Kathiresan, S., Leren, T. P., Rabes, J. P., Boileau, C., and Varret, M. (2013) Description of a large family with autosomal dominant hypercholesterolemia associated with the APOE p.Leu167del mutation, *Hum Mutat 34*, 83-87.
61. Pahan, K. (2006) Lipid-lowering drugs, *Cell Mol Life Sci 63*, 1165-1178.
62. Couture, P., and Lamarche, B. (2013) Ezetimibe and bile acid sequestrants: impact on lipoprotein metabolism and beyond, *Curr Opin Lipidol 24*, 227-232.
63. 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.
64. Zhang, L., Reue, K., Fong, L. G., Young, S. G., and Tontonoz, P. (2012) Feedback regulation of cholesterol uptake by the LXR-IDOL-LDLR axis, *Arterioscler Thromb Vasc Biol 32*, 2541-2546.
65. Gragnoli, C. (2011) Hypercholesterolemia and a candidate gene within the 12q24 locus, *Cardiovasc Diabetol 10*, 38.
66. Brouwers, M. C., van Greevenbroek, M. M., Stehouwer, C. D., de Graaf, J., and Stalenhoef, A. F. (2012) The genetics of familial combined hyperlipidaemia, *Nature reviews.Endocrinology*.
67. Hopkins, P. N., Heiss, G., Ellison, R. C., Province, M. A., Pankow, J. S., Eckfeldt, J. H., and Hunt, S. C. (2003) Coronary artery disease risk in familial combined hyperlipidemia and familial hypertriglyceridemia: a case-control comparison from the National Heart, Lung, and Blood Institute Family Heart Study, *Circulation 108*, 519-523.
68. 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*, 1544-1568.
69. Brouwers, M. C., van Greevenbroek, M. M., Stehouwer, C. D., de Graaf, J., and Stalenhoef, A. F. (2012) The genetics of familial combined hyperlipidaemia, *Nature reviews.Endocrinology 8*, 352-362.
70. 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*, 1916-1926.
71. Johansen, C. T., Wang, J., Lanktree, M. B., Cao, H., McIntyre, A. D., Ban, M. R., Martins, R. A., Kennedy, B. A., Hassell, R. G., Visser, M. E., 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., and Hegele, R. A. (2010) Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia, *Nat Genet 42*, 684-687.

72. Mouzeyan, A., Choi, J., Allayee, H., Wang, X., Sinsheimer, J., Phan, J., Castellani, L. W., Reue, K., Lusis, A. J., and Davis, R. C. (2000) A locus conferring resistance to diet-induced hypercholesterolemia and atherosclerosis on mouse chromosome 2, *J Lipid Res 41*, 573-582.
73. Vergnes, L., Lee, J. M., Chin, R. G., Auwerx, J., and Reue, K. (2013) Diet1 Functions in the FGF15/19 Enterohepatic Signaling Axis to Modulate Bile Acid and Lipid Levels, *Cell Metab 17*, 916-928.
74. Dawson, P. A. (2013) Novel Regulator of Enterohepatic Bile Acid Signaling Protects against Hypercholesterolemia, *Cell Metab 17*, 816-818.
75. Johansen, C. T., Wang, J., McIntyre, A. D., Martins, R. A., Ban, M. R., Lanktree, M. B., Huff, M. W., Peterfy, M., Mehrabian, M., Lusis, A. J., Kathiresan, S., Anand, S. S., Yusuf, S., Lee, A. H., Glimcher, L. H., Cao, H., and Hegele, R. A. (2012) Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia, *Circ Cardiovasc Genet 5*, 66-72.
76. Pulst, S. M. (1999) GEnetic linkage analysis, *Archives of Neurology 56*, 667-672.
77. Strachan, T., and Read, A. P. (1999) Human Molecular Genetics, 2nd Edition, 2010/12/07 ed.
78. Altshuler, D., Daly, M. J., and Lander, E. S. (2008) Genetic mapping in human disease, *Science 322*, 881-888.
79. Bush, W. S., and Moore, J. H. (2012) Chapter 11: Genome-wide association studies, *PLoS Comput Biol 8*, e1002822.
80. Tamari, M., Tanaka, S., and Hirota, T. (2013) Genome-wide association studies of allergic diseases, *Allergol Int 62*, 21-28.
81. Ng, P. C., and Henikoff, S. (2001) Predicting deleterious amino acid substitutions, *Genome Res 11*, 863-874.
82. Ng, P. C., and Henikoff, S. (2003) SIFT: Predicting amino acid changes that affect protein function, *Nucleic Acids Res 31*, 3812-3814.
83. 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*.
84. 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*, 248-249.
85. Kumar, P., Henikoff, S., and Ng, P. C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm, *Nature Protocols 4*, 1073-1082.
86. Do, C. B., and Katoh, K. (2008) Protein multiple sequence alignment, *Methods Mol Biol 484*, 379-413.
87. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W. Z., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Molecular Systems Biology 7*.
88. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench, *Bioinformatics 25*, 1189-1191.

89. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., and Sham, P. C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses, *American Journal of Human Genetics 81*, 559-575.
90. Anand, S. S., Yusuf, S., Vuksan, V., Devanesen, S., Teo, K. K., Montague, P. A., Kelemen, L., Yi, C., Lonn, E., Gerstein, H., Hegele, R. A., and McQueen, M. (2000) Differences in risk factors, atherosclerosis, and cardiovascular disease between ethnic groups in Canada: the Study of Health Assessment and Risk in Ethnic groups (SHARE), *Lancet 356*, 279-284.
91. Kathiresan, S., Voight, B. F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P. M., Anand, S., Engert, J. C., Samani, N. J., Schunkert, H., Erdmann, J., Reilly, M. P., Rader, D. J., Morgan, T., Spertus, J. A., Stoll, M., Girelli, D., McKeown, P. P., Patterson, C. C., Siscovick, D. S., O'Donnell, C. J., Elosua, R., Peltonen, L., Salomaa, V., Schwartz, S. M., Melander, O., Altshuler, D., Merlini, P. A., Berzuini, C., Bernardinelli, L., Peyvandi, F., Tubaro, M., Celli, P., Ferrario, M., Fetiveau, R., Marziliano, N., Casari, G., Galli, M., Ribichini, F., Rossi, M., Bernardi, F., Zonzin, P., Piazza, A., Yee, J., Friedlander, Y., Marrugat, J., Lucas, G., Subirana, I., Sala, J., Ramos, R., Meigs, J. B., Williams, G., Nathan, D. M., MacRae, C. A., Havulinna, A. S., Berglund, G., Hirschhorn, J. N., Asselta, R., Duga, S., Spreafico, M., Daly, M. J., Nemesh, J., Korn, J. M., McCarroll, S. A., Surti, A., Guiducci, C., Gianniny, L., Mirel, D., Parkin, M., Burtt, N., Gabriel, S. B., Thompson, J. R., Braund, P. S., Wright, B. J., Balmforth, A. J., Ball, S. G., Hall, A., Linsel-Nitschke, P., Lieb, W., Ziegler, A., Konig, I., Hengstenberg, C., Fischer, M., Stark, K., Grosshennig, A., Preuss, M., Wichmann, H. E., Schreiber, S., Ouwehand, W., Deloukas, P., Scholz, M., Cambien, F., Li, M., Chen, Z., Wilensky, R., Matthai, W., Qasim, A., Hakonarson, H. H., Devaney, J., Burnett, M. S., Pichard, A. D., Kent, K. M., Satler, L., Lindsay, J. M., Waksman, R., Knouff, C. W., Waterworth, D. M., Walker, M. C., Mooser, V., Epstein, S. E., Scheffold, T., Berger, K., Huge, A., Martinelli, N., Olivieri, O., Corrocher, R., McKeown, P., Erdmann, E., Konig, I. R., Holm, H., Thorleifsson, G., Thorsteinsdottir, U., Stefansson, K., Do, R., Xie, C., and Siscovick, D. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants, *Nat Genet 41*, 334-341.
92. 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*, 644-649.
93. Rozen, S., and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers, *Methods Mol Biol 132*, 365-386.
94. Durinck, S., Bullard, J., Spellman, P. T., and Dudoit, S. (2009) GenomeGraphs: integrated genomic data visualization with R, *BMC Bioinformatics 10*, 2.
95. Team}, R. D. C. (2009) *{R: A language and environment for statistical computing}*, R Foundation for Statistical Computing, Vienna, Austria.
96. Birch, D. E. (1996) Simplified hot start PCR, *Nature 381*, 445-446.
97. Sanger, F., Nicklen, S., and Coulson, A. R. (1992) DNA sequencing with chain-terminating inhibitors. 1977, *Biotechnology 24*, 104-108.

98. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc Natl Acad Sci U S A 74*, 5463-5467.
99. Dupont, W. D., and Plummer, W. D., Jr. (1990) Power and sample size calculations. A review and computer program, *Control Clin Trials 11*, 116-128.
100. 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*, 1300-1311.
101. Bewick, V., Cheek, L., and Ball, J. (2005) Statistics review 14: Logistic regression, *Crit Care 9*, 112-118.
102. Lee, J. H., Tate, C. M., You, J. S., and Skalnik, D. G. (2007) Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex, *J Biol Chem 282*, 13419-13428.
103. Goldstein, J. L., Schrott, H. G., Hazzard, W. R., Bierman, E. L., and Motulsky, A. G. (1973) Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia, *J Clin Invest 52*, 1544-1568.
104. Hokanson, J. E., Austin, M. A., Zambon, A., and Brunzell, J. D. (1993) Plasma triglyceride and LDL heterogeneity in familial combined hyperlipidemia, *Arterioscler Thromb 13*, 427-434.
105. Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F., and Petersen, G. B. (1982) Nucleotide sequence of bacteriophage lambda DNA, *J Mol Biol 162*, 729-773.
106. Loux, N., Saint-Jore, B., Collod, G., Dairou, F., Benlian, P., Truffert, J., Dastugue, B., Douste-Blazy, P., de Gennes, J. L., Junien, C., and et al. (1992) Screening for new mutations in the LDL receptor gene in seven French familial hypercholesterolemia families by the single strand conformation polymorphism method, *Hum Mutat 1*, 325-332.
107. Brusgaard, K., Jordan, P., Hansen, H., Hansen, A. B., and Horder, M. (2006) Molecular genetic analysis of 1053 Danish individuals with clinical signs of familial hypercholesterolemia, *Clin Genet 69*, 277-283.
108. Graham, C. A., Wright, W. T., McIlhatton, B. P., Young, I. S., and Nicholls, D. P. (2006) The LDLR variant T705I does not cause the typical phenotype of familial hypercholesterolaemia, *Atherosclerosis 188*, 218-219.
109. Alharbi, K. K., Aldahmesh, M. A., Spanakis, E., Haddad, L., Whittall, R. A., Chen, X. H., Rassoulian, H., Smith, M. J., Sillibourne, J., Ball, N. J., Graham, N. J., Briggs, P. J., Simpson, I. A., Phillips, D. I., Lawlor, D. A., Ye, S., Humphries, S. E., Cooper, C., Smith, G. D., Ebrahim, S., Eccles, D. M., and Day, I. N. (2005) Mutation scanning by meltMADGE: validations using BRCA1 and LDLR, and demonstration of the potential to identify severe, moderate, silent, rare, and paucimorphic mutations in the general population, *Genome Res 15*, 967-977.
110. Leigh, S. E., Foster, A. H., Whittall, R. A., Hubbart, C. S., and Humphries, S. E. (2008) Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database, *Ann Hum Genet 72*, 485-498.
111. Amsellem, S., Briffaut, D., Carrie, A., Rabes, J. P., Girardet, J. P., Fredenrich, A., Moulin, P., Krempf, M., Reznik, Y., Vialettes, B., de Gennes, J. L., Brukert, E.,

and Benlian, P. (2002) Intronic mutations outside of Alu-repeat-rich domains of the LDL receptor gene are a cause of familial hypercholesterolemia, *Hum Genet 111*, 501-510.

112. Hobbs, H. H., Russell, D. W., Brown, M. S., and Goldstein, J. L. (1990) The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein, *Annu Rev Genet 24*, 133-170.
113. Romano, M., Di Taranto, M. D., Mirabelli, P., D'Agostino, M. N., Iannuzzi, A., Marotta, G., Gentile, M., Raia, M., Di Noto, R., Del Vecchio, L., Rubba, P., and Fortunato, G. (2011) An improved method on stimulated T-lymphocytes to functionally characterize novel and known LDLR mutations, *J Lipid Res 52*, 2095-2100.
114. Fouchier, S. W., Kastelein, J. J., and Defesche, J. C. (2005) Update of the molecular basis of familial hypercholesterolemia in The Netherlands, *Hum Mutat 26*, 550-556.
115. Hobbs, H. H., Brown, M. S., and Goldstein, J. L. (1992) Molecular genetics of the LDL receptor gene in familial hypercholesterolemia, *Human mutation 1*, 445-466.
116. Liljedahl, U., Fredriksson, M., Dahlgren, A., and Syvanen, A. C. (2004) Detecting imbalanced expression of SNP alleles by minisequencing on microarrays, *BMC Biotechnol 4*, 24.
117. Andreotti, G., Chen, J., Gao, Y. T., Rashid, A., Chen, B. E., Rosenberg, P., Sakoda, L. C., Deng, J., Shen, M. C., Wang, B. S., Han, T. Q., Zhang, B. H., Yeager, M., Welch, R., Chanock, S., Fraumeni, J. F., Jr., and Hsing, A. W. (2008) Polymorphisms of genes in the lipid metabolism pathway and risk of biliary tract cancers and stones: a population-based case-control study in Shanghai, China, *Cancer Epidemiol Biomarkers Prev 17*, 525-534.
118. Wang, J., Huff, E., Janecka, L., and Hegele, R. A. (2001) Low density lipoprotein receptor (LDLR) gene mutations in Canadian subjects with familial hypercholesterolemia, but not of French descent, *Hum Mutat 18*, 359.
119. Lombardi, M. P., Redeker, E. J., Defesche, J. C., Kamerling, S. W., Trip, M. D., Mannens, M. M., Havekes, L. M., and Kastelein, J. J. (2000) Molecular genetic testing for familial hypercholesterolemia: spectrum of LDL receptor gene mutations in The Netherlands, *Clin Genet 57*, 116-124.
120. Zakharova, F. M., Damgaard, D., Mandelshtam, M. Y., Golubkov, V. I., Nissen, P. H., Nilsen, G. G., Stenderup, A., Lipovetsky, B. M., Konstantinov, V. O., Denisenko, A. D., Vasilyev, V. B., and Faergeman, O. (2005) Familial hypercholesterolemia in St-Petersburg: the known and novel mutations found in the low density lipoprotein receptor gene in Russia, *BMC Med Genet 6*, 6.
121. Laurie, A. D., and George, P. M. (2009) Evaluation of high-resolution melting analysis for screening the LDL receptor gene, *Clin Biochem 42*, 528-535.
122. Huijgen, R., Kindt, I., Fouchier, S. W., Defesche, J. C., Hutten, B. A., Kastelein, J. J., and Vissers, M. N. (2010) Functionality of sequence variants in the genes coding for the low-density lipoprotein receptor and apolipoprotein B in individuals with inherited hypercholesterolemia, *Hum Mutat 31*, 752-760.
123. Defesche, J. C., Schuurman, E. J., Klaaijsen, L. N., Khoo, K. L., Wiegman, A., and Stalenhoef, A. F. (2008) Silent exonic mutations in the low-density

lipoprotein receptor gene that cause familial hypercholesterolemia by affecting mRNA splicing, *Clin Genet 73*, 573-578.
124. Fouchier, S. W., Defesche, J. C., Umans-Eckenhausen, M. W., and Kastelein, J. P. (2001) The molecular basis of familial hypercholesterolemia in The Netherlands, *Hum Genet 109*, 602-615.
125. Lombardi, P., Sijbrands, E. J., Kamerling, S., Leuven, J. A., and Havekes, L. M. (1997) The T705I mutation of the low density lipoprotein receptor gene (FH Paris-9) does not cause familial hypercholesterolemia, *Hum Genet 99*, 106-107.
126. Jensen, H. K., Jensen, L. G., Hansen, P. S., Petersen, L. S., Gerdes, L. U., Bolund, L., Faergeman, O., and Gregersen, N. (1994) An alanine29-serine variant in exon 2 of the low density lipoprotein receptor gene: no association with hypercholesterolemia, *Clin Genet 46*, 214-215.
127. Weiss, N., Binder, G., and Keller, C. (1998) Heterozygosity for the missense mutation Ala370-->Thr in exon 8 of the low density lipoprotein receptor gene does not cause hypercholesterolemia, *Eur J Med Res 3*, 20-24.
128. Huijgen, R., Kindt, I., Defesche, J. C., and Kastelein, J. J. (2012) Cardiovascular risk in relation to functionality of sequence variants in the gene coding for the low-density lipoprotein receptor: a study among 29,365 individuals tested for 64 specific low-density lipoprotein-receptor sequence variants, *Eur Heart J 33*, 2325-2330.
129. Sun, X. M., Patel, D. D., Knight, B. L., and Soutar, A. K. (1998) Influence of genotype at the low density lipoprotein (LDL) receptor gene locus on the clinical phenotype and response to lipid-lowering drug therapy in heterozygous familial hypercholesterolaemia. The Familial Hypercholesterolaemia Regression Study Group, *Atherosclerosis 136*, 175-185.
130. Sun, X. M., Patel, D. D., Knight, B. L., and Soutar, A. K. (1997) Comparison of the genetic defect with LDL-receptor activity in cultured cells from patients with a clinical diagnosis of heterozygous familial hypercholesterolemia. The Familial Hypercholesterolaemia Regression Study Group, *Arterioscler Thromb Vasc Biol 17*, 3092-3101.
131. Pullinger, C. R., Love, J. A., Liu, W., Hennessy, L. K., Ghassemzadeh, M., Newcomb, K. C., Illingworth, D. R., and Kane, J. P. (1996) The apolipoprotein B Q3405E polymorphism has no effect on its low-density-lipoprotein receptor binding affinity, *Hum Genet 98*, 678-680.
132. Gaffney, D., Hoffs, M. S., Cameron, I. M., Stewart, G., O'Reilly, D. S., and Packard, C. J. (1998) Influence of polymorphism Q3405E and mutation A3371V in the apolipoprotein B gene on LDL receptor binding, *Atherosclerosis 137*, 167-174.
133. Akram, O. N., Bernier, A., Petrides, F., Wong, G., and Lambert, G. (2010) Beyond LDL cholesterol, a new role for PCSK9, *Arterioscler Thromb Vasc Biol 30*, 1279-1281.
134. Mukhopadhyay, A., Nikopoulos, K., Maugeri, A., de Brouwer, A. P., van Nouhuys, C. E., Boon, C. J., Perveen, R., Zegers, H. A., Wittebol-Post, D., van den Biesen, P. R., van der Velde-Visser, S. D., Brunner, H. G., Black, G. C., Hoyng, C. B., and Cremers, F. P. (2006) Erosive vitreoretinopathy and wagner

disease are caused by intronic mutations in CSPG2/Versican that result in an imbalance of splice variants, *Invest Ophthalmol Vis Sci 47*, 3565-3572.

135. Pennisi, E. (2012) Genomics. ENCODE project writes eulogy for junk DNA, *Science 337*, 1159, 1161.
136. Li, M. X., Kwan, J. S., Bao, S. Y., Yang, W., Ho, S. L., Song, Y. Q., and Sham, P. C. (2013) Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies, *PLoS Genet 9*, e1003143.
137. Musunuru, K., Pirruccello, J. P., Do, R., Peloso, G. M., Guiducci, C., Sougnez, C., Garimella, K. V., Fisher, S., Abreu, J., Barry, A. J., Fennell, T., Banks, E., Ambrogio, L., Cibulskis, K., Kernytsky, A., Gonzalez, E., Rudzicz, N., Engert, J. C., DePristo, M. A., Daly, M. J., Cohen, J. C., Hobbs, H. H., Altshuler, D., Schonfeld, G., Gabriel, S. B., Yue, P., and Kathiresan, S. (2010) Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia, *N Engl J Med 363*, 2220-2227.
138. Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., Shendure, J., and Bamshad, M. J. (2010) Exome sequencing identifies the cause of a mendelian disorder, *Nat Genet 42*, 30-35.
139. Tchernitchko, D., Goossens, M., and Wajcman, H. (2004) In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics, *Clin Chem 50*, 1974-1978.
140. Civeira, F., Jarauta, E., Cenarro, A., Garcia-Otin, A. L., Tejedor, D., Zambon, D., Mallen, M., Ros, E., and Pocovi, M. (2008) Frequency of low-density lipoprotein receptor gene mutations in patients with a clinical diagnosis of familial combined hyperlipidemia in a clinical setting, *Journal of the American College of Cardiology 52*, 1546-1553.

# Appendix

**Table 1. List and description of all variants identified in all 18 exons and promoter of *LDLR* and the minor allele counts in cases and controls**

| Variant | SNP | MAF | Cases (n=138) | Controls (n=94) |
|---|---|---|---|---|
| p.G-2R | rs147509697 | 0.002165 | 1 | 0 |
| p.C6C | rs2228671 | 0.09307 | 29 | 14 |
| p.T41M | NR | 0.002155 | 0 | 1 |
| c.190,+56G/A | rs3745677 | 0.07576 | 22 | 13 |
| c.313,+69C/T | rs56084625 | 0.01082 | 2 | 3 |
| c.314,-50T/C | rs10423288 | 0.002155 | 0 | 1 |
| c.940,+16G/A | rs72658859 | 0.002155 | 0 | 1 |
| c.940,+36G/A | rs13306513 | 0.02851 | 8 | 5 |
| c.941,-39C/T | rs55792959 | 0.01078 | 2 | 3 |
| p.G314S | NR | 0.002165 | 1 | 0 |
| c.1060,+7T/C | rs2738442 | 0.002155 | 0 | 1 |
| c.1060,+10G/C | rs12710260 | 0.4286 | 113 | 85 |
| c.1060,+49C/T | NR | 0.002155 | 0 | 1 |
| c.1060,+59A/C | rs55642005 | 0.002155 | 0 | 1 |
| c.1061,-82G/C | rs41301947 | 0.002155 | 0 | 1 |
| c.1061-8T/C | rs72658861 | 0.002174 | 1 | 0 |
| p.D333V | NR | 0.002174 | 1 | 0 |
| p.C347C | rs113669610 | 0.002174 | 1 | 0 |
| p.A370T | rs11669576 | 0.05 | 12 | 11 |
| c.1373,+29C/A | NR | 0.002174 | 1 | 0 |
| p.N407N | NR | 0.00431 | 0 | 2 |
| c.1359,-54C/T | rs6413505 | 0.00431 | 1 | 1 |
| c.1359,-30C/T | rs1003723 | 0.4353 | 117 | 85 |
| p.R450R | rs5930 | 0.4095 | 165 | 109 |
| c.1774,-87G/A | NR | 0.002155 | 0 | 1 |
| p.P518P | rs5929 | 0.05435 | 20 | 5 |
| c.1705,+56C/T | rs4508523 | 0.1354 | 40 | 22 |
| c.1706,-81C/T | rs41307025 | 0.006466 | 1 | 2 |
| c.1706,-69G/T | rs7259278 | 0.1358 | 41 | 22 |
| c.1706,-55A/C | rs2738447 | 0.4203 | 162 | 107 |
| c.1706,-10G/A | rs17248882 | 0.006466 | 2 | 1 |
| p.L554L | rs1799898 | 0.1293 | 35 | 25 |
| p.L561P | NR | 0.002155 | 1 | 0 |
| p.N570N | rs688 | 0.4353 | 118 | 84 |

| | | | | |
|---|---|---|---|---|
| p.T576T | NR | 0.002155 | 1 | 0 |
| p.A585S | rs72658865 | 0.002155 | 0 | 1 |
| c.1846,-78C/G | rs116959285 | 0.04762 | 13 | 9 |
| p.N619N | rs5926 | 0.00431 | 0 | 2 |
| p.V632V | rs5925 | 0.4351 | 118 | 83 |
| p.C677G | NR | 0.002183 | 1 | 0 |
| c.2140,+5G/A | rs72658867 | 0.00655 | 1 | 2 |
| p.T705I | rs45508991 | 0.006466 | 1 | 2 |
| p.R723R | rs5927 | 0.263 | 203 | 136 |
| c.2312,-71G/A | rs17249358 | 0.006466 | 1 | 2 |
| c.2312,-47G/A | rs41306974 | 0.03097 | 6 | 8 |
| c.2312,-28G/A | NR | 0.002155 | 0 | 1 |
| c.2389+41C/A | rs72658868 | 0.006637 | 1 | 2 |
| c.2389+46C/T | rs2738460 | 0.2633 | 76 | 43 |
| c.2389+47G/A | rs13306501 | 0.03319 | 8 | 7 |
| c.2389,+51C/T | rs145293532 | 0.002165 | 1 | 0 |
| p.V806I | rs137853964 | 0.002165 | 1 | 0 |
| c.2548,-53G/A | rs6413503 | 0.006466 | 1 | 2 |
| c.2548,-42A/G | rs6413504 | 0.4697 | 130 | 87 |
| 3'UT+19G/A | rs56270417 | 0.004329 | 2 | 0 |
| 3'UT+52G/A | rs14158 | 0.2294 | 62 | 44 |

**Table 2. List and description of all variants identified in exons 26 and 29 *APOB* in and minor allele count in cases and controls**

| Variant | SNP identifier | MAF | Cases (n=138) | Controls (n=94) |
|---|---|---|---|---|
| p.T4533 | rs72654427 | 0.002155 | 1 | 0 |
| p.T4457M | rs12713450 | 0.006466 | 3 | 0 |
| p.A4454T | NR | 0.04095 | 8 | 11 |
| p.S4403T | rs72654426 | 0.002155 | 0 | 1 |
| p.S4311N | rs1042034 | 0.194 | 44 | 46 |
| p.M4293V | NR | 0.002155 | 1 | 0 |
| p.I4287V | rs72654423 | 0.01293 | 5 | 1 |
| p.R4243T | rs1801702 | 0.0194 | 7 | 2 |
| p.V4238A | NR | 0.01078 | 2 | 3 |
| p.E4154K | rs1042031 | 0.1724 | 46 | 34 |
| p.V4101M | rs1801703 | 0.002155 | 0 | 1 |
| p.Y4089 | NR | 0.002155 | 1 | 0 |
| c.11788+150C>T | rs12713523 | 0.002155 | 0 | 1 |
| p.T3799M | NR | 0.002155 | 1 | 0 |
| p.R3611Q | rs1801701 | 0.06034 | 15 | 13 |
| p.T3540 | rs12713558 | 0.002155 | 1 | 0 |
| p.R3500W | Reported | 0.002155 | 1 | 0 |
| p.Q3405E | rs1042023 | 0.00431 | 0 | 2 |
| p.L3350 | rs1799812 | 0.002155 | 0 | 1 |
| p. S3267P | rs12720855 | 0.002155 | 0 | 1 |
| p.S3252G | rs12720854 | 0.00431 | 2 | 0 |
| p.Y3071 | NR | 0.002155 | 0 | 1 |
| p.T3020R | rs61742323 | 0.002155 | 1 | 0 |
| p.N3008 | NR | 0.002155 | 0 | 1 |
| p.P2794L | rs72653095 | 0.01078 | 4 | 1 |
| p.P2712L | rs676210 | 0.1897 | 43 | 45 |
| p.I2689 | rs6413458 | 0.01724 | 5 | 3 |
| p.L2594 | NR | 0.002155 | 1 | 0 |
| p.E2539D | NR | 0.002155 | 0 | 1 |
| p.E2539K | rs1801696 | 0.002155 | 1 | 0 |
| p.L2511 | rs72653093 | 0.00431 | 1 | 1 |
| p.T2488 | rs693 | 0.4871 | 124 | 102 |
| p.M2331I | NR | 0.002155 | 1 | 0 |
| p.V2286I | rs584542 | 0.002155 | 1 | 0 |
| p.D2285 | NR | 0.4935 | 127 | 102 |

| | | | | |
|---|---|---|---|---|
| p.R2192C | NR | 0.002155 | 0 | 1 |
| p.H2040 | rs143222685 | 0.002155 | 0 | 1 |
| p.H1896R | rs533617 | 0.02586 | 5 | 7 |
| p.N1887S | rs1801699 | 0.01293 | 3 | 3 |
| p.P1875 | NR | 0.002155 | 1 | 0 |
| p.R1662H | NR | 0.002155 | 1 | 0 |
| p.K1615R | | 0.002165 | 0 | 1 |
| p.S1586T | rs61742247 | 0.002165 | 0 | 1 |
| p. F1428 | rs12720847 | 0.002155 | 1 | 0 |
| p.C1395Y | rs568413 | 0.002155 | 1 | 0 |

**Table 3. List and description of all variants identified in the 7 exons of *IDOL* in and minor allele count in cases and controls**

| Variant | SNP identifier | MAF | Cases (n=138) | Controls (n=94) |
|---|---|---|---|---|
| 5'UTR,-56G/T | rs3765234 | 0.04565 | 17 | 4 |
| c.87+60G/A | rs1076632 | 0.004348 | 1 | 1 |
| p.C31Y | New | 0.002174 | 0 | 1 |
| c.279,-31A/G | rs34444721 | 0.002174 | 1 | 0 |
| c.464,+80T/C | rs143656216 | 0.01957 | 1 | 8 |
| p.I202L | rs79992066 | 0.3978 | 92 | 185 |
| c.663,-33A/G | rs2072783 | 0.1413 | 38 | 27 |
| p.V339I | rs142124143 | 0.004348 | 1 | 1 |
| p.N342S | rs9370867 | 0.4696 | 121 | 95 |
| p.R372W | rs141183183 | 0.002174 | 1 | 0 |
| p.C391 | rs1060901 | 0.09565 | 19 | 25 |

**Table 4. Normality tests for LDL-C genetic risk score for (a) FH Mutation negative patients and (b) FH Mutation positive patients**

**a**

| Normality Test | p-value |
|---|---|
| Shapiro-Wilk | <0.0001 |
| Kolmogorov-Smirnov | 0.0112 |
| Cramer-von Mises | <0.0050 |
| Anderson-Darling | <0.0050 |

**b**

| Normality Test | p-value |
|---|---|
| Shapiro-Wilk | 0.0004 |
| Kolmogorov-Smirnov | <0.0100 |
| Cramer-von Mises | <0.0050 |
| Anderson-Darling | <0.0050 |

Normality tests was done on the LDL-C genetic risk score for Familial Hypercholesterolemia (FH) mutation negative patients (a) and FH mutation positive patients (b). In Normality tests, p-value less than 0.05 means that the null hypothesis (which states that the data are normally distributed) means that the data are not normally distributed). So, Normality tests showed that the LDL-C genetic risk scores were not normally distributed.

Western Research

Research Ethics

**Use of Human Participants - Ethics Approval Notice**

**Principal Investigator:** Dr. Robert Hegele
**File Number:**379
**Review Level:**Delegated
**Approved Local Adult Participants:**1840
**Approved Local Minor Participants:**0
**Protocol Title:**Candidate gene sequencing, genetic and genomic analysis for identification of new genetic determinants of intermediate traits of atherosclerosis, dyslipidemia, diabetes, obesity, hypertension, lipodystrophy and other rare metabolic or cardiovascular disorders in the human population. 07920E
**Department & Institution:**Science\Vascular Biology,Robarts Research Institute
**Sponsor:**Genome Canada
Heart and Stroke Foundation of Canada
Canadian Institutes of Health Research

**Ethics Approval Date:**March 28, 2013 **Expiry Date:**December 31, 2015
**Documents Reviewed & Approved & Documents Received for Information:**

| Document Name | Comments | Version Date |
|---|---|---|
| Other Regulatory Correspondence | | |

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

**Ethics Officer to Contact for Further Information**

| | | |
|---|---|---|
| Janice Sutherland (jsutherl@uwo.ca) | ✓ Grace Kelly (grace.kelly@uwo.ca) | Shantel Walcott (swalcott@uwo.ca) |

*This is an official document. Please retain the original in your files.*

# Curriculum Vitae

## Post-Secondary Education

### University of Western Ontario, London, Ontario, Canada

1. MSc Candidate in Biochemistry (2011-present)
2. Bachelor of Medical Science (BMSc) in Honors Specialization in Clinical Biochemistry (Western Scholars) "with Distinction" (2008-2011).

### Edexcel, London Examinations, United Kingdom

General Certificate of Education Advanced levels (GCE-A-level) (2006-2008)

## Academic Honours, Awards and Achievements

1. Finalist in University of Western Ontario's inaugural 3-Minute Thesis (3MT) competition ($500) (2012)
2. Graduated as 'Western Scholar' and 'with Distinction' in BMSc
3. Dean's Honour's list every year throughout undergraduate program (2008-2011)
4. Exempted from first year university by undergraduate program because of GCE A-level results

## Other Awards

1. Canadian Institutes of Health Research (CIHR) Strategic Training Program (STP) Fellowship (2011-2103)
2. The 'International Student Bursary'($600) (2010)
3. The inaugural 'Igal Holtzer and Family International Bursary' ($1,500)(2009)
4.

## Professional Development

### Certificates from School of Graduate and Postdoctoral Studies, University of Western Ontario, London, Ontario, Canada,

1. Western Certificate in University Teaching and Learning (2012)
2. Language of Teaching in Sciences (2012)
3. Teaching Assistant Training Program (2012)

# Presentations

## University of Western Ontario, London, Ontario, Canada

1. Lawson Health Research Day (LHRD) poster presentation (2013)
2. Department of Biochemistry Graduate Student Seminar (2103)
3. Department of Biochemistry Graduate Student Poster (2012)

# Conferences

1. Atherosclerosis, Thrombosis and Vascular Biology (ATVB) 2013 Scientific Sessions, Orlando, Florida (International Conference) (2 hour poster presentation) (2013)
2. National Atherosclerosis and Cardiometabolic Forum (2012) in Toronto, Ontario (attendance) (2012)
3. Canadian Lipoprotein Conference (CLC) Halifax, Nova Scotia (attendance) (2011)

# Work Experience

## University of Western Ontario, London, Ontario, Canada (January1st 2013-April 30th 2013)

Lecturer for Introductory Genetics, Department of Biology, Faculty of Science

Responsibilities:

i. Deliver 2-hour lectures in tutorials a week every week to a class of approximately 30 upper year students. Lecture material was examinable material not covered in main lecture.
ii. Mark and grade quizzes
iii. Answer student's questions pertaining to course material
iv. Minimum of 2 hours of review session before each exam
v. Attend monthly Teaching Assistant meetings

# Publication

### (Submitted Journal Articles)

Zuhier Awan , Hong Y. Choi, Nathan Stitziel , Isabelle Ruel , Mary Aderayo Bamimore , Regina Husa, Marie-Helene Gagnon, Rui-Hao L. Wang , Gina M. Peloso , Robert A. Hegele , Nabil G. Seidah , Sekar Kathiresan , Jacques Genest . *APOE* p.Leu167del mutation in Familial Hypercholesterolemia *Atherosclerosis*