`

Genetic functional studies of low density lipoprotein-cholesterol (LDL-C) associated variants and the genetic spectrum of familial hypercholesterolemia in different ethnic groups

Roaa Hani Fairoozy

University College London

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

Centre for Cardiovascular Genetics

UCL Institute of Cardiovascular Science

Declaration of work

I, Roaa Hani Fairoozy, confirm that the work presented in this thesis is my own and I have generated it from my original research. Where the information has been obtained from other sources, this has been indicated in the text. My role in the work presented in each chapter is outlined below.

In Chapters 3 and 4, Targeted-next generation sequencing (NGS) was carried out by Dr. Marta Futema, a member of the cardiovascular genetics laboratory (CVG lab) prior to the start of the project. I carried out NGS data analysis and mutation confirmation using Sanger sequencing. Also, I performed all bioinformatics and statistical analyses as well as co-segregation analysis of some novel mutations identified in this thesis.

In Chapter 5, I was responsible for genotyping NPHSII for the SNP rs6511720. Genotype data statistical analysis was carried out by Ms. Jackie Cooper (UCL). The meta-analysis of UCLEB data and GWAS data were performed by Dr. Jon White (UCL). I carried out all bioinformatics analysis including identification of SNPs in strong LD, analysis of gene regulatory markers associated with these SNPs using Genome browsers Haploreg and MatInspector, STRING and selection of candidate SNPs. Also, I performed all molecular functional studies including cloning experiments, transfection experiments, luciferase assays, electrophoretic mobility shift assays (EMSAs), and multiplex competitor EMSAs.

In Chapter 6, I carried out all bioinformatics analyses including identifying SNPs in the R1-domain of ANXA2 using the HapMap project data, selecting the candidate SNPs for genetic

analysis and identifying SNPs in strong LD. Also, I was responsible for genotyping NPHSII and WHII for the SNP rs17845226. Genotype data were analyzed by Ms. Jackie Cooper.

In Chapter 7, I was responsible for genotyping NPHSII for the SNP rs11633032. Genotype NPHSII data were analyzed by Ms. Jackie Cooper. The meta-analysis of UCLEB-imputed SNPs, rs11633032, rs12900101, and rs17191344 were performed by Dr. Jon White. I carried out all bioinformatics analyses including identification of SNPs in strong LD, analysis of gene regulatory markers associated with these SNPs using Genome browsers HaploReg and MatInspector. I used the Gene-tissue expression (GTEx) browser to investigate eQTLs. I performed all molecular functional studies including cloning experiments, transfection experiments, luciferase assays, EMSAs, multiplex competitor EMSAs and super-shift EMSAs. GTEx unpublished data were provided by Dr. Claudia Giambartolomei (Icahn School of Medicine at Mount Sinai, New York, USA). The eQTL analysis of the Advanced Study of Aortic Pathology (ASAP) study was performed by Dr. Lasse Folkersen (The Karolinska Institute, Stockholm, Sweden).

In Chapter 8, I carried out all bioinformatics analyses using miRNA target prediction tools including miRWalk, mirSNP and miRanda. I performed all molecular functional studies including TaqMan gene expression assays, cloning experiments, transfection experiments and luciferase assays. My understanding and design of the protocol was developed with the advice of Dr. Anastasia Z. Kalea (UCL).

Acknowledgments

Firstly, I would like to acknowledge all participants of the studies involved in this project. Without them, none of the work presented in this thesis would have been possible. It is through their willingness to donate various samples for research that new medical and scientific knowledge has been gained.

I would like to thank my supervisor, Professor Steve Humphries, for his guidance and encouragement throughout my project, which helped me to bring out the best in myself and enjoy different areas of genetics. Also, I would like to thank my secondary supervisor, Dr.Anastasia Kalea for her support many times throughout my PhD. Sincere thanks to Jackie Cooper and Jon White, who have helped me with statistical methods and analysis; they were always happy to answer my questions. Thank you to all project collaborators, Tomas Grewal (Australia), Claudia Giambartolomei (USA), Lasse Folkersen (Sweden), Amirhossein Sahebkar (Iran) and Inessa Tracey (UK), who have given me a great deal of advice and together we have brought the scientific knowledge gained in this project to the scientific community by publishing in high impact journals.

Thank you to all past and present members of CVG who have made the laboratory such an enjoyable environment to work in and for being a welcoming and friendly. I would like to especially thank Katherine for always being supportive during some hard times, Dauda for being my technology advisor and Ann for academic encouragement and advice.

Thank you to my family, who have given me a great deal of support. I must also extend my most sincere gratitude to my parents, brother and sister. Although my father passed away years ago, he is always my idol, he taught me how to be strong, positive, ambitious, creative and humble. Thanks for the financial support that enabled me to continue my studies in the best universities in the world. My mother has always given me tremendous support both emotionally and spiritually.

Finally, thank you to the Kingdom of Saudi Arabia government and the King Abdullah Medical City in Makkah for providing valuable funding for my PhD and for past and future studies. “Thank you” does not seem enough.

Publication resulting from the work in this thesis

1. Fairoozy, R. H., White, J., Palmen, J., Kalea, A. Z. & Humphries, S. E. 2016. Identification of the Functional Variant(s) that Explain the Low-Density Lipoprotein Receptor (LDLR) GWAS SNP rs6511720 Association with Lower LDL-C and Risk of CHD. *PLOS ONE,* 11**,** e0167676.

2. Fairoozy, R. H., Cooper, J., White, J., Giambartolomei, C., Folkersen, L., Wannamethee, S. G., Jefferis, B. J., Whincup, P., Ben-Shlomo, Y., Kumari, M., Kivimaki, M., Wong, A., Hardy, R., Kuh, D., Gaunt, T. R., Casas, J. P., Mclachlan, S., Price, J. F., Hingorani, A., Franco-Cereceda, A., Grewal, T., Kalea, A. Z. & Humphries, S. E. 2017. Identifying low density lipoprotein cholesterol associated variants in the Annexin A2 (ANXA2) gene. *Atherosclerosis,* 261**,** 60-68.

3. Fairoozy, R. H., Futema, M., Vakili, R., Abbaszadegan, M. R., Hosseini, S., Minzadeh, M., Zaeri, H., Mobini, M., Humphries, S. E., Sahebkar, A. 2017. The Genetic Spectrum of Familial Hypercholesterolemia (FH) in the Iranian Population. *Scientific Reports*, 7(1):17087.

4. Inessa, T., Fairoozy, R. H., Futema, M., Elizabeth A. Hughes1, E.A., Humphries, S. E.2017. Case report: An interesting south Asian family with familial defective APOB. the Annals of Clinical Biochemistry. Submitted November 2017.

ABSTRACT

Since the end of the 20th century, cardiovascular disease has been the major cause of death worldwide. Hypercholesterolemia is major risk factor for CVD. Family-based linkage analyses in patients with familial hypercholesterolemia (FH) have revealed the major loci that are involved in cholesterol hemostasis. These include low density lipoprotein receptor (*LDLR*), proprotein convertase subtilisin/kexin type-9 protein (*PCSK9*), and apolipoprotein B (*APOB*). However, these discovered genes only explain disease etiology in a proportion of patients with the clinical phenotype, and roughly 60% of patients who are diagnosed with FH are negative for any FH-causative mutation. This suggests that (1) there are undetected new mutations in known genes, (2) that disease-causing mutations might occur in unidentified gene(s), (3) disease may occur as a result of an accumulation of common small-effect LDL-C raising variants.

My project has three main aims. First, the identification of mutations in patients clinically diagnosed with FH in two ethnic groups: UK (n=69) and Iran (n=16). Several novel mutations were identified in *LDLR* [UK: p.(Cys392Tyr), p.(Tyr553Ser) and p.(Tyr553*); Iran*:* p.(Leu479Gln) and p.(Glu668*)] and *PCSK9* [p.(Arg357Cys)] using next generation sequencing technology. The second aim was to investigate the functional role explaining the Genome-Wide Association (GWAS) *LDLR* "hit" Single Nucleotide Polymorphism (SNP) rs6511720. This SNP and one other in strong linkage disequilibrium (LD) (rs57217136) were both found to act as a cis-regulatory element, where the sequence around the rare alleles of the SNPs is a target for proteins that enhance gene transcription. The third aim was studying the Annexin A2 (*ANXA2)* gene. AnxA2 recently was found to be involved in the LDL-R pathway through PCSK9. The study aimed to identify and determine the functional role of common

SNPs that are associated with LDL-C. I showed that the *ANXA2* common SNP rs17845226 (V98L) minor allele, was associated with significantly higher levels of LDL-C and a higher risk of CHD in a large prospective study of healthy UK men (NPHSII). The SNP shows strong LD with SNPs in the intragenic region, and I showed that the minor alleles of rs17191344 and rs11633032 are targets for proteins that repress gene transcription and the subsequent lower levels of AnxA2 protein means that there will be higher levels of PSCK9-mediated degradation of the LDL-R and this will lead to an increase in LDL-C levels. Finally, I identified that rs116928563 in the 3'UTR of *ANXA2* was a potential site for a micro-RNA (miRNA-155*) binding, however, the study failed to result in any definite conclusions about the influence of miRNA-155 in *ANXA2* expression and LDL-C levels. However, it gives insight into how a SNP also may affect gene expression at the post-transcription level by creating or destroying micro-RNA binding sites.

CONTENTS

List of Figures

List of Tables

List of Abbreviations

3'UTR	3' Untranslated Regions
AAA	Abdominal Aortic Aneurysm
ADAM	Advanced Detection and Accurate Measurement
ADH	Autosomal Dominant Hypercholesterolemia
Alu repeats	Arthrobacter luteus repeats
ANXA2	Annexin A2 gene
AnxA2	Annexin A2 protein
APOB	Apolipoprotein B gene
ApoB-100	Apolipoprotein B-100
ApoB-48	Apolipoprotein B-48
ApoC	Apolipoprotein C
ApoE	Apolipoprotein E
ARMS	Amplification-refractory mutation system
ASO	Antisense oligonucleotides
BMI	Body Mass Index
BPV-1	Bovine Papillomavirus Type 1
CAT	Catalytic domain of PCSK9
CELSR2	Cadherin EGF LAG Seven-Pass G-Type Receptor 2
CETP	Cholesteryl ester transfer protein
CF	Cystic fibrosis
CHD	Coronary Heart Disease
CHRD	Cysteine-Histidine-Rich Domain
CTCF	CCCTC-binding factor
CTD	C-terminal of PCSK9
CTFBs	Conserved Transcription Factor Binding Sites
CVD	Cardiovascular Disease
CYP7A1	Cytochrome P450 Family 7 Subfamily A Member 1

DFH	Definite Familial Hypercholesterolemia
DLCNC	Dutch Lipid Clinic Network Criteria
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal Growth Factor
EGFR	Epidermal growth factor receptor
Egr1	Early growth response protein 1
EMSA	Electrophoretic Mobility Shift Assay
ENCODE	Encyclopedia of DNA Elements
eQTL	Quantitative Trait Loci
ER	Endoplasmic Reticulum
FAIRE	Formaldehyde assisted isolation of regulatory elements
FBS	Foetal Bovine Serum
FH	Familial Hypercholesterolemia
FOXA	Forkhead box protein A
FOXB1	Forkhead box protein B1
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GATA	GATA binding protein
GLGC	Global Lipid Genetic Consortium's
GWAS	Genome-Wide Association Studies
HapMap	Haplotype map
HDACs	Histone deacetylases
HDL	High Density Lipoproteins
HeFH	Heterozygous Familial Hypercholesterolemia
HGP	Human genome project
HMG-CoA-R	Hydroxymethylglutaryl-CoA Reductase
HoFH	Homozygous Familial Hypercholesterolemia
HRM	High-resolution melting

HWE	Hardy-Weinberg equilibrium
IDL	Intermediate Density Lipoproteins
IHD	Ischemic Heart Disease
LAR II	Luciferase Assay Reagent II
LB	Luria Broth
LB/ampicillin	Luria Broth/ Ampicillin
LD	Linkage Disequilibrium
LDL	Low Density Lipoproteins
LDL apheresis	LDL-A
LDL-C	Low Density Lipoprotein-Cholesterol
LDLR	Low Density Lipoprotein Receptor Gene
LDL-R	Low Density Lipoprotein Receptor protein
LDLRAP	Low-Density Lipoprotein Receptor Adaptor Protein
LIPC	HepaticLipase
LPL	Lipoprotein Lipase
LXRα	Liver X receptor alpha
MAF	Minor Allele Frequency
MC-EMSA	Multiplexed Competitor-Electrophoretic Mobility Shift Assay
MGB	Minor Groove Binder
MI	Myocardial Infarction
miRNA	MicroRNA
mirSVR	MiRNA support vector regression
MLPA	Multiplex ligation-dependent probe amplification
MYOCD	Myocardin
NARG2	NMDA regulated 2
NCOR	Nuclear receptor co-repressors
NFQ	Non-Fluorescent Quencher
NGS	Next generation sequencing
NICE	Clinical Health and Excellence

NPHS II	Second-Northwick-Park Heart Study
NTC	No Template Control
oxLDL	Oxidized LDL
PCR-RFLP	PCR-Restriction Fragment Length Polymorphism
PCR-SSCP	PCR-single-strand conformation polymorphism
PCSK9	Proprotein Convertase Subtilisin/Kexin type-9 gene
PCSK9	Proprotein Convertase Subtilisin/Kexin type-9 protein
PFH	Possible Familial Hypercholesterolemia
PPAR	Peroxisome Proliferator Activated Receptor
PSRC1	Proline/serine-rich coiled-coil protein 1
RA	Retinoic acid
RAR	Retinoic Acid Receptor
RARA	Retinoic Acid Receptor Alpha
RAREs	Retinoic acid response elements
RISD	RNA induced silencing complex
RNA	Ribonucleic acid
RNA pol II	RNA polymerase II
RORA	RAR- related orphan receptor alpha
RXR	Retinoid X receptor
SCAP	SREBP Cleavage-Activating Protein
SDM	Site-Directed Mutagenesis
SIE	Sis-inducible element
SMART	Silencing mediator for retinoid and thyroid hormone receptor
SMCs	Smooth muscle cells
SNP	Single Nucleotide Polymorphism
SORT1	Sortilin 1gene
SORT1	Sortilin 1 protein
SP1	Specificity Protein 1
SRE-1	Sterol Regulatory Element-1

SREBP	Sterol Regulatory Element Binding Protein
SREs	Steroid Response Elements
SSD	Sterol-Sensing Domain
STAP1	Signal-transducing adaptor protein 1 gene
STAP1	Signal-transducing adaptor protein 1
STAT	Signal transducer and activator of transcription
T2D	Type 2 diabetes
TC	Total Cholesterol
TEMED	N,N,N′,N′-Tetramethylethylenediamine
TFBs	Transcription Factor Binding Sites
TG	Triglycerides
TX	Tendon xanthomas
UCLEB	UCL-LSHTM-Edinburgh-Bristol consortium
UCSC	University of California Santa Cruz
UTR	Untranslated Region
VLDL	Very Low Density Lipoproteins
WHII	Whitehall II study
YY1	Yin Yang 1 (transcription factor)

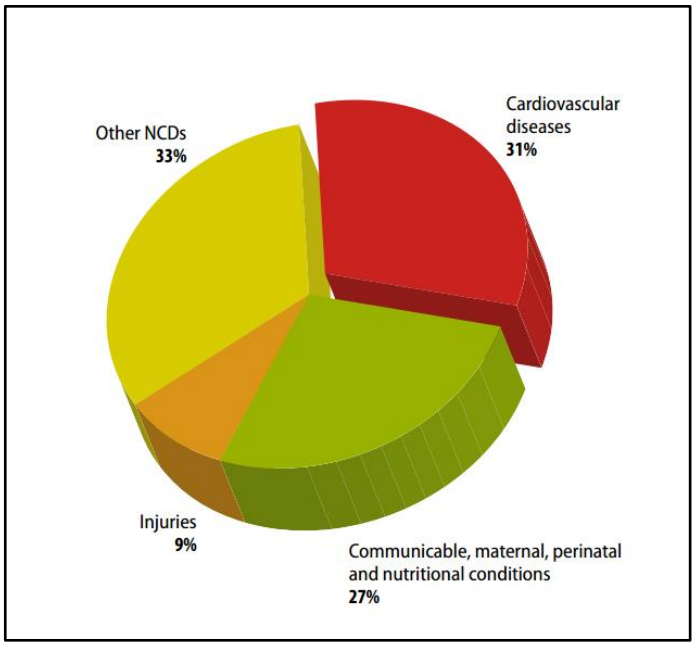
Chapter 1 : Introduction

1.1. Cardiovascular disease (CVD) overview

Cardiovascular disease (CVD) is a disease of the heart and vascular system, comprising coronary heart disease (CHD), stroke, peripheral arterial disease and aortic disease. It causes 30% of mortality world-wide (Figure1-1 A) (1, 2). In 2012, the British Heart Foundation published a report that CVD was the main cause of death in the UK, accounting for almost 180,000 deaths, around one in three of all deaths in that year. In order to prevent or reduce morbidity and mortality as a result of CVD, it is important to diagnose at-risk individuals at an early stage by considering CVD risk factors.

Figure 1-1: Main causes of death worldwide

The pie chart shows that the main causes of mortality globally. Non-communicable diseases (NCDs) are the leading causes of death globally 33%, followed by, cardiovascular diseases 31%, communicable diseases, maternal and parental conditions, and nutritional deficiencies 27%, and injuries 9%. Adopted from http://www.world-heart-federation.org/fileadmin/user_upload/images/CVD_Health/Global_CVD_Atlas.pdf and http://www.who. int/nmh/publications/ncd_report_full_en.pdf) Prasad et al. (2010).



1.1.1. The risk factors of CVD

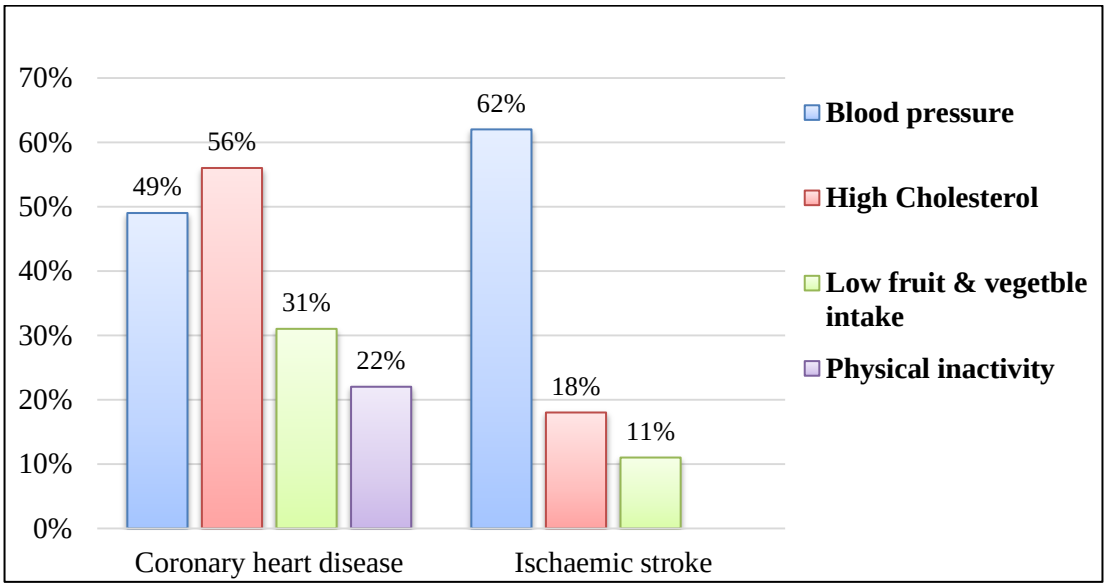
Several risk factors are known to influence the likelihood of developing CVD. These factors are classified into two types, non-modifiable and modifiable. The non-modifiable risk factors are factors that cannot change including age, gender, family history and ethnicity. Age is the strongest risk predictor of CVD (3), where there is a positive correlation between age and the risk of the disease. Gender is also another strong risk factor; men tend to develop CVD about 7 to10 years earlier than women (4). Also, both family history (5) and ethnicity (6) are important non-modifiable risk factors of CVD, which indicate the role of genetics in the risk of CVD.

The modifiable CVD risk factors, also known as "lifestyle" risk factors, include hypertension, blood lipid levels, diabetes, obesity and smoking. Prasad et al. (2010) have stated that studies by the World Health Organization (2004) showed that a high cholesterol level is a major risk factor for CHD (56%) (Figure 1-2). Another important CVD risk factor is hypertension, which links to about 50% of CHD and 62% of ischemic stroke (World Health Organization, 2004, Prasad et al., 2010). Diabetes mellitus also is a risk factor of CVD; it was reported that diabetic patients are at three times higher risk of a heart attack (7). Obesity has been classified as a modifiable CVD risk factor, which has influence on known risk factors such as dyslipidemia, hypertension and glucose intolerance (8). Smoking is a behavioral lifetime risk of many life-threatening diseases. In 2000, 11% of deaths worldwide from CVD were linked to smoking (9). It has been reported that smoking promotes CVD through several mechanisms: it increases low-density lipoprotein cholesterol (LDL-C) and is involved in the oxidative modification of LDL-C (10), it increases the risk of damage of the endothelium layer of the blood vessels and plaque rupture (10, 11), it increases the risk of clotting via

elevating the levels of clotting factors such as fibrinogen (12), it increases the risk of high blood pressure and heart rate through by delivering high arterial plasma levels of nicotine (13), and it increases the risk of diabetes mellitus (14).

Figure 1-2: Percentage contribution of modifiable CVD risk factors

The bar chart shows the percentage contributory risk factors to CHD and ischemic stroke. The major risk factor for CHD is high cholesterol levels followed by high blood pressure. In contrast, the major risk factor for ischemic stroke is high blood pressure followed by high cholesterol levels. Adapted from Prasad et al. (2010).



1.1.2. Managing risk factors

Lifestyle risk factors can be managed by quitting smoking, and adopting good nutrition and regular physical activity. Diet affects CVD in a variety of ways, such as, energy imbalance, where more energy is taken in than expended which can result in weight gain and obesity, high saturated fat raises cholesterol levels, high salt intake can raise blood pressure, and low intake of fiber, fruit and vegetables can lead to a greater susceptibility to CVD. However, persuading people to make changes to their lifestyle to reduce risk can be difficult. Therefore,

procedures to identify people with high likelihood of developing CVD, and interventions to reduce risk are essential to disease prevention.

A number of CVD risk prediction algorithms have been developed to calculate a person's relative risk of cardiovascular events. Currently, two common algorithms are used, the Framingham (15) and QRISK (16) risk scores, which include conventional risk factors (age, sex, smoking habit, physical activity, blood pressure, cholesterol level, body mass index (BMI), presence of diabetes, and family history). These algorithms are used to estimate the patient's 10-year risk of developing CVD, in order to identify high-risk individuals (a total CVD risk of ≥20%) for primary prevention (16). Although these risk assessment tools help to identify high-risk people, they underestimate the true total risk of CVD. Nearly 15% of individuals who develop CVD are considered to be low-risk, based on conventional risk assessment models (17). Identifying higher-risk patients from the many who are considered intermediate-risk or low-risk could help practitioners' decisions to start primary prevention process at an early stage. Recently, there has been debate about the clinical utility of adding a genetic risk score (GRS) to risk models to improve predictive performance and risk classification. GRS takes advantage of the findings of genome-wide association studies (GWASs), which have identified a correlation between single nucleotide polymorphisms (SNPs) and clinical phenotypes of complex diseases; adding SNP associated diseases to risk assessment models could enhance the predictability of CVD (17).

In order to reduce the risk of CVD, statins are used as primary and secondary prevention methods (18). These drugs reduce total cholesterol (TC) and LDL-C levels in circulation via inhibiting hydroxymethylglutaryl-CoA reductase (HMG-CoA-R) enzyme synthesis (19).

Although statins are effective and convenient with limited side-effects, patients have variable potency statins and the LDL-C reduction outcome range varies from 30% to 60% (20). In addition, statins increase proprotein convertase subtilisin/kexin type-9 (PCSK9) levels (21) because the *PCSK9* promoter also contains a sterol responsive element. The higher levels of PCSK9 then bind to the LDL-receptor (LDL-R) and targets it for degradation, thus mitigating to some extent the statin induction of LDL-Rs. Statins also have been shown to increase the risk of diabetes (22). Mendelian randomization study suggested a direct consequence between inhibition of HMG-CoA-R and type 2 diabetes (T2D) risk (23). However, the benefits of statin treatment for CVD prevention heavily outweigh the small increase in T2D risk.

It is essential to do more intervention studies to determine the best strategies to lower plasma LDL-C levels in order to prevent CVD. In the last few years, PCSK9 has become a proposed target of intervention to reduce risk of CVD; several approaches are proposed which are discussed later.

1.2. Lipoproteins and cardiovascular disease (CVD)

Cholesterol is a lipophilic molecule, which is transported through the blood in lipoproteins. The type of lipoprotein is determined by its density, composition, and protein content. Lipoproteins are classified into five classes: Chylomicrons, Very low density lipoproteins (VLDL), intermediate density lipoproteins (IDL), low density lipoprotein-cholesterol (LDL-C), and high density lipoproteins-cholesterol (HDL-C).

Lipoprotein trafficking is divided into three pathways: the exogenous pathway, the endogenous pathway, and the reverse cholesterol pathway.

1.2.1. Exogenous pathway

The exogenous pathway provides efficient transport of dietary lipids. Dietary triglycerides (TG) hydrolyzed by lipases to fatty acids and monoacylglycerol are readily absorbed by the small intestine where they are re-esterified to form triacylglycerol. In the enterocyte, cholesterol is esterified to form cholesteryl esters, and then longer-chain fatty acids are incorporated into TG and packaged with apolipoprotein B-48 (ApoB-48), cholesteryl esters, retinyl esters, phospholipids and cholesterol to form chylomicrons. Nascent chylomicrons enter the blood circulation via the lymphatic system, where they acquire apolipoproteins E and C (ApoE, ApoC-I, ApoC-II, ApoC-III) from HDL-C. ApoC-II, acts as a co-factor for lipoprotein lipase (LPL), which hydrolyzes the TG of chylomicrons, and free fatty acids are released. The released free fatty acids are taken up by other tissues such as muscle and adipose tissue (Figure 1-3). The chylomicron particle progressively shrinks in size and creates a new lipoprotein called a chylomicron remnant. Chylomicron remnants are enriched with cholesteryl ester and fat-soluble vitamins and contain apoB-48 and apoE. They are rapidly

removed from the circulation by receptors in the liver including the LDL-R, an ApoB/E receptor, and the LDL-R related protein (24-27).

1.2.2. Endogenous pathway

The endogenous pathway refers to the hepatic secretion and metabolism of apoB-containing lipoproteins (VLDL, IDL, and LDL-C) (27). To maintain and regulate the energy supply for tissues, the liver needs TG, derived predominantly from the esterification of long-chain fatty acids in the liver. Hepatic TG are packaged with other components including apoB-100, cholesteryl esters, phospholipids, and vitamin E to form VLDL particles. The nascent VLDLs enter the circulation, where they acquire multiple copies of apoE and apoC from HDL-C. Once nascent VLDLs are loaded with apoC-II and apoE, they are considered mature. As with chylomicrons, the TG of VLDL are hydrolyzed by LPL, and free fatty acids are released and taken up by peripheral tissues. The cholesteryl ester transfer protein (CETP) controls the exchange between TG from VLDLs and cholesteryl esters from mature HDLs, which enriches VLDLs in cholesterol esters which then form intermediate density lipoproteins (IDLs). The liver removes IDL by LDL-Rs via binding to apoE or remodeling IDL by hepatic lipase to form LDL-C. During this process, most of the triglyceride in the particle is hydrolyzed, and all apolipoproteins are removed except apoB-100. The apoB-100 in circulating LDL-C is capable of being bound to LDL-Rs and internalized (Figure 1-3). The internalized LDL-C particles undergo degradation and the cholesterol is metabolized to produce bile (25-27).

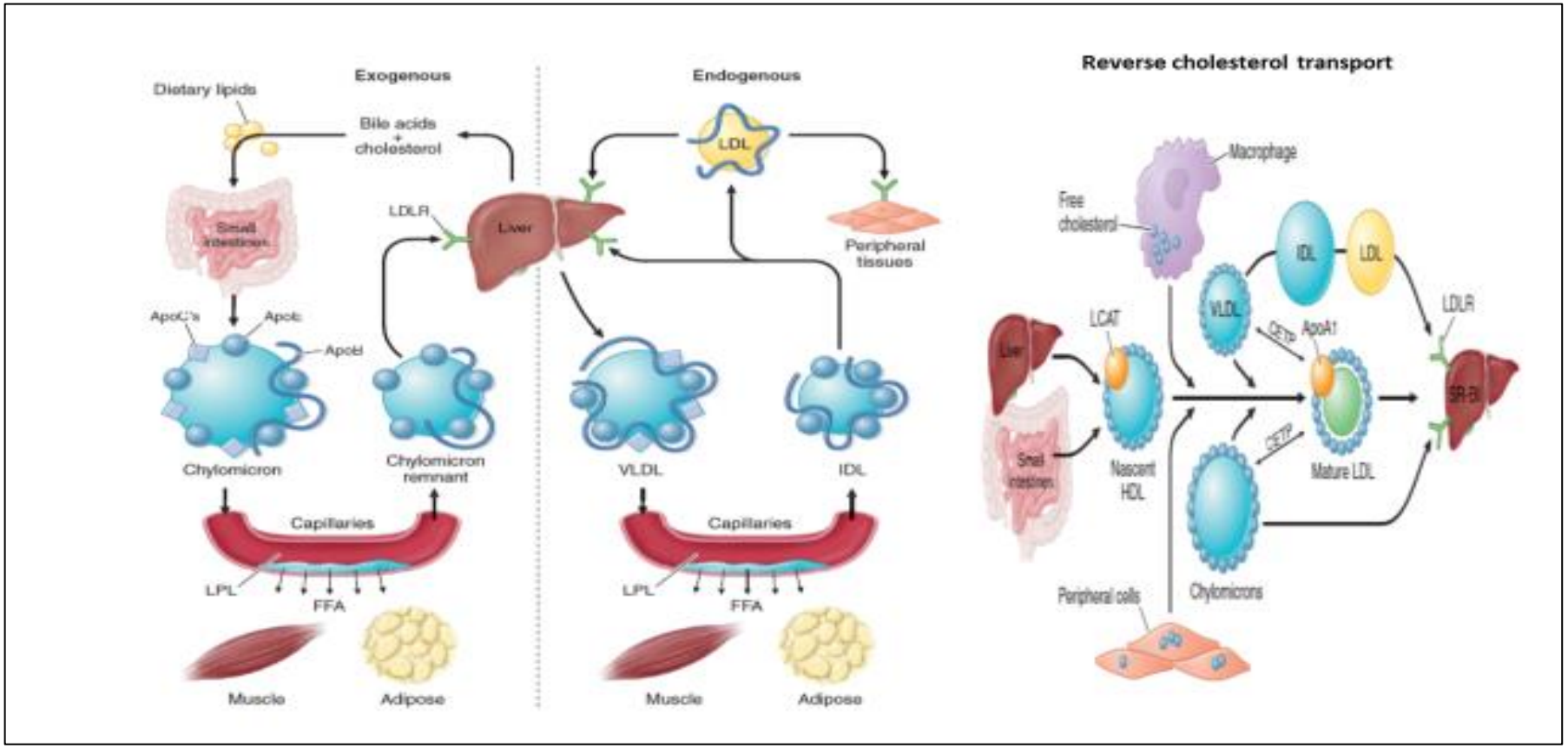
1.2.3. Reverse cholesterol transport

Cholesterol in peripheral cells is transported from the plasma membranes of peripheral cells to the liver through a process termed "reverse cholesterol transport" that is facilitated by HDL-C.

ApoA-I, synthesized by the liver, rapidly acquires phospholipids and unesterified cholesterol from its site of synthesis via the membrane protein ATP-binding cassette protein A1 (ABCA1) in hepatocytes. This process results in the formation of pre-beta-HDL particles. Further free cholesterol is acquired from macrophages and other peripheral cells and esterified by lecithin-cholesterol acyltransferase (LCAT). LCAT converts free cholesterol to cholesteryl ester, thereby moving the esterified cholesterol inside the HDL-C particle and forming mature HDL-C. HDL-C can be selectively taken up by the liver via the scavenger receptor class BI (SR-BI). Alternatively, HDL-C cholesteryl ester can be transferred by CETP from HDL-Cs to VLDLs, chylomicrons and LDLs, which can then be taken up by the liver (Figure 1-3) (25, 28).

Figure 1-3: Lipoprotein trafficking pathways

A cartoon showing the three lipoprotein trafficking pathways: the exogenous pathway transports dietary lipids to the periphery and the liver. The endogenous pathway transports hepatic lipids to the peripheral tissues. HDL metabolism and reverse cholesterol transport pathway transport excess cholesterol from the peripheral tissues back to the liver for excretion in the bile. Adapted from Rader and Hobbs (2005).

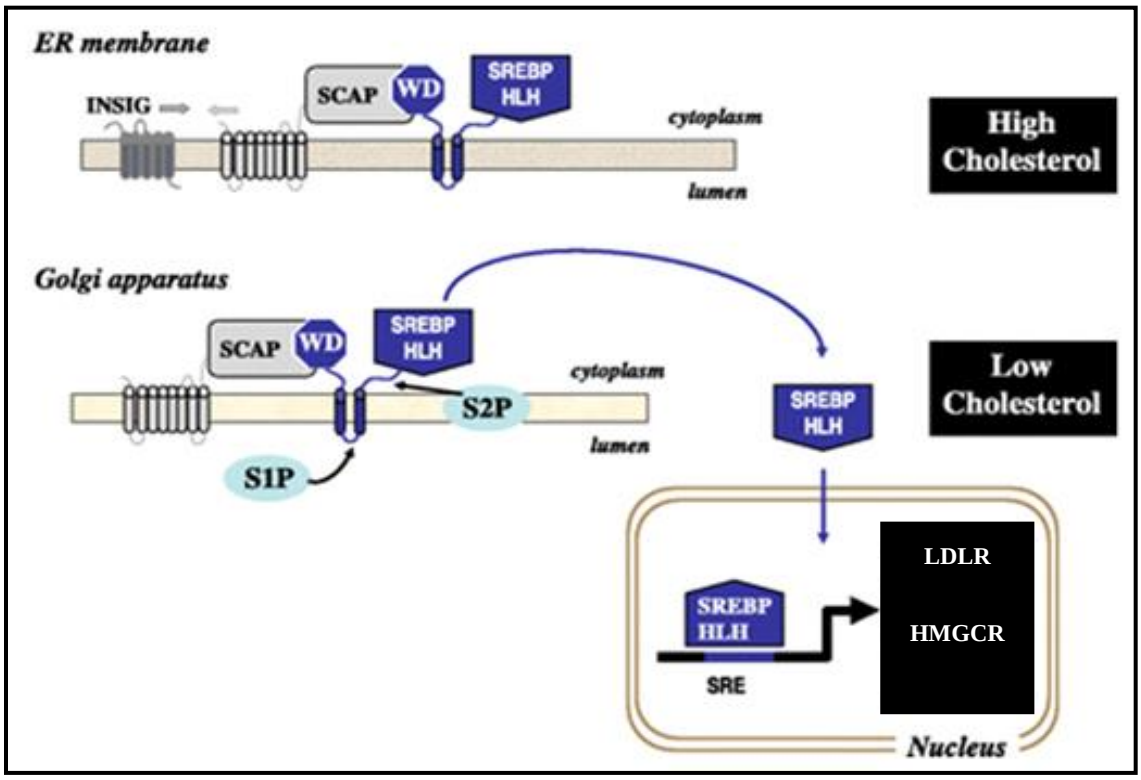


1.3. Cholesterol homeostasis

Cholesterol biosynthesis is regulated by the intracellular concentration of free cholesterol in hepatic cells. Intercellular cholesterol homeostasis is controlled by steroid response element binding protein (SREBP). SREBP is located in the endoplasmic reticulum (ER) in a complex with SREBP cleavage-activating protein (SCAP). SCAP has a cholesterol-sensing domain in its N-terminal region, when cells are loaded with sterol, SCAP transport SREBP from the ER to Golgi bodies. Once in the Golgi, SREBP is cleaved in order to release part of the protein from the membrane by two specific proteases (site-1 and site-2 proteases). Site-1 protease (S1P), cleaves SREBP in the luminal loop between two membrane-spanning sequences. SCAP allows site-1 cleavage to be activated when cells are deprived of sterols. Once the two halves of SREBP are separated, a second protease, site-2 protease (S2P), cleaves the NH_2-terminal bHLH-Zip domain of SREBP at a site located within the membrane-spanning region. After the second cleavage, the NH_2-terminal bHLH-Zip domain leaves the membrane (29, 30). This new form of SREBP enters the nucleus where it binds to steroid response elements (SREs) in the enhancer/ promoter region of many target genes involved in cholesterol hemostasis and activates their transcription, including HMG-CoA-R and LDL-R. HMG-CoA-R is the rate-limiting enzyme for cholesterol synthesis, whereas LDL-R is responsible for LDL-C uptake from plasma (see Figure 1-4). By contrast, an accumulation of free cholesterol inactivates the SREBP pathway, where a sterol-sensing domain (SSD) protein SCAP binds to the ER retention protein Insig. Insig binding prevents the transport of SCAP-SREBP complex, thus the transcription of SRE regulated genes is not turned on. Cholesterol-lowering drugs such as statins are used to activate the cholesterol synthesis pathway (31-34).

Figure 1-4: The steroid response element binding protein (SREBP) pathway in cholesterol regulation

A diagram showing the SREBP pathway in both high and low levels of cholesterol. At lower cellular cholesterol levels, SCAP binds SREBP in the ER. The SCAP-SREBP complex then moves to the Golgi, where two specific proteases (site-1 and site-2 proteases) cleave the SREBP. The cleaved SREBP then enters the nucleus and activates the transcription of the genes coding for LDL-R and for HMG-CoA-R. This in turn stimulates the rate of cholesterol uptake and synthesis. Conversely, when cellular cholesterol levels are higher, Insig ER protein binds to the SCAP-SREBP complex and prevents SREBP transport, therefore, the enzyme required for gene transcription is not produced and uptake and synthesis of cholesterol is not enhanced. Adapted (34) with modification.



1.4. Atherosclerosis

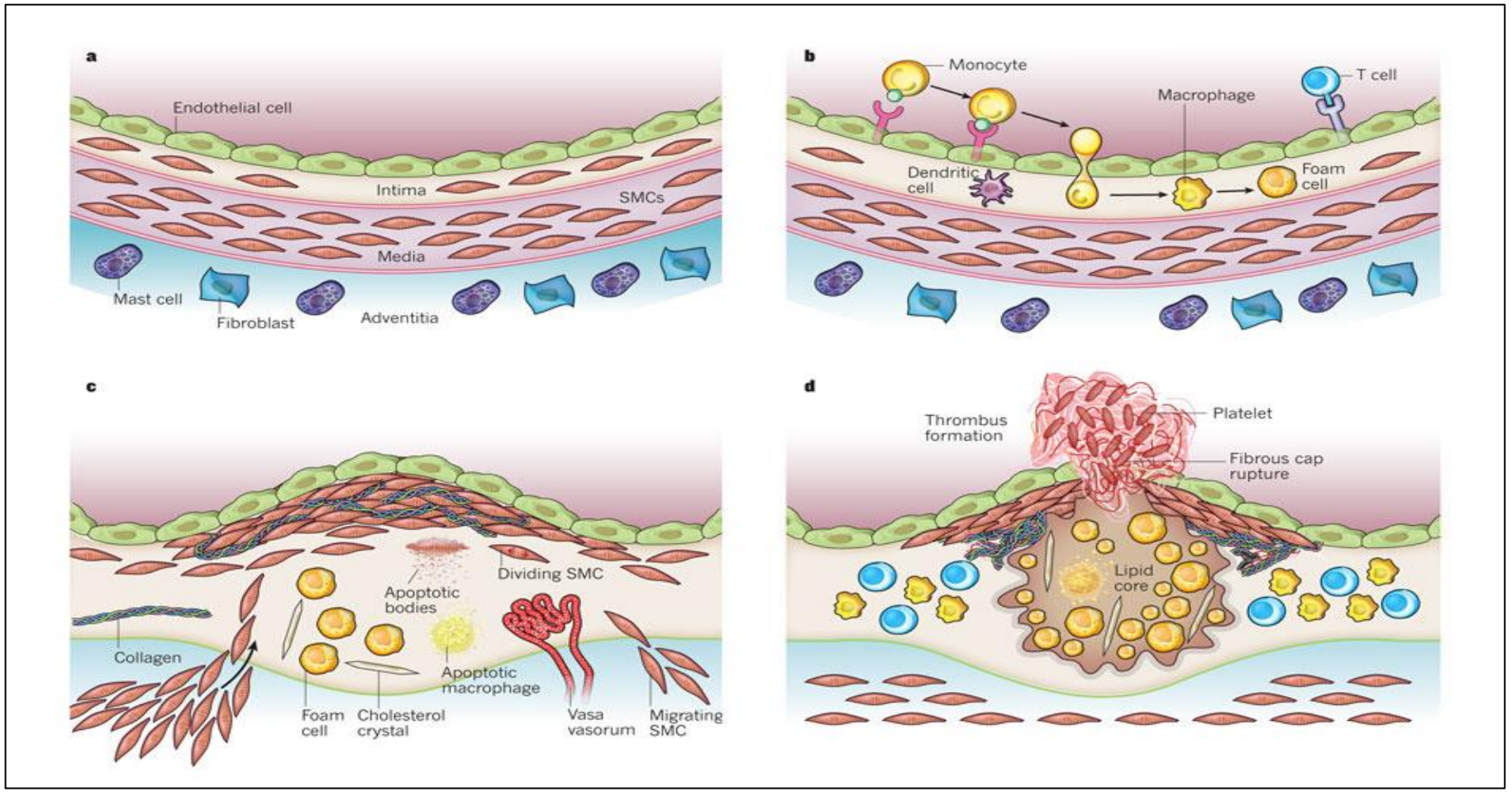
CVD arises from atherosclerosis. Atherosclerosis is a multifactorial disease that has both modifiable and non-modifiable risk factors. In the majority of patients, atherosclerosis develops through a combination of different factors, including an interaction between genes and an interaction between genes and the environment.

An accumulation of cholesterol particles in the walls of arteries is the primary driver of atherosclerosis. These particles initiate the inflammatory process, which leads to atherosclerotic blood vessels. Atherosclerosis is a progressive disorder marked by several stages, with a varying extent of lesions marking each stage. It is initiated by accumulation of LDL-C in the sub-endothelium, which leads to remodeling of LDL-C to oxidized LDL (oxLDL). ApoB-100 binds to negatively charged extracellular matrix proteoglycans, which leads to retention of LDL-C particles in the intima, where they are susceptible to oxidative modification by reactive oxygen species, or enzymes such as myeloperoxidase or lipoxygenases released from inflammatory cells. An accumulation of oxLDL promotes the recruitment of monocytes and lymphocytes and an increase in the production of cytokines and growth factors. Monocytes and lymphocytes are converted to macrophages that take up oxLDL and form foam cells (Figure 1-5 (b)). Foam cells form early 'fatty-streak' lesions, which activate inflammation. The early 'fatty-streak' lesion consists of T cells and monocyte-derived macrophage-like foam cells loaded with lipids. Sequential accumulation of apoptotic cells, debris and cholesterol crystals forms a necrotic core, which leads to infiltration and proliferation of smooth muscle cells (SMCs) and synthesis of extracellular matrix containing collagen, elastin and proteoglycans to form fibrous plaques (Figure 1-5 (c)). Ruptured plaques with thin fibrous caps result from degradation of the matrix by proteinases such as

collagenases, gelatinases, stromelysin and cathepsins, and by inhibition of matrix secretion (Figure 1-5 (d)). Thus, various factors may destabilize plaques and promote thrombosis and blocking of the artery, which can lead to myocardial infarction (MI) and sudden death (35-37).

Figure 1-5: Stages in the development of atherosclerotic lesions

A cartoon showing the atherosclerotic lesions development stages: a) normal blood vessel layers, b) OxLDL accumulation initiates primary inflammation response via forming foam cells; c) foam cells activate an inflammatory cascade with forms a fibrous plaque, and d) fibrous plaque and thrombosis formation, which leads to blocking of the artery. Adapted from Libby et al. (2011).



1.5. Hypercholesterolemia

Elevated plasma cholesterol levels are a major risk factor for atherosclerosis and CVD. They are caused by a disturbed balance between cholesterol secretion into the blood and uptake resulting in a disease called hypercholesterolemia. Hypercholesterolemia is defined as a multifactorial disease due to a combination of environmental and genetic factors. There are two types of disease, one is developed from inappropriate life style (smoking, diet, lack of physical activity, obesity, and alcohol) and additive small effects of multiple genes "polygenic hypercholesterolemia". The other type is a Mendelian disorder called familial hypercholesterolemia (FH) which is caused by a defect in a single gene.

Although hypercholesterolemia is a multifactorial disease, studies of FH have helped unravel the pathways that regulate plasma cholesterol metabolism, knowledge of which is important for the development of cholesterol-lowering drugs. In the next section general knowledge about FH including disease discovery, types, genes, diagnostic criteria and treatment are provided in order to understand the genetic power behind hypercholesterolemia.

1.6. Familial hypercholesterolemia (FH)

1.6.1. Overview

In the early 20th century, Archibald Garrod published his 'Inborn Errors of Metabolism' theory, where he said that a single defect in one gene could produce a block in a biochemical pathway, the so called "one gene-one enzyme" hypothesis (38). Based on Garrod's hypothesis, researchers have focused on understanding biochemical pathways in order to uncover the gene causing a particular disorder. One of the studied disorders was the inherited metabolic disorder FH, which is an autosomal dominant disease characterized by high concentrations of plasma LDL-C.

Between 1972 and 1985, Brown and Goldstein's laboratory did extensive research focused on dissecting the genetic architecture of FH. FH patients have high levels of cholesterol and suffer from CHD at an early age. Based on that, they postulated that FH occurs as a result of a defect in cholesterol clearance. Two-thirds of cholesterol in human plasma is contained within LDL-C particles, which are key players in the progression of atherosclerosis that leads to premature CHD. Therefore, they focused on understanding the regulatory mechanism of LDL-C (38). In 1973-74, Goldstein and Brown showed that LDL-C could bind to high affinity surface receptor sites on cells and a defect in this process represented the primary genetic abnormality in FH (39, 40). Later on in 1979, they discovered that LDL-C uptake from blood was caused by a cell surface receptor, called the LDL-R and then described the sequential steps of LDL-C uptake (41). Binding of LDL-C to LDL-R initiates internalization of LDL-C≡LDL-R complexes to lysosomes. In lysosomes, LDL-C is hydrolyzed which leads to release of unesterified cholesterol into the lysosome, while the LDL-Rs recycle to the cell

surface. In 1985, Brown and Goldstein were jointly awarded the Nobel Prize in physiology and medicine for their discoveries concerning the regulation of cholesterol metabolism.

There are two forms of FH – heterozygous and homozygous – heterozygous FH (HeFH) is caused by one of the two inherited alleles being affected, which leads to a two-to-three fold increase in LDL-C plasma levels than normal. The estimated prevalence of heterozygous FH is one in 500 in most European populations (42). However, the prevalence is known to be higher in some populations, for example, in Denmark the frequency of FH is 1/137 (43), and French Canadians (44). Individuals with heterozygous FH often develop premature CHD after the age of 35. Homozygous FH (HoFH) is caused by both inherited alleles being affected, which leads to a six-to-eight-fold increase in LDL-C plasma levels than normal. The frequency of the homozygous form is 1 in 1,000,000 in most populations. Individuals with homozygous FH develop CHD in the early stage of their lives and often die before the age of 20 (42).

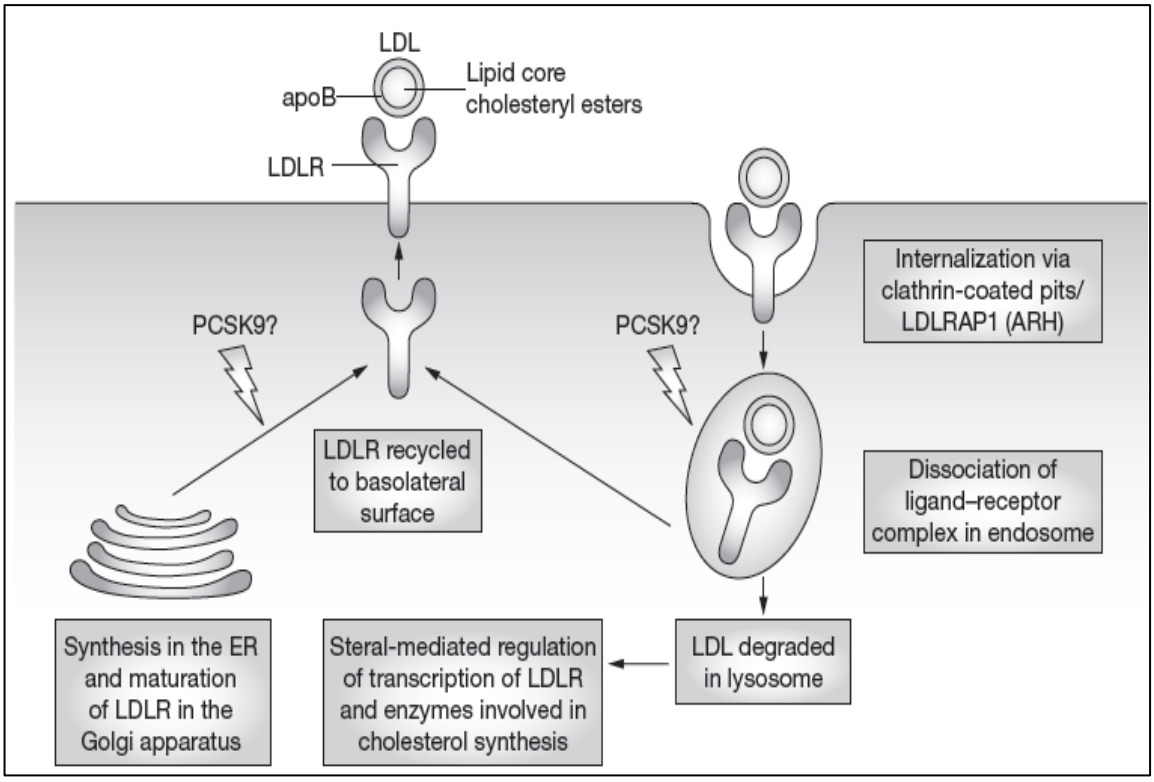
1.6.2. FH genes

FH is considered a monogenic disorder, which is caused by a defect in one of three genes that regulate LDL-C concentration: the *LDLR*, *PCSK9*, *APOB* genes. Rarely is an inheritance pattern of FH of autosomal recessive hypercholesterolemia seen, which occurs when an individual inherits two recessive variants in the low-density lipoprotein receptor adaptor protein 1 (*LDLRAP1*) gene. UK studies have shown that about 93% of identified mutations are found in the *LDLR* gene, a further 5% are in *APOB*, and about 2% are in the *PCSK9* gene (45, 46). All identified FH-causing genes and the involvement of their protein products in the LDL-C clearance pathway are summarized in Figure 1-6. More recently, a potential new FH

gene called signal-transducing adaptor protein 1 (*STAP1*) was discovered, which is the fourth gene associated with hypercholesterolemia (47, 48), but its role in the LDL-C clearance pathway is unknown yet.

Figure 1-6: Molecular pathways of cholesterol regulation

A cartoon showing the molecular pathway of cholesterol clearance. The cholesterol particles are taken-up from the blood by LDL-R protein. The LDL-R is processed through the rough endoplasmic reticulum (ER), matured in the Golgi and transported to the cell surface membrane. LDL-Rs bind specifically to particles containing ApoE or ApoB-100. A clathrin-coated pit is then formed and the receptor–ligand complex is internalized to endosomes through interactions involving LDLRAP1, which causes internalization of the complex to the endosome. Changes in the acidic environment of early endosomes to late endosomes causes dissociation of the receptor–ligand complex. The receptor is recycled to the cell membrane while the LDL-C particle is degraded in the lysosome. Cholesteryl ester in the core of LDL-C is hydrolyzed and becomes free cholesterol in the cell. PCSK9 works as a post-transcriptional inhibitor of LDL-R. Some LDL-R is bound to circulating PCSK9 before endocytosis, which leads to the degradation of the receptor in the endosome, resulting in a decreased number of receptors on the surface of the cells. Adapted from Soutar and Naoumova (2007).



1.6.2.1. Low-density lipoprotein receptor (LDLR)

1.6.2.1.1. *LDLR* ontogeny and structure

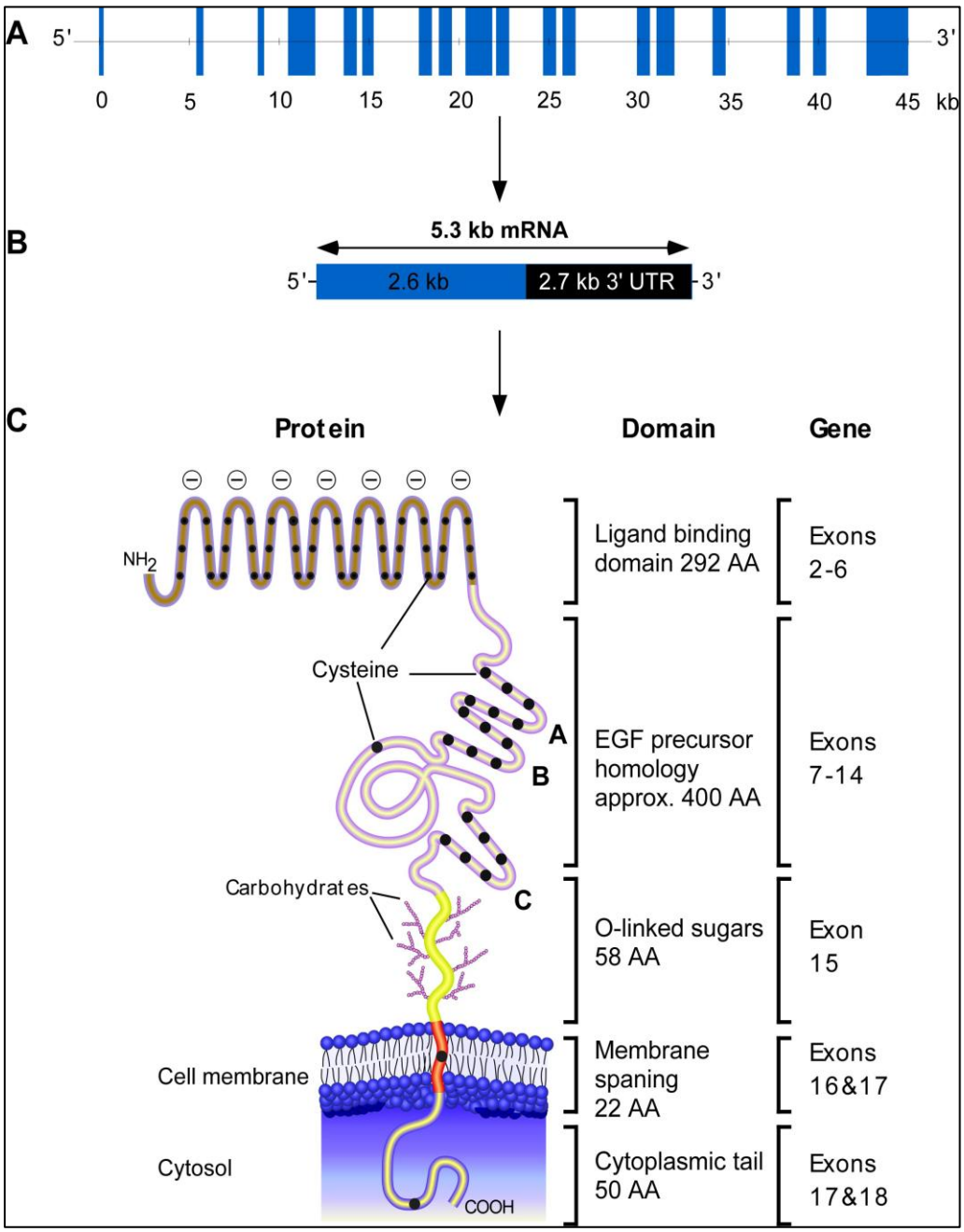
LDLR is located on chromosome 19 at p13.1-p13.3 and it encodes a cell surface glycoprotein predominantly expressed in hepatocytes that mediates the removal of cholesterol carrying LDL-C particles via ApoB-100 from blood (Bruno, et al, 2000). The gene's length of 45 kb comprises 18 exons and 17 introns (42). The mRNA transcript is 5.3 kb in length and encodes a protein of 860 amino acids. About half of the mRNA constitutes a long 3' untranslated region (3'UTR) that contains two-and-a-half copies of the Arthrobacter luteus (Alu) family of middle repetitive DNAs (49, 50).

The mature human LDL-R of 160 kDa is composed of five domains (Figure 1-7), Exon 1 encodes a short 5' untranslated region (5'UTR) and 21 hydrophobic amino acids that are not present in the mature protein. This sequence functions as a signal peptide to direct the receptor synthesizing ribosomes to the ER membrane (50). Exon 2-6 encodes the binding ligands region that are composed of seven imperfect 40 amino acid repeats that are characterized by six conserved cysteine residues and a cluster of negatively charged amino acids near the carboxy terminus of each repeat. Mutations in the ligand-binding domain lead to a LDL-R that is defective in its interaction with apoE and apoB and/or in its transport to the cell surface (51). The ligand-binding domain is followed by a domain of about 400 amino acids (Exon 7-14) that show homology to other proteins such as the C9 component (52), epidermal growth factor (EGF) precursor (53), blood coagulation factor IX, factor X (FX) and protein C (54, 55). Alterations in this domain can lead to reduced transport of the LDL-R to the cell surface or an inability of the LDL-R to bind to ligands or be recycled. This domain contains three 40-amino acid cysteine-rich growth factor repeats (designated A, B, and C). The A and B repeats

(encoded by exons 7 and 8) are separated from the C repeat (encoded by exon14) by five copies of a 40-60 amino acid repeat (encoded by exons 9-13). Each of these five repeats contains a conserved motif, Tyr-Trp-Thr-Asp (YWTD) (42). Exon 15 encodes 58 amino acids that are the short serine and threonine-rich segment that becomes O-glycosylated during maturation of the protein. Exon 16-17 encodes a transmembrane domain of 22 amino acids, while exon 18 encodes a 50 amino acids cytoplasmic tail (51). Deletion or substitution of conserved residues in the cytoplasmic tail results in a receptor that can neither localize in clathrin-coated pits nor mediate the internalization of ligands (51).

Figure 1-7: Schematic of the human *LDLR* gene and LDL-R protein

A cartoon showing the LDL-R structure. A) *LDLR* gene 45 kb encodes 18 exons, B) 5.3 kb *LDLR* mRNA transcript encodes 860 amino acids. 2.7 kb of the mRNA encodes 3' untranslated region (3'UTR), and C) LDL-R protein consists of five domains including ligand binding domain, epidermal growth factor (EGF) precursor domain, O-glycosylated domain, transmembrane domain, and cytoplasmic tail domain. Adapted from Al-Allaf et al. (2010).

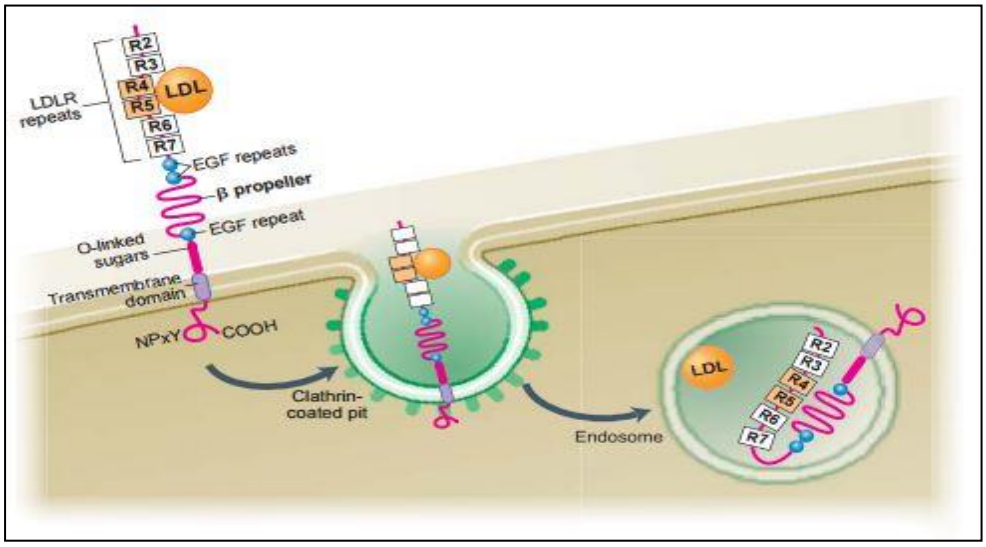


1.6.2.1.2. LDL-R biosynthesis and function

The LDL-R is processed through the ER, matured in Golgi and transported to the cell surface membrane. As illustrated in Figure 1-8, the LDL-C particle binds to the ligand-binding domain of LDL-R at the cell surface. A clathrin-coated pit is then formed and the receptor–ligand complex is internalized to endosomes. The acidity milieu of the endosome leads to release of LDL-C from the LDL-R (57), where then LDL-C is confined to the vesicular part of the endosome, which is finally degraded in the lysosome (58). Concomitant with the release of LDL-C in the endosome, the LDL-R folds back on itself to obtain a closed conformation whereby ligand-binding repeats 4 (residues Cys127 to Cys163) and 5 (Cys176 to Cys210) interact with the β-propeller (Ile377 to Gly642), then the LDL-R recycles back to the cell surface (59).

Figure 1-8: A cartoon diagram showing catch and release of LDL-C from LDL-R

The LDL-C particle binds to LDL-R at ligand-binding repeats R4 and R5 at the cell surface. A clathrin-coated pit is then formed and the receptor–ligand complex is internalized to endosomes. The difference of acidity environment between early endosomes pH 6.5 and late endosomes pH 5.3 causes dissociation of the LDL-C and structural change of the extracellular domain of the human LDL-R. A new structure of the LDL-R leads to ligand-binding repeats R4 and R5 interacting with the β-propeller region of the EGF precursor domain. This interaction displaces LDL-C from the receptor in acidic endosomes. Adapted from Innerarity (2002)

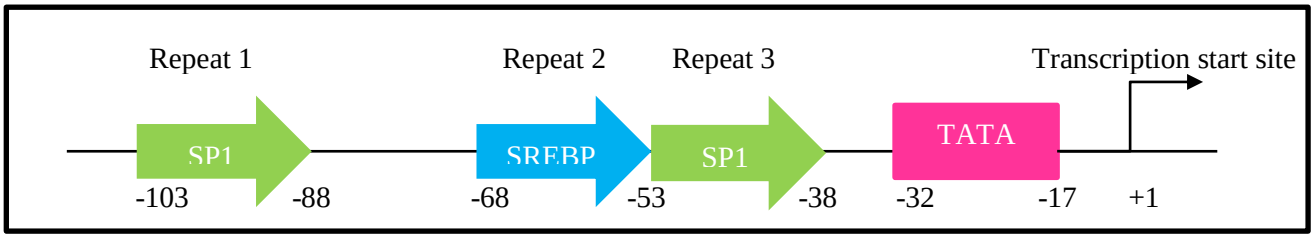


1.6.2.1.3. *LDLR* promoter

The *LDLR* promoter region contains the majority of cis-acting DNA sequences that are responsible for gene expression (Hobbs, et al, 1990). Within 200 base-pairs (bp) of the initiator methionine codon are two TATA-like sequences of 7bp each, and three imperfect direct repeats of 16bp located upstream of a TATA-like sequence (Figure 1-9) (51, 60). Repeats 1 and 3 contain binding sites for the transcription factor specificity protein 1 (Sp1) and contribute to the basal expression of the gene, while repeat 2 contains the sterol regulatory element (SRE-1) (60). At nucleotide position -103 the 5' end of repeat 1 is located, which is considered the major transcription initiation site, while repeats 2 and 3 are located immediately adjacent to each other, beginning at nucleotide position -68 (61). Sequence variations in any one of these three repeats repress transcription. Sterol-dependent repression of transcription is mediated by a 42bp region encompassing repeats 2 and 3 (60). Repeat 3 is a positive element that binds Sp1. However, in the presence of sterols, repeat 2 provides strong repression of repeat 3 (61). Sp1 and SREBP, which bind to a 10bp sequence within repeat 2, cooperate in the repression of repeat 3 (62, 63).

Figure 1-9: Schematic of the *LDLR* promoter

The diagram represents the *LDLR* promoter. The *LDLR* promoter is within 200bp upstream of the *LDLR* transcription start site. The promoter contains a TATA-like sequence domain, and three imperfect direct repeats located upstream of a TATA-like sequence, where repeats 1 and 3 contain binding sites for the transcription factor of Sp1 and repeat 2 contains the SREBP.

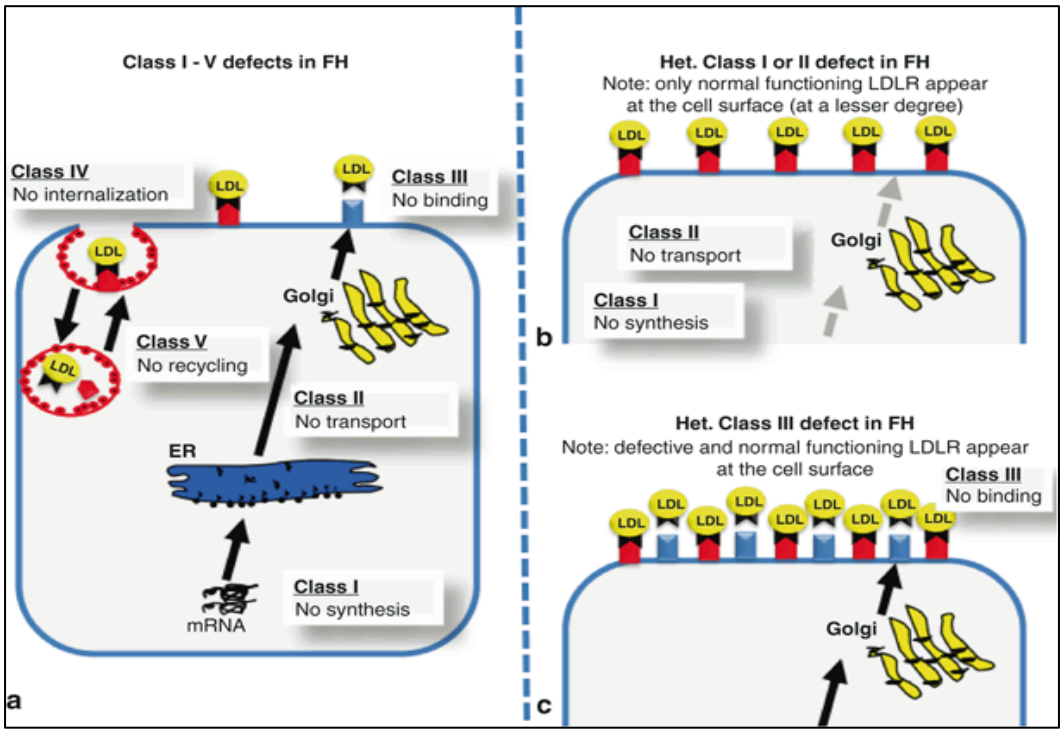


1.6.2.1.4. *LDLR* mutation classes

LDLR mutations are classified into 5 classes according to the effect they have on LDL-R protein function (see Figure 1-10) (51, 64). In class 1 mutations, the LDL-R protein is not synthesized which may be due to a mutation in exon 1. In class 2 mutations, the transport of LDL-R between ER and the Golgi complex is blocked either partially or completely. In class 3 mutations, the LDL-R does not properly bind with LDL particles. In class 4 mutations, bound surface receptors are not internalized. In class 5 mutations, the internalized LDL-C particles are not released in the endosome. Most mutations identified to date are class 2 or class 3 mutations occurring in the ligand-binding and epidermal growth factor precursor regions of the gene. To date, over 1700 *LDLR* mutations have been found to cause FH (65).

Figure 1-10: *LDLR* mutations classification

a) Illustrates the five different classes for defects of LDL receptor function. **b**) The consequences of the different defects on the cell surface for class I and II defects. **c**) Class III defects. Adapted from Schaefer et al. (2012).



1.6.2.2. Apolipoprotein B-100 (ApoB-100)

As mentioned before, LDL-C particles are removed from the bloodstream by the liver and other tissues via their binding to LDL-R via apoB. The *APOB* gene lies on human chromosome 2 p23-p24 and is 42 kb long consisting of 29 exons and 28 introns (66, 67). ApoB mRNA is 14.5 kb long and encodes two main isoforms: apoB-48 and apoB-100. A single peptide of apoB-100, found in each VLDL, IDL and LDL-C particle, acts as a ligand for the LDL-R on the cell surface and results in receptor-cholesterol homeostasis (68). ApoB-48 is a truncated version of apoB and has a different motif to apoB-100; therefore apoB48 is not recognized by LDL receptors. Full length apoB-100 mRNA encodes five structural domains (Figure 1-11 A): NH3-βα1-β1-α2-β2-α3-COOH (69). The β2 domain is the functional region for LDL-R binding that is required for the uptake of lipoproteins containing apoB-100 (70). The interaction between apoB-100 and LDL-R occurs in residues 3359-3367 (site B) in the β2 domain of apoB-100 (71).

Mutations in apoB-100 cause defective binding of apoB-100 to the LDL-R that leads to the failure of hepatic LDL-C clearance and accumulation of LDL-C particles in the circulation (72). This form of FH is called familial defective apo-B100 (FDB). A study by the Soria group (1989) showed that FDB usually occurs due to transition of G to A at nucleotide c.10580, in exon 26 of *APOB*, which leads to the substitution of arginine 3527 to glutamine (p. R3527Q, previously p. R3500Q). This substitution alters the affinity between apoB-100 and LDL-R (73). The p. R3527Q variant in *APOB* is most commonly associated with FDB in Northern Europeans and US Caucasians (74-79). A second mutation has been reported at the same codon (p.3527) resulting in a substitution of tryptophan (W) in place of arginine, which is associated with FDB in East and South

Asians (80-82). An additional mutation at p.3527, R3527L, has also been reported in the Netherlands but has appeared much less frequently than R3527Q or R3527 W (83). Large-scale exomic databases, such as that of the Broad Institute Exome Aggregation Consortium (ExAC), have reported the prevalence of R3500Q and R3500W across populations (see Table 1-1 below).

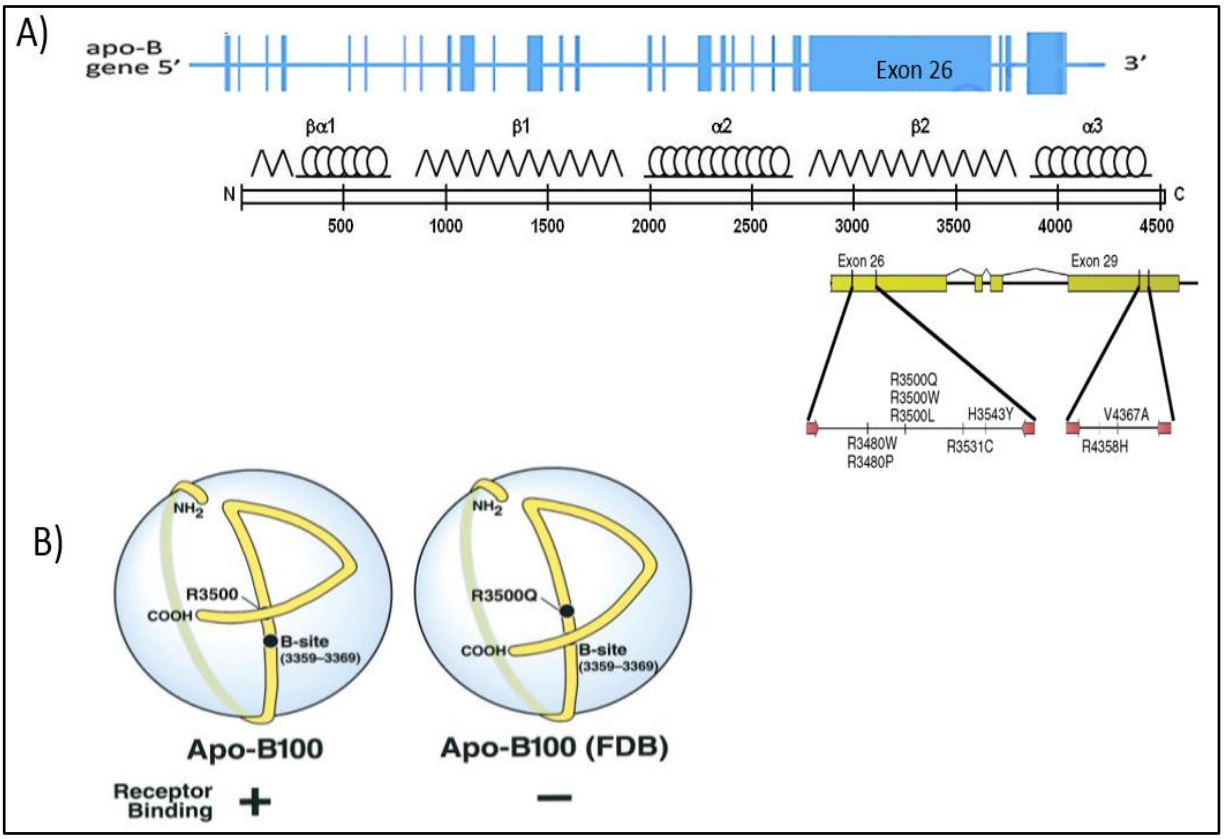
Table 1-1: APOB- R3500Q and R3500W minor allele frequencies (MAFs) across represented subpopulations

Population	R3527Q MAF (%)	R3527W MAF (%)
African	**0**	**0**
East Asian	**0**	**0.127**
European (Finnish)	**0**	**0**
European (non-Finnish)	**0.038**	**0.006**
Latino	**0.009**	**0**
Other	**0.221**	**0**
South Asian	**0**	**0.030**
Total	**0.0321**	**0.01651**

APOB p.3527 is known as a mutation "hotspot", where a mutation at this position alters the stabilizing interaction between arginine residue (wild-type) and a tryptophan residue at amino acid site 4369 (p. Trp4396) near the carboxyl terminus of the protein (Figure 1-11 B), resulting in a conformational change that reduces binding affinity with the LDL-R (84). The substitution of arginine 3527 to glutamine (p. R3527Q) affects LDL-C protein conformation and particle size: the conformational change causes inefficient binding to LDL-R. R3527Q has smaller LDL-C particle size than wild-type ApoB-100, which consequentially affects LDL-C uptake and concentration. Studies of FDB homozygotes have shown that FDB-LDL binding activity is 20% of normal, because the larger LDL-C particles and VLDL retaining ApoE moieties can still bind to LDL-R (85, 86).

Figure 1-11: Schematic of the human *APOB* gene

A) Schematic of the human *APOB* gene and reported *APOB-100* FH ligand-defective mutations. Reproduced with modifications from Sundaram and Yao (2012), and Liyanage et al. (2008). B) Schematic of the organization of apo-B100 on the LDL-C particle. The C-terminus folds back over the preceding structure; an arginine (residue 3500) binds to a tryptophan (residue 4396), which prevents the C-terminus from sliding over the binding site of the LDL receptor between residues 3359 and 3369. Mutation of the arginine (R) (residue 3500) results in reduced binding to LDL-R because the arginine–tryptophan interaction is broken. Adapted from Boren et al. (1998).



Several other variants in *APOB* associated with FH have been discovered, including variants in exon 26 p.(Arg3507Trp (89), p.(Arg3558Cys) (90), in exon 22 p.(Arg1164Thr) (89)and in exon 29 p.(Trp4396Tyr), p. (Gln4494del) (89, 91); however, these variants are less frequent than R3527Q. Functional studies have shown that these variants have similar

but usually a milder effect to R3527Q, where the variants lead to somewhat reduced the binding and uptake of LDL-C (89, 91).

1.6.2.3. Proprotein convertase subtilisin/kexin type-9 (PCSK9)

The third gene causing FH is *PCSK9*, which works as a post-transcriptional inhibitor of *LDLR*.

1.6.2.3.1. PCSK9 ontogeny, biosynthesis and structure

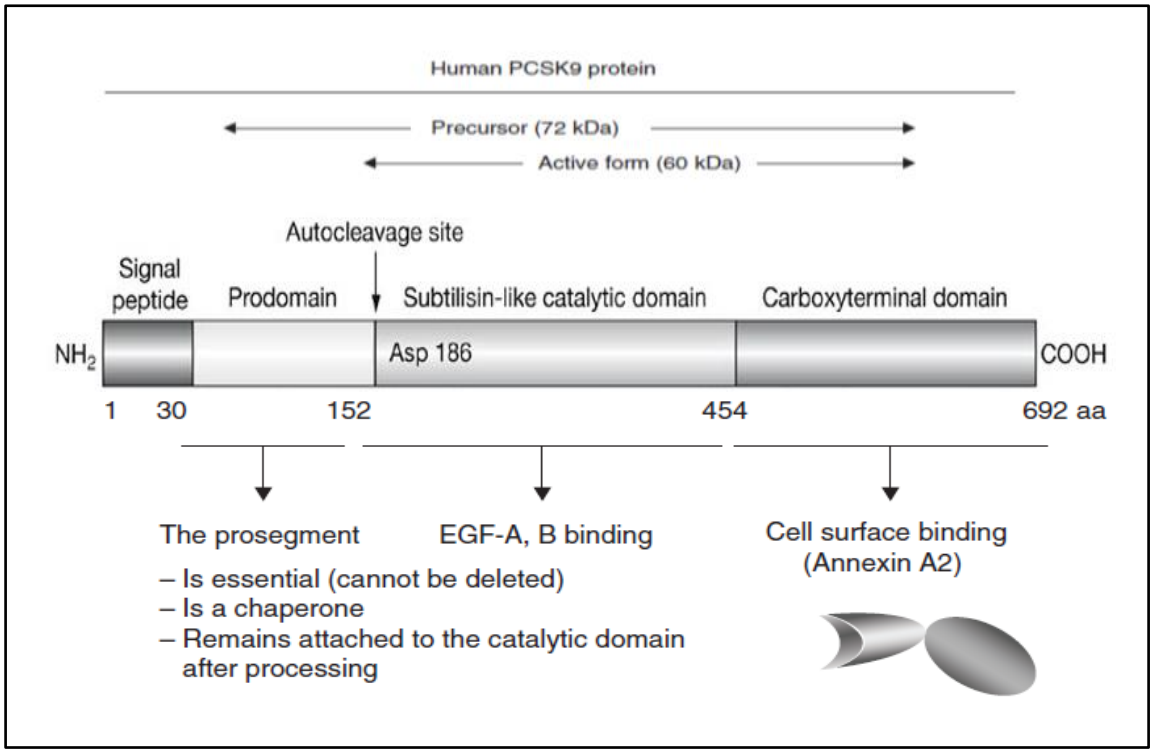
The human *PCSK9* gene is located on chromosome 1p32.3, contains 12 exons and encodes a 692 amino acid glycoprotein belonging to the family of protein convertases (92, 93). It is highly expressed in liver hepatocytes and less in small intestine and kidney. PCSK9 is comprised of four domains (Figure 1-12). The N-terminal signal peptide (aa 1-30) necessary for secretion which is followed by a prosegment domain (pro-domain) (aa 32-152) that inhibits the catalytic activity when cleaved. The third domain is a catalytic domain (CAT) (aa 153-421), which is responsible for autocatalytic intramolecular processing in the ER and is followed by a C-terminal domain (CTD) (aa 440-692): A small 18aa hinge region (aa 422-436) links the catalytic domain to cysteine-histidine-rich domain (CHRD) of CTD. The CHRD is composed of three modules: M1 (aa 453-531), M2 (aa 530-605)and M3 (aa 604-692), and the most histidine residues (9/14 His) are found in the M2 domain (94).

After cleavage of the signal peptide (aa1-30), PCSK9 is synthesized as a ~75kDa soluble zymogen (pro-PCSK9) that undergoes autocatalytic cleavage at position 152 in the ER to release the propeptide (14kDa) from the N-terminus resulting in a mature enzyme of about 60kDa (94, 95). After intracellular autocatalytic cleavage in the ER, the cleaved prosegment

domain is non-covalently bound to the catalytic domain of mature PCSK9 to form a prodomain≡PCSK9 complex and the cleaved pro-domain acts as a chaperone to promote proper folding and leaving ER (93). Thus, the prodomain≡PCSK9 complex is secreted as an enzymatically inactive protein. PCSK9 activity is related to specific target proteins that bind to the catalytic domain and/or CHRD. The first PSCK9 target protein identified was the LDL-R (96-98), where the EGF-A domain of the LDL-R Arg_{194} and Asp_{238} bound PSCK9 in the catalytic domain at Ser_{153}-Ile-Pro-Trp_{156} to direct the complex towards intercellular degradation (99). The common FH-causing gain-of-function variant Asp374Tyr increases the interaction affinity by 25% (100). Also, Annexin A2 (AnxA2) has been identified as a protein that binds to the CHRD of PCSK9 and inhibits its action on LDL-R (101, 102).

Figure 1-12: Schematic of the various domains of human proprotein convertase subtilisin/kexin type-9 protein (PCSK9).

Schematic of the human PCSK9 protein, which is auto-catalytically processed in the ER into a 540-aa (302-aa catalytic subunit and 238-aa C-terminal domain) that remains tightly bound to its 122-aa prosegment (prodomain). The catalytic subunit interacts with the EGF-A,B domain of the LDL-R to activate PCSK9 action on LDL-R; while the CHRD binds to AnxA2 to prevent PCSK9 activity on LDL-R. Adapted from Seidah (2009) with modification.

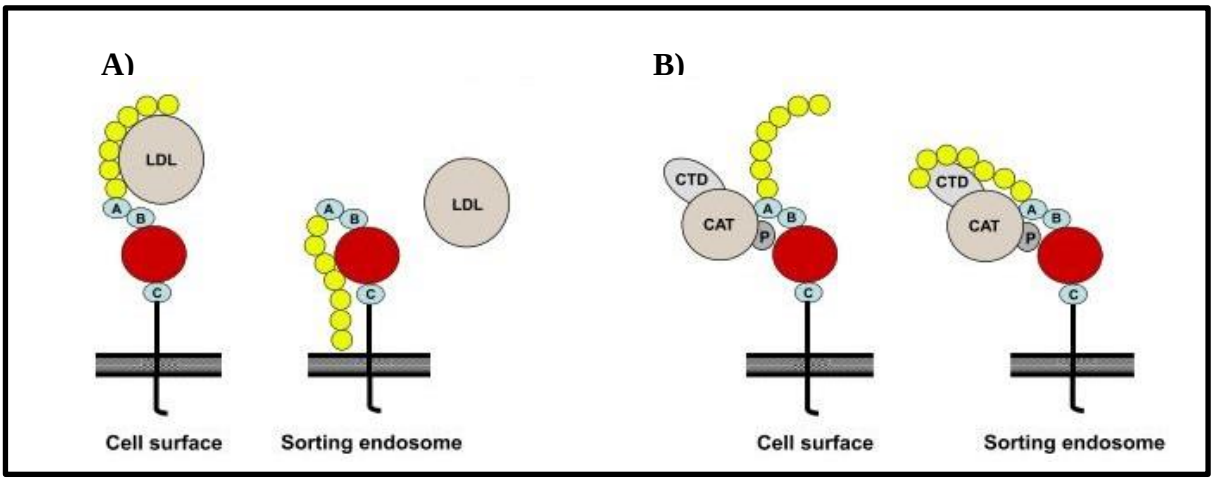


1.6.2.3.2. PCSK9 and LDL-R degradation

At the cell surface, PCSK9 binds the EGF-A domain of the LDL-R via its catalytic domain. Once the non-covalent complex PCSK9≡LDL-R is formed, it is internalized by endocytosis (103, 104). The change of acidity in the sorting endosome environment leads to increased interaction affinity between PCSK9 and LDL-R by 120-150 fold, where the C-terminal domain of PCSK9 obtains an increased positive charge and electrostatically attracts the negatively charged ligand-binding domain of the LDL-R (105). This interaction leads to both the catalytic and C-terminal domain of PCSK9 to bind the LDL-R, which consequently prevents the LDL-R from obtaining a closed confirmation in order to recycle to the cell surface (Figure 1-13) and direct the receptor to degradation.

Figure 1-13: PCSK9 mediated LDL-R degradation

A cartoon showing the function of LDL-R in the presence and absence of PCSK9. A) The normal function of the LDL-R. At the cell surface the LDL-R binds LDL-C through its ligand-binding domain (left). At the acidic pH of the sorting endosome, LDL-C is released from the LDL-R and the LDL-R adopts a closed conformation whereby ligand-binding repeats 4 and 5 interact with the β-propeller (right). B) The abnormal function of the LDL-R due to PCSK9 binding. PCSK9 binds to the EGF-A repeat of the LDL-R at the cell surface through its catalytic domain. In the acidic environment of the sorting endosome, the C-terminal domain of PCSK9 attracts with ligand-binding domain of the LDL-R. This prevents the LDL-R from adopting a closed conformation at acidic pH in the sorting endosome. The ligand-binding domain of the LDL-R with its seven ligand-binding repeats is shown in yellow. The EGF-A, EGF-B and EGF-C repeats of the EGF precursor homology domain are shown in blue and the β-propeller is shown in red. The O-linked sugar domain, the transmembrane domain and the cytoplasmic domain are indicated by a solid black line. The cell membrane is shown as a horizontal structure. PCSK9 pro-domain (P), PCSK9 catalytic domain (CAT) and PCSK9 C-terminal domain (CTD). Adapted from Leren (2014).

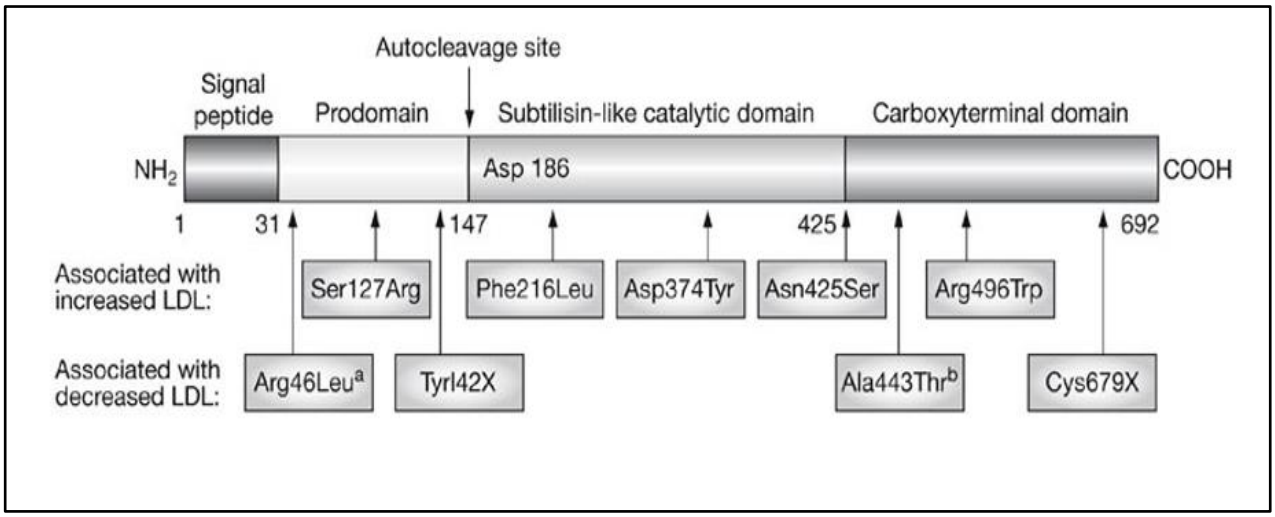


1.6.2.3.3. PCSK9 mutations

Both gain and loss of function mutations have been reported for *PCSK9* resulting in hypercholesterolemia and hypocholesterolemia, respectively (Table 1-2). Gain-of-function (GOF) mutations in *PCSK9* produce FH and increase the risk of CVD, whereas loss-of-function (LOF) mutations have been associated with reduced plasma LDL-C concentrations from birth and decreased CVD risk (92, 106, 107).

Figure 1-14 illustrates the most reported *PCSK9* variants that have been associated with effects on plasma LDL-C levels. The reported LOF mutations, Tyr142X and Cys679X, which are two nonsense mutations (108), and Ala443Thr a missense mutation (109) in the *PCSK9* gene are associated with low plasma levels of LDL-C in African-Americans. The functional disruption of PCSK9 associated with hypocholesterolemia is supported by the finding of reduced cholesterol levels in mice lacking PCSK9 (110). In Caucasians, Arg46Leu is the most common *PCSK9* variant (108), which reduces LDL-C levels by 14% and CHD by 47% (103). It was suggested that Arg46 could affect binding of PCSK9 to the LDL-R in an indirect manner and the presence of the variant may affect PCSK9≡LDL-R binding (111, 112). Interestingly, when the Arg46Leu mutation is found in patients with a clinical diagnosis of FH; the mutation causes mild reduction (non-significant <9%) in LDL-C plasma levels in HeFH patients compared to those without Arg46Leu mutation, which suggests that the effect of Arg46Leu on cholesterol levels may partially hide the strong impact of FH causing mutations in HeFH (112, 113). Another LOF mutation that has been linked to low levels of LDL-C is Q554E in the CHRD domain of PCSK9 which increases the interaction affinity to AnxA2 (101). AnxA2 is a protein that binds to PCSK9 and acts on LDL-R (discussed in detail later).

Figure 1-14: Schematic of some *PCSK9* variants associated with plasma LDL-C levels
Adapted from Soutar and Naoumova (2007).



Several *PCSK9* GOF mutations have been reported: missense mutations at residues 374 and 127 (Asp374Try, Ser127Arg) are found in patients with FH. Molecular studies have shown that the catalytic domain GOF mutation Asp374Try has a great impact and increases the interaction affinity to LDL-R (114), while the pro-domain Ser127Arg is located away from the LDL-R–binding site and has weak or no impact on PCSK9's affinity for the LDL-R, but potently reduces LDL-R numbers (115). It is suggested that Ser127Arg could modulate interactions with the cellular machinery that sorts LDL-R to lysosomes, because this amino acid substitution delays autocatalytic cleavage and secretion (96), and there is only a modest increase in affinity for the LDL-R (100, 116). More reported *PCSK9* variants are listed in Table 1-2.

Table 1-2: *PCSK9* allelic variants with known functional property.

Adopted from (117)

Functional domain	DNA allele	Mutation	Function	References
5′UTR, 2.5-fold increase of transcriptional activity	C−332A		Gain	(118)
Signal peptide	G10A	Val4Ile	Gain	(119, 120)
Prodomain	G94A	Glu32Lys	Gain	(120)
Prodomain	A161C	Glu54Ala	Gain	(120)
Generation of a truncated peptide ($Ala^{68}fsLeu^{82}X$)	202del G	Ala68ProfsX15	Loss	(121)
Prodomain	C230T	Thr77Ile	Loss	(121)
Prodomain	C277T	Arg93Cys	Loss	(120)
Prodomain	del GCC	Arg97del	Loss	(122)
Prodomain	C310T	Arg104Cys	Gain	(120)
Prodomain	G316A	Gly106Arg	Loss	(123)
Prodomain	T341C	Val114Ala	Loss	(121, 124)
Prodomain	T381A	Ser127Arg	Gain	(111, 125, 126)
Prodomain	G385A	Asp129Asn	Gain	(121)
Prodomain	A386G	Asp129Gly	Gain	(111)
Generation of a truncated peptide	C426G	Tyr142X	Loss	(108)
Catalytic	C503A	Ala168Glu	No effect	(111)
Catalytic	G644A	Arg215His	Gain	(125)
Catalytic	T646C	Phe216Leu	Gain	(126)
Catalytic	A654T	Arg218Ser	Gain	(127)
Catalytic	C655G	Gln219Glu	Loss	(120)
Catalytic	G706A	Gly236Ser	Loss	(125)
Catalytic	C716A	Ala239Asp	Loss	(120)
Catalytic	C757T	Leu253Phe	Loss	(128)
Catalytic	A1061T	Asn354Ile	Loss	(125)
Catalytic	G1070A	Arg357His	Gain	(127)
Catalytic	G1120T	Asp374Tyr	Gain	(114, 129, 130)
Catalytic	G1120C	Asp374His	Gain	(131)
Catalytic	C1171A	His391Asn	Loss	(128)
Catalytic	A1274G	Asn425Ser	Gain	(128, 132)
Catalytic	G1284A	Trp428X	Loss	(120)

C-terminal domain	C1300T	Arg434Trp	Loss	(133)
C-terminal domain	G1355A	Gly452Asp	Loss	(120)
Disrupts normal folding of the C-terminal domain	T1384C	Ser462Pro	Loss	(125)
C-terminal domain	C1405T	Arg469Trp	Gain	(120, 127)
C-terminal domain	C1486T	Arg496Trp	Gain	(132)
C-terminal domain	G1540A	Ala514Thr	Gain	(120)
C-terminal domain	G1564A	Ala522Thr	Gain	(121)
C-terminal domain	A1658G	His553Arg	Gain	(128)
C-terminal domain	C1660G	Gln554Glu	Loss	(128)
C-terminal domain	C1847T	Pro616Leu	Loss	(121)
C-terminal domain	G1870A	Val624Met	Gain	(120)
C-terminal domain	C2004A	Ser668Arg	Loss	(120)
Generation of a truncated peptide retained in the ER	C2037A	Cys679X	Loss	(108)

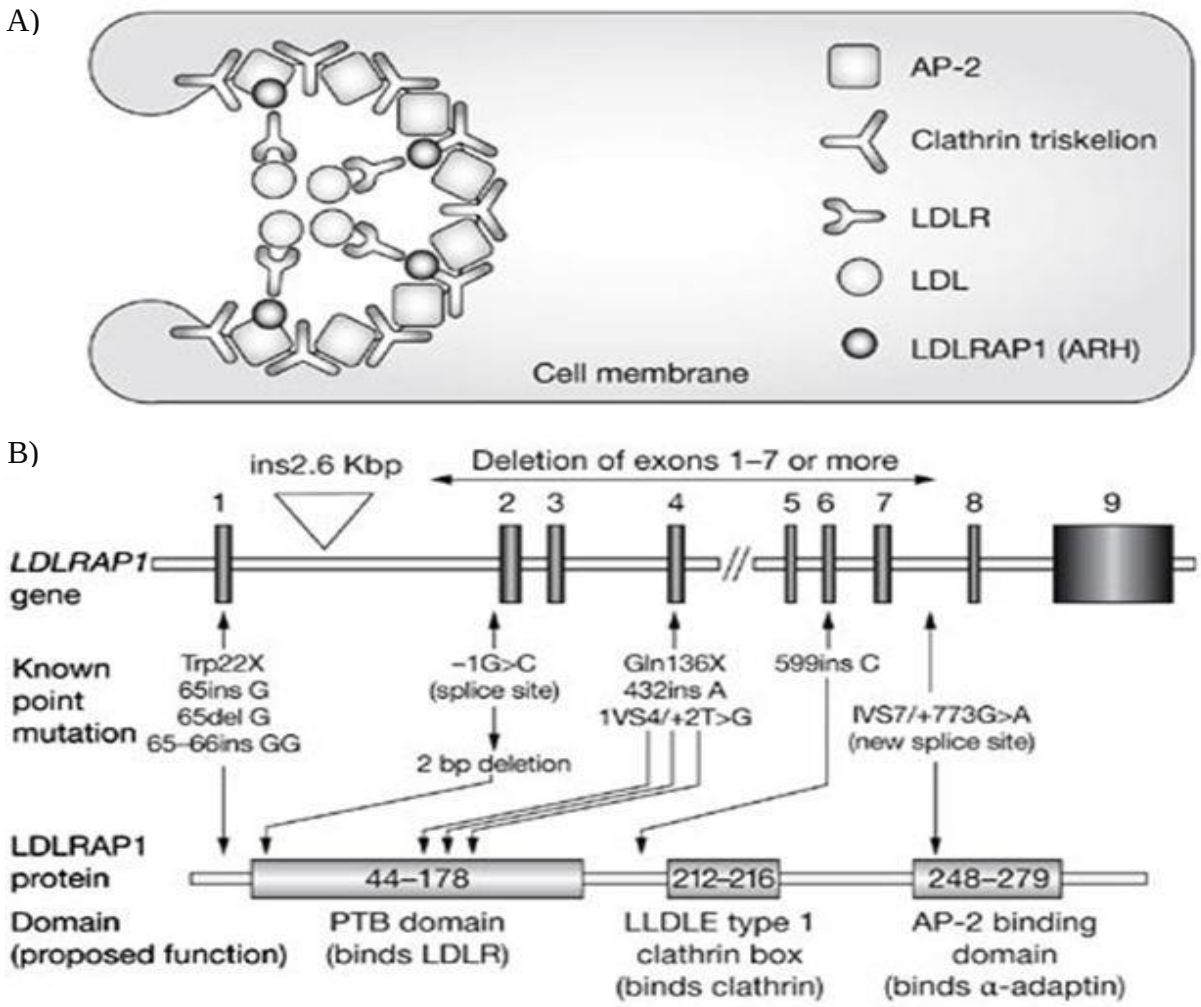
1.6.2.4. Low-density lipoprotein receptor adaptor protein 1 (LDLRAP1)

LDLRAP1 encodes an adaptor protein, which is necessary for internalization of the LDL-C≡LDL-R complex and for efficient binding of LDL-C to its receptor (Michaely, Li et al. 2004). A study by Zuliani and colleagues (134) reported two families of Sardinian origin having very similar HoFH characteristics, including severely elevated plasma LDL-C, tuberous and tendon xanthomata, and premature atherosclerosis, but they showed a recessive inheritance pattern in contrast to the usual dominant pattern. This new disorder is called autosomal recessive hypercholesterolaemia (ARH). Mapping the ARH locus revealed that it occurs when two defective copies of *LDLRAP1* are inherited, which leads to failure in the internalization of the LDL-C≡LDL-R complex and in LDL-C clearance (135).

The *LDLRAP1* gene, located on human chromosome 1p35–36.1, consists of 9 exons and 8 introns (136). It contains a phosphotyrosine binding (PTB) domain (Figure 1-15), which binds to NPXY motifs in the cytoplasmic tails of cell surface receptors, including the LDL-R, which promotes LDL-R endocytosis through the clathrin-coated pits on the cell surface (135, 137). Usually functional *LDLRAP1* mutations introduce a premature stop codon and results in a truncated non-functional protein.

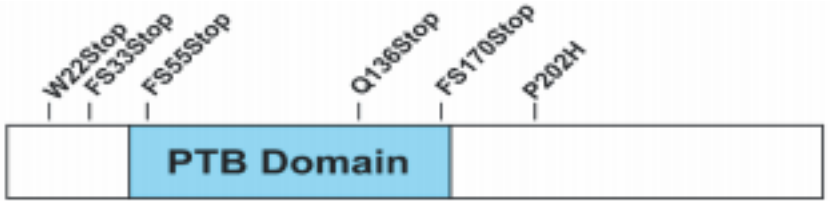
Figure 1-15: Schematic of LDLRAP1 structure and function.

A) LDLRAP1 functions to engage the LDL-R using clathrin-coated pits for endocytosis. B) Diagram of the *LDLRAP1* gene. Exons are shown as vertical bars numbered above, with the position of known point mutations shown below and two large gene rearrangements. All mutations are 'null' and would result in a truncated protein, as indicated in the protein diagram. The different domains of the LDLRAP1 protein and their proposed function on the basis of studies *in vitro* are shown. Adapted from Soutar and Naoumova (2007).



The Garcia group (2001) studied six families of Sardinian, Lebanese, American and Iranian origin with ARH and described six different mutations (Figure 1-16). Two different mutations were identified in the Sardinian families that caused premature termination of translation. One family was homozygous for a single base-pair insertion in exon 4 (c.432insA, p. (170Stop)) that introduces a premature termination codon at amino acid 170, truncating the protein in the terminal portion of the PTB domain, while the other family was homozygous for a nonsense mutation at codon 22 (c.65G>A, p. (W22X)). In Lebanese families, one family was homozygous for a nonsense mutation in codon 136 (c.406C>T, p. (Q136X)), which stops translation in the terminal region of the PTB binding domain. The second family was homozygous for a missense mutation substituting a histidine for proline at amino acid 202 (c.605C>A, p.(P202H)), which is outside the PTB domain. Two other families from USA, and Iran were homozygous for different frameshift mutations located in a string of seven guanine residues in exon 1 (c.72insG. p.(33Stop) and c.71delG, p.(55Stop) respectively). Both mutations predicted truncation of the protein near the NH2-terminus.

Figure 1-16: Schematic representation of ARH showing the location of the mutations identified by the Garcia group (2001) study

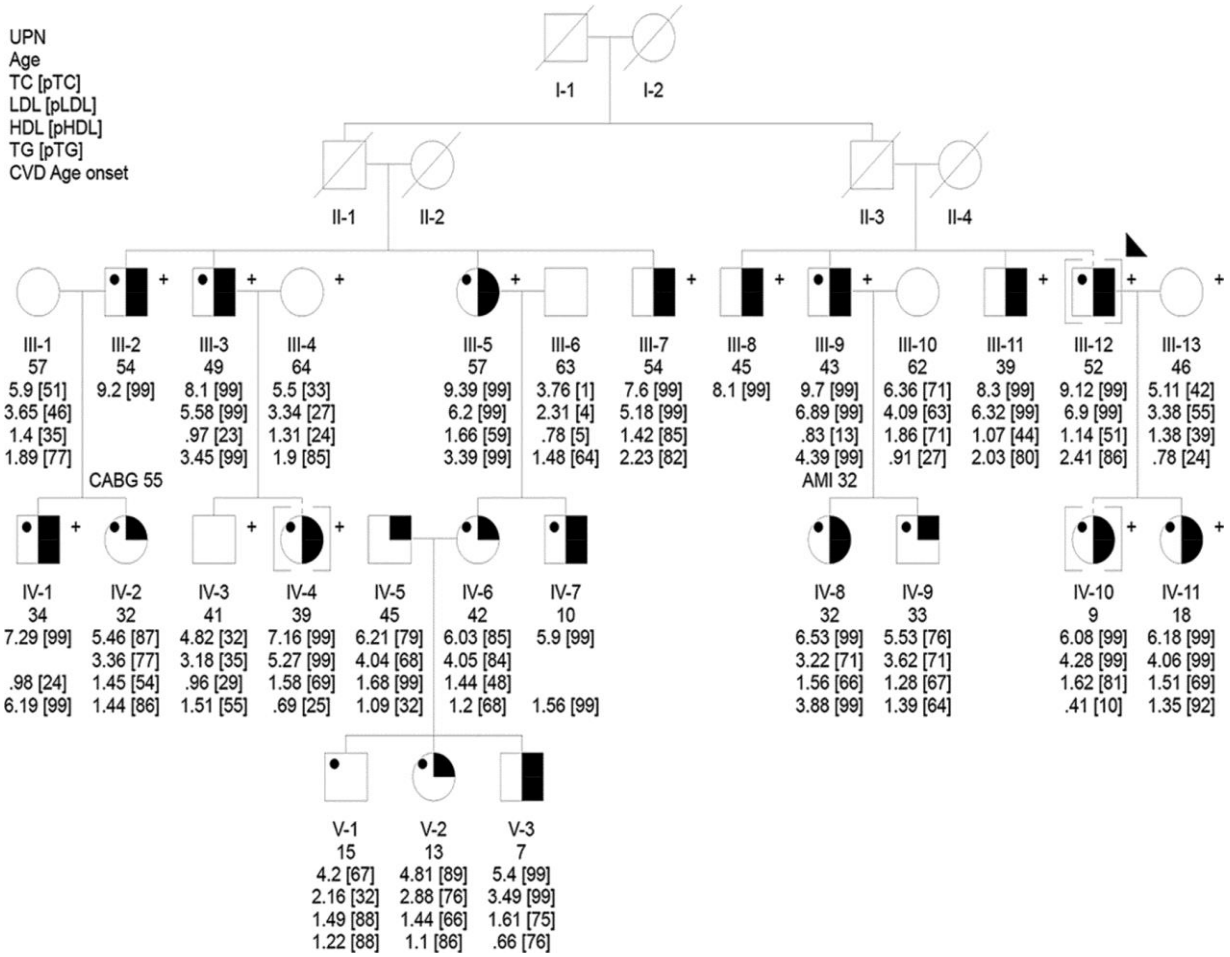


1.5.2.5. Signal-transducing adaptor protein 1 (STAP1)

More recently, a new gene, *STAP1*, was identified as the fourth gene associated with autosomal dominant hypercholesterolemia (ADH). Fouchier and his group (2014) identified a Dutch family who had FH phenotype but were mutation-negative for *LDLR/APOB/PCSK9*. Exome sequencing of chromosome 4 resulted in identification of the mutation p. Glu97Asp in exon 3 of *STAP1*. Co-segregation analysis of the mutation was carried out by screening three generations of the family. Results showed that 11 members were affected (TC or LDL-C levels > 95th percentile), 4 members were possibly affected (75th <TC or LDL-C levels <95th percentile), and one member was unaffected (TC or LDL-C levels <75th percentile (Figure 1-17). Further analysis of 400 unrelated FH subjects resulted in the identification of four rare variants in *STAP1* (Glu97Asp, Leu69Ser, Ile71Thr, and Asp207Asn) in ADH patients with no FH mutations (47). It was observed that patients with rare variants of *STAP1* have significantly higher TC and LDL-C plasma levels compared to those who are *STAP1* mutation negative. Although the lipid phenotype of *STAP1* patients is less severe than *LDLR* mutation patients, it is similar to FDB patients. The identified mutations are located in exon 3 and 6, the affected amino acids are highly conserved and were predicted to be pathogenic.

Figure 1-17: Co-segregation analysis of the *STAP1* mutation p. Glu97Asp

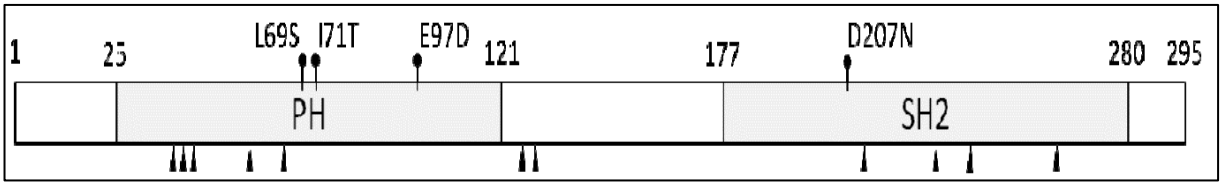
Half-blackened symbols indicate affected members (>95th percentile), quarter-blackened symbols indicate possibly affected members (between 75th and 95th percentile), and white symbols indicate unaffected members (<75th percentile). A dot indicates carriers of the p.Glu97Asp variant. The triangle represents the proband. Age (in years), lipid measurement[total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG)] in mmol/L, and history and age onset of coronary vascular disease are given. Percentiles matched for age and sex, pTC, pLDL, pHDL, and pTG are given in brackets, a percentile indicated by 99 resembles a >95th percentile. Individuals with mapping data are marked with +. For individuals with exome data, symbols are marked with brackets. AMI indicates acute myocardial infarction; CABG, coronary artery bypass graft; CVD, coronary vascular disease; and UPN, unique personal number. Adapted from Fouchier et al. (2014).



The *STAP1* gene, located on human chromosome 4p13.2, consists of 9 exons and 8 introns and contains a Pleckstrin homology domain, a Src homology 2 domain, and several tyrosine phosphorylation sites (Figure 1-18) (138). Fouchier et al. (2014) briefly described STAP1 as having three variants – Glu97Asp, Leu69Ser, and Ile71Thr – located in the Pleckstrin homology domain. This is believed to be a phosphoinositide-binding domain that facilitates the association of STAP1 with membranes, thus alterations in this domain may affect the interaction of STAP1 with membrane or membrane proteins. The fourth variant, Asp207Asn, is located in the Src homology 2 domain, which is the domain responsible for signal transduction emanating from upstream receptor tyrosine kinase which includes *c-kit* and *c-fms*. The STAP1 functioning downstream of *c-kit* controls cholesterol levels. An *in vivo* study showed that a LOF mutation in *c-kit* leads to increased plasma cholesterol levels (139, 140). Fouchier et al. (2014) suggested that STAP1 might be involved in cholesterol homeostasis.

Figure 1-18: *STAP1* gene structure

PH: Pleckstrin homology domain, SH2: Src homology 2 domain. The tyrosine residue positions are indicated by arrowheads.



1.6.3. Diagnosis of FH

1.6.3.1. Clinical diagnosis of FH

Clinical signs and family history are important in diagnosing FH. Currently, there are two clinical diagnostic criteria that are widely used. First, the Simon Broome Register Criteria in the UK distinguishes two types of the disease: definite FH (DFH), and possible FH (PFH). Both DFH and PFH patients should have a total cholesterol greater than 7.5 mmol/l, and/or a LDL-C greater than 4.9 mmol/l (Table 1-3). A positive family history of MI is also a criterion. In addition, DFH patients have observable tendon xanthoma around their joints and/or arcus cornealis (Figure 1-19). Presence of a mutation in one of the known FH genes also classifies a patient as DFH (141). In 2008, the National Institute for Clinical Health and Excellence (NICE) guidelines for FH identification and management recommended that all patients with clinical and biomarkers of hypercholesterolemia should be offered molecular diagnosis to confirm their diagnosis(142).

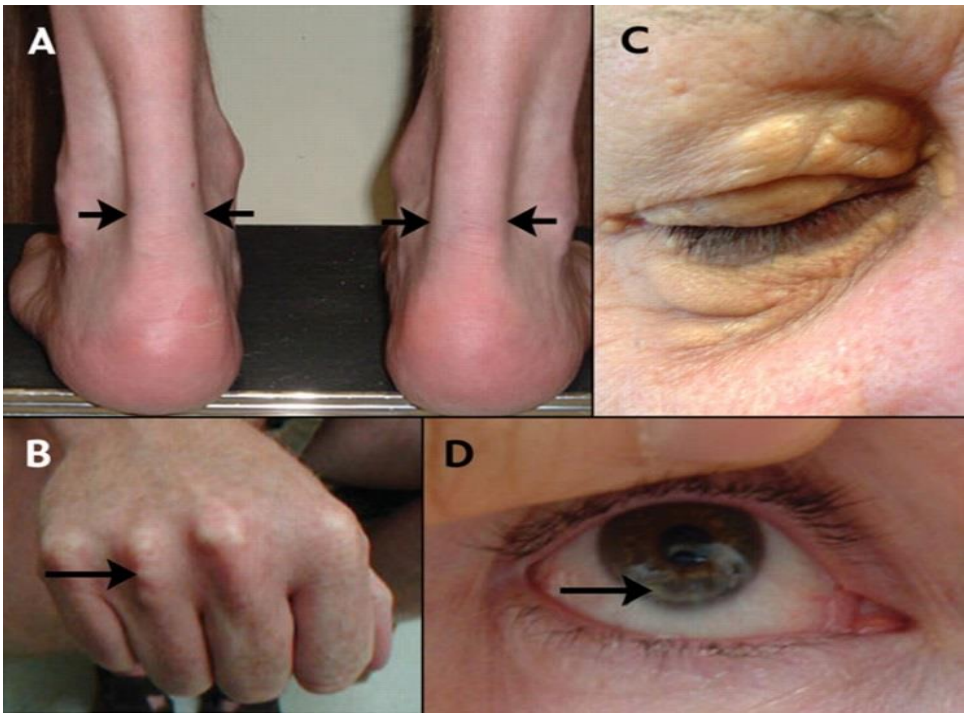
Table 1-3: Simon Broome Register Criteria to diagnose familial hypercholesterolemia in adults and children.

Adapted from Marks et al. (2003).

Simon Broome Register Criteria	
1	TC >7.5 mmol/L or LDL-C>4.9 mmol/l in adult TC >6.7 mmol/L or LDL-C>4 mmol/l in children
2	Tendon xanthomas are present in the patient, first-degree relative or second-degree relative
3	DNA-based evidence of a mutation in *LDLR*, *APOB* or *PCSK9*
4	Family history of premature myocardial infarction (<50 years in second-degree relative or <60 years in first-degree relative
5	Family history of elevated TC (>7.5 mmol/l in first-degree relative or second-degree relative
	Definite FH = 1+2 or 3 Possible FH = 1+4 or 5

Figure 1-19: Physical signs of heterozygous familial hypercholesterolemia.

Tendon xanthoma can occur in several places, they are shown with arrows. A: lateral borders of thickened Achilles, B: in hand, C: in eyelids, D: in arcus corneal. Adapted from Yuan et al. (2006).



The Dutch Lipid Clinic Network Criteria (DLCNC) is the second diagnostic tool used to diagnose FH. It has similar features to the Simon Broome criteria. In addition to family history of premature CHD, LDL-C levels, presence of tendon xanthoma, and DNA analysis, DLCNC diagnosis adds to the calculation of a numeric score (WHO-Human Genetic, 1999) to distinguish between three classes of patients: definite, probable, and possible. A diagnosis is considered definite if the score is greater than 8, considered probable if the score is between 6 and 7 points, and considered possible when the score is between 3 and 5 points. If the score is below 3 points, a diagnosis of FH is not made (144). The DLCNC criteria for FH are shown in Table 1-4.

Table 1-4: Dutch Lipid Clinic Network diagnosis score of FH

Adapted from Marks et al. (2003).

Criteria			Score
A- Family history			
First-degree relative with known premature (>/55 years men; >/60 years women) coronary and vascular disease.			1
First-degree relative with known LDL-C _/95th percentile			
First degree relative with tendon xanthomata and/or arcus cornealis			2
Children below 18 years with LDL-C_/95th percentile			
B- Clinical history			
Patient has premature (>/55 years men; >/60 years women) coronary artery disease			2
Patient has premature (>/55 years men; >/60 years women) cerebral or peripheral vascular disease			1
C- Physical examination			
Tendon xanthomata			6
Arcus cornealis below 45 years			4
D- Laboratory analysis			
	mmol/l	**mg/dl**	
LDL-C	>8.5	>330	8
LDL-C	6.5-8.4	250-329	5
LDL-C	5.0-6.4	190-249	3
LDL-C	4.0-4.9	155-198	1
E- DNA analysis			
Functional mutation in the low-density lipoprotein receptor present			8
	Total Score		
Definite FH	>8		
Probable FH	6-8		
Possible FH	3-5		

1.6.3.2. Genetic diagnostic of FH

A DNA-based test can provide a definite diagnosis of FH and is a useful tool to identify affected relatives of a patient. Different screening methods have been adopted due to mutation variety among populations. First, the amplification-refractory mutation system (ARMS) commercial kit (Elucigene FH20) used in the UK, was designed to detect the most common UK FH mutations including 18 *LDLR* mutations, *APOB* p.Arg3527Gln and *PCSK9* p.Asp374Tyr (145) . It was reported that the ARMS assay detected over 44% of the UK mutations, with a prevalence of 12% for *APOB* and 1.7% for *PCSK9*. Because of the wide FH mutation spectrum in the UK another method was developed, high-resolution melting (HRM) of the coding region of *LDLR* followed by Sanger sequencing. It is an effective method for detecting mutation involving single base changes or small deletions, however, its sensitivity decreases in some regions of the gene depending on the nucleotide composition of the fragment (146). The third method is the multiplex ligation-dependent probe amplification (MLPA) assay (SALSA P062 from MRC-Holland), which was developed to detect large deletions, and duplication (147). Although the technique is highly reliable, and allows detection of large rearrangements, which account for up 10% of UK FH mutations (148), it is time-consuming and a costly addition to sequencing. LIPOchip (Progenika Biopharma, Spain) is another commercially available kit to diagnose FH, which is designed to detect 251 FH mutations and can detect point mutations and copy number changes (149). It was reported that LIPOchip detected 66% of FH mutations in Spanish FH diagnosed patients. Although Elucigene FH20 and LIPOchip have been widely used as simple diagnostic tools, in 2012, the National Institute for Health Technology Assessment reported that the tests certainty and sensitivity to detecting FH-causing mutations in patients with a clinical diagnosis of FH were variable between studies. It found that Elucigene FH20 sensitivity ranged from 44% to 52% while LIPOchip ranged from 33.3% to 94.5%. In addition, these tools as targeted tests are

designed to detect a limited number of genetic mutations, thus they cannot detect all FH mutations, where over 1700 *LDLR FH* mutations have been reported (65, 150) and Elucigene FH20 and LIPOchip can only detect 20% of FH mutations. It became important for a new method to emerge to diagnose FH that considers cost-effectiveness, specificity and sensitivity. Next Generation Sequencing (NGS) technology has revolutionized the genomic field. Implementing NGS into molecular diagnostic labs has become essential due to time efficiency, cost-effectiveness, and specificity. NGS and targeted NGS (exome sequencing) have been used as diagnostic tools of FH (151-154), with 100% specificity and 98-100% sensitivity (155).

1.7. Polygenic hypercholesterolemia

Hypercholesterolemia is a quantitative phenotype, and is complex in etiology, with multiple environmental and genetic causes. The genetic contribution is usually considered as a single gene-familial disorder, and in FH patients, mutations can be detected in 60-80% of those with a clinical diagnosis of DFH and 20-30% with clinical PFH in the three key genes, *LDLR*, *APOB*, and *PCSK9 (46, 145, 148)*. These data show that roughly 60% of patients who are diagnosed with FH are negative for any FH-causative mutation. Failing to identify a new FH causing mutation in clinical FH patients suggests that FH could be also a polygenic disease. Polygenic hypercholesterolemia is a condition of uncertain etiology. It may occur when individuals have a negative family history of disease as a result of an unclear mode of inheritance, or relatives may be undiagnosed, or it could be the result of a new autosomal dominant mutation, autosomal recessive inheritance, and incomplete penetrance in autosomal dominant inheritance. In addition, it may occur in a presence of multiple gene variants with mild affect, where accumulations of these common variants lead to an observed phenotype.

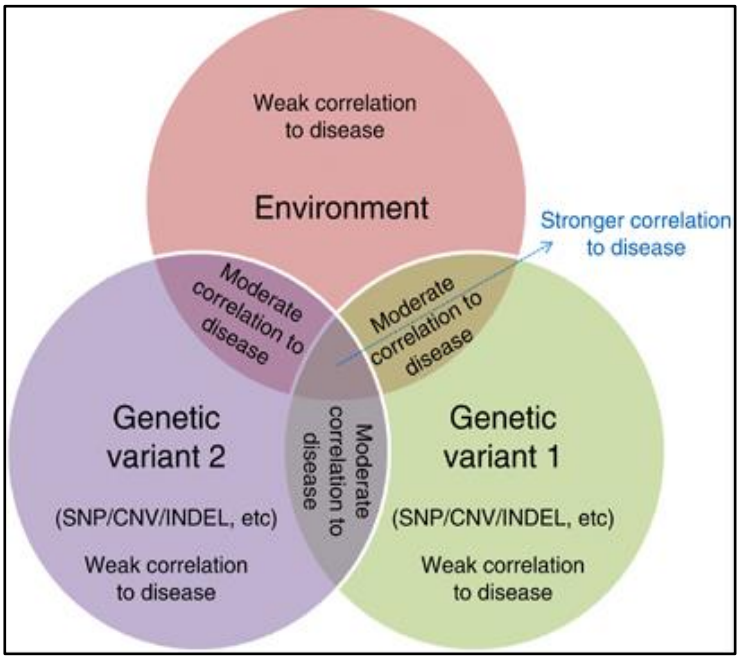
Also, it may occur as a result of the interaction between variants with mild affect and environmental factors which together cause the disease or trait (40).

There are several possible explanations for the clinical finding of a patient with a diagnosis of FH without any FH causative mutation. First, it may occur due to misinterpretation of clinical diagnosis, where patients with 'phenocopy' caused disease are included (156). Secondly, it is possible there are undetected new mutations in known genes, which may be due to low sensitivity levels of current FH-screening methods in some region of the gene(s) or may be due to classical genetic diagnostic methods not covering non-coding regions (introns or upstream or downstream transcription regulatory regions), where causal mutation could be present. The third possible explanation is that the disease-causing mutations might occur in unidentified gene(s). The classical FH mutation screening methods (HRM, MLPA and Sanger sequencing) that were used in clinics until recently only focus on the three known FH genes: *LDLR*, *APOB*, and *PCSK9*. However, identifying FH causative mutations in known genes and new genes becomes possible, faster and cost-effective by using the high-throughput NGS technology (155, 157). The fourth possibly explanation of no detected mutation in some clinical FH patients is the presence of multiple gene common variants with small effect, together leading to the FH phenotype. It has been hypothesized that common complex diseases (such as hypercholesterolemia) are caused by common genetic variants (158), each with small effect, through additive and interactive effects of these variants in conjunction with environment stress (159). Figure 1-20 shows the genetic variations and susceptibility to polygenic diseases. In the last 10 years a new genetic approach called the GWAS has become a major contributor to genetic knowledge, where the GWAS examines the association between variant genotypes and phenotypes in large populations. In 2010 and 2013, large GWAS meta-analysis identified many novel loci associated with lipid traits (160, 161). These

studies reported that SNPs in *LDLR*, *APOB*, *PCSK9* and *LDLRAP1* have significant association with LDL-C, therefore, it was suggested that other genes, within which common variants are associated with moderate effect on LDL-C, may also contain rare FH causing variants. However, to date no new rare variants in GWAS hit loci for LDL-C have been reported.

Figure 1-20: Genetic variations and susceptibility to polygenic diseases

A schematic showing several genetic variants (SNPs) and possible involvement of environmental factors in defining susceptibility to multifactorial diseases. A strong correlation to disease susceptibility can be explained by taking into account the various factors affecting susceptibility to diseases such as environment factors, different types of genomic variants (SNP, CNVs, small insertions and deletions (indels), etc). Adapted from Almal and Padh (2012) (162).

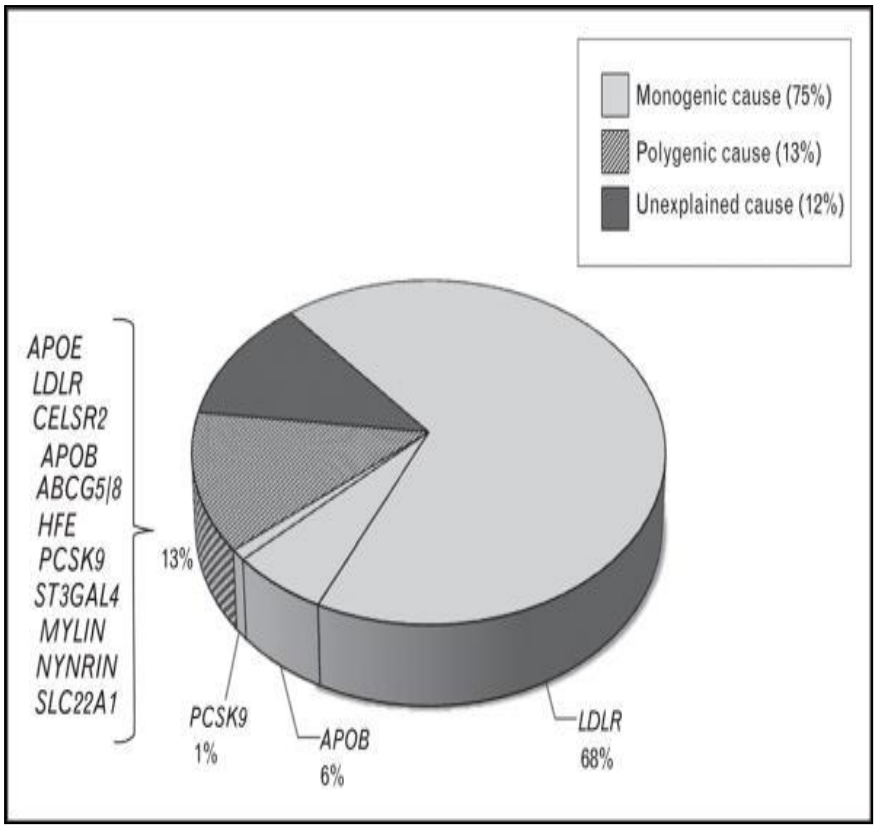


It has been reported that a proportion of individuals carrying several LDL-C-raising SNPs have LDL-C levels that exceed the FH LDL-C diagnostic cut-off 4.9mmol/l (163). In 2013, the same group were used to gene-score in a case/control study to test whether adding a LDL-C associated SNP could help to improve the FH diagnostic. They selected 12 common LDL-

C-raising SNPs identified by Teslovich et al., (2012) to construct a weighted LDL-C-raising gene score, and compared the gene-score distribution among patients with FH and with mutation positive, mutation negative, and controls from the Whitehall II study (WHII). The major finding of the study was that the SNP score for both mutation negative and mutation positive patients with a clinical diagnosisof FH had significantly higher SNP scores than the control population, and the difference between mutation negative and mutation positive also was significant (156, 161, 164). High SNP scores explained hypercholesterolemia in at least 13-20% of patients with no identified mutation; however, ~12% of FH patients who were mutation negative remained with unexplained disease (Figure 1-21). Those patients may carry very rare mutations in known genes or in new gene(s) affecting LDL-C plasma homeostasis.

Figure 1-21: Genetic causes of definite familial hypercholesterolaemia (DFH) in a genetically heterogeneous population.

A familial hypercholesterolaemia-causing mutation can be found in roughly 75% of DFH patients (68% in *LDLR*, 6% in *APOB*, and 1%in *PCSK9*). Approximately half of the patients with no mutation in known FH genes are thought to have a polygenic cause due to common variants in LDL-C-modifying genes, while the other half remained with an unexplained cause. Adapted from Talmud et al. (2014).



1.8. Annexin A2 (AnxA2) a new protein links to LDL-C levels

PCSK9 is the third gene involved in cholesterol homeostasis. It binds to LDL-Rs thus promoting receptor degradation, and thereby modulating LDL-C plasma levels. Several *PCSK9* mutations have been reported; the reported GOF variants have been associated with increased risk of high LDL-C and of CVD. This suggests that PCSK9 is an attractive therapeutic target for lowering circulating LDL-C, especially for hypercholesterolemia patients who under the standard care of statins, fail to reduce their LDL-C levels to intended target levels. One reason could be because statins increase both *LDLR* and *PCSK9* mRNA expression (21). It has been reported that high-dose atorvastatin (80 mg) leads to increased PCSK9 levels in serum over time (165). These findings have shown that PCSK9 is a viable target to reduce LDL-C and thereby reduce the risk of CVD (166, 167).

Pharmaceutical companies have become interested in the introduction of a drug that would target the PCSK9 molecule. PCSK9 monoclonal antibodies have been shown effective in reducing LDL-C levels (168), although their widespread use may be limited by practical issues such as subcutaneous administration and high costs (169). Other drugs that target *PCSK9* mRNAs (ASO, siRNAs, and LNA) are developing. So far, it is unknown whether injection of these compounds is safe for humans, and/or can be produced at a reasonable cost (168). However, Mayer (2008) and his group have identified a small molecule called Annexin A2 (AnxA2), which binds to PCSK9 and inhibits PCSK9 mediated LDL-R degradation. AnxA2, as an endogenous inhibitor for PCSK9, has an advantage as a therapeutic target. An overview of AnxA2 is given below.

1.8.1. ANXA2 ontogeny, biosynthesis and structure

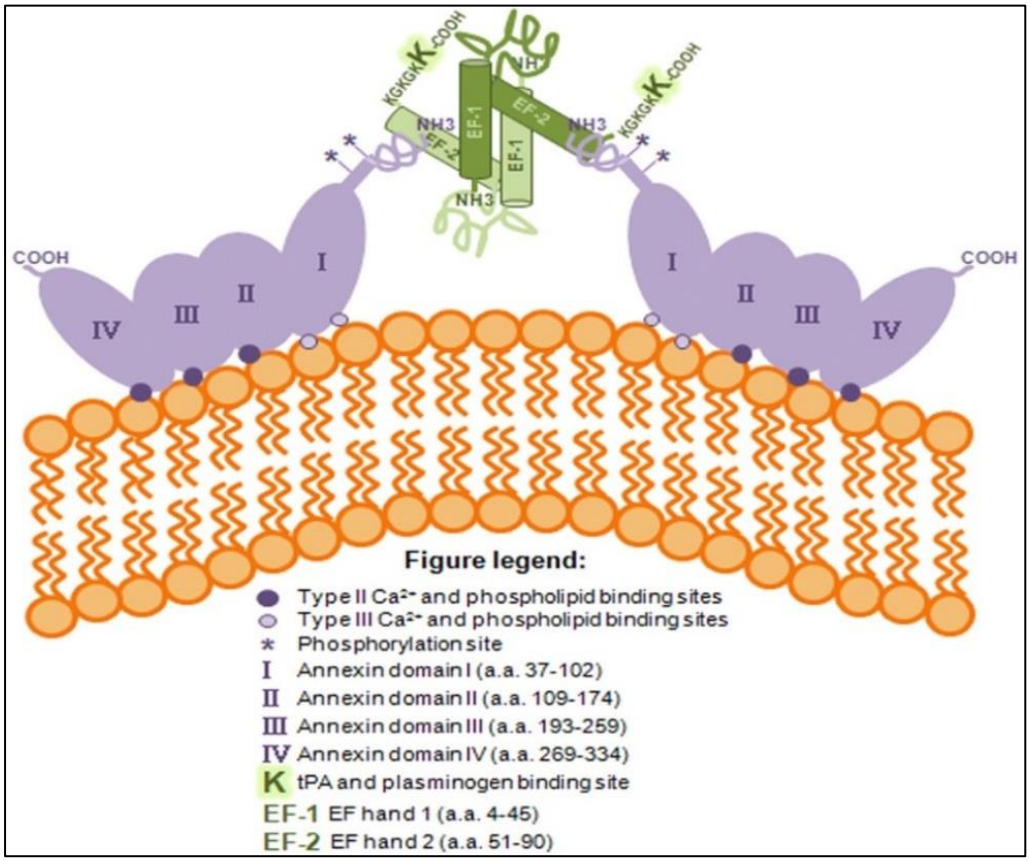
The *ANXA2* gene is located on chromosome 15q22.2 and consists of 13 exons (170). AnxA2 belongs to the Annexin family of phospholipid and calcium-binding proteins, it is a 36kDa protein produced by endothelial cells, monocytes, macrophages, trophoblasts, and some tumor cells, which is found free in the cytoplasm and in association with intracellular and plasma membrane surfaces (171, 172). It exists as a monomer and as a heterotetrameric complex with the plasminogen receptor protein p11 (also known as S100A10) (173, 174). AnxA2 as a member of the Annexins binds to anionic phospholipids in a calcium (Ca^{2+})–dependent manner. It has three distinct functional regions (Figure 1-22): first, a N-terminal region containing tissue plasminogen activator (tPA) (175) and a p11-binding site (176); second, a C-terminal region containing F-actin (177), heparin (178) and plasminogen-binding sites (179); and third, in the core region of AnxA2 has calcium and phospholipid-binding sites. The core domain contains four repeat sequences of 70 amino acid residues composed of five alpha-helices (A–E). The AB and DE helices are connected by loops. Also, it contains several Ca^{2+} binding sites which are present in the AB loops (180, 181).

To combine two molecules of AnxA2 (Figure 1-20), the first 10 amino acids of the N-terminal of AnxA2 form an alpha-helix, the hydrophobic surface of which binds to p11. Four hydrophobic amino acids of the amino terminus of AnxA2 (V3, I6, L7 and L10) form seven points of contact with helix HI of one p11 monomer. Two points of contact with the hinge region of p11 and nine points of contact with helix HIV of the other p11 monomer form a total of 19 contact points between AnxA2 and p11 (182-185). The AnxA2 heterotetramer is a planar, curved molecule with opposing convex and concave sides. The convex side of AnxA2 faces the cellular membrane and contains the Ca^{2+} and phospholipid binding sites, while, the

concave side is oriented away from the membrane and contains both the N- and C-terminal regions (183, 186).

Figure 1-22: Schematic of the cell surface of AnxA2 heterotetramer

AnxA2 complex, the predominant form of AnxA2 and p11 on the cell surface, is composed of two molecules of AnxA2 linked together by a dimer of p11, which is located at the center, with an AnxA2 molecule on each side. The first 10 amino acids of AnxA2 form a helix, which lies in a hydrophobic cleft formed by loop L2 and helix HIV of one molecule of p11 and helix HI of the other p11. The binding of tPA and plasminogen to the complex occurs on the C-terminal lysine of the p11 subunit. Adapted from Madureira (2011).

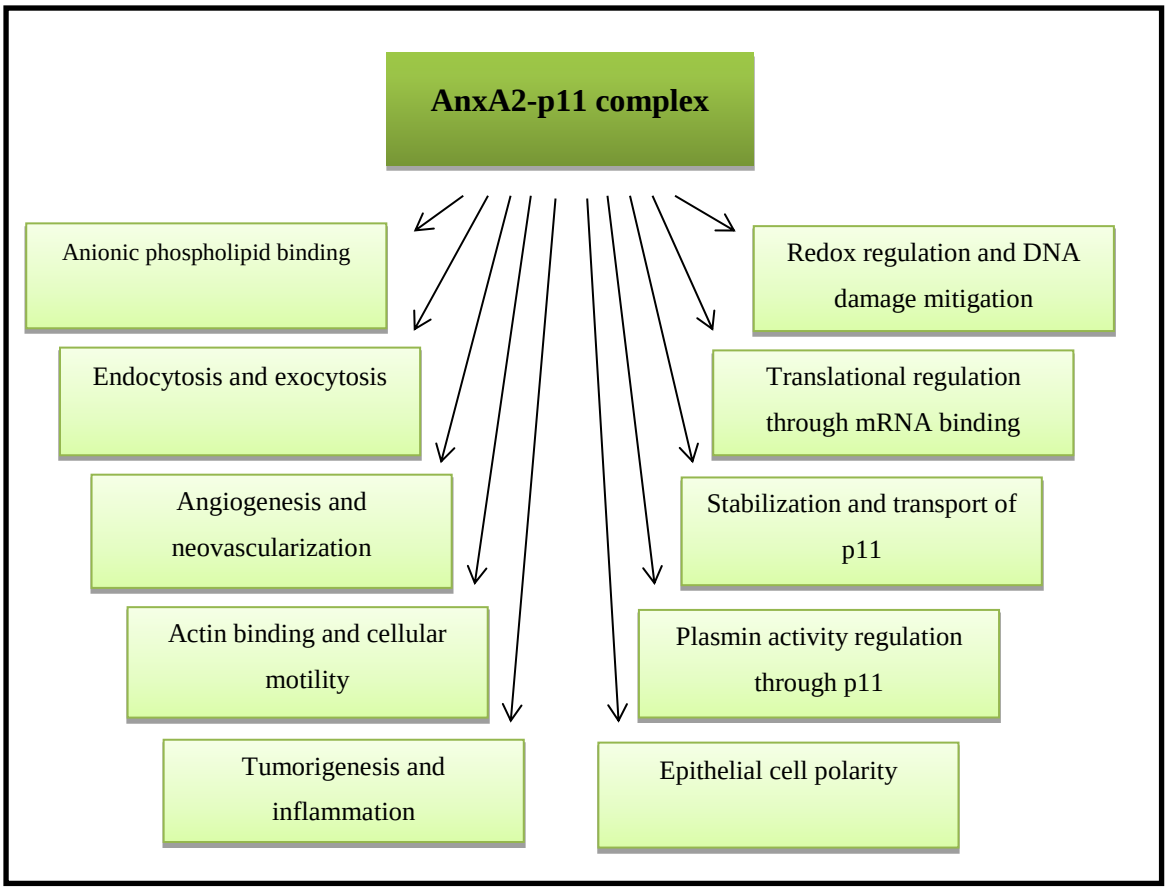


1.8.2. ANXA2 Functions

AnxA2 has been reported to play a key role in many processes including exocytosis, endocytosis, membrane organization, ion channel conductance, membrane reorganization,

vesicular trafficking and redox regulation. Anxa2 impacts over 60 genes as well as control of the cellular levels of other proteins such as an p11 (188). AnxA2 functions are summarized in Figure 1-23.

Figure 1-23: Schematic summary of the functions of AnxA2 heterotetramer.
Adapted from Bharadwaj et al. (2013) with modification.



1.8.3. The relationship between ANXA2, PCSK9 and LDL-R

Studies have shown that PCSK9 activity is related to specific target proteins that bind to the catalytic domain and/or the CHRD. It is known that the catalytic domain of PCSK9 is bound to LDL-R and degrades the receptor. In addition, it has been reported that the CHRD of

PCSK9 is important for intracellular and extracellular LDL-R degradation induced by PCSK9 (101, 189-191). Interestingly in 2008, the Mayer group identified AnxA2 as a protein that binds to the CHRD of PSCK9, where the N-terminal repeat one (R1) of AnxA2 (102) binds to M1 and M2 domains of PCSK9 CHRD (192). An *in vitro* study has shown that AnxA2-derived peptides prevent the PCSK9≡LDL-R interaction while *in vivo* adenoviral AnxA2 overexpression increased hepatic LDL-R levels (102). In addition, it was found that AnxA2 has a role in PCSK9 translation, where cytosolic AnxA2 inhibits PCSK9 mRNA translation and enhanced the sorting of the endocytosed Dil-LDL≡LDL-R complex to lysosomes (192). From these findings AnxA2 is proposed as an endogenous regulator of LDL-R degradation, mostly in extra-hepatic tissue (102).

1.9. Clinical management of FH and hypercholesterolemia

1.9.1. Lifestyle intervention

Healthcare professionals provide lifestyle advice to hypercholesterolemia and FH patients as part of the clinical management of the disease. They ask their patients to improve their lifestyle by having a low-fat diet, increasing physical activity, quitting smoking, losing weight (for overweight patients), and controlling alcohol consumption.

1.9.2. Treatments

1.9.2.1. Statins

The main target of physicians is to reduce LDL-C levels in their hypercholesterolemia and FH patients. The most common cholesterol-lowering drugs are statins, which are used as the primary intervention method to prevent CVD. Statins stimulate expression of the *LDLR* gene by inhibiting HMG-CoA-Reductase activity. Statin therapy reduces LDL-C by 30% to 60%; this variation is caused by the variable potency of the statin used, the type of statin used, the dose, and the type of *LDLR* mutation (20). Table 1-5 shows statin types, dose and reduction levels of LDL-C.

Table 1-5: Statin types, doses and reduction of LDL-C

Adapted with modification from Sharma et al. (2009) and Robinson (2013).

Statins	Low Dose, mg/d	High Dose, mg/d	Reduction levels of LDL-C
Atrovastatin	5, 10, and/or 20	40 and/or 80	30-50%
Simvastatin*	5, 10, and/or 20	40 and/or 80	30-40%
Rousuvastatin	5 and/or 10	20, 40 and/or 80	30-60%
Pravastatin	5, 10, 20 and/or 40	80	30-40%
Fluvastatin	5, 10, 20 and/or 40	80	30-40%
Lovastatin	5, 10, 20 and/or 40	80	30-40%

mg/d = milligram/ day

*Simvastatin has a better result when combined with ezetimibe > 50%

Despite the powerful effect of statins, often the target LDL-C concentrations cannot be achieved with monotherapy in many patients with FH and hypercholesterolemia. In such cases the goal of treatment should be the maximum LDL-C reduction attainable with an appropriate combination of treatment at tolerated doses, for example, use of a statin combined with a cholesterol absorption inhibitor (ezetimibe), a bile acid–binding agent (bile acid sequestrants), nicotinic acid, and fibrates, which are fibric acid derivatives. Ezetimibe combined with simvastatin decreases LDL-C levels by a further 16.5% compared with statin monotherapy (194). The bile acid sequestrants also upregulate the expression of *LDLR* and increase the uptake of LDLs from the circulation, by upregulating Cytochrome P450 Family 7 Subfamily A Member 1 (CYP7A1), an enzyme that is involved in the bile acid synthesis pathway (195). Nicotinic acid decreases LDL-C and triglyceride levels while increasing HDL-C levels in plasma (196). Fibrates decrease triglyceride, increase HDL-C levels in plasma, but have no effect on LDL-C levels directly (197). However, in patients with combined dyslipidemia,

combination therapy of a statin with fenofibric acid show a huge reduction in triglyceride and LDL-C levels and an increase in HDL-C levels than with any monotherapy medication (198).

Ezetimibe, bile acid sequestrants or nicotinic acid can be used as monotherapeutic agents inHypercholesterolemia and FH patients who show intolerance to statins. Young adults and women of childbearing age should not take statins because they comprise molecules that may alter the signaling pathway of the Sonic Hedgehog gene, which could affect embryogenesis and development. Bile acids (e.g., colesevelam) are the alternative treatment recommended during pregnancy (199).

1.9.2.2. Other treatment options

One of challenges that face these conventional and combination therapies is that sometimes they are unable to reduce LDL-C levels to recommended targets in patients with HoFH. The severity of HoFH is based on LDL-C levels (LDL-C >13 mmol/L) (200), which results from the mutation types, heterogeneity [true homozygous (have two identical mutations of the *LDLR* gene), compound heterozygous patients (have different mutations on each copy of the *LDLR*), and double heterozygous (have mutations on two different genes)] and their effects on LDL-C metabolism (200-204). Some mutations in the *LDLR* gene completely eliminate LDL-R function (<2% of normal activity; receptor-negative or null mutations), and others reduce function by up to 75% (2–25% of normal activity; receptor-defective mutations) (205). Patients with receptor-negative HoFH generally have higher levels of LDL-C, respond poorly to conventional therapy, and have more accelerated disease than patients with receptor-defective HoFH (206, 207). Conventional therapies generally are not effective in HoFH

because they predominantly upregulate LDL-R, and HoFH patients lack fully functioning LDLR.

Several new drugs that target different mechanisms of action in addition to LDL-R upregulation are being developed. They effectively reduce LDL-C levels and reduce the risk of CVD events in HoFH patients. These treatment options are discussed below.

1.9.2.2.1. LDL-apheresis

LDL-apheresis (LDL-A) is a blood purification therapy that removes LDL-C from the circulating blood and rapidly reduces the plasma cholesterol level. There are various LDL-A techniques used to reduce LDL-C levels by up to 76%, including dextran sulphate cellulose adsorption (DSA), heparin-induced extracorporeal LDL-C precipitation (HELP), immunoadsorption, double filtration plasmapheresis (DFPP, lipid filtration), and direct adsorption of lipoproteins (DALI) (208, 209). Although the DFPP and DSA systems are highly selective and effective because they eliminate all ApoB-containing lipoproteins, they are obsolete technology. The simple technology of the DALI method effectively adsorbs TC, LDL-C, Lp(a), and TG, is available in different sizes and has good selectivity. Clinical trials have shown that different LDL-A methods are in agreement with the findings of reducing serum lipid concentrations (210-216) and all systems are safe and show equivalent clinical efficacy even during long-term treatment (209).

1.9.2.2.2. Mipomersen

Mipomersen is an antisense oligonucleotide directed against the messenger RNA (mRNA) of apoB, which blocks the hepatic synthesis of apoB protein, thereby reducing secretion of LDL-C (217-219). In a placebo-controlled double-blind trial in 51 patients with HoFH, mipomersen 200 mg administered once/week by subcutaneous injection was added to standard lipid lowering therapies (excluding lipoprotein apheresis) (217). When compared to a placebo, mean reductions of plasma LDL-C, apoB, and lipoprotein(a) (Lp(a)) levels [24.7% ($p = 0.0003$), 26.8% ($p < 0.0001$), and 32% ($p = 0.0013$) respectively] were observed at 26 weeks. An open-label study in 2015 enrolled 141 patients (including 38 with HoFH), who were treated with mipomersen in three different trials. It showed that long-term treatment with mipomersen for up to 104 weeks provided sustained reductions in all atherosclerotic lipoproteins measured, including plasma LDL-C, apoB, and lipoprotein(a) levels [28% , 31% ($p < 0.001$), and 14% ($p < 0.001$) respectively] (218). Both studies (217, 218) showed common adverse events included injection site reactions, flu-like symptoms and elevations in alanine aminotransaminase levels. The second study measured median liver fat and found an incremental increase in the initial 6–12 months that appeared to diminish with continued mipomersen exposure beyond 1 year and returned towards baseline 24 weeks after last drug dose, which was suggestive of adaptation (217).

1.9.2.2.3. Lomitapide

Lomitapide is an oral inhibitor of the microsomal triglyceride transport protein (MTP), which transfers triglycerides and phospholipids onto apoB, thereby playing an important role in the assembly and secretion of apoB-containing lipoproteins (such as VLDL and chylomicrons) in the liver and intestine (220, 221). By inhibiting MTP, lomitapide induces post-translational

degradation of apoB, thereby reducing the secretion of lipoproteins into the blood, resulting in lower serum cholesterol and triglyceride levels (220, 222).

In 2013, lomitapide (5 to 60 mg/day) was added to standard care, including LDL-A, in an open-label, dose-escalation study of 29 patients with HoFH (202) . Of the 23 patients who completed the 26 week treatment (median lomitapide dose, 40 mg/day), mean LDL-C, apoB, and Lp(a) levels were reduced by 50% ($p<0.0001$), 49% ($p<0.0001$), and 15% ($p = 0.0003$), respectively. A similar observation was reported for LDL-C and apoB at weeks 52 and 78, but Lp(a) reduction was not sustained, returning to baseline at 78 weeks. The most common adverse events, reported in 93% of patients, were gastrointestinal in nature, accumulation of liver fat, and elevations in alanine aminotransaminase levels.

1.9.2.2.4. Inhibition of PCSK9

Although statins are the most effective LDL-C lowering drug, they increase PCSK9 levels and are in some patients, are unable to reduce LDL-C levels to recommended levels in patients with severe FH and HoFH (21). As a result, PCSK9 has emerged as a major target for intervention to lower plasma LDL-C levels and prevent or treat atherosclerosis and CVD. In the last few years, several pharmacologic approaches to PCSK9 inhibition have been proposed. Table 1-6 summarizes the proposed therapeutic approaches and their present stage of development.

Table 1-6: Approach and mechanism of PCSK9 inhibition

Adapted from Pitts, R. and Eckle, R. (2014).

Approach	Mechanism	Example
Gene Silencing	• Antisense oligonucleotide • Small interfering RNA (siRNA)	• ISIS pharmaceutical (223) • Santaris pharma (224) • Alnylam pharmaceutical (225)
Mimetic Peptides	• Peptide that mimics EGFA domain of LDLR and binds PCSK9 (226, 227) • Peptide that mimics PCSK9 and binds LDLR (228)	• Genentech
Monoclonal antibodies	• Targeting catalytic domain of PCSK9	• Amgen AMG145 (evolocumab) • Regeneron SAR236553/REGN727 • Pfizer RN316 (bococizumab) • Novartis LGT209 • Genentech MPSK3169A
Small molecule inhibitor	• Target binds to PCSK9	• Natural inhibitor Annexin (101, 102) • Berberine (229, 230) • BMS-962476 (231)

The most clinically advanced mechanism to inhibit PCSK9 is the use of monoclonal antibodies, which prevent the binding of PCSK9 to LDL-R. The first successful development of a monoclonal antibody against PCSK9 was reported in 2009 by Chan et al., when a human monoclonal antibody (mAB) bound to both the catalytic and prodomain sites prevented PCSK9 binding to LDL-R. The mAB was shown to increase LDL-R levels by 1.7 to 2.2-fold. Phase I studies showed significant efficacy in reducing LDL-C levels with minimal side-effects (232, 233). With time, multiple monoclonal antibodies have been developed with completion of phase II trials and ongoing and completed phase III clinical trials (234-236). In 2015, Evolocumab, an anti-PCSK9 human monoclonal antibody, was approved by European Medicine Agency for the treatment of adults and adolescents aged 12 years and over with HoFH, in combination with other lipid-lowering therapies. The combination of Evolocumab and statins lowered LDL-C levels by 59% [(0.78 mmol/l) ($p<0.001$)] and reduced the risk of cardiovascular events 15-20% compared to the "usual care" group who only received statin

therapy (237). Also, another study suggested that combination of Evolocumab with other agents such as Mipomersen and Lomitapide have the potential to further lower LDL-C levels in HoFH patients (207). A recent study in an atherosclerotic mouse model (APOE*3Leiden.CETP mouse model) (238) showed that active immunization against PCSK9 with the so-called AT04A peptide-based vaccine elicited antibodies that effectively bind and remove PCSK9 from the circulation and reduce circulating TC, LDL-C, TG, and biomarkers of inflammation.

Other PCSK9 therapeutic approaches are also proposed, but many are still in the preclinical development stage such as mimetic peptides which bind either to the PCSK9 or LDL-R binding domains, where both reduce or prevent LDL-R degradation. Furthermore, several antisense RNA inhibitors have been used to prevent the action of PCSK9 on LDL-R including antisense oligonucleotides (ASO) (226), small interference RNA (siRNA) (228), and locked nucleic acid antisense oligonucleotide (LNA) (227). In addition, small molecules such as AnxA2 and berberine have been identified, which inhibit PCSK9 action on LDL-R. Berberine, a naturally occurring plant alkaloid, has been shown to decrease PCSK9 mRNA expression and increase LDL-R in both *in vitro* and *in vivo* studies (229, 230). AnxA2, a natural inhibitor, is expressed in many tissues and inhibits PCSK9 degradation of LDL-R (101, 102).

A recently developed small interfering RNA (siRNA) molecule, Inclisiran (ALN-PCSsc) is a long-acting RNA interference (RNAi) therapeutic agent that inhibits the synthesis of PCSK9 (239). The siRNAs bind intracellularly to the RNA-induced silencing complex (RISC), enabling it to cleave mRNA encoding PCSK9 specifically. The cleaved mRNA is degraded

and thus unavailable for protein translation, which results in decreased levels of the PCSK9 protein. In 2016, a study in patients at high risk of CVD due to high LDL-C levels showed that a single subcutaneous injection of inclisiran 300 mg led to a 50% reduction in LDL-C after 90 days. And with a further injection of the same dose at 90 days after the first injection, LDL-C had decreased by 57% after 180 days. Inclisiran is considered to have good overall tolerability and safety (240).

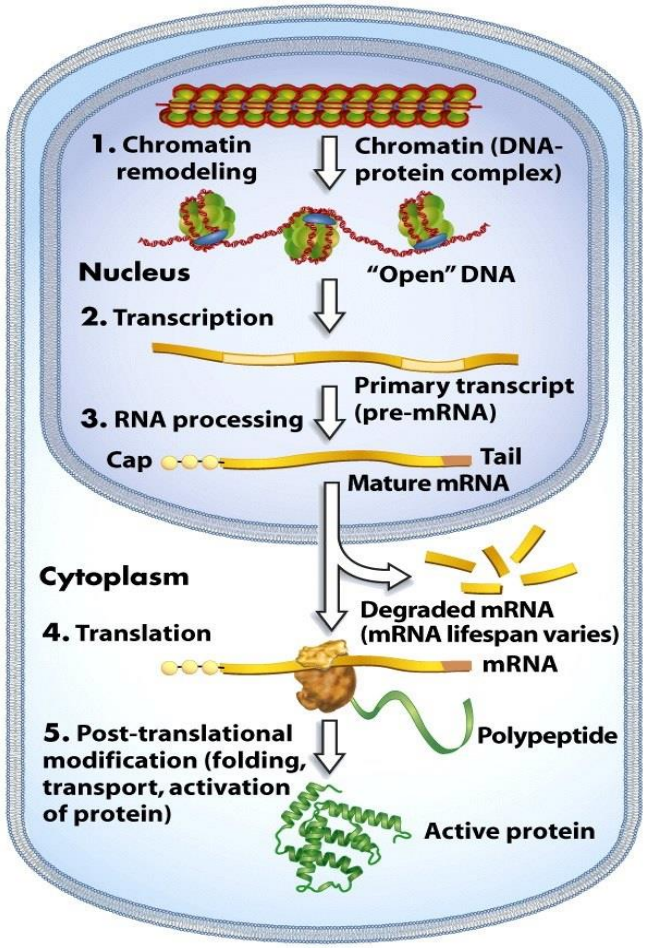
Genome-editing technology has also been used in PCSK9 drug development. A study showed the efficacy and safety of CRISPR-Cas9 therapy targeting the human PCSK9 gene in human hepatocytes in mice using adenovirus or adeno-associated virus (AAV) for delivery. The data showed that post-treatment blood levels of human PCSK9 protein were reduced on average by 52% compared with pretreatment levels ($p=0.007$). This study supports further development of the therapy, with the goal of eventual clinical use for long-term protection against CVD (241).

1.10. Gene regulation

Only 2% of the human genome is coding DNA, while 98% is non-coding DNA. Some non-coding DNA is transcribed into functional non-coding RNA molecules, while others are involved in gene expression regulation. Expression of a gene can be controlled at many levels, including transcription, post-transcription, translation and post-translation. Figure 1-24 shows gene regulation targets in eukaryotes.

Figure 1-24: Gene regulation targets in eukaryotes

A cartoon showing the five different targets of gene regulation: chromatin-remodeling, transcription, post-transcription (RNA processing), translation, and post-translation. Adapted from http://www.uic.edu/classes/bios/bios100/lectures/genetic_control.htm



1.10.1. Transcriptional regulation

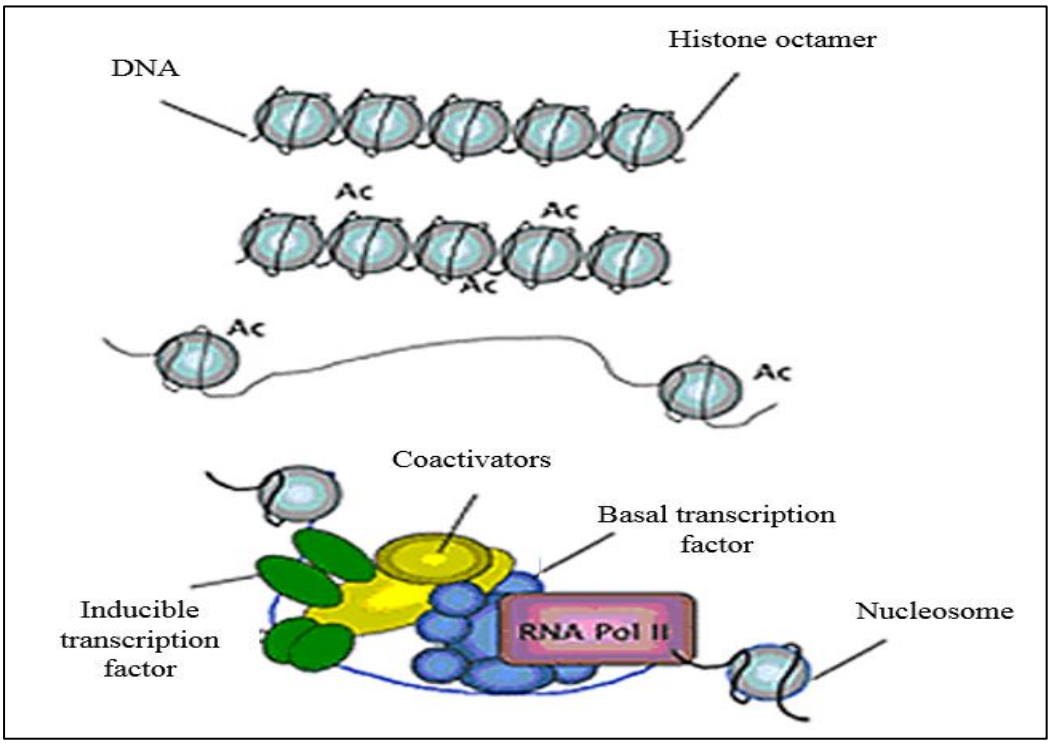
Gene transcription is a process that converts DNA to RNA in order to produce a protein that is involved in biological pathways. The transcriptional machinery is regulated by a combination of three elements including chromatin regulators, Cis-regulatory elements, and transcription factors (TFs). These elements together are the components of the “gene regulation network” (242).

1.10.1.1. Chromatin regulation

In eukaryotic cells, DNA is tightly wound into a complex called chromatin in order to be enfolded and packaged inside the nucleus. Chromatin comprised of octamers of histone proteins, and chromatin structure regulates gene expression through control of the accessibility of RNA polymerase II (RNA pol II) and TFs to the gene. A condensed form of chromatin prevents the transcriptional machinery from accessing the DNA, and consequently, the gene is repressed (243). Therefore, chromatin must have an "open structure" in order for gene expression to take place and this process is called chromatin remodeling (Figure 1-25). Chromatin remodeling occurs as a result of enzymatic modifications of the amino terminal tails of histones which leads to an opening up of the structure. For example, histone acetylation correlates with gene expression and silencing, where certain lysine residues in histones are acetylated leading to open chromatin. In contrast, histone deacetylases close chromatin structure and prevent transcriptional machinery from operating (244).

Figure 1-25: Cartoon of chromatin remodeling.

Chromatin consists of octamers of histone proteins around which DNA is wrapped and packaged inside the nucleus. Enzymatic modifications of histones, e.g. acetylation (Ac) result in remodeling of the chromatin and the opening up of the structure, which is now called euchromatin. The euchromatin allows the transcription machinery to access the DNA and regulate transcription. Adapted from http://www.lerner.ccf.org/pathobio/labs/ stylianou/



1.10.1.2. Transcription factor (TFs)

TFs play a key role in gene expression regulation, but it is important to distinguish between two types of TFs: the basal transcription apparatus and the regulatory factors. The basal transcription apparatus requires that the initiating transcription machinery including RNA pol II and other general TFs are directed to the promoter region, while regulatory factors interact either directly or indirectly with cis-regulatory elements and control the efficiency of transcription via links to promoters (245). Regulatory transcription factors can be divided into two categories; sequence-specific regulators and co-regulators. The sequence-specific regulators interact directly with cis-regulatory elements in a sequence-specific manner. They

comprise two types of regulators – activator and repressor – depending on their role in promoter utilization (246, 247). In contrast, the co-regulators do not interact with DNA directly; they interact with regulatory elements via protein-protein interactions with sequence-specific regulators. Similar to the sequence-specific regulators, co-regulators can be subdivided into co-activator and co-repressor, depending on their role in gene regulation (245, 248). Regulatory transcription factor binding sites are usually clustered and form regulatory modules. Regulatory modules are directed to regulatory elements and are either enhancer or silencer, based on their role in transcription regulation. Many genes have multiple enhancer and/or silencer elements distributed over large regions (100kb or more). The interaction between these elements and regulatory modules are affected by time (development stage), tissue type and environment (245).

1.10.1.3. Cis-regulatory element

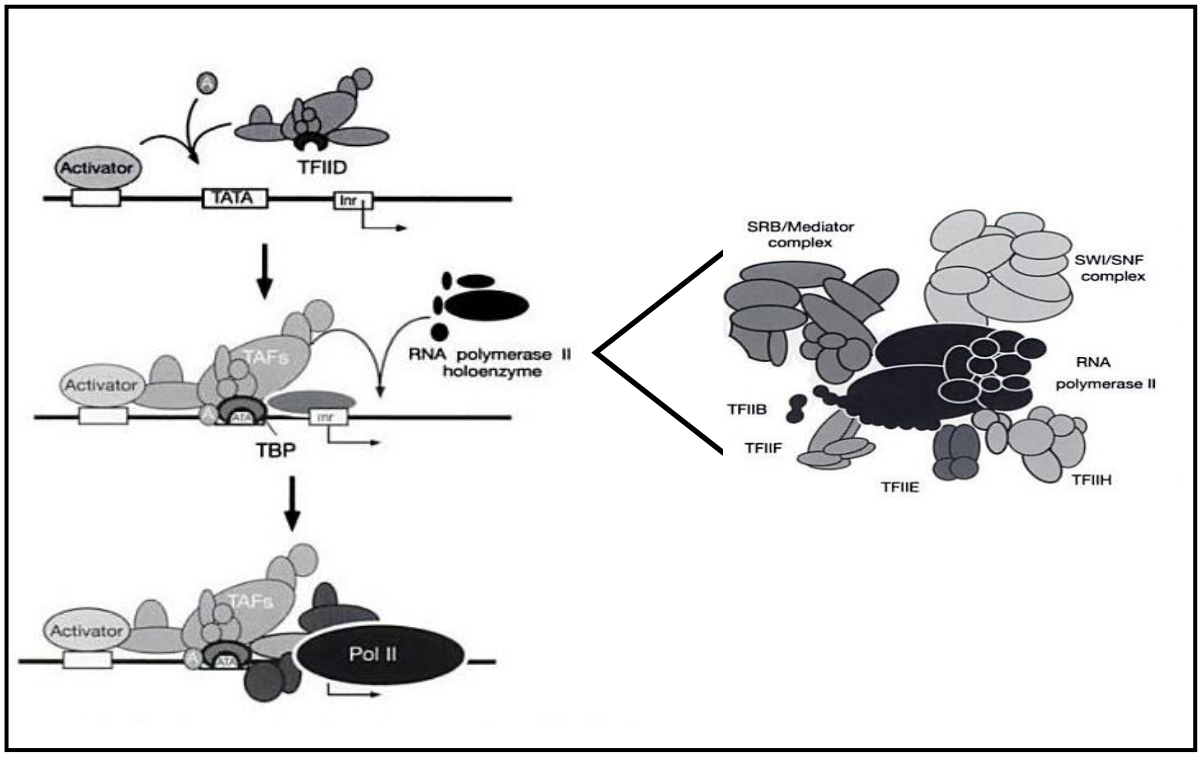
The cis-regulatory elements are located in non-coding DNA regions. They are recognized by sequence-specific DNA binding motifs and are located in the 3'UTR, 5'UTR, and within the introns, and control the transcription of genes (249, 250). The *cis*-regulatory elements including promoter and enhancer sequences have a role in the transcription levels of genes.

The promoter region of a gene contains sequence-specific DNA binding transcription elements that work in the complex to initiate DNA transcription. Promoter regions provide a platform for the assembly of a basal transcription apparatus complex that includes TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and RNA pol II, which work together to specify the transcription start site. Transcription complex formation usually begins with a TATA-binding protein (TBP), a subunit of transcription factor TFIID binding to the TATA box. This recruits

other TFs and RNA pol II holoenzyme (including a variety of distinct complexes that contain the core RNA pol II, SRB/Mediator complex, the chromatin remodeling SWI/SNF complex, and general initiation factors TFIIB, TFIIE, TFIIF, and TFIIH) which together form a promoter-bound complex (251) (Figure 1-25).

Figure 1-26: Promoter regions of a gene and transcription machinery

There are two stages of assembly of the RNA polymerase II transcription complex. Transactional activators bind to the activator sequence of the promoter and recruits transcription factors TFIID and TFIIA to the TATA box. The combined action of this complex recruits the RNA polymerase II holoenzyme, which includes a variety of distinct complexes that contain the core RNA pol II, SRB/Mediator complex, the chromatin remodeling SWI/SNF complex, and general initiation factors TFIIB, TFIIE, TFIIF, and TFIIH. Adapted from Locker (2001).



However, a promoter-bound complex alone is not enough to initiate transcription, as another cis-regulatory element is required, which is called the enhancer element (245). Enhancer elements can be bound with "enhancer-binding proteins" to activate gene

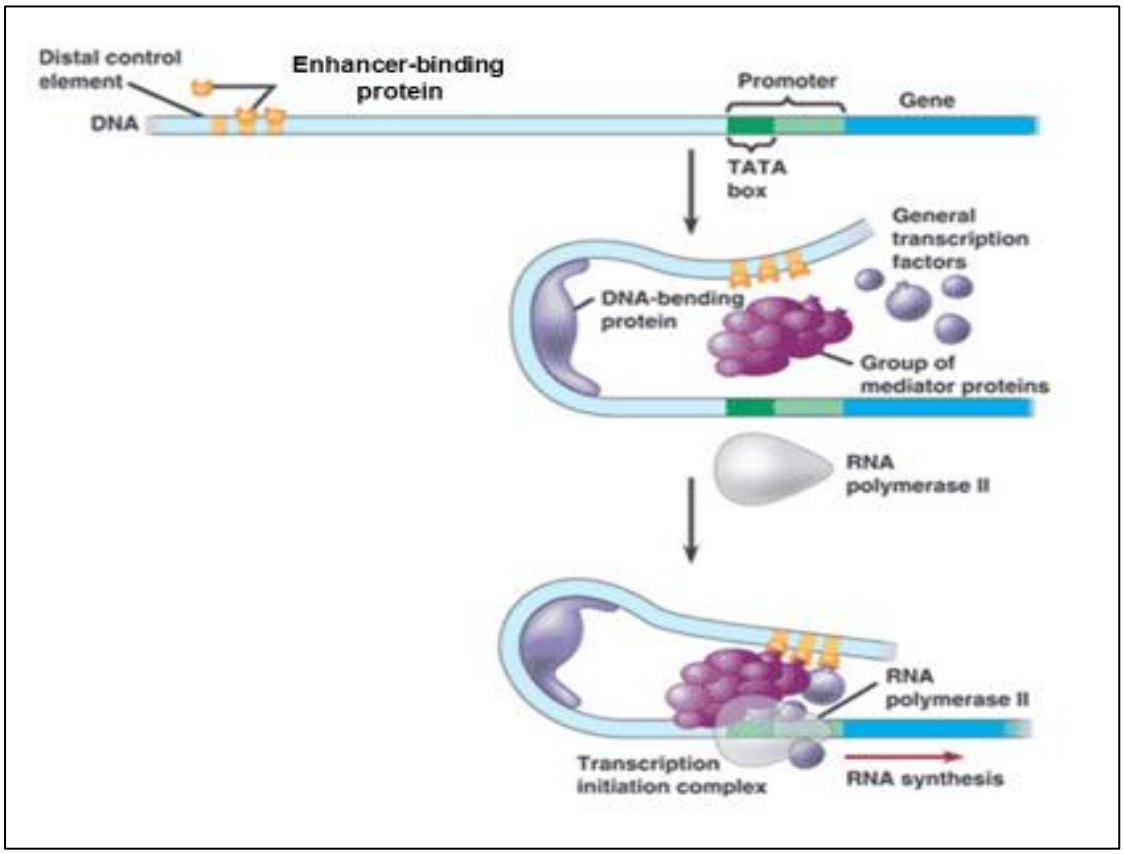
transcription (252). Enhancers can be located upstream, downstream, or within genes they control, and they can function in the normal or inverted orientation (249, 250). Enhancer DNA sequences bind to regulatory transcription factors, which also have sites that bind to a basal transcription apparatus located in the promoter of the gene, and they together activate transcription.

1.10.1.4. Landscape regulatory

Transcription regulation is mediated by "DNA looping", which is generated by a protein or complex of proteins that simultaneously binds to two different sites on the DNA (253). When loops form in the DNA structure, proteins that are bound a distance away from the genes they control can be brought to the promoter. DNA looping brings promoters, enhancers, transcription factors and RNA processing factors together to regulate gene expression, and the DNA loop is stabilized by a complex of TFs (Figure 1-27). For example, the bovine papillomavirus type 1 (BPV-1) gene requires an enhancer protein E2 to initiate gene transcription. However, the BPV-1 promoter has no binding site for this protein, but the promoter has a binding site for Spl which interacts with the distance enhancer protein E2 through DNA looping; this complex in turn initiates gene expression (254).

Figure 1-27: Landscape transcriptional regulation

Schematic showing the transduction of regulatory information from a gene activator protein bound to DNA enhancer element to the pol II transcription machinery at a promoter by mediator proteins. Adapted from http://bio1151.nicerweb.com/Locked/media/ch18/activator.html



TFs and chromatin regulators coordinate the regulation of complex transcriptional networks, and a selective recruitment of these factors to enhancer elements has an essential role in transcription activation (255). Thus, TFs and chromatin regulators are useful tools in identifying enhancer elements. Gene expression patterns can be disrupted by perturbation of TFs. Variants or mutation in the regulatory DNA element including the promoter and enhancer can contribute to disease and changes in trait and then phenotype. GWAS have identified trait-associated SNPs, and has been suggested that a genetic variant can contribute

to disease and/or traits by affecting gene expression activity. For example, the presence of a SNP in an enhancer element may create or disrupt a TF binding site; consequently, enhancer activity and target gene expression are changed. For example, GWAS identified a new locus Sortilin 1 (*SORT1)* that is associated with LDL-C and MI. It was found that the rs12740374 SNP creates enhancer-binding protein for C/EBP transcription factors, which leads to increased *SORT1* expression levels and reduced LDL-C and VLDL levels (256). The C/EBPα is a liver-enriched transcriptional factor that regulates the expression of several hepatic genes involved in metabolic processes (257).

1.10.2. Post-transcription: micro-RNA mediated regulation

In eukaryotes, gene regulation has complex machinery whereby transcription and translation are processed in separate compartments, the nucleus and the cytoplasm, respectively. This additional level of regulation is called post-transcription regulation, where pre-mRNA is processed to produce a mature mRNA. The post-transcriptional process includes capping, splicing, addition of a poly (A) tail, RNA editing, export and mRNA stability. Then, when the mature mRNA is in the cytoplasm, it is translated into protein. In the post-transcriptional stage, a member of a class of non-coding RNAs, termed micro-RNAs (miRNAs), are involved in regulating gene expression.

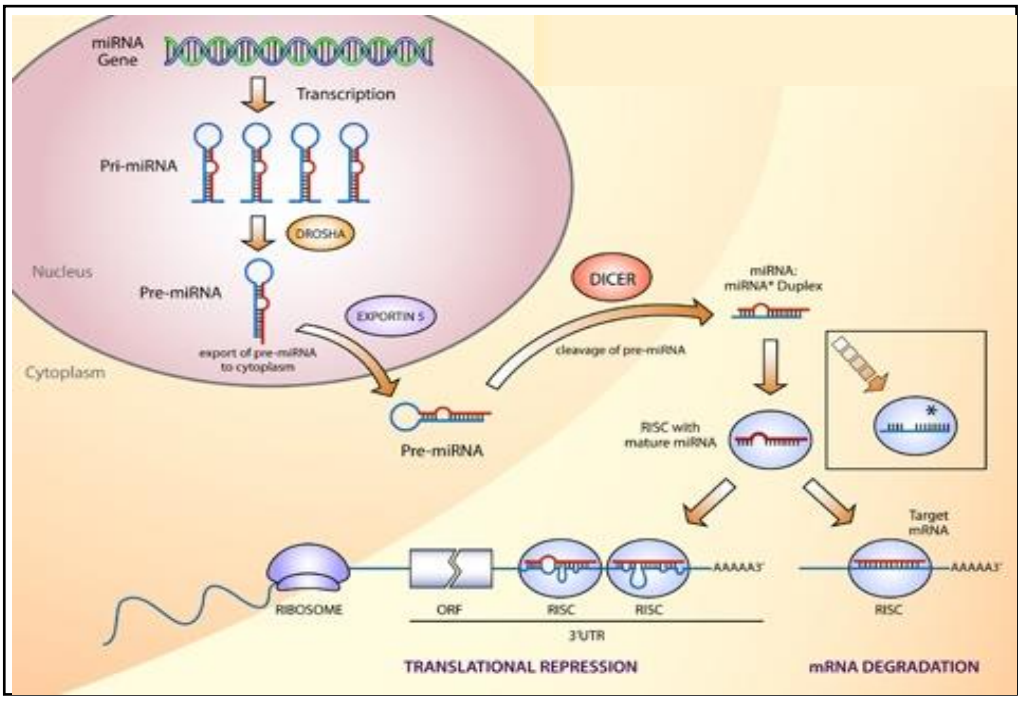
1.10.2.1. Micro-RNA biogenesis

MiRNAs are small non-coding single strands of RNAs about 22 nucleotides in length. In eukaryotes, miRNA biogenesis begins inside the nucleus (Figure 1-28), and then processing and maturation take place in the cytoplasm. MiRNAs are transcribed by RNA polymerase II as a long primary transcript (pri-miRNAs) characterized by multiple hairpin loop structures.

Pri-miRNAs are processed by the enzyme Drosha complex to produce smaller precursor molecules of nearly 70-nucleotides (pre-miRNAs). These precursors are then exported to the cytoplasm by the Exportin 5/Ran-GTP complex (258). In the cytoplasm, the pre-miRNAs are processed by RNAse III Dicer, which generates double-stranded-RNAs called duplex miRNA/miRNA of 22-24 nucleotides. The duplex miRNAs are converted to mature single-stranded miRNAs by the RNA-Induced Silencing Complex (RISC) that creates a large protein complex; the complex then interacts with the 3'UTR of the targeted mRNA. Mature miRNAs target specific mRNA by Watson-Crick pairing (259, 260), where a complementary level between the miRNA 'seeding region' and the sequence of 3'UTR of mRNA controls the expression. If the complementarity between the 2-7 nucleotide of the miRNA seeding region and the 3'UTR is perfect, the mRNA is cleaved by RISC and the mRNA is degraded, whereas if the matching is imperfect then translational repression occurs (261).

Figure 1-28: miRNA biogenesis

The synthesis of miRNA begins within the nucleus, where is transcribed by polymerase II to pri-miRNA. Then, Drosha enzyme converts the pri-miRNA to pre-miRNA through cleavage. Once pre-miRNA is formed, the molecule is exported into the cytoplasm by 5/Ran-GTP complex. Through cleavage and the splitting of phosphodiester bonds, a second RNAase III, Dicer, converts pre-miRNA to miRNA duplex, which then is processed by RISC to generate RISC that is bound to 3' UTR of the targeted mRNA. Adapted with modification from He and Hannon (2004).



1.10.2.2. Micro-RNA and cholesterol metabolism regulation

MiRNAs have been identified as post-transcriptional regulators of lipid metabolism related genes. MiR-122, for instance, was the first identified miRNA to regulate lipid metabolism. It is highly expressed in liver (262) and is associated with plasma cholesterol levels (263, 264), where *in vivo* studies have shown that miR-122 antisense oligonucleotides (ASO) treated mice have 25-35% lower levels of LDL-C and HDL (265). Studies showed that the miR-122 downregulated several genes, some of which, such as Fatty acid synthetase (FASN), and Acetyl-CoA Carboxylase 1 and 2 (ACC1 and ACC2), are involved in fatty acid synthesis and

oxidation (265, 266). In addition, miR-122 has indirect impact on targeted gene expression by interacting with other miRNAs such as miR-370, where the miR-370 upregulates the expression of miR-122 leading to an increased expression of lipogenic genes including Cpt1α ,SREBP1c and DGAT2 (267).

A variant such as a SNP in either the target gene seeding region or in the miRNA itself could disrupt mRNA/miRNA binding, which may affect gene expression and disease phenotype (268). Identifying functional miRNA-associated SNPs is challenging because miRNAs usually have imperfect complementarity with the target gene (263). The presence of SNPs in the target gene seeding region have direct and indirect impact of miRNAs in gene regulation.

1.11. Human genome sequence projects

In time, and with genetic technology improvement, it has become easier, faster and cheaper to study the human genome. From the end of last century to date, several projects have been published to analyze the human genome, which are having a great influence on human health. Here, I highlight some of the human genome related projects.

1.11.1. The Human Genome Project (HGP)

The human genome project (HGP) is an international, collaborative research program. The HGP ran between 1990 and 2003, and had two goals: the mapping of the human genome to help in the study of inherited disease and provide scaffold for genome assembly; and the sequencing of the human genome to serve as a foundation for future genomic studies. In 2003, the final phase of the project reported that the human genome sequence comprises about 2.85 billion nucleotides and covered ~ 99% of the euchromatic genome with an error rate of ~1 event per 100,000 bases (269). The HGP has had, and continues to have, a great impact on biomedical research. It provides the essential foundation data for follow-up genome studies such as HapMap and ENCODE.

1.11.2. The HapMap project

The HapMap project is an international organization aimed to develop a haplotype map (HapMap) that describes the common patterns of DNA sequence variation in the human genome (https://hapmap.ncbi.nlm.nih.gov/index.html.en) (270). Individuals share approximately 99.5%

of their DNA sequence; however, their genome differs at specific nucleotide sites; these sites are known SNPs. The HapMap project was interested only in common SNPs that occur in >1% of the population. To determine the common patterns of DNA sequence in the human genome, the HapMap project used genotype technology to characterize sequence variants, their frequencies, and correlations between them, in DNA samples from populations with different ancestry from Japan, the United Kingdom, Canada, China, Nigeria, and the United States. HapMap has three phases: the complete data obtained in Phase I was published in 2005 (271), the Phase II dataset analysis was published in 2007 (272), and the Phase III dataset was released in 2010 (273).

1.11.3.The 1000 genome project

The 1000 Genomes project was set up in 2008 and final phase data were published in 2015. The 1000 Genomes project extended the data from the international HapMap project. The goal of the 1000 Genomes project was to provide the largest public catalogue of human variation and genotype and their haplotype contexts. By the end of the project, the genomes of 2,504 individuals from 26 populations were reconstructed using a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping. It characterized over 88 million variants (84.7 million SNPs, 3.6 million short insertions/deletions (indels), and 60,000 structural variants) (274). The 1000 Genomes Project data can be viewed in Ensembl (http://www.ensembl.org/ index.html).

1.11.4. The Encyclopedia of DNA Elements (ENCODE)

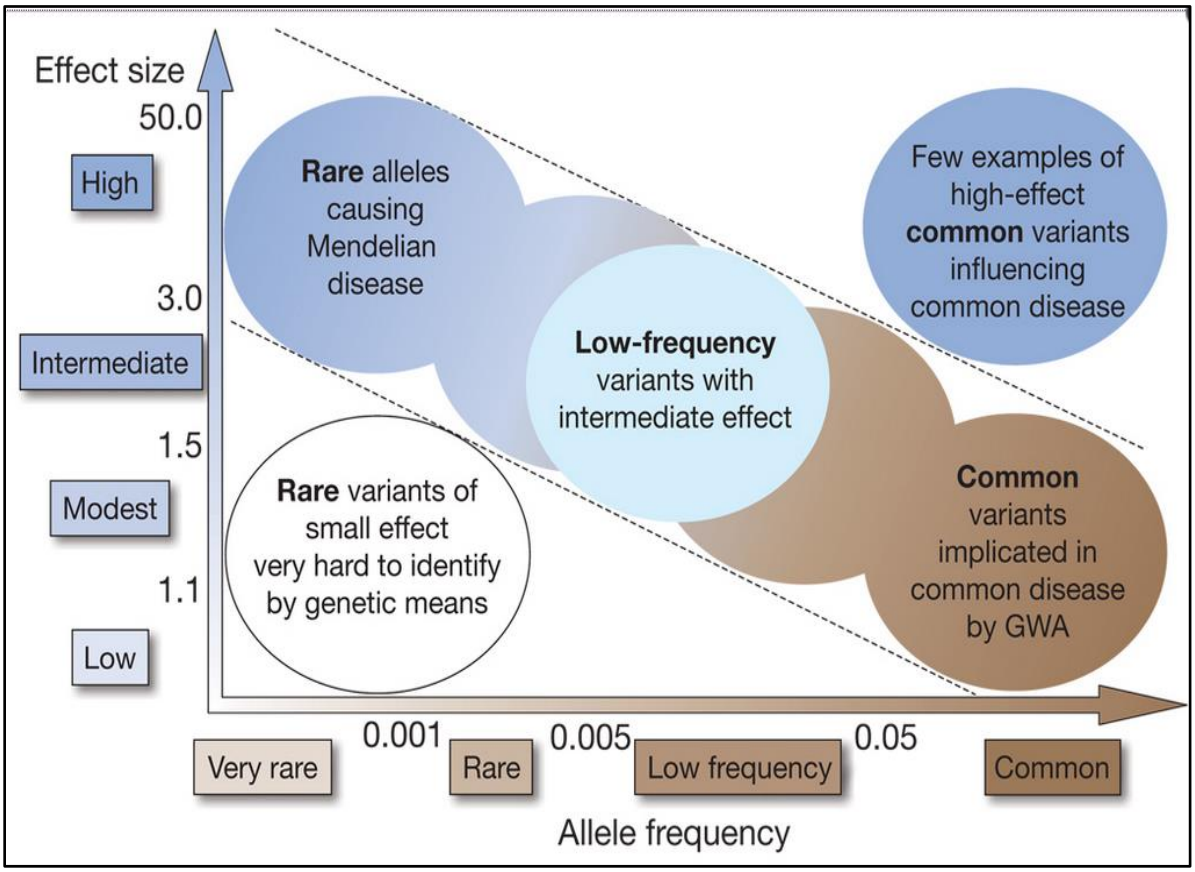
In 2003, the ENCODE project was launched as a follow-up project of the HGP to interpret the human genome sequence. The human genome has approximately 20,000 genes that encode protein, but these genes represent only 1% of genome, while 99% of genome encodes non-gene sequences of unknown function. The ENCODE project aimed to identify and locate all of the genome's protein-coding genes, non-protein coding genes, and other sequence-based functional elements (275). ENCODE has developed methods and performed a large number of sequence-based studies to map functional elements across the human genome including RNA transcribed regions (RNA-seq, CAGE, RNA-PET, and manual annotation), protein-coding regions (mass spectrometry), transcriptional factor-binding sites (ChIP-seq and DNase-seq), chromatin structure (DNase-seq, formaldehyde assisted isolation of regulatory elements (FAIRE-seq), histone ChIP-seq and MNase-seq), and DNA methylation sites (RRBS assay. These have led to systematically mapped regions of transcription, transcription factor association, chromatin structure, and histone modification (276). ENCODE has reported that 80% of the non-coding region of the genome has biochemical functions and the majority of GWAS identified SNPs are residing in or near these ENCODE-defined functional regions (277). In fact, the ENCODE project has provided a comprehensive catalogue of the structural and functional components encoded in the human genome sequence, which has had a great impact in identifying biological pathways influenced by variants which led to disease phenotypes.

1.11.5.The Genome-Wide Association Study (GWAS)

To study a Mendelian disorder, family-based linkage analyses are used, where a large family with a penetrant phenotype is required (278). Linkage studies are well-powered to detect variants with large effects and high penetrance. However, by contrast, a common complex disease is caused by a combination of multivariants with small effects as well as environment. Therefore, identifying trait or disease associated variants requires a new experimental regime and a new statistical method (279). GWAS is a statistical method that used the success of the international HapMap project to annotate the location of millions of the common SNPs and apply an association approach in large phenotyped cohorts (case-control study genotype them all with SNP-Chip) (279). This method has had a great impact on identifying common disease-associated variants that have minor allele frequency (MAF) of more than 5% with small genetic effects (278). Figure 1-29 shows the feasibility of identifying genetic variants by risk-allele frequency and strength of genetic effect. Since 2007, GWAS has been successful in identifying a huge number of genetic variations that contribute to the risk of common and complex diseases such as cancer, diabetes, blood pressure, heart disease and mental disorders. In 2012, the Genome-Wide Associations published at $p \leq 5X10\text{-}8$ for 17-trait categories http://image.sciencenet.cn/home/093802b 2f182 zmjuuzap84.jpg.

Figure 1-29: Feasibility of identifying genetic variants by risk-allele frequency and strength of genetic effect

Schematic showing the relationship between the frequency of susceptible alleles and the strength of their phenotypic effect (effect size). Here variants are categorized as common (MAF > 5%), low frequency (5% > MAF > 0.05%), and rare (MAF < 0.05%). The severity of the mutation (penetrance) can be expressed as the phenotypic effect (effect size) to minor allele carriers and can be distributed as high (effect size > 3.0), or low (effect size < 1.1). The causative variants of Mendelian inheritance disorders confer a high risk of disease (e.g. FH), most of which are caused by rare mutations. By contrast, genetic variants identified by GWASs are mainly common alleles that are associated with a low risk of disease. Currently, low-frequency alleles that are associated with intermediate levels of risk and rare variants that are associated with a low risk of disease can only be identified (whole-genome or targeted) by the resequencing of large cohorts of patients. Adapted from Manolio et al. (2009).



GWAS have examined the correlation between phenotype and variant genotype in large samples of the general population, assuming that individuals of similar phenotype share the same risk variant (279). GWAS statistical power is a function of sample size, effect size, and causal/marker allele frequency. Thus, large population samples are required to detect variants of small effects. In the past few years, GWAS has moved along very rapidly. Combined data from different GWASs of large-scale meta-analyses have resulted in an impressive number of robustly-replicated associations between common variants and specific traits. An example of a large-scale GWAS consortia is the Global Lipid Genetic Consortium's (GLGC) (http://csg.sph.umich.edu/) meta-analysis which has identified over 150 human loci associated with lipid levels (161, 281), and CARDIoGRAMplusC4D (http://www. cardiogramplusc4d.org/) whose GWAS meta-analysis has identified >50 risk loci for CVDs (282)

Although GWAS has been a major success in biomedical discovery, there is some skepticism regarding its small effect size and the problem of "missing" heritability, signals in non-coding regions and linkage disequilibrium (LD). Herein, I discussed some of these issues. First**,** small effect size and missing heritability, the effect sizes (odds ratios) of GWAS "hit" SNPs that are small, often less than 1.3, and even in combination they explain less than 10% of complex trait heritability (283). Some of this 'missing heritability' may result from the number of variants that are responsible for the phenotype being significantly underestimated, thus there are likely to be many more variants with even smaller effect size that need to be discovered. Another possible explanation is that GWAS arrays lack the ability to detect structural variants such as insertions, deletions,

inversions and copy number variants, which commonly occur in the human genome and have strong association with several diseases and/or the arrays miss some loci. There is also a possibility that missing heritability arises from rare variants with a large effect size, which are undetectable by GWAS SNP arrays (278, 283).

One of the major issues associated with GWAS findings is that the majority of common variants (80% or more) have been discovered in non-coding regions, and their functional implications are unknown (278, 283). It then became clear that non-coding regions are involved in gene regulation, and a variant in these regions could turn the gene OFF or ON; however, interpretation of the molecular mechanisms of non-coding variants is a huge challenge because of the diversity of non-coding functions, including transcription, mRNA splicing and translation (284). Extensive functional studies are needed to identify the precise molecular mechanisms of SNPs that associate with a trait or disease. LD is another obstacle of the GWAS findings (278, 283). LD is the nonrandom association between alleles at different loci because of recombination events at meiosis. On GWAS arrays, marker SNPs serve as proxies for much of the surrounding genetic variation with which it is inherited. The problem is identifying the true "causal" variant, and how do we prove that a particular variant *causes* the observed phenotype, when all SNPs in the LD with the functional SNP may carry some or all of the association with the trait of interest, although they have no relevant function?. This requires substantial biological research.

1.11.6. The Genotype-Tissue Expression (GTEx) project

GWASs have identified common SNPs in thousands of loci that are associated with common diseases, however, the majority of common variants have been discovered in non-coding regions, and the mechanism underlying disease susceptibility remains unknown (278, 283). It is suggested that such common variants can influence gene expression through alterations in splicing, non-coding RNA expression and RNA stability. The Genotype-Tissue Expression (GTEx) project aimed to create a reference catalogue for genetic variant and gene expression association in multiple reference human tissues (285, 286). The GTEx project is a powerful tool for interpreting GWAS findings, and provides data and resources about expression quantitative trait loci (eQTL) in a wide range of tissues of relevance to many diseases.

Chapter 2 : Materials and Methods

2.1. Materials

2.1.1. Reagent/stock/kit:

Table 2-1: Reagent/stock/kit list

Reagent/stock/kit	suppliers	Contents
ABI PRISM ® 384 well clear optical reaction plate with barcode	Applied Biosystems/ Life Technologies Cat No 4309849	50 plates
ACC65I	New England BioLab Inc Cat No R0599L	10,000 units
AccuMelt HRM SuperMix (2x)	Quanta Biosciences	
AccuStain kit	NanoEnTek Inc Cat No: ADR-1000	12.5 ml AccuStain solution T, and 4 12.5 ml AccuStain solution N
Acrylamide to Bisacrylamide Stabilized Solution40% (W/V)19:1	National diagnostic , USA Cat No EC-850	1L
Agarose gel (1.2% (w/v))	In house	1.2g agarose, 10ml 10X Tris-Borate-EDTA (TBE) buffer with 0.01µl ethidium-bromide dye (EtBr); 90 ml dH2O
Agarose gel loading buffer (6X)	In house	10mM Tris-HCL(PH7.6),0.03% Bromophenol Blue, 0.03% Xylene Cyanol CFF, 60% glycerol (C3H8O3), 60mM Ethylenediaminetetraacetic acid (EDTA)
Ammonium Persulphate (APS)	Sigma-Aldrich Cat No A3678	100g
Ampicillin D-(−)-α-Aminobenzylpenicillin sodium salt	Sigma-Aldrich Cat No A9518-25G	25 g
Annealing buffer (10X)	In house	200 mM Tris pH 7.6, 100mM $MgCl_2$, 500mM NaCl
ANXA2 rs17191344 TaqMan® SNP Genotyping Assay (40X)	Applied Biosystems Cat No 4351379	188ul
ANXA2 rs17845226 TaqMan® SNP Genotyping Assay (40X)	Applied Biosystems Cat No 4351379	625ul
Bacto Agar	BD Diagnostics, Maryland, USA Cat NO 214010	454g
BamH I	New England BioLab Inc Cat No R0136S	10,000 units
Biotin-11-dUTP	Ferments Life Science Cat No R0081-50 nmol	To get a 5µM stock (=5pmol/µl) dilute 1:200 in 10mM Tris-HCl pH 7.6, 1mM EDTA
Bovine Serum Albumin (BSA)	Sigma Cat No B4287-1G	Heat shock fraction, protease free, suitable for hybridization, pH 7, ≥98%

Bromophenol Blue 0.04%	Sigma-Aldrich Cat No 318744	500ml
calf intestinal alkaline phosphatase (CIAP)	Invitrogen, Life Technologies, Cat No 18009-027	1 U/µL
Chloroform **AnalaR NORMAPUR® ACS analytical reagent**	VWR, UK Cat No 22711.324	2.5 L
CL-XPosure Film, 7 x 9.5in. (18 x 24cm)	Thermo Scientific Cat No 34089	100 films
DH5α cell	New England BioLab Inc Cat No C2988J Lot No 0291305	NEB 5-alpha Competent E. coli (Subcloning Efficiency)
Dulbecco's Modified Eagle medium	Code:E15-810 LotE81013-1793	DMEM with 4.5g/L glucose and L-Glutamine (500ml)
Eagle's Minimum Essential Medium (EMEM)	Sigma Cat No M4655	500ml
EMSA Annealing Buffer (10X)	In house	200 mM Tris pH 7.6 100 mM $MgCl_2$ 500 mM NaCl
EMSA Pre-loading dye	In house	200 µl of 2X Binding buffer, 8 µl BSA, 66.6 µl 6X dye, 125.4 µl dH2O
EMSA, 2X Binding buffer	In house	16% ficol, 40mM HEPES, 100mM KCl, 2mM EDTA, and 1mM DTT
Ethidium-bromide dye (EtBr) 10mg/ml	Invitrogen, Life Technologies, Carisbad, California, USA Cat No 15585011	10ml
Ethylenediaminetetraacetic acid (EDTA)	Cat No AM9262	1L
Ficol	Pharmacia Biotech Cat No 237655	Ficol®400
Foetal bovin serum (FBS)	Gibco, Life Technologies Cat No 10499044	500ml
GenElute™ HP Plasmid Maxiprep Kit	Sigma Cat No NA0310-1kt	25 preps
Glucose **D(+)-Glucose anhydrous, Certified AR for analysis, meets analytical specification of Ph.Eur., BP, USP**	Fisher Scientific Cat No 10141520	500g
glycerol (C3H8O3) **Glycerol, 99.5%, for molecular biology, DNAse, RNAse and Protease free**	Fisher Scientific Cat No 10692372	1 L
HepG2 cell line	European Collection of Cell Cultures Cat No: 85011430.	1 tube
High Capacity RNA to cDNA Kit	Applied Biosystems Cat No 4387406	50 rxn
HindIII	New England BioLab Inc Cat R0104S	10,000 units
Huh7 cell line	Health Protection Agency culture collections Cat No: 01042712	1 tube
Hybond-N+ membrane	VWR, UK Cat No RPN203B	20Cmx3M

Illustrs GFX PCR DNA and gel band purification kit	GE Healthcare Cat No 28-9034-70	100 purifications
In-Fusion® HD Cloning Kit	Clontech Laboratories, Inc Calfornia, USA Cat No 638918	100 rxns
Isoamyl alcohol	Fisher Scientific Cat No 10680311	500 ml
KAPA 2Xmastermix	KAPA Bioscience Cat No KK4604	5ml
Klear Kall 2x mastermix	K Bioscience Cat No KBS-1002-003	2.5ml
LightShift™ Chemiluminescent EMSA Kit EMSA Kit	Thermo Scientific Cat No PI-20148	-
Lipofectamine™ 2000	Invitrogen, Life Technologies, Carisbad, Calfornia, USA Cat No: 11668-019	1.5ml
Luciferase Reporter Assay System	Promega Lot No 0000078853	-
Luria Broth (LB)	In house	2.5g of Tryptone powder, 1.25g of Yeast, 2.5g of NaCl, 0.5g of Glucose, 2.50µl of 5mol NaOH, and 250ml dH2O
Luria Broth (LB) /ampicillin agar	In house	2.5g of Tryptone powder, 1.25g of Yeast, 2.5g of NaCl, 0.5g of Glucose, 2.50µl of 5mol NaOH, 3.75g of Agar , 250ml dH2O, and 250µl of 100mg/ml ampicillin)
MgCl2	BDH- Biochemical Lot No TA753889	50µm/µl
mirVana[a] miRNA Mimic miR-1 Positive Control	Applied Biosystems Cat No 4464062	5nmol
mirVana™ miRNA mimic (hsa-miR-155-3p)	Life Technologies Cat No 4464066 (MC12634)	5nmol
mirVana™ miRNA Mimic Negative Control #1	Applied Biosystems Cat No 4464058	5nmol
N,N,N′,N′-Tetramethylethylenediamine (TEMED)	BOH Electron Cat No 54824214.742	100ml
Nuclei extract buffer A	In house	10mM HEPES (pH 7.9), 1.5mM MgCl2, and 10mM KCl
Nuclei extract buffer C	In house	20mM HEPES (pH 7.9), 25% v/v glycerol, 0.42M NaCl, 1.5mM MgCl2, and 0.2mM EDTA
Opti-MEM® Medium	Gibco, Life Technologies, Carisbad, Calfornia, USA Cat No: 31985-062	100ml
Oragene•DNA (OG-500) kit	DNAgenoTek, Ottawa, ON, Canada Cat No: (OG-500)	2ml
Paraffin	BDH Cat No 29436511	20 L
pGL3- Basic Vector	Promega, Madison, Wisconsin, USA, Cat No E1751	20µg

pGL3- Promoter Vector	Promega, Madison, Wisconsin, USA, Cat No E1761	20µg
pGL3-Control Vector	Promega, Madison, Wisconsin, USA, Cat No E1741	20µg
pGL3-Enhancer Vector	Promega, Madison, Wisconsin, USA, Cat No E1771	20µg
Phosphate-buffered saline (PBS)	Gibco Cat No: 18912014	100 tablets
Phusion®High-Fidelity PCR Kit	New England BioLab Inc Cat No E0553L	-
Plus DNA ladder (1Kb)	Thermo Fisher Scientific Cat No 10787-026	-
Poly(deoxyinosinic-deoxycytidylic) acid sodium salt	Qiagen Cat No 338162	1µg/µl
Poly(deoxyinosinic-deoxycytidylic) acid sodium salt (poly dI.dC)	Sigma Aldrich Cat No 118578-37-3	1µg/µl
Potassium chloride (KCl)	AnalaR BDH Limited Poole England Cat No 1019 Lot No 512542H	-
prepIT®•L2P reagent	DNAgenoTek, Ottawa, ON, Canada Cat No: PT-L2P-5	0.5mL
pRL-TK	Promega, Madison, Wisconsin, USA Cat No:E2241	20µg
Protease inhibitor cocktail (100X)	Sigma Cat No P8340	1ml
puc 18 plasmid	New England Biolabs, Ipswich, Ma, USA Cat No: 78410	20µg
QIAamp DNA Mini Kits	QIAGEN Cat No 51306	50 QIAamp
QIAprep Spin Miniprep Ki	QIAGEN Cat No 27104	50 QIAprep
QuikChange Lighting Site-Directed Mutagenesis Kit	Agilent Technologies Cat No 210519	30 reactions
Rabbit monoclonal [EPR17362] to GATA1	Abcam , Cambridge, UK Cat No: ab181544	10µl
Rabbit monoclonal [EPR17874] to GATA2 + GATA3	Abcam , Cambridge, UK Cat No: ab182747	10µl
Rabbit monoclonal [EPR5014(2)] to Egr1	Abcam , Cambridge, UK Cat No: ab133695	10µl
Rabbit monoclonal [EPR7313] to CTCF	Abcam , Cambridge, UK Cat No: ab128909	100µl
SalI	New England BioLab Inc Cat No R0138L	10,000 units
siPORT™ NeoFX™ Transfection Agent with Manual	Applied Biosystems Cat No AM4510M	400ul
Sodium chloride (NaCl)	Fisher Scientific Cat No 117128	-

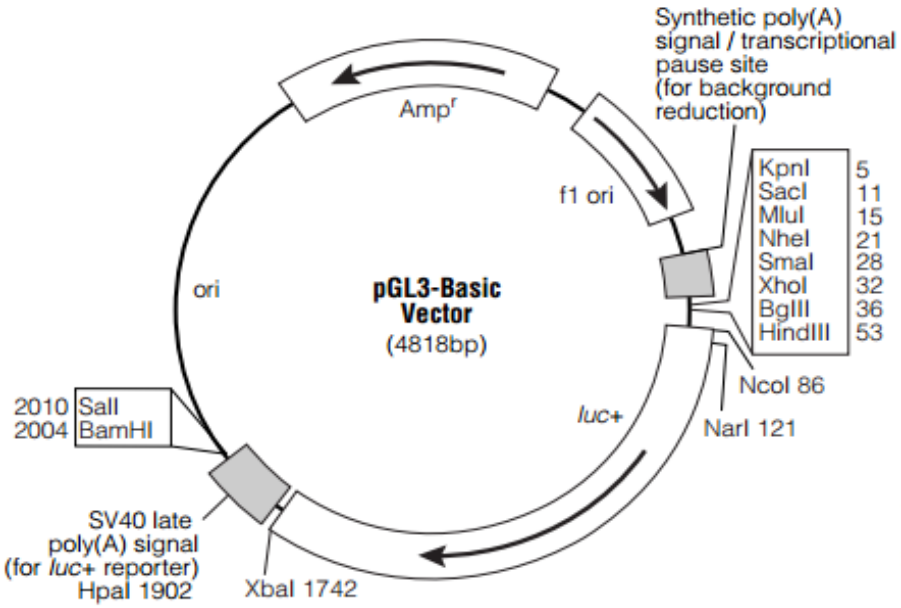
Sodium Acetate (NaAc)	AnalaR BDH Limited Cat No 100192-416	3 M-pH5.2 100ml
Sodium Hydroxide (NaOH)	Sigma-ALDRICH Cat No BCBH 9037V	-
SSC 20X	In house	175.32g NaCl + 88.23g Na3Citrate → in 1L dH2O
T4 DNA ligation	New England BioLab Inc Cat No M0202S	20,000 units
TaqMan Expression assay: UBC	Applied Biosystems Cat No 4331182	250 rxn
TaqMan Gene Expression assay ANXA2 (50RXN/150PCR runs) Hs00743063_s1	Applied Biosystems Cat No 4331182	250 rxn
TaqMan miRNA Expression assay : GAPDH	Applied Biosystems Cat No 4331182	250 rxn
TaqMan miRNA Expression assay : hsa-miR-155-3p or miR-155*	Applied Biosystems Cat No 4427975	50RXN/150PCR runs
TaqMan miRNA Expression assay : hsa-miR-155-5p	Applied Biosystems Cat No 4427975	50RXN/150PCR runs
TaqMan Universal Master Mix II-no UNG	Applied Biosystems Cat No 4440040	5ml
TaqMan® MicroRNA Assays: RNU6B	Applied Biosystems Cat No 4427975 (001093)	50RXN/150PCR runs
TaqMan® MicroRNA Reverse Transcription Kit, 200 rxn	Applied Biosystems Cat No PN 4366596	200 rxn
TaqMan® Universal PCR Master Mix, No AmpErase® UNG, 1-Pack (1 x 5 mL)	Applied Biosystems Cat No PN 4324018	200 rxn
Tris Base	Fisher Scientific Cat No 117128	-
Tris- Borate -EDTA (TBE) 1X	Fisher BioReagents Cat No BP2430-4	5L
Tris-HCL	Fisher Scientific Cat No BP153-1	1Kg
TRIzol reagent	Thermo Fisher scientific (Ambion™) Cat No 15596026	100ml
Tropix TF microplate 96 well Assay plate, no lid,flat bottom,non-treated,non-sterile, white polystyrene	Appleton Woods Cat No CC704	25/pack
TruSeq Custom Amplicon (TSCA) v1.5 kit	Illumina, San Diego, CA Cat No: FC-130-1001	96 samples
Trypsin-EDTA solution	Sigma Cat No T4049-100ML	0.25% Trypsin-EDTA 100mL
Tryptone powder	Oxoid Cat No LP043B	500g
Ultra pure agarose	Invitrogen, Life Technologies, Cat No 16500-500	500g
Whatman paper	GE Healthcare life Sciences Cat No: 3030-700	Grade 3MM Chr Blotting Paper, roll, 23 cm × 100 m 3MM Chr cellulose blotting sheets
X-gal	Fisher Scientific Cat No 10234923	-

Yeast Extract Powder	Oxoid Cat No LP0021B	500g
β-Gal fixation solution	In house	100 µl 10XPBS, 50 µl Formalin (40%formaldehyde Sol), 2 µl Glutaraldehyde, 850 µl dH2O
β-Gal staining solution	In house	100 µl 10XPBS, 2µl MgCl2 1M, 100 µl potassium Ferricyanide (50mM),100 µl potassium Ferrocyanide (50mM), 700 µl dH2O
β-Galactosidase (β-Gal)	Promega, Madison, Wisconsin, USA, Cat No:E1081	20µg

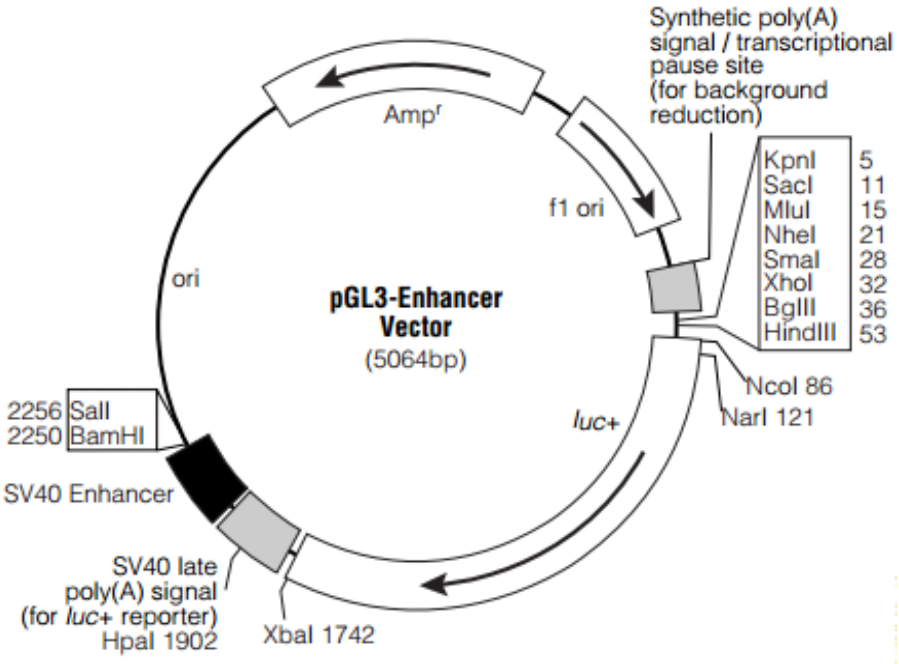
Figure 2-1: pGL3 Vectors circle maps

A)The **pGL3-Basic Vector** lacks eukaryotic promoter and enhancer sequences, allowing maximum flexibility in cloning putative regulatory sequences. Expression of luciferase activity in cells transfected with this plasmid depends on insertion of a functional promoter upstream from luc+. Potential enhancer elements can also be inserted upstream of the promoter or in the BamHI or SalI sites downstream of the luc+ gene. B) The **pGL3-Enhancer Vector** contains an SV40 enhancer located downstream of luc+ and a poly (A) signal. This aids in the verification of functional promoter elements because the presence of an enhancer will often result in transcription of luc+ at higher levels. C) The **pGL3-Promoter Vector** contains an SV40 promoter upstream of the luciferase gene. DNA fragments containing putative enhancer elements can be inserted either upstream or downstream of the promoter-luc+ transcriptional unit. D) The **pGL3-Control Vector** contains SV40 promoter and enhancer sequences, resulting in strong expression of luc+ in many types of mammalian cells. This plasmid is useful in monitoring transfection efficiency, in general, and is a convenient internal standard for promoter and enhancer activities expressed by pGL3 recombinants. luc+ = cDNA encoding the modified firefly luciferase; Ampr = gene conferring ampicillin resistance in E. coli; f1 ori = origin of replication derived from filamentous phage; ori = origin of plasmid replication in E. coli. Arrows within luc+ and the Ampr gene indicate the direction of transcription; the arrow in f1 ori indicates the direction of ssDNA strand synthesis. Adopted from https://www.promega.co.uk/resources/vector-sequences/reporter-vectors/pgl3-luciferase-reporter-vectors/

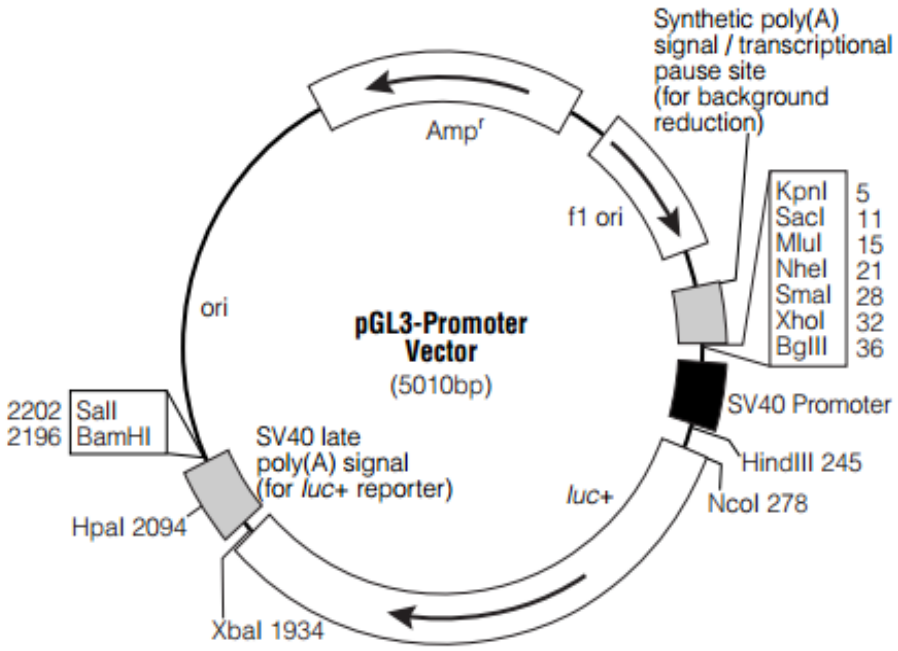
A) pGL3-Basic Vector



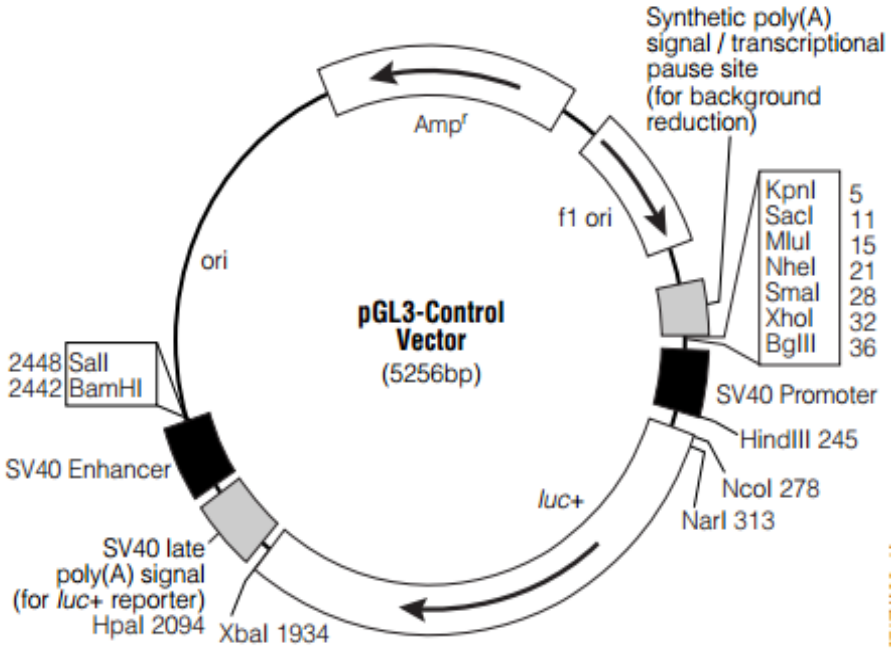
B) pGL3-Enhancer Vector



C) pGL3-Promoter Vector



D) pGL3-Control Vector



2.1.2. Equipment:

Table 2-2: Equipment

Equipment	Company
BioRad C1000^{Tm} Thermocycler	
7900HT fast Real-Time PCR System	Applied Biosystems
Syngene Gel Documentation system and Genesnap software version 6.04	Syngene
Syngene G:BOX Documentation system	Syngene
NanoDrop 8000	Thermo Scientific, ThermoFisher Scientific, USA
Advanced Detection and Accurate Measurement (ADAM) cell counter	Digital Bio
Maxi Electrophoresis Modular Systems (code CS2221)	Appleton Woods
UV Stratalinker 2400	Stratagene
X-ray film developer	Photon imaging system
Rotor-Gene6000	Qiagen Ltd, Crawley, West Sussex, UK

2.1.3. Websites:

Table 2-3: Websites

Database	URL
International HapMap project	http://www.hapmap.org
UCSC Genome Browser	http://genome.ucsc.edu/
Ensembl	http://www.ensembl.org/
HaploReg	http://www.broadinstitute.org/mammals/haploreg/haploreg.php
MatInspector	http://www.genomatix.de/matinspector.html
miRWalk	http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/
MirSNP	http://cmbi.bjmu.edu.cn/mirsnp
miRanda	http://www.microrna.org/
PolyPhen 2	http://genetics.bwh. harvard.edu/pph2/
SIFT	http://sift.jcvi.org/
STRING	http://string-db.org/
Applied-Biosystems	http://www.thermofisher.com/uk/en/home/brands/applied-biosystems.html
R version 3.2.0	http://www.R-project.org/.
GTEx portal	http://www.gtexportal.org/home/
Gene network	http://genenetwork.nl/bloodeqtlbrowser/
GE Healthcare/ Life Sciences	https://www.gelifesciences.com
Source Bioscience Life Sciences	http:// www.lifesciences. sourcebioscience.com/genomic-services/saner-sequencing-service.aspx
CLUSTALW sequence alignments	http://www.genome.jp/tools/clustalw/
In-Fusion primers convert	http://bioinfo.clontech.com/infusion/convertPcrPrimersInit.do
Primer3 design program version 4.0	http://primer3.ut.ee/

2.2. Bioinformatics tools

I used several bioinformatics databases that helped me to annotate genes and SNPs of interest as well as annotating protein interaction.

2.2.1. Genomic annotations

International HapMap project: HapMap Genome Browser release #28 (http://www.hapmap.org) was chosen to identify gene FASTA sequence (287). The gene name or SNP ID number was typed into the search box and the standard search parameters were kept and the 'Download Decorated FASTA File' Report & Analysis function was selected to view results.

University of California Santa Cruz (UCSC) Genome Browser: The UCSC Human Genome Browser (http://genome.ucsc.edu/) was chosen to annotate genes and SNPs, regulatory domains and regulatory features, including transcribed regions and sequences that interact with protein, RNA and other regions of the genome to affect gene activation and chromatin structure (288). The gene name or SNP rs number was typed into the search box. The regulation track was regulated by specifying the cell line in all sub-tracks including regulation from ENCODE tracks where transcription factor criteria were selected, ENCODE open chromatin by DNaseI HS and FAIRE tracks, ENCODE transcription factor binding tracks, and HMR conserved transcription factor binding sites (CTFBS). When all tracks were optimized, the data were submitted to view the results. This bioinformatics tool also was used to identify the transcription factor binding sites (TFBS) domain within the gene sequence by selecting the view option from the main menu,

then DNA and extended DNA, and case/color options were configured to enable discrimination of TFBS from the whole sequence.

Ensembl: The Ensembl project (http://www.ensembl.org/) was chosen because it facilitates genome annotation by providing links to gene view, transcript view, protein view, and variation view. These pages contain additional information about the gene or a region of the gene (289).

The gene name or SNP rs number was typed into the search box and then the annotation of interest was selected. Menus provided the ability to select specific annotations for display.

2.2.2. Tools for annotating the non-coding genome

HaploReg: HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php) is a tool that was used for annotations of the non-coding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease-associated loci. This tool uses LD information from the 1000 Genomes project, LD SNPs along with their predicted chromatin state, their sequence conservation across mammals, and their effect on gene regulation such as promoter and enhancer related histone marks, DNase hypersensitivity, binding proteins and altered transcription factor motifs (290). The SNP rs number was typed into the search box and the standard parameters were kept to view the results.

MatInspector: MatInspector software (http://www.genomatix.de/matinspector.html) was utilized to predict transcription factor binding sites in specific DNA sequences *in silico* (291).

Consensus FASTA sequences were selected from the Ensembl Human Genome Browser and manually copied into the MatInspector tool. The software uses a large library of matrix descriptions for transcription factor binding sites, locating matches within the FASTA sequence and assigns a quality rating to each match. I selected each FASTA sequence to include either the common or the rare variant of a SNP of interest, to assess whether there is a possibility of differential TFBs between alleles. The parameters selected for tests were as follows: sequence added manually using 'your sequence' function, *Homo sapiens* option selected for 'search corresponding promoters' and 'Transcription Factor Binding (weight matrices)' selected for the library option.

2.2.3. Tools for annotating 3'UTR and miRNAs predicting binding sites

Several miRSearch tools are available online, which are designed to identify miRNAs that target specific gene transcripts. In my thesis I used computational prediction methods to predicate miRNA target. Several predication algorithms were used for this purpose.

miRWalk

miRWalk is an algorithm search for seeds based on Watson–Crick complementarity. It is walking on the complete sequence of a gene searching for a heptamer (seven nucleotides) seed, and as soon as it identifies a heptamer with perfect base-pairing, it immediately extends the length of the miRNA seed until a mismatch arises (292). The miRWalk database provides predicted and validated information on miRNA-target interaction (http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/).

For predicted miRNA-target interaction, the "gene miRNA target" option was selected, human gene Ensembl ID was entered (ANXA2: ENSG00000182718). In addition, a compression tool was used, two other popular miRSearch databases, miRanda and Targetscan, were selected, then the search was run. MiRWalk provided a table of putative miRNA binding site predictions within the 3'UTR region. The table showed information on miRNA-target interaction including the starting position of miRNA seeding, seeding length, seed start, end position, and p-value. Also, miRWalk provided a table of a comparative platform of miRNA binding site predictions with the 3'UTR region resulting from different algorithms (miRWalk, miRanda and Targetscan). These algorithms are used for seeding the complementarity method.

Moreover, miRWalk provides experimentally verified miRNA-target interactions. For this the "validated target module" option was selected from the main menu, then the "gene miRNA target" option was selected, and human gene Ensembl ID was entered (ANXA2: ENSG00000182718), and then the search was run. The output data showed validated gene-miRNA interactions.

MirSNP

MirSNP is a collection of human SNPs at predicted miRNA-mRNA binding sites (293) and is publicly available online (http://cmbi.bjmu.edu.cn/mirsnp). The MirSNP algorithm was used to predict miRNA: ANXA2 3'UTR SNP seed match. The gene name was typed in the single search box and the search was run. The output data showed the gene name, the miRNA name, the SNP ID, the mirSVR score of a binding site, the considered alleles for each SNP, the predicted score

of miRNA-mRNA binding by miRanda software (the higher the score, the more stable the binding), the free energy of the miRNA-mRNA duplex, and the conservation (phastCons score). An example of a miRNA-related SNP search results is shown in Figure (2-2).

Figure 2-2: An example of miRNA-related SNP search results using the MirSNP algorithm

The figure shows the MirSNP output data. The G-allele of the *ANXA2* SNP rs184828726 may promote the binding of the *ANXA2* gene and hsa-mir-4305, and where the interaction is stable (mirSVR= -0.249), leads to produce free energy (-12.52), and is conserved. However, the C-allele of the SNP prevents the interaction between the mature miRNA and the *ANXA2* gene.

Gene	miRNA	Snp	mirSVR	Effect	Allele	Score	Energy	Conservation
ANXA2	hsa-miR-4305	rs184828726	-0.249	break	G	140.00	-12.52	0.287
					C			

MiRanda

MiRanda software is freely available and facilitates evaluation of predicted targets and design of experiments. It provides both high-scoring targets as strong candidates for validation experiments, and lower-scoring targets which may have a role in gene regulation. In addition to expression information for miRNAs and mRNAs, target sites are ranking by complementarity and evolutionary conservation (294).

At the home page (http://www.microrna.org/) the target mRNA was typed (ANXA2) in the search box, species type was selected (Homo sapiens), and the search was run. For miRNA-ANXA2 alignment, all miRNA target sites with all mirSVR score options were selected. The

output data showed a road map of ANXA2-3'URT with both validated and predicted miRNA target sites. Simply click on the miRNA mature name to visualize the miRNA-ANXA2 complementary alignment.

2.2.4. Tools for annotating the coding genome

PolyPhen 2: The Polymorphism Phenotyping V2 (PolyPhen 2) software (http://genetics.bwh.harvard.edu/pph2/) was used to predict the possible impact of missense mutations on the stability and function of human proteins using eight sequence-based and three structure-based predictive features (295). A protein sequence in FASTA format was entered and amino acid substitutions were specified.

SIFT: The Sorting Intolerant From Tolerant (SIFT) algorithm (http://sift.jcvi.org/) was used to predict whether an amino acid substitution affects protein function (296). The SIFT/PROVEAN Human SNP database was selected and input format was used as recommended.

Mutation Taster2: This is the other *in silico* tool (http://www.mutationtaster.org/) (297) that was used to estimate the impact of the variant on the gene product / protein. Ensembl transcript ID of gene of interest was used as input format (starting with ENST).

2.2.5. Protein annotation

STRING: The STRING database (http://string-db.org/) version 9.1 was used to identify protein-protein networks (298). A protein name was typed into the search box, then the *Homo sapiens* option selected.

2.3. Population background

Genotype-phenotype analysis was performed on several studies. A general overview of these studies is given below.

The Second-Northwick-Park Heart Study (NPHSII)

The NPHSII consists of 3,052 healthy middle-aged men (50–61 years) who were recruited in 1989 from nine general medical practices in the UK (299). Information on lifestyle habits, height, weight, blood pressure, and family history of CHD was recorded at baseline and on subsequent prospective follow-up (300). Exclusion criteria were as follows: participants with unstable angina, history of MI, evidence of silent infarcts, coronary surgery, anti-coagulant drugs (including aspirin), cerebrovascular disease, malignancy and any other condition or disease preventing the attainment of written, informed consent or long-term follow-up. Participants did not fast for the study, but they had been instructed to avoid heavy meals to ensure that triglyceride concentration and clotting factors influenced by plasma lipids would be neither atypically low (fasting) nor high (heavy lipemia) at examination.

DNA was obtained from 2,778 of the men at the time of recruitment. Interviews and repeat measurements were conducted annually for surviving participants. CHD end points up to 15

years follow-up were as follows: (1) acute CHD events: sudden coronary death, fatal acute MI, and nonfatal acute MI, the clinical history, ECGs, cardiac enzymes, and pathology were assessed by independent review according to the World Health Organization criteria (1976) (71) and normal limits for cardiac enzymes were those for the reporting laboratory; (2) a new major Q wave on an ECG after 5 years of follow-up (87); and (3) surgery for angina pectoris with CHD angiographically demonstrated (301). Further details about recruitment, measurement, follow-up and incident disease definition have been described elsewhere (5, 299, 302, 303).

Whitehall II study (WHII)

The WHII is a cohort study of 10,308 UK participants (3,413 women and 6,895 men) recruited between 1985 and 1989 from 20 London-based Civil service departments for consideration to participate in a study that investigated the effect of socioeconomic differences in health and mortality. The participants were invited to the research clinic at 5-year intervals (1985-2009). The study has 9 phases; only phases 1, 3, 5 and 7 included both medical and non-medical examinations. Details of the WH-II cohort profile have been described elsewhere (304). Blood samples for DNA were collected from more than 6,000 participants. They attended the clinics in a fasting state, had blood drawn, drank a standard glucose load, and had blood drawn again precisely two hours later—a mammoth undertaking. This blood sample collection method was performed at the Phase 3 (5-year examination), repeated at Phase 5 (10-year examination), and again at Phase 7 (15 years) (305). Standard assays for plasma lipids were carried out as described (306). LDL-C concentration was calculated with the Friedewald formula (LDL-C (mmol/L) = TC – HDL-C – TG/2.2). Details of the design, participants, and lipid measurements have been described in Talmud et al. (2009).

The UCL-LSHTM-Edinburgh-Bristol (UCLEB) consortium

The UCLEB consortium was established to allow interrogation of genetic associations of risk factors for CVD through Metabochip-wide association analyses, as well as associations with measures of subclinical disease such as carotid intima media thickness (cIMT), and with clinical events (307). The consortium consists of 30,000 participants from 12 well-established UK studies (participants are almost exclusively of European ancestry): NPHS II, WHII, British Regional Heart Study (BRHS), English Longitudinal Study of Ageing (ELSA), MRC National Survey of Health and Development (MRC NSHD), 1958 Birth Cohort (1958BC), Edinburgh Artery Study (EAS), Edinburgh Type 2 Diabetes Study (ET2DS), Edinburgh Heart Disease Prevention Study (EHDPS), Aspirin for Asymptomatic Atherosclerosis Trial (AAAT), Caerphilly Prospective Study (CaPS) and the British Women's Heart and Health Study (BWHHS).

Around 21,000 individuals across the cohorts have been typed using Metabochip, a genotyping platform consisting of 200,000 SNPs, which cover loci identified by GWAS in cardiometabolic diseases, and rare variants from the 1000 Genomes project (307). Also, the study extended coverage from 200,000 SNPs to approximately 1 million SNPs by imputation of data from 1000 Genomes project when the r^2 is >0.8. The principle of imputation is that by exploiting patterns of LD within populations, the genotypes of unobserved sites can be inferred. Details of the consortium have been described in Shah et al. (2013).

2.4. TaqMan Genotyping

Each Applied Biosystems® TaqMan® SNP Genotyping Assay (40X) contains sequence-specific forward and reverse primers to amplify a polymorphic sequence of interest, and two TaqMan minor groove binder (MGB) probes. The TaqMan® MGB Probes consist of target-specific oligonucleotides with a fluorescence reporter dye at the 5´ end of each probe; one probe labeled with VIC® dye detects the allele 1 sequence, and the other probe labeled with FAM™ dye detects the allele 2 sequences. The presence of VIC dye fluorescence indicates homozygosity for allele 1, while the presence of FAM dye fluorescence indicates homozygosity for allele 2. The presence of both dyes in a reaction allows genotyping of the two possible variant alleles (heterozygosity) at the SNP site. The 3´ end of the MGB probe has a non-fluorescent quencher (NFQ).

For a TaqMan genotyping reaction, 5ng of genomic DNA was pipetted into a barcoded 384 well polymerase chain reaction (PCR) plate (Applied Biosystems/ Life Technology) using Biomek 2000 laboratory automation workstation. Then the plate was centrifuged at 3000rpm for 1 minute, followed by air-drying for 24 hours at room temperature. Each plate contained at least two no template control (NTC) wells.

The genotyping reaction was carried out as per the genotyping mastermix manufacturer's instructions. Two genotyping mastermixes were used; Klear Kall 2x mastermix and KAPA mastermix (Fast qPCR kits 2x Master mix), the ROX passive reference was supplemented with the Klear Kall 2x mastermix kit; while the ROX passive reference was added to the KAPA mix as recommended by the manufacturers. Components of both mastermix reactions are shown in Table 2-4. In both cases, 2µl of the reaction mix was added to each well. The

plate was then centrifuged at 3000rpm for 1 minute and the PCR reaction carried out using the BioRad C1000^{Tm} Thermocycler. The standard PCR reaction conditions were as follows; started at 50^{0}C for 2 minutes, then 95^{0}C for 10 minutes to activate DNA polymerase followed by 40 cycles of denaturing at 92^{0}C for 15 seconds and annealing/extending at 60^{0}C for 1 minute.

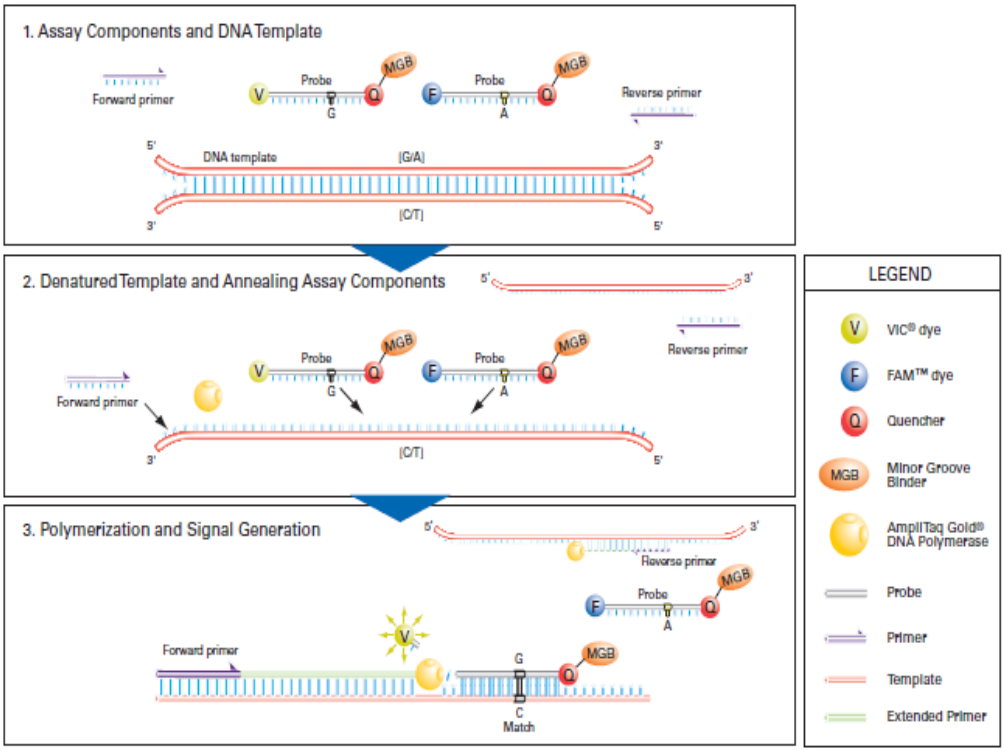
Table 2-4: Master Mix reaction for TaqMan genotyping (384 well PCR plate)

	KAPA	Klear Kall
Sigma Water (Molecular Biology Reagent)	384	389.5
2x Master mix	410	410
ROX	426.4	-
X40 SNP	20.5	20.5

Each PCR reaction had several phases: when genomic DNA is introduced into a reaction mixture consisting of TaqMan genotyping mastermix, forward and reverse primers and two TaqMan MGB probes, each TaqMan MGB probe anneals specifically to a complementary sequence between the forward and reverse primer sites. When the probe is intact, cleavage separates the reporter dye from the quencher dye, increasing fluorescence by the reporter, while the proximity of the quencher dye to the reporter dye suppresses the reporter fluorescence. The increase in fluorescence occurs only if the amplified target sequence is complementary to the probe. Thus, the fluorescence signal generated by PCR amplification indicates which alleles are in the sample (Figure 2-3).

Figure 2-3: A cartoon of the process of TaqMan assay genotyping during PCR.

Adapted from http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/general documents/cms_042998.pdf



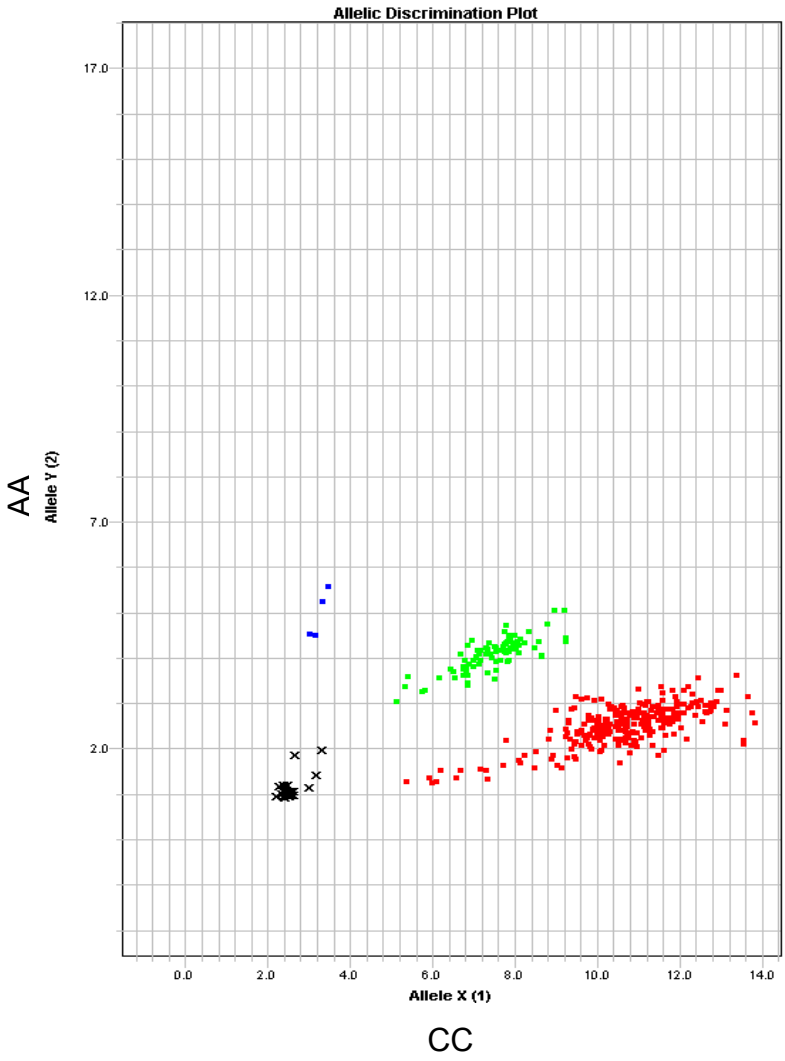
TaqMan signal detection

Fluorescence was detected using 7900HT fast Real-Time PCR System with Sequence Detection System (SDS) V2.4 software (http://tools.lifetechnologies.com/content/sfs/manuals/cms042114. pdf). Standard allelic discrimination was carried out using SDS software which induced an automatic cutoff signal detector. When the PCR plate was inserted into the 7900HT fast Real-Time PCR system, the standard allelic discrimination program assigned genotypes automatically to a plate coordinate in a text output file and in the form of an allelic discrimination plot based on the fluorescence signal detection. The light source from the cycler excites the reporter dyes by fluorescence resonance energy transfer. VIC and FAM have significantly different excitation and mission wavelengths, which allows them to

be differentially detected. Each dot on the scatter plot from results is shown in Figure 2-4. Undefined samples are represented by 'X'. These samples did not reach the cut-off fluorescence signal level (95%) to be defined as a particular genotype.

Figure 2-4: Allelic discrimination plot

This plot represents the results of genotyping *ANXA2* SNP rs17845226 C>A. Samples with significant VIC signal are called homozygous for common allele C (red cluster). Samples with significant FAM signal are called homozygous for common allele A (blue cluster). A sample with an even distribution of VIC and FAM signals is called heterozygous (green cluster). Any sample that has not reached the cut-off signal level for any group is indicated by 'X'.



2.4. 1. Genotype - phenotype association statistical analyses

Using an in-house spreadsheet template, the MAF [MAF= $(2XN_{11}) + N_{12})/N_{total}$], the Hardy–Weinberg equilibrium (HWE) (Chi-square test [$x^2 = \sum \frac{(Observed\ value - Expected\ value)2}{Expected\ value}$], frequency = $p^2 + 2pq + q^2 = 1$). P-value significance cut-off of 0.05 was used to determine whether the genotypes were in HWE.

Table 2-5: Two-allele polymorphism genotype statistical analysis: Hardy–Weinberg equilibrium

The table shows an example of Hardy–Weinberg equilibrium assessment output in the NPHS-II study for the *ANXA2* SNP rs17845226. Observed genotype (O), Expected genotype (E), major allele homozygote (11), heterozygote (12), minor allele homozygote (22).

Stats for a 2-allele polymorphism: rs17845226

	Observed			
	rs17845226	rs17845226	rs17845226	
Genotypes	**11**	**12**	**22**	Total
Frequency	1889	570	31	2490

p	**q**	**p+q**
0.873	0.127	1

	Expected			
	rs17845226	rs17845226	rs17845226	
Genotypes	**11**	**12**	**22**	Total
Frequency	1898.1	551.8	40.1	2490

Chi^2 test and level of significance				
O	**E**	**(O-E)^2/e**	**Chi^2**	**Sig level**
1889	1898.1	0.0		
570	551.8	0.6		
31	40.1	2.1		
			2.711	0.0997

To assess the association between SNP genotype and lipid trait and CHD risk, statistical analyses were performed on several studies, which were conducted using R version 3.2.0 (Dr. Jon White) and STATA data analysis and statistical software (Ms. Jackie Cooper).

The *LDLR* SNP rs6511720 was genotyped in the UCLEB consortium (307), meta-analyses and statistical analyses was carried out by Dr. Jon White. Patient levels of LDL-C and TC were first adjusted to take account of the individual's cholesterol lowering drug status (LDL-C * 1.352; TC * 1.271). The adjusted values were further adjusted to take account of the fixed effect of the study. Patient adjusted lipid measurements were then partitioned according to patient gender and genotype. The patient genotype was represented as number of copies of the minor allele (0, 1, and 2 (additive model)). The levels of LDL-C and TC data were normally disrupted and mean and standard deviation (SD) were determined. The per allele effect was estimated using a linear model in terms of effect size (ES) and 95% confidence intervals (95% CI). The per allele effect of the SNP on CHD risk was estimated in a logistic regression in terms of Odds Ratio (OR) and 95% CI.

The ANXA2 SNPs rs17845226 and rs17191344 were genotyped in NPHSII, and statistical analysis was carried out by Ms. Jackie Cooper. To determine the relationships between SNP genotype and the continuous variables (plasma lipid levels) under selected genotypic models, a summary measure for each group was generated based on data distributions. For normal distributed data, mean and SD was calculated. For skewed data, the median and inter-quartile range (IQR) was calculated. In highly skewed distribution data, a log transformation was

performed in order to make data normally distributed. To test the hypothesis a one-way ANOVA was used for either an additive or recessive model. To determine whether a patient's risk of having CHD as a health outcome is increased or decreased by the presence of exposure (SNP genotype), the Cox proportional hazards models or the logistic regression models were used. In cohort studies, the Cox proportional hazards models were used to estimate hazard ratios (HR) and 95% CI to determine how many times more likely the exposed group (specific SNP genotype) was to develop CHD than the unexposed group. Stepwise models were used to determine which of the combined genotype groups (1-9) of the rs17191344 and rs17845226 SNPs were having an effect on lipids and CHD. The groups were fitted using 8 dummy variables. Models used forward selection with a significance level of 5%.

Meta-analyses and statistical analyses were carried out on the ANXA2 SNPs rs11633032, rs12900101, and 116255697 by Dr. Jon White, using the R statistical software. The *ANXA2* SNPs rs11633032, rs12900101, were inputted into the UCLEB consortium (307) r^2= 0.60 to 0.65. Studies included in this analysis were: MRC1946, WH-II, ET2DS, ELSA, EAS, CAPS, BRHS, and BWHHS. The genotypes were phased using MACH1 http://csg.sph.umich.edu//abecasis/MaCH/tour/imputation.html, and missing SNPs were imputed using the 1000G Europeans as a reference haplotype population (308) (Minimac software was used: http://genome.sph.umich.edu/wiki/Minimac): the imputation strategy has been described in Howiw et al. (2012). For imputed genotypes, a probabilistic model is used which can be expressed either in terms of the probabilities of the three genotypes: P (homozygous minor allele), P (heterozygous), P (homozygous major allele) or as the "dose" of a stated allele. "Dose" is given by 2*P (homozygous for allele) + P (Heterozygous), when the probabilities are 1 or 0; it

gives three possible genotypes 0, 1, 2 which correspond to the additive model. In the case of imputed genotypes, the genotypes include non-integer values in the range 0-2. A recessive genotypic model was used in the *ANXA2* SNPs statistical analyses, thus imputed genotype data were converted to the recessive model. The patient genotype was represented as a binary character, the imputed dose <= 1.5 was called the genotype 0 (heterozygotes+homozygous major allele) otherwise, the genotype was set to 1 (homozygous minor allele genotype). Patients' lipid levels were first adjusted to take account of age, sex, the fixed effect of the study and genotype, then lipid measurements were partitioned according to patient genotype. The per-allele effect on the continuous variable was estimated using a linear model, where the variables were log-transformed before analysis due to skewed distribution. A p-value ≤ 0.05 was considered as significant.

2.5. Gene expression: the expression quantitative trait loci (eQTL)

The eQTL applies a direct association between genetic variation and gene expression levels of mRNA. Genetic studies of gene expression have identified thousands of eQTLs in different tissue types for the majority of human genes. The eQTL catalogue provides an important resource for identifying the functional impact of human genetic variation. In my project, I used three data sets as shown below.

GTEx portal

The Genotype-Tissue Expression (GTEx) project is a public resource database. It provides data and resources on eQTLs in a wide range of tissues of relevance to many diseases (309). On the home page (http://www.gtexportal.org/home/) under the gene association section, the Test your own SNP option was selected in order to find the association between SNPs and the tissue of interest. In the text entry box, the SNP ID, gene name, and tissue name were typed; the data output showed eQTL box plot, effect size, standard error, and p-value. In addition, Dr. Claudia Giambartolomei provided further analysis for *ANXA2* intergenic SNPs in tissues of interest from unpublished data.

The Advanced Study of Aortic Pathology (ASAP)

The ASAP uses the gene expression level quantification method to identify cardiovascular and metabolic associated-SNPs in relevant tissues including liver, aorta (medial and adventitial portions), mammary artery, and carotid plaque. Full description of the study is found in Folkersen et al. (2010). The ASAP study's statistical analyses of microarray expression data was

carried out by Dr. Per Eriksson and Dr. Lasse Folkersen (the Karolinska Institute, Stockholm, Sweden) using R version 2.2.3. (310). Genotypes for relevant SNPs were extracted and re-coded according to a linear additive model encoding genotypes numerically (0, 1, 2). Association with gene expression levels was then calculated with linear regression and the log2 transformed gene expression levels, using the *lm* function in R. Calculations were performed with or without gender as covariate on all samples or gender-specific samples as indicated.

eQTL meta-analysis in blood for lipid-regulatory and immune-mediated disease variants

Westra et al. (2013) established a data set, which is a large blood eQTL meta-analysis for lipid-regulatory and immune-mediated disease variants. This data set is available online (http://genenetwork.nl/bloodeqtlbrowser/). On the home page, the SNP ID was typed in the search box.

2.6. DNA amplification and purification

The Phusion®High-Fidelity PCR Kit **was used to amplify DNA fragments of interest.** DNA fragment primers were designed using Primer3 design program version 4.0 (http://primer3.ut.ee/) and are listed in Table 2-6. The PCR reaction and conditions are shown in Tables 2-7 and 2-8, respectively.

To amplify the *ANXA2 rs12900101* DNA fragment, the nested PCR approach was used. Nested PCR involves two sets of primers, used in two sequential runs of PCR, the second set intended to amplify a secondary target within the first run product.

Table 2-6: DNA fragment primers

DNA Fragment	primer	Sequence	Fragment size
WT *LDLR* promoter	Forward	5'-TTTGGAATAACAATGTAAAAACAATAT3'	594 bp
	Reverse	5'-CAGCCGCTGCAGGCAGTGTC3'	
***LDLR* rs6511720**	Forward	5'- GCCACTCAGTTTTACAAAAGAAA -3'	883bp
	Reverse	5'- TTGA TCAGTGCTTCTTTCCAA -3'	
***LDLR* rs141787760 + rs60173709**	Forward	5'- TGACTGGATGGAGAGCTTTGA -3'	643bp
	Reverse	5'- GCTTGCGGTTCAGATTCCAA -3'	
***LDLR* rs57217136**	Forward	5'-TGATCTGCCCAAAAGTGCTG -3'	814bp
	Reverse	5'-TGAAGATCTCGGCTGCAGAC -3'	
ANXA2 Promoter	Forward	5'-TTTTCGCTCTTGTTGCCCAG-3'	1323 bp
	Reverse	5'-GTTGGTCTCAGCAACTAGGC-3'	
ANXA2 rs11633032	Forward	5'-TGTGTGCTCTTTGGTTCAAGG-3'	585 bp
	Reverse	5'-CTACTTGGGAGGCTGAGGTG	
ANXA2 rs12900101 Nested PCR	Forward	5'-GGAGACCTCAGGATCATGGC-3'	2012 bp
	Reverse	5'-CCGCTTTTCCTTTGCTCTCA-3'	
ANXA2 rs12900101 target	Forward	5'- CCATGGTCTTAGGCAGCTCT -3'	369 bp
	Reverse	5'- GTGGGTGCACAGAAGTCAAG -3'	
ANXA2 rs17191344	Forward	5'-AACCACTTCAGAGGCGATCA3'	767bp
	Reverse	5'-AGGCACTCACCCAAAGAAGT3'	
ANXA2 rs116928563	Forward	5'- GCTGCCCCTGATGCTTATTT3'	409bp
	Reverse	5'- GTACTTTGTGGCCCTGCTTT3'	
ANXA2 rs17845226	Forward	5'- AAGCGTCATACTGAGCAGGT3'	253bp
	Reverse	5'- GGGTCTGTTGAAGGGAATGC3'	

Table 2-7: The Phusion®High-Fidelity PCR reaction

Component	20 µl reaction	Final concentration
Nuclease –free water	10.5 µl	-
5X Phusion GC buffer	4 µl	1X
10mM dNTPs	0.4 µl	200 mM
10 µM Forward primer	1 µl	0.5 mM
10 µM Reverse primer	1 µl	0.5 mM
70ng Template DNA	2 µl	<250ng
DMSO	0.6 µl	3%
Phusion DNA polymerase	0.2 µl	0.4 units/20 µl PCR

Table 2-8: The Phusion®High-Fidelity PCR conditions

Cycle step	Cycle	Temperature	Time
Initial Denaturation	1	98°C	30 seconds
Denaturation **Annealing** **Extension**	30	98°C 45-72°C 72°C	10 seconds 30 seconds 30 seconds
Final Extension	1	72°C	10 minutes
Hold	1	4°C	10 minutes

Agarose gel was commonly used either to confirm DNA size or to purify the PCR product. A standard 1.2% (w/v) agarose gel (Table 2-1) was used for showing products in the range of 500 bp-1kb. The agarose and water were added together and heated in a microwave on medium power for 3 minutes. Then, the TBE/EtBr was added to the agarose mix, which was poured into a pre-assembled module after it had cooled for a few minutes. A comb was used to form wells and it was left at room temperature for 30 minutes to solidify.

Usually, 3µl of 6X agarose gel loading buffer (Table 2-1) was added to 10µl of PCR product. Five microliters of 1Kb Plus DNA ladder was loaded into the well of each gel in order to determine fragment size. Gel electrophoresis was carried out in 1X TBE buffer at 120V for 45-60 minutes or longer until loading buffer had moved down two-thirds of the gel. Gels were analysed and photographed under UV illumination using the Syngene Gel Documentation system and Genesnap software version 6.04 or Syngene G:BOX Documentation system. For optimal gel image viewing, the standard parameters were maintained with brightness and zoom for each gel.

To purify a DNA fragment from the gel, the DNA bands were visualized in darkness using a UV light transilluminator and then were cut out of the gel and kept in individual micro-centrifuge tubes. The DNA gel bands were purified using the Illustra GFX PCR DNA and gel band purification kit (https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1314774443672/litdoc28951562AA_20110831111511.pdf). Five-hundred microliters of capture buffer type 3 was added to the gel slice, mixed by inversion and incubated at 60°C for 15–30 minutes until the agarose was completely dissolved. Once the agarose had completely dissolved and the capture buffer sample mix was yellow or pale orange in color, the mix was centrifuged briefly to collect the supernatant. Then, the supernatant was transferred into the assembled GFX MicroSpin column and collection tube, incubated at room temperature for 1 minute, spun at 16000×g for 1 minute and the flow through discarded. Then 500µl of wash buffer type 1 was added to the GFX MicroSpin column, spun at 16000×g for 1 minute, and the GFX MicroSpin column was transferred to a fresh DNase-free 1.5ml micro-centrifuge tube. Finally, 10–50µl of Nuclease–free

water was added to the center of the membrane in the assembled GFX MicroSpin column and sample collection tube, incubated at room temperature for 1 minute, and spun at 16000×g for 1 minute to recover the purified DNA. The DNA concentration was measured using a NanoDrop 8000. The purified DNA was kept at 4°C for short-term storage and -20°C for long-term storage

The standard Sanger sequencing (in the forward and reverse direction) was used as the confirmational method. The service is provided by Source Bioscience Life Sciences. The primers used in sequencing are listed in Table 2-2 in addition to pGL3 3'UTR and pGL3 3'UTR primers (listed in Table 2-9)

Table 2-9: Sanger sequencing primers

pGL3 3'UTR	sense primer	5'-TCAGGTTCAGGGGGAGGT-3'
	antisense primer	5´-GACGATAGTCATGCCCCGCG-3´
pGL3	Forward	5´-CTAGCAAAATAGGCTGTCCC-3´
	Reverse	5´-CTTTATGTTTTTGGCGTCTTCCA-3´

2.7. Cell culture

Two human liver hepatocellular carcinoma cell lines, Huh7and HepG2 were used. The Huh7 cell line was cultured in Dulbecco's Modified Eagle's Medium (DMEM with 4.5g/L glucose and L-Glutamine) supplemented with 10% fetal bovine serum (FBS), while the HepG2 cell line was cultured in Eagle's Minimum Essential Medium (EMEM) supplemented with 10% FBS and 0.5% essential amino acids. The cells were cultured and maintained in a 5% CO_2 humidified atmosphere at 37°C. The cells were grown to a confluence of about 80-90%.

To passage cells, the monolayer was briefly rinsed with warm 1x phosphate-buffered saline (PBS), and then pre-warmed (37°C) 0.05% Trypsin-EDTA solution was added and the mix incubated for 2-3 minutes in a 5% CO_2 humidified atmosphere at 37°C. When the cells detached from the flask surface, 12ml of complete growth medium with 10% FBS was added, and transferred to 15ml Falcon tubes.

Cell viability was estimated on an Advanced Detection and Accurate Measurement (ADAM) cell counter. Twenty microliters of cells were added to each of two micro-centrifuge tubes, followed by adding an equal volume of AccuStain Solution T to the first tube for total cell counting, and 20µl AccuStain Solution N to the second tube for non-viable cell counting. Both tubes were incubated at room temperature for 2 minutes, then 15µl of each mix was added to a designated place in a chip, which was then inserted in the ADAM cell counter.

DNA purification from cell cultured cells

QIAamp DNA Mini Kits was used to purify DNA from the cultured cells. In a 175T flask, cells were grown in a monolayer to 90-95% confluence, detached from the culture flask by trypsinization. A quantity of $5x10^6$ cells was transferred to a 1.5ml micro-centrifuge tube, centrifuged for 5 minutes at 300xg, and the supernatant removed completely without disturbing the cell pellet. Then, the cell pellet was re-suspended in PBS to a final volume of 200µl. Twenty microliters of QIAGEN Proteinase K and 200µl Buffer Al were added to the sample and mixed by pulse-vortexing, incubated at 56°C for 10 minutes, and briefly centrifuged. Then 200µl 100% ethanol was added to the sample, vortexed for 15 seconds, and briefly centrifuged. Then, the mixture was applied to the QIAamp Spin Column, centrifuged at 6000xg for 1 min. The QIAamp Spin Column was placed in a clean 2ml collection tube, 500µl Buffer AW1 was added, and centrifuged at 6000 x g for 1 minute. The QIAamp Spin Column was placed in a new clean 2 ml collection tube, 500µl Buffer AW2 was added and centrifuged at 6000xg for 3 minutes. Finally, The QIAamp Spin Column was placed in a clean 1.5 ml micro-centrifuge, 200µl dH_2O was added, incubated at room temperature for 1 minute, and then centrifuged at 6000xg for 1 minute.

The DNA was ready for downstream experiments including PCR amplification and sequencing. This method was used to confirm *ANXA2* rs116928563 genotype in the Huh7 cell line (Chapter 6), where DNA was purified from Huh7 cells and then sequenced.

.

2.8. Reporter gene expression assay

A Dual Luciferase assay was used to study eukaryotic gene expression by determining whether different genotypes affect luciferase expression **(**https://www.promega. co.uk/~/media/Files/Resources/Protocols/Technical%20Manuals/0/Dual%20Luciferase%20Reporter%20Assay%20System%20Protocol.pdf). A DNA fragment that contained the SNP of interest was inserted into a pGL3-basic vector or pGL3-promoter vector (vector construct is shown in Figure 2-1). Dual reporter refers to using two reporter enzymes: experimental and co-transfected reporters, the latter of which is used as an internal control. This system helps to normalize the activity of the experimental reporter to the activity of the internal control, which minimizes experimental variability caused by differences in cell viability or transfection efficiency. The pRL-TK vector was used as the co-transfected reporter. The pRL-TK vector contains a cDNA (Rluc) encoding Renilla luciferase and a herpes simplex virus thymidine kinase (HSV-TK) promoter to provide low to moderate levels of firefly luciferase expression in co-transfected mammalian cells.

2.8.1 Cloning

2.8.1.1. Preparation of pGL3-basic vector and DNA fragments for cloning

For cloning, the In-Fusion® HD Cloning Kit was used. This system requires 15bp overlap between DNA fragment and vector. To achieve this, DNA fragment primers were modified to have a 15bp overlap with the vector.

Primer modification

The DNA fragments end was modified using synthetic linkers, by a PCR amplification using primers containing sites for appropriate restriction enzymes and vector. To convert PCR Primers

into In-Fusion® primers, the bioinformatics website was used to design the fusion primers **(**http://bioinfo.clontech.com/infusion/convertPcrPrimersInit.do). The following criteria were considered for designing the In-Fusion PCR primers: A) the 5' end of the primer must contain 15 bases that are homologous to 15 bases at one end of the vector; and B) the 3' end of the primer must contain a sequence that is specific to the target gene, between 18–25 bases in length, and have a Guanine-Cytosine content between 40–60%, and a melting temperature (Tm) between 58 and 65°C. The Tm difference between the forward and reverse primers should be ≤ 4°C, and the last five nucleotides at the 3' end of each primer should not contain more than two guanines (G) or cytosines (C).

When In-Fusion® primers were designed, the pGL3-basic vector or pGL3-promoter vector and appropriate restriction enzymes were selected from the primer designing program menu. The primers that were used to amplify DNA fragments of interest are listed in Table 2-10. DNA fragments were amplified using the Phusion®High-Fidelity PCR kit and were followed by DNA purification as described previously.

Table 2-10: In-fusion primers sequence to amplify a target from human genome DNA

The restriction enzymes site is highlighted in yellow and the vector site highlighted in green.

polymorphism	Enzyme	Vector type	Orientation	Primer sequence	PCR product size
WT *LDLR* promoter	ACC65I HindIII	pGL3-basic	Forward	5'- TCTATCGATAGGTACCTTTGGAATAACAATGTAAAAACAATAT -3'	594bp
			Reverse	5'- CCGGAATGCCAAGCTTCAGCCGCTGCAGGCAGTGTC -3'	
***LDLR* rs6511720 T allele**	BamH I and Sal I	pGL3-basic + *LDLR* promoter	Forward	5'- AAATCGATAAGGATCCTGCCACTCAGTTTTACAAAAGAAA -3'	883bp
			Reverse	5'- AAGGGCATCGGTCGACTTGA TCAGTGCTTCTTTCCAA -3'	
***LDLR* rs141787760 C allele + rs60173709 T allele**	BamH I and Sal I	pGL3-basic + *LDLR* promoter	Forward	5'- AAATCGATAAGGATCCTGACTGGATGGAGAGCTTTGA -3'	643bp
			Reverse	5'- AAGGGCATCGGTCGACGCTTGCGGTTCAGATTCCAA -3'	
***LDLR* rs57217136 T allele**	BamH I and Sal I	pGL3-basic + *LDLR* promoter	Forward	5'-AAATCGATAAGGATCCTGATCTGCCCAAAAGTGCTG -3'	814bp
			Reverse	5'-AAGGGCATCGGTCGACTGAAGATCTCGGCTGCAGAC -3'	
***ANXA2* rs11633032 G allele**	BamH I and Sal I	pGL3-promoter	Forward	5'-AAATCGATAAGGATCCTGTGTGCTCTTTGGTTCAAGG-3'	585bp
			Reverse	5'-AAGGGCATCGGTCGACCACCTCAGCCTCCCAAGTAG-3'	
***ANXA2* rs12900101 C allele**	BamH I and Sal I	pGL3-promoter	Forward	5'-AAATCGATAAGGATCCCCATGGTCTTAGGCAGCTCT-3'	369bp
			Reverse	5'-AAGGGCATCGGTCGACGTGGGTGCACAGAAGTCAAG-3'	
***ANXA2* rs17191344 A allele**	Sal I	pGL3-promoter	Forward	5'-ATAAGGATCCGTCGACAACCACTTCAGAGGCGATCA-3'	767bp
			Reverse	5'-AAGGGCATCGGTCGACAGGCACTCACCCAAAGAAGT-3'	
***ANXA2* rs116928563 T allele**	BamH I and Sal I	pGL3-basic	Forward	5'- AAATCGATAAGGATCCGCTGCCCCTGATGCTTATTT-3'	409bp
			Reverse	5'- AAGGGCATCGGTCGACGTACTTTGTGGCCCTGCTTT-3'	

PGL3 vector digestion

The pGL3-basic vector or the pGL3-promoter was digested with restriction enzymes that will generate compatible ends for cloning. One microgram of plasmid DNA, 2.5µl of restriction enzyme A, 2.5µl of restriction enzyme B, and 5µl of 10X buffer 3 were added in a 50µl total reaction volume. The reaction was incubated in a water bath for 3 hours at 37^0C. Then, the vector digestion confirmation was performed using gel electrophoresis and followed by DNA purification as described previously.

2.8.1.2. Ligation

To ligate the compatible 'Sticky' ends of the linearized vector polylinker site and amplified DNA fragment, the In-fusion® HD cloning kit was used. For each fragment, the reaction was set up according to Table 2-11; volumes of vector and inserted DNA were varied according to their concentration after gel band purification. Each 10µl total reaction mix was set up and gently pipetted up and down, briefly centrifuged and incubated at 50^0C for 15 minutes and then the ligation mixture was placed on ice.

Table 2-11: In-fusion ligation reaction

Reagent	Volume (µl)
5× In-Fusion HD Enzyme Premix	2µl
Linearized vector	To 50-100 ng
DNA fragment	To 10-50 ng
dH2O	To 10µl
Total	10µl

2.8.1.3. Transforming vector/insert complex into NEB 5-alpha competent cells

The ligation mix was transformed into NEB 5-alpha competent *E.coli* cells (DH5α) to maintain the vector/insert complex. An amount of 2.5µl of ligation product was transformed to 50µl of DH5α using the heat shock method. This was achieved by incubating the transformation mixture on ice for 30 minutes, then in a water bath at 42^0C for 1 minute, and immediately incubated for 2 minutes on ice.

To grow the transformed cells, 500µl of warmed Luria Broth (LB) bacterial culture medium (Table 2-1) was added to the transformed cells and incubated at 37^0C for 1 hour on a 220rpm shaker. After an hour, the cells were plated onto an LB/ampicillin agar plate (Table 2-1) and incubated at 37^0C overnight. Since homologous recombination is restored on the second day, only corrected cells survive in a medium containing ampicillin.

2.8.1.4. Plasmid DNA purification

2.8.1.4.1. Mini-prep plasmid purification

One colony was picked from the agar plate and grown in 5ml LB/ampicillin medium (Table 2-1) at 37^0C overnight on a 220rpm shaker. The plasmid DNA was purified using QIAprep Spin Miniprep Kit (http://www.qiagen.com/Products/ Catalog/Sample-technologies/DNA-Sample-Technologies/Plasmid-DNA/QIAprep-Spin-Miniprep-kit#resources).Fifteen-hundred microliters of bacterial cells were transferred to 2ml micro-centrifuge and centrifuged for 2 minutes at 13,000rpm. Then, the bacterial cell pellet was re-suspended in 250µl Buffer P1, vortexed briefly. Two-hundred and fifty microliters of Buffer P2 was added and mixed by

inverting the tube 4–6 times. Three-hundred and fifty microliters of Buffer N3 was added and mixed immediately by inverting the tube 4–6 times, centrifuged for 10 minutes at 13,000rpm. Then the supernatant was applied to the QIAprep spin column, centrifuged for 30–60 seconds, and the flow-through discarded. Next the QIAprep spin column was washed by adding 500µl Buffer PB, centrifuged for 30–60 seconds, and the flow-through discarded. The QIAprep spin column was washed again using 750µl Buffer PE, centrifuged for 30–60 seconds, the flow-through discarded, and the spin column centrifuged for an additional 1 minute to remove residual wash buffer. Finally, the QIAprep column was transferred into a clean 1.5ml micro-centrifuge tube, and 20-30µl dH_2O was added to the centre of each QIAprep spin column to elute DNA, incubated at room temperature for 1 min, and centrifuged for 1 minute. Then, the sample underwent plasmid DNA screening in order to confirm the generated pDNA construct had the fragment of interest.

2.8.1.4.2. Plasmid DNA screening

For clone screening, the purified plasmid DNA construct was digested to confirm DNA fragment size for vector and fragments. The diagnostic digestion reaction required 2µl of DNA, 0.5µl of each restriction enzyme that was used in the In-fusion protocol, 1µl of 10X buffer 3, and dH_2O to a total volume of 10µl. The diagnostic digest reaction was incubated in a water bath for 2 hours at 37^0C. After that, the digested product underwent gel electrophoresis as described earlier.

In addition, standard Sanger sequencing (in the forward and reverse direction) was used as the second confirmational method. The pDNA samples were prepared for sequencing as recommended by the service provider. The sequencing primers used pGL3 3'UTR (Table 2-

5). All sequencing results were analyzed manually on FinchTV (Geospiza Inc), and aligned with a reference sequence using multiple sequence alignment by CLUSTALW (http://www.genome.jp/tools/clustalw/). After the fragment sequence had been confirmed, the bacterial glycerol stock was made to store the plasmid for long-term by adding 0.5 ml of culture to 0.5 ml of glycerol (50%) and stored at -80^{0}C for future experiments.

2.8.1.4.3. Maxi-prep plasmid purification

When the cloned sequence was confirmed, 15-20µl of mini-prep bacterial cell product was plated onto a LB/ampicillin agar plate and incubated at 37^{0}C overnight, then one colony was picked from the agar plate and inoculated into 150ml LB/ampicillin medium and incubated at 37^{0}C overnight on a 220rpm shaker. Then, bacterial cells were harvested and DNA purified using the GenElute™ HP Plasmid Maxiprep Kit. The bench protocol was performed as recommended by the manufacturer (http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Bulletin/2/na0200bul.pdf). First, the column was prepared by adding 12ml of column preparation solution to the column and allowing it to pass through. Second, a filter syringe was prepared by removing the plunger and placing the barrel in an upright position. Third, the pDNA purification process was performed using 150ml of an overnight culture pelleted at 5740rpm for 10 minutes and the supernatant discarded. Then, the cells were re-suspended in 12 ml of Resuspension/RNase Solution and vortexed. Twelve milliliters of lysis solution was added, the tube was gently inverted 6–8 times to mix, and incubated at room temperature for 3-5 minutes. Twelve milliliters of neutralization solution was added to the lysed cells that were gently inverted 6–8 times to mix. Then 9ml of binding solution was added and gently inverted 1–2 times to mix. Immediately, the mix was added to the barrel of the filter syringe and was let to sit for 5 minutes. The filter syringe was held over the column and the mix gently inserted to expel the cleared lysate, then centrifuged at 15000rpm for 2

minutes, and the flow-through discarded. Twelve milliliters of wash solution 1 was added to the column, centrifuged at 15000rpm for 2 minutes, the flow-through discarded. Then, 12 ml of wash solution 2 was added to the column, centrifuged at 15000rpm for 5 minutes, and the flow-through discarded. Finally, the column was transferred to a new collection tube, where 3 ml of Elution Solution was added to the binding column, and then the column/collection tube unit was centrifuged in a swinging bucket rotor at 15000rpm for 5 minutes.

For pDNA precipitation, 300µl of 3M NaAc (pH=6.5) and 2.2ml of isopropanol were added in a tube, mixed, and incubated at -20^{0}C for overnight. The mix was centrifuged at 15000 rpm for 45 minutes, and the isopropanol mix discarded. Then, 2ml of 70% ethanol was added to the pellet, centrifuged at 15000rpm for 20 minutes, the ethanol discarded, and the pellet air-dried for 15 minutes. The pDNA pellet was re-suspended in 150-200µl dH_2O. The pDNA construct was saved at 4^{0}C for short-term storage and at -20^{0}C for long-term storage.

2.8.2. Site-Directed Mutagenesis

To create variants, a site-directed mutagenesis (SDM) was performed using QuikChange Lightning Site-Directed Mutagenesis Kit. The SDM primers were designed (Table 2-12) and bench protocol was performed as recommended by the manufacturer (http://pef.aibn.uq.edu.au/wordpress/wp-content/blogs.dir/1/files/Support/MolecularCloning/Manuals/QuikChange_Lightning_SDM.pdf). The SDM reaction was prepared as shown in Table 2-13, and the reaction mix overlaid with paraffin. The SDM reaction underwent PCR according to the conditions shown in Table 2-14. When the PCR was completed, 2µl of the *Dpn* I restriction enzyme was directly added to amplification reaction, the reaction mixture was mixed by pipetting the solution up and down several times, spun down briefly and immediately

incubated at 37°C for 5 minutes to digest the parental supercoiled pDNA (the non-mutated). At the same time, 2µl of β-mercaptoethanol was added to 45µl of XL10-Gold ultracompetent cells, spun briefly, and the cells were incubated on ice for 10 minutes. Next, 2µl of the *Dpn* I-treated DNA was transferred to the ultracompetent cells, spun briefly, and the reaction was incubated on ice for 30 minutes, placed in a water-bath at 42^{o}C for 30 seconds, and on ice for 2 minutes. Finally, the SDM mixture was added to 500µl of warmed LB medium and was left for 1 hour at 37^{o}C on a shaker; then the total cell mixture was plated onto LB/Ampicillin agar. On the next day, a colony was collected and plasmid purification was performed as described previously.

Table 2-12: The QuikChange Lightning Site-Directed Mutagenesis (SDM) primer sequences

SNP ID	Primer	Sequence
***LDLR* rs6511720 G allele**	Forward	5'-CCAATCAACCTCTTAAGAGAAAATGTTAAGGAAGTCCTTAGG-3'
	Reverse	5'- CCTAAGACTTCCTTAACATTTTCTCTTAAGAGGTTGATTGG-3'
***LDLR* rs141787760 C allele deletion**	Forward	5'-GAGCCACCGCGCCCACCGGGGA-3'
	Reverse	5'-TCCCCGGTGGGCGCGGTGGCTC-3'
***LDLR* rs60173709 T allele deletion**	Forward	5'-GTAACTAACACACCAGTTCCTCTGGGAGGCCA-3'
	Reverse	5'-TGGCCTCCCAGAGGAACTGGTGTGTTAGTTAC-3'
***LDLR* rs57217136 T allele**	Forward	5'-CATTGGGTGTAGTCCAACAGGCCACCCTCG-3'
	Reverse	5'-CGAGGGTGGCCTGTTGGACTACACCCAATG-3'
***LDLR* rs57217136 G allele deletion**	Forward	5'-CATTGGGTGTAGTCCAACAGGCCACCCTCG-3'
	Reverse	5'-CGAGGGTGGCCTGTTGGACTACACCCAATG-3'
***ANXA2* rs11633032 A allele**	Forward	5'-TGCCCTTCGCTTGCTGTCTGAGCTGCTTC-3'
	Reverse	5'-GAAGCAGCTCAGACAGCAAGCGAAGGGCA-3'
***ANXA2* rs12900101 G allele**	Forward	5'-CACAGTGCCCCTCCTGGGGCACTGC-3'
	Reverse	5'-GCAGTGCCCCAGGAGGGGCACTGTG-3'
***ANXA2* rs17191344 G allele**	Forward	5'-GACTCCTCCACGACGGCCCTGCTCTCA-3'
	Reverse	5'-TGAGAGCAGGGCCGTCGTGGAGGAGTC-3'
***ANXA2* rs116928563 A allele**	Forward	5'-TGAGGCTGTCCCTGTTGGAAGAAAGCTCTGG-3'
	Reverse	5'-CCAGAGCTTTCTTCCAACAGGGACAGCCTCA-3'

Table 2-13: The QuikChange Lightning SDM reaction

Reagent	Volume
10x reaction buffer	2.5 µl
400ng of template pDNA	1.5 µl
100ng For SDM primer	1.5 µl
100ng Rev SDM primer	1.5 µl
dNTP mix	0.5 µl
QuikSolution	0.75 µl
dH_2O	16.25 µl
QuikChange Lightning Enzyme	0.5 µl
Total	25 µl

Table 2-14: The QuikChange Lightning SDM PCR conditions

Segment	Cycles	Temperature	Time
Initial Denaturation	1	95°C	2 minutes
Denaturation **Annealing** **Extension**	18	95°C 60°C 68°C	20 seconds 10 seconds 30 seconds/kb
Final Extension	1	68°C	5 minutes

For experimental efficiency purposes, the SDM pDNA products were re-cloned into a fresh vector in order to avoid any change within the vector that could have occurred during the SDM procedure. The digestion reaction for re-cloning is shown in Table 2-15. The digestion reaction was incubated in a water bath at 37°C overnight, the digestion process and gel electrophoresis were carried out and the DNA was purified as described earlier.

Table 2-15: Re-cloned SDM pDNA reaction

SDM plasmid construct		Vector	
SDM miniprep, up to 1µg	xµl	Vector, up to 1µg	5µl (400ng/ul)
Enzyme A	2.5µl	Enzyme A	2.5µl
Enzyme B	2.5µl	Enzyme B	2.5µl
10x buffer	5µl	10x buffer	5µl
dH_2O	xµl	dH_2O	35µl
Total	50µl	Total	50µl

T4 DNA ligation was used in re-cloning. The T4 ligation reaction component was the vector with an insert ratio of 3:1; 1µl of T4 ligase buffer, 1µl of T4 ligase and dH_2O to a total volume of 10µl was mixed gently, centrifuged briefly, and then incubated in a thermo-cycler block overnight via the program shown in Table 2-16. The next day, 2µl of the ligation reaction was transformed to 50µl of DHα5 competent cells. The transformed product was plated on LB/Ampicillin agar and after 24 hours, a colony was collected and plasmid purification was performed as described earlier.

Table 2-16: T4 ligation program

Stage	Temperature	Time
1	22^0C	10 minutes
2	16^0C	10 minutes
3	Go to 1 X5	
4	4^0C	30 seconds
5	27^0C	30 seconds
6	13^0C	30 seconds
7	Go to 4 X50	
8	4^0C	30 30 seconds

2.8.3. Transfection to mammalian cell

The original and mutated plasmid DNA was transfected into Huh7 cells. When the Huh7 cells were 85-95% confluent, they were plated into a 96-well plate ($2X10^4$ cells/ 100µl) and incubated in a humidified CO_2 atmosphere at 37°C for 24 hours. For each luciferase assay carried out in this study the plate included one row containing the puc18 plasmid as a negative control, one containing pGL3 promoter as a positive control, and one containing β-Gal as a transfection efficiency detector. Detection of β-Gal was by X-gal, when X-gal was cleaved by active β-Gal enzymes. This reaction leads to color the transfected cells blue, which was easily identified by light microscopy.

Transfection was carried out with Opti-MEM® reduced serum medium and Lipofectamine 2000. Two mastermixes: pRL-TK mastermix and lipid mastermix were prepared for 96-well plates. The first was pRL-TK mastermix (0.1µl of pRL-TK vector (10ng/µl), and 2,928µl of Opti-MEM® Medium). The pRL-TK activated the firefly luciferase gene in the pDNA and this measurement represented the co-transfected control reporter enzyme activity. The second was Lipid mastermix (66µl of Lipofectamine™ 2000 (1mg/ml) and 3,243µl of Opti-MEM® medium). Lipofectamine™ 2000 is a transfection reagent that contains lipid subunits capable of forming liposomes in an aqueous environment. In the presence of free DNA, liposomes encapsulate the nucleic acids to create an efficient delivery system. The liposome-nucleic acid complex is able to interact with the cell membrane of the Huh7 cells because of charge differences and then endocytosis subsequently releases the nucleic acids into the Huh7 cells cytoplasm.

The transfection procedure was performed as follows. In a micro-centrifuge tube 343µl of the pRL-TK mix was added to 7µl pDNA (400ng/µl), except in the case of puc18 and β-Gal. The samples were incubated at room temperature for 5 minutes. Then, lipid mastermix was prepared and incubated for 5 minutes at room temperature before adding 350µl of the mix to each reaction tube. After all mixes were added to pDNA constructs, the samples were incubated at room temperature for 20 minutes. To each well of the 96-well plate of Huh7 cells, 50µl of transfection mix was added and the plate was incubated in a humidified 5% CO_2 atmosphere at 37°C for 48 hours.

2.8.4. Assay of luciferase activity

2.8.4.1. Cell lysate and β-Gal

For the β-Gal assay, the medium was removed from the cells and the cells washed with 150µl of 1X PBS. Then 50µl of fixation solution (Table 2-1) was added per well and incubated at room temperature for 5 minutes. The fixation solution was removed and cells were washed twice with 150µl of 1X PBS. Next, 50µl of staining solution (Table 2-1) per well was added, incubated for 20 minutes at room temperature and then examined for transfection efficiency under the microscope. When transfection efficiency ≥30%, the transfected cells were eligible for the luciferase activity assay.

To perform the luciferase activity assay, cells were lysed by washing them twice with 150µl of 1xPBS. Then, 20µl of 1x lysis buffer (Luciferase Reporter Assay System kit) was added to each well and the plate was incubated for 20 minutes at room temperature on a shaker.

2.8.4.2. Dual-Luciferase Reporter

The activity of firefly luciferase and renilla luciferase in cell lysate was measured using the Dual Luciferase Reporter Assay System. The firefly luciferase was measured by the addition of Luciferase Assay Reagent II (LAR II) to generate a stabilized luminescent signal. After quantitation of the firefly luminescence, this reaction was quenched, and the renilla luciferase reaction was initiated by adding Stop & Glo reagent to the same well. The Stop & Glo reagent also produced a stabilized signal from renilla luciferase, which decays slowly over the course of the measurement.

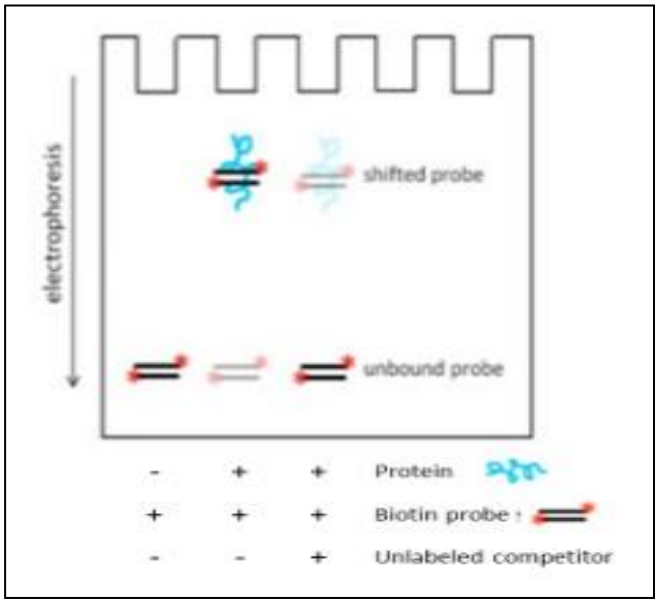
The Dual Luciferase Reporter Assays were performed using the Tropix TF microplate luminometer (Applied Biosystems/ Life Technologyies) and WinGlow software. The equipment injectors were set to dispense 50 µl of LAR II and Stop & Glo. All measurements were set for 2 second delay and 10 second read time. A 10µl aliquot of lysed cells from each well was transferred to the white Tropix TF microplate and placed in the luminometer. LAR II was dispensed to each well in the plate and the firefly luciferase activity was measured and recorded. Then Stop & Glo was dispensed and the renilla luciferase activity was measured and recorded. The results were recorded using WinGlow software. The ratio of firefly luciferase to renilla luciferase activity was used to compare the different expression between samples and plates. Twelve wells were assayed for each construct and transfection experiments were repeated at least 4 times.

2.9. Electrophoretic mobility shift assay (EMSA)

EMSA, also known as band shift assay, is a technique that is used to characterize nuclear protein-DNA binding interactions (311). The conventional EMSA protocol was performed to determine whether different alleles of a SNP of interest cause differential protein binding. The assay concept is that when protein is added to a SNP oligonucleotide sequence, and protein complex generates, the fragment becomes heavier and runs more slowly on the gel and is 'shifted' up the gel (Figure 2-5 (lane 2)). In my thesis, labeled and unlabeled (specific competitor) oligonucleotides were used to determine a SNP specific band. The specific competitor had identical sequence as labeled oligonucleotide, thus, the competitor competed out any specific interactions with labeled oligonucleotide, eliminating or reducing any positive shift result (Figure 2-5 (lane 3)).

Once DNA-protein binding is determined, the search will narrow to identify the transcription factor that could regulate the transcription of the gene of interest. To achieve this multiplexed-competitor-EMSA (MC-EMSA) was perform, seven sets of cocktails were used with each set having 10 consensus sequence oligonucleotides of known transcription factors. When one set competed out binding, the individual consensus sequences from this set were run in a further reaction (312).

Figure 2-5: Overview of the electrophoretic mobility shift assay (EMSA) method.
Adapted from http://www.licor.com/bio/applications/emsa/



2.9.1. Nuclear extraction

Mammalian eukaryotic cell nuclear extract containing proteins was isolated from a Huh7 cell line that had grown to 95% confluency in a T175 flask. First, the Huh7 cells were washed with 25ml 1XPBS, then 5 ml of Trypsin-EDTA solution (0.05%) was added to detach growing cells from the flask surface by incubation at 37^0C for 3 minutes. Second, 25ml of DMEM supplemented with 10% FBS was added to the cells, which were transferred to falcon tubes and centrifuged at 4^0C for 5 minutes at 1500rpm. Third, the cell pellet was re-suspended in 5ml ice-cold nuclear extract Buffer A (Table 2-1) with 50µl of 100X protease inhibitor cocktail, vortexed, then incubated for 10 minutes on ice, and centrifuged at 4^0C for 5 minutes at 1500rpm. The cell pellet was re-suspended in 2ml ice-cold nuclear extract Buffer A with 20µl of 100X protease inhibitor cocktail, vortexed and incubated in ice for 10 minutes, and then centrifuged at 4^0C for 5 minutes at 13000rpm. Finally, the cell pellet was re-suspended in 800µl ice-cold nuclear extract Buffer C (Table 2-1) with 16µl of 100X protease inhibitor cocktail, vortexed, then incubated on ice for 10 minutes (the vortex/incubation process was

repeated 4 times in total), and then centrifuged at 4^0C for 45 minutes at 13000rpm. The nuclear product (supernatant) was aliquoted into 50 µl volumes and stored at -80^0C.

2.9.2. Designing probes, biotin labeling and annealing probes

Forward and reverse probes were designed to be between 27-31bp. When the probe was designed the following criteria were considered; the forward and reverse probes were 100% complementary, and the probes length was between 27-31 nucleotides with SNP in the middle. EMSA probes that were used in this thesis are shown in Table 2-17.

Table 2-17: Electrophoretic mobility shift assay (EMSA) probes

LDLR rs6511720 (G>T)	
Forward wild type	ACCTCTTCCTTAAGA**G**AAAATGTTAAGGAAG
Reverse wild type	CTTCCTTAACATTTT**C**TCTTAAGGAAGAGGT
Forward rare variant	ACCTCTTCCTTAAGA**T**AAAATGTTAAGGAAG
Reverse rare variant	CTTCCTTAACATTTT**A**TCTTAAGGAAGAGGT
LDLR rs57217136 (T>C)	
Forward wild type	AATTCCATTGGGTG**T**AGTCCAACAGGCCAC
Reverse wild type	GTGGCCTGTTGGACT**A**CACCCAATGGAATTC
Forward rare variant	GAATTCCATTGGGTG**C**AGTCCAACAGGCCAC
Reverse rare variant	GTGGCCTGTTGGACT**G**CACCCAATGGAATTC
LDLR rs60173709 (T>-)	
Forward wild type	TAACTAACACACCAG**T**TTCCTCTGGGAGGCC
Reverse wild type	GGCCTCCCAGAGGAA**A**CTGGTGTGTTAGTTA
Forward rare variant	TAACTAACACACCAG*TTCCTCTGGGAGGCC
Reverse rare variant	GGCCTCCCAGAGGAA*CTGGTGTGTTAGTTA
LDLR rs141787760 (C>-)	
Forward wild type	GAGTGAGCCACCGCG**C**CCCACCGGGGATGAT
Reverse wild type	ATCATCCCCGGTGGG**G**CGCGGTGGCTCACTC
Forward rare variant	GAGTGAGCCACCGCG*CCCACCGGGGATGAT
Reverse rare variant	ATCATCCCCGGTGGG*CGCGGTGGCTCACTC
ANXA2 rs12900101 (C>G)	
Forward wild type	TAGGCAGTGCCCCAG**C**AGGGGCACTGTGCGG
Reverse wild type	CCGCACAGTGCCCCT**G**CTGGGGCACTGCCTA
Forward rare variant	TAGGCAGTGCCCCAG**G**AGGGGCACTGTGCGG
Reverse rare variant	CCGCACAGTGCCCCT**C**CTGGGGCACTGCCTA

ANXA2 rs11633032 (G>A)	
Forward wild type	TGAAGCAGCTCAGAC**G**GCAAGCGAAGGGCAA
Reverse wild type	TTGCCCTTCGCTTGC**C**GTCTGAGCTGCTTCA
Forward rare variant	TGAAGCAGCTCAGAC**A**GCAAGCGAAGGGCAA
Reverse rare variant	TTGCCCTTCGCTTGC**T**GTCTGAGCTGCTTCA
ANXA2 rs116255697 (A>G)	
Forward wild type	GTTCAAAGTACCACA**A**ATCTCTAAGGCAGGG
Reverse wild type	CCCTGCCTTAGAGAT**T**TGTGGTACTTTGAAC
Forward rare variant	GTTCAAAGTACCACA**G**ATCTCTAAGGCAGGG
Reverse rare variant	CCCTGCCTTAGAGAT**C**TGTGGTACTTTGAAC
ANXA2 rs17191344 (A>G)	
Forward wild type	TGAGAGCAGGGCC**A**TCGTGGAGGAGTC
Reverse wild type	GACTCCTCCACGA**T**GGCCCTGCTCTCA
Forward rare variant	TGAGAGCAGGGCC**G**TCGTGGAGGAGTC
Reverse rare variant	GACTCCTCCACGA**C**GGCCCTGCTCTCA

*Nucleotide deletion

2.8.2.1. Biotin labeling

EMSA probes were labeled with biotin using the Thermo Scientific Biotin 3´ End DNA Labeling Kit. This kit uses terminal deoxynucleotidyl transferase (TdT) enzyme to incorporate 1-3 biotinylated ribonucleotides onto the 3´-OH end of DNA (https://www.piercenet.com/instructions/2161290.pdf).

A standard 50µl biotin labeling reaction used 10µl of 5X TdT Buffer, 5µl of Biotin-11 DUTP, 0.5 µl TdT, and 5µl of primer. Biotin labeling mixes were incubated at 37^0C for 90 minutes then at 70^0C for 10 minutes to stop the reaction. Then, an equal volume of chloroform: isoamyl alcohol (24:1) was added to extract TdT; this mix was vortexed briefly and centrifuged for 2 minutes at 13000rpm. The labeled primer layer (top layer) was kept for the annealing step.

2.8.2.2. Annealing primers

Both labeled and unlabeled forward and reverse probes were annealed to create double

stranded probes. Equal amounts of labeled (5pmol/µl) or unlabeled (undiluted, at 100pmol/µl) complementary probes were mixed with $1/10^{th}$ total volume 10X annealing buffer (Table 2-1). The annealing mix was vortexed briefly and underwent an annealing action on the BioRad thermocycler: 95^0C for 5 minutes, then the temperature was gradually reduced per cycle 95-35^0C (-5^0C/cycle) for 3 minutes and then held at 4^0C. Probes were measured into 50µl aliquots and stored at -80^0C.

2.9.3. Polyacrylamide gel preparation

EMSA reactions were run on a 5% polyacrylamide gel (1.5 mm thickness). A 5% polyacrylamide gel was prepared using 6.25ml acrylamide 40% (w/v) (19:1), 2.5ml TBE 10X, 250µl ammonium persulphate 10%, 50µl TEMD, and dH_2O to a total volume of 50ml. A gel glass was placed on a frame holder of the Electrophoresis Modular System**.** In the 1.5 mm gap between the glass plates, the polyacrylamide mix was poured immediately and a comb was placed into the top of the glass through the polyacrylamide mix to create wells. The polyacrylamide mix was left overnight at room temperature to polymerize. When the polyacrylamide gel had set, the comb was removed and the gel wells were washed with 0.5X TBE. Then the glass holder was placed into a running tank containing 5L pre-cooled 0.5X TBE. The middle bay between glass plates was filled with 0.5X TBE to a height several millimeters above the top of the wells. The outside of the tank was filled with 0.5X TBE to above the wells, which reduced heat during electrophoresis. A 5µl EMSA pre-loading dye (Table 2-1) was used and the gel was pre-electrophoresed at 120V for 1 hour at 4^0C.

2.9.4. EMSA binding reaction and electrophoresis

Two EMSA reactions were performed, one has competitor probes while the other did not. The following mix was used for both reactions: 10 µl of 2X binding buffer (Table 2-1), 0.8 µl of poly dI.dC (1µg/ µl), 2µl of $MgCl_2$ (50mM), 0.8µl of BSA and 3µl of Huh7 cells nuclear extract and dH_2O to 20µl. For the competitor reactions 1µl of unlabeled probe was added and incubated at 4^0C for 30 minutes. For both reactions 1.5µl of biotin labeled primers were added, and then reaction tubes were incubated at 25^0C for 30 minutes, followed by adding 4µl of 6X gel loading dye to each reaction, then the reactions were loaded to pre-prepared polyacrylamide gel and run at 120V at 4^0C for 4 hours. The samples were electrophoresed until the bromophenol blue dye had migrated approximately 2/3 to 3/4 down the length of the gel. The free biotin-DNA duplex migrated just behind the bromophenol blue in the polyacrylamide gel.

2.9.5. EMSA blotting membrane

Once the gel electrophoresis was completed, the protein-DNA was blotted onto a nucleic acid nylon membrane in order to visualize the binding bands. Four pieces of Whatman paper and a piece of Hybond-N+ membrane were cut to about 15X10 cm and soaked for 10 minutes in 10X SSC buffer.

The blotting tank contained 20X SSC buffer (Table 2-1) with a perspex block in the middle. Over the perspex block, a piece of Whatman paper was laid with the ends submerged in the buffer. Over the prepared perspex block the blotting membrane assembly included two pieces of 10X SSC soaked Whatman paper, then on top of these the gel was placed with its edges covered with clingfilm. The Hybond-N+ membrane was placed on top of the gel and finally

the remaining two pieces of Whatman paper were placed on top of the membrane, and were flattened with a roller to ensure no air bubbles were trapped. The Whatman paper was essential for keeping all layers of the blotting membrane assembly moist. To create pressure on the blotting apparatus, a large stack of paper hand-towel and weights were added. This was left to blot overnight. After 18 hours, the membrane was retrieved from the blotting assembly and cross-linked using a UV Stratalinker 2400. The autocross link program (254nm bulbs for 60 seconds' exposure) was chosen to fix the protein-DNA complexes to the membrane.

2.9.6. Lightshift chemiluminescent EMSA detection system

The Thermo Scientific LightShift Chemiluminescent EMSA Kit was used to detect binding bands of protein-DNA complexes (https://www.piercenet.com/instructions/2160919.pdf). First, the Blocking Buffer and the 4X wash buffer were gently warmed to 37-50°C in a water bath until all particulates were dissolved. Then, 1X wash solution was prepared by adding 40mL of 4X wash buffer to 120mL of dH_2O. Second, the membrane was blocked by adding 20mL of Blocking Buffer and incubated for 20 minutes with gentle shaking. Next, 66.7µL Stabilized Streptavidin-Horseradish Peroxidase Conjugate was added to 20mL Blocking Buffer (1:300 dilution), and incubated for 20 minutes with gentle shaking. The membrane was transferred into a new container and washed with 20mL of 1X wash solution for 10 minutes with gentle shaking (the washing process was repeated 4 times in total). Then, 30mL of Substrate Equilibration Buffer was added to the membrane, and incubated for 10 minutes with gentle shaking. The membrane then was removed and the edge of membrane was carefully blotted on a paper towel to removed excess buffer. The membrane was placed on a clean glass plate. Substrate Working Solution (6mL Luminol/Enhancer Solution and 6mL Stable Peroxide Solution) was poured onto the membrane, incubated for 7 minutes, the membrane removed from the Working Solution and the edge of the membrane blotted on a paper towel for 2-5

seconds to remove excess buffer, and wrapped in plastic wrap, avoiding bubbles and wrinkles. Finally, to visualize the protein-DNA complex, the membrane was exposed to X-ray film or Syngene G:BOX Documentation system at various time points. The film was developed according to the manufacturer's instructions.

2.9.7. Multiplexed Competitor Electrophoretic Mobility Shift Assay (MC-EMSA)

The MC-EMSA concept uses seven cocktails (Table 2-18), each cocktail containing 10 competitors. These 70 competitors are consensus sequences for known transcription factor binding sites. When these unlabeled competitor sequences are incubated with the nuclear extract, a protein which binds to the SNP of interest instead binds to the relevant competitor probe, which causes band elimination (312).

The SNPs that showed a band shift in conventional EMSA were subjected to MC-EMSA analysis in order to identify the transcription factors. The cocktail of DNA competitors were made by mixing 50µl of forward and reverse primer sets (100pmol/µl), adding $1/10^{th}$ total volume of 10X annealing buffer, then vortexing the tubes gently. The multiplex competitor sequences used and their corresponding transcription factors are listed in Table 2-19. The primers were run on the 'anneal' program as described earlier. For the MC-EMSA, the reaction conditions were identical, except that the DNA competitor is a cocktail. If a particular set of competitor sequences is able to compete out the SNP of interest, then an individual competitor of this set undergoing EMSA separately will identify the precise competitor(s).

Table 2-18: MC-EMSA Cocktails

Unlabeled Multiplexed DNA-based competitors to commonly studied DNA-binding protein

Cocktail 0	Cocktail 1	Cocktail 2	Cocktail 3	Cocktail 4	Cocktail 5	Cocktail 6
AP1	E2F1	HIF1a	NFATc	RAR	Stat4	VDR
AP2	Egr	ISRE	NFkB	RXR	Stat5	YY1
AR	ER	HNF4	NR5A2	SIE	Stat56	ZEB
Brn3	Ets	IRF1	Oct-01	Smad	Tbet	HNF1
CBP	Ets1	MEF1	p53	Smad34	TFE3	ARP1
CDP	FAST1	MEF2	Pax5	Smuc	TFEB	NFY
CEBP	GAS	MIBP1	Pbx1	Sp1	TFIID	HNF3
cMyb	GATA	MycMax	Pit1	SRE	TGIF	BARP
CREB	Gfi1	NF1	PPAR	Stat1	TR	SREBP1
CTCF	GR	NFE2	PR	Stat3	USF1	HSF1

Table 2-19: Consensus sequence for competitor cocktails

Oligo name		consensus sequence
AP1	Forward	CGCTTGATGACTCAGCCGGAA
	Reverse	TTCCGGCTGAGTCATCAAGCG
AP2a	Forward	GATCGAACTGACCGCCCGCGGCCCGT
	Reverse	ACGGGCCGCGGGCGGTCAGTTCGATC
AR	Forward	GAAGTCTGGTACAGGGTGTTCTTTTTG
	Reverse	CAAAAAGAACACCCTGTACCAGACTTC
Brn3	Forward	CACAGCTCATTAACGCGC
	Reverse	GCGCGTTAATGAGCTGTG
CBP	Forward	AGACCGTACGTGATTGGTTAATCTCTT
	Reverse	AAGAGATTAACCAATCACGTACGGTCT
CDP	Forward	ACCCAATGATTATTAGCCAATTTCTGA
	Reverse	TCAGAAATTGGCTAATAATCATTGGGT
CEBP_F_R	Forward/ Reverse	TGCAGATTGCGCAATCTGCA
cMyb	Forward	TACAGGCATAACGGTTCCGTAGTGA
	Reverse	TCACTACGGAACCGTTATGCCTGTA
CREB	Forward	AGAGATTGCCTGACGTCAGAGAGCTAG
	Reverse	CTAGCTCTCTGACGTCAGGCAATCTCT
CTCF	Forward	GGCGGCGCCGCTAGGGGTCTCTCT
	Reverse	AGAGAGACCCCTAGCGGCGCCGCC
E2F1	Forward	ATTTAAGTTTCGCGCCCTTTCTCAA
	Reverse	TTGAGAAAGGGCGCGAAACTTAAAT
Egr	Forward	GGATCCAGCGGGGGCGAGCGGGGGCGA
	Reverse	TCGCCCCCGCTCGCCCCCGCTGGATCC
ER	Forward	GGATCTAGGTCACTGTGACCCCGGATC
	Reverse	GATCCGGGGTCACAGTGACCTAGATCC
Ets	Forward	GGGCTGCTTGAGGAAGTATAAGAAT
	Reverse	ATTCTTATACTTCCTCAAGCAGCCC
Ets1	Forward	GATCTCGAGCAGGAAGTTCGA
	Reverse	TCGAACTTCCTGCTCGAGATC
FAST1	Forward	TGTGTATTCA
	Reverse	TGAATACACA
GAS	Forward	AAGTACTTTCAGTTTCATATTACTCTA
	Reverse	TAGAGTAATATGAAACTGAAAGTACTT

GATA	Forward	CACTTGATAACAGAAAGTGATAACTCT
	Reverse	AGAGTTATCACTTTCTGTTATCAAGTG
Gfi1	Forward	TAAATCACTGC
	Reverse	GCAGTGATTTA
GR	Forward	AGAGGATCTGTACAGGATGTTCTAGAT
	Reverse	ATCTAGAACATCCTGTACAGATCCTCT
HIF1a	Forward	TCTGTACGTGACCACACTCACCTC
	Reverse	GAGGTGAGTGTGGTCACGTACAGA
ISRE	Forward	AAGTACTTTCAGTTTCATATTACTCTA
	Reverse	TAGAGTAATATGAAACTGAAAGTACTT
HNF4	Forward	CTCAGCTTGTACTTTGGTACAACTA
	Reverse	TAGTTGTACCAAAGTACAAGCTGAG
IRF1	Forward	GGAAGCGAAAATGAAATTGACT
	Reverse	AGTCAATTTCATTTTCGCTTCC
MEF1	Forward	GATCCCCCCAACACCTGCTGCCTGA
	Reverse	TCAGGCAGCAGGTGTTGGGGGGATC
MEF2	Forward	GATCGCTCTAAAAATAACCCTGTCG
	Reverse	CGACAGGGTTATTTTTAGAGCGATC
MIBP1	Forward	TCTTTTCCCA
	Reverse	TGGGAAAAGA
MycMax	Forward/ Reverse	GGAAGCAGACCACGTGGTCTGCTTCC
NF1	Forward	TTTTGGATTGAAGCCAATATGATAA
	Reverse	TTATCATATTGGCTTCAATCCAAAA
NFE2	Forward	TGGGGAACCTGTGCTGAGTCACTGGAG
	Reverse	CTCCAGTGACTCAGCACAGGTTCCCCA
NFATc	Forward	CGCCCAAAGAGGAAAATTTGTTTCATA
	Reverse	TATGAAACAAATTTTCCTCTTTGGGCG
NFkB	Forward	AGTTGAGGGGACTTTCCCAGGC
	Reverse	GCCTGGGAAAGTCCCCTCAACT
NR5A2	Forward	GATCAACGACCGACCTTGAG
	Reverse	CTCAAGGTCGGTCGTTGATC
OCT1	Forward	TGTCGAATGCAAATCACTAGAA
	Reverse	TTCTAGTGATTTGCATTCGACA
p53	Forward	TACAGAACATGTCTAAGCATGCTGGGG
	Reverse	CCCCAGCATGCTTAGACATGTTCTGTA
Pax5	Forward	GAATGGGGCACTGAGGCGTGACCACCG
	Reverse	CGGTGGTCACGCCTCAGTGCCCCATTC
Pbx1	Forward	CTCCAATTAGTGCATCAATCAATTCG
	Reverse	CGAATTGATTGATGCACTAATTGGAG
Pit1	Forward	TGTCTTCCTGAATATGAATAAGAAATA
	Reverse	TATTTCTTATTCATATTCAGGAAGACA
PPAR	Forward	AGGTCAAAGGTCA
	Reverse	TGACCTTTGACCT
PR	Forward	GATCCTGTACAGGATGTTCTAGCTACA
	Reverse	TGTAGCTAGAACATCCTGTACAGGATC
RAR	Forward	AGGGTAGGGTTCACCGAAAGTTCACTC
	Reverse	GAGTGAACTTTCGGTGAACCCTACCCT
RXR	Forward	AGCTTCAGGTCAGAGGTCAGAGAGCT
	Reverse	AGCTCTCTGACCTCTGACCTGAAGCT
SIE	Forward	GTGCATTTCCCGTAAATCTTGTCTACA
	Reverse	TGTAGACAAGATTTACGGGAAATGCAC
Smad	Forward	GTCTAGACCA
	Reverse	TGGTCTAGAC
Smad34	Forward	TCGAGAGCCAGACAAAAAGCCAGACATTTAGCCAGACAC
	Reverse	GTGTCTGGCTAAATGTCTGGCTTTTTGTCTGGCTCTCGA
Smuc	Forward	GGATCCCCCAACACCTGCTGCCTGA
	Reverse	TCAGGCAGCAGGTGTTGGGGGATCC

Sp1	Forward	ATTCGATCGGGGCGGGGCGAGC
	Reverse	GCTCGCCCCGCCCCGATCGAAT
SRE	Forward	GGATGTCCATATTAGGACATCT
	Reverse	AGATGTCCTAATATGGACATCC
Stat1	Forward	CATGTTATGCATATTCCTGTAAGTG
	Reverse	CACTTACAGGAATATGCATAACATG
Stat3	Forward	GATCCTTCTGGGAATTCCTAGATC
	Reverse	GATCTAGGAATTCCCAGAAGGATC
Stat4	Forward	GAGCCTGATTTCCCCGAAATGATGAGC
	Reverse	GCTCATCATTTCGGGGAAATCAGGCTC
Stat5	Forward	AGATTTCTAGGAATTCAATCC
	Reverse	GGATTGAATTCCTAGAAATCT
Stat56	Forward	GTATTTCCCAGAAAAGGAAC
	Reverse	GTTCCTTTTCTGGGAAATAC
Tbet_F_R	Forward/ Reverse	AATTTCACACCTAGGTGTGAAATT
TFE3	Forward	GATCTGGTCATGTGGCAAGGC
	Reverse	GCCTTGCCACATGACCAGATC
TFEB	Forward	CACGTG
	Reverse	CACGTG
TFIID	Forward	GCAGAGCATATAAAATGAGGTAGGA
	Reverse	TCCTACCTCATTTTATATGCTCTGC
TGIF	Forward	ACTCTGCCTGTCAAGCGAGG
	Reverse	CCTCGCTTGACAGGCAGAGT
TR	Forward	AGCTTCAGGTCACAGGAGGTCAGAGAG
	Reverse	CTCTCTGACCTCCTGTGACCTGAAGCT
USF1	Forward	CACCCGGTCACGTGGCCTACACC
	Reverse	GGTGTAGGCCACGTGACCGGGTG
VDR	Forward	AGCTTCAGGTCAAGGAGGTCAGAGAGC
	Reverse	GCTCTCTGACCTCCTTGACCTGAAGCT
YY1	Forward	CGCTCCCCGGGCCATCTTGGCGGCTGGT
	Reverse	ACCAGCCGCCAAGATGGCCGGGGAGCG
ZEB ZEB	Forward	GATCTGGCCAAAGGTGCAGGATC
	Reverse	GATCCTGCACCTTTGGCCAGATC
HNF1	Forward	GTTAATGATTAAC
	Reverse	GTTAATCATTAAC
ARP1	Forward	AGGTGACCTTTGCCCA
	Reverse	TGGGCAAAGGTCACCT
NFY	Forward	ATCAGCCAATCAGAGC
	Reverse	GCTCTGATTGGCTGAT
HNF3	Forward	GCCCATTGTTTGTTTTAAGCC
	Reverse	GGCTTAAAACAAACAATGGGC
BARP	Forward	TCACTCAAGTTCAAGTTATT
	Reverse	AATAACTTGAACTTGAGTGA
SREBP1	Forward	TTTGAAAATCACCCCATGCAAACTC
	Reverse	GAGTTTGCATGGGGTGATTTTCAAA
HSF1_F	Forward	GATCTCGGCTGGAATATTCCCGACCTGGCAGCCGA
		TCGGCTGCCAGGTCGGGAATATTCCAGCCGAGATC
CTCF motifs consensus sequence		
CTCF1	Forward	CCAGCCTTCAGCAGGTGGCACTGTTG
	Reverse	CAACAGTGCCACCTGCTGAAGGCTGG
CTCF2	Forward	TGTTTCAGCCACCTGCTGGGCAGCAG
	Reverse	CTGCTGCCCAGCAGGTGGCTGAAACA
CTCF3	Forward	CAGCCCTGCCCCCTGCCGGGCCCTGTG
	Reverse	CACAGGGCCCGGCAGGGGGCAGGGCTG
CTCF4	Forward	ATGATGGCCAGCAGAGGGCGCATGGC
	Reverse	ATGATGGCCAGCAGAGGGCGCATGGC

CTCF5	Forward	AGACAGCGCCCTCTGTTGGAATTAGCC
	Reverse	GGCTAATTCCAACAGAGGGCGCTGTCT
CTCF6	Forward	CTGATTTCCACTAGGGGGCCGTGGTG
	Reverse	CACCACGGCCCCCTAGTGGAAATCAG
CTCF7	Forward	CAACGGCGCCCCTTCGTGGTGATAAT
	Reverse	ATTATCACCACGAAGGGGCGCCGTTG
CTCF8	Forward	TCCTGGCCCCCTCTGCTGGACATCCT
	Reverse	AGGATGTCCAGCAGAGGGGGCCAGGA
CTCF10	Forward	CCAAGGCTGCCCCTGCTGGCAAAGGC
	Reverse	GCCTTTGCCAGCAGGGGCAGCCTTGG
CTCF11	Forward	AGCTACACCGACAGGGGGCGCCAGCC
	Reverse	GGCTGGCGCCCCCTGTCGGTGTAGCT
CTCF12	Forward	CTTGAGTGCCCTCTGGTGGGCAATAG
	Reverse	CTATTGCCCACCAGAGGGCACTCAAG

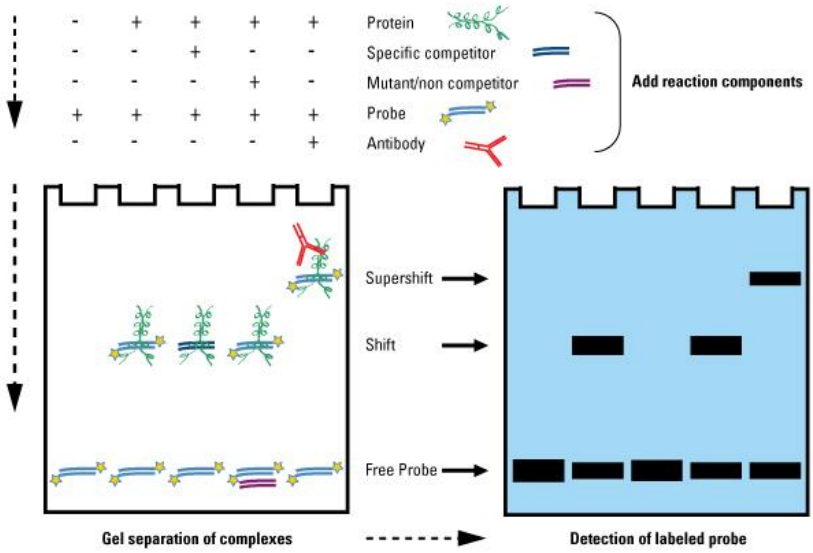
2.9.8. Supershift EMSA

To verify protein specificity, a supershift assay is often carried out using a specific monoclonal antibody to a putative transcription factor. For the supershift assay, the reaction conditions were identical to conventional EMSA, except that the DNA competitor was replaced with 1.5µl of monoclonal antibody to the transcription factor of interest. Anti-CTCF was added to rs17191344 reaction mix and anti-GATA1, anti-GATA2+3 and Egr1 (Table 2-1) were added to rs11633032 reaction mix individually.

Antibodies and binding partners may require a specific order of addition. Some antibodies may block binding of a protein to DNA if added before the complex is formed, thus adding the antibody after labeled probe is required. Thus, two antibody addition protocols were used, before and after addition of labeled probed to the nuclear extract. If the antibody recognizes the target protein two results are possible. First, when the antibody binds to the protein which then binds to the labelled probe, this creates a larger molecular weight complex antibody-protein-DNA, and this will move slower through the gel which will be observed as supershift on the gel (Figure 2-6). Alternatively, the antibody may prevent DNA/ protein interaction, where the antibody binds to essential site on the protein required for DNA binding, resulting

in a loss of the specific complex and absence on the gel. The disappearance of specific band is a negative result, but it can support the identification of the protein of interest.

Figure 2-6: Overview of the Supershift EMSA Method.
Adapted from http://www.piercenet.com/method/gel-shift-assays-emsa



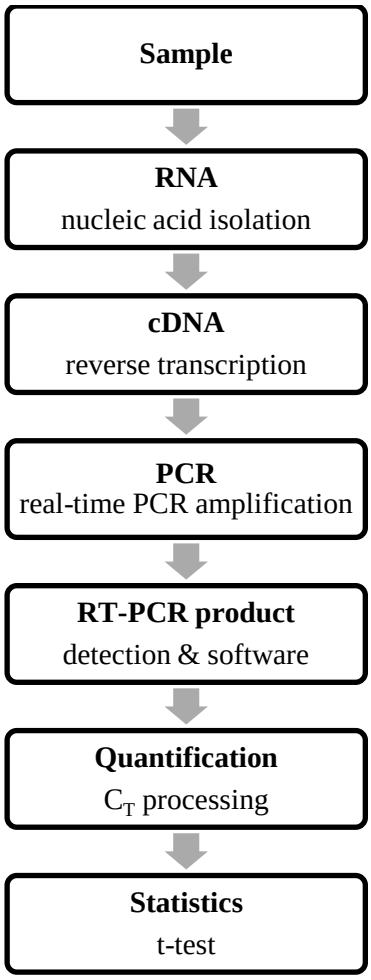
2.10. TaqMan Gene Expression

TaqMan Gene Expression was used to perform real-time reverse transcription-PCR (real-time RT-PCR), to measure the amount of mRNA expression associated with the target sequence and allele of interest in the presence and absence of treatment (miRNA).

The Real-Time PCR was performed with two-step RT-PCR, using two separate reactions. In the first reaction, total RNA is reverse transcribed into complementary DNA (cDNA),

and in the second reaction, the cDNA is amplified by PCR. Gene expression was quantified using the Relative Quantitation (RQ) method. The workflow of mRNA quantification using real-time RT-PCR is summarized in Figure 2-7.

Figure 2-7: The work flow of mRNA quantification using real-time RT-PCR



2.10.1. RNA isolation and DNase I treatment

Total RNAs were isolated from culture cells by TRIzol. First, the growth medium was removed from the culture dish, then 1ml of TRIzol was added to the culture dish per 10 cm^2 of dish surface, incubated for 5 minutes at room temperature. The cells were lysed in a culture dish by pipetting the cells up and down several times, and then the cell lysate was transferred to a 2ml micro-centrifuge tube, incubated at -80^oC for between 1 hour and overnight. To separate phases, 200µl of chloroform was added and incubated for 3 minutes at room temperature, and centrifuged at 4°C for 15 minutes at 12000xg, to separate the solution into an aqueous phase and an organic phase (RNA remains exclusively in the aqueous phase). The aqueous phase was transferred into a fresh tube; the RNA was recovered by precipitation with 500µl of 100% isopropyl alcohol, incubated at room temperature for 10 minutes, and centrifuged at 4°C for 15 minutes at 12000xg. The RNA pellet then was washed with 1ml of 75% ethanol and centrifuged at 4°C for 10 minutes at 7500xg. The washing step was repeated twice, and then the sample was stored at -80°C overnight. On the next day, the washing step was repeated once and the supernatant discarded. The RNA pellet then was re-suspended with 50µl of DEPC-H_2O and 5µl of NaAc RNase-free (3M), mixed, 150µl of 100% ethanol was added, incubated at -80°C for at least 60 minutes, centrifuged at 4°C for 10 minutes at 12000xg, and the supernatant was discarded. The RNA pellet was washed with 1ml of 75% ethanol and centrifuged at 4°C for 10 minutes at 12000xg. Residual ethanol was removed and the RNA pellet air-dried and then dissolved in 20µl of Nuclease-free water. RNA concentration and quality were examined using a Nanodrop spectrophotometer device and absorbance was measured at 260 nm and 280 nm.

To remove the genomic DNA from the RNA sample, it was treated with DNase I amplification grade. One microgram of RNA sample was added to 1µl 10X$MgCl_2$, 1µl DNase I to a total volume of 10µl. followed by incubation at 37°C for 30 minutes. Then 2µl of 250M EDTA was added to the mixture and incubated at 65°C for 10 minutes. The prepared RNA was stored at -80°C, to use as a template for reverse transcriptase.

2.10.2. Reverse transcription (RT)

In the first step of a two-step RT-PCR protocol, total RNA is reverse transcribed into cDNA. Two different protocols were performed: High Capacity cDNA and TaqMan® MicroRNA Reverse Transcription.

High Capacity RNA-to-cDNA KIT was used to synthesize single-stranded complementary DNA (cDNA) from total RNA. For a 20µl reaction: 10µl 2XRT buffers, 1µl RT Enzyme Mix, and 9µl RNA template (1µg) was added to a thin micro-tube followed by gentle mixing, briefly centrifuged and incubated on ice for 5 minutes. The RT thermal cycler condition was 60 minutes at 37°C, 5 minutes at 95°C, and 10 minutes at 4°C.

To synthesize single-stranded cDNA of Mir-155, Mir-155* and a small nuclear RNA U6 (used as endogenous control for miRNAs) from total RNA, the TaqMan® MicroRNA Reverse Transcription kit was used with specific miRNA primers. For TaqMan® MicroRNA Reverse Transcription the RNA sample was first diluted 1:50 as recommended by the manufacturer. The miRNA RT reaction consisted of 5µl of diluted RNA, 0.15µl dNTP mix, 1µl Multiscribe^{TM} RT enzyme,1.5 µl 10XRT buffer, 0.19 RNase inhibitor, 3µl

5XRT specific primers (mir155, mir155*, and U6), and 4.16 µl nuclease-free water. Then, the reaction tube was gently mixed, briefly centrifuged, and incubated on ice for 5 minutes. The RT thermal cycler condition was 30 minutes at 16°C, 30 minutes at 42°C, 5 minutes at 85°C, and 10 minutes at 4°C.

2.10.2.1. TaqMan Gene Expression Assays (20X)

TaqMan Gene Expression Assays (20X) (Table 2.1) are targeting protein-coding transcripts. All assays used in this thesis were of human species. The assays used for the work were either Inventoried assays or Made-to-Order assays, which were predesigned real-time PCR assays previously manufactured and either stored as stock or manufactured at the time of order.

TaqMan endogenous controls are a collection of predesigned assays for candidate controlled genes used to normalize for differences in RNA samples. TaqMan Gene Expression Assays and Endogenous Controls (20X) contain two unlabeled primers (18 µM per primer) and one 6-FAM dye-labeled TaqMan MGB probe (5 µM). A FAM dye reporter dye is linked to the 5’ end of the probe and a MGB at the 3’ end of the probe, while a nonfluorescent quencher (NFQ) is linked at the 3’ end of the probe.

2.10.2.2. The real-time PCR reaction and conditions

In the second step of a two-step RT-PCR protocol, the TaqMan® Universal Master Mix II was used. In 384-well TaqMan PCR reaction plate, the 20µl PCR reaction comprising 10µl of master mix II, 1µl TaqMan Gene Expression Assays (20X) (ANXA2, GAPDH,

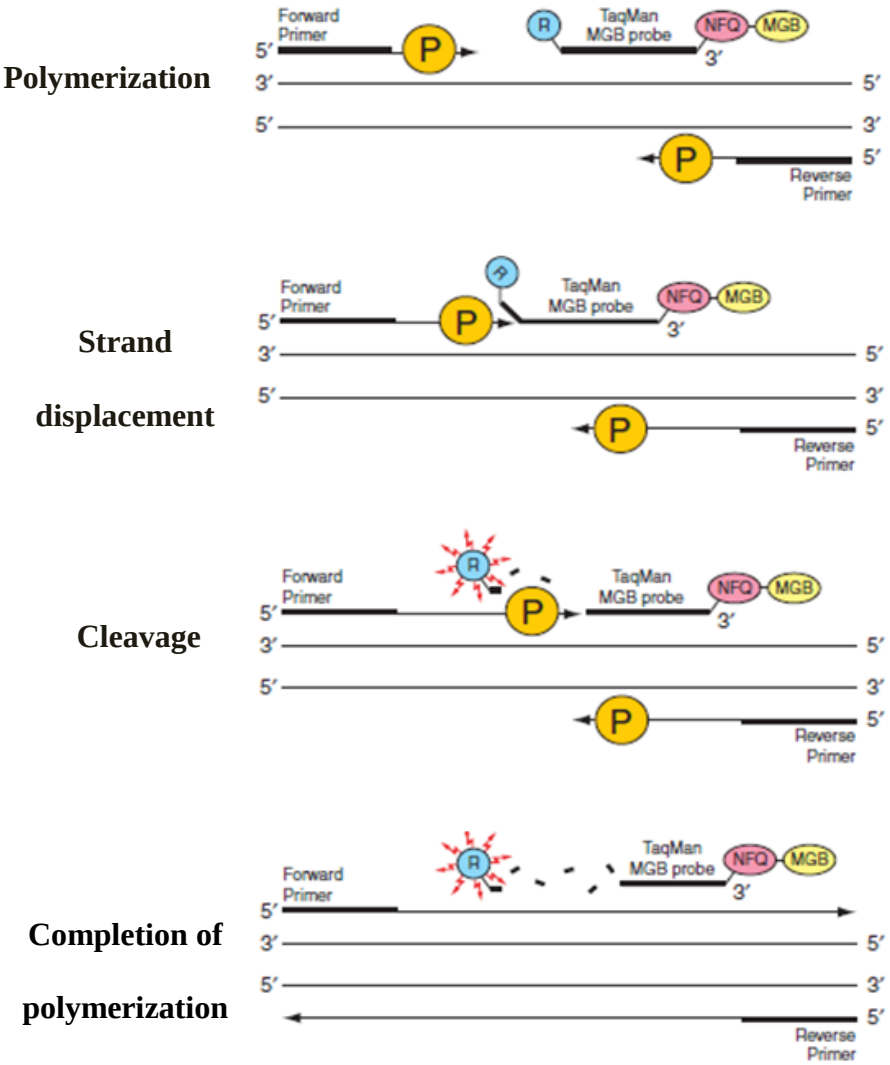
UBC, MiR155*, MiR155, U6), and cDNA product 2µl or 3µl (ANXA2, MiRNAs, respectively) was placed. The plate was covered by a TaqMan PCR adhesive cover and inverted several times to mix the reaction components and centrifuged briefly. Each sample had three replicates per experiment. A Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and ubiquitin C (UBC) were used to normalize the mRNA levels of the target gene (*ANXA2*).

The Real-Time PCR was performed in The TaqMan 7900HT fast Real-Time PCR System. The standard PCR reaction conditions were as follows: the reaction was incubated at 50°C for 2 minutes, then DNA polymerase was activated at 95°C for 10 minutes, followed by 40 cycles of denaturing at 95°C for 15 seconds for each and annealing/extension at 60°C for 1 minute.

TaqMan Gene Expression Assays are similar to TaqMan genotyping; there are several phases that occur during the PCR reaction. The TaqMan MGB probe anneals to a specific complementary sequence between the forward and reverse primer sites. When the probe is intact, the proximity of the reporter dye to the non-fluorescent quencher leads to suppression of the reporter fluorescence (Figure 2-8). Next, the DNA polymerase 5’ to 3’ activity along the DNA template specifically cleaves the probes that were hybridized to their target. When the cleavage occurred, the reporter dye separates from the non-fluorescent quencher which leads to increased fluorescence by the reporter dye and the probe fragment becomes displaced from the target. Polymerization of the strand continues.. During PCR process, ROX dye was the passive reference that acted as an internal reference to normalize the reporter dye signal.

Figure 2-8: TaqMan Gene Expression PCR stages.

The schematic shows the stages of the cycle of the TaqMan Gene Expression PCR. The TaqMan MGB probe is hybridized to the complementary sequence of the target between the forward and reverse primer sites. The primers bind to the DNA template followed by extension by DNA polymerase. When the probe hits with the primer, the reporter dye separates from the quencher dye, thus fluorescence is increased by the reporter. The increase in fluorescence occurs only if the target sequence is complementary to the probe and has been amplified during PCR, thus nonspecific amplification is not detected. NFQ= Non-fluorescent quencher, MGB = Minor groove binder, R=Reporter, and P= DNA polymerase. Adapted from https://tools.thermofisher.com/content/sfs/manuals/cms04128 0.pdf



2.10.2.3. TaqMan Gene Expression signal detection

Fluorescence was detected using the 7900HT fast Real-Time PCR System with Sequence Detection System (SDS) V2.4 software (http://tools.lifetechnologies.com/content/sfs/manuals/cms042114. pdf). Standard relative quantification (RQ) $\Delta\Delta C_T$ was carried out using the SDS software which induced an automatic threshold signal detector. When the PCR plate was inserted into the 7900HT fast Real-Time PCR System, the $\Delta\Delta C_T$ (RQ) program assigned real-time PCR to a plate coordinate in a text output file and in the form of real-time C_T plot based on the fluorescence signal detection.

In Real-Time PCR, reactions are characterized by the point in time during cycling when amplification of a target is first detected. The quantitative endpoint for real-time PCR is the threshold cycle (C_T). The C_T is defined as the fractional cycle number at which the fluorescence signal of the reporter dye passes the threshold. The relative quantification method compares the C_T of the target gene related to the passive reference (ROX).

Normalized reporter (R_n) was automatically calculated using the SDS relative quantification function. R_n is the ratio of the fluorescence emission intensity of the reporter dye to the fluorescence emission intensity of the passive reference dye (ROX) at two time points (baseline and R_n). The baseline is determined from the initial cycles of PCR in which there is little change in fluorescence signal, while R_n is fluorescence signal at the endpoint of PCR (40 cycles). Delta Rn (ΔR_n) is the magnitude of the signal generated by the specified set of PCR conditions ($\Delta R_n = R_n$ - baseline). The threshold is a level of ΔR_n used for C_T determination in real-time assays; this level is set to be above the baseline and sufficiently low to be within the exponential growth region of the

amplification curve (Figure 2-9) (https://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042115.pdf). Table 2-10 shows an example of a TaqMan gene expression output plot.

Figure 2-9: The schematic shows how TaqMan gene expression amplification plot is interpreted and the calculation of C_T.

The graph represents how ΔR_n is calculated and indicates the point of significant exponential growth that determines C_T based on the increase in fluorescence signal detected with PCR amplification.

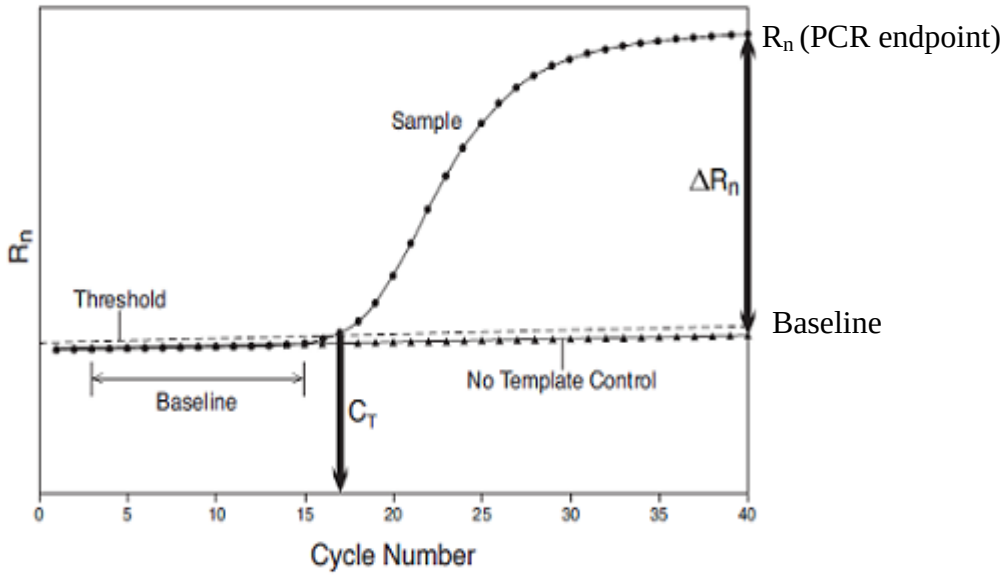
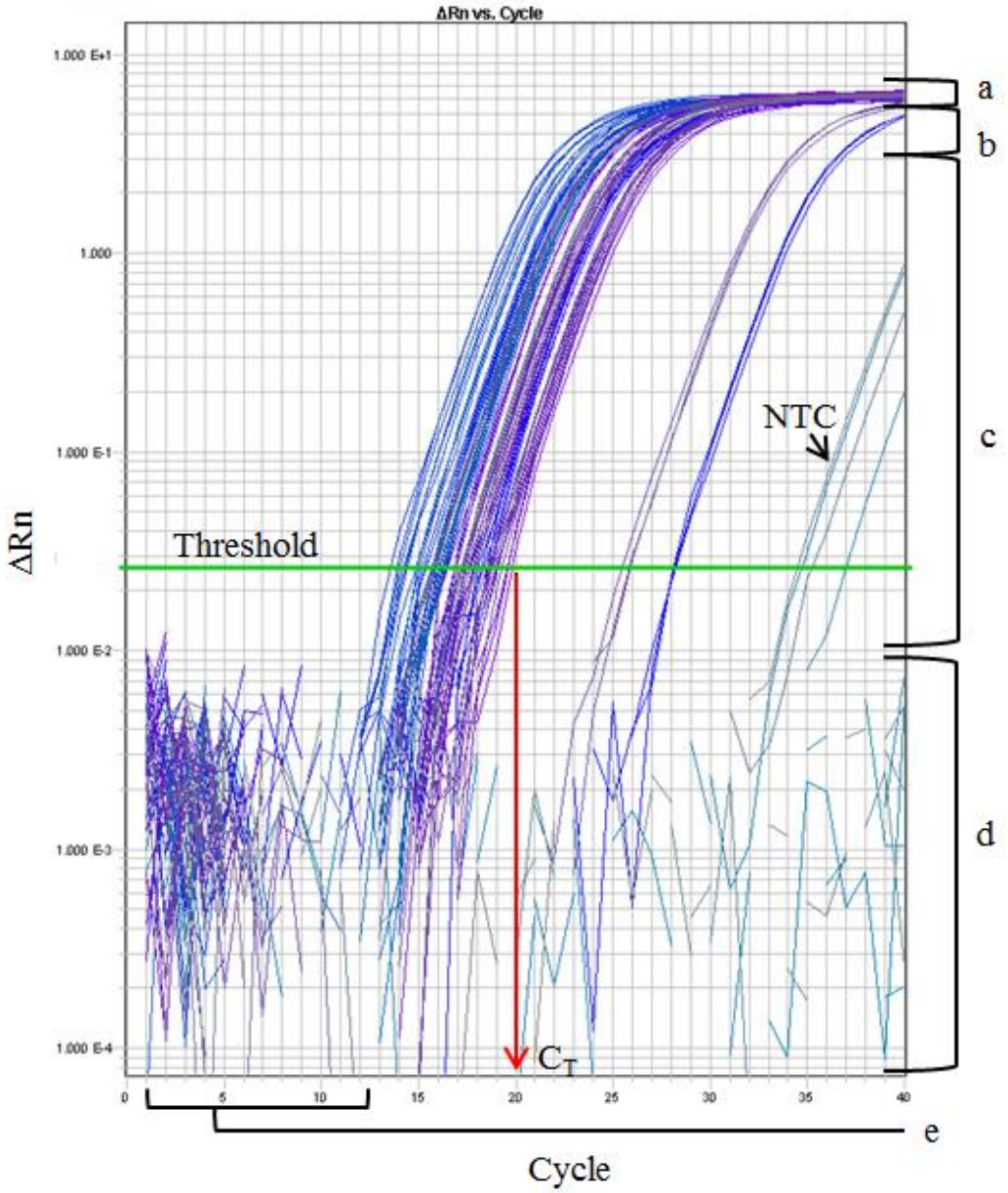


Figure 2-10: TaqMan gene expression output plot (amplification plot)
The figure is an example of a typical results plot for *ANXA2* TaqMan gene expression using SDS software. The plot allows interpretation of whether the assay has worked. The PCR was run for 40 cycles. The point at which the curve intersects the threshold (horizontal green line) is the $C_{T.}$ In this example, the samples reached the threshold between 13-18 cycles, while the No template control (NTC) samples did not reach the threshold before 40 cycles. a) plateau phase, b) linear phase, c) exponential phase, d) background, and e) baseline.



2.10.3. Transfection to mammalian cells: *ANXA2* rs116928563 (T>A) and mir-155-3p

To determine the effect of miR-155 on both alleles of the *ANXA2* rs116928563 SNP, constructs that have the common allele (T) and the rare allele (A) were created using a cloning method and transfected to mammalian cells using two-step transfection: plasmid DNA transfection and MiRNA transfection.

2.10.3.1. Cloning

The 409bp of 3'UTR of human *ANXA2* was amplified using Phusion®High-Fidelity PCR Kit as described earlier, and inserted into the enhancer site of pGL3-basic luciferase reporter vector after the SV40 polyadenylation signal at Sal I and BamH I. The *ANXA2* rs116928563 variant was generated using a QuikChange Site-Directed Mutagenesis kit as described earlier. All constructs were confirmed by sequencing.

2.10.3.2. Plasmid DNA transfection

The original and mutated plasmid DNA were transfected into Huh7 cells. When the Huh7 cells were 90-95% confluent, they were plated into a 20 cm^2 cell culture dish ($3X10^6$ cells/ 5ml) and incubated in a humidified CO_2 atmosphere at 37°C for 24 hours. For each TaqMan expression carried out in this study, three dishes were prepared; one contained the *ANXA2* rs116928563 common allele plasmid, another contained the *ANXA2* rs116928563 rare allele plasmid, and the third contained the β-Gal as transfection efficiency detector.

Transfection was carried out with Opti-MEM® reduced serum medium and Lipofectamine 2000. The transfection procedure was performed as follows; in micro-centrifuge tube 500µl of Opti-MEM® was added to 20µl of pDNA (400ng/µl). Then, lipid mastermix was prepared (60µl of Lipofectamine™ 2000 (1mg/ml) and 1440µl of Opti-MEM® medium) and incubated for 5 minutes at room temperature before adding 500µl of the mix to each reaction tube, which was incubated at room temperature for 20 minutes. To each dish of the Huh7 cells, 1000µl of transfection mix was added and the plate was incubated in a humidified 5% CO_2 atmosphere for 24 hours at 37°C.

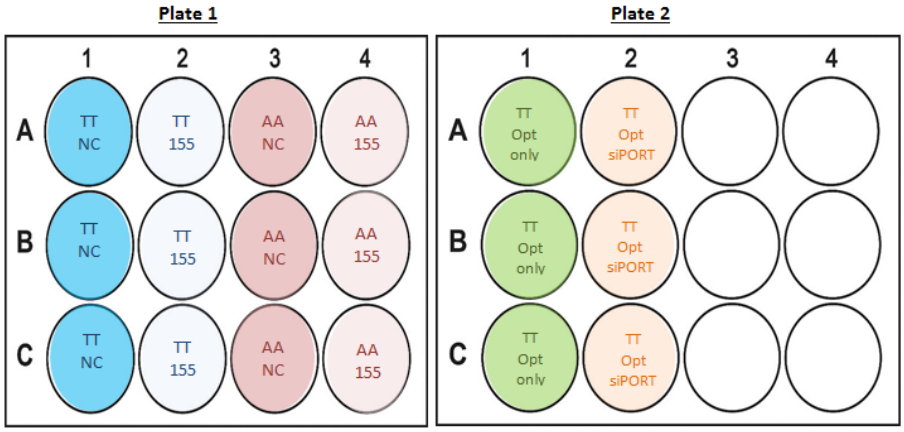
For the β-Gal assay, the medium was removed from the cells and washed with 5 ml of 1XPBS. Then 500µl of fixation solution was added to the cells and incubated for 5 minutes at room temperature. The fixation solution was removed and the cells were washed twice with 150µl of 1XPBS. Next, 500µl of staining solution was added, incubated for 20 minutes at room temperature and then examined for transfection efficiency under the microscope. When the transfection efficiency of the cell was ≥30%, the transfected cells were eligible for the second transfection.

2.10.3.3. MiRNA-155* transfection

Human mir-155* (mir-155-3p) was delivered to Huh7 cells that contained *ANXA2* rs116928563 SNP constructs using has-mir-155-3p mimic. Transfection was performed using siPORT NeoFX Transfection Agent. This Transfection Agent adopts the reverse transfection method, which involves simultaneous transfection and plating. For each TaqMan expression carried out in this study, two 12 well plates (6 column) would include one column containing the rs116928563 T allele transfected with negative control miRNA mimic (NC), one

containing the rs116928563 T allele transfected with mir-155* mimic, one containing the rs116928563 A allele transfected with NC, one containing the rs116928563 A allele transfected with mir-155* mimic, columns five and six contained Opti-MEM® medium, and Opti-MEM and siPORT NeoFX Transfection Agent, respectively, which are used as negative controls. Plate design is shown in Figure 2-11.

Figure 2-11: MiRNA-155* transfection plate design



The transfection procedure was performed as follows (the volume stated is per well); first, adherent cells of the first transfection was trypsinized and diluted in normal growth medium ($1.6X10^5$ per well/1ml growth medium), and incubated at room temperature for 10 minutes. Second, 3µL of siPORT™ NeoFX™ Transfection Agent was diluted in 50µL of Opti-MEM medium, and incubated 10 minutes at room temperature. Three microlitres of miRNA (10 µM) was diluted in 50µL of Opti-MEM medium. Third, diluted siPORT™ NeoFX™ Transfection Agent and diluted RNA were combined, mixed by pipetting up and down and incubated for 10 minutes at room temperature. Next RNA/siPORT™ NeoFX™ Transfection Agent transfection complexes were dispensed into wells of a culture plate. Finally, 1ml of cell

suspension was overlaid onto the transfection complexes and the plate gently tilted to mix, then incubated in normal cell culture conditions for 24 hours.

Total RNAs were isolated from culture cells by TRIzol as previously described, and TaqMan expression assay was performed.

2.10.4. TaqMan gene expression analysis for ANXA2 and MiR-155 isoforms

To assess whether the *ANXA2*, the *MiR-155** and *MiR-155* (*MiR155-5p)* are expressed in two of the human liver hepatocellular carcinoma cell lines (Huh7 and HepG2), TaqMan gene expression was carried out.

Into 6 well plates, $2X10^5$ cells were added per well (one plate for Huh7, another for HepG2). When the cells were 90% confluent, RNA was isolated and the expression assay performed as described above. They were compared and normalized to housekeeping genes: *GAPDH* for *ANXA2*, and *U6* for MiR-*155** and *MiR155.* Each sample was run in triplicate for each assay on an ABI 7900HT fast Real-Time PCR System.

2.10.5. TaqMan gene expression analysis for ANXA2 rs116928563 (T>A)

To determine the effect of miR-155 on both alleles of the *ANXA2* rs116928563 SNP, human mir-155* was delivered to Huh7 cells that contained *ANXA2* rs116928563 SNP as described above. Each allele of SNP was transfected with both mir-negative control (mir-NC) and mir-155* mimic. Each sample was run in triplicate for each assay, and each assay had two

negative controls: the SNP common allele (T allele) and Opti-MEM medium, and the SNP common allele and siPORT NeoFX Transfection Agent.

Huh7 cells were first transfected with *ANXA2* rs116928563 SNP constructs (WT and mutant). After 24 hours the constructs were transfected with either mir-155* mimic or mir-NC and incubated for another 24 hours. RNA was isolated and the expression assay performed as described above. The samples expression were compared and normalized to housekeeping genes: *GAPDH* or *UBC* for *ANXA2*, and U6 for MiR-155* and MiR155. Each sample was run in triplicate for each assay on an ABI 7900HT fast Real-Time PCR System.

2.10.6. Quantitative real-time PCR data analysis

To analyze quantitative real time PCR data, the relative quantification method was used which describes the change in expression of the target gene relative to a reference group or control. ΔCt and ΔΔCt were calculated using the following mathematical formula: $\Delta C_T = C_{T\ target} - C_{T\ reference}$; $\Delta\Delta Ct = \Delta C_{T\ target} - \Delta C_{T\ control}$. ΔC_T was calculated separately for each well. These ΔCt values were then averaged before proceeding with the $2^{\wedge}\Delta\Delta C_T$ calculation. The mean relative expression difference between variant genotype was determined by a two-sample *t*-test. Results were analyzed using a spreadsheet.

2.11. Familial hypercholesterolemia (FH) mutation screening

2.11.1. FH subjects

2.11.1.1. UK- FH

Sixty-nine samples were collected from several UK clinics. Participants' information about lifestyle habits, ethnicity, height, weight, blood pressure, treatment and family history of CHD were recorded. Standard assays for plasma lipids were carried out on samples providerd. In this study there was no clinical data for 18 subjects, however the FH mutation screening was performed based on physicians requests.

2.1.1.1. Iran- FH

Ninety-five samples were collected from 16 probands with clinical FH and their families from different pediatric and endocrinology clinics in five different cities (Ahvaz, Bandar Abbas, Esfarayen, Mashhad and Torbat Heydariye), which represent three regions of Iran (Figure 2-12). Clinical diagnosis was made according to the Dutch Lipid Clinic Network Criteria (DLCNC) as previously described (144).The probands were diagnosed with clinical definite FH (DFH) if they had a total cholesterol >6.7 mmol/l or LDL-C >4.0 mmol/l if <16 years and/or TX in a first or second degree relative. They were diagnosed with clinical probable FH (PFH) if they had a high TC and LDL-C as for DFH and a family history of myocardial infarction (MI) in a first degree relative<60 years or a family history of total cholesterol >7.5 mmol/l in a first or second degree relative. Dr Amir Hossein Sahebkar provided the DNA samples and participants' clinical information on lipid plasma levels and family history of CHD.

Figure 2-12: Map of the Islamic Republic of Iran

The map shows the five regions of Iran and their capital cities (Tehran, Isfahan, Tabriz, Kermanshah, and Mashhad). The investigated samples were collected from five cities shown by a red dot and these cities occur in three regions of Iran. Provided by Dr Amir Hossein

2.11.2. DNA sample collection, purification and quantification

The Oragene•DNA (OG-500) kit (http://www.dnagenotek.com/ROW/pdf/PD-BR-017.pdf) was used to collect saliva samples from the UK participants as recommended by the manufacturer, and prepIT®•L2P (PT-L2P) reagent was used for the purification of genomic DNA. Each sample was mixed in the DNA Genotek kit by inversion and gentle shaking for a few seconds and incubated at 50°C in a water bath for 1 hour or in an air incubator for a minimum of 2 hours. Five-hundred microliters of the mixed sample was transferred to a 1.5mL micro-centrifuge tube. Then, 20µL of PT-L2P was added and the sample was mixed by vortexing for a few seconds and incubated on ice for 10 minutes followed by centrifugation at room temperature for 5 minutes at 15000xg. The clear supernatant was transferred into a fresh micro-centrifuge tube. To 500µL of supernatant, 600µL of 100% ethanol was added and mixed gently by inversion 10 times. The sample was left at room temperature for 10 minutes to allow the DNA to fully precipitate and then centrifuged at room temperature for 2 minutes at 15000xg. The supernatant was discarded and the DNA pellet was washed with 250µL of 70% ethanol, the sample left at room temperature for 1 minute, and the supernatant discarded. One-hundred microliters of TE solution was added to dissolve the DNA pellet and vortexed for at least 5 seconds and incubated at 50°C for 1 hour with occasional vortexing. The sample was stored at -20°C for long-term storage and in TE at 4°C for up to 2 months (http://www.dnagenotek.com/US/pdf/PD-PR-006.pdf).

For Iranian participants, blood samples (3 mL) was obtained from all patients and their available parents and were collected in EDTA tubes. Genomic DNA was extracted from

peripheral blood samples using the standard salting-out method (313) with the High Pure PCR Template Preparation Kit (version 20).

Genomic DNA was quantified using a Nanodrop spectrophotometer device. To assess DNA quality, the absorbance ratio was measured at 260 nm and 280 nm. The optimized DNA absorbance ratio values A260/A280 = 1.8-2.0. The DNA concentration was optimized to be 50ng/ul for targeting sequencing and 30ng/ul for high resolution melting (HRM) using dH_2O and TE buffer.

2.11.3. Targeted Sequencing

Targeted sequencing was performed by Dr. Marta Futema and she provided sequencing data as a BAM file. Primers to amplify coding regions (±25 base pairs (bp)) of the three autosomal dominant FH genes (*LDLR*, *APOB*, *PCSK9*), and the autosomal recessive FH gene (*LDLRAP1*), and six LDL-C-associated SNPs [SNPs ID (nearby gene): rs629301 (*CELSR2*), rs1367117 (*APOB*), rs6544713 (proxy of rs4299376, *ABCG5/8*), rs6511720 (*LDLR*), rs429358 (*APOE*), and rs7412 (*APOE*)] for targeted sequencing were designed using the Illumina DesignStudio. The length of amplicons was set to 250bp. The library preparation was performed using the TruSeq Custom Amplicon (TSCA) v1.5 kit (http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistrydocumentation/sampleprepstruseq/truseqcustomamplicon/truseq-custom-amplicon-15-reference-guide-15027983-02.pdf) and sequencing (in both directions) was done using the Illumina MiSeq platform.

The criteria for standard variant calling pipeline were: coverage ≥30×, minimum of five reads for an altered allele, Phred quality ≥20, and a strand bias filter. To ensure that variants were not missed a sensitive pipeline was used (coverage ≥15×, minimum of two reads for an altered allele, Phred quality ≥ zero, no strand bias filter.

2.11.4. Filtering of the variants

The BAM file provide all variants in different target genes transcripts, thus only variants that were reported in reference gene transcripts (*LDLR*: ENTS00000558518, *APOB*: ENTS00000233242, *PCSK9*: ENTS00000302118, and *LDLRAP1*: ENTS 00000374338) were nominated. The nominated variant also was filtered based on its frequency and functional affects. Variants were flagged as rare (frequency<0.005) and novel (frequency=0) according to their frequency in publicly available databases including the 1000 Genomes project. Also, variants were flagged as functional, when they were most likely to affect a protein's function, that is, non-synonymous, stop gain, stop loss, frame-shift deletions and insertions, and splice site variants.

2.11.5. Analysis of the variants

2.11.5.1. Bioinformatics

Variants within FH known genes (*LDLR*, *APOB*, *PCSK9*, *LDLRAP1*) were assessed on the basis of their frequency and functionality using *in-silico* prediction tools (SIFT, PolyPhen 2 and mutation Taster), Also, the variants were described as causing FH if they were reported on the UCL-FH mutation database (http://www.ucl.ac.uk/ldlr/Current/)

(65, 314), which is supported by evidence in the literature. Variants were described as possible FH-causing if they were unreported in the database but were predicted to be pathogenic *in-silico* and highly conserved (the UCSC genome browser used to analyze the conservation level of the variant position).

2.11.5.2. Sanger sequencing

Sanger sequencing was used to confirm all identified mutations by NGS. Oligonucleotide primers for PCR were designed to cover, where appropriate, intron–exon junctions and up to 60bp of the introns (146). Primers were designed to amplify the promoter and *LDLR* exons, a part of exon 26 of *APOB* that included the p.R3527Q mutation and p.D374Y (exon 7) of the *PCSK9* gene for screening (Table 2-20). PCR was carried out in the Rotor-Gene6000. Exon 4 of the *LDLR* was screened by four overlapping PCR fragments (4.1, 4.2, 4.3and 4.4) and exon 10 by two overlapping fragments (10.1 and 10.2).

Table 2-20: PCR primers used to amplify DNA fragments of *LDLR, APOB,* and *PCSK9*

Target	Orientation	Sequences
LDLR promoter	Forward	CAGCTCTTCACCGGAGACCC
	Reverse	ACCTGCTGTGTCCTAGCTGG
LDLR- Exon 1	Forward	ACTCCTCCCCCTGCTAGAAACCTCA
	Reverse	CTATTCTGGCGCCTGGAGCAAGCC
LDLR- Exon 2	Forward	TTGAGAGACCCTTTCTCCTTTTCC
	Reverse	GCATATCATGCCCAAAGGGG
LDLR- Exon 3	Forward	TCAGTGGGTCTTTCCTTTGAG
	Reverse	CAGGACCCCGTAGAGACAAA
LDLR- Exon 4.1	Forward	TGGTGTTGGGAGACTTCACA
	Reverse	CACTCATCCGAGCCATCTTC
LDLR- Exon 4.2	Forward	AAGTGCATCTCTCGGCAGTT
	Reverse	CCCCTTGGAACACGTAAAGA
LDLR- Exon 4.3	Forward	AGCTTCCAGTGCAACAGCTC
	Reverse	CATACCGCAGTTTTCCTCGT
LDLR- Exon 4.4	Forward	TGTTCCAAGGGGACAGTAGC
	Reverse	AAATCACTGCATGTCCCACA
LDLR- Exon 5	Forward	AGAAAATCAACACACTCTGTCCTG
	Reverse	GGAAAACCAGATGGCCAGCG
LDLR- Exon 6	Forward	TCCTCCTTCCTCTCTCTGGC
	Reverse	TCTGCAAGCCGCCTGCACCG
LDLR- Exon 7	Forward	GGCGAAGGGATGGGTAGGGG
	Reverse	GTTGCCATGTCAGGAAGCGC
LDLR- Exon 8	Forward	CATTGGGGAAGAGCCTCCCC
	Reverse	GCCTGCAAGGGGTGAGGCCG
LDLR- Exon 9	Forward	TCCATCGACGGGTCCCCTCTGACCC
	Reverse	AGCCCTCATCTCACCTGCGGGCCAA
LDLR- Exon 10.1	Forward	AGATGAGGGCTCCTGGTGCGATGCC
	Reverse	GCCCTTGGTATCCGCAACAGAGACA
LDLR- Exon 10.2	Forward	GATCCACAGCAACATCTACTGGACC
	Reverse	AGCCCTCAGCGTCGTGGATA
LDLR- Exon 11	Forward	TCCTCCCCCGCCCTCCAGCC
	Reverse	GCTGGGACGGCTGTCCTGCG
LDLR- Exon 12	Forward	GCACGTGACCTCTCCTTATCCACTT
	Reverse	CACCTAAGTGCTTCGATCTCGTACG
LDLR- Exon 13	Forward	GTCATCTTCCTTGCTGCCTG
	Reverse	GTTTCCACAAGGAGGTTTCAAGGTT
LDLR- Exon 14	Forward	GAATCTTCTGGTATAGCTGAT
	Reverse	GCAGAGAGAGGCTCAGGAGG
LDLR- Exon 15	Forward	GAAGGGCCTGCAGGCACGTGGCACT
	Reverse	GTGTGGTGGCGGGCCCAGTCTTT
LDLR- Exon 16	Forward	CCTTCCTTTAGACCTGGGCC
	Reverse	CATAGCGGGAGGCTGTGACC
LDLR- Exon 17	Forward	GGGTCTCTGGTCTCGGGCGC
	Reverse	GGCTCTGGCTTTCTAGAGAGGG
LDLR- Exon 18	Forward	GCCTGTTTCCTGAGTGCTGG
	Reverse	TCTCAGGAAGGGTTCTGGGC
APOB- Exon 26	Forward	TGTCAAGGGTTCGGTTCTTT
	Reverse	GGGTGGCTTTGCTTGTATGT
PCSK9- Exon7	Forward	CCCTCTCTTGGGCTCCTTTCT
	Reverse	AAAGGGGCTGTTAGCATCACG

The PCR reaction mix is comprised of 1.5ng of DNA, 0.05µl of each primer, 5µl AccuMelt HRM SuperMix (2x), and dH_2O to a total volume 10µl. The PCR conditions were as follows; started at 95^0C for 5 minutes, followed by 40-45 cycles of denaturing at 95^0C for 5 seconds, annealing at 60^0C for 10 seconds, and extending at 70^0C for 20 seconds. Negative and positive controls with defined mutations were used for each run to compare with samples and evaluate run performance. The amplified fragment was sequenced using Sanger sequencing. The DNA sequence was assessed manually.

Chapter 3 : Mutation detection and spectrum of familial hypercholesterolemia of UK patients

3.1. Introduction

FH is an autosomal dominant disorder resulting in elevated plasma LDL-C levels. The inherited form of heterozygous FH (HeFH) is the most common, with an estimated prevalence of 1/500 (42). This is even higher in some populations such as those of Denmark (43) and the Netherlands (201), the Afrikaners in South Africa (315) and French Canadians (44). The plasma LDL-C levels in HeFH are two-to-three fold higher than normal. Individuals with this form of FH are more likely to develop premature CHD in the second or third decade of their lives (42). If an individual inherits two defective genes they have homozygous FH (HoFH), which has an estimated prevalence of 1 in 1,000,000 in most populations. Individuals with homozygous FH have six-to-eight-fold higher levels of LDL-C plasma than normal and develop CHD in the first decade of life and often die of CHD before the age of 20 (42). FH can be diagnosed using well-established clinical criteria such as The Simon Broome Register Criteria (141) in the UK, and The Dutch Lipid Clinic Network (DLCNC) (144) in other European countries. FH patients are treated with statins, however the treatment outcome ranges from 30% to 60% reduction in LDL-C, because it depends on statin potency, type and dose, and the FH mutation type (20). Identifying and treating FH patients in childhood through to early adulthood is critical in order to prevent or reduce morbidity and mortality from premature CHD (316).

FH is considered a monogenic disorder, which is caused by a defect in one of three known genes *LDLR*, *APOB*, and *PCSK9*. Inheritance of FH as an autosomal recessive hypercholesterolemia (ARH) has rarely been reported (134); it occurs when an individual

inherits two variants in the *LDLRAP1* gene. In the UK, the vast majority (93%) of identified mutations are found in the *LDLR* gene, a further 5% in *APOB*, and about 2% in the *PCSK9* gene (45, 46). To date 1,700 different *LDLR* mutations have been reported (65), but only one common *APOB* mutation p.(Arg3527Gln) and one common PCSK9 mutation p.(Asp374Tyr) are found in the UK. The FH mutation spectrum varies between European countries. In Greece, for instance, six mutations have been reported in 60% of FH patients in that country (317), while in the Netherlands two mutations are common (318). On the other hand, the UK has a wide FH mutation spectrum, where over 200 common FH mutations have been reported (150). The majority of *LDLR* mutations (76%) are single nucleotide substitutions (missense, nonsenses, and splice site), 21% small insertions or deletions, and 5-6% large insertions or deletions.

The different types of FH mutations have led to the development of different FH screening methods such as ARMS, HRM, MLPA and commercial chips (LIPOchip, Elucigene FH20) (46, 146, 147, 149). However, these methods are labor-intensive and costly. Recently, next-generation sequencing (NGS) has been used to identify FH causing mutations (153, 157); the improved specificity and sensitivity of the method has led to the identification of new mutations in well-known FH genes and potential new FH genes (153, 155, 157, 319). In this study, a NGS-based method and Sanger sequencing were used to screen FH mutations in 69 subjects from the UK either with a clinical diagnosis of definite FH (DFH) or possible FH (PFH). Also, a family segregation analysis was performed in this study on a reported case of familial defective apo-B100 (FDB), which

provided an early diagnosis of disease in members of the family and led to appropriate treatment.

3.2. Results

3.1.1. Patient characteristics

Sixty-nine samples were collected from suspected FH subjects from several UK clinics. The subjects participating were clinically diagnosed with HeFH. Participants' clinical information and family history of FH and CHD were provided. Baseline characteristics are shown in Table 3-1. Forty-four subjects were children and young adults (≤20 years old), and 25 subjects were adults (>20 years old). The Simon Broome Register Criteria used for the diagnosis of FH in the UK are: (a) total cholesterol >7.5 mmol/l or LDL-C >4.9 mmol/l if >16 years and total cholesterol >6.7 mmol/l or LDL-C >4.0 mmol/l if <16 years; (b) tendon xanthoma (TX) in the patient or in a first or second degree relative; (c) family history of MI <60 years in a first degree relative or family history of MI <50 years in a second degree relative; and (d) family history of total cholesterol >7.5 mmol/l in first or second degree relative. Diagnosis is classified into two classes – DFH where both a+b must be present, or PFH where both a+c or a+d must be observed (141). In this study, 3 subjects were diagnosed with DFH, and 19 subjects were diagnosed with PFH. However, 49 subjects had suspected FH because they had high levels of TC and LDL-C, but no TX or a reported family history of MI (60% children). Indeed, pediatric subjects diagnosed with HeFH may initially present with lipid levels within the normal range, or borderline, but they are highly likely to develop hypercholesterolemia in adulthood (320). It is common that parents with hypercholesterolemia follow a low-fat family diet, which

leads to borderline lipid measurements in their children. In many of these cases a family history of MI may not be evident due to the parents' young age. Also, children with HeFH rarely present with TX in childhood and this also may be the case for their young parents. Therefore, elevated lipid levels in a family member with a positive history of premature CAD are suggestive of HeFH in children (317). Table 3-1 shows that the TC, LDL-C and TG plasma levels were significantly different between adult and children, which is expected.

Table 3-1: Familial hypercholesterolemia (FH) patient characteristics in UK study

	Adult	Children	p-value
Sample size	25 (36.2%)	44 (63.8%)	-
Age (years)	46.26 (±12.0)	11.43 (±3.5)	0.0001
Male (%)	12 (48%)	22 (50%)	0.017
TC (mmol/l)	9.8 (±2.77)	7.12 (±0.91)	0.0001
LDL-C (mmol/l)	7.66 (±2.0)	5.16 (±0.94)	0.0001
HDL-C (mmol/l)	1.27 (±0.29)	1.4 (±0.33)	0.120
TG (mmol/l)	2.27 (±1.17)	1.08 (±0.44)	0.0001
Tendon xanthomata (%)	2 (8.0. %)	1 (2.2%)	0.24
Family history of MI (%)	7 (28%)	11 (25%)	0.79

Mean and standard deviation, where appropriate, are shown. P-value was determined using ANOVA test for numeric variables and chi-squared test for categorical variables.

3.1.1. Identified mutations in this study

An FH-causing variant was found in 29 out of 69 patients (42.1%) examined, the detection rate in the children was not different (21/44 = 48%) compared to the adult samples (8/25 = 32%), p= 0.41. As shown in Table 3-2, there was no significant difference in cholesterol plasma levels between mutation positive and mutation negative groups (TC: p= 0.23, and LDL-C: p= 0.16) in adult groups. However, in the children's group cholesterol plasma levels were significantly different between mutation positive and mutation negative groups (TC: p= 0.002, and LDL-C: p= 0.001). Seven children with FH mutation had a family history of MI.

Table 3-2: Familial hypercholesterolemia (FH) patient characteristics: mutation positive and negative groups in the UK study

	Adult (25)			**Children (44)**		
Variable	Mutation positive	Mutation negative	p-value	Mutation positive	Mutation negative	p-value
Sample size	8	17	-	21	23	-
TC (mmol/l)	10.1 (±2.7)	8.5 (±2.8)	0.23	7.6 (±0.99)	6.7 (±0.6)	0.002
LDL-C (mmol/l)	8.7 (±2.7)	7.2 (±1.4)	0.16	5.7 (±1.0)	4.7 (±0.6)	0.001
HDL-C (mmol/l)	1.1 (±0.2)	1.4 (±0.3)	0.89	1.4 (±0.3)	1.5 (±0.3)	0.55
TG (mmol/l)	2.9 (±1.5)	1.9 (±0.9)	0.12	1.0 (±0.3)	1.1 (±0.5)	0.63
Tendon xanthomata	2	0	0.06	1	0	0.3
Family history of MI	3	0	0.02	7	0	0.001

Mean and standard deviation, where appropriate, are shown. P-value was determined using ANOVA test for numeric variables and chi-squared test for categorical variables.

3.1.2. FH mutation positive in UK study

Mutations were found in all three FH causing genes (Table 3-3); 18 different *LDLR* mutations (3 novel) were found in 23 patients and accounted for 79% of all observed mutations, 2 different mutations were reported in the *APOB* gene in 5 subjects (3 probands) (18%), one novel mutation in *PSCK9* gene (3%), and no mutations was detected in *LDLRAP1*.

3.1.2.1. *LDLR* mutations

Twenty-three patients clinically diagnosed with FH carried *LDLR* causing variants; the 19 different mutations identified have been reported in the literature, and three were novel (Figure 3-1). The identified mutations are located in 10 different exons of *LDLR* (2, 3, 4, 8, 9, 10, 11, 12, 13, and 14) (Table 3-3). Of the identified mutations that have been reported in the UCL-FH mutation database, three mutations were most common: the nonsense mutations in exon 3 [c.226 G>A, p.(Cys89*)] was detected in two patients, the missense mutation in exon 4 [c.136 T>G, p.(Cys197Tyr)] was identified in two patients, another a common missense mutation was identified in exon 14 [c.2054 C>T, p.(Pro685Leu)] in three patients.

Other reported missense mutations were identified in exon 2 [c.136 T>G, p.(Cys46Gly)], in exon 4 [c.434 T>A, p.(Val145Glu); c.463 T>A, p.(Cys155Ser)], in exon 8 [c.1073 G>A, p.(Cys358Tyr), c.1133 A>C, p.(Gln378Pro)], in exon 9 [c.1246 C>T, p.(Arg416Trp)], in exon 10 [c.1436 T>C, p.(Leu479Pro)], in exon 12 [c.1745 T>C,

p.(Leu582Pro)], in exon 13 [c.1860 G>T , p.(Trp620Cys)], and in exon 14 [c.2042 G>C , p.(Cys681Ser)]. Furthermore, two previously reported nonsense mutations were identified in exon 8 [c.1150 C>T, p.(Gln384*)] and in exon 11 [c.1686 G>A, p.(Trp562*)].

Figure 3-1: Diagram of *LDLR* gene showing the mutations identified in this study.

Of 19 mutations identified in UK subjects in this study, three are novel mutations (indicated in red). Exons are shown as vertical boxes and introns as the lines connecting them.

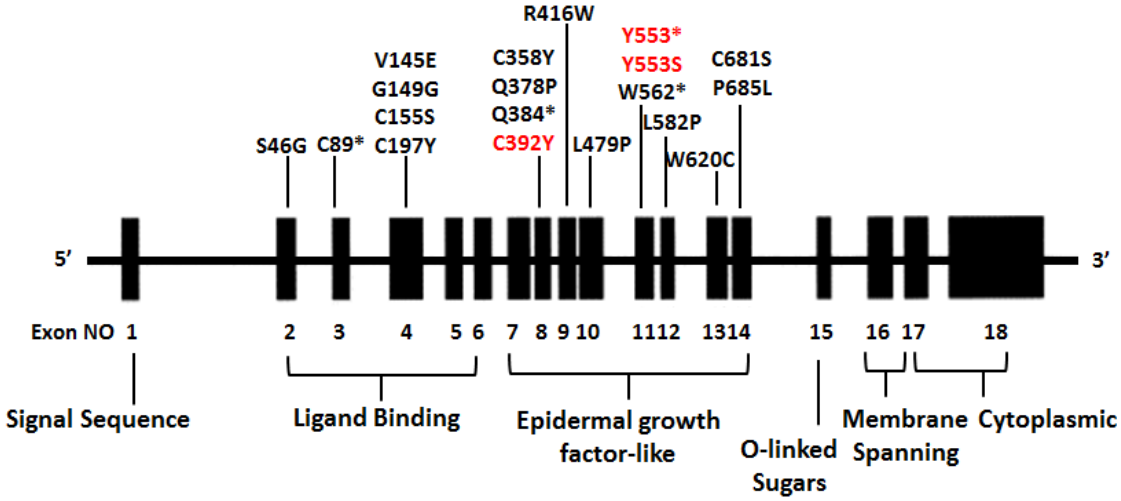


Table 3-3: Identified *LDLR* mutations in UK familial hypercholesterolemia (FH) patients

Gene	Location	Base pair change	Amino acid change	PolyPhen	SIFT	Mutation Assessor	Reported FH caussing	Number of carriers (number of probands)
LDLR	Exon 10	19:c.1436 T>C	p.(Leu479Pro)	D	D	D	Yes	1 (1)
	Exon 11	19:c.1659 C>G	**p.(Tyr553*)**	N/A	N/A	N/A	No	1 (1)
	Exon 11	19:c.1686 G>A	p.(Trp562*)	N/A	N/A	N/A	Yes	1 (1)
	Exon 11	19.c.1658 A>C	**p.(Tyr553Ser)**	D	D	P	No	1 (1)
	Exon 12	19:c.1745 T>C	p.(Leu582Pro)	D	D	D	Yes	1 (1)
	Exon 13	19:c.1860 G>T	p.(Trp620Cys)	D	D	D	Yes	1 (1)
	Exon 14	2:c.2042 G>C	p.(Cys681Ser)	D	D	D	Yes	1 (1)
	Exon 14	19:c.2054 C>T	p.(Pro685Leu)	D	D	D	Yes	3 (3)
	Exon 2	19:c.136 T>G	p.(Cys46Gly)	D	D	D	Yes	1 (1)
	Exon 3	19:c.226 G>A	p.(Cys89*)	N/A	N/A	N/A	Yes	2 (2)
	Exon 4	19:c.590 G>A	p.(Cys197Tyr)	D	D	D	Yes	2 (2)
	Exon 4	19:c.434 T>A	p.(Val145Glu)	N/A	N/A	N/A	Yes	1 (1)
	Exon 4	19:c.463 T>A	p.(Cys155Ser)	N/A	N/A	N/A	Yes	1 (1)
	Exon 8	19:c.1150 C>T	p.(Gln384*)	N/A	N/A	N/A	Yes	1 (1)
	Exon 8	19:c.1073 G>A	p.(Cys358Tyr)	D	D	D	Yes	1 (1)
	Exon 8	19:c.1133 A>C	p.(Gln378Pro)	D	P	P	Yes	1 (1)
	Exon 8	19:c.1175 G>A	p.(Cys392Tyr)	D	D	D	No	1 (1)
	Exon 9	19:c.1246 C>T	p.(Arg416Trp)	D	D	D	Yes	1 (1)
APOB	Exon 26	2:c.10580 G>A	p.(Arg3527Gln)	D	D	D	Yes	2 (2)
	Exon 26	2:c.10579 C>T	p.(Arg3527Trp)	D	D	D	Yes	3(1)
PCSK9	Exon 7	1:c.1070 C>T	p.(Arg357Cys)	P	P	P	No	1 (1)

D: Deleterious (PolyPhen2, SIFT), portion function (Mutation Assessor), P: Possibly damaging, B: benign, N/A: not applicable

3.1.2.2. *APOB* mutations

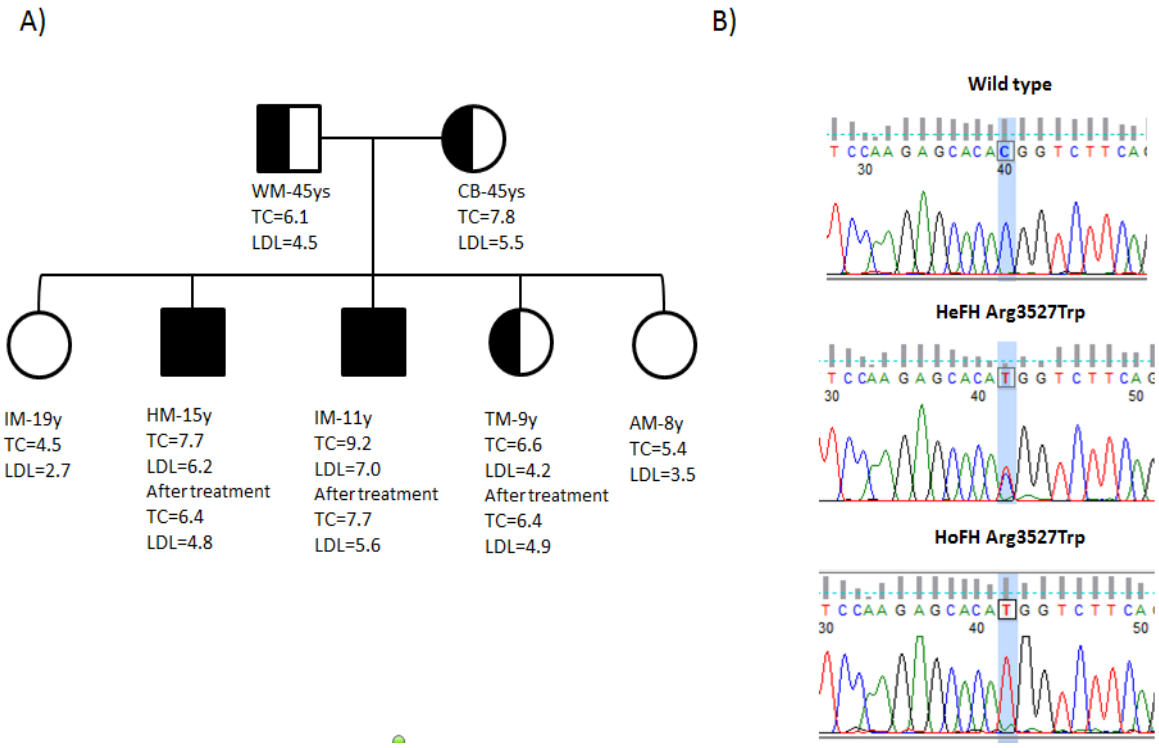
Five patients (3 probands) carried the two different *APOB* missense mutations at position p.3527, two patients carried p.(Arg3527Gln) and three patients carried the p.(Arg3527Trp) variant.

Family segregation analysis of the *APOB* Arg3527Trp affected subjects'

The *APOB* p.(Arg3527Trp) was identified in a proband patient (TM) who had suspected FH. A 9 year-old South Asian girl (Pakistan) with a family history of MI had a TC = 6.6 mmol/l, LDL-C = 4.2 mmol/l is heterozygotes for the p.(Arg3527Trp) mutation. To analyze mutation heritability, family members were genetically screened. Figure 3-2 shows family segregation analysis; there is a consanguineous marriage of the parents, where the parents' mothers are sisters (cholesterol levels unknown) and they share the same grandparents. The parents are heterozygotes for the mutation, and the mother has a raised TC level (7.5 mmol/l) and LDL-C level (5.5 mmol/l). Of their five children, not only is the proband affected but also two boys, aged 11 and 15 years are homozygous for the mutation [TC = 9.2 mmol/l, and LDL-C = 7.0 mmol/l; TC = 7.7 mmol/l, and LDL-C = 6.2 mmol/ respectively], while the other two girls aged 8 and 19 years are mutation negative and have normal cholesterol levels.

Figure 3-2: Family segregation of the c.10579 C>T, p.(Arg3527Trp) APOB mutation.

(A) A family pedigree of the proband patient (TM) with the mutation including age (years), TC level (mmol/l) and LDL-C level (mmol/l). Six members of the family (WM, CB, IM, HM, IM and AM) were screened and sequenced for the mutation. Four members of the family were found to carry the variant. (B) APOB exon 26 sequencing for wild type and base change in patients (appropriate base highlighted).

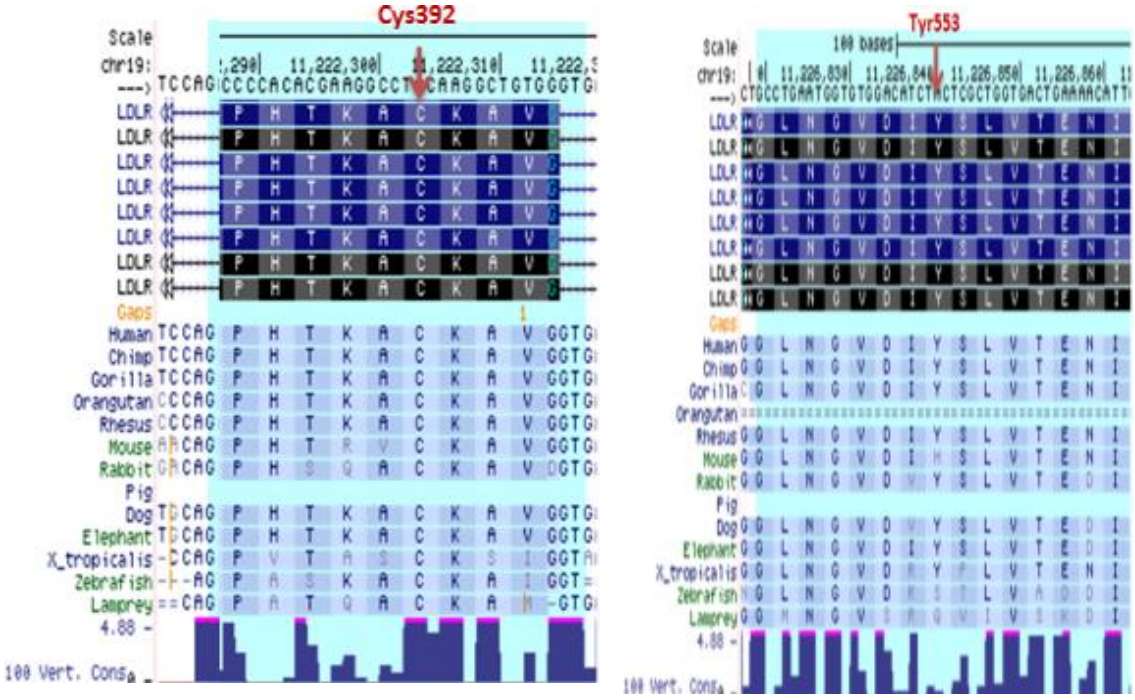


Full-filled shape= Homozygote for the *APOB* p.(Arg3527Trp) mutation (HoFH)
Half-filled shape= Heterozygote for the *APOB p.*(Arg3527Trp) mutation (HeFH)
No filled shape = no mutation

3.1.2.3. Novel mutations in the UK-FH study

Three novel mutations were found in the *LDLR* gene and they were predicted to be pathogenic, however, family members of these patients were not available for segregation analysis. Missense mutations were found in exon 8 [c.1175 G>A, p.(Cys392Tyr) and in exon 11 [c.1658 A>C, p.(Tyr553Ser)] and one nonsense mutations was found in exon 11 [c.1659 C>G, p.(Tyr553*)]. The three mutations are highly conserved a cross species (Figure 3-3) and are encoded in the epidermal growth factor-like domain (EGF-like domain), where a variant in this domain blocks receptor dissociation in endocytosis and receptor recycle to the cell surface machinery, both of which are highly likely to cause a disease phenotype. A 35 year-old White Caucasian woman was found to be heterozygous for Cys392Tyr (Figure 3-4 (A)); she had raised levels of TC=9.1 mmol/l, and LDL-C= 6.5 mmol/l, but no TX or MI family history. Tyr553Ser (Figure 3-4 (B)) was identified in a 38 year-old White Caucasian man, who had raised levels of TC=8.5 mmol/l, and LDL-C= 7.2 mmol/l, but no TX or MI family history. Peptide truncated mutation Tyr553* ((Figure 3-4 (C)) was identified in a 12 year-old White Caucasian girl, but unfortunately no clinical data were available.

Figure 3-3: Sequence conservation of the two *LDLR* novel mutations in the UK study across-species

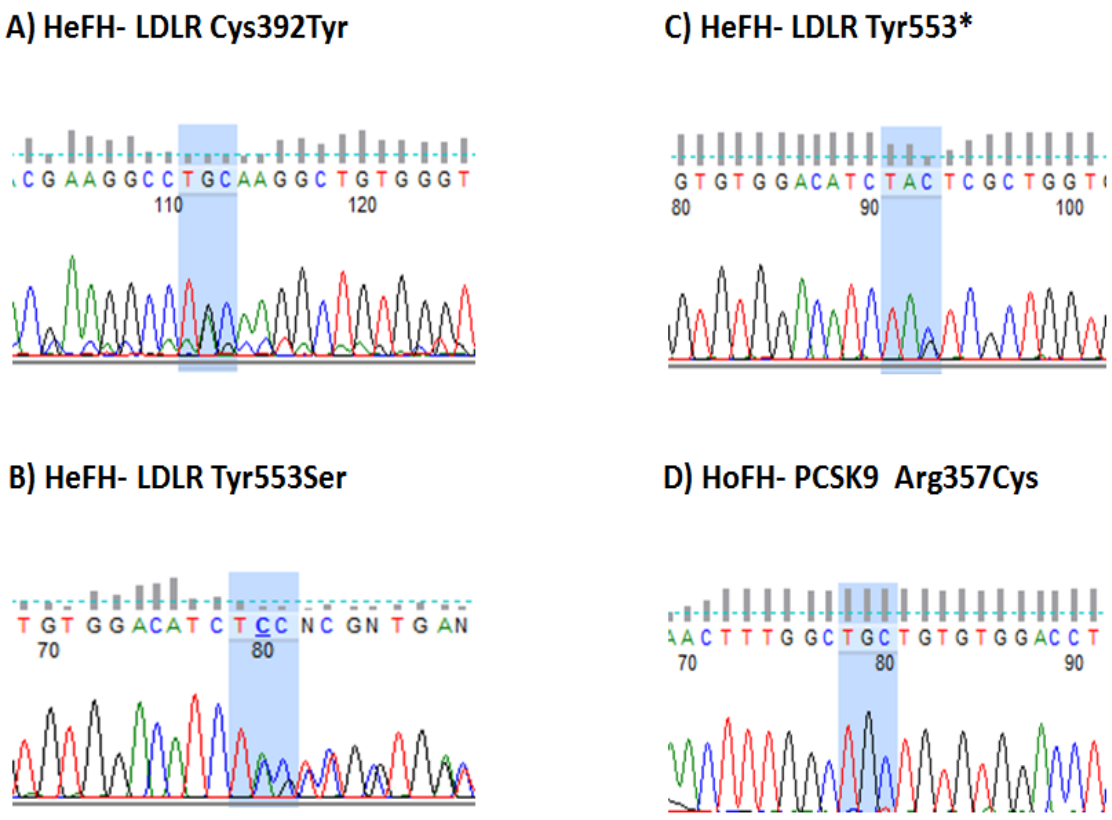


Another novel variant p. Arg357Cys was identified in the *PCSK9* gene in a 38 year-old white Caucasian man who was homozygous for the variant (Figure 3-3 (D)); the Arg 357 is conserved, and substitution at this position is predicted to be pathogenic. He had raised levels of TC=8 mmol/l and LDL-C= 10.1 mmol/l and a family history of MI in a first degree relative (father). The variant is within the catalytic domain of PCSK9 (which is the domain that interacts with LDL-R to mediate receptor degradation). Also, it is located close to the oxyanion hole (Asn317), which has functional importance in catalysis (Russel et al., 1987; Carter et al., 1988). Replacement of an arginine by cysteine (p.(Arg357Cys)) may disturb hydrogen bond formation between residues (Betzel et al., 2001). In addition, arginine 357 is located next to cytosine 358 which is predicted to be

involved in a disulfide bond in the rat Narc-1 protein (Naureckiene et al., 2003). Interestingly, another variant has been identified in a French FH patient at the same codon (p.357) resulting in a substitution of histidine (H) in place of arginine. The patient has a strong family history of hypercholesterolemia through five generations in both sides of his family and history of CVD in the paternal side (127).

Figure 3-4: Novel mutations identified in the UK familial hypercholesterolemia study

The figures show the DNA sequence of novel identified mutations. A) LDLR mutation in exon 8 c.1175 G>A, p.(Cys392Tyr). B) LDLR mutation in exon 11 c.1658 A>C, p.(Tyr553Ser). C) LDLR mutation in exon 11 [c.1659 C>G, p.(Tyr553*). D) PCSK9 mutation in exon 7 c.1068 C>T p.(Arg357Cys).



3.3. Discussion

This study has identified a number of previously reported and novel mutations among patients with FH attending various UK clinics. Amongst 69 patients clinically diagnosed with HeFH, mutations were found in 29 patients (42%). Different mutations were identified using a target NGS method and were confirmed by Sanger sequencing: 18 mutations in *LDLR* (3 novel mutations: p.(Cys392Tyr), p.(Tyr553Ser), and p.(Tyr553*)), one novel mutation in *PCSK9* p.(Arg357Cys), and two mutations in *APOB* (p.(Arg3527Gln) and p.(Arg3527Trp)).

The overall mutation detection rate in this study is 42% which is higher than that reported in previous UK studies: 35.53% (145), and 20% (148). This may be due to an increase in the sensitivity and specificity of the NGS mutation detection method. The Taylor (2010b) and Futema (2013) groups used ARMS, MLPA and HRM screening methods, which may include false negative results due to human error in data and sample handling (157). Also the ARMS kit was designed to detect only the 18 most common *LDLR* mutations in the UK (44% of UK mutations), *APOB* p.(Arg3527Gln) and *PCSK9* p.(Asp374Tyr) (145), while the MLPA kit was developed to detect large deletions, and duplications which account for 10% of UK mutations (147). The HRM method has low coverage in some regions which may affect method sensitivity (146). In contrast, the NGS method has been used recently as a diagnostic tool for FH in studies of different FH populations (151-154), and has shown 100% specificity and 98-100% sensitivity (155). All of these factors in addition to time efficiency and cost-effectiveness, have led to identification of novel mutations either in the three related FH genes (155, 157) or in new

genes that are suggested to be associated with polygenic FH (151, 157). It is also possible that the detection rate is higher because a large proportion of the subjects here are children with a clinical diagnosis of HeFH and as such it would be expected that fewer have a polygenic etiology than a cohort of adults with a clinical diagnosis of HeFH

In this study, the mutation spectrum was similar to the results of previous UK studies overall, where 80% of identified mutations were reported in *LDLR*, 17% in *APOB*, and 3% in *PCSK9*. The *LDLR* identified mutations are located in 10 out of 18 exons of *LDLR*; there is a correlation between exons and the functional domain of the LDL-R protein. The identified mutations are encoded in two domains: the ligand binding domain and the EGF-like domain. The reported mutations could be classified into three classes of *LDLR* according to the 5 proposed *LDLR* mutation classes (51). Class 1 mutations are usually caused by nonsense mutations where the variants create stop codons and hence truncated proteins which prevent LDL-R protein synthesis. Class 3 mutations result in defective LDL binding to the receptor. Class 5 mutations occur due to a mutation in the EGF-like domain, where a mutation in this domain leads to blocking receptor dissociation in endocytosis and machinery for receptor recycling to the cell surface. Both *LDLR* mutation type and heterogeneity have an influence on FH phenotype. Studies have shown that the *LDLR* mutation types have an influence on the lipid profile and the response to lipid-lowering therapy (321, 322). Also, pharmacogenomic studies have reported that FH patients harboring a 'null mutation' (including classes 1 and 2, and large rearrangements or deletions leading to a frameshift and a premature stop coding) are at higher risk of premature CHD than patients with a defective mutation (322). This raises the importance

of identifying FH mutations and their type in order to establish lipid-lowering treatments at an early stage of the disease.

In addition to 15 *LDLR* identified mutations that have been previously reported and linked to FH, there were three novel mutations all reported in the EGF-like domain. The first *LDLR* novel mutation reported was the substitution of cysteine 392 to tyrosine. Indeed, the 392 position looks like a hotspot for a FH causing mutations. It has been reported previously as a site for defective and null mutations: p.(Cys392Arg) (323) and p.(Cys392*) respectively (324-327). The other two novel mutations were reported at position 553; one defective mutation p.(Tyr553Ser) and one null mutation p.(Tyr553*). The literature has reported that the substitution of tyrosine 533 to cysteine is an FH causing mutation (83). The three novel mutations are encoded positions that have been reported for FH causing mutations, and are predicted to be pathogenic, which suggests that the new identified mutations are highly likely to be functional.

Also, a novel mutation in *PCSK9* p.(Arg357Cys) was identified. The identified mutation is encoded in the catalytic domain, which is the part of PCSK9 the binds to LDL-R and mediates receptor degradation. Thus, mutation in this domain is highly likely to be functional. Indeed, patient clinical data, *in silico* functional prediction tools and mutation screening together have demonstrated that the identified mutation is highly likely to be an FH causing mutation. Usually, family segregation analysis would provide a definitive conclusion; however, family members were not available. Interestingly, Allard et al.

(2005) reported an FH case with p.(Arg357His) (high TC and LDL-C levels, and family history of hyperlipidemia and CHD, but no data was available for variant heterogeneity). They suggested that substitution of arginine 357 to histidine may disturb the hydrogen bond formation between residues Asp, His and Ser of the catalytic triad, and consequently the structure of PCSK9 (127).

Five patients (3 probands) were found with familial defective apoB-100 (FDB), which accounted for 17% of the identified mutations in this sample. Two unrelated white Caucasian children carried the most common *APOB* mutation p.(Arg3527Gln). The substitution of arginine to glutamine at position 3527 reduces the affinity between apoB-100 and LDL-R, where the highly conserved receptor binding-site is stabilized by the interaction of Arg3527 with Trp4396, thus replacement of arginine 3527 by a glutamine impairs receptor recognition, affects hepatic LDL-C clearance and accumulation of LDL-C particles in the circulation (73, 328). The affected probands are heterozygote for the variant which reflects the FH phenotype (TC > 7mmol/l, LDL-C > 5mmol/l, family history of MI in a first degree relative and no TX reported).

The substitution of arginine 3527 with tryptophan p.(Arg3527Trp) was identified in a South Asian proband and another four members of the family. The novel finding here was that two patients were homozygous for the *APOB* variant p.(Arg3527Trp) and have milder phenotype than those homozygous for *LDLR*. There are three possible explanations of the lower plasma lipid levels found in FDB patients. First, studies in

FDB homozygotes showed that the binding affinity of LDL to receptors in homozygous FDB was 10-20 % of normal affinity (329), which reduce clearance of LDL via the receptor. Secondly, lower plasma lipoprotein levels in FDB can partially be caused by an increased uptake of VLDL remnants via LDL-Rs (85). Thirdly, the increased uptake of apoE containing particles from the plasma results in a decrease of apoB-100 production rate as well as of LDL-C particles (85). It has been reported that the p.(Arg3527Trp) mutation has arisen independently in Asians (75, 80-82), while the p.(Arg3527Gln) mutation is inherited from a common ancestor in several Western populations (74-79, 330).

Fifty-eight percent of subjects examined in this study were FH-mutation negative. It has been demonstrated that the elevated LDL-C levels in mutation negative subjects with a clinical diagnosis of FH can be explained by the "polygenic" contribution of 12 common LDL-C raising variants in genes that were identified by GWAS (164). Although 6 of these LDL-C associated variants were included in the NGS amplicon design, they were excluded in data analyses due to the filtration criteria, where the common variants with MAF > 0.05% are excluded. Thus, it is possible that the elevated levels of LDL-C seen in the FH-mutation negative patients have a polygenic and not a monogenic explanation. For future work, it is worth examining whether FH is caused by an accumulation of common small-effect LDL-C raising variants in the mutation negative group. Such examination may help to explain the FH phenotype seen in the FH-mutation negative group.

3.4. Limitations

There were limitations in this study. One was the small sample size and another, the unavailability of lipid measurements for 18 (26%) of the 69 participants, due to loss of participant communication after they had submitted FH mutation screening applications. The samples were sent for screening because the patients had met the FH criteria and it would not have been ethical to stop screening. For the same reason it was impossible to perform segregation analysis for all patients with novel variants. Another limitation was that polygenic FH was not examined in the mutation negative group due to time constraints.

3.5. Conclusion

NGS is a powerful technology that has the potential to address clinically relevant phenotypes related to genetic conditions of FH. FH monogenic mutations were identified in 42% of FH clinically diagnosed subjects. The mutations were identified in autosomal dominant hypercholesterolemia (ADH) genes including *LDLR*, *APOB*, and *PCSK9;* among 21 mutations identified in this study, four mutations were novel. It is recommended that FH-mutation negative subjects with a clinical diagnosis of FH are examined for 12 common "polygenic" LDL-C raising variants.

Chapter 4 : Mutation spectrum of familial hypercholesterolemia in Iranian patients

4.1. Introduction

CHD is a leading cause of mortality, morbidity, and disability in the Iranian population, which accounts for nearly 50% of total deaths per year, and elevated LDL-C levels is the major risk factor, being present in ~45% of subjects (331). The prevalence of hypercholesterolemia (defined as having TC>90th percentile, 170mg/dl) in Iranian children and adolescents varies between 3% and 48% across the country (332). The causes of this are likely to be due to differences in both environmental factors and genetic background, for example due to the difference in frequency of common cholesterol-raising variants in different regions. The prevalence of the monogenic disorder FH may also contribute to this, and the frequency and spectrum of FH-causing mutations in the Iranian population is essentially unknown, with only a handful of studies reporting FH patient characteristics in Iran. Cases of homozygote FH (HoFH) have been reported in children with severe septum symptoms (TC≥ 440 mg/dl, LDL-≥ 390 mg/dl, HDL≥ 76 mg/dl) corneal arcus in both eyes and tendon xanthomas in the dorsum of both hands, both elbows, both knees, and Achilles tendons, and with early CHD (coronary artery stenosis and treated with coronary artery bypass graft surgery) (333, 334). Genetic studies by the Fard-Esfahani group in 2005, who examined 30 patients with clinical possible FH, identified one novel mutation in exon 4 of the *LDLR* (p.(Gly445Cys0 G>T) using the PCR-single-strand conformation polymorphism (PCR-SSCP) method (335), but no carriers of the common mutation in the *APOB* gene (exon 26) were found (336). In another study in 2011, 30 Iranian subjects who had been clinically diagnosed with FH, were screened for FH mutations using PCR-Restriction Fragment Length Polymorphism (PCR-RFLP) and PCR-SSCP methods. No mutations were reported either in a promoter or coding region of *LDLR* or in exon 26 of *APOB*, but four common *LDLR* polymorphisms were reported in 17% of Iranians with possible FH, sensitivity of the method demonstrated by the identification of common *LDLR* polymorphisms (337, 338). Targeted

PCR methods are designed to detect a limited number of genetic mutations, thus they cannot detect all FH mutations; in addition, specificity and sensitivity issues are associated with the methods. Recently, NGS methods have been used as diagnostic tools for FH (151-154), and have shown 100% specificity and 98-100% sensitivity (155).

In 2015, clinical practice guidelines for the diagnosis and treatment of FH in the Middle East were published. These guidelines noted some factors that are not applicable in Western countries such as the high prevalence of consanguineous marriage, limited treatment accessibility and the difficulties of cascade testing (339), which make FH management and CHD prevention difficult. Historically, HeFH prevalence in most Western populations is thought to be 1:500 (42). By contrast, HoFH has a much lower frequency of 1:1,000,000 in most Western countries (42) and 1:300,000 in the Netherlands (201). HoFH may have a higher prevalence in Iran than in Western countries due to consanguineous marriage, which is more common in Iranian culture (35-54%) (340). To be given a clinical diagnosis of HoFH patients should have LDL-C >13 mmol/L (500 mg/dL) if untreated or LDL-C >8 mmol/L (300 mg/dL) if treated, and the presence of tendon xanthomas (TX) before the age of 10 years, or both the patients' parents should have been diagnosed with HeFH (200). However, the range of LDL-C levels reported in patients with HoFH is broad and can overlap with ranges found in other types of hypercholesterolemia (200, 339). In Iran, it is essential to determine the type of FH patient in order to reduce morbidity and mortality due to CHD. Also, identifying the mutation spectrum and frequency of this region with help to improve disease management in the country.

As mentioned earlier, some countries in the Middle East have limited access to the new technology that is now used in FH mutation screening. In this study, my lab collaborated with Dr. Amirhossein Sahebkar from Iran, who provided samples from patients clinically diagnosed with HoFH, which were collected across several pediatric clinics in Iran. This study aimed to assess the FH mutation spectrum and frequency in Iran using the NGS-based screening method followed by Sanger sequencing to confirm identified mutations.

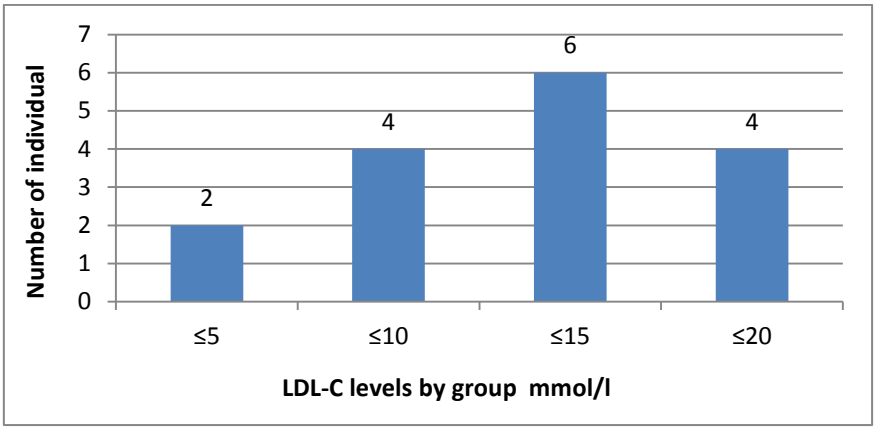
4.2.Results

Fifty-seven samples were received from 16 unrelated families, and 16 children probands were selected to undergo mutation screening; of these, 14 probands were diagnosed with clinical HoFH, with an LDL-C >13 mmol/L (untreated) or LDL-C >8 mmol/L (treated), TX, and a positive family history of hyperlipidemia or MI. Two probands (FH9-P1 and FH10-P2) were suspected of having HeFH, although their cholesterol levels (LDL-C = 3.3 and 4.9 mmol/l, respectively) did not exceed the FH diagnostic cut-off, but they reported a strong family history of hyperlipidemia and/or MI. Thus, these probands also underwent analysis for *LDLR*, *APOB* and *PCSK9* mutations because if the child was from a family with PFH, there is a high probability that he/she has FH. For this reason, samples were collected from the proband and any first degree relatives. Baseline characteristics of the 16 selected cases are shown in Table 4.1 and Figure 4-1shows LDL-C level distribution in Iranian FH probends.

Table 4-1: Baseline characteristics (mean ± SD) used in the Iranian FH study

Variable	Probands
Sample size	16
Age (in years)	7.1 (±4.8)
Total Cholesterol (mmol/l)	16.3 (±5.7)
LDL-Cholesterol (mmol/l)	12.0 (±4.6)
Tendon xanthomata	14 (87.5%)
Family history of CHD	6.0 (37.5%)
Family history of hyperlipidemia	16 (100%)
Taking lipid-lowering medication	16 (100%)

Figure 4-1: The distribution LDL-C levels in Iranian FH probands



4.1.1. Identified mutations in this study

Out of the 16 patients examined, a FH-causing variant was found in 9 patients (56%), while no mutation was detected in 7 individuals (44%), 5 clinically diagnosed with FH and 2 suspected of FH. All identified mutations were found in *LDLR*, and no mutation was found in *APOB* or *PCSK9*. As shown in Table 4-2, although TC and LDL-C levels were higher in the mutation positive group compared to the mutation negative groups, in this small sample these differences were not statistically significant (TC p= 0.10 and LDL-C p= 0.09).

Table 4-2: Familial hypercholesterolemia (FH) patient characteristics: mutation positive and negative groups in the Iran study

Variable	Mutation positive	Mutation negative	p-value
Sample size	9	7	
Age (in years)	7.8 (±5.6)	6.3 (±3.8)	0.56
Total Cholesterol (mmol/l)	18.4 (±4.6)	13.7 (±6.7)	0.10
LDL-Cholesterol (mmol/l)	13.6 (±3.8)	9.6 (±5.0)	0.09
Tendon xanthomata	9 (100%)	5(71.4%)	0.16
Family history of CHD	5 (55.5%)	1 (14.3%)	0.12
Family history of hyperlipidemia	9 (100%)	7 (100%)	-
Taking lipid-lowering medication	9 (100%)	7 (100%)	-

Mean and standard deviation (SD), where appropriate, are shown. P-value was determined using ANOVA for numeric variables and chi-squared test for categorical variables.

Nine patients clinically diagnosed with HoFH carried *LDLR* causing variants; five different mutations were identified that have been previously reported in the literature, and two were novel (Table 4.3, Figure 4.2). The identified mutations are located in six different exons of *LDLR* (4, 10, 11, 12, 14, and 17). Of the identified mutations reported in the UCL-FH mutation database, two were nonsense mutations [c.389C>G, p.(Ser130*) in exon 4, and c.1599G>A, p. (Trp533*) in exon 11], and two were missense mutations [c.1729T>C p.(Trp577Arg) and c. 2001T>G p.(Cys667Trp) in exon 12 and 14, respectively]. The fifth mutation was a frame-shift mutation in exon 17, c.2416_2417insG p.(Val806GlyfsTer11), where an insertion of guanine leads to a change in the amino acid from valine to glycine, which shifted the reading frame and created a stop codon downstream. Of the five identified mutations, two were common in this study: the missense mutation p.(Trp577Arg) was detected in three patients and the frame-shift mutation p.(Val806GlyfsTer11) was identified in two patients.

The missense mutation's functionality was assessed by *in-silico* mutation prediction tools, including PolyPhen2, SIFT, and Mutation Taster, and were predicted to be functional (Table 4.3). As the nonsense mutations and insertion frame-shift mutation led to peptide truncation, they were predicted to be functional.

Also, Lipoprotein(a) (Lp(a)) levels was screened in two FH suspected probands (FH mutation negative), and found that they had high levels of (Lp(a) [Lp(a) > 100 mg/dL but with LDL-C levels within the normal range (~3.3 mmol/l) when undergoing lipid-lowering drug therapy]

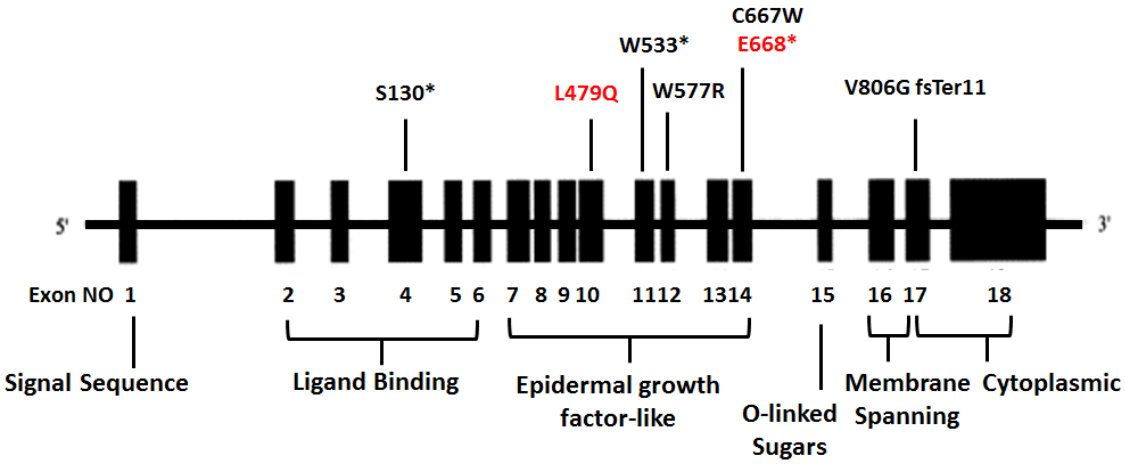
Table 4-3: Identified *LDLR* mutation in Iranian patients with familial hypercholesterolemia (FH)

ID	TC mmol/l	LDL-C mmol/l	Exon	Base pair change	Amino acid change	PolyPhen	SIFT	Mutation Assessor	Reported FH causing	Country report
FH15-P	18.6	13.0	Exon 4	c.389C>G	p.(Ser130*)	N/A	N/A	N/A	(341)	Denmark
FH14-P	16.5	15.0	Exon 10	c.1436 T>A	p. (Leu479Gln)	D	D	D	Novel	Iran
FH8-P	21.7	19.4	Exon 11	c.1599G>A	p.(Trp533*)	D	D	D	(342)	Subject origin unknown
FH1-P	11.4	10.0	Exon 12	c.1729T>C	p.(Trp577Arg)	D	D	D	(343, 344)	Turkey
FH2-P	18.6	14.2								
FH7-P	22.5	17.5								
FH17-P	15.3	13.6	Exon 14	c.2001T>G	p.(Cys667Trp)	D	D	D	(345)	France
				c.2002G>T	p.(Glu668*)	N/A	N/A	N/A	Novel	Iran
FH3-P	21.5	16.1	Exon 17	c.2416_2417insG	p.(Val806Glyfs Ter11)	N/A	N/A	N/A	(346), (347), (348), (325) and (349).	US South-Sweden Czech Netherlands Japan
FH5-P	15.5	7.8								

D: Deleterious (PolyPhen2, SIFT), portion function (Mutation Assessor), P: Possibly damaging, B: benign, N/A: not applicable

Figure 4-2: Diagram of *LDLR* gene showing the mutations identified in the Iranian study.

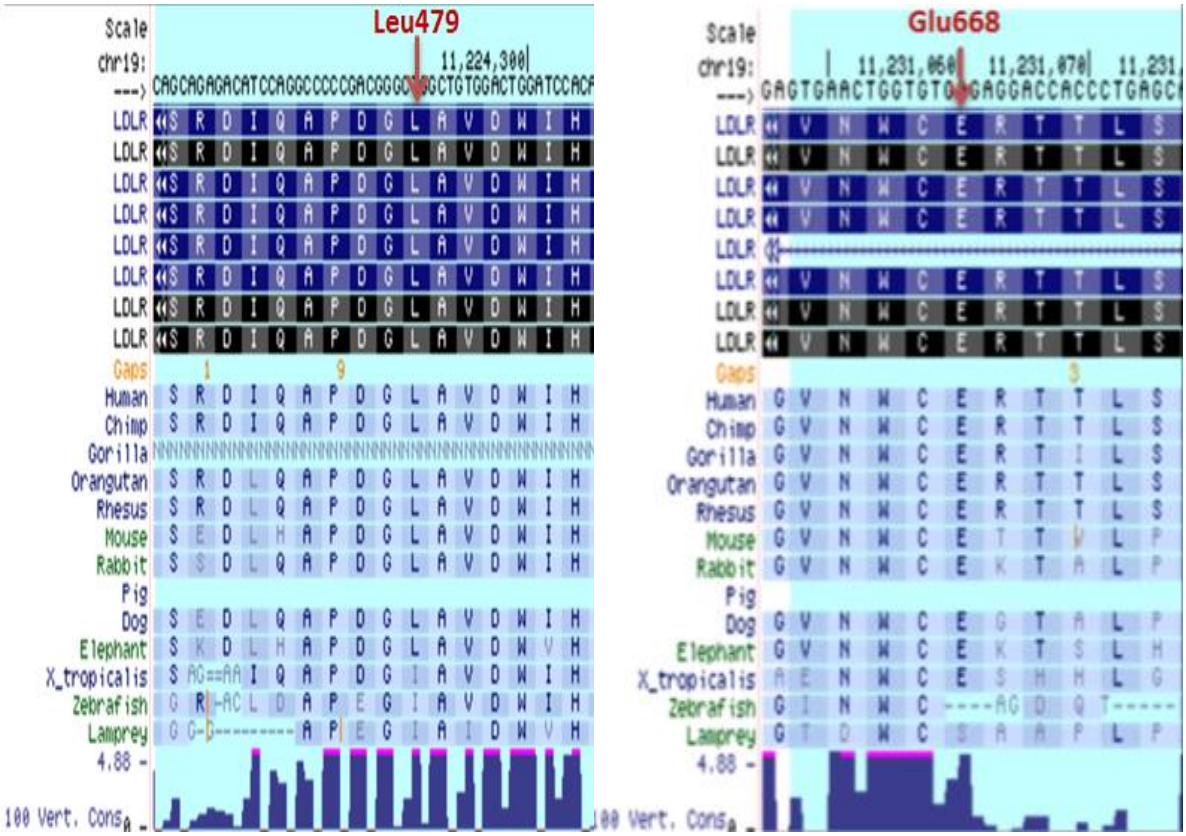
Seven mutations were identified in this study: two novel mutations are indicated in red. Exons are shown as vertical boxes and introns as the lines connecting them.



4.1.1. Novel mutations

Two novel mutations in *LDLR* were found in two probands with HoFH, which were not reported in the UCL-FH mutation database, one in exon 10 c.1436T>A p.(Leu479Gln) and another in exon 14 c.2002G>T p.(Glu668*). Both mutations are highly conserved a cross species (Figure 4-3) and are encoded in the epidermal growth factor-like domain (EGF-like domain), where a variant in this domain blocks receptor dissociation in endocytosis and the receptor recycle to the cell surface machinery, both of which are highly likely to cause a disease phenotype.

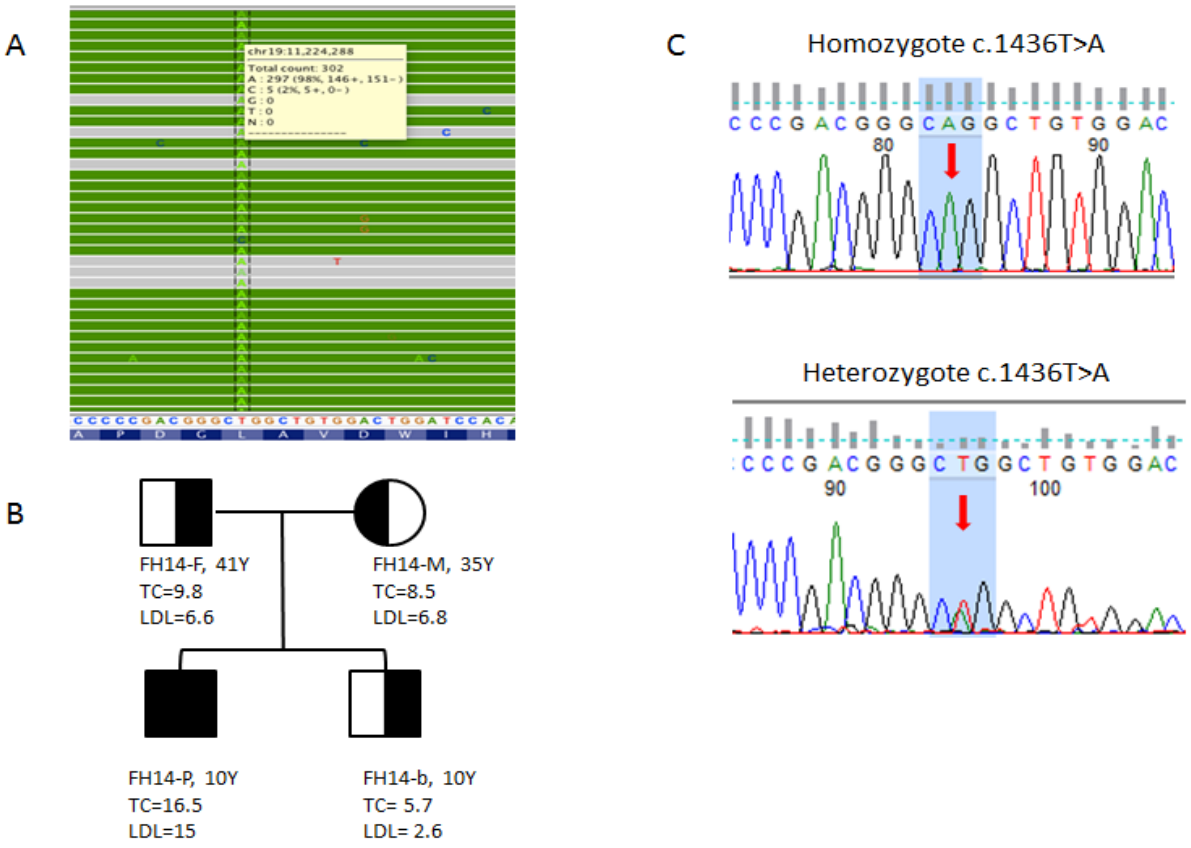
Figure 4-3: Sequence conservation of the two *LDLR* novel mutations in the Iranian study a cross-species



The homozygous p.(Leu479Gln) variant was found in a 10 year-old boy (FH14-P) who had high levels of TC (16.5 mmol/l) and LDL-C (15 mmol/l). Co-segregation analyses showed that this mutation is an FH-causing mutation and was present in all members of the family including both parents and a sibling (Figure 4-4). They were all heterozygous for this mutation; both parents had raised cholesterol levels [father (FH14-F): TC = 9.8mmol/l, LDL-C = 6.6mmol/l; mother (FH14-M): TC = 8.5mmol/l, LDL-C = 6.8mmol/l], but not the sibling.

Figure 4-4: Co-segregation analysis of a family with novel mutation at exon 10 of *LDLR*

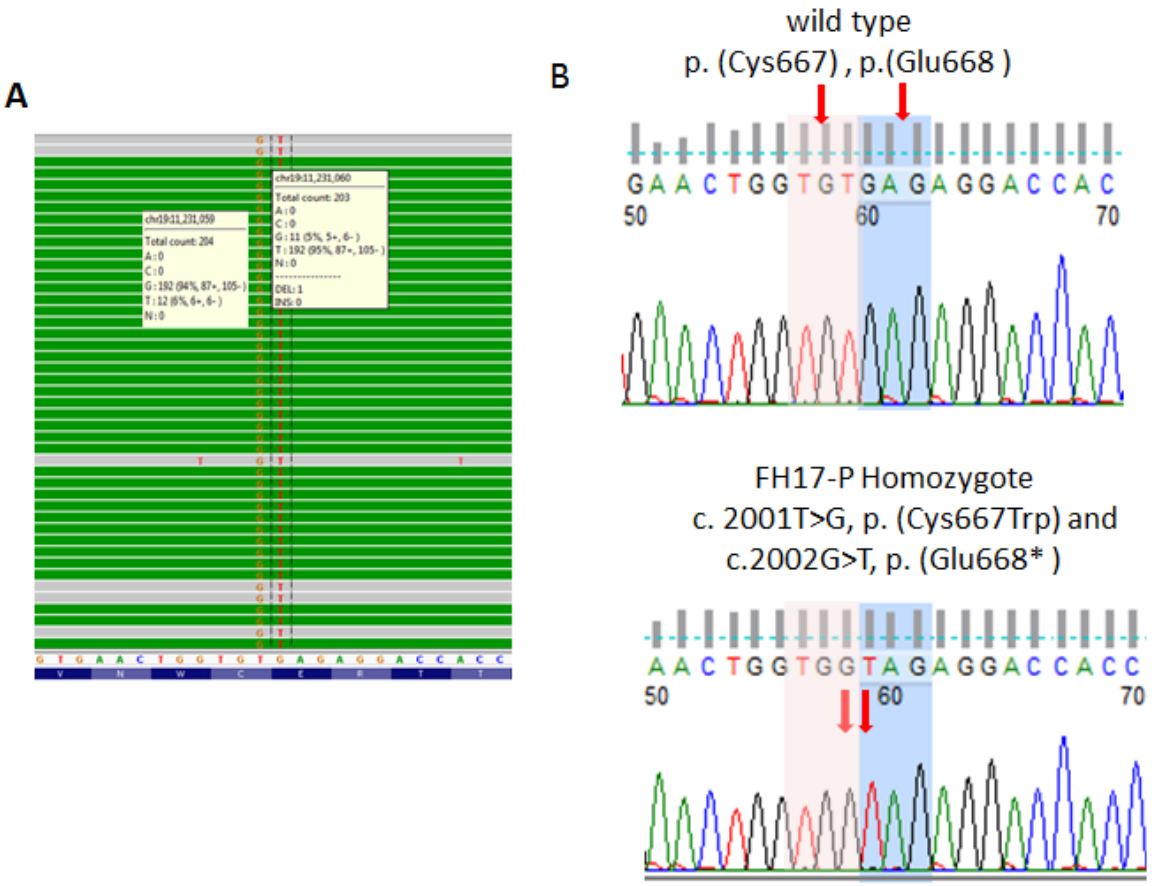
A) NGS sequencing output for proband FH14-P reports a variant at exon 10 of the *LDLR* gene [c.1436T>A, p.(Leu479Gln)], the call rate A allele of the variant is 95% which indicates the proband is a homozygous for the variant. B) Co-segregation analysis of the FH14 family shows the 3 other family members are heterozygous for the variant. C) Sanger sequencing shows a base change between homozygote and heterozygote inherited modes.



The second novel mutation, p.(Glu668*), was found in a 4 year-old girl ((FH17-P), Figure 4-5) who had high levels of TC (15.0mmol/l) and LDL-C (13.6mmol/l), but no family history of either hyperlipidemia or MI. It was also found that the proband was homozygous for another mutation p.(Cys667Trp). Family members of the proband were not available for segregation analysis.

Figure 4-5: a novel mutation at exon 14 of *LDLR* sequence data

A) NGS sequencing output for proband FH17-P reports two variants at exon 14 of the *LDLR* gene [c. 2001T>G, p.(Cys667Trp) and c.2002G>T, p.(Glu668*)], the call rate of minor alleles of the variants are ≥94% which indicate the proband is a homozygous for the variant. B) Sanger sequencing shows a base change between wild type and mutant homozygote inherited mode.



4.2. Discussion

This study has identified a number of previously reported as well as novel mutations among patients with a clinical diagnosis of HoFH and HeFH attending various Iranian clinics. Amongst the 16 child probands clinically diagnosed with HoFH, mutations were found in 9 patients (57%). Different mutations were identified using the targeted NGS method and were confirmed by Sanger sequencing, all in *LDLR*, where five had been previously reported as FH-causing mutations, but were not previously reported in the Iranian population, and two, p.(Leu479Gln) and p.(Glu668*) that were first reported in this study.

This study has found two common mutations in people of Persian descent. The most common mutation was p.(Trp577Arg), which was seen in three patients with HoFH, explaining 37.5% of HoFH *LDLR* identified mutations. This was previously reported as very common in the Turkish population explaining 21.4% of the predicted defective *LDLR* alleles (343, 344). For populations that are geographically close, genetic and geographic distances are often highly correlated (350). Since there is close geographical and cultural links between Turkey and Iran they may share some common founder mutations. The three probands were from Esfarayen (FH1-P, FH2-P) and Guchan (Fh7-P) which are in the north-east of Iran on the border with Turkmenistan (Figure 4-6) and 74% of the Turkmenistan population are Turkmen (Turkish background).

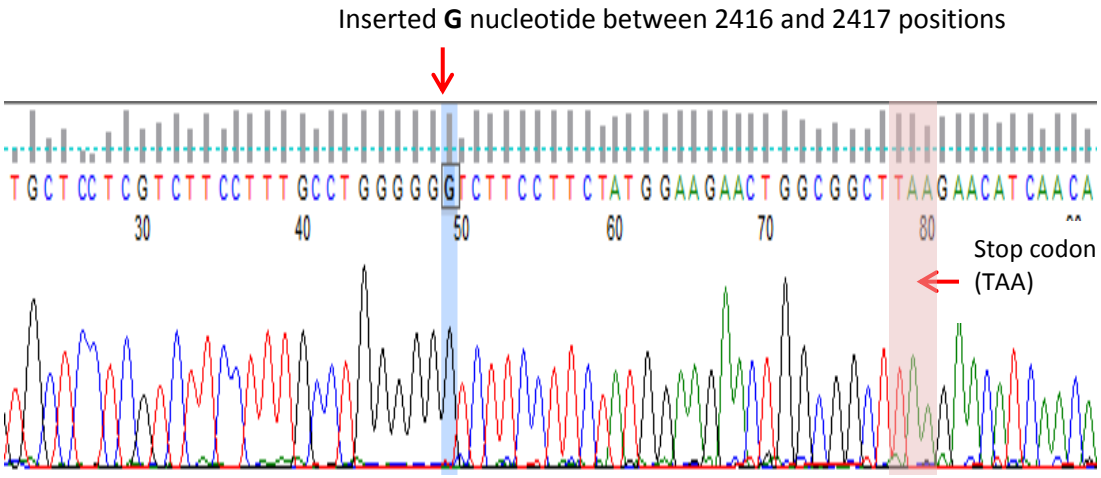
Figure 4-6: Map of Iran representing the region of *LDLR* p.(Trp577Arg) affected probands

The second common mutation was a frame-shift mutation p.(Val806GlyfsTer11) in exon 17, which was found in two patients with HoFH. This mutation has been reported in different populations in the USA (346), South Sweden (347), Czech (348), Netherlands (325) and Japan (349). There are two possible explanations for the presence of the same mutation in different populations. First, there may be an ancestral mutation occurring in an individual living somewhere in the European continent, which has then spread throughout the world due to human migration. The FDB mutation (R3527Q), for example, is an ancestral mutation found in western Europe 6000-7000 years ago, and distributed to the East and West via human migration in the past 2000-3000 years (330).

This possibility could be tested by determining if the mutation occurs on the same SNP haplotype, but data are not available to carry this out.

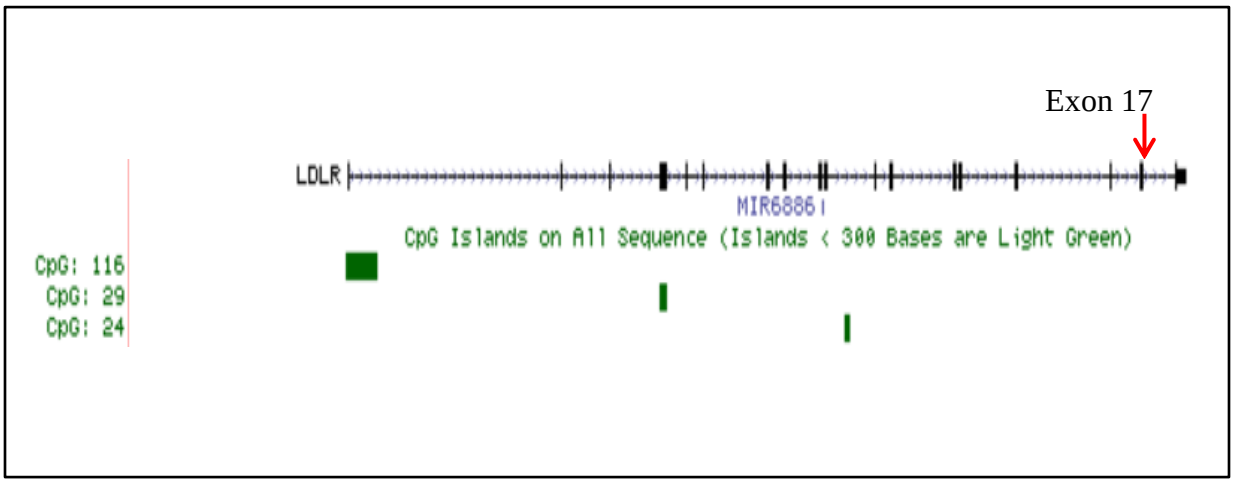
The second possible explanation is that the mutation may be located in a CpG (5'—C—phosphate—G—3') mutation hotspot (351). Cytosine in the CpG dinucleotide carries a methyl group which is added via methyltransferase to form 5-methylcytosine, which causes DNA methylation and affects gene transcription via preventing transcription factors and other proteins from binding (352). However, by looking for the sequence around the point mutation and the CpG island map of the *LDLR* gene, the insertion of guanine (G) nucleotide (c.2416_2417insG) does not create a CpG sequence (Figure 4-7).

Figure 4-7: Electropherogram for DNA sequence analysis of 19: c.2416_2417insG (*LDLR*, p.(Val806GlyfsTer11)



As shown in the *LDLR* CpG island map, exon 17 was not included in any CpG island (figure 4-8). The literature reported 12 CpG in the *LDLR* promoter in atherosclerosis patients (353). Thus at the present time it is not possible to distinguish between these two possibilities.

Figure 4-8: The CpG island map of *LDLR* gene by ENCODE, displayed using UCSC genome Browser



Green boxes are indicating CpG islands. CpG island map shows that *LDLR* gene has three CpG islands, in the promoter and 5'UTR (CpG: 116), in exon 4 (CpG: 29), and in intron 10 (CpG: 24).

Five of the identified mutations in this study have been reported in the UCL-FH mutation database, and are predicted to be pathogenic. Two identified mutations are nonsense mutations: p.(Ser130*) (exon 4) is located in repeat three of the ligand binding domain of the receptor and p.(Trp533*) (exon 11) encodes the EGF-like domain. Both mutations create a premature stop codon resulting in peptide truncation and nonfunctional protein products. Also two missense mutations p.(Trp577Arg) and p.(Cys667Trp,) were identified that encode the EGF-like domain. A variant in this domain blocks both receptor dissociation in endocytosis and this influences the machinery responsible for recycling

receptors to the cell surface, which causes the disease phenotype. The exon 12 mutation p.(Trp577Arg) is located in the third Y-repeat (five repeats containing a YWTD motif) of the EGF-like domain of the LDL-R, where other missense mutations have been functionally characterized to be class 2 mutations (transport-defective alleles) (42, 354). Thus we propose that the p.(Trp577Arg) mutation reported here would cause the same class of deficiency due to its location. The exon 14 mutation p. (Cys667Trp) encodes the third cysteine-rich EGF-like growth factor repeat of the LDL-R, which is a region that interacts with other proteins (42)., Six other mutations affecting this codon have been described in the UCL-FH mutation database, one of which, p.(Cys667Tyr) (FH French Canadian-2) (355), results in a transport defective protein (mutation class 2). It was postulated that the p.(Cys667Tyr) mutation identified here would have a similar impact; however, due to the presence of the sequence change creating the nonsense mutation at position 668, we designate the pathogenic mutation in this subject as p.(Glu668*). The fifth mutation was identified in Exon 17 p.(Val806GlyfsTer11) which encodes membrane-anchoring and cytoplasmic tail domains which are vital for locating the receptor correctly in the cell membrane, thus the truncated peptide at this position affects this machinery and leads to the disease phenotype.

Two novel mutations have been identified in *LDLR* that were not reported in the UCL-FH mutation database: a missense mutation in exon 10 p.(Leu479Gln) and truncated peptide in exon 14 p.(Glu668*). Both mutations are located in the EGF-like domain of *LDLR*, which is known to be important for receptor dissociation in endocytosis and receptor recycling to the cell surface, and are thus pathogenic. Analysis of the proband family

members carrying p.(Leu479Gln) showed that the mutation is inheritable and confirms that it is pathogenic and causing FH.

Forty-three percent of subjects examined in this study were FH-mutation negative. It has been demonstrated that the elevated LDL-C levels in mutation negative subjects with a clinical diagnosis of FH may be due to accumulation of LDL-C raising "polygenic FH" (164). There are 12 common "polygenic" LDL-C raising variants described in the European population (MAF > 0.05%), however, the frequency of these variants in the Iranian population has not yet been studied. However, based on data from other countries, it is likely that the elevated levels of LDL-C seen in patients who are FH-mutation negative have a polygenic and not a monogenic explanation, and this could be examined in the future. The other possible explanation for the lack of a mutation in the three known FH genes would be that the hypercholesterolaemia is due to a mutation in an undiscovered gene(s).

Two probands (7 and 12 years) suspected of having FH were included in this study, although their cholesterol levels did not exceed the FH cut-off, but they reported a strong family history of hyperlipidemia and MI. No mutations were found in the three FH genes. In both patients the presence of TX, a family history of CHD and hyperlipidaemia led to testing with a further lipoprotein screen: high levels of Lp(a) were found in both probands (Lp(a) > 100 mg/dL but with LDL-C levels within the normal range (~3.3 mmol/l) when undergoing lipid-lowering drug therapy). An elevated plasma concentration of Lp(a) has

been reported as an independent risk factor for CVD (356-358). The clinical phenotype of Lp(a) hyperlipidemia (Lp(a)-HLP) shows great variability, as does FH, however the diagnosis of Lp(a)-HLP can be determined by the plasma concentration of Lp(a) (359). Currently there is no effective and approved therapeutic drug for Lp(a)-lowering. A cohort study in 2009 showed that the long-time use of Lp(a) apheresis with maximal tolerated doses of lipid-lowering drugs reduced the plasma concentrations of Lp(a) by over 70%, and decreased the risk of CHD events (360). The HEART-UK working group has recommended that patients with progressive CVD, plasma concentrations of Lp(a) above 60 mg/dL, and plasma LDL cholesterol above 3.2 mmol/l (despite maximal lipid-lowering drug therapy) should be considered for LDL apheresis (361). However, other clinicians have disputed whether Lp(a)-HLP should be treated by Lp(a) apheresis (359). There is no clear guidance for the indication of Lp(a) apheresis treatment, however most apheresis centers use the criteria of (a) plasma Lp(a) of 60 mg/dL as the cut-off, and (b) history of advanced and/or progressive CVD, or alternatively (c) presence of additional risk factors for CVD that accelerate the course of the disease (362). Further clinical and experimental studies are required in order to establish diagnoses, screening and therapeutic guidelines for Lp(a)-HLP and combined Lp(a)-HLP/ FH. Several genetic variants have been reported in the *Apo(a)* gene including a coding variant in the kringle domains (e.g. R560S), and a SNP in the promoter region of the gene (362). Indeed, the development of genetics and the availability of high throughput screening technology provide a great opportunity to identify gene variants that may explain disease phenotypes.

4.3. Limitations

There were limitations in this study. One was the small sample size and another was that only patients with HoFH were included in this study. It was not possible to perform segregation analysis for all patients with novel variants due to communication difficulties. Another limitation was that polygenic FH was not examined in the mutation negative group due to lack of reports about polymorphism frequency in the Iranian population, and *Apo(a)* was not screened because of grant and time limitations.

4.4. Conclusion

Although more than 1,700 variants have been reported in *LDLR* (65), there are still novel variants being found, proving the heterogeneity of FH. This is the first study in an Iranian population integrating both clinical and molecular-based techniques in order to elucidate FH heterogeneity and mutation spectrum. FH monogenic mutations occurred in *LDLR* and were identified in 57% of patients with FH clinically diagnosed with HoFH. Among the seven mutations identified in this study, two mutations were novel. It is recommended that FH-mutation negative subjects with a clinical diagnosis of FH are examined for 12 common "polygenic" LDL-C raising variants Also, a genetic screening for the *Apo(a)* gene in patients with Lp(a)-HLP is recommended. Further and larger studies are needed to elucidate FH frequency in Iran and to identify common variants and/ or unique variants in subjects of Persian ancestry.

Chapter 5 : Identification of functional variant(s) behind the GWAS *LDLR* SNP (rs6511720)

5.1.Introduction

Elevated levels of LDL-C is a major risk factor for CHD. LDL-C is mainly cleared by LDL-R, and mutations in *LDLR* are the most common cause of FH (93%), with more than 1,700 variants identified within the gene (65, 314, 363). Most of these mutations are located in the exonic regions, and thus affect protein structure and function. Ten percent of mutations are in the intronic region (exon boundary) and are predicted to affect splicing, while 2% are found in the promoter region. Thirty-six SNPs have been identified within the *LDLR* promoter region, which are predicted to prevent gene transcription via altering transcription factors that have an essential role in *LDLR* expression: they include SREBP, SP1 and CREB (311, 364-366). This finding encourages researchers to identify SNPs in regulatory regions of the gene and determine their pathogenicity and functionality.

GWAS have revealed new genes and variants that are associated with traits and/or disease phenotypes. A large-scale GWAS has identified numerous loci (95 loci) that harbor common SNPs which have relatively small effects on lipid traits, 22 of which are associated with LDL-C levels and CHD events, including SNPs at the *LDLR, PCSK9, APOB* and *APOE* loci (160). However, the majority of common variants that have been discovered in GWAS are in non-coding regions and their functional implications are unknown (367). Interpretation of the molecular mechanisms of non-coding variants is a huge challenge because of linkage disequilibrium (LD), and the diversity of non-coding functions, including transcription, mRNA splicing and control of translation (368, 369).

The T allele of the *LDLR* SNP rs6511720 (G>T) [MAF = 0.10 in the European population, (1000 Genomes Project Phase 3)] has been identified as being associated with lower plasma

levels of LDL-C and a lower risk of CAD in several studies. Teslovich et al. (2010) study showed that the *LDLR* SNP rs6511720 (G>T) is associated with LDL-C ($p = 5 \times 10^{-9}$) and CAD in people with European ethnicity (>100,000 individuals). The study also found this applied to Non-Europeans including East Asian and African American populations (total $p = 4 \times 10^{-117}$) for LDL-C when combined with the European population. This finding supports previous GWAS results, which are summarized in Table 5-1. The *LDLR* SNP rs6511720 (G>T) minor allele frequency is 10-13% in the general population. The SNP has been identified as being associated with lower LDL-C levels (effect size estimates differ between -0.15 and -0.26 mmol/l), and lower risk of CHD. More recently, data from a large meta-analysis, the global lipids genetics consortium, confirmed this finding, where the SNP was associated with LDL-C [β (SE) = -0.227 (0.008), $p= 2.79 \times 10^{-151}$, N = 83,042] (161). Anand et al. (2009) also found that *LDLR* rs6511720 is associated with myocardial infarction (MI); the minor allele is associated with a lower ApoB and ApoB/A1 as well as a lower risk of MI (OR (95% CI) = 0.86 (0.77-0.95); p= 0.011). Furthermore, it is significantly associated with lower risk of Abdominal Aortic Aneurysm (AAA) (OR (95% CI) = 0.76 (0.70-0.83); $p= 2.08 \times 10^{-10}$) (370). Between-study similarities have provided confidence that the *LDLR* SNP rs6511720 is either functional or may be a marker for a functional variant elsewhere in the gene.

Table 5-1: Summary of genome-wide association study results identifying *LDLR* SNP rs6511720

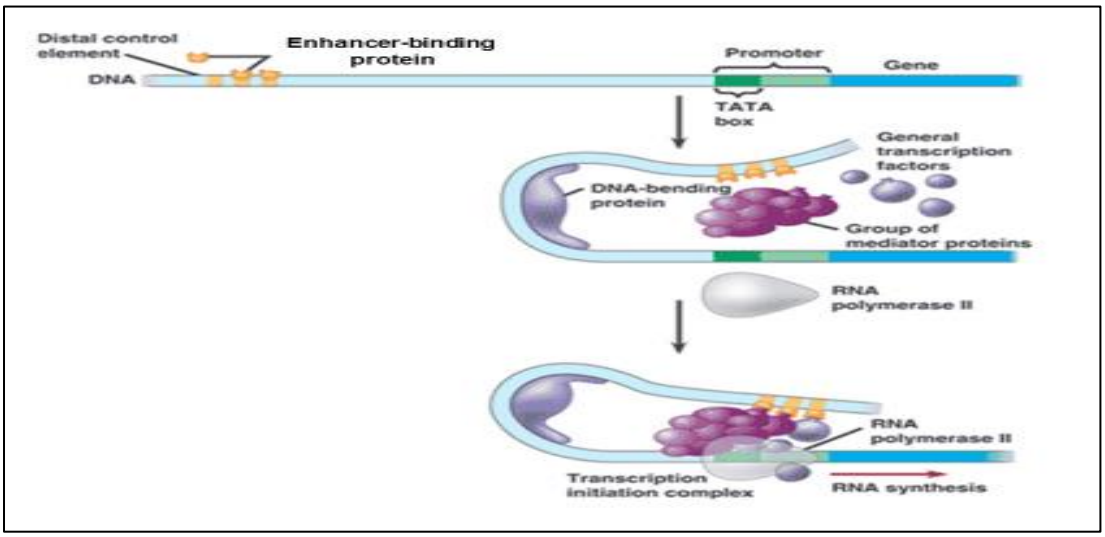
Publication	Population	Sample size	MAF	Effect size	LDL-C p- value	CHD
Bradley et al. (2013) (370)	European	cases (AAA)=5,138 controls=39,273	Cases=0.062 Controls=0.088	-	2.08 $x10^{-10}$	YES (AAA)
Trompet et al. (2011) (371)	European	5244	0.13	-0.19 mmol/l	5.2 $x10^{-15}$	NA
Teslovich et al. (2010) (160)	European, East Asian and African- American	>100,000 European >15,000 East Asians >9,000 South Asians and >8,000 African Americans	0.11	-6.99 mg/dl (-0.18 mmol/l)	4 $x10^{-117}$	YES
Kathiresan et al. (2009) (372)	European	19,648	0.10	-0.26 mmol/l	2 $x10^{-26}$	NA
Anand et al. (2009) (373)	5 ethnic groups (South Asian, European, Arab, Irani, and Nepalese)	6.9% South Asian (n=2346), 9.4% European (n=3666), 14.1% Arab (n=1498), 11.2% Irani (n=402), and 2.1% Nepalese (n=122) Total=8,795	0.096	NA	ApoB/AI= 1.0 $x10^{-10}$	YES (MI)
Kathiresan et al. (2008) (374)	European	8,816	0.10	-0.26 mmol/l	2 $x10^{-51}$	NA
Willer et al. (2008) (375)	European	8,816	0.10	-8.03 mg/dl (-0.20mmol/l)	6.8 $x10^{-10}$	YES
Aulchenko et al. (2008) (376)	European	17,797	0.122	-0.15 mmol/l	6.84 $x10^{-18}$	YES
Miljkovic et al. (2010) (377)	African	1,750	NA	NA	0.0014	NA

MI: Myocardial Infarction, and AAA: Abdominal Aortic Aneurysm

The rs6511720 SNP is located in intron-1 of the *LDLR* gene, where cis-acting gene regulatory sites are commonly found (378). Cis-regulatory elements physically interact with the promoter region of a gene to initiate DNA transcription (Figure 5-1) (253, 365, 379, 380). It was hypothesized that rs6511720 could be a part of an enhancer element which modifies a transcription factor-binding site (TFBS) that in turn recruits co-activators and chromatin regulators to facilitate the transcription of the *LDLR* gene (252). However, analysis of the genetic function of such variants is complex because of the LD between SNPs, which are co-inherited with a causal variant. Thus, all SNPs in LD with the functional variant may be responsible for some or all the associations with the trait of interest, although they may not have any relevant function. Studying all SNPs in an LD block is time and resource costly, thus several bioinformatics databases were used in order to select and prioritize the LD SNPs with the strongest regulatory signs. The selected LD SNPs, in addition to the GWAS hit SNP rs6511720, were functionally analyzed.

Figure 5-1: The role of The cis-regulatory elements in transcriptional regulation

The cis-regulatory elements including promoter and enhancer sequences interact with each other by "DNA looping". DNA looping brings promoters, enhancers, transcription factors and RNA processing factors (RNA polymerase II transcription complex) together to regulate gene expression, and the DNA loop is stabilized by a complex of proteins. Adapted from http://bio1151.nicerweb.com/Locked/media/ch18/activat or.html



5.2. Results

5.2.1. Association of *LDLR* rs6511720 with LDL-C and CHD

5.2.1.1. The rs6511720 genotype-phenotype association

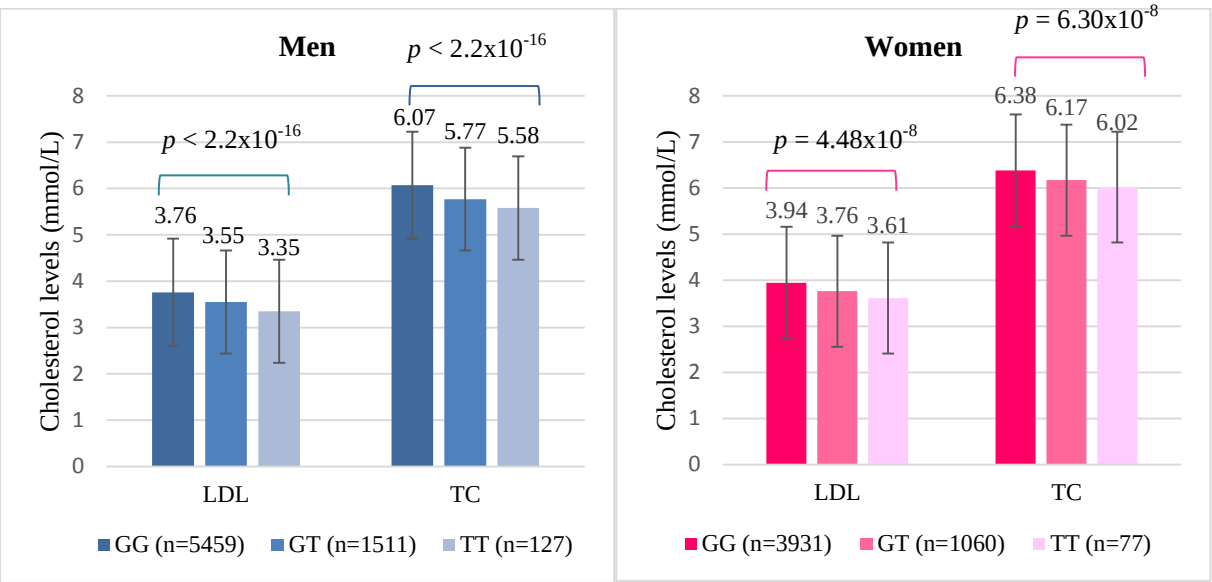
The UCL-LSHTM-Edinburgh-Bristol (UCLEB) consortium replicated an association between *LDLR* SNP rs6511720 (G>T) and cholesterol levels. The UCLEB consortium baseline characteristics and CHD incident rates of the subjects in the individual studies are presented in Table 5-2. Of the 12,395 subjects (range of follow-up 4-20 years, median 10 years), 2,131 (17.19%) developed CHD, but the incidence rate differed among studies due to variations in mean age and duration of follow-up. Genotype was obtained for 12,165 samples and association analyses of *LDLR* SNP rs6511720 with lipid traits were performed. As shown in Figure 5-2, the minor T allele was significantly associated with lower levels of TC (≈7%) and LDL-C (≈10%) in both genders ($p = <2.2x10^{-16}$), TC [Effect size (95%CI) = -0.239 mmol/l (±0.0437), $p < 2.2x10^{-16}$) and LDL-C [Effect size (95%CI) = -0.193 mmol/l (±0.037), $p < 2.2x10^{-16}$). The minor allele was also associated significantly with a lower risk of CHD [OR (95% CI) = 0.883 (0.794 – 0.982); p = 0.02]. This suggests that individuals carrying one or more copies of the minor LDL-C lowering T allele of rs6511720 would have a lower lifetime risk of CHD.

Table 5-2: The baseline characteristics of seven studies in UCL-LSHTM-Edinburgh-Bristol (UCLEB) consortium.

	Studies						
	BRHS	**BWHHS**	**CaPS**	**EAS**	**ELSA**	**ET2DS**	**WHII**
Number included in analysis	2342	1980	1341	764	1883	1007	3078
Number with incident CHD (%)	608	328	345	165	297	214	174
Length of follow-up (years)	20	7	15.5		4	4	10
Age, years	69 (5.65)	71.5 (5.3)	61.9 (5.0)	69.9 (5.6)	73.6 (9.5)	67.9 (4.2)	60.8 (5.98)
Sex, % male	100%	0%	100%	48.4%	52.7%	51.1%	76.7%
BMI, kg/m2	26.8 (3.6)	27.7 (5.0)	26.9 (3.7)	26.2 (4.2)	27.5 (4.5)	31.4 (5.7)	26.8 (4.2)
TC, mmol/l	6.4 (1.1)	6.6 (1.2)	6.2 (1.1)	-	5.7 (1.3)	4.3 (0.9)	5.7 (1.0)
HDL-C, mmol/l	1.15 (0.2)	1.6 (0.4)	-	-	1.5 (0.4)	1.3 (0.4)	1.6 (0.4)
LDL-C, mmol/l	3.9 (1.0)	4.1 (1.1)	-	-	3.4 (1.1)	3.0 (0.9)	3.5 (0.918)
TG, mmol/l	2.1 (1.2)	1.9 (1.0)	1.9 (1.1)	-	1.8 (1.1)	-	1.4 (0.9)
SBP, mmHg	81.8 (12.8)	80.1 (11.8)	81.7 (12.0)	82.3 (12.1)	73.0 (11.5)	69.1 (9.0)	74.5 (10.4)
DBP, mmHg	144.1 (20.0)	150.0 (25.4)	145.0 (22.3)	148.2 (24.6)	139.0 (19.6)	133.3 (16.5)	127.9 (16.4)

- The seven studies are the British Regional Heart Study (BRHS), British Women's Heart and Health Study (BWHHS), the Caerphilly Prospective Study (CaPS), the Edinburgh Artery Study (EAS), the English Longitudinal Study of Aging (ELSA), the Edinburgh Type 2 Diabetes Study (ET2DS), and the Whitehall II study (WHII).
- Body mass index (BMI), total cholesterol (TC) high density lipoproteins-cholesterol (HDL-C), low density lipoproteins-cholesterol (LDL-C), Triglycerides (TG), systolic blood pressure (SBP), and diastolic blood pressure (DBP).
- Mean and standard deviation (SD), where appropriate, are shown.

Figure 5-2: The UCLEB consortium: the *LDLR* rs6511720 genotype and cholesterol levels for men and women



5.2.1.2. Meta-analysis of GWAS

The published GWAS data were also used to identify the association between rs6511720 and LDL-C and CHD. The summary estimates for the log odds Ratio [log(OR)] and its standard error for the relationship between rs6511720 and CHD were taken from CARDIoGRAM (381) and C4D (382) (http://www.cardiogramplus c4d.org/data-downloads/) and were combined by fixed effects meta-analysis. The meta-analysis showed that the rs6511720 minor allele was associated with lower risk of CHD [log (OR) = −0.1155; SE = 0.0217; p = 1.04 x10^{-7}] (Table 5-3) (383).

Table 5-3: Association of rs6511720 genotype and CHD risk in CARDIoGRAM and C4D

SNP(allele)	log(OR) with CHD per-allele	SE (log(OR)0	Sample numbers
rs6511720 (T~minor allele) CARDIoGRAM	- 0.1253	0.03321	8948 cases 4747 controls
rs6511720 (T~minor allele) C4D	- 0.1082	0.0287	15393 cases 15036 controls
fixed effects meta-analysis	- 0.1155	0.0217	24341 cases 19783 controls

Also, the regression coefficient for rs6511720 on LDL-C was estimated from the Global Lipids Genetics Consortium (GLGC) data (161) (http://csg.sph.umich.edu/abecasis/public/lipids2013/). The GLGC (N = 170,607) showed that the rs6511720 minor allele is associated with lower levels of LDL-C [beta = −0.2209; SE = 0.0061; p = 3.85 x10^{-262}] (Table 5-4). These support the earlier finding, where individuals carrying one or more copies of the T allele of rs6511720 have lower LDL-C levels and would have a lower lifetime risk of CHD.

Table 5-4: Association of rs6511720 genotype and lipid traits from data from the Global Lipids Genetics Consortium (GLGC)

	Reference allele	Beta	Se	p-value	Sample size
TC	T	-0.1851	0.0059	5.43 x 10^{-202}	184764
LDL-C	T	-0.2209	0.0061	3.85 x 10^{-262}	170608
HDL-C	T	0.0249	0.0057	6.32 x 10^{-5}	184617
TG	T	-0.0084	0.0056	0.1043	175280

5.2.2. SNP annotation

5.2.2.1. Linkage disequilibrium (LD)

Several SNP annotation databases were used to identify possible SNP functions. First, the Locuszoom (384) (http://csg.sph.umich. edu/locuszoom/), and the HaploReg V2-V4 (290, 385) (http://compbio.mit.edu/Haplo Reg) were used to identify LD SNPs with the *LDLR* GWAS hit SNP rs6511720. It was found that the SNP rs6511720 has strong LD ($r^2 \geq 0.8$) with 48 other SNPs: three SNPs are located in intron-1 and the others are located ≥1.5kb upstream of the *LDLR* locus (Figures 5-3 and 5-4).

Figure 5-3: Linkage disequilibrium plot of the *LDLR* SNP rs6511720

An LD plot was generated using Locuszoom (http://csg.sph.umich.edu/locuszoom/). SNPs are plotted with the meta-analysis p value of LDL-C association (as –log10 values) as a function of genomic position. The lead SNP (rs6511720) is represented by a diamond, while LD SNPs are represented by circles. The LD SNPs are color-coded to represent the r-squared (r2) between SNPs and the putative associated variant, where red indicates strong LD r2≥0.8 and dark blue indicates weak LD r2≤0.2. The blue line indicates estimated recombination rates and the dark blue arrows indicate gene annotations. LD and recombination rates are based on HapMap Phase II (the Centre d'Etude du Polymorphisme Humain collected in Utah, USA, with ancestry from northern and western Europe (CEU)).

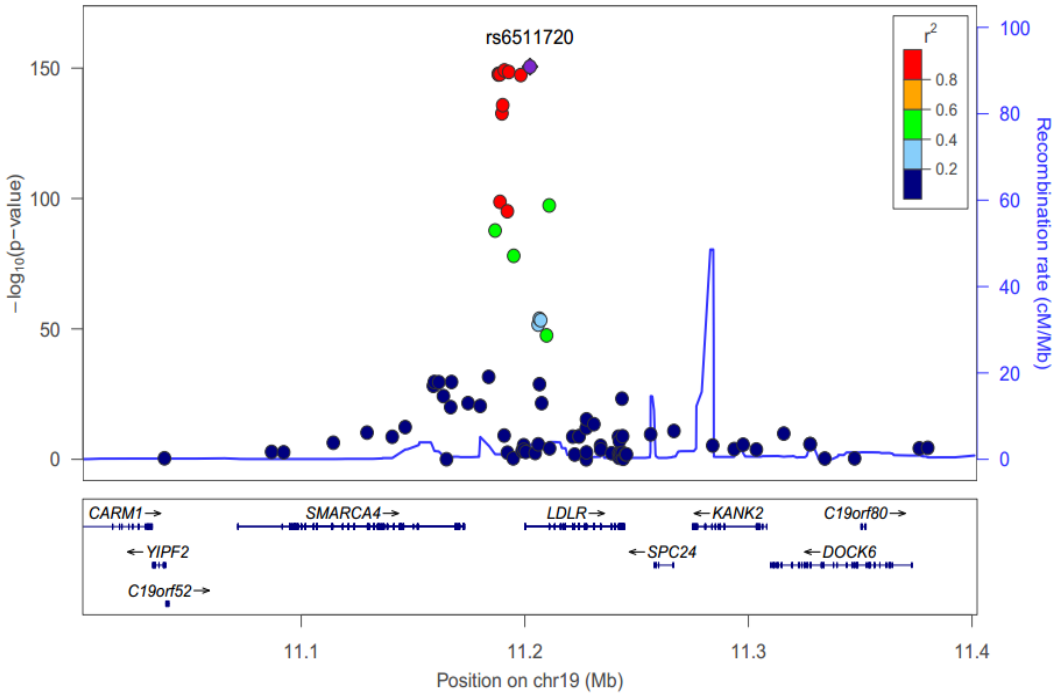


Figure 5-4: The linkage disequilibrium predictions for *LDLR* intron-1 SNP rs6511720 with r^2>0.8

The figure shows the HaploReg reported LD SNPs, 48 SNPs have strong LD (r2≥0.8) with rs16511720, which are reported according to LD information from the 1000 Genomes Project. The HaploReg annotation tools also provide information that relates regulatory elements for the SNP of interest and LD SNPs including, promoter and enhancer markers, DNAse, predicted portion binding, predicted motif changes and their location within the gene. General information about the SNP is also provided; chromosome number, SNP position in the gene, LD value by deviation (D) and frequency (r^2), SNP ID (variant), major allele (Ref) and minor allele (Alt), MAF in four populations: African American, American, Asian and European.

Query SNP: rs6511720 and variants with r^2 >= 0.8

chr	pos (hg19)	LD (r^2)	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	eQTL tissues	Motifs changed	GENCODE genes	dbSNP func annot
19	11187324	0.93	0.98	rs143020224	C	G	0.19	0.11	0.01	0.10			NHLF				4 altered motifs	13kb 5' of LDLR	
19	11187358	0.91	0.98	rs144826254	T	G	0.12	0.10	0.01	0.10			NHLF				5 altered motifs	13kb 5' of LDLR	
19	11187422	0.93	0.98	rs112736558	T	C	0.20	0.10	0.01	0.10			HepG2					13kb 5' of LDLR	
19	11187611	0.88	0.98	rs111989435	A	G	0.20	0.11	0.01	0.11			HepG2		CEBPB		14 altered motifs	12kb 5' of LDLR	
19	11188117	0.94	0.98	rs55997232	C	T	0.12	0.10	0.01	0.10			NHEK, HMEC	HPDE6-E6E7,HEEpiC			Irf	12kb 5' of LDLR	
19	11188153	0.93	0.97	rs55791371	A	C	0.12	0.10	0.01	0.10			NHEK, HMEC	HPDE6-E6E7			5 altered motifs	12kb 5' of LDLR	
19	11188164	0.91	0.96	rs56125973	T	C	0.18	0.11	0.01	0.10			NHEK, HMEC	HPDE6-E6E7			Egr-1,PU.1,Sin3Ak-20	12kb 5' of LDLR	
19	11188247	0.91	0.98	rs56289821	G	A	0.10	0.10	0.01	0.10			HMEC, NHEK				4 altered motifs	12kb 5' of LDLR	
19	11188313	0.94	0.98	rs145329186	C	CTTTA	0.19	0.11	0.01	0.10			HMEC, NHEK				15 altered motifs	12kb 5' of LDLR	
19	11188850	0.93	0.97	rs112898275	T	C	0.18	0.10	0.01	0.10							TAL1,ZEB1	11kb 5' of LDLR	
19	11188899	0.94	0.98	rs112374545	C	T	0.20	0.11	0.02	0.10							11 altered motifs	11kb 5' of LDLR	
19	11189047	0.89	0.95	rs201579954	TA	T	0.42	0.15	0.04	0.10							17 altered motifs	11kb 5' of LDLR	
19	11189205	0.94	0.98	rs148898583	C	G	0.12	0.10	0.01	0.10							CACD,Rad21	11kb 5' of LDLR	
19	11189272	0.94	0.98	rs113722226	T	C	0.18	0.11	0.01	0.10							4 altered motifs	11kb 5' of LDLR	
19	11189298	0.84	0.97	rs200495339	CG	C	0.15	0.10	0.01	0.09							6 altered motifs	11kb 5' of LDLR	
19	11189366	0.88	0.97	rs201408139	C	CAG	0.18	0.10	0.01	0.09							RXRA	11kb 5' of LDLR	
19	11189764	0.94	0.98	rs73015011	T	C	0.19	0.11	0.01	0.10			K562, HepG2	HepG2,K562,HAEpiC	FOXA1,HNF4A,MAFK		4 altered motifs	10kb 5' of LDLR	
19	11189937	0.96	0.99	rs114821903	T	A	0.20	0.10	0.01	0.10			HepG2				4 altered motifs	10kb 5' of LDLR	
19	11189980	0.96	0.99	rs138175288	C	A	0.18	0.10	0.01	0.10			HepG2				6 altered motifs	10kb 5' of LDLR	
19	11190074	0.96	0.99	rs112107114	G	A	0.20	0.10	0.01	0.10							5 altered motifs	10kb 5' of LDLR	
19	11190110	0.96	0.99	rs115594766	A	G	0.19	0.10	0.01	0.10								9.9kb 5' of LDLR	
19	11190292	0.94	0.98	rs112032422	T	C	0.10	0.10	0.01	0.10							Mef2	9.7kb 5' of LDLR	
19	11190481	0.96	0.99	rs77265569	G	T	0.20	0.10	0.01	0.10							ERalpha-a,Pax-4,ZBRK1	9.6kb 5' of LDLR	
19	11190534	0.93	0.97	rs142158911	G	A	0.17	0.10	0.01	0.10								9.5kb 5' of LDLR	
19	11190544	0.88	0.97	rs118068660	C	T	0.17	0.10	0.01	0.09							Ets	9.5kb 5' of LDLR	
19	11190549	0.93	0.97	rs145960625	G	A	0.16	0.10	0.01	0.10							Egr-1,Ets,Pou2f2	9.5kb 5' of LDLR	
19	11190556	0.93	0.97	rs139853365	T	C	0.16	0.10	0.01	0.10							TCF12	9.5kb 5' of LDLR	
19	11190652	0.93	0.98	rs142130958	G	A	0.14	0.11	0.01	0.10							8 altered motifs	9.4kb 5' of LDLR	
19	11190873	0.94	0.98	rs73015013	C	T	0.20	0.11	0.01	0.10			K562	LNCaP,Melano			6 altered motifs	9.2kb 5' of LDLR	
19	11191197	0.94	0.98	rs114846969	G	A	0.14	0.11	0.01	0.10							Nanog,RXRA,ZID	8.8kb 5' of LDLR	
19	11191300	0.94	0.98	rs73015016	G	A	0.29	0.18	0.04	0.10							ELF1,Myc,YY1	8.7kb 5' of LDLR	
19	11191677	0.94	0.98	rs10402112	T	A	0.35	0.17	0.04	0.10							DEC,Nkx2	8.4kb 5' of LDLR	
19	11191729	0.94	0.98	rs138294113	C	T	0.14	0.11	0.01	0.10							SZF1-1,Spz1,ZBTB33	8.3kb 5' of LDLR	
19	11192193	0.94	0.98	rs61194703	A	T	0.23	0.17	0.04	0.10				NHEK,Medullo,Stellate			EWSR1-FLI1	7.8kb 5' of LDLR	
19	11192550	0.93	0.98	rs73015020	G	A	0.24	0.17	0.04	0.10							Hic1	7.5kb 5' of LDLR	
19	11192831	0.94	0.98	rs77140532	A	G	0.34	0.18	0.04	0.10							Zfp410	7.2kb 5' of LDLR	
19	11192915	0.96	0.99	rs73015021	A	G	0.32	0.18	0.04	0.10							Irx,TATA	7.1kb 5' of LDLR	
19	11193091	0.96	0.99	rs112552009	T	G	0.10	0.10	0.01	0.10							4 altered motifs	6.9kb 5' of LDLR	
19	11193949	0.93	0.98	rs10412048	A	G	0.43	0.19	0.04	0.10							CHD2,NRSF,Rad21	6.1kb 5' of LDLR	
19	11196356	0.96	0.99	rs139306531	AC	A	0.13	0.10	0.01	0.10			8 cell types	60 cell types	14 bound proteins		5 altered motifs	3.7kb 5' of LDLR	
19	11196886	0.94	0.98	rs8106503	T	C	0.29	0.17	0.04	0.10			7 cell types	HeLa-S3,Ishikawa			BCL,GATA,TAL1	3.2kb 5' of LDLR	
19	11197261	0.93	0.98	rs12151108	G	A	0.15	0.10	0.01	0.10			4 cell types				Rad21	2.8kb 5' of LDLR	
19	11197598	0.94	0.98	rs73015024	G	T	0.14	0.11	0.01	0.10			K562, NHLF, HMEC				CDP	2.4kb 5' of LDLR	
19	11198187	0.91	0.96	rs17248720	C	T	0.32	0.18	0.04	0.10			4 cell types				Pou3f2,Pou6f1	1.9kb 5' of LDLR	
19	11198502	0.91	0.98	rs17248727	T	C	0.27	0.11	0.01	0.10			4 cell types				9 altered motifs	1.5kb 5' of LDLR	
19	11201124	1	1	rs57217136	T	C	0.29	0.11	0.01	0.10		9 cell types	GM12878, H1, Huvec	77 cell types	17 bound proteins		SP1	LDLR	intronic
19	11201988	1	1	rs60173709	GT	G	0.12	0.10	0.01	0.10		7 cell types	HSMM, Huvec	5 cell types	POL2		8 altered motifs	LDLR	intronic
19	11202194	1	1	rs141787760	GC	G	0.13	0.10	0.01	0.10		7 cell types	HSMM, Huvec		POL2		13 altered motifs	LDLR	intronic
19	11202306	1	1	rs6511720	G	T	0.13	0.10	0.01	0.10		4 cell types	5 cell types	A549,MCF-7	POL2			LDLR	intronic

5.2.2.2. SNP prioritization and selection

For prioritizing and selecting SNPs for functional studies, three bioinformatics resources were used: UCSC Genome Browser (386), MatInspector (291) and HaploReg (290, 385). Using UCSC Genome Browser, genome-wide maps of the chromatin state of relevant cell types (HepG2, human hepatocytes) were examined. Variant position was evaluated for evidence of histone modification markers, H3K4me1, H3K27Ac and H3K4me3, as well as for DNase I hypersensitivity sites and for formaldehyde assisted isolation of regulatory elements (FAIRE). The post-translational chromatin markers H3K4me1, and H3K27Ac are often associated with enhancer regions (387), while H3K4me3 is associated with promoter regions (388, 389). DNase I and FAIRE are established methods for the identification of nucleosome regulatory regions (390). In addition, transcription factor occupancy in chromatin was assessed using genome wide ChIP-seq data sets (276).

It was found that the *LDLR* intron-1 SNP rs6511720 is located in an open chromatin zone, which has strong markers for histone modification markers H3K4me1, H3K27Ac and FAIRE and the region is occupied with transcription factors (Figure 5-5). The same criteria was applied to evaluate the LD SNPs which found that of the 48 variants meeting the LD threshold (r^2<0.8), 38 did not show a regulatory effect on a hepatic cell line. Further, seven SNPs (rs8106503, rs139306531, rs138175288, rs114821903, rs73015011, rs111989435, rs112736558) are located >3kb upstream of *LDLR*, while the 5'UTR functional region of the *LDLR* promoter starts -250bp upstream of the translation start site (311, 366).These seven SNPs also have a low regulatory markers score and are not predicted to bind to regulatory proteins. Therefore, these 45 SNPs were excluded from this study, However, three SNPs were found to have strong chromatin

signals [rs57217136 (T>C), rs141787760 (C>deletion) and rs60173709 (T>deletion] and are in complete LD (r^2=1) with GWAS hit SNP rs6511720. This GWAS SNP and the further three selected SNPs are located in intron-1 of *LDLR* within ≈1200bp of each other. Figure 5-6 shows the *LDLR* selected SNPs location within intron-1 and possible haplotypes.

Figure 5-5: Genome road map of chromatin state of *LDLR* intron-1.

Schematic presentation of the *LDLR* intron-1 chromatin status (https://genome-euro.ucsc.edu). The intron-1 of the *LDLR* has a strong histone marker and FAIRE signal. It is located at DNaseI hypersensitivity sites and within an area that harbors transcription factor binding sites. The area of interest of intron-1 is highlighted in light-blue, where the four SNPs of interest (rs6511720, rs141787760, rs60173709, and rs57217136) are located (383).

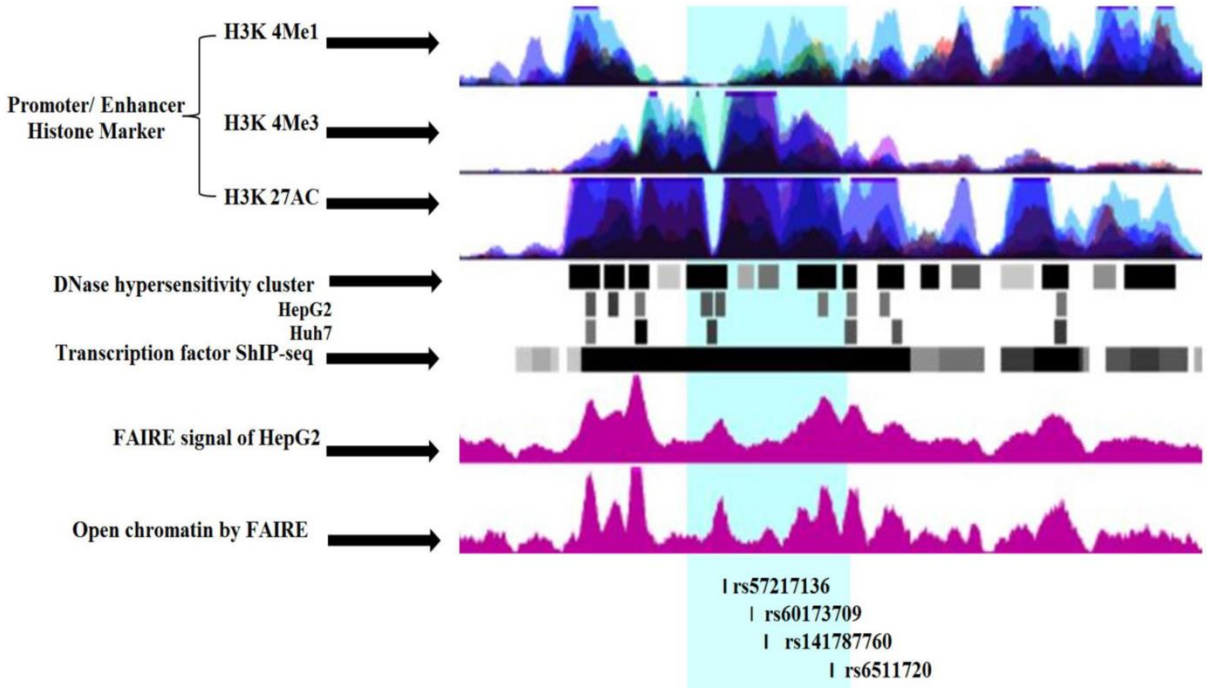
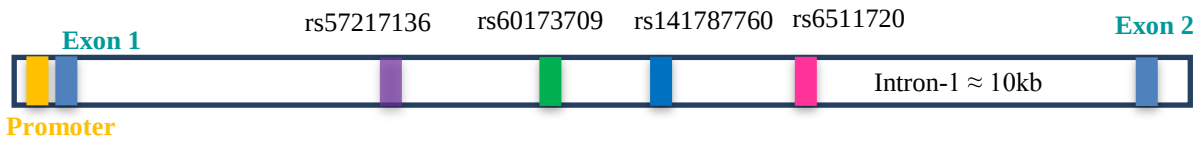


Figure 5-6: *LDLR* intron-1 SNPs location and possible phenotypes

The figure shows the schemes of *LDLR* intron-1and the four potential functional candidate SNPs, [rs6511720 (G>T) rs57217136 (T>C), rs141787760 (C>deletion) and rs60173709 (T>deletion)]. These SNPs are in complete LD (r^2=1), thus two haplotypes are possible; all major alleles together or all minor alleles together.



MatInspector and HaploReg tools, showed different DNA-protein binding profiles for these SNPs (Table 5-5); the *LDLR* SNP annotated results suggest that rs6511720 is a strong enhancer, while the other three SNPs are involved in gene expression by activating the *LDLR* promoter [Encode data (391)]. In addition, MatInspector (*in silico*) and HaploReg (ChIP-seq data) showed different DNA-protein binding profiles for these SNPs. The MatInspector software predicted a sequence around the minor allele of rs6511720 that would bind to GATA and the snRNA activating protein/ proximal sequence element (SNAP/PSE) complex, but HaploReg did not report any protein binding around this SNP. SNAP/PSE has a role in gene transcription initiation (392-394). MatInspector did not predict any binding sites for either rs141787760 or rs60173709, while HaploReg software showed that the minor allele of rs141787760 would bind to transcription factors that have roles in chromatin modulation through general transcription regulation (such as CHD2, CTCF, Egr-1and p300) and a known *LDLR* transcription regulator specificity protein 1 (SP1). The rs60173709 SNP was also reported as a binding site for some proteins, but none of them have a role either in chromatin modulation or in *LDLR* gene

regulation. For rs57217136, HaploReg showed that the major allele would bind to TFs such as forkhead box protein A (FOXA1, FOXA2), which are liver transcription activators, and was also predicted to bind to sterol regulatory element-binding protein 1 (SREBP1), while the minor allele would bind to SP1.

Table 5-5: Predicted regulatory element and protein binding of the *LDLR* selected SNPs

SNP IDs	Position (hg19)	Regulatory element [30]	Allele	Predicted protein binding and motif change (Haploreg)	Predicted protein binding (MatInspector)
rs6511720	11202306	Strong Enhancer	G	-	BLMP
			T*	-	GATA1, SNAP/PSE
rs141787760	11202194	Active Promoter	C	CDP	-
			deletion*	CHD2, CTCF, ELF1, Egr-1, SP1, Zic, and p300	-
rs60173709	11201988	Active Promoter	T	EBF, Myf, Pax-5	-
			deletion *	GR and MAZ	-
rs57217136	11201124	Active Promoter	T	FOXA1, FOXA2, and SREBP1	-
			C*	SP1	OAZF/ROAZ

* = SNP minor allele

5.2.3. Allele-specific protein binding of *LDLR* intron-1 SNPs in Huh7 cells

The SNP annotation analysis suggested that the *LDLR* intron-1 selected SNPs are sites for transcription factor binding. To assess whether the alleles of the SNPs differentially affect DNA-protein binding *in vitro*, EMSAs for the four intron-1 SNPs were carried out. This technique is based on the fact that DNA-protein complexes migrate slower than non-bound DNA in polyacrylamide gel electrophoresis, resulting in a "shift" in migration of the labeled DNA band. By using a non-isotopic method to detect DNA-protein interactions, biotin end-labeled DNA containing the binding site of interest was incubated with a Huh7 nuclear extract. Each set of biotin labeled probes was run alongside an unlabeled probe that was added to the reaction before adding the biotin labeled probe, which led to the unlabeled probe competing with the labeled probe to bind to protein. Multiple conventional EMSAs were performed for both alleles of four *LDLR* intron-1 SNPs: rs6511720, rs57217136, rs60173709, and rs141787760 (Figure 5-7). The SREBP1 was used as a control. The alteration of binding intensity or band shifting between alleles are a sign of a change in DNA-protein binding that may be involved in *LDLR* regulation (383).

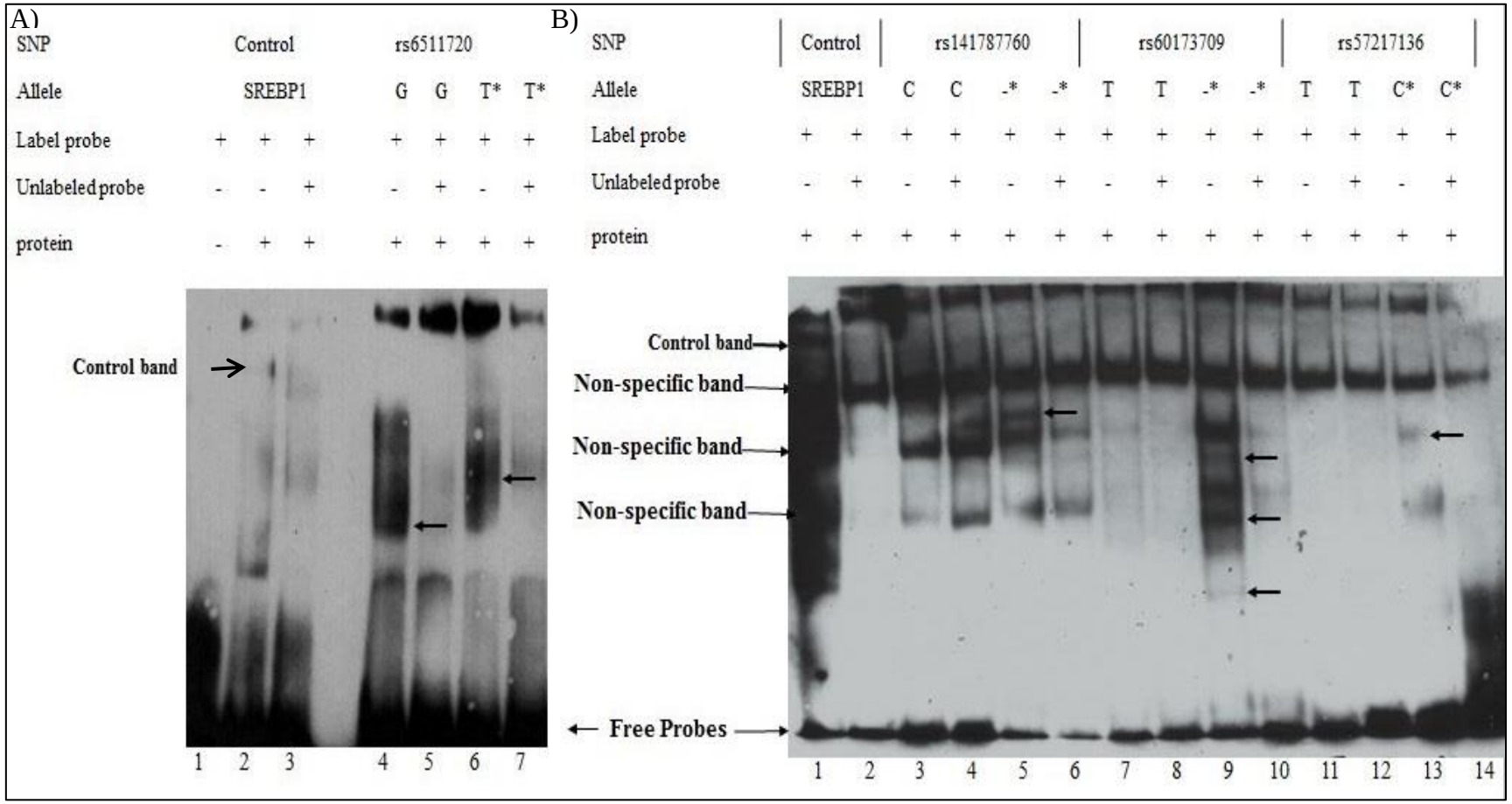
In Figure 5-7 (A), gel A, SREBP1 was used as the control (lanes 1-3). The negative control contained no nuclear extract and no bands were visible (lane 1), whereas the positive control contained a labeled probe that showed a band indicated by arrow (lane 2). When the unlabeled probe was added the specific band disappeared (lane 3). For the rs6511720 SNP (lines 4-7), it was observed that the minor T allele of the SNP in lane 6 formed a different protein complex which moved slower than the major G allele complex in lane 4 (bands shift indicated by the arrows): also there were different binding affinities between alleles. These allele protein

complexes were eliminated in the presence of unlabeled probes (lanes 5 and 7), which indicate the protein complexes observed in lanes 4 and 6 are allele-specific (383).

In Figure 5-7 (B), gel B shows the EMSA results of three *LDLR* intron-1 SNPs that have complete LD with rs6511720. Lanes 1 and 2 contained SREBP1 as the positive control while lanes 3 to 6 have rs141787760 allele probes: the major and minor alleles showed three non-specific bands, but the rare allele showed a new band (lane 5), which is indicated by an arrow, this specific band was eliminated when an unlabeled probe was added (lane 6) (383). The rs60173709 alleles were loaded in lanes 7-10; the major allele (T) showed two weak bands where the non-specific bands are located (lane 7), while the minor allele showed three new bands (lane 9) that were eliminated by the unlabeled probe (lane 10). The rs57217136 (lanes 11-14) minor allele (C) showed two bands which were eliminated by the unlabeled probe (lanes 13-14). These two bands were located where unspecific bands were found, but they had 90% less intensity than the unspecific bands observed in other SNPs, which suggest that perhaps the C allele migrates alongside non-specific bands. The major allele (T) did not show any protein binding (lanes 11-12).

Figure 5-7: DNA binding properties of *LDLR* intron-1 SNPs

The conventional EMSA analysis of the *LDLR* intron-1 SNPs (rs6511720, rs141687760, rs60173709, and rs57217136). The biotin-labeled probes were run alongside unlabeled probes. The minor allele is labeled with (*). (A) EMSA for the *LDLR* GWAS hit SNP rs6511720. (B) The *LDLR* complete LD SNPs with GWAS hit SNP: rs141787760, rs60173709, and rs57217136. These four SNPs have allele-specific binding, indicated by arrows. (-) = deletion, (*) = minor allele. SREBP1 was used as a positive control.

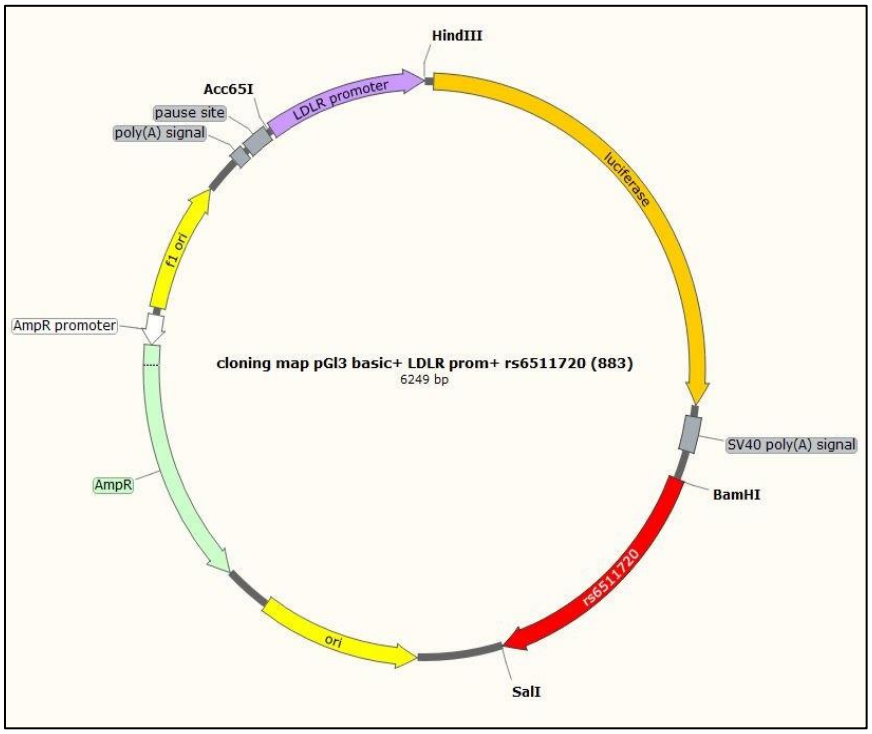


5.2.4. Allele-specific enhancer activity of *LDLR* intron-1 SNPs in Huh7 cells

To determine whether *LDLR* intron-1 SNPs have an influence on gene expression, reporter gene expression assays were performed, where the *LDLR* promoter 594bp DNA fragment (full *LDLR* promoter) was inserted into a pGL3-basic luciferase reporter vector to generate p*LDLR* promoter. The insertion was upstream of the luciferase gene (*luc*+) at the promoter site. Then the *LDLR* SNP sequences encompassing each SNP allele were individually added to the p*LDLR* promoter construct, and inserted into the enhancer site of the p*LDLR* promoter after the SV40 polyadenylation signal (Figure 5-8).

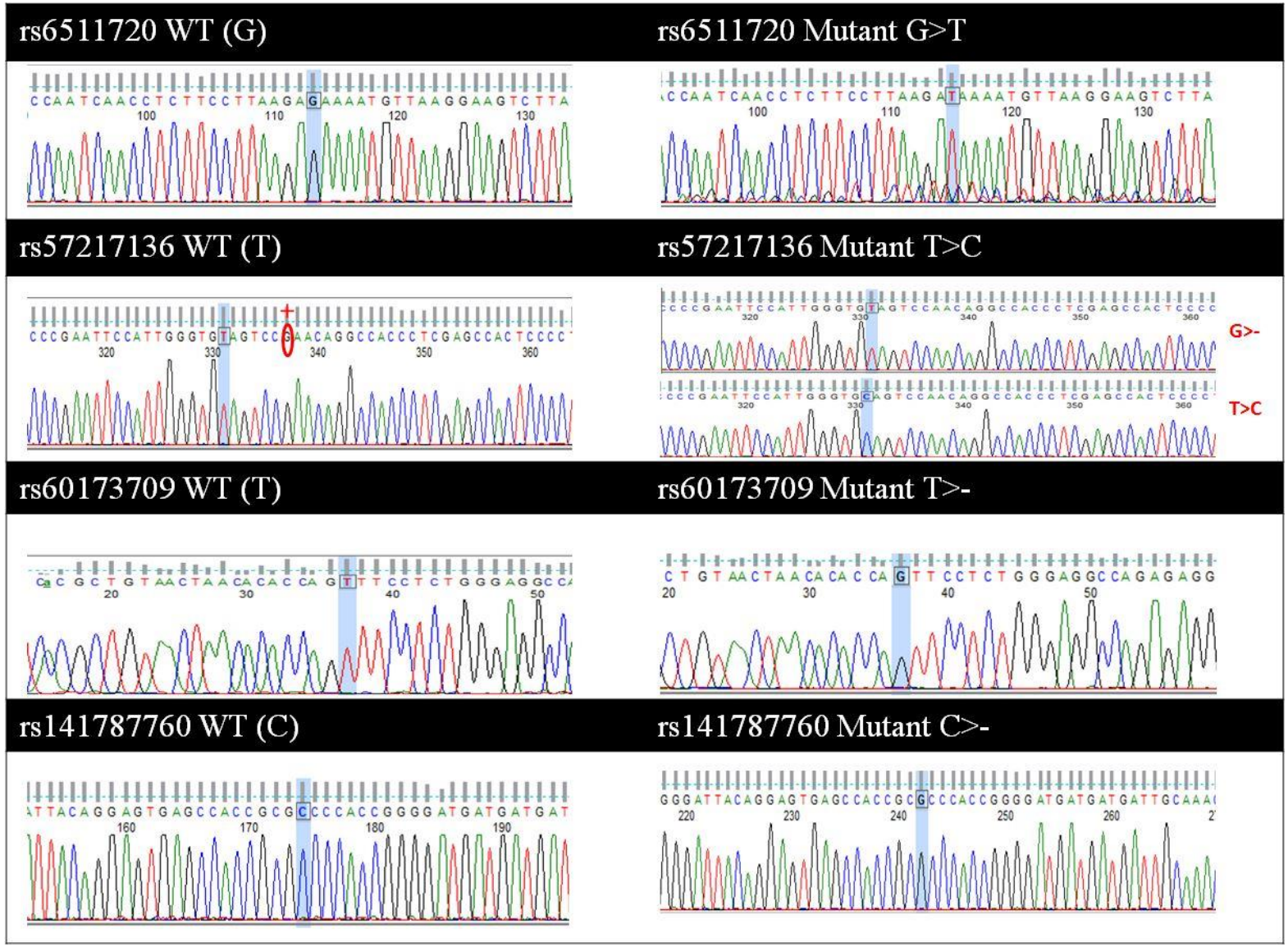
Figure 5-8: An example of *LDLR* SNP cloning construct map

This cloning construct map shows where 594bp of the *LDLR* promoter fits into the promoter site of the pGL3-basic luciferase reporter vector, and the *LDLR* SNP (e.g., rs6511720) fits into the enhancer site of the vector. The map was generated by SnapGene (http://www.snapgene.com/)



I tried to amplify a fragment that contained all four SNPs, however, it was impossible due to Arthrobacter luteus repeats (Alu repeats): the *LDLR* intron-1 has 20 Alu repeats (395) which prevent DNA amplification. Therefore, this fragment was divided into three fragments to exclude Alu repeats; the first fragment contained rs6511720 (884bp), the second fragment (643bp) contained both SNPs rs60173709 and rs141787760, and the third fragment contained (814bp) rs57217136. The SNP variants were generated using site-directed mutagenesis (see Figure 5-9 for base change sequence of *LDLR* intron-1 SNPs). Seven constructs were generated, which were then grown in E-coli cells, and then transfected into Huh7 cells along with the control plasmid pRL-TK vector that encodes the firefly luciferase gene to normalize for transfection efficiency. The pGL3-*LDLR* promoter WT was used as a normalizing factor. The luciferase assays were carried out in four replicates, and the statistical analysis mean relative expression difference between wild type and variant was determined by a t-test.

Figure 5-9: Base change of *LDLR* intron-1 SNPs by Site-Directed Mutagenesis



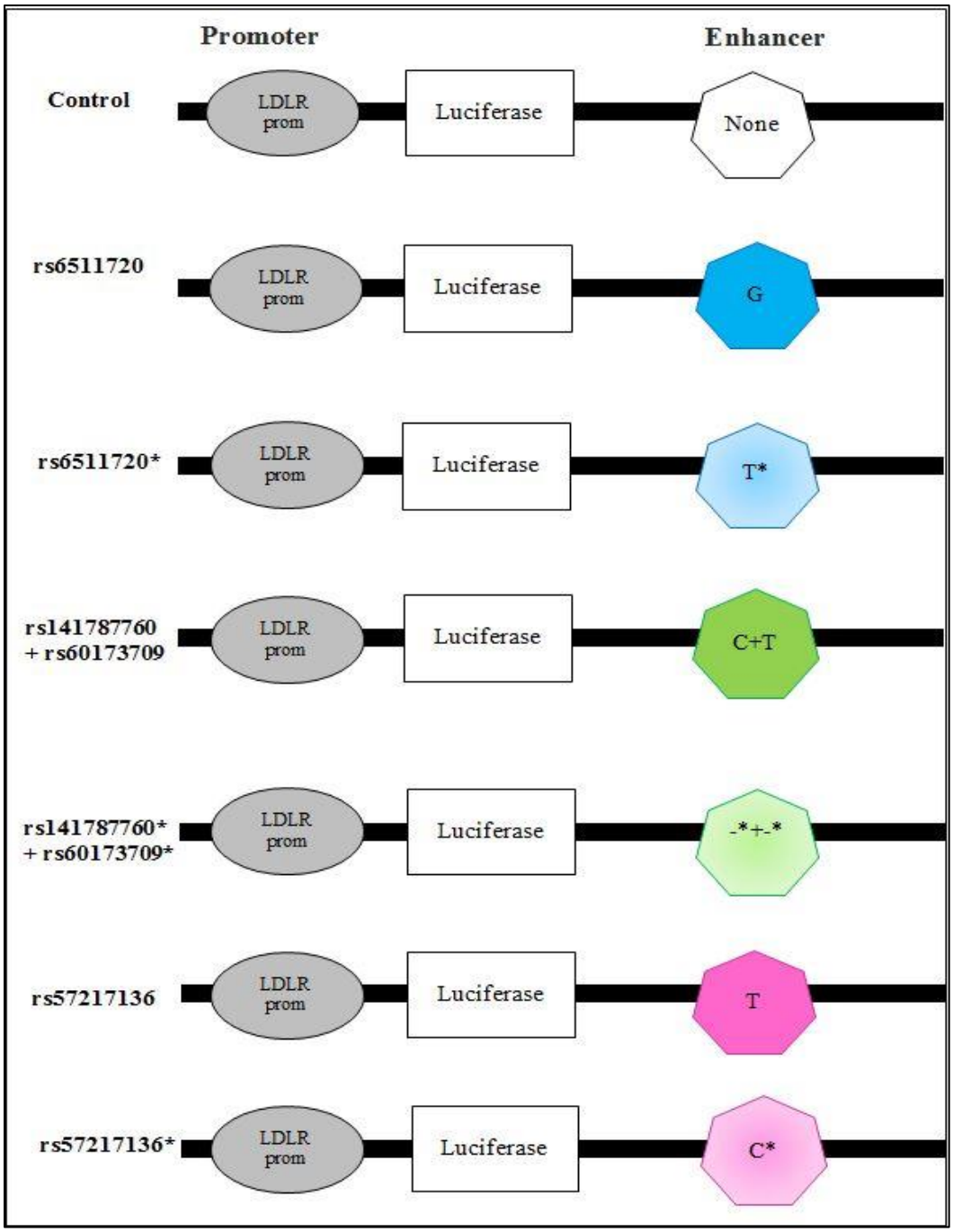
rs57217136 WT sequences had an extra G nucleotide at position 337. The first SDM was done to delete G, and then the T allele was substituted for the C allele

To assess the effect of the *LDLR* SNPs transfected construct on *LDLR* promoter activity, the relative expressions were normalized to the promoter construct alone (Figure 5-10 A & B). Luciferase activity measurements showed that although the rs141787760/rs60173709 transfected construct led to a significant reduction in *LDLR* promoter expression when compared to the promoter alone: -24% ($p = 0.006$) for the major allele and -29% ($p = 4.0$x10-8) for the minor allele, there was no significant difference between the alleles ($p = 0.28$). In contrast, both major and minor alleles of rs6511720 appear to function as enhancers, and increased *LDLR* promoter expression by +54% and +89% ($p = 1.88\text{x}10^{-6}$, and $p = 3.70\text{x}10^{-13}$), respectively. The rs57217136 SNP only had enhancer activity in the presence of the minor allele, which increased *LDLR* promoter expression by +28%. These data suggest that the minor alleles of rs6511720 and rs57217136 function as enhancers. The minor alleles of both rs6511720 and rs57217136 increased *LDLR* promoter expression activity significantly by +29% and +24% ($p = 0.026$, and $p = 0.002$), respectively, when compared to the major alleles. Since higher expression of *LDLR* would lead to greater number of LDL-R on the surface of the liver, this would increase clearance of LDL-C from the blood, and would explain the lower LDL-C levels reported in individuals with the T allele of rs6511720 (383).

Figure 5-10: *LDLR* luciferase constructs map and SNP luciferase activity in Huh7 cell line

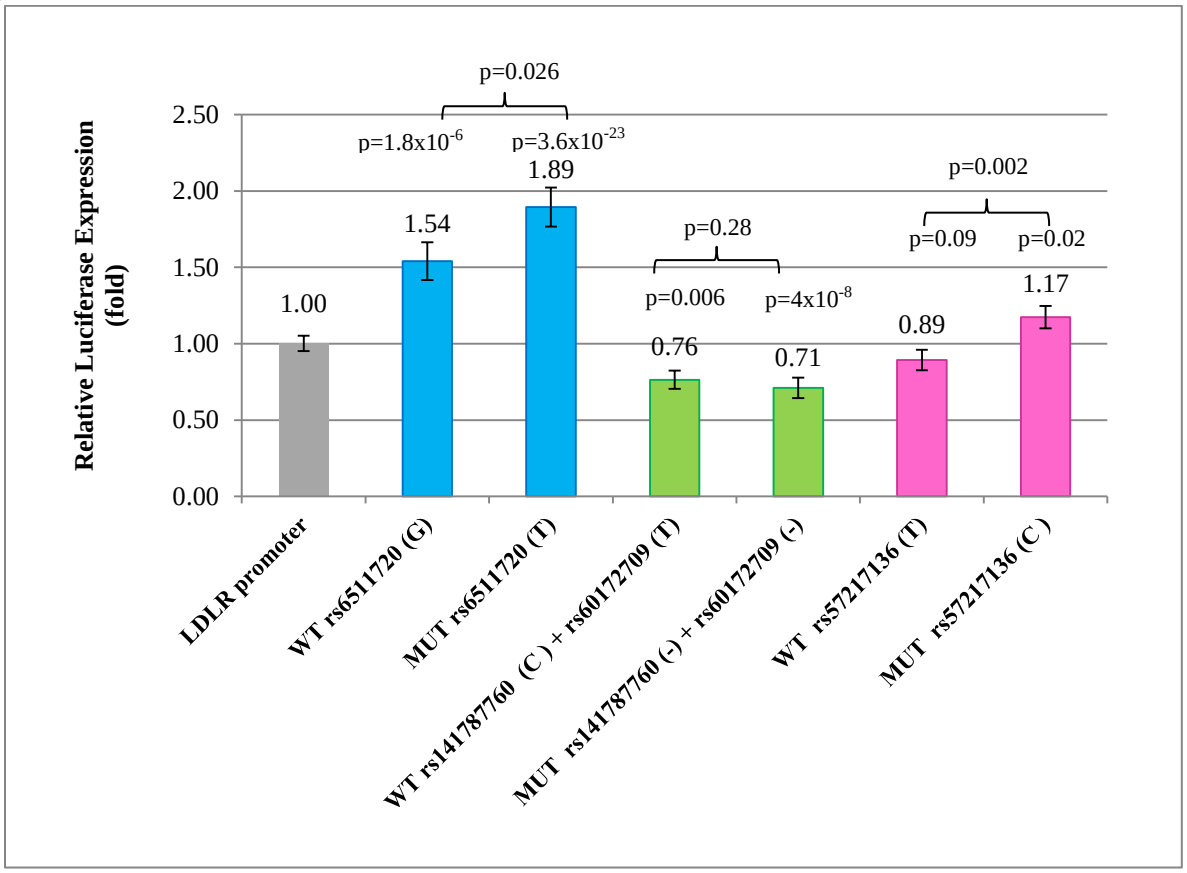
(A) Schematic presentation of LDLR-luciferase-construct (promoter only) and LDLR-luciferase-enhancer-constructs. (B) Results of luciferase reporter assays showing relative expression of LDLR-luciferase-enhancer constructs of LDLR SNPs relative to the LDLR-luciferase (no enhancer) construct.

A)



(-) = deletion and (*) = minor allele

B)


Relative Luciferase Expression (fold)
2.50
2.00
1.50
1.00
0.50
0.00
p=0.026
p=1.8x10-6
p=3.6x10-23
p=0.28
p=0.006
p=4x10-8
p=0.002
p=0.09
p=0.02
1.00
1.54
1.89
0.76
0.71
0.89
1.17
LDLR promoter
WT rs6511720 (G)
MUT rs6511720 (T)
WT rs141787760 (C) + rs60172709 (T)
MUT rs141787760 (-) + rs60172709 (-)
WT rs57217136 (T)
MUT rs57217136 (C)

5.2.5. Multiplexed competitor electrophoretic mobility shift assay (MC-EMSA)

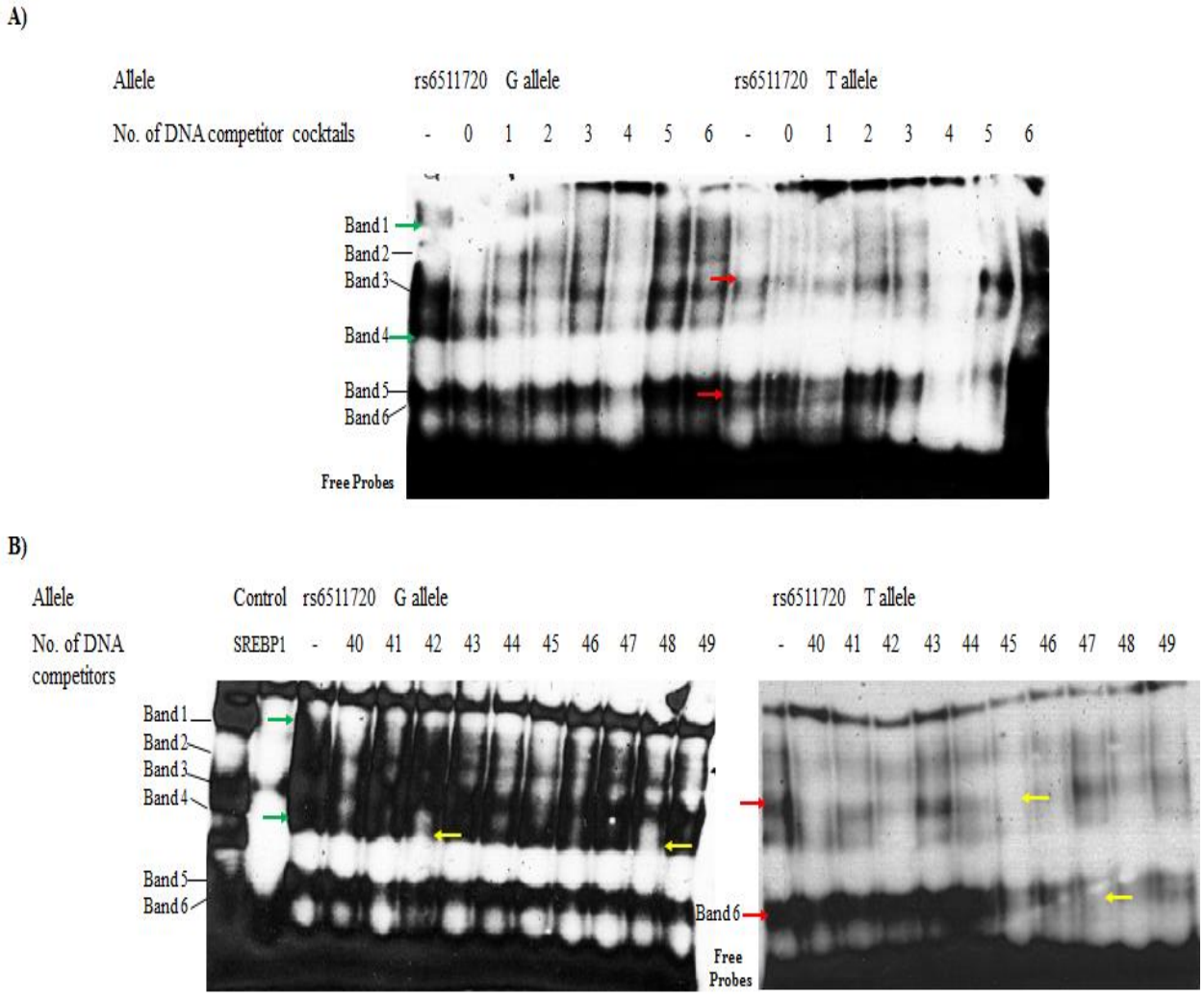
To identify the transcription factors that bind to *LDLR* intron-1 variants and cause differential expression in the luciferase assay, MC-EMSAs were carried out. Multiple MC-EMSAs were performed using seven sets of cocktails, with each set having 10 unlabeled dsDNA consensus sequences for well-characterized proteins. These proteins were multiplexed and added to the binding reaction comprising Huh7 nuclear extract. When a particular set of a multiplexed competitors eliminated the allele specific band(s), the individual competitor from this set were examined further using EMSA. The two SNPs, rs6511720 and rs57217136, showed a different DNA-protein banding in conventional EMSA and an increase in gene expression of luciferase reporter gene in MC-EMSA.

As shown in Figure 5-11(A), both alleles of the *LDLR* SNP rs6511720 were run with the seven cocktails. For the rs6511720 G allele specific bands (Bands 1 & 4) were competed out by 'cocktail 4' competitors, where the bands' intensity reduced by ≈90%, while the rs6511720 T allele specific bands (Bands 3 & 6) were eliminated by 'cocktail 4' competitors. Further EMSAs were carried out on individual proteins of 'cocktail 4' in order to determine the transcription factors that mediate *LDLR* gene expression (see Figure 3-10(B)). It was observed that two proteins bind to the sequence around the rs6511720 G allele. The rs6511720 G allele specific band (Band 4) was eliminated by sis-inducible element (SIE) and signal transducer and activator of transcription 1 (STAT1) (competitors #42 and 48, respectively). However, the other G allele specific band (Band 1) was competed out by most competitors, which indicates none of them are specific to the G allele. On the other hand, the rs6511720 T allele specific band 6 was eliminated

by sterol response element (SRE) (competitor #47), while the T allele specific band 3 was eliminated by several competitors.

Figure 5-11: DNA binding and expression of the transcription factors of the *LDLR* SNP rs6511720 (G>T)

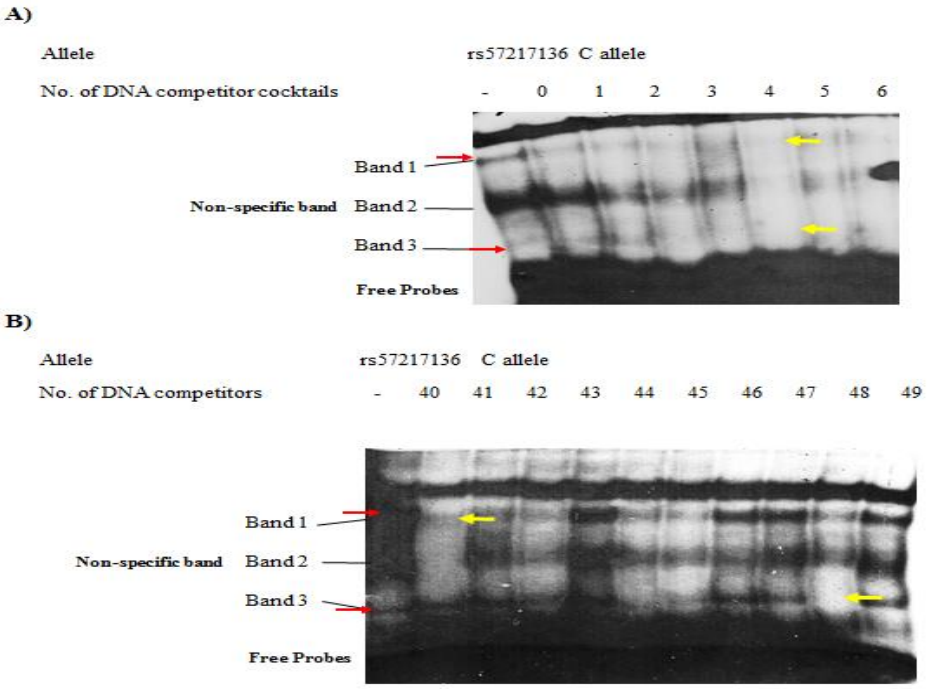
MC-EMSA analysis. (A) Nuclear proteins from the Huh7 cell line were incubated with 7 cocktails of unlabelled DNA competitors. (B) The single competitors of cocktail 4 were run individually in a further EMSA for both alleles of rs6511720. Bands 1 and 4 are specific for rs6511720 G allele, bands 2 and 5 are non-specific bands, while bands 3 and 6 are specific for rs6511720 T allele and migrate alongside non-specific band 5. The green arrows show G allele specific bands, the red arrows show T allele specific bands, and the yellow arrows indicate positive competitors.



Conventional EMSA showed that the presence of rs57217136 C allele led to creation of a protein binding site to the sequence around the allele, in contrast to the T allele of the SNP that had no specific DNA-protein interaction. Thus, when the rs57217136 C allele was run with the seven cocktails, it was eliminated by cocktail 4 competitors (Figure 5-12 A). In Figure 5-12 B, the individual competitors of 'cocktail 4' showed that the C allele specific bands (bands 1 and 3) competed with Retinoic Acid Receptor (RAR) (competitor #40) and STAT1 (competitor #48), respectively.

Figure 5-12: DNA binding and expression of the transcription factors of the *LDLR* SNP rs57217136 C allele

MC-EMSA analysis. A) Nuclear proteins from the Huh7 cell line incubated with seven cocktails of unlabeled DNA competitors in the presence of rs57217136 C biotinylated labeled probes. The rs57217136 T allele specific bands 1 and 3 were eliminated by 'cocktail 4', while band 2 was non-specific. B) EMSA of 10 individual competitors of 'cocktail 4'. The specific band 1 was eliminated by competitor #40, while Band 3 competed with protein #48. The red arrows show C allele specific bands, and the yellow arrows indicate positive competitors.



5.2.6. Protein-protein network identification

Gene transcription involves complex machinery; several proteins network together to create a transcription complex that interacts with the gene promoter and initiates gene transcription. The functional SNPs rs6511720 and rs57217136 are predicted to act as enhancer and promoter activator elements respectively, thus co-operate with proteins to regulate genes. For better understanding of how such machinery works, the STRING protein network software provides a visual summary of proteins that may interact with a protein of interest. STRING is a useful tool to identify which kind of regulatory proteins cooperate with known proteins that regulate the gene.

Figure 5-13 shows the protein network of identified allele-specific proteins. There are three key proteins identified: SRE, STAT1, and RAR. SRE has a short recognition sequence commonly found in the 5' of the site of transcription initiation of many genes, which is bound to serum response factor (SRF) in order to regulate gene transcription (396). Figure 5-13 A shows that SRF interacts with some transcription regulatory factors such as myocardin (MYOCD), GATA binding protein 4 (GATA4), and SMAD family member 2 (SMAD2).

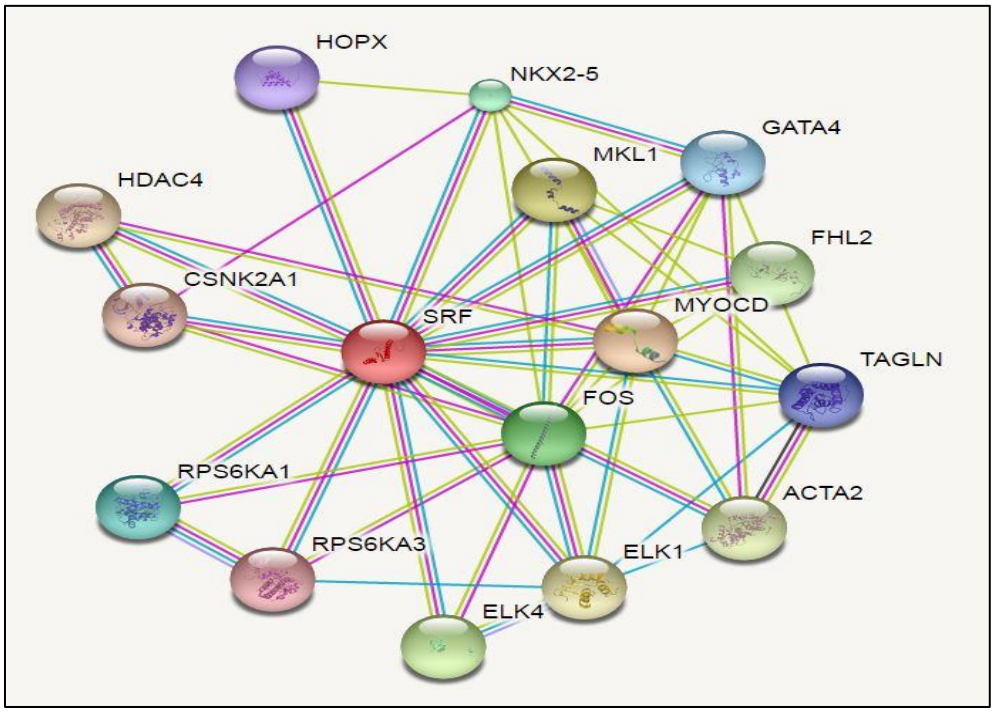
The second key protein is STAT1, which interacts with transcription regulator and co-regulator proteins (Figure 5-13 B) such as STAT2, STAT3, protein inhibitor of activated STAT1 (PIAS1) and interferon regulatory factor 1(IRF1). STAT1 also interacts with histone regulatory proteins CREBBP (CBP/p300), E1A binding protein p300 (EP300), and proteins linked to lipid metabolism – Janus kinase 2 (JAK1) and epidermal growth factor receptor (EGFR).

The third key protein is RAR, RARA isoform is able to interact with a wide variety of proteins (see Figure 5-13 C) that have a role in gene transcription regulation including nuclear receptor co-repressors (NCOR1and NCOR2), nuclear receptor co-activators (NCOA1, NCOA2, and NCOA2), nuclear receptor interacting protein 1 (NRP1), mediator complex subunit (MED1), histone regulatory proteins (EP300, HDAC, and STAT3), and retinoid X receptor gamma (RXRG) which links to lipid homeostasis.

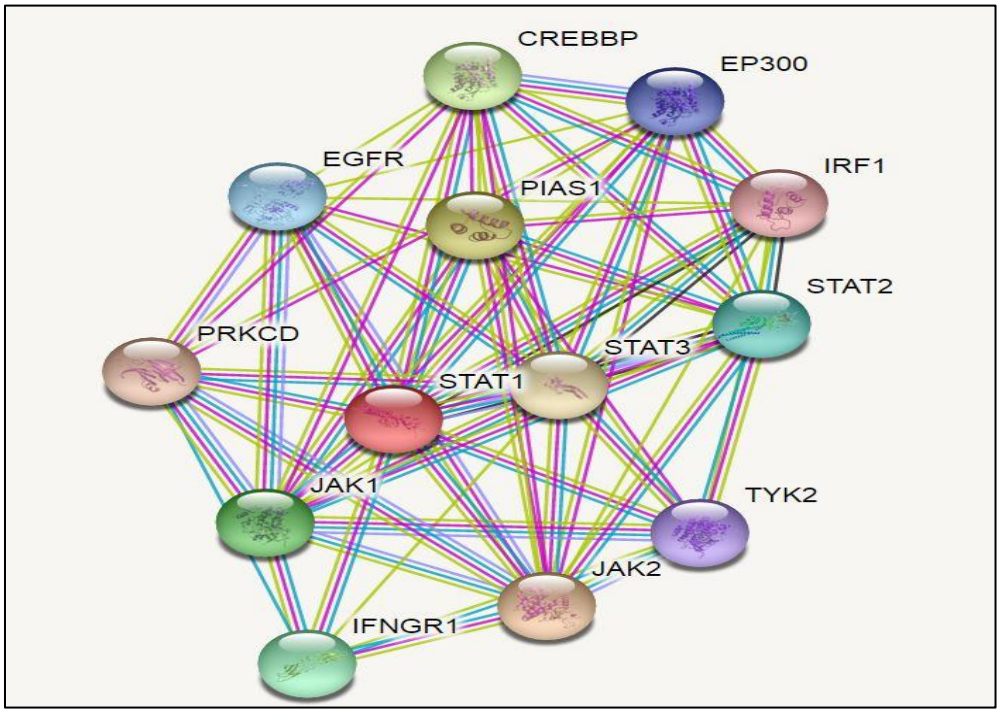
Figure 5-13: the *LDLR* intron-1 functional SNPs: allele-specific protein network

These figures illustrate the rs6511720 and rs57217136 minor allele-specific protein network, where rs6511720 binds to SRE, which is a DNA sequence element for SRF, and rs57217136 binds to STAT1 and RARA. A) The protein network of SRE. B) The protein network of STAT1. C) The protein network of RARA.

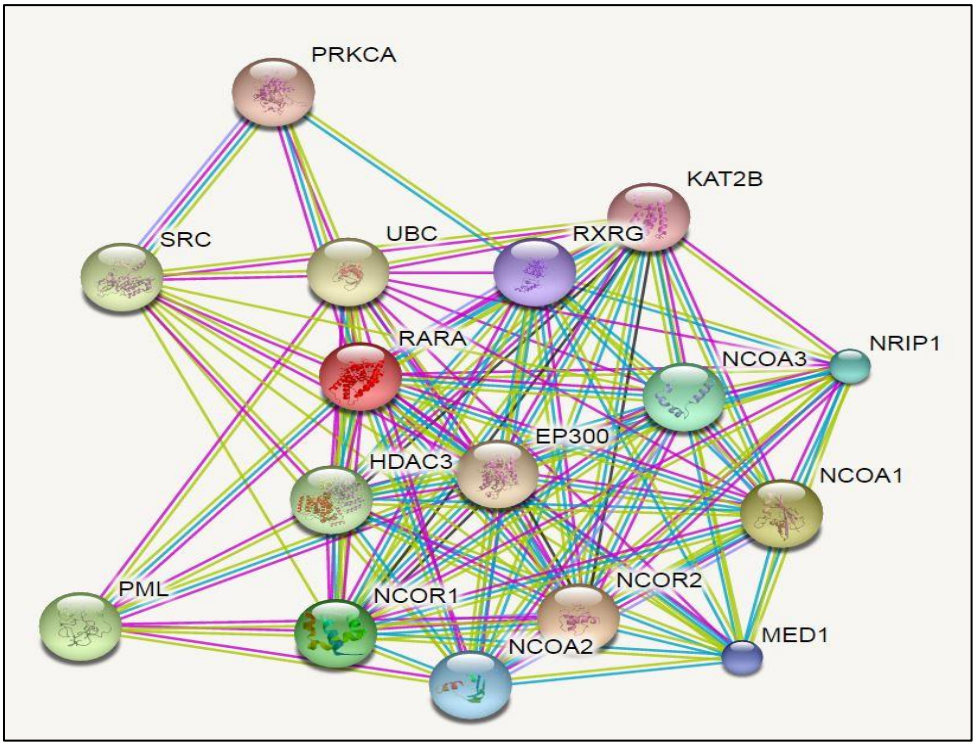
A) SRE protein network



B) STAT1 protein network



C) RARA protein network



5.3. Discussion

LDL-R is the primary protein involved in clearance of cholesterol from the blood, and mutations or variants in the gene lead to FH (39, 40). Studies also demonstrate that the co-inheritance of many trait-associated SNPs can explain the disease phenotype. GWAS has reported 22 SNPs associated with LDL-C. The *LDLR* SNP rs6511720 is considered a protective SNP. Several studies using GWAS data have reported that the minor allele of the SNP is associated with lower plasma levels of LDL-C and a lower risk of CHD (160, 374, 376). I have confirmed this effect in the UCLEB consortium data, which demonstrated that individuals carrying one or more copies of the T allele of rs6511720 have lower LDL-C levels and have a lower lifetime risk of CHD. Furthermore, data from the GLGC consortium confirmed that the *LDLR* rs6511720 minor allele (T), which is carried by approximately 10% of the population, is strongly associated with a protective effect, being associated with lower levels of LDL-C. This association is thus consistent and of sufficient strength to suggest that common variation in *LDLR* has implications for health, and that determining the precise molecular mechanism of the effect is relevant.

The GWAS hit SNP rs6511720 is located in intron-1 where the cis-acting gene regulating sites may commonly be found (378). Assessing SNP functionality is complex due to the LD between genetic loci which creates a noise background around functional loci, because all SNPs in LD with functional loci may carry some or all of the association with the trait of interest, although they have no relevant function. The rs6511720 has strong LD with 48 SNPs, but which of these SNPs has the highest likelihood of being functional? To prioritize and select candidate variants for functional study, several bioinformatics databases, which are useful resources to identify variants differentially affecting transcriptional activity, were used. It was found that four

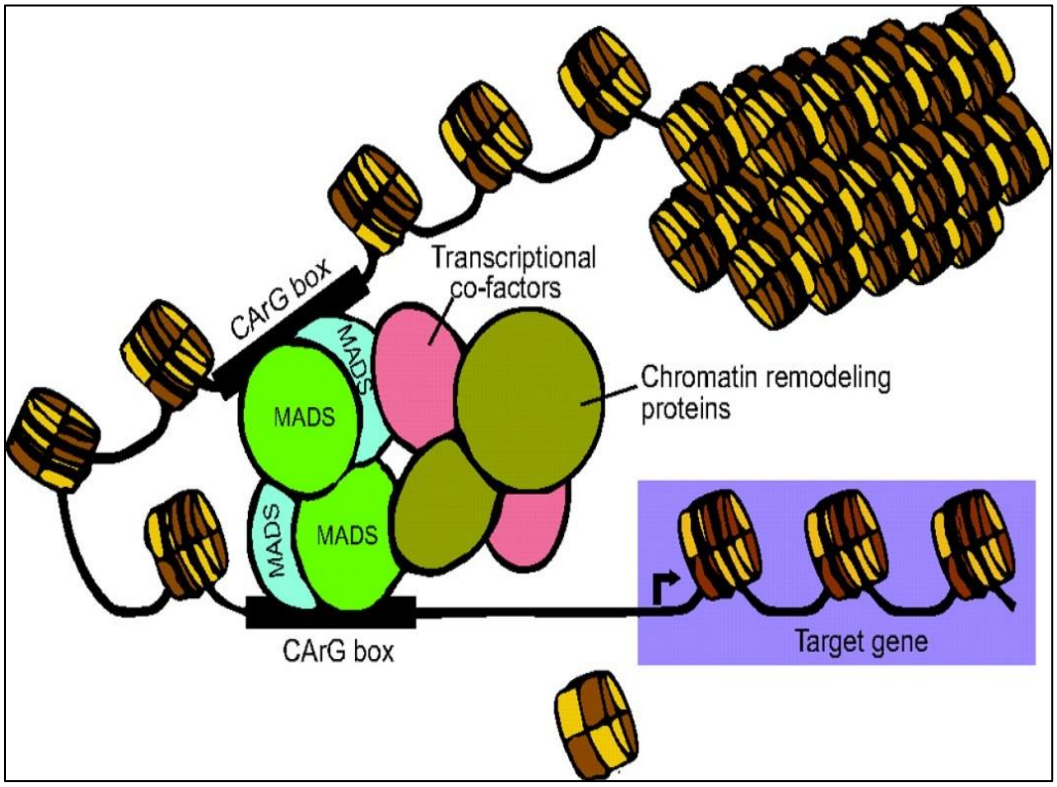
promoter-proximal intron-1 variants (rs6511720, rs57217136, rs141787760 rs60173709) had strong chromatin signals in liver cells: these SNPs are in complete LD (r^2=1) with the GWAS 'hit' SNP. Although bioinformatics tools were useful to nominate potential functional SNPs, MatInspector and HaploReg showed different DNA-protein binding profiles of these SNPs. The inconsistent prediction between bioinformatics tools prevented finding an interesting TF to test, however, the data shed light on candidate SNPs that have a strong regulatory profile, and which were worthy of further investigation using functional assays.

EMSAs confirmed the bioinformatics finding, where the four SNPs all showed allele-specific protein binding. This finding suggests that these proteins may have an influence on regulating *LDLR* expression. To determine whether the different genotypes have an influence on *LDLR* expression, gene reporter assays were used. Data from the luciferase reporter assays was consistent with *LDLR* intron-1 SNPs affecting *LDLR* gene expression, as the minor alleles of rs6511720 and rs57217136 showed significantly higher expression, while rs141787760 and rs60173709 showed significantly lower expression but there was no difference between the alleles. It is important to consider the combined effect on expression of these four SNPs because all these four rare alleles are always present together. It is likely that the lowering expression effect of the minor alleles of rs141787760 and rs60173709 partially repress the elevating effect of the rare alleles of rs6511720 and rs57217136 SNPs. If the combined effect of the minor alleles of four *LDLR* intron-1 SNPs was estimated by summation, we would predict that the haplotype will show ~29% higher expression. This would lead to up-regulation of LDL-R numbers on the surface of hepatocytes and this in turn would explain the ≈10% lower levels of LDL-C and ≈12% lower risk of CHD in individuals carrying the rs6511720 minor allele haplotype. The actual

mechanism of this effect is only partially unraveled by this work. The data suggest that both minor alleles of SNPs rs6511720 and rs57217136 are likely to create enhancer-binding sites for transcription factors that would contribute to increased *LDLR* expression. MC-EMSAs showed the sequence around rs6511720 minor allele (T) contains the SRE, which has a CArG box sequence that is a target binding site for the SRF and the SRE-SRF controls gene expression of many genes (397). It has been reported that SRE has a role in *LDLR* gene expression stimulation in HepG2 cells through sterol-independent pathways (398). The SRF is a member of the MCM1, Agamous, Deficiens, and SRF (MADS) box superfamily of transcription factors and, SRF has been implicated in the expression of many genes through its association with a diverse set of co-activators in different tissues (399) in order to mediate transcriptional regulation of the target gene (see Figure 4-13). In addition, SRF could co-operate with chromatin remodeling proteins, which relax the chromatin structure at the target gene transcription start site and/or enhancer sites allowing for the initiation of gene transcription (400). It has been reported that there are three sets of SRF binding proteins that are involved in gene regulation of muscle and cardiac specific genes: the ternary complex factor (TCF) family proteins that include Elk1, SAP1 and SAP2 (401, 402), the myocardin-related MKL family proteins that include Megakaryoblastic leukemia-1/2 (MKL1 and MKL2) (399, 403, 404), Nkx2-5 and GATA4 (405). Also, it has been reported that the SRF regulates expression of several genes with roles in hepatic function, but their molecular mechanisms are unclear (406).

Figure 5-14: Model for the action of MADS-domain protein complexes.

The figure shows a model of MADS-domain protein complex formation and its gene regulation mechanism. MADS-domain proteins (green and blue) interact with two DNA binding sites (CArG boxes; black) in close proximity, resulting in DNA looping. Subsequently, MADS-domain proteins recruit transcriptional co-factors (pink) and chromatin remodeling proteins (brown), which in the complex mediate transcriptional regulation. Adapted from Smaczniak et al. (2012).

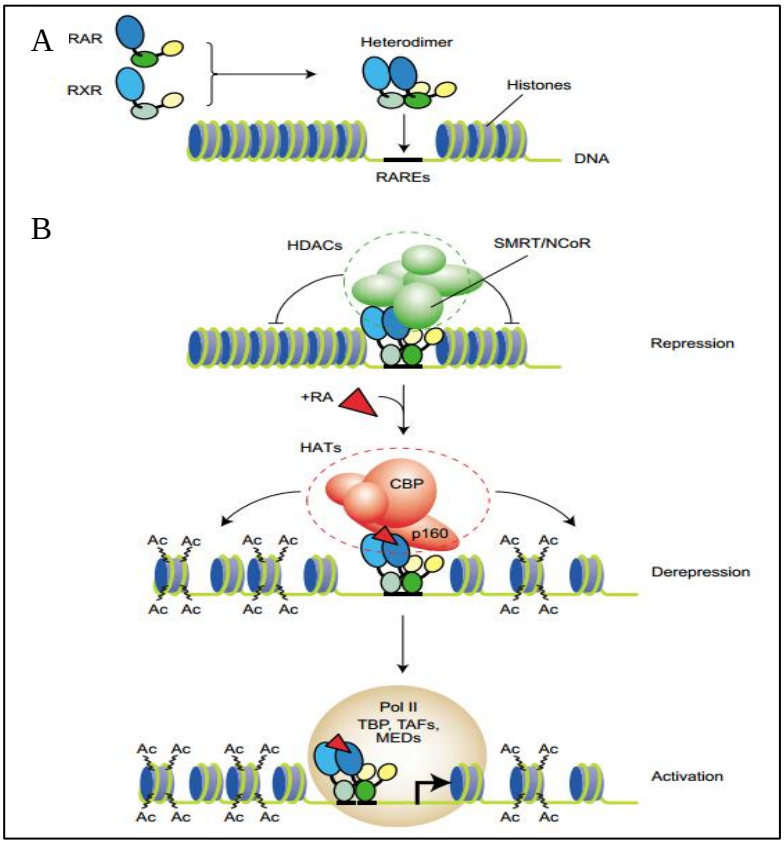


The data also suggest that rs57217136 is bound to RAR and STAT1. RAR is a member of a family of nuclear receptor proteins actively involved in retinoic acid-mediated transcriptional regulation of genes that control lipid metabolism, through dimerization with other proteins to initiate transcription (407). RAR proteins are not active in their monomeric form, but they can be dimerized with another family of nuclear receptor proteins called retinoid X receptors (RXR). The RAR≡RXR dimer binds to hormone response elements known as retinoic acid response elements (RAREs) to create an apo-RAR≡RXR heterodimer complex. Transcriptional activation of genes occurs when the ligands, including trans retinoic acid or 9-cis retinoic acid (RA) bind to the ligand binding sites on RAR≡RXR (408), however, gene expression is repressed in the absence of the ligand. In the absence of the ligands, the apo-RAR≡RXR heterodimer recruits co-repressors, the nuclear receptor co-repressor (NCoR) and the silencing mediator for retinoid and thyroid hormone receptor (SMRT), that inhibit transcription by binding to the ligand-binding domains of RAR≡RXR. Then, the co-repressor recruits the heterodimer to HDACs complex containing Sin3, which removes the acetyl groups from nucleosomal histones, leading to chromatin condensation and gene repression or silencing (409-411). . However, in the presence of an RAR≡RXR ligand (e.g. RA), the ligand is able to recruit histone acetyltransferase (HAT) co-activators, such as CBP, p300 and p160 (412) (see Figure 5-15). Therefore, the chromatin relaxes due to histone acetylation and is able to interact with the basal transcription apparatus: RNA polymerase II holoenzyme, together with the TATA-binding protein (TBP) and TBP-associated factors (TAFs), and mediator complexes (MEDs), to initiate gene transcription (409). Also it has been reported that the apo-RAR≡RXR heterodimer recruits co-activators NCoA/ SRC-1 which in turn recruit other co-activator factors such as CBP/p300, and TIF-1 to P/CIP to activate gene transcription (413). Binding of RAR to the sequence around the rs57217136 T

allele may initiate such pathways and facilitate an open chromatin structure surrounding intron-1, which allows further access to DNA-binding protein that facilitates gene transcription in a complex.

Figure 5-15: Mode of action of RAR≡RXR heterodimers

This figure shows the action mode of RAR≡RXR heterodimers. (A) Retinoic acid receptors (RARs) and retinoid X receptors (RXRs) form heterodimers that bind within the regulatory region of target genes through the retinoic acid response elements (RAREs). (B) In the absence of a ligand (e.g. RA), RAR–RXR heterodimers (apo-heterodimers) are bound to RAREs of target genes together with transcriptional co-repressors (N-CoR or SMRT), which then recruit HDAC complex and repress gene expression through chromatin condensation. However, binding of the ligand induces the release of the HDAC complex and results in the recruitment of histone acetyltransferase (HAT) co-activators, such as CBP and p160. The subsequent chromatin relaxing is due to this RAR tone acetylation ('Ac'). In the last step, the RNA polymerase II holoenzyme, together with the TATA-binding protein (TBP) and TBP-associated factors (TAFs), and mediator complexes (MEDs) are recruited, which initiates the transcription machinery. Adapted from Clarke et al. (2004).



The rs57217136 T allele also binds to STAT1, a transcription factor that is involved in regulating expression (414) and regulating lipid-related genes. The transcriptional activity role of STAT1 is dependent on biological conditions, where it drives the expression of many genes, but also suppresses the transcription of others (415). Studies have shown that phosphorylated STAT1 is inter-nucleus and binds to a specific sequence of DNA and regulates gene expression (416, 417). It has been reported that STAT1 and STAT3 are involved in lipid metabolism via JAK/STATs through several pathways. JAK/STAT1 proteins, for instance, bind to the promoter of peroxisome proliferator-activated receptor-gamma 2 (*PPAR-γ2*) and *LPL*, which leads to repression of *PPAR-γ2* and *LPL* expression, which promotes adipogenesis and inhibits lipolysis, respectively, in mature fat cells (418-420).

On the other hand, studies have shown that the positive function of STAT1 in transcriptional regulation is due to its binding to transcriptional co-activators such as CERB complex (CBP/p300), pCIP, MCM5, Nmi, and BRCA1 (421-424). CBP/p300 is histone acetyl transferase, which relaxes the chromatin near to transcription start sites or enhancer sites to provide a platform for several associated proteins (423). pCIP interacts with CBP/p300 to activate STAT1, while MCM5 and BRCA1 co-operate with STAT1 for maximal transcription activity (422). Another interesting co-activator transcription factor that co-operates with STAT1 is SP1, where the ENCODE project consortium (2011) ChIP-Seq data showed that SP1 binds to the rs57217136 minor allele (276). SP1 is an *LDLR* associated protein, which binds to repeats 1 and 3 in *LDLR* promoter and initiates gene transcription (60).

5.4. Limitations

This study has a number of limitations. This study suggests that a cis-regulatory element near rs6511720 and rs57217136 SNPs acts in the liver cell line. However, I did not examine other LD distal SNPs that may have a role in transcription regulation and, therefore, further studies are needed in this area. Also, since I studied the effect of these SNPs only in liver cells I cannot determine whether or not they also influence *LDLR* gene expression in other tissues. However, since the major site of expression of *LDLR* is the liver, where clearance of LDL-C from the plasma occurs, this is not a major limitation.

Another limitation in this study was that I could not design one luciferase construct that contains all SNPs of interest because of the presence of Alu repeats in the region of interest, and therefore I could only estimate the likely combined effect *in vitro* of the effects of all SNPs acting in concert.

In this study, luciferase reporter assays and EMSAs were used, which are techniques that measure the difference of allelic expression and determine DNA-protein interaction *in vitro.* As such, both luciferase reporter assays and EMSA are *in vitro* techniques which cannot account for regulatory landscape elements and gene spatial organization. Therefore, more studies are needed to define the spatial organization of the gene, which has a fundamental role in controlling gene expression (425-428), by the chromatin looping that brings enhancers and promoters into close spatial proximity to interact and initiate transcription. In recent years a number of chromosome conformation capture methods have been developed such as chromosome conformation capture (3C) (429, 430) and chromosome conformation capture carbon copy (5C) (431) in order to study DNA interactions in native states.

The concepts of 3C are to study whether two distant genomic sequences physically interact (e.g. promoter – enhancer) by the looping interaction. The DNA binding proteins are formaldehyde cross-linked to the DNA followed by the use of appropriate restriction enzymes to solubilize the cross-linked DNA. The cross-linked DNA is then ligated creating "ligation junctions". The cross-links are removed, the DNA purified and selected ligation products quantified using PCR with primers specific to the genomic loci being studied (429, 430). Also, it is possible to use the 5C method, which is a similar protocol to 3C, except that the 3C library converts into a 5C library by annealing and ligating 5C oligonucleotides in a multiplex setting, in order to use high-throughput detection methods including microarrays and quantitative DNA sequencing (431). In the case of *LDLR*, the proximal enhancer sequence of rs6511720 and rs57217136 is cross-linked to the *LDLR* promoter.

5.5. Conclusion

In conclusion, this study provides a partial explanation of the reason behind the protective effect of GWAS *LDLR* SNP rs6511720 minor allele on CHD. Integration of bioinformatics with GWAS disease-associated variants helps to elucidate gene-regulatory variants underlying association signals. Both rs6511720 and rs57217136 were identified as part of a cis-regulatory complex in a liver cell line that altered transcriptional activity and thus may increase *LDLR* expression, and consequently reduce LDL-C levels. *In vitro* data suggests that the rs57217136 binds to RARA and STAT1, which are both transcription factors requiring different histone acetyltransferase activities to activate transcription, which may interact with rs6511720 binding protein (SRE-SRF) and create a complex 'enhancer model' that co-operates with the basal transcription apparatus to maximize the enhancing activity of the core

transcription machinery. For future work, the use of chromosome conformation capture methods may be helpful in order to understand the role of spatial organization of the identified regulatory elements in gene regulation.

Chapter 6 : Identification of an LDL-C associated variant in the coding region of the *ANXA2* gene

6.1. Introduction

Hypercholesterolemia is a major risk factor for atherosclerosis and CHD, most often caused by an individual having a greater than average number of common lipid-raising SNPs. The Global Lipids Genetics consortium (2013) has identified 157 novel loci associated with lipid levels, 15 of which are known to influence plasma levels of low density lipoprotein cholesterol (LDL-C) (161). By contrast, the monogenic disorder, FH is known to be caused by a genetic defect at one of three loci that code for proteins involved in LDL-C clearance: *LDLR*, *APOB*, and *PCSK9*. These three genes are GWAS hits for LDL-C.

PCSK9 plays an important role in cholesterol clearance from the blood. It is highly expressed in the liver (93), and binds to the epidermal growth factor domain A (EGF-A) of the LDL-R via its catalytic domain either intracellularly or at the cell surface (432). Once the PCSK9≡LDL-R complex is formed, it is internalized by endocytosis and degraded (103, 104). Gain-of-function mutations in *PCSK9* strongly promote LDL-R degradation and lead to FH, whereas loss-of-function mutations of *PCSK9* are unable to enhance LDL-R down regulation and therefore result in lower levels of LDL-C (128). It has been proposed that PCSK9 is a viable therapeutic target to reduce LDL-C and protect against atherosclerosis and CAD (166, 167). Several monoclonal antibodies have been developed and showed significant efficacy in reducing LDL-C levels with minimal side-effects (234-236). In the last few years, a new protein called AnxA2 has been identified in animal and cellular models as an endogenous inhibitor of PCSK9, and thus influences LDL-R and plasma cholesterol levels (92, 101, 102).

AnxA2 is highly expressed in the lungs, pancreas, colon, ileum and adrenal tissues, but not in the spleen, testis, kidney and liver (102). AnxA2 belongs to the conserved annexin family of phospholipid and calcium-binding proteins. AnxA2 exists as a monomer, yet the majority of AnxA2 forms a heterotetramer with the S100 protein p11 (S100A10) both in intra- and extracellular locations (188, 433). Intracellularly, AnxA2 regulates a spectrum of functions related to membrane organization and trafficking (172, 188, 434). Extracellular, AnxA2 has several activities (435) including involvement in cholesterol clearance metabolism either as a monomer or complexed with p11 via the binding of its R1-domain to an CHRD of PCSK9 at the cell surface, which inhibits PCSK9-mediated degradation of LDL-R. This helps to maintain LDL-R levels at the cell surface with the subsequent greater clearance of LDL-C (101, 102). An *in vitro* study reported that a mutation Q554E in the CHRD of PCSK9 increased the binding affinity between PCSK9 and AnxA2, which in turn led to a loss-of-function of PCSK9 towards LDL-R degradation (101). Also, *in vivo* studies showed that AnxA2 knockout mice had higher levels of plasma PCSK9 and LDL-C, which correlated with a reduction in LDL-R protein levels, mostly in extrahepatic tissues (102). Moreover, adenoviral AnxA2 overexpression in mouse liver was significantly associated with higher hepatic LDL-R levels (102).

Molecular network and protein interaction are responsible for cell function, thus a mutation in any gene involved in a diseased molecular network may lead to the observed phenotype of the "Guilt by Association" (GBA) mechanism (436-438). It was hypothesized that a mutation in the *ANXA2*-R1 domain could also affect LDL-C levels. The *ANXA2* locus is located on chromosome 15q22.2 and consists of 13 exons (170), and

its expression is regulated at both the transcriptional and translational levels (188). In this study, in the *ANXA2*-R1 domain, a common coding SNP (rs17845226), which changes Valine to Leucine at position 98, was selected to determine whether the SNP influences LDL-C levels. In a preliminary study including only 43 subjects, this SNP was implicated in the regulation of circulating PCSK9 levels (102), but a thorough analysis of its association with LDL-C and CHD in larger cohorts has not yet been performed. In this chapter I use bioinformatics and genotype-phenotype association analyses to determine whether the SNP is associated with differences in LDL-C levels and probably Anex2 function. I show that this common coding SNP appears unlikely to affect the protein's functional structure, but it could be a marker for other proxy variants at the locus.

6.2. Results

6.2.1. Bioinformatics

The R1-domain of *ANXA2* is encoded by Exons 4-6, which has eight reported SNPs (Table 6-1). The missense variant, rs17845226, was selected for further study as it has been validated by HapMap and the 1000 Genomes project, and is the only SNP that has a minor allele frequency (MAF) ≥ 0.05. The SNP is located in exon 6 of the gene and causes a valine/leucine amino acid change; the MAF is 12% in European populations (1000 Genomes Project Phase 3)]. Although the affected amino acid (valine) is highly conserved in vertebrate species (Figure 6-1); the change is predicted to be non-pathogenic by the SIFT, Polyphen-2, and Mutation Taster predicting tools (Table 6-2). Interestingly, rs17845226 shows modest LD (r2 >= 0.4) with 34 SNPs, all located downstream of the *ANXA2* gene-coding region in the long intergenic region (Figure 6.2) (439)

Table 6-1: The R1-domain of *ANXA2* reported SNPs by 1000 Genomes Project

Exon	SNP	Ref/Alt	Amino acid Position	MAF	Validated*
Exon 4	rs41315968	C/T	p.(Lys28 Lys)	NA	No
	rs11553791	A/G	p.(Phe33 Phe)	NA	No
	rs75993598	C/T	p.(Glu43 Lys)	NA	Yes
Exon 5	rs11553794	G/A	p.(Arg63Ser)	< 0.01	Yes
Exon 6	rs17845226	C/A	p.(Val98Leu)	0.05	Yes
	rs17854749	A/G	P.(Leu102Leu)	NA	No
	rs144035126	G/A	p.(Thr97Met)	< 0.01	Yes
	rs147297902	G/A	p.(Ala90Val)	<0.01	Yes

*- The variant was discovered in the1000 Genomes project, and the Exome Sequencing project

Table 6-2: Functional prediction of ANXA2 coding variant p.(Val98Leu) rs17845226 by different prediction tools

Gene_ species_ variant	SIFT	Polyphen-2	Mutation Taster
ANXA2_HUMAN V98L	neutral	neutral	neutral

Figure 6-1: The *ANXA2* exon 6 amino acid sequence in different species.

Schematic presentation of the *ANXA2* exon 6 amino acid sequence in different species (https://genome-euro.ucsc.edu). The *ANXA2* coding SNP rs17845226 p.(Val98Leu) is highly conserved across species.

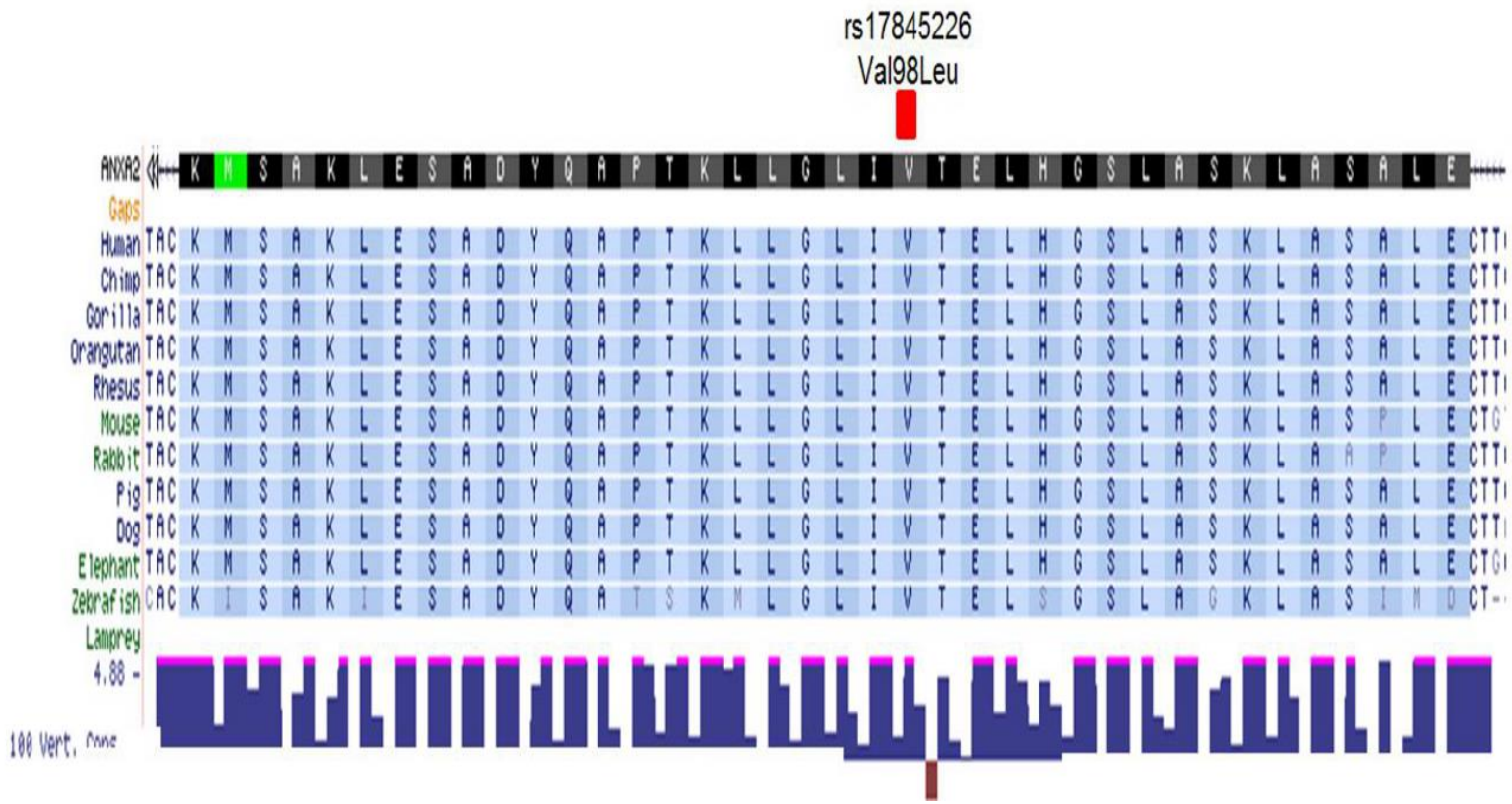


Figure 6-2 *ANXA2* coding SNP rs17845226 LD SNP table

The figure shows modest LD r2≥0.4 SNPs for query SNP rs17845226. The summary detail for each SNP is given including chromosome number (chr), variant piston according to human genome 38 (pos(hg38)), LD r2 score and LD D' score*, variant ID, reference allele (Ref) and alternative allele (Alt), and the minor allele frequency in four populations including African, American, Asian and European. http://archive.broadinstitute.org/mammals/haploreg/haploreg.php

Query SNP: rs17845226 and variants with r^2 >= 0.4

chr	pos (hg38)	LD (r^2)	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq
15	60268173	0.4	0.74	rs72742094	A	G	0.02	0.10	0.00	0.16
15	60272427	0.41	0.77	rs6494184	T	C	0.07	0.10	0.08	0.17
15	60274767	0.41	0.77	rs8029328	C	G	0.21	0.11	0.08	0.17
15	60279505	0.41	0.77	rs7166569	G	C	0.08	0.10	0.08	0.17
15	60280106	0.42	0.78	rs72744815	C	G	0.02	0.10	0.08	0.17
15	60280124	0.42	0.78	rs72744816	G	A	0.02	0.10	0.08	0.17
15	60284063	0.43	0.79	rs17191323	G	A	0.03	0.10	0.07	0.17
15	60284387	0.43	0.79	rs142261085	ATCCT	A	0.03	0.10	0.07	0.17
15	60286628	0.43	0.79	rs72744820	G	A	0.03	0.10	0.07	0.17
15	60287699	0.43	0.79	rs142661682	G	C	0.02	0.10	0.07	0.17
15	60288382	0.45	0.78	rs56140011	G	A	0.02	0.10	0.00	0.16
15	60292850	0.44	0.8	rs1396949	G	A	0.03	0.10	0.07	0.17
15	60293005	0.44	0.8	rs7171001	A	G	0.06	0.11	0.07	0.17
15	60293061	0.44	0.8	rs10519006	C	T	0.05	0.10	0.07	0.17
15	60295115	0.41	0.79	rs12900101	C	G	0.11	0.12	0.12	0.17
15	60295810	0.42	0.8	rs116255697	A	G	0.11	0.12	0.12	0.18
15	60296709	0.42	0.8	rs6494187	G	C	0.27	0.13	0.12	0.18
15	60298581	0.4	0.78	rs11633032	G	A	0.11	0.12	0.11	0.17
15	60299110	0.42	0.8	rs1509798	G	A	0.11	0.12	0.11	0.18
15	60299456	0.46	0.81	rs72744839	T	C	0.02	0.10	0.06	0.16
15	60299923	0.45	0.8	rs17191344	A	G	0.02	0.10	0.06	0.16
15	60300236	0.43	0.8	rs10152183	T	C	0.12	0.12	0.11	0.17
15	60302572	0.41	0.77	rs57027904	G	A	0.12	0.11	0.11	0.16
15	60303436	0.47	0.82	rs11629902	A	C	0.02	0.10	0.06	0.16
15	60303973	0.44	0.82	rs66871125	G	C	0.12	0.12	0.11	0.17
15	60304173	0.42	0.78	rs68129504	T	A	0.17	0.12	0.11	0.17
15	60306912	0.41	0.78	rs60322655	A	G	0.17	0.12	0.11	0.17
15	60307019	0.41	0.78	rs57881710	T	C	0.12	0.12	0.11	0.17
15	60309553	0.41	0.78	rs113335202	A	C	0.12	0.12	0.11	0.17
15	60310360	0.41	0.78	rs67379639	G	T	0.12	0.12	0.11	0.17
15	60311213	0.44	0.82	rs67701283	T	G	0.17	0.12	0.11	0.17
15	60315590	0.42	0.76	rs16942463	G	C	0.05	0.10	0.03	0.16
15	60316183	0.42	0.69	rs11629692	A	G	0.01	0.09	0.00	0.13
15	60317094	0.55	0.76	rs61381915	C	T	0.06	0.10	0.00	0.13
15	60361006	1	1	rs17845226	C	A	0.01	0.10	0.00	0.12

*LD *D'* is the ratio of *D* to its maximum possible absolute value, given the allele frequencies, where *D* is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (pAB) and the product of the frequencies of those alleles (pA and pB) [DAB = pAB – pApB]. LD r^2 is the square of the correlation coefficient between two indicator variables (A and B**)** [r2 = D2/ pA (1 – pA) pB (1 – pB] (440). In literature r^2 is mostly common used.

6.2.2. Sequencing

ANXA2, located at 15q22.2, has three pseudogene sequences found on chromosomes 4 (*ANXA2P1)*, 9 (*ANXA2P2*) and 10 (*ANXA2P3*). There is a possibility that the off-the-shelf genotyping commercial primers may have amplified one of these pseudogene sequences instead of the target sequence on chromosome 15. Therefore, unique primers were designed to amplify a fragment containing rs17845226 on chromosome 15. Random samples from each genotype were sequenced and the sequencing result compared to the genotype results to verify genotyping specificity.

Figure 6-3A shows the reference sequence around *ANAX2* SNP rs17845226 in chromosome 15 and the three pseudogenes on chromosomes 4, 9 and 10 aligned to the reference sequence. The result shows that the reference sequence and the other three pseudogenes share some part of the sequence, however, the sequence around the SNP of interest is different. The sequencing results show 100% alignment between the *ANXA2* chromosome-15 reference sequence and sample sequences (Figure 6-3 B). Sanger sequencing verified the TaqMan genotyping results which proved the primer's specificity.

Figure 6-3: *ANXA2* pseudogene sequences

A) Using CLUSTALW multiple sequence alignment tool (http://www.genome.jp/tools/clustalw/) the reference sequence around *ANAX2* SNP rs17845226 on chromosome 15 was aligned to the three pseudogenes on chromosomes 4, 9 and 10. (B) TaqMan sequence results of a random sample selection - an individual whose genotype is homozygous for the common allele (C allele); blue highlighted nucleotide is the *ANXA2* SNP rs17845226 on both strands.

A)

rs17845226

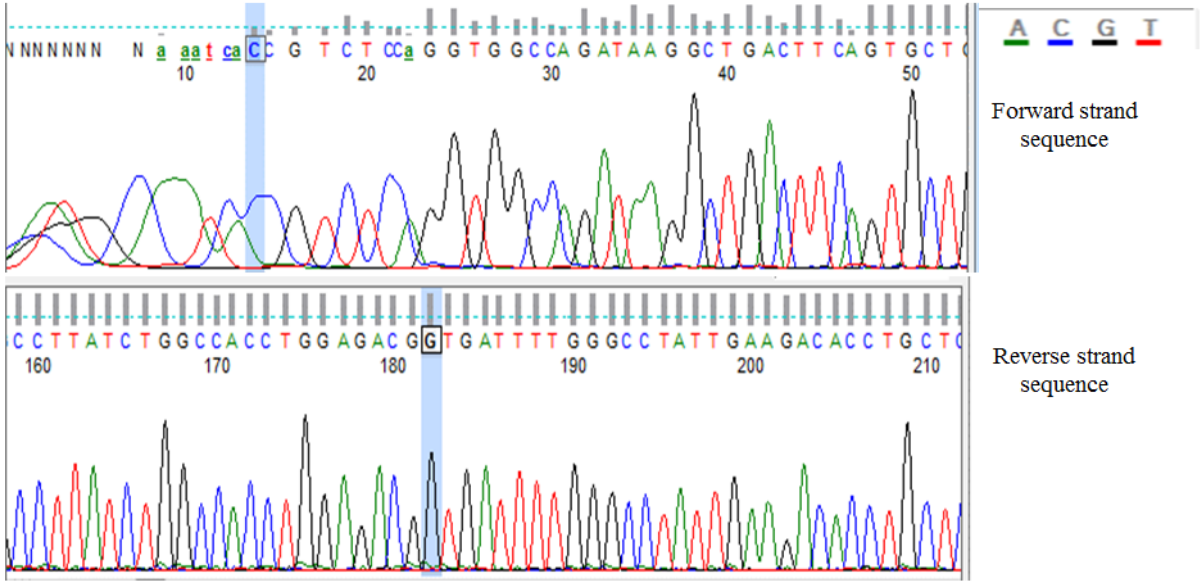
ANXA2 (Ch15) AATAGGCCCAAAATCACCGTCTCCAGGTGGCCAGATAAGGCTGACTTCAGTGCTGATGCA

ANXA2 (Ch10) AACA-------TGAAAGGACCATC----AGGACAAAAGGTGTGGGTAAGTTCACCATTGTCAAC

ANXA2 (Ch9) AGCTGAGATCATACCATTGCACTCCAGGCAGCTGGGCGA---CAGGGTGAAACTCTGTCTCAAA

ANXA2 (Ch4) G AGGGGCATAAAATAATATATTTTCAGCTCAGCATCTGGGGACGCTCTCAGGTCTTGGTGC

B)



6.2.3. Genotype-phenotype association

The association of rs17845266 with LDL-C and CHD was investigated in the NPHSII and WHII cohorts. NPHSII consists of 3,052 healthy middle-aged men (50–61 years) who were recruited from nine general medical practices in the UK (299), while WHII consists of 3,413 women and 6,895 men who were recruited from 20 London-based civil service departments in the UK; both studies' participants were followed for 15 years. Baseline study characteristics are summarized in Table 6-4. TaqMan genotyping of rs17845266 was successful with a call rate of ≥90%. The minor allele frequency of rs17845266 in NPHSII was 0.13 which is similar to that seen in European populations ($\chi 2$=1.722, p=0.19), while in the WHII the SNP call rate was 99%, and the minor allele frequency was 0.138 ($\chi 2$=2.585, p=0.11).

Table 6-3: The Second-Northwick-Park Heart Study (NPHSII) and Whitehall II (WHII) cohort baseline characteristics. Values shown are mean (±SD)

	NPHSII	WHII
Sample size	2463	4414
Age (years)	56.1 (3.4)	49.1 (6.0)
Male (%)	2463 (100%)	3344 (75.7%)
Current or ex-Smoker (%)	1694 (68.8)	518 (12.3%)
Body mass index (kg/m^2)	26.5 (3.5)	25.1 (3.6)
Type 2 diabetes (%)	-	81 (2.0)
Systolic blood pressure (mmHg)	138.7 (19.2)	120.6 (13.1)
Diastolic blood pressure (mmHg)	84.6 (11.2)	79.7 (9.2)
Total Cholesterol (mmol/l)	5.75 (1.01)	6.45 (1.13)
LDL-Cholesterol (mmol/l)	3.10 (1.00)	4.38 (1.01)
HDL-Cholesterol (mmol/l)	0.80 (0.24)	1.42 (0.41)
Triglycerides (mmol/l)[#]	1.80 (0.95)	1.23 (0.68)

#: Geometric mean (approximate SD)

6.2.3.1. Association of *ANXA2* SNP rs17845226 with lipid traits and CHD in NPHSII

The *ANXA2*-R1 domain SNP rs17845226 (G>A) showed an association with LDL-C levels and CHD risk under both inheritance modes (additive and recessive). Table 6-5 shows that individuals who are homozygous for the minor allele (A) have significantly higher levels of total cholesterol (TC) ≈8.4% and LDL-C ≈18.8% (p=0.007 and 0.003, respectively), and have a twofold higher risk of CHD (HR (95% CI): 2.17 (1.03 - 4.60), p=0.04)). The most likely cause of the CHD association observed is that it is due to these subjects having a higher level of LDL-C, since when the CHD association was adjusted for LDL-C the effect was reduced to 1.25 (0.49-3.20) and was no longer statistically significant (p=0.60). This suggests that the main mechanism of this SNP influencing the risk of CHD is via its effect on LDL-C levels.

Table 6-4: An association between *ANXA2* rs17845226 (Val98Leu) genotype and lipid risk factors and coronary heart disease risk in The Second-Northwick-Park Heart Study (NPHS-II).

Genotype	CC (n=1868)	CA (n=561)	AA (n=34)	P-value §	P-value*
TC (mmol/l)	5.73 (1.00)	5.80 (1.04)	6.21 (1.05)	0.014	0.007
LDL-C (mmol/l)	3.08 (0.99)	3.16 (1.04)	3.66 (0.94)	0.006	0.003
HDL-C (mmol/l)	0.81 (0.24)	0.80 (0.26)	0.74 (0.23)	0.456	0.201
TG (mmol/l)#	1.80 (0.95)	1.79 (0.96)	1.88 (0.93))	0.970	0.648
CHD risk by genotype					
No. without CHD	1672 (75.8)	507 (23.0)	26 (1.2)	-	-
No. with CHD	196 (76.0)	54 (20.9)	8 (3.1)	-	-
HR (95% CI) Model 1	1.00	0.93 (0.68-1.26)	2.17 (1.03-4.60)	0.113	0.04
HR (95% CI) Model 2	1.00	0.84 (0.59-1.20)	1.25 (0.49-3.20)	0.787	0.60

Values shown are mean (± SD). #: Geometric mean (approximate SD), p-value*: recessive genotypic model, p-value§: additive genotypic model, Model 1: age and practice adjusted , Model 2: adjusted for age, practice and LDL-C.

6.2.3.2. Association of *ANXA2* SNP rs17845226 with lipid traits and CHD in WHII

No association was found in the WHII between the SNP and lipid traits in either sex (Table 6-5). Also, a lack of trend between the rs17845226 genotype and plasma lipid concentrations in this cohort was noted. On the other hand, the minor allele was associated with a higher risk of CHD in both genders [OR (95% CI): 2.41 (1.28 - 4.54), p=0.02]. Although women with the minor allele were at higher risk of CHD [OR (95% CI): 2.85 (0.80 - 10.10), p=0.35] than men [OR (95% CI): 2.28 (1.10 -4.74), p=0.05], the difference between groups was not significant (p=0.77) (439).

6.2.3.3. *ANXA2* SNPs rs17845226 meta-analysis for association with LDL-C and CHD

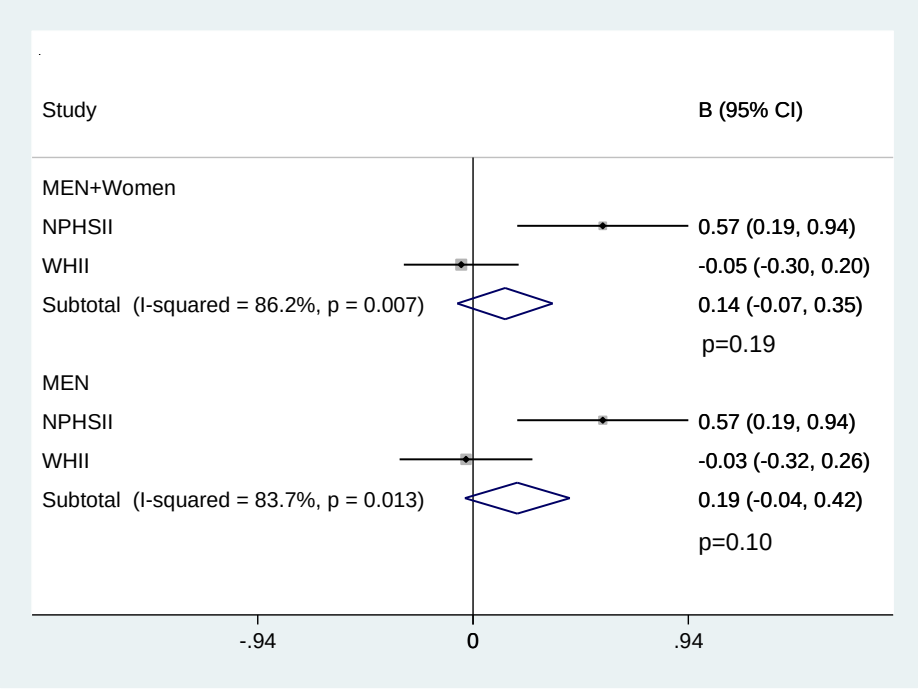
Figure 6-4 shows that the minor allele of the SNP had the opposite effect in the two studies (NPHSII and WHII); the overall effect of the SNP in the LDL-C is non-significant [Men: beta-coefficient β (95%CI) = 0.19 (-0.04 - 0.42), p=0.10; Men and women: β (95%CI) = 0.14 (-0.07 - 0.35), p=0.19]. Due to the inconsistency of the SNP effect on LDL-C levels, the heterogeneity of outcome data was tested using I^2 and 'substantial' heterogeneity was found between the two studies [I^2 =86.2%, p=0.007]. In contrast, as shown in Figure 6-6, the risk of CHD in both studies was consistent [I squared=0.0%, p=0.85]. The minor allele of the SNP was associated with higher risk of CHD in both studies; the overall risk in both sexes OR (95%CI) = 2.92 (1.41 - 3.72), p=0.001.

Table 6-5: An association between ANXA2 rs17845226 (Val98Leu) genotype and lipid risk factors and CHD risk in WHII

Sex	Men				Women				Men+ women			
Genotype	**CC (n=2474)**	**CA (n=818)**	**AA (n=52)**	**p-value* (ANOVA)**	**CC (n=793)**	**CA (n=258)**	**AA (n=19)**	**P-value* (ANOVA)**	**CC (n=3267)**	**CA (n=1076)**	**AA (n=71)**	**p-value*# (ANOVA)**
TC (mmol/l)	6.45 (1.11)	6.45 (1.12)	6.42 (1.17)	0.86	6.44 (1.19)	6.53 (1.16)	6.27 (1.04)	0.49	6.45 (1.13)	6.47 (1.13)	6.38 (1.13)	0.60
LDL-C (mmol/l)	4.42 (0.98)	4.44 (1.00)	4.40 (1.11)	0.83	4.21 (1.09)	4.28 (1.030	4.15 (1.07)	0.77	4.37 (1.01)	4.40 (1.01)	4.33 (1.10)	0.73
HDL-C (mmol/l)	1.29 (0.34)	1.29 (0.32)	1.23 (0.35)	0.28	1.66 (0.42)	1.67 (0.45)	1.61 (0.38)	0.56	1.37 (0.39)	1.37 (0.38)	1.33 (0.39)	0.22
TG (mmol/l)	1.32 (0.74)	1.29 (0.70)	1.39 (0.81)	0.48	1.00 (0.47)	1.01 (0.51)	0.95 (0.38)	0.68	1.23 (0.683)	1.22 (0.66)	1.25 (0.70)	0.67
CHD risk by genotype												
No. without CHD (%)	2266 (74.5)	734 (24.1)	43 (1.4)	-	744 (74.3)	241 (24.1)	16 (1.6)	-	3010 (74.4)	975 (24.1)	59 (1.5)	-
No. with CHD (%)	208 (69.1)	84 (27.9)	9 (2.99)	-	49 (71.0)	17 (24.6)	3 (4.4)	-	257 (69.5)	101 (27.3)	12 (3.2)	-
OR (95% CI)	1.00	1.25 (0.95-1.63)	2.15 (1.04-4.46)	0.04	1.00	1.07 (0.61-1.89)	2.80 (0.80-9.85)	0.11	1.00	1.21 (0.95-1.54)	2.41 (1.28-4.54)	0.01

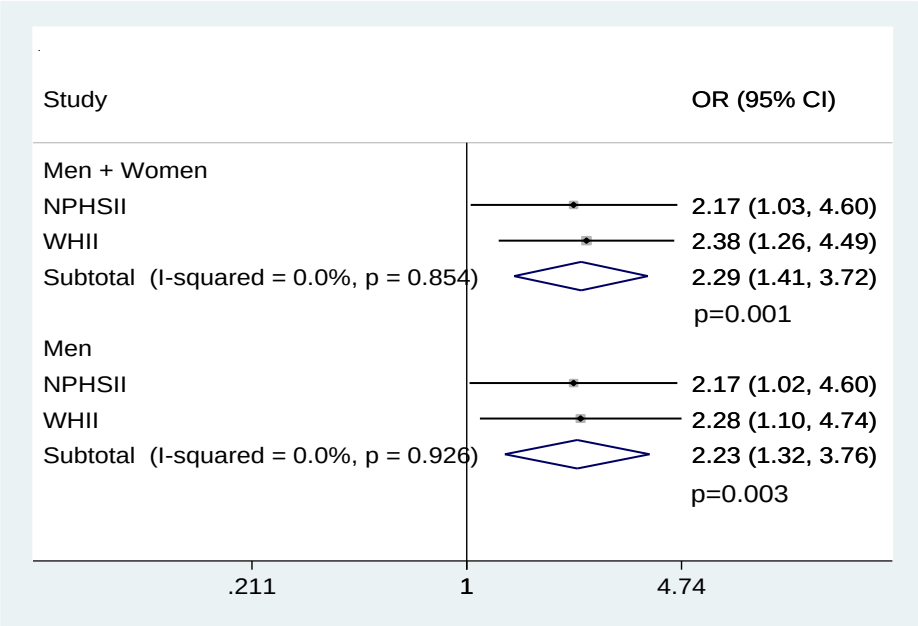
Values shown are mean (±SD)
p-value*: recessive genotypic model
#: Adjusted for sex

Figure 6-4: Forest plot for the meta-analysis of LDL-C and carriage of the minor allele of rs17945226



B= beta-coefficient, Subtotal= fixed effect.

Figure 6-5: Forest plot for the meta-analysis of CHD and carriage of the minor allele of rs17945226



6.3. Discussion

In vivo and *in vitro* biological studies have demonstrated that AnxA2 prevents PCSK9-mediated degradation of LDL-R (101, 102). AnxA2 mediates this inhibitory effect via the interaction of its R1-domain with the CHRD of PCSK9. It was suggested that a mutation in the *ANXA2* R1-domain could affect LDL-C levels in humans. Out of eight reported SNPs in the R1-domain, the rs17845226 missense variant p.(Val98Leu) was selected for further analysis because this variant had previously been associated with lower circulating PCSK9 levels (102) and the MAF in the European population is high at 13%. Genotype-phenotype analysis of NPHSII data suggested that this missense mutation had a recessive effect on LDL-C levels, with significantly higher levels seen only in those carrying two copies of the minor allele. Indeed, the activity of the leucine variant may depend on the fact that AnxA2 exists as a dimer with two R1-domains, and thus in an individual heterozygous for the variant, where one of the two alleles carried is less effective, the other is functional and is still able to bind to PCSK9. However, if both inherited alleles are less effective in their ability to interact with PCSK9, this may allow PCSK9 to act on LDL-R. It is important to note that the majority of AnxA2 exists as a heterotetramer complexed with p11, and it is as yet unclear how the rs17845226 variant may affect the AnxA2/p11 complex formation or the ratio of monomeric vs complexed AnxA2 proteins, all of which may have consequences for PCSK9-mediated LDL-R down regulation.

Unfortunately, the identified association with LDL-C levels in the NPHSII was not replicated in the WHII, but in both studies the SNP was associated with CHD risk. The

inconstancy between NPHSII and WHII results may be due to confounding factors such as sampling and measurements, where sample collection and lipid measurements were different between studies. The NPHSII samples were collected from non-fasting subjects but they were under food restrictions of having only a light breakfast, also the study only had one set of LDL-C and HDL-C directly measured at year-6 in the majority of subjects, where the means were within the range expected for middle aged males. On the other hand, the WHII samples were collected from fasting subjects and lipid markers were measured in three phases of the study (3, 5, and 7), where the mean LDL-C varied considerably between years (~ 4.4, 3.8 and 3.5 mmol/l respectively). In this study phase 3 data was used for analysis, which was the most informative data (the majority of subjects had data).

The inconstancy between NPHSII and WHII results, together with bioinformatics results of non-pathogenicity of the SNP and the presence of modest proxy SNPs in the non-coding region, prevented a definite conclusion about SNP functionality. However, it also suggests that the Val98Leu SNP may be acting as a marker for one or more SNPs in the non-coding region that affects gene expression, and further analyses will be needed to disentangle this. Chapter 7 investigates the likelihood of the *ANXA2* coding SNP rs17845228 (Val98Leu) acting as a marker for a functional SNP elsewhere at the locus.

Chapter 7 : Identification of an LDL-C associated variant in the non-coding region of *ANXA2* and determination of the functional role of identified variants

7.1. Introduction

Recently, *in vivo* studies have supported *in vitro* studies that implicate AnxA2 in preventing PCSK9-mediated degradation of LDL-R (101, 102). AnxA2 mediates this inhibitory effect via the interaction of its R1-domain with a CHRD of PCSK9. In Chapter 6, it was found that the *ANXA2*-R1 missense mutation p.(Val98Leu) was associated with high levels of LDL-C in the NPHSII but not in the WHII study, but was predicted to be non-pathogenic and had modest LD with SNPs downstream of the *ANXA2* gene in the long intergenic region. *ANXA2* expression is regulated at both the transcriptional and translational levels (188), and the long intergenic regions are known to have key regulators for transcriptional and translational machinery of genes (441). For instance, an ETS-related gene (ERG) binds to the DNA sequence upstream of *ANXA2* gene in the long intergenic region, and acts as a repressor to down-regulate *ANXA2* expression in prostate epithelial cells (442).

It was hypothesized that the *ANXA2* coding SNP, rs17845226 p.(Val98Leu) may act as a marker for proxy functional SNPs. The aims were a) to select the potential functional candidate SNPs, b) to assess the association between the candidate SNPs and cholesterol levels and CHD risk, and c) to elucidate the molecular mechanism behind the effect identified. Association analyses, and bioinformatics and *in vitro* functional assays were used in order to identify the functional SNPs at this locus.

7.2. Results

7.2.1. Bioinformatics analysis

The *ANXA2*-R1 rs17845226 SNP has a modest LD (r^2 >= 0.4) with 34 SNPs (Figure 6-2). All of these located downstream of the *ANXA2* gene-coding region in the long intergenic region between two genes, *FOXB* and *ANXA2* on chromosome 15, and near the *RORA* and *LIPC* loci which play roles in lipid metabolism and atherosclerosis (Figure 7-1). Out of the 34 SNPs, the rs17191344 SNP (r^2=0.45, MAF= 16% in European population) has the strongest regulatory profile as is highly conserved and has strong enhancer signs in 13 tissues including liver. The SNP is located in an open chromatin region, where the markers of DNAase hypersensitivity, FAIRE, and transcription factor binding by ChIP-seq are strong (Figure 7-2). In addition, both ENCODE and the ElDorado *in silico* tool have shown that the G allele of the rs17191344 SNP creates a binding site for CTCF, and this SNP also changes a YY1 binding motif . The YY1 transcription factor can associate with CTCF and regulate gene expression (443). The rs17191344 SNP has strong LD ($r^2 \geq 0.8$) with 66 SNPs, all in the intergenic region (Figure 7-3), of which SNPs rs11633032 and rs12900101 (MAF= 17%) are predicted to bind to regulatory transcription factors as shown in the ElDorado data. AP-1 and NRSF are predicted to bind to rs11633032 in the presence of the minor allele of the SNP, while MAZR and PLAG1 are predicted to bind to the minor allele of rs12900101. These two SNPs are also in modest LD with the *ANXA2*-coding SNP rs17845226 (r^2=0.40 and 0.41 respectively).

Figure 7-1: Diagram of chromosome 15 locus

The diagram shows *ANXA2* and nearby genes. Single nucleotide polymorphisms (SNPs) selected for the study are shown: one coding SNP rs1785226 and two intergenic SNPs, rs17191344 and rs11633032, showing modest LD (r^2=0.45).

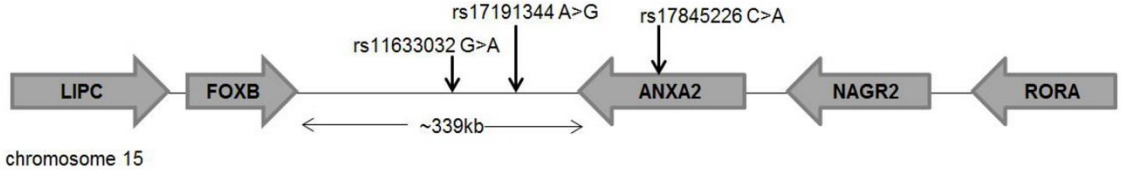


Figure 7-2: Genome-wide maps of chromatin state of *ANXA2* rs17191344 SNP.

Schematic presentation of the rs17191344 chromatin status (https://genome-euro.ucsc.edu). The area of interest is highlighted in light-blue. FAIRE = formaldehyde assisted isolation of regulatory elements.

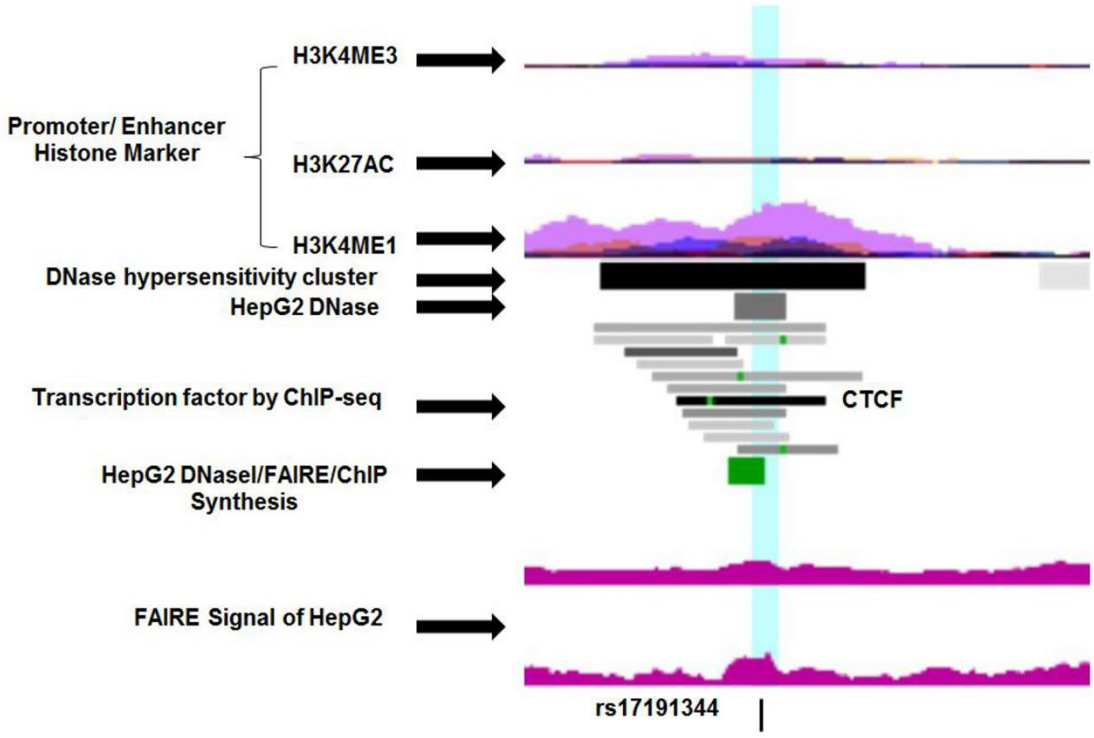
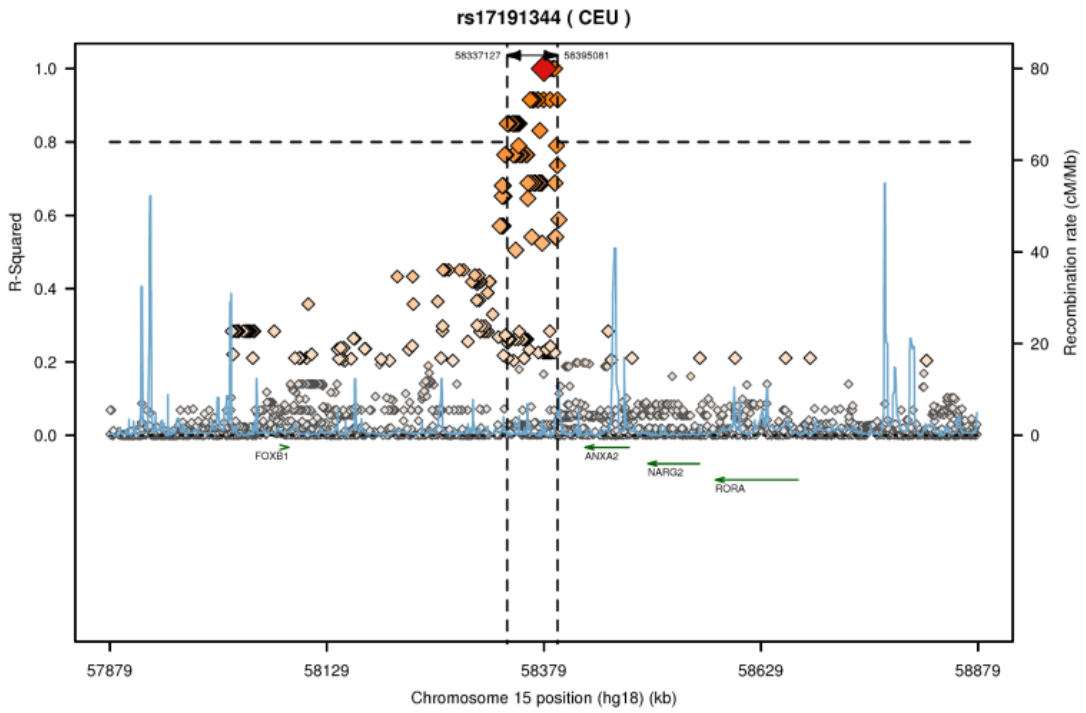


Figure 7-3: *ANXA2* rs17191344 linkage disequilibrium (LD) plot.

An LD plot was generated using SNAP V2.2 (Broad Institute; https://www.broadinstitute.org/mpg/snap/ldplot.php). The LD between the query SNP and each corresponding proxy SNP are represented by diamonds, the query SNP (rs17191344) is represented by a red diamond, while LD SNPs are represented by orange diamonds. R-squared value is shown on the left Y-axis, where $r^2 \geq 0.8$ indicates a strong LD, and $r^2 \leq 0.2$ indicates a weak LD. A blue line indicates estimated recombination rates and their values are presented on the right Y-axis. LD and recombination rates are based on HapMap Phase II in a population of Utah residents with Northern and Western European ancestry (CEU).



7.2.2. Genotype-Phenotype Association

To determine whether the *ANXA2* intergenic SNPs are associated with LDL-C or CHD, I looked at the *ANXA2* intergenic SNPs selected (rs17191344, rs11633032 and rs12900101) or proxy SNPs selected in large meta-analysis studies including the Global Lipids Genetics Consortium (http://csg.sph.umich.edu/abecasis/public/lipids2013/) and

the CARDIoGRAMplusC4D consortium (http://www.cardiogramplusc4d.org/data-downloads/). However, the SNPs were not in these published data. Therefore the *ANXA2* SNP rs17191344 was genotyped in the NPHSII and the proxy SNPs rs11633032 and rs12900101were imputed in the UCLEB consortium.

7.2.2.1. Association of *ANXA2* SNPs rs17845226 and rs17191344 with lipid traits and CHD in NPHSII

The association of rs17191344 with LDL-C and CHD was investigated in the NPHSII cohort. Baseline study characteristics are summarized in Table 6-3. TaqMan genotyping of rs17191344 was successful with a call rate of ≥90%. The minor allele frequency of rs17191344 in the NPHSII was 0.145 which is similar to the frequency (0.155) seen in European populations ($\chi 2$=0.09, p=0.77).

The modest LD SNP rs17191344 (A>G) shows the same trend as the lead coding SNP (rs17845226) that was reported in chapter six. The rs17191344 showed an association with LDL-C and CHD under a recessive mode of inheritance (Table 7-1), where individuals who carried two copies of the minor allele had higher levels of LDL-C (p=0.05) and a higher risk of CHD [HR (95%CI) = 1.86 (1.02-3.41), (p=0.05)]. The data also showed that a rising level of LDL-C is most likely the risk factor that explains the higher risk of CHD (439).

Table 7-1: An association between rs17191144 genotype and lipid risk factors and CHD risk in the Second-Northwick-Park Heart Study (NPHSII).

Genotype	AA (n=1956)	AG (n=659)	GG (n=57)	P-value § (ANOVA)	p-value* (ANOVA)
TC (mmol/l)	5.72 (1.01)	5.74 (1.03)	5.96 (1.04)	0.205	0.085
LDL-C (mmol/l)	3.09 (1.00)	3.09 (1.03)	3.38 (1.03)	0.302	0.051
HDL-C (mmol/l)	0.81 (0.24)	0.79 (0.25)	0.81 (0.26)	0.494	0.786
TG (mmol/l)#	1.79 (0.93)	1.84 (0.97)	1.79 (0.95)	0.465	0.604
CHD risk by genotype					
No. without CHD (%)	1749 (73.5)	585 (24.6)	45 (1.9)	-	-
No. with CHD (%)	207 (70.7)	74 (25.3)	12 (4.1)	-	-
HR (95% CI) Model 1	1.00	1.10 (0.83-1.44)	1.86 (1.02-3.41)	0.113	0.052
HR (95% CI) Model 2	1.00	0.98 (0.72-1.34)	1.29 (0.60-2.79)	0.787	0.504

Values shown are mean (±SD), # Geometric mean (approximate SD), *P-value according to recessive model, § P-value according to additive model, Model 1: age and practice adjusted, Model 2: adjusted for age, practice and LDL-C.

It is known that a single gene polymorphism with a modest effect on a trait will have minimal effects in the end-point disease phenotype. However, with a trait such as LDL-C, where genetic and environment factors affect the trait, the presence of multiple SNPs with small effect can lead to a more severe disease or trait phenotype. In order to assess whether the intergenic SNP rs17191344 is responsible for observed phenotypes (LDL-C and CHD risk), or whether both SNPs have an impact on the phenotype where the risk is increased per allele, the association between 9 combined genotype groups (1-9) of rs17191344 and rs17845226 and cholesterol levels were examined. Table 7-2 and Figure 7-4 show that the LDL-C levels increase per minor allele of both SNPs. The individuals who have two copies of the minor allele of both SNPs (G and A, respectively) have

significantly higher levels of LDL-C (p=0.007) and a two-fold higher risk of CHD [HR (95% CI) = 2.75 (1.18 - 6.39), (p=0.02)] (439).

Stepwise models were also applied to determine which of the combined genotype groups (1-9) were having an effect on lipids and CHD. Stepwise models indicated that only those with two copies of the minor allele of both SNPs had significantly raised levels of the lipids: TC [β (se) = 0.48 (0.22), p=0.03] and LDL-C [β (se) = 0.65 (0.24), p=0.008] (Table 7-3). This confirms the results shown in Table 7-2, where only the GG/AA group has significantly different levels from the other groups for these lipid traits. For CHD, the GG/CC group has significantly higher risk compared to the other groups [HR (95% CI) = 6.53 (1.62-26.27), p=0.008] (439).

Table 7-2: The rs17191344 and rs17845226 SNPs combined genotype association with cholesterol traits and CHD in the Second-Northwick-Park Heart Study (NPHSII).

		Lipids					CHD			
rs17191344	**rs17845226**	**N**	**TC mmol/l**	**P-value**	**LDL-C mmol/l**	**P-value**	**N**	**CHD % (N)**	**HR (95% CI)**	**p-value**
AA	CC	1658	5.73 (1.00)	-	3.09 (0.99)	-	1658	10.7 (177)	1.00	-
	CA	75	5.81 (1.05)	0.49	3.26 (1.06)	0.21	75	6.7 (5)	0.53 (0.21-1.33)	0.18
	AA	3	5.57 (0.45)	0.78	3.18 (0.72)	0.88	3	0 (0)	-	-
AG	CC	137	5.67 (0.97)	0.49	2.97 (1.00)	0.21	137	12.4 (17)	1.32 (0.79-2.20)	0.29
	CA	443	5.78 (1.04)	0.33	3.13 (1.04)	0.48	443	10.6 (47)	1.03 (0.74-1.43)	0.88
	AA	9	6.32 (0.92)	0.08	3.63 (1.08)	0.15	9	22.2 (2)	3.17 (0.78-12.92)	0.11
GG	CC	4	6.18 (0.36)	0.38	3.72 (0.77)	0.28	4	50.0 (2)	8.60 (2.09-35.34)	0.003
	CA	23	5.90 (0.94)	0.43	3.22 (1.04)	0.56	23	13.0 (3)	1.35 (0.43-4.26)	0.604
	AA	21	6.22 (1.17)	0.025	3.75 (0.96)	0.007	21	28.6 (6)	2.75 (1.18-6.39)	0.019

Adjusted for age
Values shown are mean (±SD)

Figure 7-4: The association of 9 combined genotype groups of rs17191344 and rs17845226 with LDL-C levels in the Second-Northwick-Park Heart Study (NPHSII)

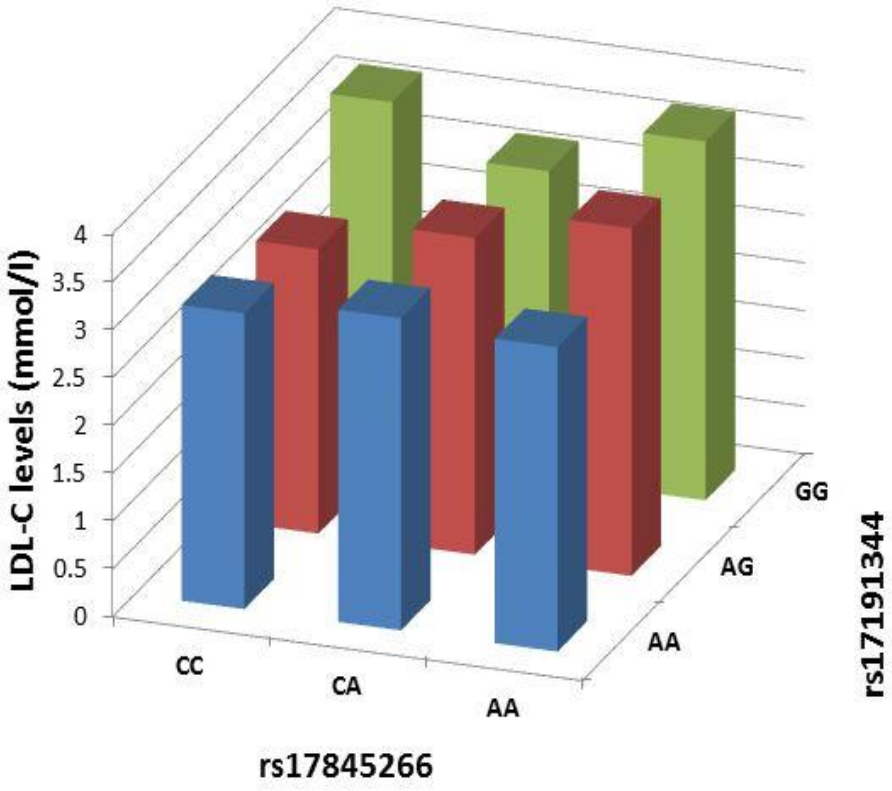


Table 7-3: Stepwise regression analysis: Second-Northwick-Park Heart Study (NPHSII)

The stepwise regression of the combined genotype groups (1-9) of the rs17191344 and rs17845226 SNPs for each outcome variable: TC, LDL-C and CHD. The reference group is the group that carries major alleles of both SNPs (AA/CC) respectively.

Group	TC B (se)	p-value	LDL-C B (se)	p-value	CHD HR (95% CI)	P-value
Reference	0		0		1.00	
GG/AA	0.482 (0.221) R^2=0.002	0.029	0.649 (0.244) R^2=0.004	0.008	-	
GG/CC					6.53 (1.62-26.27)	0.008

Adjusted for age and practice

For TC and LDL-C, only those with two copies of the minor allele of both SNPs differed significantly from the group that carried the major alleles of both SNPs (AA/CC), and for the analysis above the reference group used is the combination of AA/CC and all other groups. For CHD, only those with the genotype GG/CC differed significantly from the group that carried the major alleles of both SNPs (AA/CC), and for this analysis the reference group used is the combination of AA/CC and all other groups.

7.2.2.2. Association of *ANXA2* intergenic SNPs with lipid traits and CHD in the UCLEB consortium

The UCLEB consortium, comprising ~14,600 subjects from the UK general population, was used for replication. Study characteristics are summarized in Table 7-4. It was not possible to impute rs17191344, but two SNPs, rs11633032 (G>A) and rs12900101(C>G) [having a strong LD (r^2=0.98) with rs17191344], had a high average value of imputation genotype data from the Metabochip (r^2= 0.63) (Table 7-5).

As shown in Table 7-6 and Figure 7-5, the minor alleles of rs11633032 and rs12900101 were associated with significantly higher levels of LDL-C in men with the recessive inherited mode [effect size = 0.21 mmol/l, p=0.018 and 0.19 mmol/l, p=0.036, respectively]; but not in women, which may be due to different sample sizes between the sexes. To assess whether the effect is different between the sexes, we tested the estimated difference between sexes, and found no difference (p=0.29) (Table 7-7). Overall, the minor alleles of rs11633032 and rs12900101 were associated with significantly higher levels of LDL-C [effect size = 0.16 mmol/l, p=0.03 and 0.14 mmol/l, p=0.05 respectively for the recessive model] with similar effects in both genders (Table 7-7). This result confirms the results seen above in the NPHSII subjects, who were all men.

Table 7-4: The UCL-LSHTM-Edinburgh-Bristol (UCLEB) consortium baseline characteristics for eight studies included in this study.

	BRHS	BWHHS	CaPS	EAS	ELSA	ET2DS	WHII	MRC1946
Number included in analysis	2387	1922	1381	850	1934	1049	3311	2139
Age (years)	68.9 (5.61)	71.5 (5.28)	61.90 (5.05)	64.46 (5.64)	73.64 (9.55)	67.90 (4.20)	49.00 (5.96)	50 (0)
Sex (% male)	100	0	100	49	53	52	76	50
BMI (kg/m^2)	26.83 (3.63)	27.67 (4.99)	NA	26.2 (4.2)	27.48 (4.49)	31.39 (5.66)	25.21 (3.54)	27.63 (4.63)
TC (mmol/l)	6.36 (1.03)	6.62 (1.22)	6.20 (1.10)	7.08 (1.33)	5.71 (1.28)	4.31 (0.90)	6.44 (1.12)	6.09 (1.07)
HDL-C (mmol/l)	1.15 (0.24)	1.62 (0.45)	NA	1.45 (0.38)	1.49 (0.39)	1.29 (0.36)	1.41 (0.40)	1.67 (0.52)
LDL-C (mmol/l)	3.89 (1.00)	4.14 (1.10)	NA	5.33 (1.22)	3.42 (1.06)	NA	4.37 (1.00)	3.52 (0.97)
TG * (mmol/l)	2.05 (1.22)	1.91 (1.05)	1.87 (1.10)	1.53 (0.90)	1.80 (1.11)	NA	1.46 (1.15)	2.16 (1.51)
SBP (mmHg)	144.0 (19.95)	150.0 (25.36)	145.0 (22.32)	NA	138.9 (19.63)	133.3 (16.46)	120.5 (13.14)	136.2 (20.01)
DBP (mmHg)	81.8 (12.80)	80.1 (11.76)	81.7 (11.98)	NA	73.0 (11.46)	69.1 (9.00)	79.8 (9.12)	84.6 (12.17)

Mean and standard deviation (SD), where appropriate, are shown. * analysis carried out on log transformed and median and IQR presented.

British Regional Heart Study (BRHS), British Women's Heart and Health Study (BWHHS), Caerphilly Prospective Study (CaPS), Edinburgh Artery Study (EAS), English Longitudinal Study of Aging (ELSA), Edinburgh Type 2 Diabetes Study (ET2DS), Whitehall II study (WHII), and MRC National Survey of Health and Development 1946 (MRC 1946).

Table 7-5: Imputation quality of The UCL-LSHTM-Edinburgh-Bristol (UCLEB) consortium

Quality of imputation was measured as the ratio of observed (imputed) genetic variance to expected genetic variance (r^2) in 8 studies

SNP	WHII	BRHS	BWHHS	ET2DS	EAS	CAPS	MRC1946	ELSA
rs11633032	0.63	0.64	0.63	0.63	0.62	0.64	0.63	0.63
rs12900101	0.63	0.63	0.63	0.63	0.62	0.64	0.63	0.63

Table 7-6: Association of *ANXA2* intergenic SNPs and lipid traits in the UCL-LSHTM-Edinburgh-Bristol (UCLEB) Consortium

	AnxA2 SNP	Lipid trait	Effect Size[$]	Std. Error	p-Value*	Number of samples by genotype
Men	**rs12900101**	TC	0.147	0.086	0.088	CC=5995, CG=2824, GG=161
		LDL-C	0.188	0.090	0.036	CC=4626, CG=2190, GG=124
		HDL-C	0.010	0.031	0.734	CC=4998, CG=2361, GG=131
		Log TG	-0.007	0.047	0.876	CC=5084, CG=2419, GG=134
	rs11633032	TC	0.171	0.087	0.049	GG=5992, AG=2824, AA =158
		LDL-C	0.215	0.091	0.018	GG=4626, AG=2193, AA =121
		HDL-C	0.009	0.031	0.774	GG=5000, AG=2362, AA =128
		Log TG	-0.002	0.048	0.968	GG=5086, AG=2420, AA =131
Women	**rs12900101**	TC	0.092	0.126	0.461	CC=3714, CG=1773, GG=90
		LDL-C	0.104	0.119	0.383	CC=3302, CG= 1588, GG=84
		HDL-C	0.020	0.045	0.668	CC=3682, CG=1760, GG=90
		Log TG	0.410	0.052	0.427	CC= 3374, CG= 1626, GG=85
	rs11633032	TC	0.105	0.127	0.410	GG=3718, GA=1771, AA=88
		LDL-C	0.104	1.120	0.390	GG=3306, GA= 1586, AA=82
		HDL-C	0.029	1.046	0.532	GG=3686, GA=1758, AA=88
		Log TG	0.044	1.052	0.401	GG= 3378, GA= 1624 , AA=83
Men & Women	**rs12900101**	TC	0.117	0.072	0.104	CC=9709, CG=4597, GG=251
		LDL-C	0.143	0.072	0.048	CC=7928, CG=3778 , GG=208
		HDL-C	0.016	0.026	0.536	CC=8680, CG=4121, GG=221
		Log TG	0.007	0.035	0.836	CC=8458, CG=4045, GG=219
	rs11633032	TC	0.139	0.073	0.056	GG=9716, AG=4595, AA =246
		LDL-C	0.160	0.073	0.029	GG=7932, AG=3779, AA =203
		HDL-C	0.019	0.026	0.468	GG=8686, AG=4120, AA =216
		Log TG	0.013	0.036	0.726	GG=8464, AG=4044, AA =214

$= Effect size is the main effect of the homozygous minor allele genotype on lipid concentration estimated in a linear model containing age and study ID as covariates.
*= Recessive genotypic model

Figure 7-5: Average of the phenotype by genotype class for *ANXA2* intergenic SNPs in the UCL-LSHTM-Edinburgh-Bristol (UCLEB) Consortium.

The Genotype is presented as a binary character: class 0 (homozygous major allele) dose 0 - 0.5, class 2 (homozygous minor allele) dose 1.5 – 2, and class 1 (heterozygotes) dose 0.5-1.5. The error bars represent the 95% confidence interval for the mean.

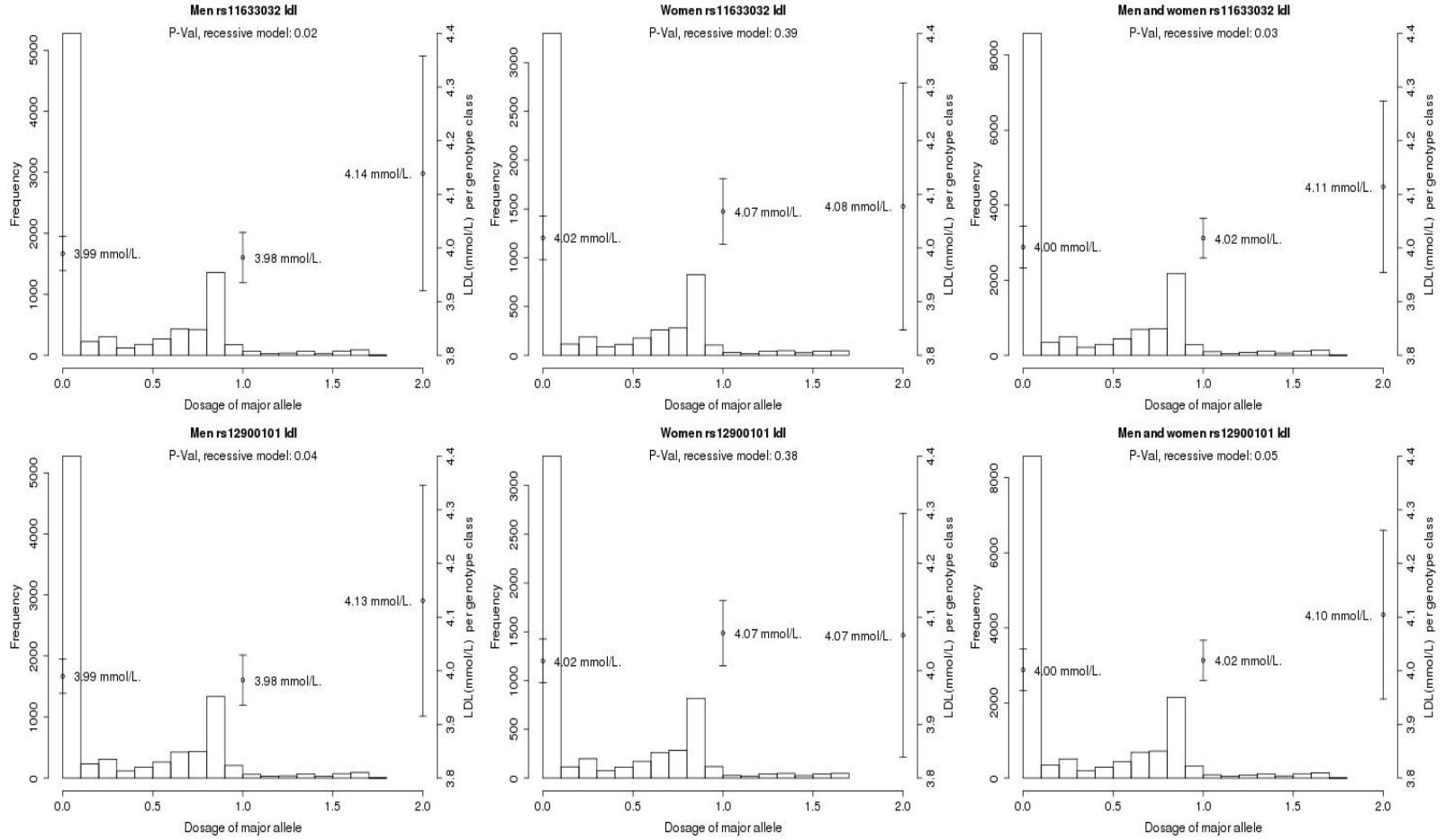


Table 7-7: Estimated difference between sexes of *ANXA2* intergenic SNPs and lipid traits in The UCL-LSHTM-Edinburgh-Bristol (UCLEB) Consortium

SNP ID	Lipid trait	Mean	Standard error	P- value
rs12900101	LDL-C	0.084	0.102	0.819
	TC	0.054	0.103	0.521
rs11633032	LDL-C	0.11	0.104	0.288
	TC	0.066	0.104	0.524

To assess whether there the effect of the SNP on TC and LDL-C was different between the sexes. The estimated difference between sexes was tested using a t-test.

7.2.3. Allele-specific protein binding of ANXA2 intergenic SNPs in Huh7 Cells

7.2.3.1. Conventional EMSA of *ANXA2* intergenic SNPs

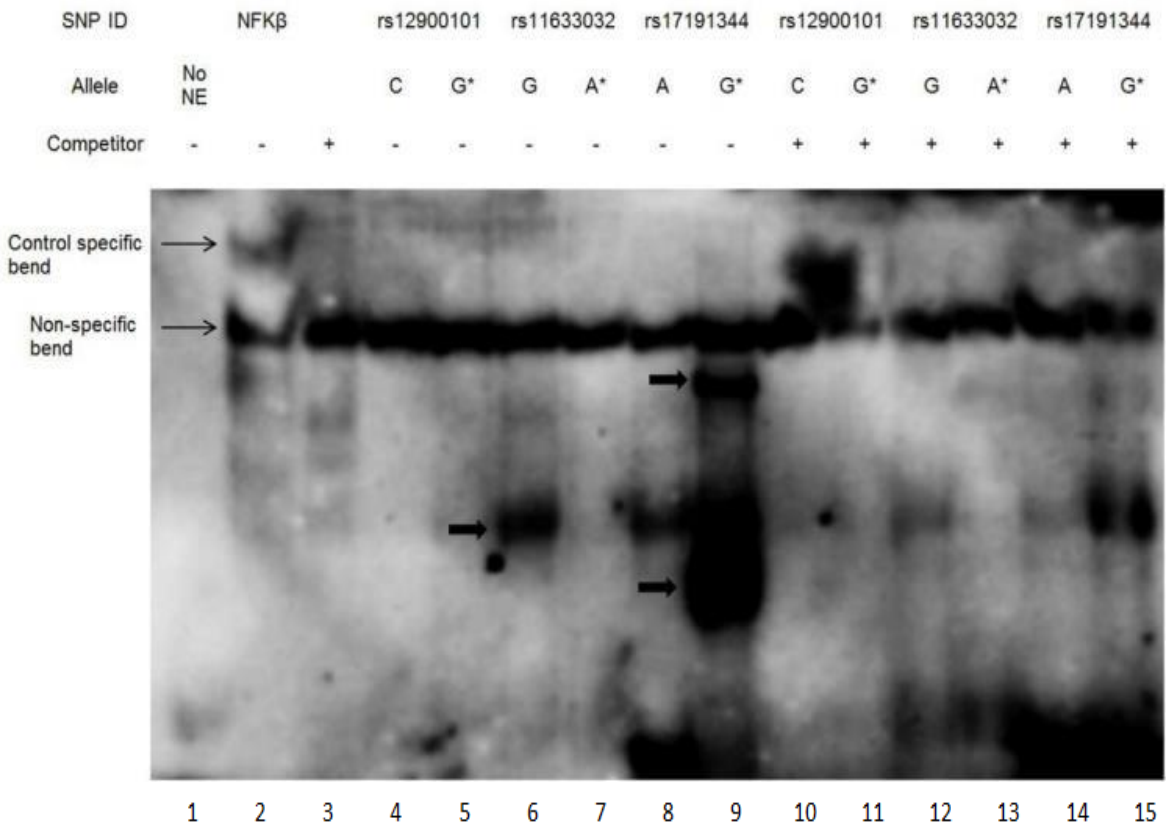
Conventional EMSAs were performed to determine whether the three *ANXA2*-intergenic SNPs, rs12900101, rs11633032 and rs17191344, are within potential regulatory elements and able to affect DNA-protein interactions. NFKβ was used as a control. Each set of biotin labeled probes was run alongside an unlabeled probe in order to identify allele specific band(s). The band shifting between alleles is a sign of a change in DNA-protein binding that may be involved in *ANXA2* regulation.

In Figure 7-6, NFKβ control lanes 1-3 comprised: the negative control (lane 1) that did not contain nuclear extract and no bands were visible, the positive control (lane 2) that contained a labeled probe that showed a NFKβ specific band indicated by arrows, and lane 3 that showed when the unlabeled probe was added the specific band disappeared. The rs12900101 SNP did not show protein binding in either alleles of the SNP (lanes 4-5 and 10-11). However, two SNPs, rs11633032 and rs17191344, demonstrated differential protein binding by allele. The rs11633032 major G allele was bound to proteins, whereas the risk A allele did not show allele-specific protein binding (lanes 6-7), and the G allele specific band was eliminated by an unlabeled probe (lanes 12-13). In contrast, the risk G allele of the rs17191344 SNP bound strongly to proteins or complexes of proteins where

two specific bands indicated by arrows (lanes 8-9) appeared; but when the unlabeled probe was added the G allele specific bands disappeared (lanes 14-15) (439).

Figure 7-6: DNA binding properties of *ANXA2*- intergenic SNPs.

Conventional EMSA analysis of the *ANXA2*-intergenic SNPs (rs12900101, rs11633032 and rs17191344). The major allele of rs11633032 has allele-specific binding that is competed out by the allele competitor probe. The minor allele of rs17191344 has allele-specific binding at two positions which are competed out by the allele competitor probe. Allele specific bands are indicated by arrows and (*) indicates minor allele.



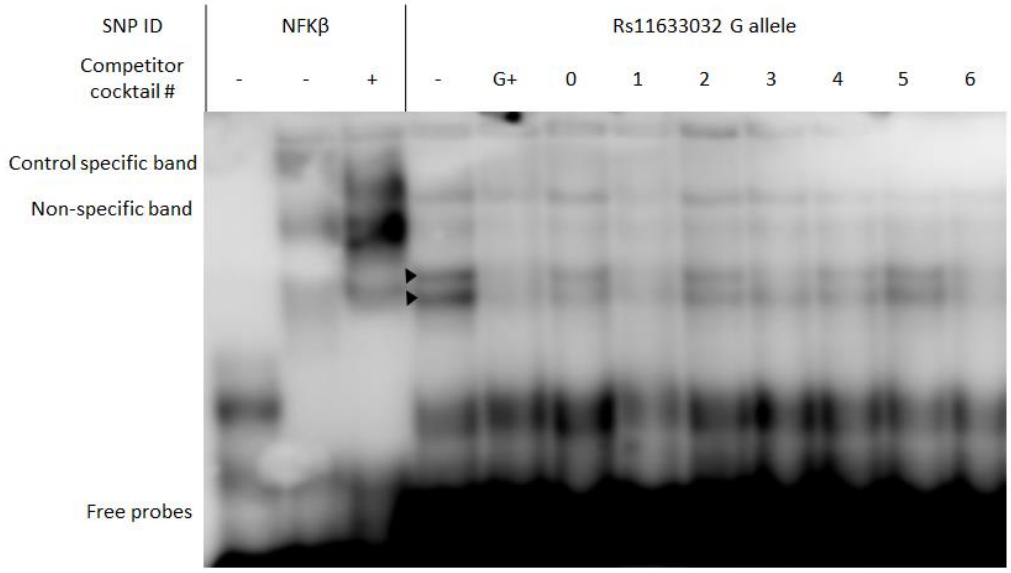
7.2.3.2. Multiplex competitor EMSAs (MC-EMSA) of *ANXA2* intergenic SNP rs11633032

MC-EMSA was performed to characterize the DNA-protein interaction of the major G allele of the rs11633032 SNP. Nuclear extract of Huh7 cells was incubated with seven sets of cocktails (0-6) with each set having 10 unlabeled dsDNA consensus sequences for well-characterized proteins (Table 2-18), before adding the rs11633032 G allele labeled probe. The results showed that the G allele specific-bands indicated by arrows, were competed out by cocktail 1 (Figure 7-7 A). Then, when each competitor of cocktail 1 was run individually, the G allele specific-bands were competed out by addition of the GATA and Egr1 consensus sequences (Figure 7-7 B) (439).

Figure 7-7: DNA binding and expression of the transcription factors of the *ANXA2*-intergenic SNP rs11633032.

MC-EMSA analysis. Nuclear proteins from the Huh7 cell line were incubated with 7 cocktails (0-6) of unlabeled DNA competitors (70 well-characterized DNA-binding proteins) for 15 minutes, then a 5' end-biotinylated allele-specific probe was added. The multiplex competitors competed out any specific interactions with a labeled probe, eliminating or reducing any positive shift result. A) *ANXA2*- intergenic SNP rs11633032 MC-EMSA for G allele of the SNP, allele specific bands indicated by arrows, were eliminated by cocktail 1. B) The single competitors from cocktail 1 were run individually in a further EMSA, showing Egr1 and GATA that resulted in competition (eliminated bands indicated by dotted arrows). G+ indicates that the unlabeled G probe was added to the reaction mixture.

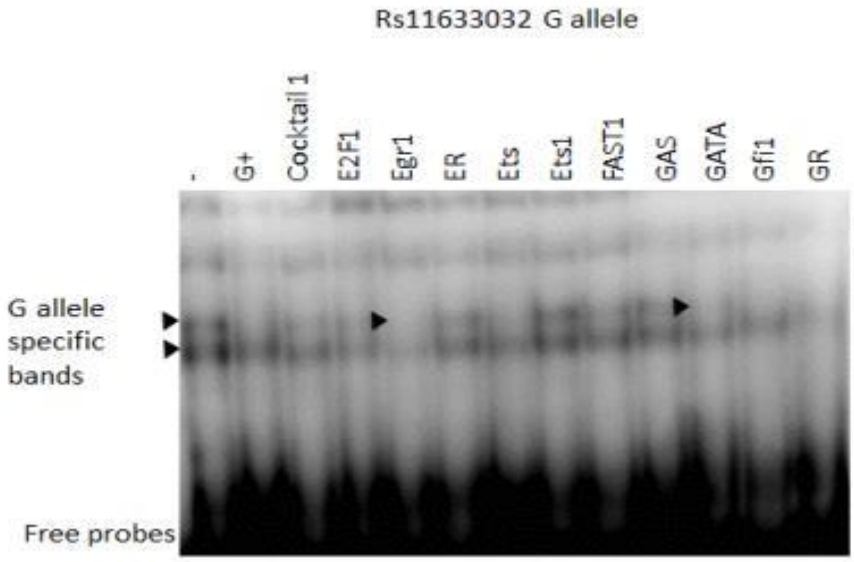
A)



Unlabeled multiplexed DNA-based competitors to commonly studied DNA-binding proteins

AP1	E2F1	HIF1a	NFATc	RAR	Stat4	VDR
AP2	Egr1	ISRE	NFkB	RXR	Stat5	YY1
AR	ER	HNF4	NR5A2	SIE	Stat56	ZEB
Brn3	Ets	IRF1	Oct-01	Smad	Tbet	HNF1
CBP	Ets1	MEF1	p53	Smad34	TFE3	ARP1
CDP	FAST1	MEF2	Pax5	Smuc	TFEB	NFY
CEBP	GAS	MIBP1	Pbx1	Sp1	TFIID	HNF3
cMyb	GATA	MycMa	Pit1	SRE	TGIF	BARP
CREB	Gfi1	NF1	PPAR	Stat1	TR	SREBP1
CTCF	GR	NFE2	PR	Stat3	USF1	HSF1

B)

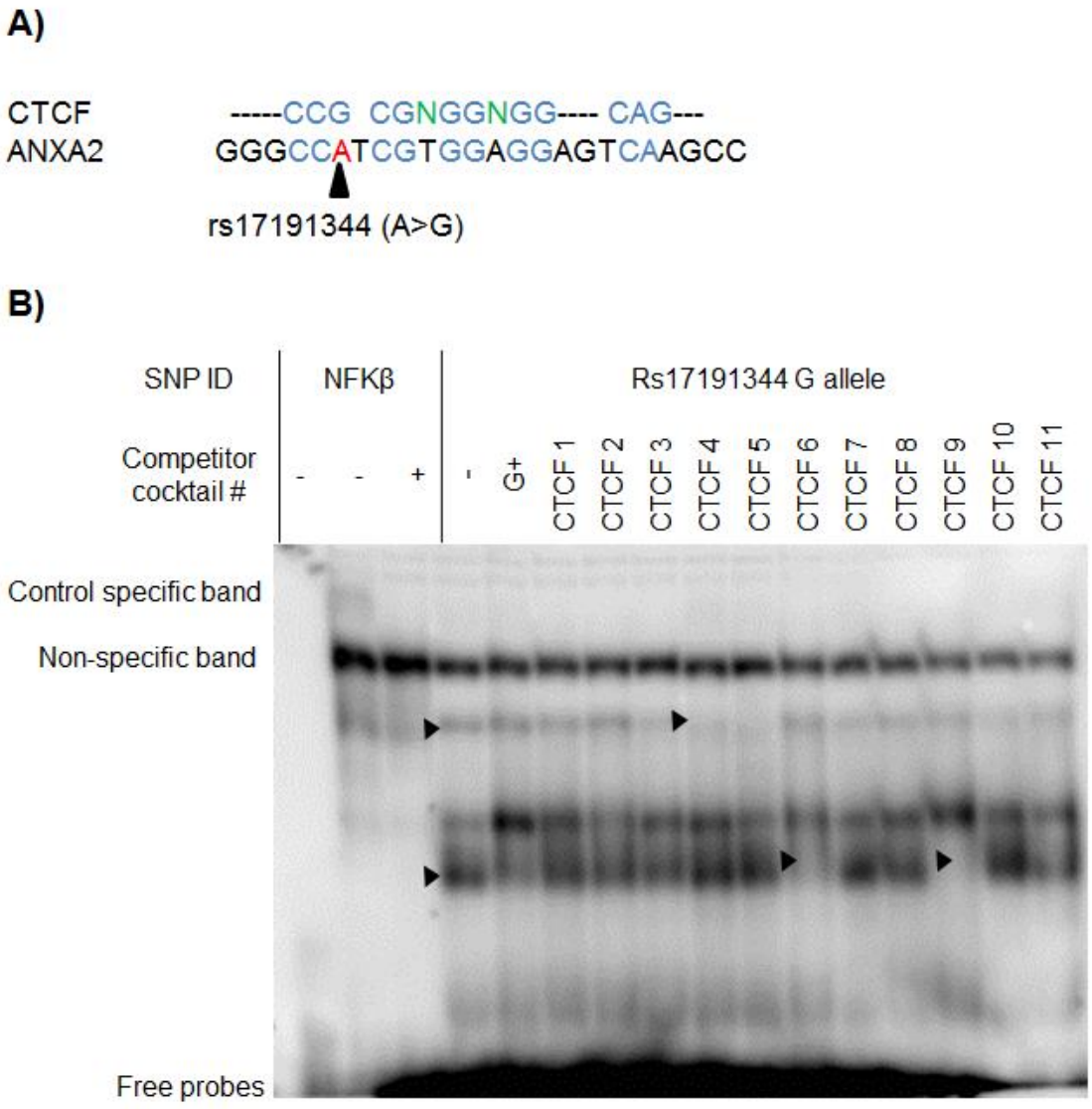


7.2.3.3. CCCTC-binding factor (CTCF) specific protein binding to *ANXA2* rs17191344

The bioinformatics analysis of the rs17191344 SNP suggests that the risk allele of the SNP is a site for CTCF protein binding. As CTCF has thousands of binding sites genome-wide, rs17191344 SNP was tested with 11 CTCF consensus sequence competitors, that are most common CTCF binding sites genome-wide (444). Comparing the CTCF-binding motif to the genomic sequence around the rs17191344 revealed that they matched up well, and the presence of the SNP risk G allele strengthened the binding motif (Figure 7-8 A). In the EMSA experiments, the rs17191344 G allele specific bands were competed out by at least three isoforms of CTCF indicated by arrows (Figure 7-8 B), suggesting that CTCF is the protein that binds to the sequence around the G allele of the SNP (439).

Figure 7-8: CTCF-specific protein binding to the *ANXA2* intergenic SNP rs17191344 in Huh7 Cells.

A) CTCF consensus sequences used in EMSAs were compared with genomic sequence around rs17191344. The single nucleotide polymorphism G allele matches more closely with the consensus sequence, which may increase the binding affinity to CTCF, suggesting CTCF may be responsible for protein binding here. B) EMSA analysis of the *ANXA2*-intergenic SNP rs17191344 G allele which is competed out by 11 isoforms of CTCF. The G allele specific bands, indicated by arrows, were eliminated by at least 3 isoforms of CTCF (eliminated bands indicated by arrows).



7.2.3.4. Supershift EMSA for confirmation of protein binding

MC-EMSAs and EMSAs implicated the CTCF protein as binding to rs17191344, while a portion of the GATA family and Egr1 bound to rs11633032. The supershift EMSA is a technique used to verify specific protein binding, where an antibody specific to putative DNA binding protein is incubated with nuclear extract before and after addition of labeled probed. If the antibody recognizes the target protein two results are possible. One possibility is that the antibody binds to the protein which then binds to the labeled probe, creating a larger molecular weight complex antibody-protein-DNA, which moves slower through the gel and will be observed as a band shift on the gel. A successful shift confirms the identity of the protein. Alternatively, the antibody may prevent DNA/ protein interaction because it binds to an essential site on the protein required for DNA binding, resulting in a loss of the specific complex and no band shift on the gel.

Results of *in-silico* and EMSA implicated the CTCF protein as binding to rs17191344 in the presence of the G allele; supershift EMSA was performed to verify protein binding. Figure 7-9 shows that the G allele probe produced a band as expected and was competed out by the unlabeled probe (lanes 4-5), however, a band shift or band absence was not observed with the addition of antibody (lanes 6-7). However, the specific bands in lanes 6-7 had higher intensity.

MC-EMSA implicated the rs11633032, which binds to Egr1 and a portion of the GATA family. The proteins in the GATA family share very similar consensus sequences, thus

MC-EMSA could not identify which proteins were binding. Three GATA proteins were examined (GATA1 and GATA2+3); these isoforms have emerged as candidate regulators of gene expression for either ANXA2 or liver associated genes (445). Supershift EMSA was performed on the rs11633032 G probe. Again the rs11633032 G allele probes produced a band as expected that was competed out by an unlabeled probe (Figure 7-9, lanes 7-8). When the GATA1 antibody was added, no differences were seen in lanes 8, 9 and 10. However, with the addition of GATA2+3 (lanes 11-12) and Egr1 (lanes 13-14) antibodies, the G allele specific band was eliminated, which indicates that the antibodies interfered with the interaction of nucleoprotein complexes, because the antibodies interact with essential sites on the protein required for DNA binding and prevent shift complex forming.

The consensus sequences of protein binding motifs were compared to the DNA sequence around rs11633032 (Figure 7-10), and the binding sites of three proteins, GATA2, GATA3 and Egr1 were found to overlap with the SNP of interest. Thus, nucleotide substitution may affect protein–DNA binding affinity. This may explain the absence of specific bands of the G allele with addition of the three antibodies. If these antibodies bound to core (A/T) GATA(A/G) nucleotide motif, then the rs11633032 G probe would be unable to bind to protein. However, it is important to note that Egr1 is most likely to be allele specific binding protein as the presence of the G allele of the SNP may increase the binding affinity to Egr1. In contrast, GATA1 binding motif binds to the sequence beside the SNP but does not overlap with the SNP itself.

Figure 7-9: Supershift EMSA of *ANXA2* intergenic SNPs: rs17191344 and rs11633032

Supershift EMSA shows the effect of adding CTCF to rs17191344, and GATA and Egr1 to rs11633032. The probe specific bands are indicated by arrows. The SNP labeled probes produced a band which was eliminated following addition of an unlabeled probe. Addition of the antibodies did not produce a band shift or band absence. * indicating the antibodies that were added after the complex formed.

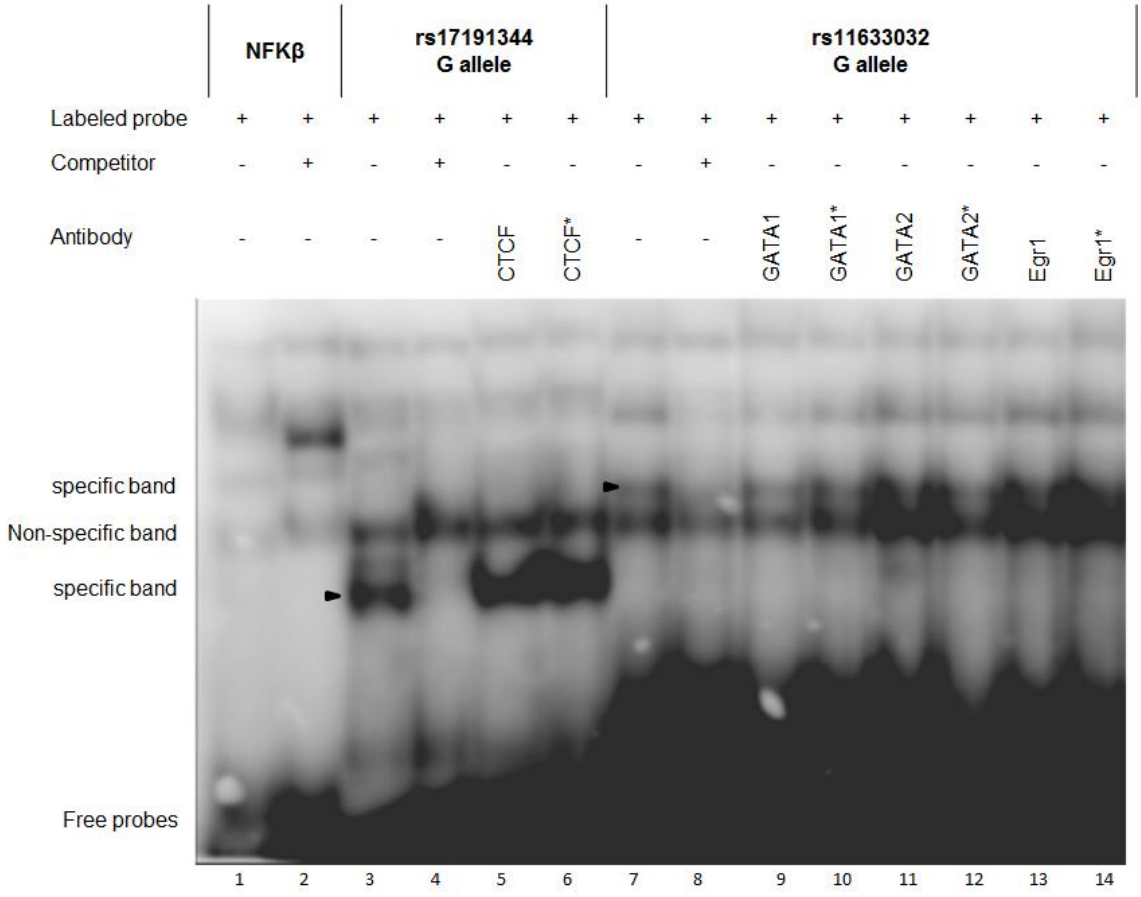
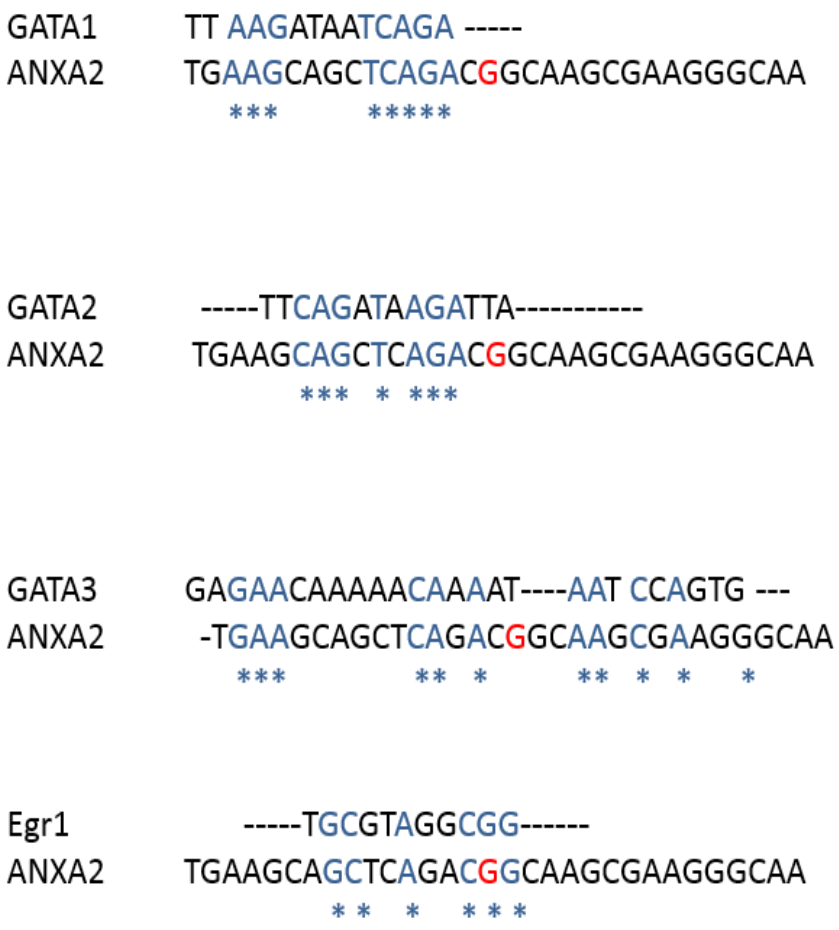


Figure 7-10: Nucleotide sequence overlap between rs11633032 G allele and suggested the specific allele protein binding

GATA protein family including GATA1, GATA2, and GATA3 and Egr1 consensus sequences of binding motif were compared with genomic sequence around rs11633032. The SNP G allele matches more closely with the consensus sequence of Egr1, which may increase the likelihood of this protein being the allele specific binding protein. The ANXA2 rs11633032 G allele is indicated in red, while the overlap sequences are indicated in blue and by stars.



7.2.4. Effect of rs17191344 and rs11633032 on reporter gene expression

Luciferase reporter assays were performed to assess whether the rs17191344 and rs11633032 SNPs genotypes affect gene expression.

7.2.4.1. *ANXA2* promoter

To determine the effect of identified intergenic SNPs on gene expression of *ANXA2*, it is ideal to use the pGL3-basic luciferase reporter vector and insert an *ANXA2* promoter DNA fragment (1520bp) upstream of the Luciferase gene (*luc*+) at the promoter site. However, amplifying a 1520bp fragment of *ANXA2* promoter was impossible due to the presence of a sequence of short tandem repeats (STRs) including one ALU repeat, and three LINE repeats, which were interspersed in 823bp (54.14%) of the promoter (Figure 7-8). These were identified using RepeatMasker software (446). In PCR amplification, primers bind to the DNA at specific loci and are extended by PCR to produce full length PCR product. However, the PCR product length is fragmented in the presence of STRs. The length of the PCR product depends on the number of repeats. *In-silico* results have suggested that the *ANXA2* promoter amplification could produce four bands: 123bp, 405bp, 250bp and 45bp (Figure 7-12), which was confirmed by PCR. The PCR product of the *ANXA2* promoter had three fragment sizes similar to what was predicted by the *in silico* tool (Figure 7-11). Thus due to the difficulty in amplifying *ANXA2* promoter, the pGL3-promoter luciferase reporter vector was used for further reporter gene expression.

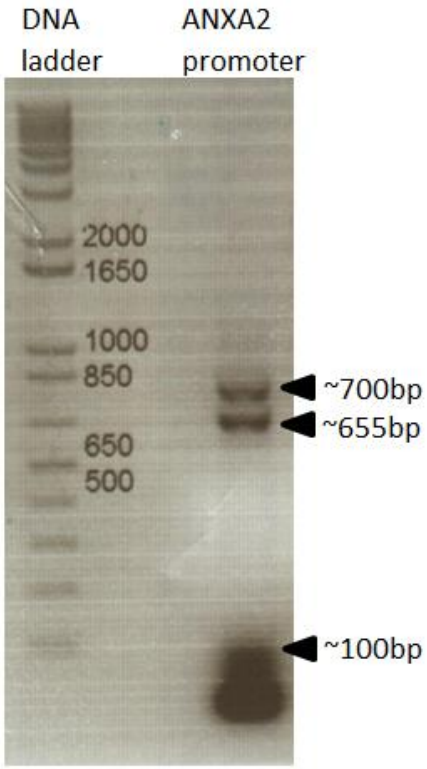
Figure 7-11: Presence of a sequence of short tandem repeats (STRs) in the *ANXA2* promoter

RepeatMasker tool predicted that the *ANXA2* promoter contains two type of STRs: SINEs (ALUs) and LINEs (LINE1, LINE2, and L3/CR1). This suggested that DNA amplification could generate four DNA fragments (123bp, 405bp, 250bp and 45bp)

```
==================================================
File name: RM2sequpload_1425982319
Sequences:                1
Total length:         1520 bp  (1520 bp excl N/X-runs)
GC level:            36.18 %
Bases masked:          823 bp  (54.14 %)
==================================================
                Number of      length      percentage
                elements*    occupied of
                              sequence
--------------------------------------------------
SINEs:                  1           123 bp     8.09 %
      ALUs              1           123 bp     8.09 %
      MIRs              0             0 bp     0.00 %

LINEs:                  3           700 bp    46.05 %
      LINE1             1           405 bp    26.64 %
      LINE2             1           250 bp    16.45 %
      L3/CR1            1            45 bp     2.96 %

LTR elements:           0             0 bp     0.00 %
      ERVL              0             0 bp     0.00 %
      ERVL-MaLRs        0             0 bp     0.00 %
      ERV_classI        0             0 bp     0.00 %
      ERV_classII       0             0 bp     0.00 %

DNA elements:           0             0 bp     0.00 %
     hAT-Charlie        0             0 bp     0.00 %
     TcMar-Tigger       0             0 bp     0.00 %

Unclassified:           0             0 bp     0.00 %

Total interspersed repeats:         823 bp    54.14 %


Small RNA:              0             0 bp     0.00 %

Satellites:             0             0 bp     0.00 %
Simple repeats:         0             0 bp     0.00 %
Low complexity:         0             0 bp     0.00 %
==================================================

* Most repeats fragmented by insertions or deletions
  have been counted as one element
```

Figure 7-12: PCR products of *ANXA2* promoter

PCR amplification of ANXA2 promoter (1520bp) shows that three short tandem repeats (STRs) were present in the sequence, which led to fragmentation of the PCR product into three fragments (bands indicated by arrows). 1kb plus a DNA ladder was used as molecular size markers.



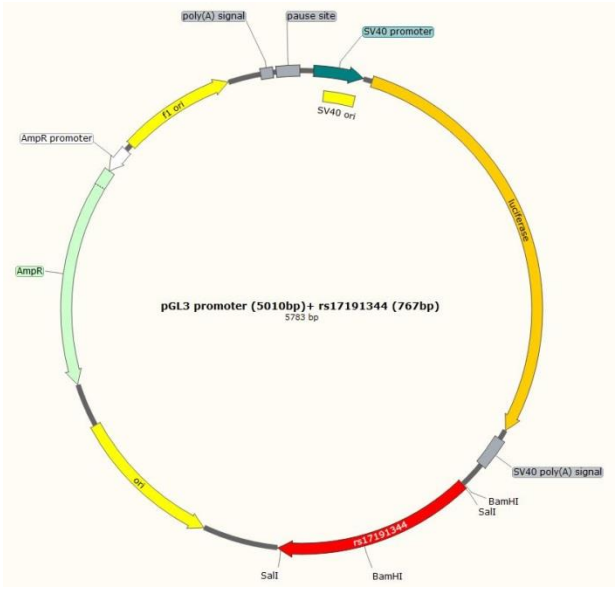
7.2.4.2. The impact of *ANXA2* SNPs rs11633032 and rs17191344 on reporter gene expression

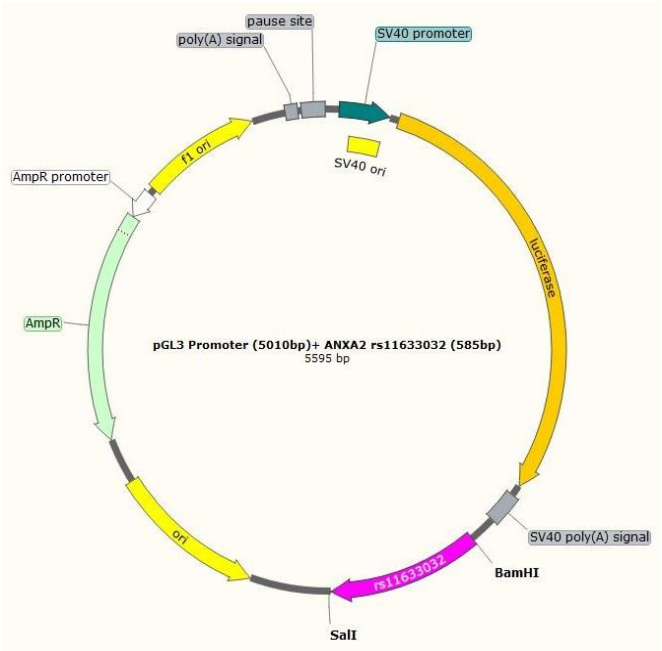
The *ANXA2* SNPs rs11633032 (593bp) and rs17191344 (776bp) fragments containing either allele of the SNP were inserted downstream of the luciferase gene in the pGL3-promoter luciferase reporter vector (Figure 7-13 A). The inserted fragment in the pGL3-promoter vector resulted in a decrease of expression compared to the control vector (pGL3-promoter) for both alleles (Figure 7-13 B). However, the presence of minor alleles caused approximately 18% further significant decrease of gene expression in rs11633032 ($p=0.91x10^{-3}$) and rs17191344 ($p=0.27x10^{-3}$). This suggests that the sequences around rs11633032 and rs17191344 are sites for repressor protein binding (439).

Figure 7-13: *ANXA2*- intergenic SNPs luciferase activity in Huh7 cell line.

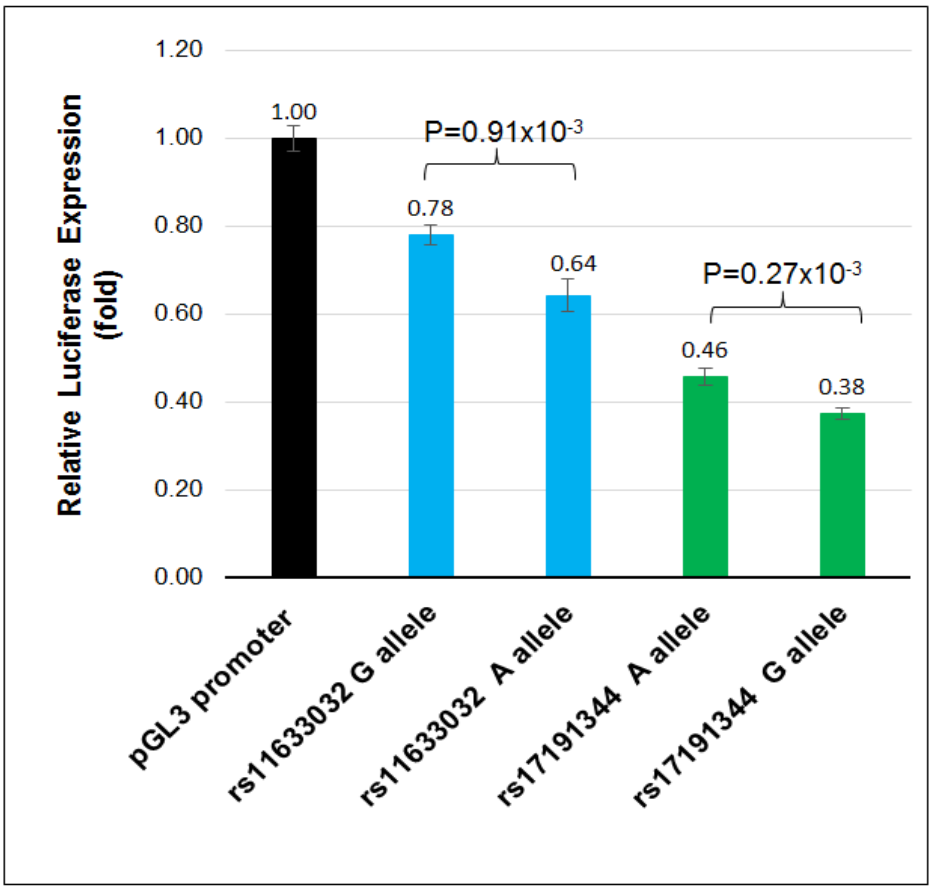
A) A map of the pGL3-promoter constructs showing the location of the added SNP fragments of rs11633032 (585bp) and rs17191344 (767bp). B) The luciferase reporter assays show relative expression of *ANXA2*- intergenic SNPs-luciferase-enhancer-constructs (sequence has SNP of interest inserted into enhancer site of the vector) relative to the control (pGL3-promoter vector).

A)





B)


Relative Luciferase Expression (fold)
1.20
1.00
0.80
0.60
0.40
0.20
0.00
1.00
0.78
0.64
0.46
0.38
P=0.91x10-3
P=0.27x10-3
pGL3 promoter
rs11633032 G allele
rs11633032 A allele
rs17191344 A allele
rs17191344 G allele

7.2.5. Expression Quantitative Trait Loci (eQTL) analysis

To assess genotype-gene expression level associations, three expression quantitative trait loci (eQTL) catalogues were used. The first was the Genotype–Tissue Expression (GTEx) portal (285, 286) which calculated the association between rs11633032 genotype and expression of the four closest genes to the SNP in different tissues. The second was the Advanced Study of Aortic Pathology (ASAP) study (310). The third catalogue was the eQTL meta-analysis of blood for lipid-regulatory and immune-mediated disease variants (447).

7.2.5.1. The Genotype–Tissue Expression (GTEx)

To determine whether *ANXA2* intergenic SNPs were associated with altered gene expression *in human*, the publicly available gene expression data set, GTEx, was used. The four genes (Forkhead box B (*FOXB), ANXA2,* RAR-related orphan receptor alpha (*RORA*) and Lipase C (*LIPC*)) near the SNPs were tested in GTEx, but no significant association was found between the SNPs' genotype and gene expression in the liver, whole blood and arteries (Table 7-8). However, subjects with one copy of the risk allele for rs11633032 had lower expression of *ANXA2* in all tissues including liver, whole blood and coronary artery (Effect size = -0.062, -0.039 and -0.20, respectively) (Figure 7-14), but this effect was not statistically significant ($p>0.05$) due to the small sample size, with only three to-six subjects being homozygous for the minor allele. However this effect was confirmed in the latest unpublished GTEx data, which showed that the rs11633032 minor allele was associated with significantly reduced *ANXA2* expression in a tibial artery sample (effect size -0.177, $p=2.9x10^{-06}$, N=285) under additive model (439).

Table 7-8: The influence of *ANXA2* rs11633032 on nearby gene expression

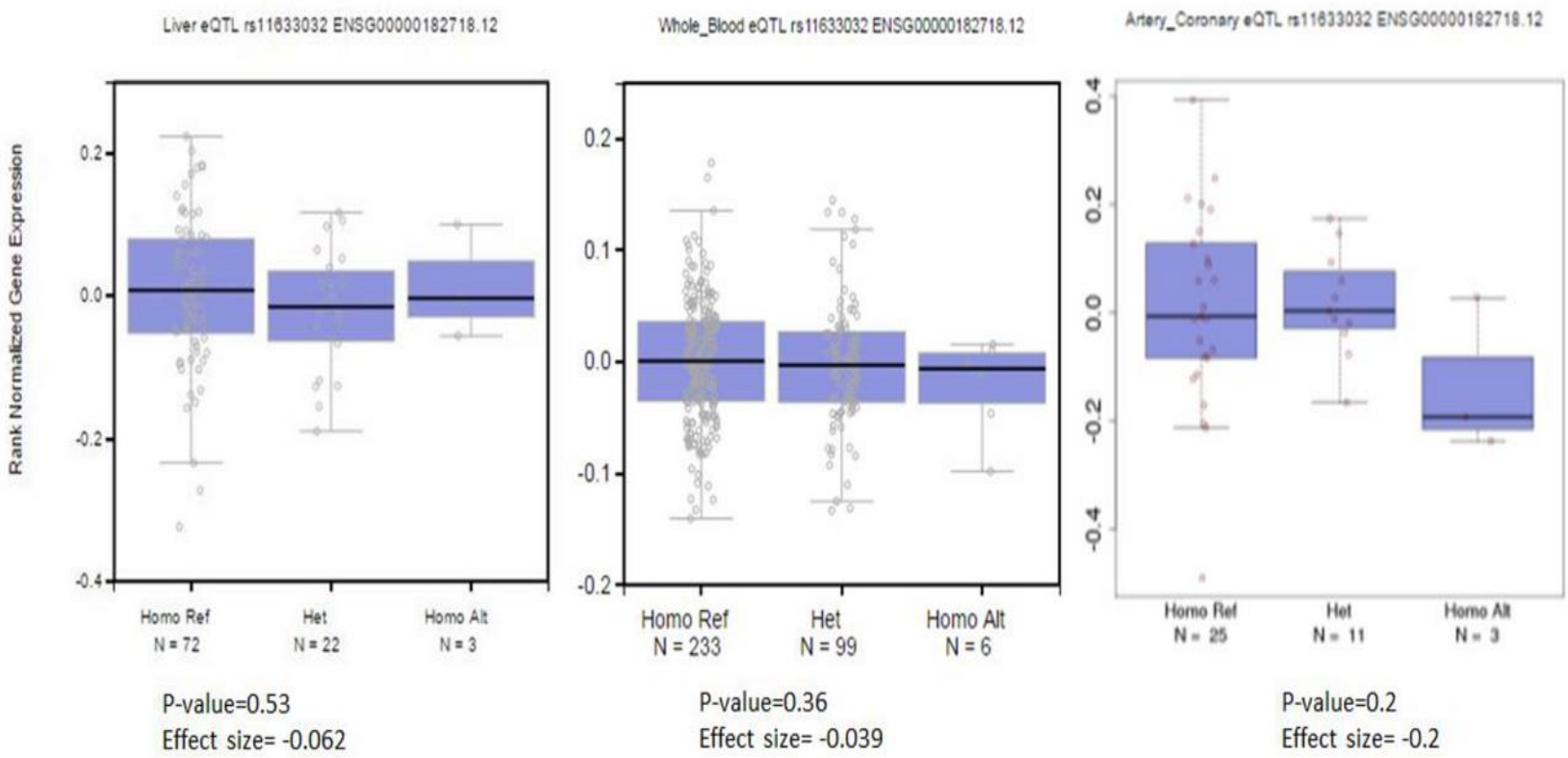
Using the GTEx public database, five genes close to rs11633032 SNP were analyzed (NARG2, *ANXA2, RORA, LIPC* and *FOXB* (no data found) in five different tissues including liver, whole blood and arteries (aortic artery, coronary artery, and tibial artery)

Gene Symbol	p-value	Effect Size	Tissue
NARG2	0.069	-0.12	Artery - Aorta
	0.35	-0.097	Artery - Heart
	0.81	-0.013	Artery - Tibia
	0.4	-0.14	Liver
	0.094	0.080	Whole blood
ANXA2	0.17	-0.094	Artery - Aorta
	0.20	-0. 2	Artery - Heart
	0.33	-0.047	Artery - Tibia
	0.53	-0.062	Liver
	0.36	-0.039	Whole blood
LIPC	0.33	-0.12	Artery - Aorta
	0.46	-0.091	Artery - Heart
	0.57	0.051	Artery - Tibia
	0.75	0.039	Liver
	0.70	0.038	Whole blood
RORA	0.51	0.048	Artery - Aorta
	0.83	-0.023	Artery - Heart
	0.64	0.024	Artery - Tibia
	0.58	0.072	Liver
	0.042	0.056	Whole blood

NMDA receptor regulated 2 (NARG2), Annexin A2 (ANXA2), Lipase C (LIPC), and RAR-related orphan receptor alpha (RORA)

Figure 7-14: *ANXA2* expression by rs11633032 genotype in human tissues.

Data and graphs from the Genotype–Tissue Expression (GTEx) Portal. The tissue-specific association between rs11633032 and genotype and expression level of *ANXA2* in three different tissues (liver, whole blood, and coronary artery) was investigated. Although no significant association was found due to sample size, the SNP genotype showed the same expression trend in all three tissues, where carriers of the minor alleles had lower gene expression.



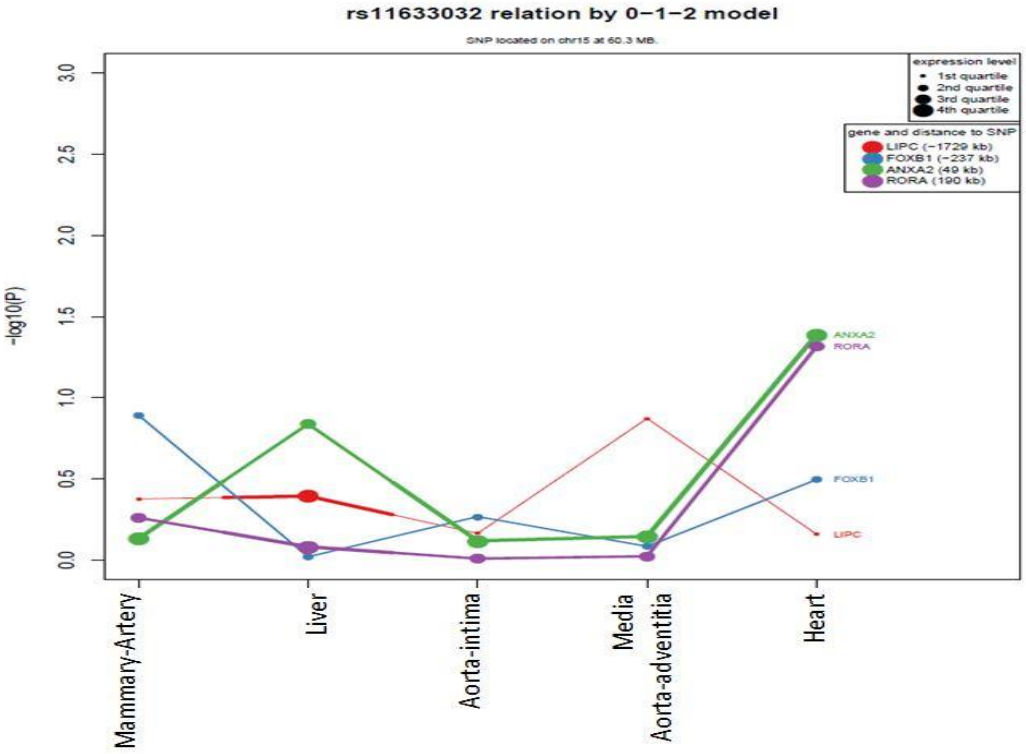
7.2.5.2. The advanced study of aortic pathology (ASAP) study

To examine further the effect of rs11633032 on ANXA2 expression in different human tissues, the ASAP study was used, where all genes within 628 kb were analyzed in cardiovascular and metabolic relevant tissues, including liver, aorta (medial and adventitial portions), mammary artery, and carotid plaque. This showed that the rs11633032 SNP genotype affected *ANXA2* expression mostly in the heart and to a moderate level in the liver (Figure 7-15 A). In liver, again the rs11633032 SNP risk allele (A) was associated with a reduced expression level of *ANXA2* (p=0.075) in the additive model (Figure 7-15 B) (439). Applying the recessive model to our study was unsuitable because of the small number of cases carrying two copies of the risk allele and the modest effect of SNP on expression. Although there was a lack of a significant association, the ASAP result was consistent with results found in other data.

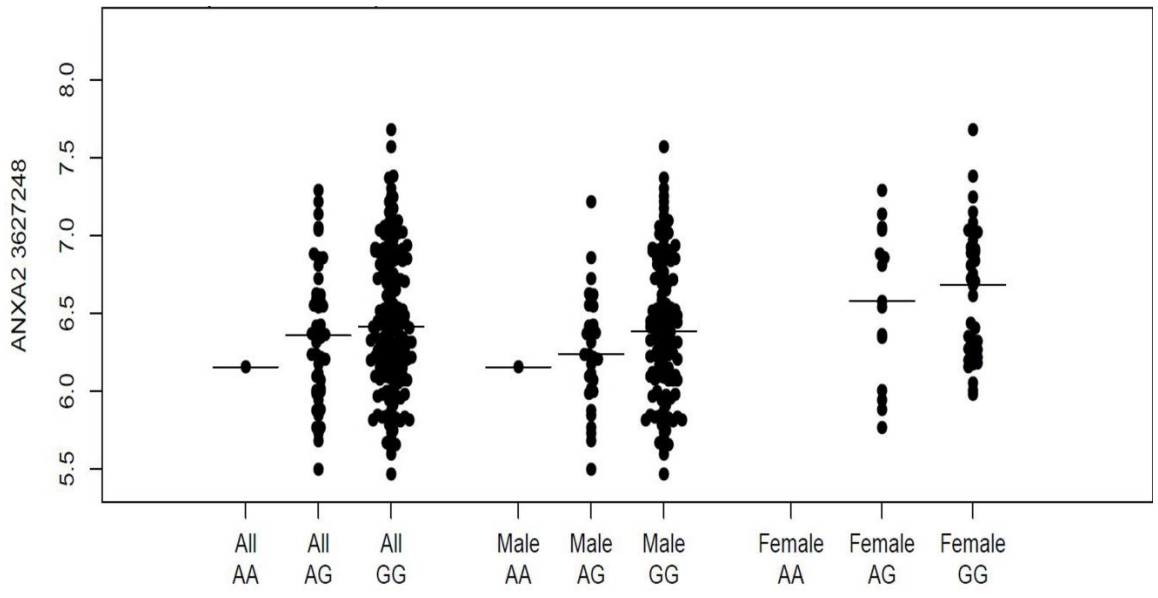
Figure 7-15: Gene expression by rs11633032 genotype in the Advanced Study of Aortic Pathology (ASAP) study

A) Tissue-specific association between the rs11633032 genotype and expression level of all genes within 628 kb of the locus. In the plot, the y-axis shows the log10 (P) calculated for the association between genotype and expression level using an additive model. The x-axis shows the five different tissues examined: Mammary-Artery, Liver, Aorta-intima, Media Aorta-adventitia, and Heart. Each colored line shows the association level for one gene across different tissue types (*ANXA2* represented by a green line). The relative expression level of a gene in the tissue type is indicated by the dot size. B) Additive linear models used to detect the association between SNP and *ANXA2* expression levels in liver tissue. Gene expression measurement in a patient with rs11633032 genotype, which can be AA, AG and GG. Expression data is shown for the whole sample, men and women.

A)


rs11633032 relation by 0−1−2 model
SNP located on chr15 at 60.3 MB.
expression level
1st quartile
2nd quartile
3rd quartile
4th quartile
gene and distance to SNP
LIPC (−1729 kb)
FOXB1 (−237 kb)
ANXA2 (49 kb)
RORA (190 kb)
−log10(P)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
ANXA2
RORA
FOXB1
LIPC
Mammary-Artery
Liver
Aorta-intima
Media
Aorta-adventitia
Heart

B)


ANXA2 3627248
5.5
6.0
6.5
7.0
7.5
8.0
All AA
All AG
All GG
Male AA
Male AG
Male GG
Female AA
Female AG
Female GG

7.2.5.3. The eQTL meta-analysis for lipid-regulation

To further verify these findings, the publicly available eQTL meta-analysis for lipid-regulation (http://genenetwork.nl/bloodeqtlbrowser/) was used which found that the proxy SNP rs9920796 (r^2=0.736) was significantly associated with lower *ANXA2*-mRNA expression levels (Z-score= -4.35, p=1.36 $x10^{-05}$). The Z-score is the number of standard deviations away from the mean. In other words, the Z-score shows the relationship to the mean in a group of values. For example, a Z-score of 0 shows that the score is identical to the mean score, but if the Z-score has a negative value (-0.0 to -4.0), it indicates a score below the mean.

7.3. Discussion

The two studies that investigated the *ANXA2*-R1 missense variant p.(Val98Leu) in chapter six showed the substantial heterogeneity of the association of the SNP with LDL-C. In addition, the SNP was predicted to be non-pathogenic and in modest LD in the non-coding region. These results suggested that the *ANXA2* coding SNP may be a marker for SNPs in non-coding region which may have an influence on gene expression. Using bioinformatics, genotype-phenotype analysis, and evidence from differential protein binding and allele-specific gene expression assays, two candidate SNPs were identified in the *ANXA2* cis-regulatory region (rs17191344 and rs11633032). Both cis-regulatory SNPs were associated with LDL-C and the risk of CHD, and showed that these SNPs affect *ANXA2* gene expression via alterations in transcription factors that bind to alleles of the SNPs. Using data from population studies, this work has confirmed for the first time the observations in cell culture and AnxA2 knockout mice, where changes in the levels of AnxA2 directly influence plasma LDL-C levels in human, and thus implicate this protein as a potential therapeutic target for LDL-C lowering.

The effect of rs17191344 and rs11633032 on cholesterol levels was determined in the NPHSII and UCLEB consortium. These SNPs were associated with LDL-C level and CHD risk with a recessive mode of inheritance, where two copies of the minor allele led to significantly higher levels of LDL-C and higher risk of CHD (439). This finding is similar to the effect seen by Val98Leu SNP. This raised the question as to whether the observed phenotype was the result of the effect of a non-coding SNP (rs17191344) or was due to the combined effect of coding SNP and non-coding SNP. This was assessed

by examining cholesterol levels and risk of CHD for these individuals with different combinations of genotypes. It was found that individuals carrying two copies of the minor allele for both SNPs had the highest LDL-C levels and the highest risk of CHD. However, only a small number of individuals carry the minor alleles, and they had modestly elevated cholesterol levels and CHD risk, suggesting that both the amino-acid change and the intronic SNPs are functional. Although we reported that the intergenic SNPs did not display a significant difference in cholesterol levels between sexes, women appeared to be less affected by these variants. It has been reported that sex and age have an influence on PCSK9 concentration and consequently LDL-C levels (448), where women have higher levels of PCSK9 and postmenopausal women exhibit even higher concentrations than premenopausal women (449, 450).

Although the GWAS reveals many novel loci associated with lipid levels (161), it is unclear why the *ANXA2* locus has not been previously identified as a GWAS hit for LDL-C (161), but there are several possibilities to consider. First, GWAS need to set stringent cut-off values for association ($p < 10^{-8}$) to reduce the possibility of false positive associations, although this could lead to false negative results and some of the disease-associated variants of modest effect could be missed. In addition, most of the GWAS published data used an additive inherited model in their statistical analyses, which may miss trait-associated variants that have recessive inheritance modes. It may also have been that this variant (or the *ANXA2* locus) was not well covered in the GWAS genotyping chips.

Bioinformatics databases are particularly powerful tools to select candidate SNPs for functional studies. The Val98Leu SNP has a modest LD with 34 SNPs, all located downstream of the *ANXA2* gene-coding region in the long intergenic region. Such regions often comprise cis-regulatory sequences that have a role in chromatin organization and transcription regulation (451). Cis-regulatory sequences could be a hit for enhancer elements that have an important role in gene transcription. Enhancers contain multiple short binding site sequences for activators and repressors that regulate gene transcription (452). One gene can have multiple enhancers in order to facilitate gene activation; an enhancer can be located anywhere within a span of 100kb upstream or downstream of a gene whose expression it regulates.

Data from ENCODE and RoadMap Epigenomics were used to evaluate whether the 34 SNPs were located in the cis-regulatory element regions and whether they had regulatory potential. A shortlist of three potentially functional variants identified SNPs rs17191344, rs11633032 and rs12900101. The regulatory markers showed a number of characteristics: a) chromatin immunoprecipitation sequencing demonstrated peaks of transcription factor binding sites at these SNPs, which suggested that an allele of the variants might affect transcription factor binding and actions; b) histone methylation and acetylation marks showed areas consistent with enhancers; c) FAIRE and DNase hypersensitivity showed regions of open chromatin, which are more likely to be involved in regulating expression; and d) the ENCODE data and ElDorado tool showed that the variants might either facilitate or block transcription factor binding.

The *in silico* predictions of protein binding to the selected SNPs were confirmed by *in vitro* assays. Using EMSA, both rs11633032 and rs17191344 showed strong allele-specific protein binding and several potential transcriptional factors involved in the regulation of *ANXA2* expression levels were identified. MC-EMSA suggested that the Egr1 and GATA proteins were implicated in binding to the sequence around the G allele of rs11633032 (439). Egr1 is a nuclear factor that regulates gene expression in a tissue-restricted manner through its binding to other regulatory transcription factors (453). The protein family GATA comprises six members (GATA1 to GATA6) all of which have a highly conserved double zinc finger domain that mediates binding to DNA and co-factors to regulate gene expression in a highly tissue-restricted fashion. GATA recruits chromatin remodeling complex and mediates either repression or activation of target genes (454, 455).

EMSA also showed that CTCF binds to the DNA sequence around rs17191344 in the presence of the G allele (439). CTCF is considered to be an insulator element and plays a critical role in transcriptional regulation. There are two possible functions of an insulator in gene regulation. First, it binds to DNA-regulatory sequences in the promoter-proximal regions where it competes for enhancer-bound activators and prevents the activation of downstream promoters (444, 456). In addition, insulators may be involved in gene regulation by facilitating the formation of separate loop domain structures, which prevent an enhancer on one loop from contacting with a promoter on a different loop (457, 458). Despite the identification of above mentioned transcription factors GATA, Egr1 and CTCF binding AnxA2 mRNA *in vitro*, future work will need to validate their

contribution to the regulation of AnxA2 expression levels in hepatic cell lines. In particular, overexpression or silencing of the transcription factor CTCF could provide further insight of whether rs11633032 and rs17191344 can modulate the deduced repressor functions of this transcription factor.

A supershift assay was performed to investigate which DNA binding proteins are involved. The order of addition of reaction components (extract, oligos, and antibodies) is critical in these assays. An antibody may interfere with the interaction of nucleoprotein complexes and can lead to band elimination, thus the antibody is added after nucleoprotein complexes are formed. In this study, the antibody was added either before or after complexes were formed, but unfortunately none of the candidate proteins were successfully identified, that is, no band shift was seen following addition of antibody. However, the addition of the CTCF antibody to the rs17191344 G probe showed 50% higher band intensity than the rs17191344 G probe alone. A possible explanation is that the larger molecular weight of the complex (antibody-protein-DNA) (459) may have affected the complex's mobility in the gel and the shift band may have been hidden in the binding band, suggesting that a lower gel concentration or softer gel may help band mobility.

The supershift assay with rs11633032 G probe was performed with 3 different antibodies (GATA1, GATA2+3, and Egr1). GATA1, GATA 2 and GATA 3 are members of the GATA family proteins, that bind to the (A/T)GATA(A/G) nucleotide motif with a high

degree of amino acid sequence identity (455). These GATA antibodies were chosen as they have been implicated in *ANXA2* expression and in liver associated gene expression. It has been reported that GATA 2 is a proximal co-regulator transcription factor of the *ANXA2* gene (http://encodenets.gersteinlab. org/index.html#; (445). GATA 2 and GATA 3 regulate liver and cholesterol associated genes (460), while GATA1 is expressed in fetal liver (461). The GATA2+3 and Egr1 antibodies eliminated the specific band which suggests that the antibodies interact with essential sites on the protein required for DNA binding, while GATA1 antibody showed no band shift. Overall, as the antibody is expected to bind to its corresponding protein, the lack of shift suggests that the antibodies used did not bind to the protein under EMSA conditions.

The luciferase reporter assay was used to assess the mechanism and action of the SNPs and evaluate how different protein binding may affect gene expression. It was found that the minor allele of both SNPs rs11633032 and rs17191344 reduced gene expression (439). Such lower gene expression in carriers of the risk A allele of rs11633032 was confirmed using *in human* expression data from GTEx, ASAP and from the large eQTL meta-analysis.

Overall these findings suggest that the sequence around the minor alleles of *ANXA2*-intergenic SNPs rs11633032 and rs17191344 are sites for transcriptional repressor proteins that lead to reduced *ANXA2*-mRNA expression and protein levels (439). However, it is as yet unknown whether with such low levels of expression and in the

presence of high-affinity AnxA2 binding partners in plasma, such as plasminogen or tPA (433), AnxA2 is still able to interact with PCSK9 and inhibit PCSK9 acting on LDL-R, leading to higher levels of LDL-R receptors in the liver and consequently, reduced plasma LDL-C levels.

7.4. Limitations

This study had limitations. The first limitation was that although the *ANXA2* coding SNP was thought to be functional, no data addressed directly whether or not the Val98Leu change affects AnxA2 structure and function. I am unaware of any data that has addressed the functional consequences of this particular mutation. One study (102) has presented preliminary evidence that V98L does not affect the binding affinity of AnxA2 for PCSK9, but it is associated with lower circulating PCSK9 and may be causing lower LDL-C levels. The authors suggest that further studies are needed to examine whether this mutation modifies the function of PCSK9 or has downstream consequences on LDL-R activity. In addition to the possibility that the coding SNP may affect protein structure, it also may be a site for a cis-regulatory element (462, 463). The sequence around the coding SNP rs17845226 is located in a DNAse I hypersensitive domain, thus it may be a site for positive-acting regulatory sequences which could interact with intergenic functional SNPs and *ANXA2* promoter to initiate gene transcription. A second limitation is that it was not possible to impute the genotype in UCLEB for rs17845226, and that the examined SNPs were imputed with a relatively low degree of precision. Since they were associated with LDL-C levels in the UCLEB cohort, it is possible that examining these SNPs with wet-lab genotyping may help to see larger mean effect on LDL-C levels.

The third limitation is *in vitro* data. These data showed a regulatory element near to rs11633032 and rs17191344 SNPs which acted as a repressor of *ANXA2* expression in the liver cell line studied, chosen because liver is the major site for LDL-C clearance from the plasma (464). However, the impact of these SNPs on gene expression in other tissues was not examined. AnxA2 levels in liver are generally considered low (102), but in line with this study using Huh7 as a model system, stable knockdown of AnxA2 expression in Huh7 resulted in PCSK9 upregulation and a marked reduction in LDL-R levels (192). I am aware that Huh7 may not be representative of all human hepatocellular carcinoma cell lines as AnxA2 depletion in HepG2 cells, which expresses approximately 5-fold less AnxA2 mRNA compared to Huh7, did not alter PCSK9 or LDL-R protein expression. However, this study elucidates a potentially new mechanism, suggesting a role for AnxA2 in the translational control of PCSK9 protein levels (192). The overall contribution of transcriptional and translational regulation of AnxA2 expression, which affects PCSK9 maturation and protein levels in the extracellular space and during PCSK9 synthesis and secretion, has yet to be determined *in vivo*. Another unavoidable limitation is that *in vitro* EMSA and luciferase assays were used to demonstrate the functional role of SNPs using the hepatoma cell line Huh7. These two *in vitro* assays can only approximate the actual gene expression occurring *in vivo*, where environment and open chromatin modification and interaction play essential roles in mediating gene expression.

The fourth limitation is linked to SNP selection. The *ANXA2* cis-regulatory SNPs, rs11633032, rs17191344 and rs12900101, which were studied, have strong LD with 64 SNPs, all in the intergenic region. Here the ENCODE data, a summary tool, HaploReg

V4 and *in silico* ElDorado data were used to select potential functional candidate SNPs. However, the other LD SNPs not examined may have a role in transcriptional regulation. This kind of limitation is unavoidable because there is no conclusive tool to rank likelihood functionality of non-coding variants. There are some annotation programs such as combined annotation dependent depletion (CADD) (465) and RegulomeDB (466) that provide their own ranking-score using growing genomic annotation data from the ENCODE and Roadmap projects, but the ranking varies between programs with no definite assumptions.

The fifth limitation was associated with the supershift assay. It is an important assay that provides another level of certainty of protein-DNA binding. However; in this study some difficulties were faced with antibodies, complex size, and EMSA conditions. Addressing these issues needs more time and resources, which were very limited during this study.

7.5. Conclusion

In summary, AnxA2 has been previously linked to cholesterol homeostasis and CHD risk via its interaction with PCSK9. Integration of bioinformatics with disease-association analysis and *in vitro* studies is a powerful tool, and has led to identification of functional SNPs in *ANXA2* and its cis-regulatory domain. Although there was a lack of consistency in the association analysis data of the *ANXA2* coding SNP in the previous chapter, further analysis of the coding SNP with a proxy SNP has shown that coding SNPs may be functional and the negative prediction of SNP functionality may be incorrect. As in many

bioinformatics prediction tools, they have a proportion of false negative results. Therefore, further biological functional studies need to verify whether the coding SNP has a role in protein structure or function.

Despite the unknown functional role of the *ANXA2* coding SNP, the SNP has LD with SNPs in the cis-regulatory region. This study used bioinformatics and functional assays to identify a potential cis-regulatory variant associated with the chromosome 15 locus and in turn with LDL-C levels and CHD risk. However, identifying the functional SNPs was another challenge due to LD. The selected SNPs of modest LD showed that these SNPs are sites for transcription protein binding and caused expression alteration, suggesting *ANXA2* is a potential therapeutic target. The risk allele reduced the expression levels of *ANXA2*, lowered *ANXA2*-mRNA expression and is thus predicted to result in less protein. Low levels of AnxA2 allow PCSK9 to increase LDL-R degradation and thus increase LDL-C levels and the risk of CHD.

Chapter 8 : Identification of functional variants in the *ANXA2*-3'UTR

8.1.Introduction

MicroRNAs (miRNAs) are small non-coding single strands of RNA about 22 nucleotides in length, which regulate gene expression through the post-transcription regulation machinery (261). MiRNA genes are transcribed via RNA polymerase II to generate pri-miRNAs that have multiple hairpin loop structures (Figure 1-28). Pri-miRNAs are processed by the Drosha complex to produce smaller precursor molecules called pre-miRNAs. Then, pre-miRNAs are exported from the nucleus via exportin-5 to the cytoplasm (467). In the cytoplasm, pre-miRNAs are processed via Dicer to generate short double-stranded miRNAs, which are then converted to mature, single-stranded miRNAs via RISC, the mature miRNAs interact with 3'UTR of the targeted mRNA (259, 260). Complementarity between the miRNA and a seeding site within the 3'UTR of a target mRNA regulates the post-transcriptional fate of the mRNA, where perfect complementarity leads to the degradation of the target mRNA, while imperfect complementarity leads to repressed translation (261). The seeding site is a target sequence in 3'UTR-mRNA where miRNAs are able to recognize and hybridize, using 6–8 nucleotides either at the 5' or 3' region of the miRNA.

A variant such as a SNP within a miRNA, or at a miRNA seed site on the targeted mRNA, may create or destroy miRNA/mRNA binding sites and thus affect the translation of the mRNA, the production of the encoded protein and ultimately a disease trait (e.g. LDL-C) or a disease phenotype endpoint (e.g. CHD) (268). It has been reported that SNPs at miRNA target sites have been associated with an increased risk of various complex diseases including cancer, and cardiac and other diseases (468). Identifying and selecting miRNA/mRNA potential candidates for experimental validation is challenging due to the nature of miRNA/mRNA target interaction

prediction tools, that usually provide a very large number of potential target sites that then need to be validated (469). miRNAs have a tissue-specific expression signature and their actions are controlled by physiological and pathological environments (470).

ANXA2 expression is regulated at both the transcriptional and translational level (188). For instance, ERG binds to a DNA sequence upstream of the *ANXA2* gene, and acts as a repressor to down-regulate *ANXA2* expression in prostate epithelial cells (442). Also, in chapter seven I showed that CTCF binds to a DNA sequence downstream of the *ANXA2* gene in the long intergenic region, and represses *ANXA2* expression. In addition, the 3'UTR of *ANXA2* plays an important role in ANXA2-mRNA transport and translation via binding to itself (471), to other mRNAs (472) and to miRNAs (473). It has been reported that miRNAs regulate *ANXA2* expression and mediate several disease-related processes. For instance, miR-21 down-regulated *ANXA2* expression and mediated endothelial dysfunction in patients with diabetes (473), while miR-206 down-regulated *ANXA2* expression and controlled the phenotypic modulation of pulmonary artery smooth muscle cells in hepatopulmonary syndrome (474). Also miR-206 (475), miR-9 (476) and miR-1 (477) have been linked to carcinogenesis in different cancer types.

AnxA2 also has been linked to the cholesterol clearance process because it influences the PCSK9-mediated degradation of LDL-R (101, 102). Previous chapters have shown variants either in the coding or regulatory element regions that could mediate *ANXA2* expression. In this chapter, I examine the hypothesis that the 3'UTR of *ANXA2*-mRNA may be a site for miRNAs that could regulate gene expression, and consequently affect cholesterol levels in plasma. My

aim was to identify common variants in the 3'UTR of *ANXA2* that may create binding sites for miRNAs and may have a differential influence on *ANXA2* expression and LDL-C levels.

8.2. Results

8.2.1. Computational prediction of miRNA targets

To identify a possible interaction between miRNAs and mRNA of the *ANXA2*, the 3'UTR of human *ANXA2* was analyzed using a computational prediction method. Three commonly used miRNA prediction algorithms were used: miRWalk (292), mirSNP (293) and miRanda (478),

8.2.1.1. The miRWalk algorithm

The miRWalk database provides predicted and validated information on miRNA-target interactions (http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/). Also, this prediction tool provides a compression option which compares its output with two other popular miRSearch databases, miRanda and Targetscan. The miRWalk algorithm selects only the miRNAs that have ≥7 nucleotides seed length which increases the possibility of interactions with its possible target sequence (*ANXA2*), and miRNAs that have probability for random matches lower than 0.05 [Poisson distribution is used to calculate p-value in Sadygov et al., 2003 and Havilio et al., 2003]. The miRWalk algorithm showed that a total of 71 miRNAs fitted the criteria and were predicted to bind to the *ANXA2*-3'UTR (Table 8-1). In addition, 53 of these miRNAs were also predicted by two other algorithms: TargetScan (479) and miRanda.

Table 8-1: The miRWalk algorithm: *ANXA2*-3'UTR predicted miRNA target sites

The table shows miRNAs that could bind to the ANXA2-3'UTR and information about miRNA-target interactions including seeding length, starting position of miRNA seeding, and end position, and p-value as predicted by miRWalk. Also, the table shows the comparative platform of miRNA binding site predictions using different algorithms (miRWalk, miRanda and TargetScan). When a specific miRNA is predicted by one algorithm, it was counted as 1, and the SUM column shows how many algorithms predicted a specific miRNA. * indicated to MiRNA different isoform.

MicroRNA	Seed length	Start	End	P-value	miRanda	miRWalk	Targetscan	SUM
hsa-miR-29b	10	1191	1182	0.0005	1	1	1	3
hsa-miR-29c	10	1191	1182	0.0005	1	1	1	3
hsa-miR-513a-5p	9	1261	1253	0.0019	1	1	1	3
hsa-miR-613	9	1360	1352	0.0019	1	1	1	3
hsa-miR-155	9	1278	1270	0.0019	1	1	1	3
hsa-miR-376b	8	1386	1379	0.0074	1	1	1	3
hsa-miR-376a	8	1386	1379	0.0074	1	1	1	3
hsa-miR-650	8	1405	1398	0.0074	1	1	1	3
hsa-miR-376a	8	1386	1379	0.0074	1	1	1	3
hsa-miR-29a	8	1191	1184	0.0074	1	1	1	3
hsa-let-7g*	8	1567	1560	0.0074	1	1	0	2
hsa-miR-92a-1*	8	1288	1281	0.0074	1	1	0	2
hsa-miR-206	8	1361	1354	0.0074	1	1	1	3
hsa-miR-1	8	1361	1354	0.0074	1	1	1	3
hsa-miR-1226	8	1524	1517	0.0074	1	1	1	3
hsa-miR-29b-1*	8	1227	1220	0.0074	1	1	0	2
hsa-miR-548k	7	1449	1443	0.0293	1	1	1	3
hsa-miR-9	7	1348	1342	0.0293	1	1	1	3
hsa-miR-488	7	1480	1474	0.0293	1	1	1	3
hsa-miR-29b-2*	7	1226	1220	0.0293	1	1	0	2
hsa-miR-1299	7	1183	1177	0.0293	1	1	1	3
hsa-miR-559	7	1450	1444	0.0293	1	1	1	3
hsa-miR-302a*	7	1602	1596	0.0293	1	1	0	2
hsa-miR-650	7	1404	1398	0.0293	1	1	1	3

hsa-miR-519d	7	1399	1393	0.0293	1	1	1	3
hsa-miR-664	7	1432	1426	0.0293	1	1	1	3
hsa-miR-574-3p	7	1177	1171	0.0293	1	1	1	3
hsa-miR-101	7	1414	1408	0.0293	1	1	1	3
hsa-miR-17	7	1399	1393	0.0293	1	1	1	3
hsa-miR-767-5p	7	1190	1184	0.0293	1	1	1	3
hsa-miR-520g	7	1466	1460	0.0293	1	1	1	3
hsa-miR-10a*	7	1443	1437	0.0293	1	1	0	2
hsa-miR-138	7	1228	1222	0.0293	1	1	0	2
hsa-miR-579	7	1616	1610	0.0293	1	1	1	3
hsa-miR-890	7	1588	1582	0.0293	1	1	1	3
hsa-miR-520g	7	1399	1393	0.0293	1	1	1	3
hsa-miR-185	7	1335	1329	0.0293	1	1	1	3
hsa-miR-579	7	1354	1348	0.0293	1	1	1	3
hsa-miR-377*	7	1289	1283	0.0293	1	1	0	2
hsa-miR-220c	7	1409	1403	0.0293	1	1	1	3
hsa-miR-520h	7	1466	1460	0.0293	1	1	1	3
hsa-let-7g*	7	1566	1560	0.0293	1	1	0	2
hsa-miR-588	7	1252	1246	0.0293	1	1	1	3
hsa-miR-425	7	1419	1413	0.0293	1	1	1	3
hsa-miR-92a-1*	7	1287	1281	0.0293	1	1	0	2
hsa-miR-220c	7	1260	1254	0.0293	1	1	1	3
hsa-miR-520h	7	1399	1393	0.0293	1	1	1	3
hsa-miR-206	7	1360	1354	0.0293	1	1	1	3
hsa-miR-590-3p	7	1499	1493	0.0293	1	1	1	3
hsa-miR-20b	7	1399	1393	0.0293	1	1	1	3
hsa-miR-93	7	1399	1393	0.0293	1	1	1	3
hsa-miR-1	7	1360	1354	0.0293	1	1	1	3
hsa-miR-600	7	1341	1335	0.0293	1	1	1	3
hsa-miR-452*	7	1429	1423	0.0293	1	1	1	3
hsa-miR-101	7	1414	1408	0.0293	1	1	1	3
hsa-miR-1226	7	1523	1517	0.0293	1	1	1	3
hsa-miR-513a-3p	7	1610	1604	0.0293	1	1	1	3
hsa-miR-27b*	7	1550	1544	0.0293	1	1	0	2

hsa-miR-1	7	1360	1354	0.0293	1	1	1	3
hsa-miR-409-3p	7	1359	1353	0.0293	1	1	0	2
hsa-miR-1206	7	1356	1350	0.0293	1	1	1	3
hsa-miR-130a*	7	1391	1385	0.0293	1	1	0	2
hsa-miR-617	7	1381	1375	0.0293	1	1	0	2
hsa-miR-412	7	1379	1373	0.0293	1	1	1	3
hsa-miR-29b-1*	7	1226	1220	0.0293	1	1	0	2
hsa-miR-548k	7	1464	1458	0.0293	1	1	1	3
hsa-miR-138	7	1228	1222	0.0293	1	1	0	2
hsa-miR-513a-3p	7	1610	1604	0.0293	1	1	1	3
hsa-miR-155*	7	1540	1534	0.0293	1	1	1	3
hsa-miR-631	7	1216	1210	0.0293	1	1	0	2

8.2.1.2. The mirSNP algorithm

The mirSNP database (293) was used to identify which of these miRNAs will bind to ANXA2-3'UTR once a SNP is present. MirSNP used GWAS and eQTLs research to identify the putative miRNA-related SNPs. The mirSNP database provides information including miRNA support vector regression (mirSVR), the effect of the presence of the SNP on miRNA binding, miRanda imperfect 7 nucleotides seed site pairing score (cutoff of ≥140), alleles of the SNP, free energy that is released due to the interaction between miRNA and mRNA, and the conservation score. The mirSVR is a method that takes in to account local and global contextual features and is used to rank microRNA target sites by a down-regulation score. The local contextual features include the A:U content and the predicted secondary structure accessibility at positions flanking the site, while global contextual features include the relative position in the 3'UTR length, and conservation (293).

The mirSNP database showed that 51 SNPs in the 3'UTR could create sites, disrupt potential binding sites, and alter miRNA binding stability (by enhancing or decreasing binding) (Table 8-2). To narrow down the potential target sites for experimental validation, only SNPs that created the binding sites and had known mirSVR scores were selected. This criteria allowed nomination of the strongest candidates: only seven (highlighted in bold in Table 8-2) out of 51 SNPs were nominated to bind 9 miRNAs with different isoforms (hsa-miR-155-3p:rs116928563, hsa-miR-185-5p:rs11553799, hsa-miR-4306:rs11553799, hsa-miR-376a-3p:rs41315062, hsa-miR-376b: rs41315062,

hsa-miR-376a-3p:rs150128007, hsa-miR-376b:rs150128007, hsa-miR-520g: rs140887396, and hsa-miR-520h:rs140887396).

From the five SNPs, miR-155:rs116928563 was selected for experimental study because the rs116928563 A allele created a binding site for miR-155-3p (miR-155*), reduced the free energy by 21 kcal/mol and produced a conserved seed match score of 0.52. In addition, the literature showed that miR-155 binds to *ANXA2*-mRNA (480, 481), and is associated with atherosclerosis and LDL-C (482).

Table 8-2: The mirSNP algorithm: SNP-associated miRNAs

The table shows that 51 SNPs were predicted to bind to miRNA. Details of SNP-associated miRNAs are given including support vector regression (mirSVR), the effect of interaction, SNP alleles, miRanda imperfect 7 nucleotides seed site pairing score (cutoff ≥140), free energy, and conservation in different mammalian species. The five nominated SNPs are highlighted in bold.

miRNA	SNP	mirSVR	Effect	Allele	Score	Free energy kcal/mol	Conservation
hsa-miR-101-3p	rs140887396	-0.777	decrease	G	145	-17.19	0.176
				C	140	-10.26	0.176
hsa-miR-155-3p	**rs116928563**	**-0.256**	**create**	**T**			
				A	**146**	**-15.34**	**0.52**
hsa-miR-185-5p	**rs11553799**	**-0.736**	**create**	**T**			
				C	**154**	**-21**	**0.404**
hsa-miR-185-5p	rs184828726	-0.736	decrease	G	154	-21	0.91
				C	146	-17.67	0.91
hsa-miR-29b-1-5p	rs1803910	-0.242	decrease	A	156	-19.72	0
				C	152	-16.75	0
hsa-miR-29b-1-5p	rs11553796	-0.242	decrease	T	154	-18.07	0
				G	152	-16.75	0
hsa-miR-3125	rs184828726		create	G			
				C	150	-15.39	0.001
hsa-miR-3160-5p	rs116928563		enhance	T	144	-13.93	0.001
				A	145	-15.97	0.001
hsa-miR-3612	rs140887396		decrease	G	145	-18.75	1
				C	141	-16.05	1
hsa-miR-3673	rs138873688		enhance	G	149	-11.41	0.961
				A	153	-11.7	0.961
hsa-miR-3684	rs11553796		create	T			
				G	144	-12.55	0.253
hsa-miR-3685	rs116928563		create	T			
				A	140	-13.12	0
hsa-miR-376a-3p	**rs41315062**	**-0.501**	**create**	**G**			
				A	**141**	**-12.44**	**1**
hsa-miR-376a-3p	**rs150128007**	**-0.501**	**create**	**G**			
				A	**141**	**-12.44**	**0.431**
hsa-miR-376b	**rs41315062**	**-0.501**	**create**	**G**			
				A	**141**	**-12.22**	**1**
hsa-miR-376b	**rs150128007**	**-0.501**	**create**	**G**			
				A	**141**	**-12.22**	**0.431**
hsa-miR-3916	rs184828726		create	G			
				C	153	-19.56	0.001
hsa-miR-4305	rs11553796		break	T	140	-10	0.304
				G			
hsa-miR-4305	rs184828726	-0.249	break	G	140	-12.52	0.287
				C			
hsa-miR-4306	**rs11553799**	**-0.732**	**create**	**T**			
				C	**154**	**-21**	**0.404**
hsa-miR-4306	rs184828726	-0.732	decrease	G	154	-21	0.91
				C	146	-15.45	0.91

hsa-miR-4459	rs140887396		enhance	G	152	-19.27	0.999
				C	157	-23.78	0.999
hsa-miR-4470	rs140887396		break	G	141	-12.57	1
				C			
hsa-miR-4476	rs3582		break	C	146	-22.94	0.875
				T			
hsa-miR-4493	rs3582		break	C	144	-13.09	0.764
				T			
hsa-miR-4503	rs11553798		break	T	151	-11.86	0
				C			
hsa-miR-452-5p	rs140887396		break	G	151	-12.45	0.999
				C			
hsa-miR-4524b-3p	rs184828726		break	G	140	-14.73	0.079
				C			
hsa-miR-4644	rs3582		decrease	C	152	-22.05	0.001
				T	151	-21.23	0.001
hsa-miR-4644	rs11553799		create	T			
				C	152	-22.05	0.404
hsa-miR-4644	rs184828726		enhance	G	152	-22.05	0.91
				C	160	-21.05	0.91
hsa-miR-4672	rs140887396		enhance	G	151	-15.69	0.546
				C	156	-19.02	0.546
hsa-miR-4676-3p	rs140887396		break	G	162	-18.28	0.999
				C			
hsa-miR-4695-5p	rs140887396		enhance	G	146	-22.73	1
				C	150	-23.28	1
hsa-miR-4720-3p	rs11553798		break	T	143	-11.42	0.002
				C			
hsa-miR-4742-5p	rs11553798		enhance	T	142	-12.76	0
				C	144	-14.8	0
hsa-miR-4753-5p	rs3582		break	C	142	-12.25	0.644
				T			
hsa-miR-4755-3p	rs11553798		enhance	T	148	-18.26	0
				C	150	-21.2	0
hsa-miR-4768-3p	rs11553799		create	T			
				C	152	-16.54	0.277
hsa-miR-4768-3p	rs184828726		enhance	G	152	-16.54	0.952
				C	160	-21.05	0.952
hsa-miR-499a-3p	rs150128007		break	G	150	-15.88	0.176
				A			
hsa-miR-499b-3p	rs150128007		break	G	143	-11.52	0.176
				A			
hsa-miR-499b-5p	rs150128007		break	G	142	-15.53	0.794
				A			
hsa-miR-520g	rs41315062	-0.868	enhance	G	165	-15.9	0.603
				A	169	-13.74	0.603
hsa-miR-520g	**rs140887396**	**-0.868**	**create**	**G**			
				C	**169**	**-13.74**	**1**
hsa-miR-520h	rs41315062	-0.868	enhance	G	162	-15.9	0.603
				A	164	-12.93	0.603
hsa-miR-520h	**rs140887396**	**-0.868**	**create**	**G**			
				C	**164**	**-12.93**	**1**
hsa-miR-548g-3p	rs140887396		break	G	147	-7.46	1
				C			
hsa-miR-548u	rs11553799		break	T	143	-8.65	0.449
				C			
hsa-miR-600	rs11553799	-0.405	decrease	T	154	-12.88	0.006
				C	153	-15	0.006
hsa-miR-613	rs138873688	-1.186	decrease	G	181	-25.16	0.996

				A	173	-19.72	0.996

8.2.1.3. The miRanda algorithm

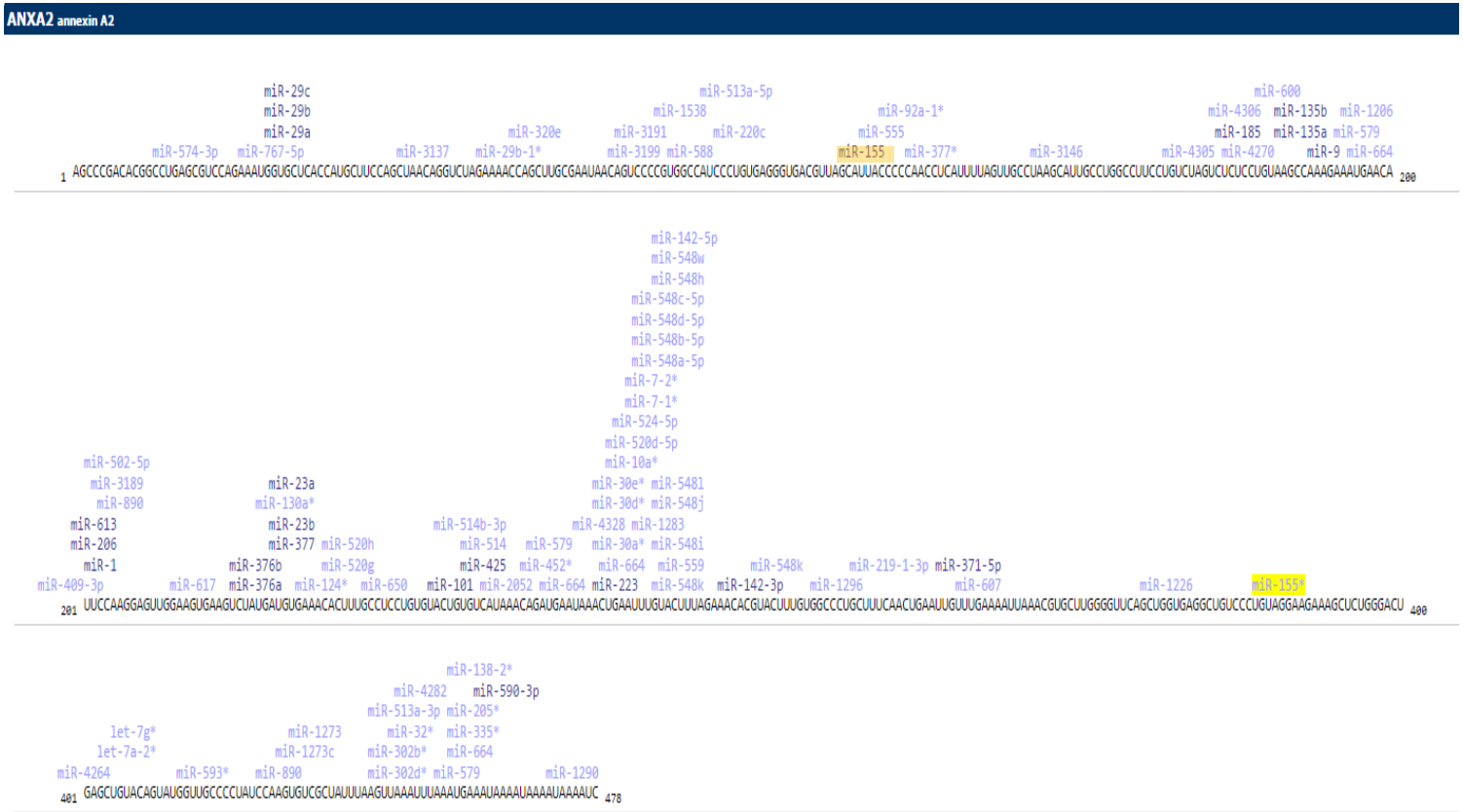
The miRanda algorithm was used to identify SNP-miRNA binding positions within *ANXA2*-3'UTR. Figure 8-1 A shows the *ANXA2*-3'UTR: miRNAs genome wide road map, where 16 miRNAs had good mirSVR scores (mirSVR≤ -0.1) and were conserved (dark blue), while 77 miRNAs had good mirSVR scores, but were less conserved (purple). Some of these miRNAs are predicted to target mRNA with more than one isoform.

The selected miR-155 binds to the *ANXA2*-3'UTR mRNA at two positions (Figure 8-1 B): the first, the miR-155 isoform, has been validated (481) to bind to *ANXA2*-3'UTR at position 96, with a mirSVR score of -0.2098 and a PhastCons score of 0.5084. The other isoform, miR-155*, has been predicted to bind to *ANXA2*-3'UTR mRNA at position 363, but only in the presence of the minor allele of rs116928563 (T>A), where the A allele of the SNP leads to a 7nt imperfect match to the 3'UTR of *ANXA2*. It is predicted that the miR-155* binds to the sequence around the A allele of the SNP; this binding creates free energy of -15.34 kcal/mol, the mirSVR score is -0.256, and the PhastCons score is 0.55, which suggests that rs116928563 may be a site for miRNA binding that has an influence on gene expression.

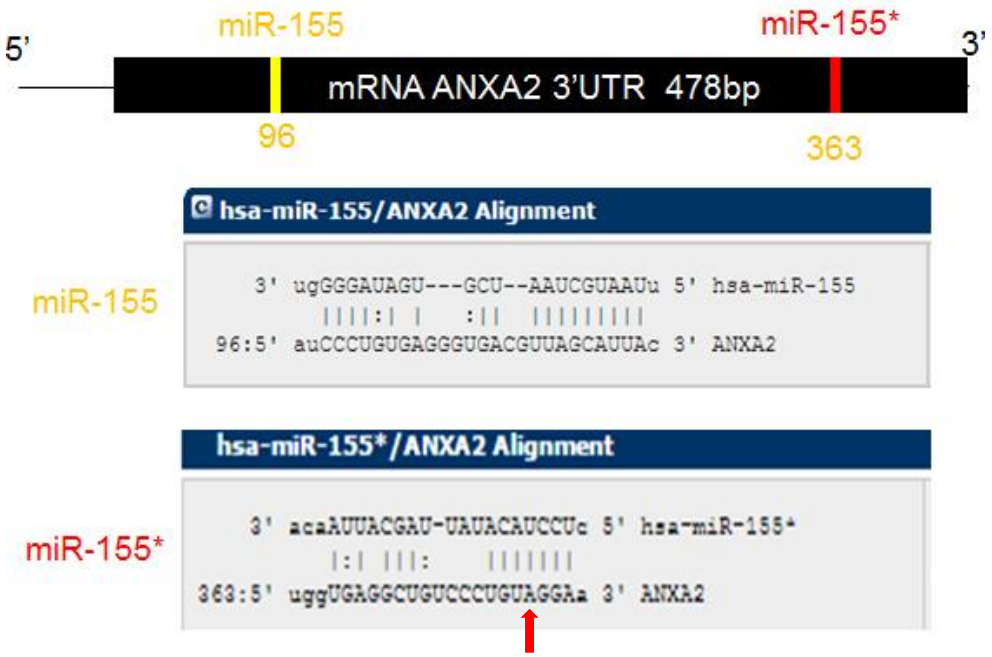
Figure 8-1: The miRanda algorithm: miRNAs binding position in the human *ANXA2*- 3'UTR

The *ANXA2*-3'UTR associated miRNAs genome wide road map. A) MiRNAs that are conserved and had a good mirSVR (mirSVR ≤ -0.1) are shown in dark blue, while miRNAs that are less conserved and had a good mirSVR are shown in purple. *ANXA2*: miR-155 is highlighted in orange and rs116928563 (A allele): miR-155* is highlighted in yellow. B) Sequence identity flanking conserved seed matches and miRNA seeds. Two isoforms of miR-155 bind to the *ANXA2*-3'UTR: miR-155 and the *ANXA2* rs116928563 (A allele): miR-155*. The road map was generated using miRanda (478). The gray dashes indicate the same analysis for conserved matches to control sequences. The colons indicate the G:U wobble pair.

A)



B)



rs116928563 A allele

8.2.2. Allele and genotype frequencies by population

The minor allele frequency (MAF) of the *ANXA2* SNP rs116928563 is variable between populations. The 1000 Genomes Project Phase 3 (Ensembl database) shows that the MAF is 4.6% in the East Asian population, although it varies between subpopulations (Table 8-3). In Hispanic and South Asian people where the SNP is rare, the MAF is 0.4% and 0.2%, respectively. In contrast, people with European and African ancestries do not carry the minor allele of the SNP.

Table 8-3: The *ANXA2* SNP rs116928563 genotype frequency by population.

Minor allele frequency (MAF) and genotype frequency in each population and subpopulation of the 1000 Genomes Project Phase 3. The data were adopted from the Ensembl database (http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core; r=15:60346749-60347749;v=rs116928563;vdb= variation;vf=124358705)

Population	MAF	Genotype frequency
East Asian	0.046	TT:0.911, AA:0.002, AT:0.087
Chinese Dai in Xishuangbanna, China	0.091	TT:0.828, AA:0.01, AT:0.161
Han Chinese in Beijing, China	0.039	TT:0.922, AT:0.078
Southern Han Chinese, China	0.010	TT:0.981, AT:0.019
Japanese in Tokyo, Japan	0.010	TT:0.981, AT:0.019
Kinh in Ho Chi Minh City, Vietnam	0.086	TT:0.828, AT:0.172
American	0.004	TT:0.991, AT:0.009
Colombian in Medellin, Colombia	0.005	TT:0.989, AT:0.011
Mexican Ancestry in Los Angeles, California	0.016	TT:0.969, AT:0.031
Peruvian in Lima, Peru	0	TT:1.000
Puerto Rican in Puerto Rico	0	TT:1.000
South Asian	0.002	TT:0.996, AT:0.004
Bengali in Bangladesh	0	TT: 1.000
Gujarati in Indians in Houston, TX	0.005	TT:0.990, AT:0.010
Indian Telugu in the UK	0	TT: 1.000
Punjabi in Lahore, Pakistan	0	TT: 1.000
Sri Lankan Tamil in the UK	0.005	TT:0.990, AT:0.010
European	0	TT: 1.000
Utah resident with Northern and Western European ancestry	0	TT: 1.000
Finnish in Finland	0	TT: 1.000
British in England and Scotland	0	TT: 1.000
Iberian populations in Spain	0	TT: 1.000
Toscani in Italy	0	TT: 1.000
African	0	TT: 1.000
African Caribbean in Barbados	0	TT: 1.000
African Ancestry in Southwest US	0	TT: 1.000
Esan in Nigeria	0	TT: 1.000
Luhya in Webuye, Kenya	0	TT: 1.000
Mandinka in The Gambia	0	TT: 1.000
Mende in Sierra Leone	0	TT: 1.000
Yoruba in Ibadan, Nigeria	0	TT: 1.000

8.2.3. *ANXA2* and miR-155 expressions in human hepatic cell lines

The mRNA expression of *ANAX2* and both isoforms of miR-155 were determined in two different human hepatic cell lines (Huh7 and HepG2) using TaqMan gene expression assays. Relative expression was calculated using the comparative threshold (Ct) values. *GAPDH* or *UBC* was used as the inner reference gene for *ANXA2*, while *U6* was used as the inner reference gene for miR-155. Relative expression was calculated using $2^{\wedge}\Delta C_T$, $\Delta C_T = C_{T\ target} - C_{T\ reference\ gene}$. Relative quantification of expression of the gene of interest was calculated using $2^{\wedge}\Delta\Delta\ C_T$ method, $\Delta\Delta\ C_T = \Delta\ C_{T\ target\ group} - \Delta\ C_{T\ control}$.

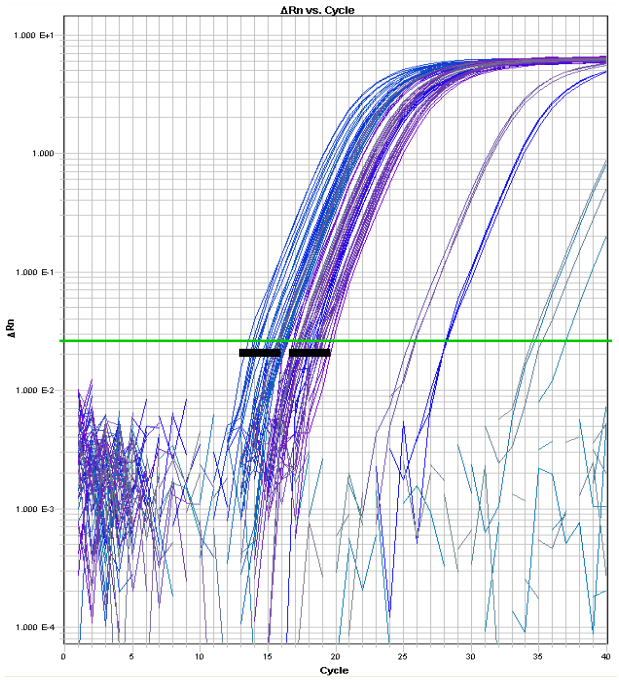
It was found that the mRNA of both the target gene (*ANXA2*) and the reference gene *GAPDH* were expressed in both cell lines (Figures 8-2 A and B). The expression level of ANXA2-mRNA was 4-fold higher in Huh7 cells than in HepG2 cells (see Figure 8-2 C), but it was not statistically significant (p=0.196), which may be due to the wide variability in levels recorded in the replicates. In contrast, miR-155 and miR-155* were not expressed in either cell line (see Figures 8-3 A and B), while U6 was expressed in both cell lines (Figure 8-3 C). Thus, no relative expression could be examined between miR-155 and U6.

Also to determine the genotype of rs116928563 in Huh7 cells, DNA was extracted from a Huh7 cell line, purified and the *ANXA2*-3'UTR sequence was checked. The sequencing data showed that Huh7 cells have the WT form of the SNP (Figure 8-4).

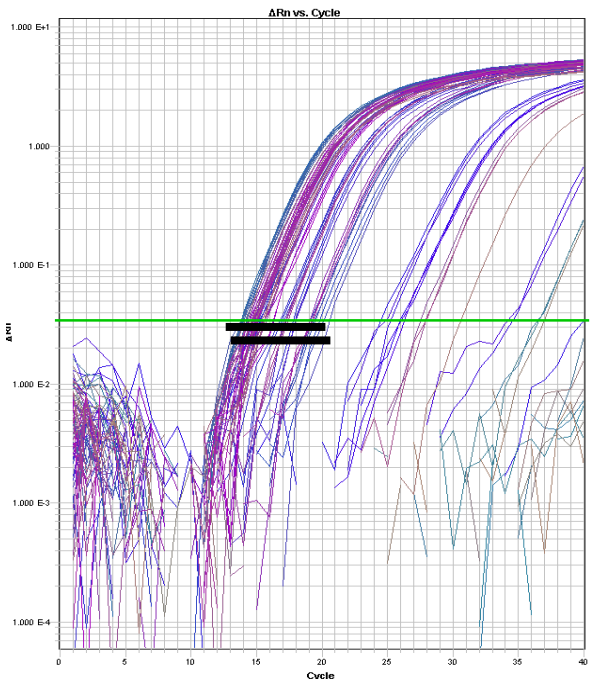
Figure 8-2: *ANXA2* and *GAPDH* mRNAs expression in liver cell lines

The figure illustrates the TaqMan gene expression plot of *ANXA2* and *GAPDH*, and the calculated relative expression. A) *ANXA2*-mRNA expression in two hepatic cell lines – Huh7 and HepG2. B) *GAPDH*-mRNA expression in Huh7 and HepG2 cell lines. The PCR was run for 40 cycles. The point at which the curve intersects the threshold (horizontal green line) is the $C_{T.}$ C) The relative expression of *ANXA2* to *GAPDH* in both cell lines – Huh7 and HepG2.

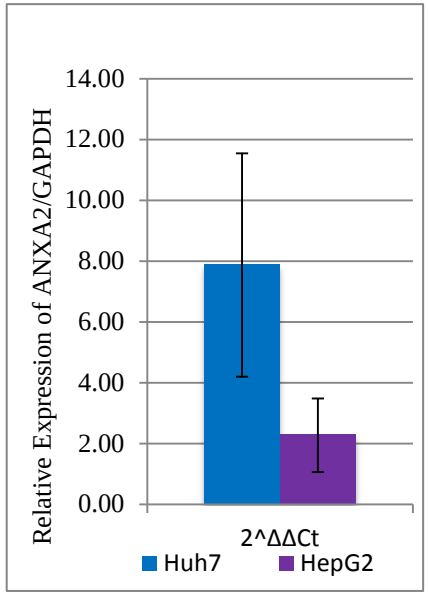
A) ***ANXA2*- mRNA expression**



B) *GAPDH*- mRNA expression



C) Relative expression of *ANXA2/ GAPDH*



Note. NTC = no template control

Figure 8-3: *MiR-155, miR-155** **and U6 expression in hepatic cell lines**

The figure illustrates plots of *miR-155, miR-155** **and U6 TaqMan gene expressions in Huh7 and HepG2 cell lines. A&B) miR-155 TaqMan gene expression plot shows that (A) *miR-155*-mRNA was not expressed in either cell line, and (B) *miR-155**** **was not expressed in either cell line. C) U6 TaqMan gene expression plot shows that U6 was expressed in both cell lines.**

A) *miR-155* expression **B) *miR-155**** **expression** **C) *U6* expression**

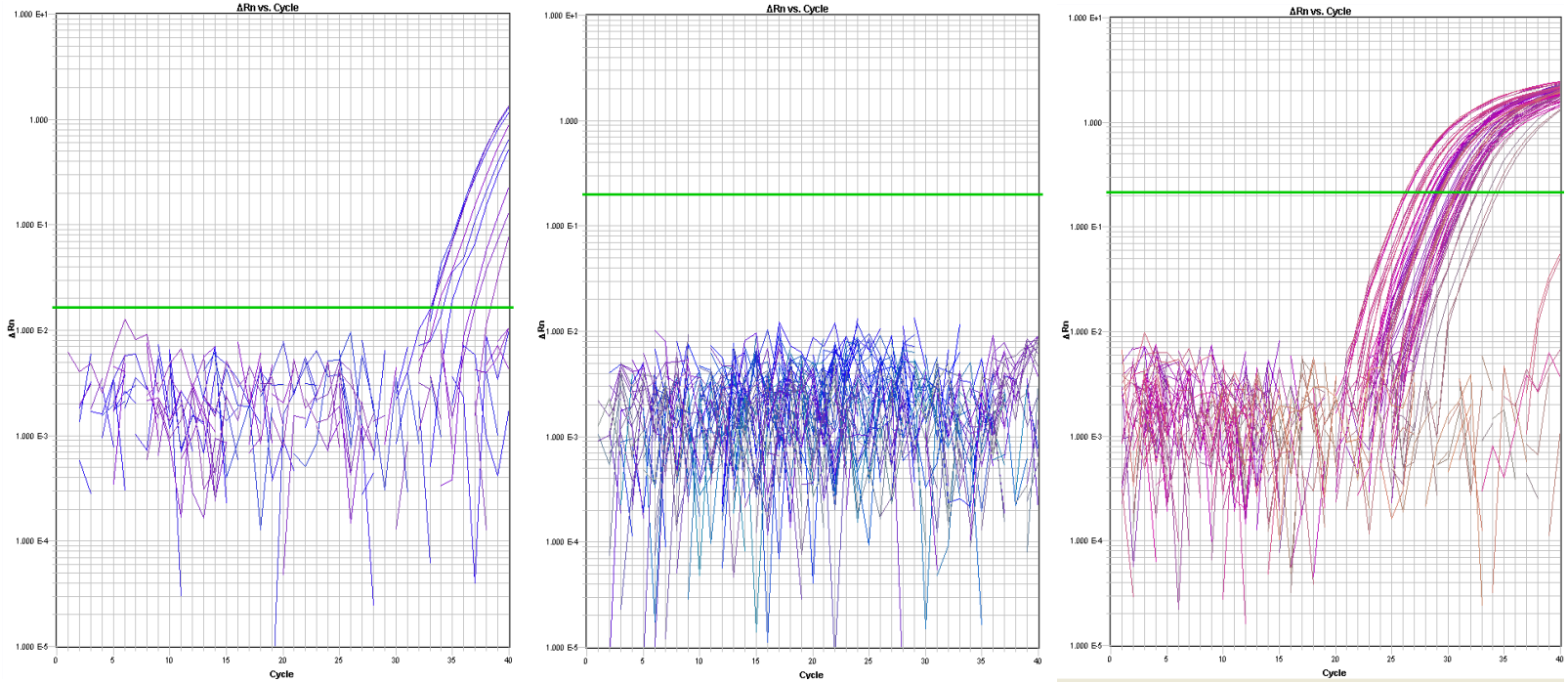
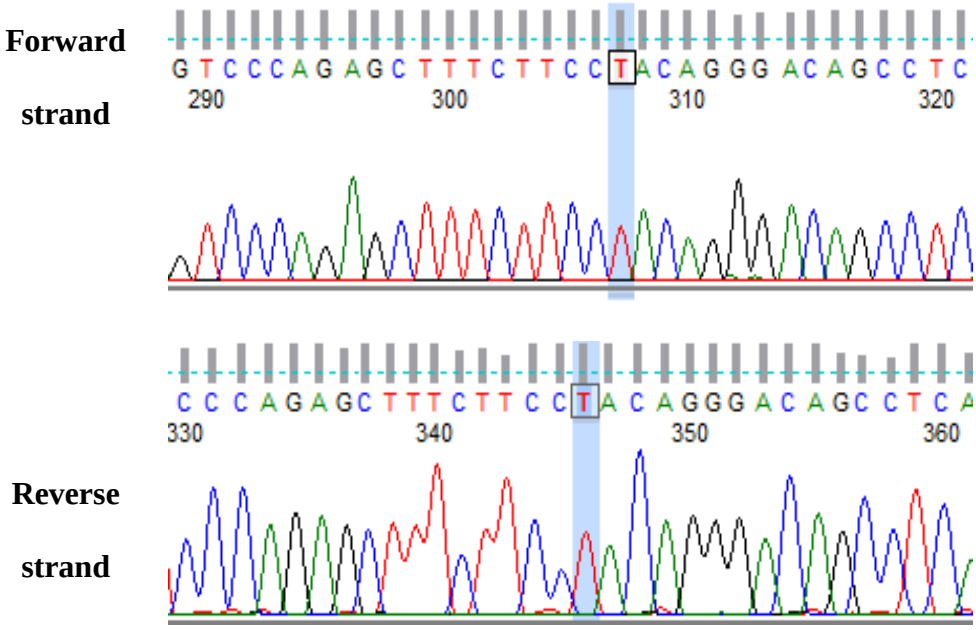


Figure 8-4: The *ANXA2* SNP rs116928563 sequencing result for the Huh7 cell line

The highlighted nucleotide is the *ANXA2* SNP rs116928563, which showed the WT genotype (T allele) in both strands.



8.2.4. The influence of miR-155* on *ANXA2* expression

To find out the effect of miR-155* on *ANXA2*-mRNA expression, the 409bp 3'UTR of human *ANXA2*, containing the SNP rs116928563, was amplified and then inserted into the enhancer site of the pGL3-basic luciferase reporter vector, after the SV40 polyadenylation signal between Sal I and BamH I (Figure 8-6 A). An *ANXA2* rs116928563 variant was then generated by using a quickchange site-directed method. All constructs were confirmed by sequencing (Figure 8-6 B).

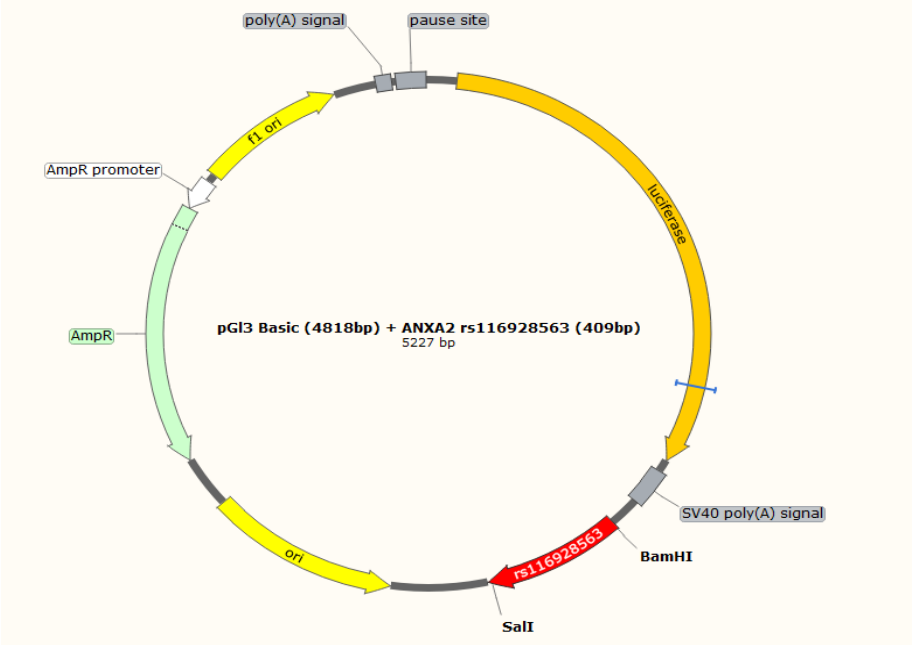
The Huh7 cell line was selected as I aimed to study the potential of miR-155 for *ANXA2* in hepatic cell lines. This cell line was ideal for such study because there was no endogenous miR155 competitor that could interfere with the experimental approach. Both alleles of the SNP constructs were first transfected into Huh7 cells, and then the cells were transfected with a synthetic miR-155* mimic after 24 hours, and gene expression quantified by real time-PCR after a further 24 hours. The efficacy of the first transfection was tested by β-gal, and when the transfection efficiency ≥ 30% the second transfection was performed. The efficacy of the second transfection was examined by assessing miR-155* expression in transfected cells with miR-155*. Figure 8-6 C only shows samples that were transfected with the miR-155* mimic and consequently expressed miR-155*, which indicates that the second transfections succeeded and no contamination occurred.

Relative expression of *ANXA2* in control samples evaluated experimental quality. No differences in relative expression of *ANXA2*-mRNA were noted between the control samples (A allele-negative control (NC), T allele-Opti-MEM medium, and T allele-siPORT transfection agent (p= 0.12, p= 0.52, p=0.14, respectively)), which were calculated relative to the T allele-NC. Also, no difference was found between the two alleles using transfected and the non-transfected miR-155* (T allele: p=0.30, A allele: p=0.78) (see Figure 8-6 D). It was noticed that miR-155* reduced the *ANXA2* expression by ≈2-fold in a presence of the A allele of the SNP, however no difference was found between the SNP alleles that were transfected with miR-155* (p=0.47), which may be due to lack of power that was caused by large variability.

Figure 8-5: *ANXA2* rs116928563 cloning construct map and the expression effect of miR-155* on rs116928563 alleles.

A) The *ANXA2* rs116928563 cloning construct map. 409bp of 3'UTR of human *ANXA2* containing rs116928563 was inserted after the SV40 polyadenylation signal between the Sal I and BamH I sites. B) Sequencing of the *ANXA2* rs116928563 WT and Mutant fragments. C) The assessment of miR-155* transfection efficiency. D) Effect of miR-155*-3p on the *ANXA2*-mRNA expression in the presence of the SNP rs116928563.

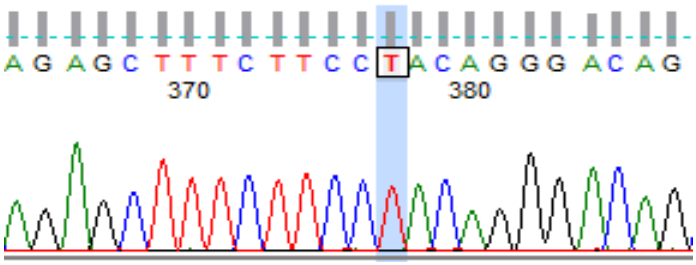
A)



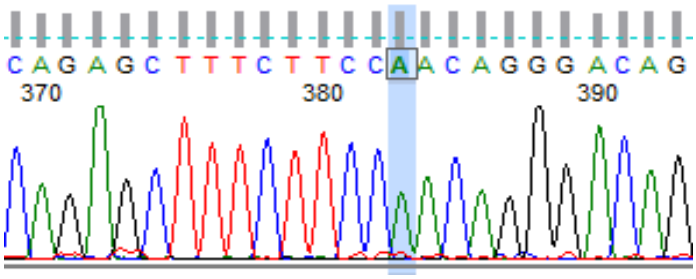
B)

rs116928563

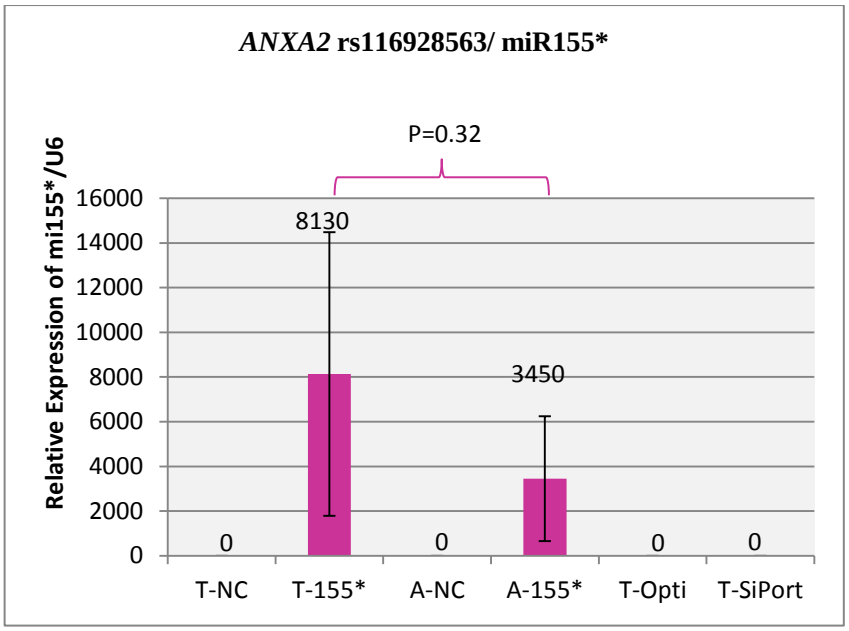
WT



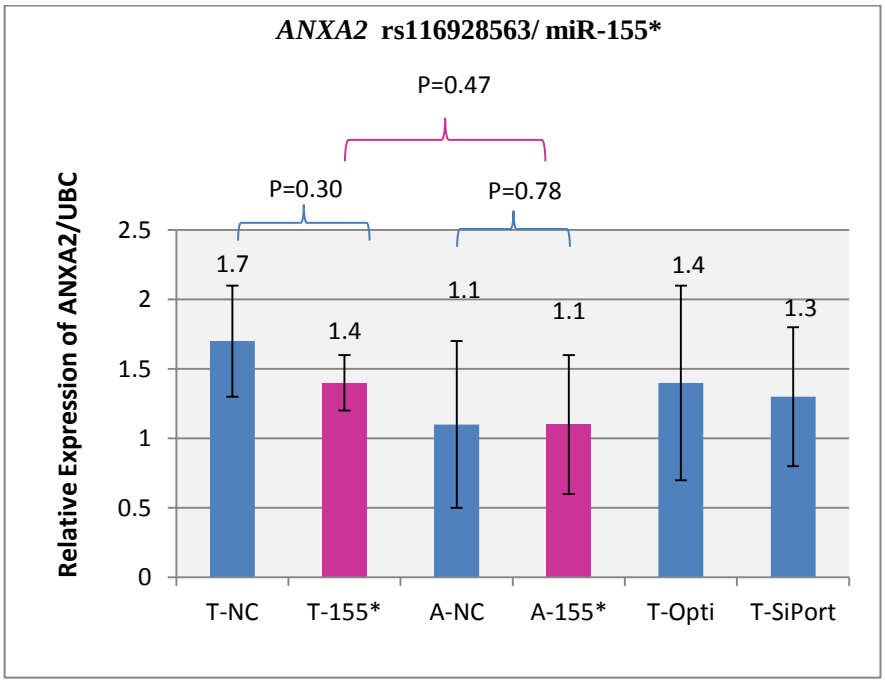
(A allele)



C)


ANXA2 rs116928563/ miR155*
P=0.32
Relative Expression of mi155*/U6
16000
14000
12000
10000
8000
6000
4000
2000
0
8130
3450
0
0
0
0
T-NC
T-155*
A-NC
A-155*
T-Opti
T-SiPort

D)


ANXA2 rs116928563/ miR-155*
P=0.47
P=0.30
P=0.78
Relative Expression of ANXA2/UBC
2.5
2
1.5
1
0.5
0
1.7
1.4
1.1
1.1
1.4
1.3
T-NC
T-155*
A-NC
A-155*
T-Opti
T-SiPort

8.3.Discussion

In silico miRNA prediction tools were useful in this study, making it possible to use three different tools to identify the miRNA of greatest potential interest (miRNA-155). The *in silico* tools predicted 71 miRNAs to bind to *ANXA2*-3'UTR; 53 of these miRNAs were predicted by more than one algorithm, of which only 9 were predicted to bind to *ANXA2*-3'UTR due to the presence of SNPs. The rs116928563: miRNA-155* was selected for experimental study. Studying the function of miRNAs and their role in gene regulation is complex because they are strongly cell type and context-dependent. This study failed to result in any definite conclusions about the influence of miRNA-155 in *ANXA2* expression and LDL-C levels. However, it was of note that rs116928563 was a site for miRNA-155* binding and miRNA-155* reduced *ANXA2* expression 2-fold (not significantly due to variation) under the conditions used. Other transfection conditions such as lipid stress and statins are suggested for future work.

It is well-known that the 3'UTR of genes, which are a target for miRNAs, have an essential role in post-transcriptional regulation. MiRNA-regulated gene expression has various biological roles including cell proliferation, cell differentiation, apoptosis, signal transduction, immune response, stress resistance, lipid metabolism, insulin secretion and hematopoiesis (467). *ANXA2* expression is regulated by miRNAs, which bind to partially complementary sites within the 3'UTR of the target *ANXA2*-mRNA and consequently lead to translational blockage or mRNA degradation (259, 260). It has been reported that miRNAs regulate *ANXA2* expression and mediate several disease-related processes (473-477). It has been suggested that lipid associated miRNAs could bind to *ANXA2*-mRNA and regulate its expression, which consequently affects LDL-C levels.

The main challenge in miRNA research is to detect the specific sequences on target genes that are fully or partially complementary to miRNAs. Both miRNA size and partial complementarity of miRNA to target sequences cause insufficient information of interaction specificity (483). A computational method is commonly used to predict miRNA/mRNA interaction. Computational methods are based on algorithms that are created by the physical properties of miRNA regulation; Watson–Crick complementarity between 2-7 nucleotides of the miRNA and the 3'UTR of mRNA. In time, other features have been added to improve prediction specificity including evolutionary conservation, free energy of RNA-miRNA duplexes, and secondary structure of 3'UTRs based on adenine–uracil **(**AU) contact (484). However, as with many computational methods these algorithms generate many false positive predictions (469). In order to overcome the complexity of predicting the seeding region and false positive predictions, several algorithms were used. Using these algorithms, 71 miRNAs were predicted to bind to the *ANXA2*-3'UTR, and of these 53 were predicted by more than one algorithm. The resulting lists of candidate target miRNAs overlapped between different algorithm predictions, which indicate the accuracy and efficacy of the prediction algorithms that were used. Five of the predicted target miRNAs have been experimentally validated, using a proteomics approach to measure changes in synthesis of proteins in response to miRNA transfection or endogenous miRNA knockdown as well as using microarrays to quantify mRNA levels, including hsa-miR-155-5p (481), hsa-miR-218-5p (485), hsa-miR-769-3p (486), hsa-miR-132-3p (486), and has-miR-133a-3p (487) which have added other level of confidence to the accuracy of these algorithms.

MiR-155 was selected for further functional study for several reasons. First, miR-155 has been experimentally validated to bind to *ANXA2*-mRNA, where miR-155 overexpression in the hCMEC/D3 cells decreased luciferase activity of *ANXA2* by ~30–40% (481). Second, it has been associated with lipid metabolism and chronic inflammatory diseases such as atherosclerosis. Miller et al. (2013) showed that miR-155 has a regulatory role in cholesterol and lipid metabolism in the liver via liver X receptor alpha (LXRα), which is a transcriptional regulator of many genes in liver metabolism. They also showed that an absence of miR-155 in mice fed a high fat diet was associated with significantly higher levels of VLDL/LDL cholesterol, and hepatic steatosis (482). It has been reported that circulating miR-155 has been significantly correlated with high levels of LDL-C (R= 0.32, p =0.021) (488). Also miR-155 has been considered as an inflammation-associated miRNA, which has influence in the pro-inflammatory environment and thus may affect atherosclerotic lesion formation (482, 489-492). Nazari-Jahantigh et al. (2012) found that miR-155 promotes atherosclerosis via its expression in pro-inflammatory macrophages and in atherosclerotic plaques and represses expression of BCL6, a transcription factor that reduces pro-inflammatory NF-κB signaling. Wagner et al. (2013) suggested that miR-155 highly bound to LDL-C may transfer to macrophages and atherosclerotic plaques.

Another important reason for selecting miR-155 was that it not only binds to the 3'UTR of WT *ANXA2*-mRNA, but also it is predicted that miR-155* isoforms bind in the presence of the minor allele of the SNP rs116928563 (T>A). It has been reported that the miRNA-associated SNPs have important roles in regulation of gene expression (468). Based on all of these, miR-155 appears to be a strong candidate for study. It was suggested that rs116928563 creates a seeding site for miR-155*, which may thus reduce *ANXA2* expression and increase LDL-C levels. Although the SNP rs116928563 is absent from the European population, the

minor allele has a frequency of 4.6% in the Asian population. Studying such a variant may give insight into the effect of miR-155 on *ANXA2* expression and LDL-C levels in this population.

AnxA2 levels in the liver are generally considered low (102), but in line with this study using Huh7 cells as a model system, stable knockdown of AnxA2 expression in Huh7 calls resulted in PCSK9 upregulation and a marked reduction in LDL-R levels (192). As the functions of miRNAs are strongly cell type- and context-dependent (470), two human hepatocellular carcinoma cell lines, Huh7 and HepG2, were selected to determine the expression of *ANXA2* and both isoforms of miR-155 (miR-155, miR-155*). *ANXA2*-mRNA expression was higher in Huh7 cells than in HepG2 cells. This finding is consistent with another study which reported that Huh7 cells showed 5-fold higher levels of *ANXA2*-mRNA than HepG2 cells (192). However, it was found that miR-155* and miR-155 were not expressed in either cell line under cell culture conditions. The expressions and functions of miRNAs are cell environment-dependent (470); they may be expressed or active under certain conditions but not expressed or are inactive in other conditions. Interestingly, it has been reported that the expression levels of miR-155 are affected by a high fat diet (482, 492), glucose (493), and statins (488).

8.4.Limitations

The two aims of this study were to identify miRNA-associated SNPs in the *ANXA2*-3'UTR and to determine the functionality of the selected SNPs. The first limitation of this study is that the miR-associated SNPs were selected in 2015 and the selection method of some miRNA algorithms has been updated since then. Therefore, there may be new potentially

useful miR-associated SNPs that could have been used in the analyses besides the SNP that was studied in this chapter.

The selected *ANXA2* SNP rs116928563 is located in a seeding region for lipid and inflammatory associated miRNA (miR-155*). However, this SNP is not present in the European population and therefore it has no local clinical advantage, although it can be studied in the Asian population. In addition, since miR-155 was not expressed in liver cell lines under the conditions that were used; it appears that miR-155 is less likely to have an influence on *ANXA2* expression. However, using alternative conditions such as lipid stress or statins may affect miR-155 expression in liver cells, which in turn may affect *ANXA2* expression. Due to time limitation it was not possible to carry out such complex experiments at this stage.

8.5. Conclusion

Bioinformatics tools such as miRNAs computational prediction algorithms are effective tools to identify miRNA associated genes. The complexity of studying miRNAs and their role in gene regulation comes from their functions, which are strongly cell type and context-dependent. Therefore, to determine their role in gene regulation, it is important to test miRNA expression under several conditions, such as in the presence of lipids or statins that are related to a trait and a disease of interest. In my opinion, although *ANXA2* rs116928563 does not exist in the European population and thus cannot bind to miR-155* and affect gene regulation, the miR-155 seems to be a good candidate for further study, especially as it also binds to WT *ANXA2*, and may affect its expression under specific circumstances. Further

investigations are required to elucidate whether miR-155 has a role in *ANXA2* expression and function.

Chapter 9 : General discussion

9.1. Overview

CHD is one of a number of common complex diseases that has been intensively studied. Genetic studies of monogenic causes of CHD aim to identify functional variants in genes that explain the phenotype, in addition to identifying therapeutic targets. For example, genetic studies of FH led to the identification of the FH-causing gain-of-function mutation in the *PCSK9* gene, and animal models and human subjects were used to successfully develop a PCSK9 monoclonal inhibitor, which was able to reduce LDL-C significantly even in subjects on maximal tolerated statin therapy (494), and have recently been shown to also reduce CHD events (495). Studies have also shown that both the affected gene and the type of mutation in the gene have roles in therapy outcomes. For example, Humphries et al. (2006) found that *LDLR* mutation positive FH patients have an 84% higher risk of early CHD than those who are mutation negative, and patients who are carriers of the *PCSK9* p.(Asp374Tyr) mutation also have a higher risk of CHD, while carriers of the *APOB* mutation are less likely to have CHD (496). Patients benefit greatly from genetic screening in terms of disease management and treatment plans. Genetic screening helps to identify pathogenic variants, the mutation spectrum and mutation classes that then contribute to disease management and prevention.

In general, CHD is more often caused by the co-inheritance of many common variants which lead to the observed phenotype, i.e. a "polygenic" cause of CHD. In the last 10 years GWASs have been successful in identifying a huge number of risk loci that are associated with common and complex diseases (suggest http://www.ebi.ac.uk/gwas/). The results from GWASs have been important in two ways: one by improving risk prediction assessment and the other by improving the understanding of human genetics. Detailed molecular analyses of the GWAS hit variants have allowed the identification of the precise functional variant at the

locus and the molecular mechanism involved, and thereby the understanding of the role of gene expression in the disease phenotype. For example, GWAS identified a new locus containing three novel genes on chromosome 1 (*SORT1, CELSR2* and *PSRC1*) that was associated with LDL-C and CHD (497). Molecular genetic studies showed that the *SORT1* SNP rs12740374 creates a new enhancer-binding protein for transcription factors, which in turn increase gene expression of *SORT1* and reduce LDL-C levels (256).

9.2. Relevance of studied variants in this thesis by allele frequency and strength of genetic affect

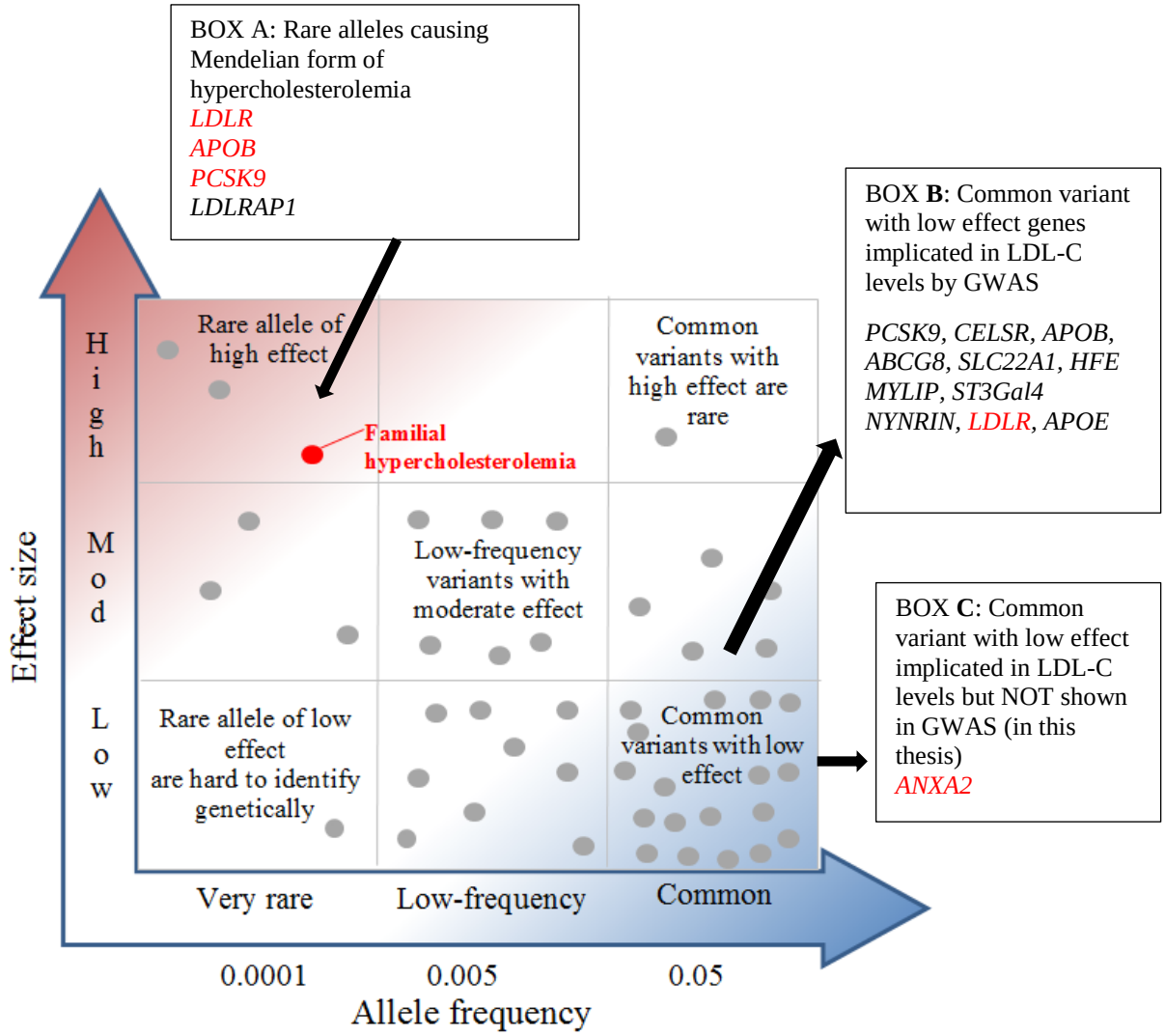
Common complex diseases have different genetic architecture to rare disorders. Cystic fibrosis (CF), for instance, is caused by genetic variants in a single gene (*CFTR*) (498), and genetic screening and linkage analysis of affected families are successful tools for identifying pathogenic genetic variants. However, applying the linkage approach to more common disorders, such as CHD and hypercholesterolemia, often does not result in clear-cut patterns of inheritance. This suggests that the common diseases have different genetic mechanisms to the rare disorders, and thus allele frequency and the disease-risk relationship hypothesis were developed and are often conceptualized in two dimensions: allele frequency and effect size (Figure 9-1). Highly penetrant alleles for Mendelian disorders (e.g. FH and CF) are rare with large effect sizes (Figure 9-1, upper left), while most GWAS findings are associations of common SNPs with small effect sizes (Figure 9-1, lower right), and both contribute to the disease phenotype.

There are many different technologies, study designs and analytical tools for studing genetic risk. Genetic effects of rare variants are more amenable to smaller family-based studies and

linkage analysis. On the other hand, effects of common variants are typical of findings from GWAS, requiring large sample sizes and a large panel of genetic markers to obtain statistically robust results. Within hypercholesterolemia, I have focused my study on rare variants with high effect size by performing cascade genetic tests for FH patients (Chapters 3 and 4), and on common variants with modest effect size by studying the GWAS common variant rs6511720 in the *LDLR* gene (Chapter 5). In addition, I studied the *ANXA2* gene, where biological studies have shown its involvement in the LDL-C clearance pathway. Common SNPs were selected for genetic association analyses in order to identify the LDL-C associated variant (Chapters 6 and 7). This study found a (low frequency) *ANXA2* coding SNP rs17845226 and a more common intergenic SNP rs11633032, which are associated with higher levels of LDL-C and a higher risk of CHD.

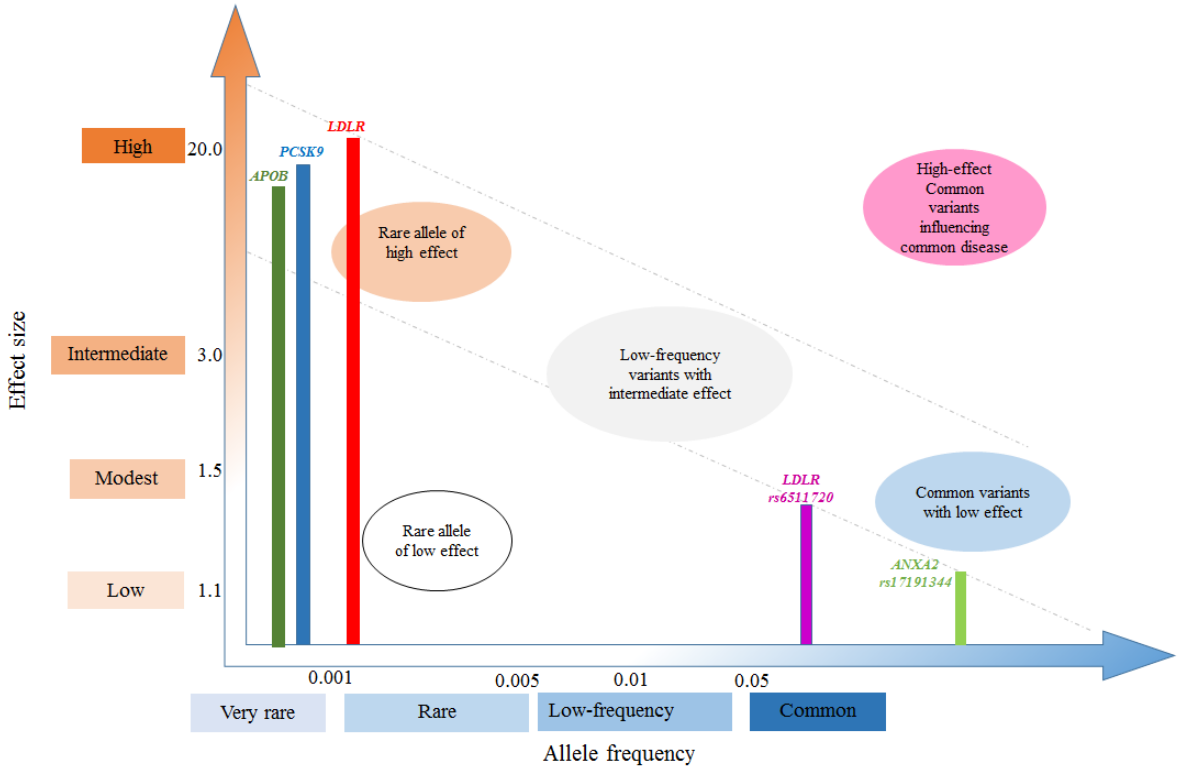
Figure 9-1: The relationship between effect size and allele frequency

Spectrum of allele frequency and effect size in the genetics of hypercholesterolemia. Identifying genetic variants by risk allele frequency (x-axis) and strength of genetic effect (odds ratio; effect size; y-axis). The hypercholesterolemia-associated genes are indicated in boxes corresponding to two categories of variants. Box A shows four genes (*LDLR, APOB PCSK9* and *LDLRAP1*) that are related to the Mendelian form of the disease (familial hypercholesterolemia). Box B shows a list of common variants identified by GWAS (160). Box C shows common variants identified in this thesis and not shown in GWAS. Genes indicated in red were studied in this thesis.



Although both SNPs studied, *LDLR* (rs6511720) and *ANXA2* (rs17191344), are common, the *ANXA2* SNP is more common (MAF=0.17) than the *LDLR* SNP (MAF=0.10), thus the latter has a higher genetic effect (modest effect size) while *ANXA2* has a lower genetic effect (low effect size) (Figure 9-2). From a scientific point of view, the *LDLR* SNP is more important than the *ANXA2* SNP because it has a dominant effect (one copy of the variant causing phenotype), a lower allele frequency and a modest effect on LDL-C levels, whereas the *ANXA2* SNP has a recessive effect (two copies of the variant required to observe phenotype), a higher allele frequency and a lower effect size.

Figure 9-2: Feasibility of studied genetic variants in this thesis by risk-allele frequency and strength of genetic effect



9.3. Common variant and functional analysis

Common alleles show heritability with low genetic effect, and multiple common alleles are required to influence disease susceptibility. Indeed, identifying common variants that are statistically associated with a disorder is not enough to allow definite conclusions of variant functionality due to the fact of chromosomal linkage, recombination events through generations (in other words the phenomenon of LD) and pattern differences between populations. It has been postulated that statistically associated variants are either functional or a marker of a functional variant somewhere in the locus. Therefore, it is important to map the precise location of the possible effective variants. The HapMap project and the 1000 Genomes project were successful in mapping 99% of human variation and identifying their haplotypes. The availability of the variant (common allele) haplotype map (270), in addition to the ENCODE project results, has provided insight about variant location and possible genetic impact (275, 276, 499). The genomic data help molecular genetics researchers to define functional variant(s) and their genetic mechanism and effect. In fact, a wide array of biomedical disciplines is cooperating in order to translate genetic findings into direct health benefits.

This thesis studied the molecular genetic mechanism of common variants associated with LDL-C and CHD that had been either identified by GWAS (*LDLR*) or identified in this study (*ANXA2).* The strategy used focused on identifying a functional SNP candidate and possible target genes using large genetic datasets combined with *in vitro* functional assays in order to identify causal variant(s). In this thesis, eight SNPs were selected for functional analysis, four *LDLR* variants and four *ANXA2* variants. This study identified that the GWAS *LDLR* "hit" SNP rs6511720 and LD SNP rs57217136 (383), and *ANXA2* common SNPs rs17191344 and

rs11633032 (439), are functional and act as cis-regulatory elements, where the sequence around the SNPs is a target for proteins that regulate gene transcription. Further study is required, especially in analyzing the interactions between DNA elements using chromatin capture techniques (429, 430) or genome editing (500) which may better elucidate the molecular mechanisms involved. Such studies build on available knowledge and can provide a fuller picture of the pathogenesis of hypercholesterolemia and the risk of CHD.

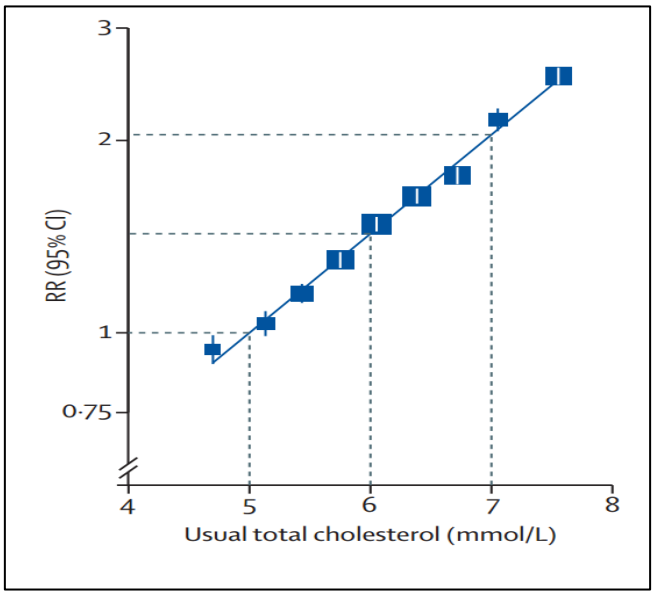
One identified *ANXA2* LDL-C associated variant, rs17845226, is located in the coding region of the gene. Therefore, a different strategy was employed to identify its functional role. However, time and resources were unavailable to carry out such analysis during the thesis's time frame. Further studies are needed to examine whether this mutation modifies the function of PCSK9 or has downstream consequences on LDL-R activity. Also, they should investigate whether AnxA2/p11 complex formation or the ratio of monomeric vs complexed AnxA2 proteins are in play, both of which may have consequences for PCSK9-mediated LDL-R down-regulation (102, 192).

9.4. Relation of LDL-C concentration to CHD risk

Elevated plasma cholesterol levels are a major risk factor for atherosclerosis and CHD (2). Studies have shown a positive association between LDL-C concentration and the rate of CHD events (501, 502) in different populations (Figure 9-3). In addition, large-scale evidence from randomized trials has shown that LDL-C lowering drug "statins" reduce the risk of major vascular events by about one-quarter for each mmol/l reduction in LDL-C during each year (502).

Figure 9-3: Association of LDL-C concentration to CHD risk

A logarithmic scale shows a linear relationship between LDL-C levels and CHD risk. The data suggest that for every 1 mmol/l change in LDL-C, the relative risk for CHD risk is changed proportionally by about 50%. The relative risk (RR) is set at 1.0 for LDL-C= 5mmol/l. (Adapted from (502))



Studies have shown that both monogenic and polygenic hypercholesterolemia have an association with CHD. It is well known that patients who have inherited an abnormal *LDLR* genetic variant from both parents (HoFH) and who have an LDL-C concentration greater than 13 mmol/l are likely to develop a CHD event before the age of 20 years (200), while those who only inherited one copy of an abnormal genetic variant from one parent (HeFH) have a LDL-C concentration greater than 8 mmol/l and are likely to have a CHD event in early middle-age (503). Moreover, several studies have provided evidence that the presence of several LDL-C common associated variants cause similar increases in LDL-C concentration as are observed in HeFH, thus smaller increases in the risk of CHD may occur (504).

Identifying individuals with moderate and high risk of CHD is important to reduce the mortality rate. The current genetic technologies provide a great opportunity to identify genetically affected individuals, which results in early diagnosis and thus allows early treatment, as well as an opportunity to identify new genetic markers such as *ANXA2* that was studied in this thesis.

9.5. FH children

A recent study that screened children for FH at the 12 month routine immunization GP appointment has shown that the FH prevalence in children in the UK is 1 in 273 (505). This high prevalence illustrates the importance of genetic screening in childhood not only to identify FH subjects, but as a means of preventative action prior to development of cardiovascular disease. In this thesis, FH genetic screening was performed in two population groups of different ethnicity (UK and Iran), using the target-NGS method. Children accounted for 70.6% (60/85) of the subjects studied, and mutations were found in 50% of children. Although the other 50% of children were FH-mutation negative, they are still at an elevated risk of CHD. FH-mutation negative children are likely to carry several common variants with low effect that may lead to the FH phenotype (506). The other possibility is that these children may carry a genetic defect in an undiscovered gene, in which case whole genome sequencing may be helpful.

The specificity and sensitivity of the screening method used in this thesis led to the identification of several novel mutations. The results showed novel *LDLR* variants (5 variants) in addition to one novel variant in *PSCK9*, all of which are predicted to be pathogenic. Three of the novel identified variants were found in children while the other three

were found in young adults <40 years. Family screening is highly recommended for all affected families, to identify members who are at high risk of CHD.

The findings from this thesis will benefit and enable researchers to build on the available knowledge of hypercholesterolemia and CHD. The results from this thesis indicate that both the rare and common variants investigated influence LDL-C levels and CHD risk, but at different levels depending on allele frequency and the affected gene. This study also showed how the use of multiple complementary tools can be used to identify novel pathways for complex diseases such as CHD from GWAS and wet lab findings. Such studies could ultimately identify novel therapeutic targets.

References

1. Organization WH. he Atlas of Heart Disease and Stroke. Geneva: World Health Organization 2004.
2. Prasad D, Kabir Z, Dash A, Das B. Cardiovascular risk factors in developing countries: A review of clinico-epidemiological evidence. CVD Prevention and Control. 2010;5(4):115-23.
3. Castelli W. Epidemiology of coronary heart disease: the Framingham study. The American journal of medicine. 1984;76(2):4-12.
4. Maas A, Appelman YEA. Gender differences in coronary heart disease. Netherlands Heart Journal. 2010;18(12):598-602.
5. Hawe E, Talmud P, Miller G, Humphries S. Family history is a coronary heart disease risk factor in the Second Northwick Park Heart Study. Annals of Human Genetics. 2003;67(2):97-106.
6. Balarajan R. Ethnic differences in mortality from ischaemic heart disease and cerebrovascular disease in England and Wales. British Medical Journal. 1991;302(6776):560-4.
7. Yusuf S, Hawken S, Ôunpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. The Lancet. 2004;364(9438):937-52.
8. Poirier P, Giles TD, Bray GA, Hong Y, Stern JS, Pi-Sunyer FX, et al. Obesity and Cardiovascular Disease: Pathophysiology, Evaluation, and Effect of Weight Loss: An Update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease From the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. Circulation. 2006;113(6):898-918.
9. Ezzati M, Henley SJ, Thun MJ, Lopez AD. Role of Smoking in Global and Regional Cardiovascular Mortality. Circulation. 2005;112(4):489-97.
10. Heitzer T, Ylä-Herttuala S, Luoma J, Kurz S, Münzel T, Just H, et al. Cigarette smoking potentiates endothelial dysfunction of forearm resistance vessels in patients with hypercholesterolemia role of oxidized LDL. Circulation. 1996;93(7):1346-53.
11. Pech-Amsellem M, Myara I, Storogenko M, Demuth K, Proust A, Moatti N. Enhanced modifications of low-density lipoproteins (LDL) by endothelial cells from smokers: a possible mechanism of smoking-related atherosclerosis. Cardiovascular research. 1996;31(6):975-83.
12. Meade T, Imeson J, Stirling Y. Effects of changes in smoking and other characteristics on clotting factors and the risk of ischaemic heart disease. The Lancet. 1987;330(8566):986-8.
13. Gourlay SG, Benowitz NL. Arteriovenous differences in plasma concentration of nicotine and catecholamines and related cardiovascular effects after smoking, nicotine nasal spray, and intravenous nicotine. Clinical Pharmacology & Therapeutics. 1997;62(4):453-63.
14. Al-Delaimy WK, Manson JE, Solomon CG, et al. SMoking and risk of coronary heart disease among women with type 2 diabetes mellitus. Archives of Internal Medicine. 2002;162(3):273-9.
15. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. American heart journal. 1991;121(1):293-8.
16. Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. The British Medical Journal. 2012;344:e4181.

17. Thanassoulis G, Peloso GM, Pencina MJ, Hoffmann U, Fox CS, Cupples LA, et al. A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium the framingham heart study. Circulation: Cardiovascular Genetics. 2012;5(1):113-21.
18. Virani SS. Statins in the Primary and Secondary Prevention of Cardiovascular Disease in Women: Facts and Myths. Texas Heart Institute Journal. 2013;40(3):288-9.
19. Endo A. The discovery and development of HMG-CoA reductase inhibitors. Journal of lipid research. 1992;33(11):1569-82.
20. Robinson JG. Management of Familial Hypercholesterolemia: A Review of the Recommendations from the National Lipid Association Expert Panel on Familial Hypercholesterolemia. J Manag Care Pharm. 2013;19(2):139-49.
21. Dong B, Wu M, Li H, Kraemer FB, Adeli K, Seidah NG, et al. Strong induction of PCSK9 gene expression through HNF1α and SREBP2: mechanism for the resistance to LDL-cholesterol lowering effect of statins in dyslipidemic hamsters. Journal of Lipid Research. 2010;51(6):1486-95.
22. Sattar N, Preiss D, Murray HM, Welsh P, Buckley BM, de Craen AJM, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. The Lancet. 2010;375(9716):735-42.
23. Swerdlow DI, Hingorani AD, Humphries SE. Genetic Risk Factors and Mendelian Randomization in Cardiovascular Disease. Current Cardiology Reports. 2015;17(5):33.
24. Cooper AD. Hepatic uptake of chylomicron remnants. Journal of Lipid Research. 1997;38(11):2173-92.
25. Westerterp M. Hyperlipidemia, inflammation and atherosclerosis: roles of apolipoprotein C1 and cholesteryl ester transfer protein: Department of General Internal Medicine, Endocrinology, and Metabolic Diseases, Medicine/Leiden University Medical Center (LUMC), Leiden University; 2007.
26. Brown M, Goldstein J. Lipoprotein receptors in the liver. Control signals for plasma cholesterol traffic. Journal of Clinical Investigation. 1983;72(3):743.
27. Rader DJ, Hobbs HH. Disorders of lipoprotein metabolism. Harrisons Principles of Internal Medicine. 2005;16(2):2286.
28. Fielding CJ, Fielding PE. Molecular physiology of reverse cholesterol transport. Journal of Lipid Research. 1995;36(2):211-28.
29. Brown MS, Goldstein JL. The SREBP Pathway: Regulation of Cholesterol Metabolism by Proteolysis of a Membrane-Bound Transcription Factor. Cell. 1997;89(3):331-40.
30. Brown MS, Goldstein JL. A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. Proceedings of the National Academy of Sciences. 1999;96(20):11041-8.
31. Sakakura Y, Shimano H, Sone H, Takahashi A, Inoue K, Toyoshima H, et al. Sterol Regulatory Element-Binding Proteins Induce an Entire Pathway of Cholesterol Synthesis. Biochemical and Biophysical Research Communications. 2001;286(1):176-83.
32. Osborne TF, Gil G, Goldstein JL, Brown MS. Operator constitutive mutation of 3-hydroxy-3-methylglutaryl coenzyme A reductase promoter abolishes protein binding to sterol regulatory element. Journal of Biological Chemistry. 1988;263(7):3380-7.
33. Horton JD, Goldstein JL, Brown MS. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. The Journal of Clinical Investigation. 2002;109(9):1125-31.
34. GB Healthwatch: The SREBP pathway in cholesterol regulation [Available from: http://www.gbhealthwatch.com/Science-Cholesterol-Lipoproteins-101.php.
35. Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis. Nature. 2011;473(7347):317-25.

36. Weber C, Noels H. Atherosclerosis: current pathogenesis and therapeutic options. Nature Medicine. 2011;17(11):1410-22.
37. Lusis AJ. Atherosclerosis. Nature. 2000;407(6801):233-41.
38. Brown MS, Goldstein JL. A Receptor-Mediated Pathway for Cholesterol Homeostasis (Nobel Lecture). Angewandte Chemie International Edition in English. 1986;25(7):583-602.
39. Goldstein JL, Brown MS. Binding and Degradation of Low Density Lipoproteins by Cultured Human Fibroblasts: Comparison of Cells From a Normal Subject and from a Patient with Homozygous Familial Hypercholesterolemia. Journal of Biological Chemistry. 1974;249(16):5153-62.
40. Goldstein JL, Schrott HG, Hazzard WR, Bierman EL, Motulsky AG. 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. 1973;52(7):1544-68.
41. Brown MS, Goldstein JL. Receptor-mediated endocytosis: insights from the lipoprotein receptor system. Proceedings of the National Academy of Sciences. 1979;76(7):3330-7.
42. Hobbs HH, Brown MS, Goldstein JL. Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. Human Mutation. 1992;1(6):445-66.
43. Benn M, Watts GF, Tybjaerg-Hansen A, Nordestgaard BG. Familial Hypercholesterolemia in the Danish General Population: Prevalence, Coronary Artery Disease, and Cholesterol-Lowering Medication. The Journal of Clinical Endocrinology & Metabolism. 2012;97(11):3956-64.
44. Moorjani S, Roy M, Gagne C, Davignon J, Brun D, Toussaint M, et al. Homozygous familial hypercholesterolemia among French Canadians in Québec province. Arteriosclerosis, Thrombosis, and Vascular Biology. 1989;9(2):211-6.
45. Soutar AK, Naoumova RP. Mechanisms of Disease: genetic causes of familial hypercholesterolemia. Nature Clinical Practice Cardiovascular Medicine. 2007;4(4):214-25.
46. Taylor A, Bayly G, Patel K, Yarram L, Williams M, Hamilton-Shield J, et al. A double heterozygote for familial hypercholesterolaemia and familial defective apolipoprotein B-100. Annals of Clinical Biochemistry. 2010;47(5):487-90.
47. Fouchier SW, Dallinga-Thie GM, Meijers JC, Zelcer N, Kastelein JJ, Defesche JC, et al. Mutations in stap1 are associated with autosomal dominant hypercholesterolemia. Circulation research. 2014;115(6):552-5.
48. Day IN. FH4= STAP1. Another Gene for Familial Hypercholesterolemia? Relevance to Cascade Testing and Drug Development? Circulation research. 2014;115(6):534-6.
49. Yamamoto T, Davis CG, Brown MS, Schneider WJ, Casey ML, Goldstein JL, et al. The human LDL receptor: A cysteine-rich protein with multiple Alu sequences in its mRNA. Cell. 1984;39(1):27-38.
50. Sudhof TC, Goldstein JL, Brown MS, Russell DW. The LDL receptor gene: a mosaic of exons shared with different proteins. Science. 1985;228(4701):815-22.
51. Hobbs HH, Russell DW, Brown MS, Goldstein JL. The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. Annual review of genetics. 1990;24(1):133-70.
52. Stanley KK, Kocher HP, Luzio JP, Jackson P, Tschopp J. The sequence and topology of human complement component C9. The European Molecular Biology Organization Journal. 1985;4(2):375-82.
53. Russell DW, Schneider WJ, Yamamoto T, Luskey KL, Brown MS, Goldstein JL. Domain map of the LDL receptor: Sequence homology with the epidermal growth factor precursor. Cell. 1984;37(2):577-85.

54. Sudhof T, Russell D, Goldstein J, Brown M, Sanchez-Pescador R, Bell G. Cassette of eight exons shared by genes for LDL receptor and EGF precursor. Science. 1985;228(4701):893-5.
55. Goldstein JL, Brown MS, Anderson RG, Russell DW, Schneider WJ. Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. Annual review of cell biology. 1985;1(1):1-39.
56. Al-Allaf F, Coutelle C, Waddington S, David A, Harbottle R, Themis M. LDLR-Gene therapy for familial hypercholesterolaemia: problems, progress, and perspectives. International Archives of Medicine. 2010;3(1):36.
57. Innerarity TL. LDL Receptor's β-Propeller Displaces LDL. Science. 2002;298(5602):2337-9.
58. Brown MS, Anderson RGW, Goldstein JL. Recycling receptors: The round-trip itinerary of migrant membrane proteins. Cell. 1983;32(3):663-7.
59. Rudenko G, Henry L, Henderson K, Ichtchenko K, Brown MS, Goldstein JL, et al. Structure of the LDL receptor extracellular domain at endosomal pH. Science. 2002;298(5602):2353-8.
60. Südhof TC, Van der Westhuyzen DR, Goldstein JL, Brown MS, Russell DW. Three direct repeats and a TATA-like sequence are required for regulated expression of the human low density lipoprotein receptor gene. Journal of Biological Chemistry. 1987;262(22):10773-9.
61. Dawson P, Hofmann S, Van Der Westhuyzen D, Südhof T, Brown M, Goldstein J. Sterol-dependent repression of low density lipoprotein receptor promoter mediated by 16-base pair sequence adjacent to binding site for transcription factor Sp1. Journal of Biological Chemistry. 1988;263(7):3372-9.
62. Sanchez HB, Yieh L, Osborne TF. Cooperation by Sterol Regulatory Element-binding Protein and Sp1 in Sterol Regulation of Low Density Lipoprotein Receptor Gene. Journal of Biological Chemistry. 1995;270(3):1161-9.
63. Smith JR, Osborne TF, Goldstein JL, Brown MS. Identification of nucleotides responsible for enhancer activity of sterol regulatory element in low density lipoprotein receptor gene. Journal of Biological Chemistry. 1990;265(4):2306-10.
64. Schaefer JR, Kurt B, Sattler A, Klaus G, Soufi M. Pharmacogenetic aspects in familial hypercholesterolemia with the special focus on FH(Marburg) (FH p.W556R). Clinical Research in Cardiology Supplements. 2012;7(Suppl 1):2-6.
65. Leigh S, Futema M, Whittall R, Taylor-Beadling A, Williams M, den Dunnen JT, et al. The UCL low-density lipoprotein receptor gene variant database: pathogenicity update. Journal of Medical Genetics. 2016.
66. Huang LS, Miller DA, Bruns GA, Breslow JL. Mapping of the human APOB gene to chromosome 2p and demonstration of a two-allele restriction fragment length polymorphism. Proceedings of the National Academy of Sciences. 1986;83(3):644-8.
67. Law SW, Lee N, Monge JC, Brewer Jr HB, Sakaguchi AY, Naylor SL. Human apoB-100 gene resides in the p23 → pter region of chromosome 2. Biochemical and Biophysical Research Communications. 1985;131(2):1003-12.
68. Mahley RW, Innerarity TL. Lipoprotein receptors and cholesterol homeostasis. Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes. 1983;737(2):197-222.
69. Segrest JP, Jones MK, De Loof H, Dashti N. Structure of apolipoprotein B-100 in low density lipoproteins. Journal of Lipid Research. 2001;42(9):1346-67.
70. Whitfield AJ, Barrett PHR, van Bockxmeer FM, Burnett JR. Lipid Disorders and Mutations in the APOB Gene. Clinical Chemistry. 2004;50(10):1725-32.
71. Boren J, Lee I, Zhu W, Arnold K, Taylor S, Innerarity TL. Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its

binding activity by the carboxyl terminus in familial defective apo-B100. The Journal of Clinical Investigation. 1998;101(5):1084-93.
72. Innerarity TL, Mahley RW, Weisgraber KH, Bersot TP, Krauss RM, Vega GL, et al. Familial defective apolipoprotein B-100: a mutation of apolipoprotein B that causes hypercholesterolemia. Journal of Lipid Research. 1990;31(8):1337-49.
73. Soria LF, Ludwig EH, Clarke HR, Vega GL, Grundy SM, McCarthy BJ. Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. Proceedings of the National Academy of Sciences. 1989;86(2):587-91.
74. Rabès JP, Varret M, Saint-Jore B, Erlich D, Jondeau G, Krempf M, et al. Familial ligand-defective apolipoprotein B-100: Simultaneous detection of the ARG3500→GLN and ARG3531→CYS mutations in a French population. Human Mutation. 1997;10(2):160-3.
75. Gaffney D, Reid JM, Cameron IM, Vass K, Caslake MJ, Shepherd J, et al. Independent Mutations at Codon 3500 of the Apolipoprotein B Gene Are Associated With Hyperlipidemia. Arteriosclerosis, Thrombosis, and Vascular Biology. 1995;15(8):1025-9.
76. Ludwig EH, McCarthy BJ. Haplotype analysis of the human apolipoprotein B mutation associated with familial defective apolipoprotein B100. American journal of human genetics. 1990;47(4):712.
77. Rauh G, Schuster H, Fischer J, Keller C, Wolfram G, Zöllner N. Familial defective apolipoprotein B-100: haplotype analysis of the arginine (3500)→ glutamine mutation. Atherosclerosis. 1991;88(2):219-26.
78. Leren T, Rosdningen O, Tonstad S, Rossby O, Urdal P, Ose L. Identification of the apo B-3500 mutation in the Norwegian population. Scandinavian journal of clinical and laboratory investigation. 1995;55(3):217-21.
79. Wenham PR, Bloomfield P, Blundell G, Penney MD, Rae PW, Walker SW. Familial defective apolipoprotein B-100: a study of patients from lipid clinics in Scotland and Wales. Annals of Clinical Biochemistry. 1996;33(5):443-50.
80. Tai D-Y, Pan J-P, Lee-Chen G-J. Identification and haplotype analysis of apolipoprotein B-100 Arg3500→ Trp mutation in hyperlipidemic Chinese. Clinical chemistry. 1998;44(8):1659-65.
81. Chiou K-R, Charng M-J. Genetic diagnosis of familial hypercholesterolemia in Han Chinese. Journal of Clinical Lipidology. 2016;10(3):490-6.
82. Choong M-L, Koay ES, Khoo K-L, Khaw M-C, Sethi SK. Denaturing gradient-gel electrophoresis screening of familial defective apolipoprotein B-100 in a mixed Asian cohort: two cases of arginine3500→ tryptophan mutation associated with a unique haplotype. Clinical chemistry. 1997;43(6):916-23.
83. Fouchier SW, Kastelein JJP, Defesche JC. Update of the molecular basis of familial hypercholesterolemia in The Netherlands. Human Mutation. 2005;26(6):550-6.
84. Borén J, Ekström U, Ågren B, Nilsson-Ehle P, Innerarity TL. The molecular mechanism for the genetic disorder familial defective apolipoprotein B100. Journal of Biological Chemistry. 2001;276(12):9214-8.
85. Schaefer JR, Scharnagl H, Baumstark MW, Schweer H, Zech LA, Seyberth H, et al. Homozygous familial defective apolipoprotein B-100. Arteriosclerosis, thrombosis, and vascular biology. 1997;17(2):348-53.
86. März W, Baumstark MW, Scharnagl H, Ruzicka V, Buxbaum S, Herwig J, et al. Accumulation of "small dense" low density lipoproteins (LDL) in a homozygous patients with familial defective apolipoprotein B-100 results from heterogenous interaction of LDL subfractions with the LDL receptor. Journal of Clinical Investigation. 1993;92(6):2922-33.
87. Sundaram M, Yao Z. Recent progress in understanding protein and lipid factors affecting hepatic VLDL assembly and secretion. Nutrition & Metabolism. 2010;7(1):35.

88. Liyanage KE, Hooper AJ, Defesche JC, Burnett JR, van Bockxmeer FM. High-resolution melting analysis for detection of familial ligand-defective apolipoprotein B-100 mutations. Annals of clinical biochemistry. 2008;45(2):170-6.
89. Alves AC, Etxebarria A, Soutar AK, Martin C, Bourbon M. Novel functional APOB mutations outside LDL-binding region causing familial hypercholesterolaemia. Human Molecular Genetics. 2014;23(7):1817-28.
90. Pullinger CR, Hennessy LK, Chatterton JE, Liu W, Love JA, Mendel CM, et al. Familial ligand-defective apolipoprotein B. Identification of a new mutation that decreases LDL receptor binding affinity. Journal of Clinical Investigation. 1995;95(3):1225-34.
91. Fernández-Higuero JA, Etxebarria A, Benito-Vicente A, Alves AC, Arrondo JLR, Ostolaza H, et al. Structural analysis of APOB variants, p.(Arg3527Gln), p.(Arg1164Thr) and p.(Gln4494del), causing Familial Hypercholesterolaemia provides novel insights into variant pathogenicity. Scientific Reports. 2015;5:18184.
92. Seidah N, Prat A. The proprotein convertases are potential targets in the treatment of dyslipidemia. J Mol Med. 2007;85(7):685-96.
93. Seidah NG, Benjannet S, Wickham L, Marcinkiewicz J, Jasmin SB, Stifani S, et al. The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): Liver regeneration and neuronal differentiation. Proceedings of the National Academy of Sciences. 2003;100(3):928-33.
94. Piper DE, Jackson S, Liu Q, Romanow WG, Shetterly S, Thibault ST, et al. The Crystal Structure of PCSK9: A Regulator of Plasma LDL-Cholesterol. Structure. 2007;15(5):545-52.
95. Naureckiene S, Ma L, Sreekumar K, Purandare U, Frederick Lo C, Huang Y, et al. Functional characterization of Narc 1, a novel proteinase related to proteinase K. Archives of Biochemistry and Biophysics. 2003;420(1):55-67.
96. Benjannet S, Rhainds D, Essalmani R, Mayne J, Wickham L, Jin W, et al. NARC-1/PCSK9 and Its Natural Mutants: Zymogen Cleavage and Effects on the Low Density Lipoprotein (LDL) Receptor And LDL Cholesterol
Journal of Biological Chemistry. 2004;279(47):48865-75.
97. Maxwell KN, Breslow JL. Adenoviral-mediated expression of Pcsk9 in mice results in a low-density lipoprotein receptor knockout phenotype. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(18):7100-5.
98. Park SW, Moon Y-A, Horton JD. Post-transcriptional Regulation of Low Density Lipoprotein Receptor Protein by Proprotein Convertase Subtilisin/Kexin Type 9a in Mouse Liver. Journal of Biological Chemistry. 2004;279(48):50630-8.
99. Seidah NG. PCSK9 as a therapeutic target of dyslipidemia. Expert Opinion on Therapeutic Targets. 2009;13(1):19-28.
100. Cunningham D, Danley DE, Geoghegan KF, Griffor MC, Hawkins JL, Subashi TA, et al. Structural and biophysical studies of PCSK9 and its mutants linked to familial hypercholesterolemia. Nature Structural & Molecular Biology. 2007;14(5):413-9.
101. Mayer G, Poirier S, Seidah NG. Annexin A2 Is a C-terminal PCSK9-binding Protein That Regulates Endogenous Low Density Lipoprotein Receptor Levels. Journal of Biological Chemistry. 2008;283(46):31791-801.
102. Seidah NG, Poirier S, Denis M, Parker R, Miao B, Mapelli C, et al. Annexin A2 is a natural extrahepatic inhibitor of the PCSK9-induced LDL receptor degradation. PloS one. 2012;7(7):e41865.
103. Kwon HJ, Lagace TA, McNutt MC, Horton JD, Deisenhofer J. Molecular basis for LDL receptor recognition by PCSK9. Proceedings of the National Academy of Sciences. 2008;105(6):1820-5.

104. Nassoury N, Blasiole DA, Tebon Oler A, Benjannet S, Hamelin J, Poupon V, et al. The Cellular Trafficking of the Secretory Proprotein Convertase PCSK9 and Its Dependence on the LDLR. Traffic. 2007;8(6):718-32.
105. Leren TP. Sorting an LDL receptor with bound PCSK9 to intracellular degradation. Atherosclerosis. 2014;237(1):76-81.
106. Tavori H, Fan D, Blakemore JL, Yancey PG, Ding L, Linton MF, et al. Serum Proprotein Convertase Subtilisin/Kexin Type 9 and Cell Surface Low-Density Lipoprotein Receptor Evidence for a Reciprocal Regulation. Circulation. 2013;127(24):2403-13.
107. Benn M, Nordestgaard BG, Grande P, Schnohr P, Tybjærg-Hansen A. PCSK9R46L, Low-Density Lipoprotein Cholesterol Levels, and Risk of Ischemic Heart Disease: 3 Independent Studies and Meta-Analyses. Journal of the American College of Cardiology. 2010;55(25):2833-42.
108. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nature genetics. 2005;37(2):161-5.
109. Mombelli G. Potential of PCSK9 as a new target for the management of LDL cholesterol. 2015.
110. Rashid S, Curtis DE, Garuti R, Anderson NN, Bashmakov Y, Ho YK, et al. Decreased plasma cholesterol and hypersensitivity to statins in mice lacking Pcsk9. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(15):5374-9.
111. Homer VM, Marais AD, Charlton F, Laurie AD, Hurndell N, Scott R, et al. Identification and characterization of two non-secreted PCSK9 mutants associated with familial hypercholesterolemia in cohorts from New Zealand and South Africa. Atherosclerosis. 2008;196(2):659-66.
112. Strøm TB, Holla ØL, Cameron J, Berge KE, Leren TP. Loss-of-function mutation R46L in the PCSK9 gene has little impact on the levels of total serum cholesterol in familial hypercholesterolemia heterozygotes. Clinica Chimica Acta. 2010;411(3–4):229-33.
113. Humphries SE, Neely RDG, Whittall RA, Troutt JS, Konrad RJ, Scartezini M, et al. Healthy Individuals Carrying the PCSK9 p.R46L Variant and Familial Hypercholesterolemia Patients Carrying PCSK9 p.D374Y Exhibit Lower Plasma Concentrations of PCSK9. Clinical Chemistry. 2009;55(12):2153.
114. Leren TP. Mutations in the PCSK9 gene in Norwegian subjects with autosomal dominant hypercholesterolemia. Clinical Genetics. 2004;65(5):419-22.
115. Pandit S, Wisniewski D, Santoro JC, Ha S, Ramakrishnan V, Cubbon RM, et al. Functional analysis of sites within PCSK9 responsible for hypercholesterolemia. Journal of Lipid Research. 2008;49(6):1333-43.
116. Fisher TS, Surdo PL, Pandit S, Mattu M, Santoro JC, Wisniewski D, et al. Effects of pH and Low Density Lipoprotein (LDL) on PCSK9-dependent LDL Receptor Regulation. Journal of Biological Chemistry. 2007;282(28):20502-12.
117. Ferri N. Proprotein Convertase Subtilisin/Kexin Type 9: From the Discovery to the Development of New Therapies for Cardiovascular Diseases. Scientifica. 2012;2012:21.
118. Blesa S, Vernia S, Garcia-Garcia A-B, Martinez-Hervas S, Ivorra C, Gonzalez-Albert V, et al. A new PCSK9 gene promoter variant affects gene expression and causes autosomal dominant hypercholesterolemia. The Journal of Clinical Endocrinology & Metabolism. 2008;93(9):3577-83.
119. Shioji K, Mannami T, Kokubo Y, Inamoto N, Takagi S, Goto Y, et al. Genetic variants in PCSK9 affect the cholesterol level in Japanese. Journal of human genetics. 2004;49(2):109-14.

120. Miyake Y, Kimura R, Kokubo Y, Okayama A, Tomoike H, Yamamura T, et al. Genetic variants in PCSK9 in the Japanese population: rare genetic variants in PCSK9 might collectively contribute to plasma LDL cholesterol levels in the general population. Atherosclerosis. 2008;196(1):29-36.
121. Fasano T, Cefalu AB, Di Leo E, Noto D, Pollaccia D, Bocchi L, et al. A novel loss of function mutation of PCSK9 gene in white subjects with low-plasma low-density lipoprotein cholesterol. Arteriosclerosis, thrombosis, and vascular biology. 2007;27(3):677-81.
122. Zhao Z, Tuakli-Wosornu Y, Lagace TA, Kinch L, Grishin NV, Horton JD, et al. Molecular characterization of loss-of-function mutations in PCSK9 and identification of a compound heterozygote. The American Journal of Human Genetics. 2006;79(3):514-23.
123. Berge KE, Ose L, Leren TP. Missense mutations in the PCSK9 gene are associated with hypocholesterolemia and possibly increased response to statin therapy. Arteriosclerosis, thrombosis, and vascular biology. 2006;26(5):1094-100.
124. Cariou B, Ouguerram K, Zaïr Y, Guerois R, Langhi C, Kourimate S, et al. PCSK9 dominant negative mutant results in increased LDL catabolic rate and familial hypobetalipoproteinemia. Arteriosclerosis, thrombosis, and vascular biology. 2009;29(12):2191-7.
125. Cameron J, Holla Ø, Laerdahl J, Kulseth M, Ranheim T, Rognes T, et al. Characterization of novel mutations in the catalytic domain of the PCSK9 gene. Journal of internal medicine. 2008;263(4):420-31.
126. Abifadel M, Varret M, Rabes J-P, Allard D, Ouguerram K, Devillers M, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nature Genetics 2003;34(2):154-6.
127. Allard D, Amsellem S, Abifadel M, Trillard M, Devillers M, Luc G, et al. Novel mutations of the PCSK9 gene cause variable phenotype of autosomal dominant hypercholesterolemia. Human Mutation. 2005;26(5):497.
128. Kotowski IK, Pertsemlidis A, Luke A, Cooper RS, Vega GL, Cohen JC, et al. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. The American Journal of Human Genetics. 2006;78(3):410-22.
129. Timms KM, Wagner S, Samuels ME, Forbey K, Goldfine H, Jammulapati S, et al. A mutation in PCSK9 causing autosomal-dominant hypercholesterolemia in a Utah pedigree. Human genetics. 2004;114(4):349-53.
130. Sun X-M, Eden ER, Tosi I, Neuwirth CK, Wile D, Naoumova RP, et al. Evidence for effect of mutant PCSK9 on apolipoprotein B secretion as the cause of unusually severe dominant hypercholesterolaemia. Human molecular genetics. 2005;14(9):1161-9.
131. Bourbon M, Alves A, Medeiros A, Silva S, Soutar A. Familial hypercholesterolaemia in Portugal. Atherosclerosis. 2008;196(2):633-42.
132. Pisciotta L, Oliva CP, Cefalù AB, Noto D, Bellocchio A, Fresa R, et al. Additive effect of mutations in LDLR and PCSK9 genes on the phenotype of familial hypercholesterolemia. Atherosclerosis. 2006;186(2):433-40.
133. Dubuc G, Tremblay M, Paré G, Jacques H, Hamelin J, Benjannet S, et al. A new method for measurement of total plasma PCSK9: clinical applications. Journal of lipid research. 2010;51(1):140-9.
134. Zuliani G, Arca M, Signore A, Bader G, Fazio S, Chianelli M, et al. Characterization of a New Form of Inherited Hypercholesterolemia: Familial Recessive Hypercholesterolemia. Arteriosclerosis, Thrombosis, and Vascular Biology. 1999;19(3):802-9.
135. Garcia CK, Wilund K, Arca M, Zuliani G, Fellin R, Maioli M, et al. Autosomal Recessive Hypercholesterolemia Caused by Mutations in a Putative LDL Receptor Adaptor Protein. Science. 2001;292(5520):1394-8.

136. Eden ER, Naoumova RP, Burden JJ, McCarthy MI, Soutar AK. Use of homozygosity mapping to identify a region on chromosome 1 bearing a defective gene that causes autosomal recessive homozygous hypercholesterolemia in two unrelated families. The American Journal of Human Genetics. 2001;68(3):653-60.
137. He G, Gupta S, Yi M, Michaely P, Hobbs HH, Cohen JC. ARH Is a Modular Adaptor Protein That Interacts with the LDL Receptor, Clathrin, and AP-2. Journal of Biological Chemistry. 2002;277(46):44044-9.
138. Masuhara M, Nagao K, Nishikawa M, Sasaki M, Yoshimura A, Osawa M. Molecular cloning of murine STAP-1, the stem-cell-specific adaptor protein containing PH and SH2 domains. Biochemical and biophysical research communications. 2000;268(3):697-703.
139. Hatanaka K, Tanishita H, Ishibashi-Ueda H, Yamamoto A. Hyperlipidemia in mast cell-deficient W/WV mice. Biochimica et Biophysica Acta - Lipids and Lipid Metabolism. 1986;878(3):440-5.
140. Smith ER, Yeasky T, Wei JQ, Miki RA, Cai KQ, Smedberg JL, et al. White spotting variant (Wv) mouse as an experimental model for ovarian aging and menopausal biology. Menopause 2012;19(5):588-96.
141. Marks D, Thorogood M, Neil HAW, Humphries SE. A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia. Atherosclerosis. 2003;168(1):1-14.
142. Wierzbicki AS, Humphries SE, Minhas R. Familial hypercholesterolaemia: summary of NICE guidance. British Medical Journal. 2008;337.
143. Yuan G, Wang J, Hegele RA. Heterozygous familial hypercholesterolemia: an underrecognized cause of early cardiovascular disease. Canadian Medical Association Journal. 2006;174(8):1124-9.
144. Liyanage KE, Burnett JR, Hooper AJ, van Bockxmeer FM. Familial hypercholesterolemia: epidemiology, Neolithic origins and modern geographic distribution. Critical Reviews in Clinical Laboratory Sciences. 2011;48(1):1-18.
145. Taylor A, Wang D, Patel K, Whittall R, Wood G, Farrer M, et al. Mutation detection rate and spectrum in familial hypercholesterolaemia patients in the UK pilot cascade project. Clinical Genetics. 2010;77.
146. Whittall RA, Scartezini M, Li K, Hubbart C, Reiner Z, Abraha A, et al. Development of a high-resolution melting method for mutation detection in familial hypercholesterolaemia patients. Annals of Clinical Biochemistry. 2010;47(1):44-55.
147. Taylor A, Martin B, Wang D, Patel K, Humphries SE, Norbury G. Multiplex ligation-dependent probe amplification analysis to screen for deletions and duplications of the LDLR gene in patients with familial hypercholesterolaemia. Clinical Genetics. 2009;76(1):69-75.
148. Futema M, Whittall RA, Kiley A, Steel LK, Cooper JA, Badmus E, et al. Analysis of the frequency and spectrum of mutations recognised to cause familial hypercholesterolaemia in routine clinical practice in a UK specialist hospital lipid clinic. Atherosclerosis. 2013;229(1):161-8.
149. Sharma P, Boyers D, Boachie C, Stewart F, Miedzybrodzka Z, Simpson W, et al. Elucigene FH20 and LIPOchip for the diagnosis of familial hypercholesterolaemia: a systematic review and economic evaluation. Health Technol Assess. 2012;16(17):1-266.
150. Leigh SE, Foster AH, Whittall RA, Hubbart CS, Humphries SE. Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. Annals of Human Genetics. 2008;72.
151. Futema M, Plagnol V, Li K, Whittall RA, Neil HAW, Seed M, et al. Whole exome sequencing of familial hypercholesterolaemia patients negative for LDLR/APOB/PCSK9 mutations. Journal of Medical Genetics. 2014;51(8):537-44.

152. Maglio C, Mancina R, Motta B, Stef M, Pirazzi C, Palacios L, et al. Genetic diagnosis of familial hypercholesterolaemia by targeted next-generation sequencing. Journal of internal medicine. 2014;276(4):396-403.
153. Radovica-Spalvina I, Latkovskis G, Silamikelis I, Fridmanis D, Elbere I, Ventins K, et al. Next-generation-sequencing-based identification of familial hypercholesterolemia-related mutations in subjects with increased LDL–C levels in a latvian population. BMC Medical Genetics. 2015;16:86.
154. Al-Allaf FA, Athar M, Abduljaleel Z, Taher MM, Khan W, Ba-hammam FA, et al. Next generation sequencing to identify novel genetic variants causative of autosomal dominant familial hypercholesterolemia associated with increased risk of coronary heart disease. Gene. 2015;565(1):76-84.
155. Vandrovcova J, Thomas ER, Atanur SS, Norsworthy PJ, Neuwirth C, Tan Y, et al. The use of next-generation sequencing in clinical diagnosis of familial hypercholesterolemia. Genetics in Medicine. 2013;15(12):948-57.
156. Talmud PJ, Futema M, Humphries SE. The genetic architecture of the familial hyperlipidaemia syndromes: rare mutations and common variants in multiple genes. Current opinion in lipidology. 2014;25(4):274-81.
157. Futema M, Plagnol V, Whittall RA, Neil HAW, Humphries SE. Use of targeted exome sequencing as a diagnostic tool for Familial Hypercholesterolaemia. Journal of medical genetics. 2012;49(10):644-9.
158. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends in genetics. 2001;17(9):502-10.
159. Murcray CE, Lewinger JP, Gauderman WJ. Gene-Environment Interaction in Genome-Wide Association Studies. American Journal of Epidemiology. 2009;169(2):219-26.
160. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010;466(7307):707-13.
161. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. Nature Genetics 2013;45(11):1274-83.
162. Almal SH, Padh H. Implications of gene copy-number variation in health and diseases. Journal of Human Genetics. 2012;57(1):6-13.
163. Talmud PJ, Drenos F, Shah S, Shah T, Palmen J, Verzilli C, et al. Gene-centric Association Signals for Lipids and Apolipoproteins Identified via the Human CVD BeadChip. American Journal of Human Genetics. 2009;85(5):628-42.
164. Talmud PJ, Shah S, Whittall R, Futema M, Howard P, Cooper JA, et al. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case–control study. Lancet. 2013;381.
165. Welder G, Zineh I, Pacanowski MA, Troutt JS, Cao G, Konrad RJ. High-dose atorvastatin causes a rapid sustained increase in human serum PCSK9 and disrupts its correlation with LDL cholesterol. Journal of Lipid Research. 2010;51(9):2714-21.
166. Ezzahti M, Sijbrands E, Mulder M, van Lennep JR. familial hypercholesterolaemia: new treatment options. 2013.
167. Stein EA, Giugliano RP, Koren MJ, Raal FJ, Roth EM, Weiss R, et al. Efficacy and safety of evolocumab (AMG 145), a fully human monoclonal antibody to PCSK9, in hyperlipidaemic patients on various background lipid therapies: pooled analysis of 1359 patients in four phase 2 trials. European Heart Journal. 2014;35(33):2249-59.
168. Foltz IN, Karow M, Wasserman SM. Evolution and Emergence of Therapeutic Monoclonal Antibodies: What Cardiologists Need to Know. Circulation. 2013;127(22):2222-30.

169. Catapano AL, Papadopoulos N. The safety of therapeutic monoclonal antibodies: Implications for cardiovascular disease and targeting the PCSK9 pathway. Atherosclerosis. 2013;228(1):18-28.
170. Spano F, Raugei G, Palla E, Colella C, Melli M. Characterization of the human lipocortin-2-encoding multigene family: its structure suggests the existence of a short amino acid unit undergoing duplication. Gene. 1990;95(2):243-51.
171. Gerke V, Moss SE. Annexins: From Structure to Function. Physiological Reviews. 2002;82(2):331-71.
172. Gerke V, Creutz CE, Moss SE. Annexins: linking Ca2+ signalling to membrane dynamics. Nature reviews Molecular cell biology. 2005;6(6):449-61.
173. He K-L, Deora AB, Xiong H, Ling Q, Weksler BB, Niesvizky R, et al. Endothelial Cell Annexin A2 Regulates Polyubiquitination and Degradation of Its Binding Partner S100A10/p11. Journal of Biological Chemistry. 2008;283(28):19192-200.
174. Ishii H, Yoshida M, Hiraoka M, Hajjar KA, Tanaka A, Yasukochi Y, et al. Recombinant Annexin II Modulates Impaired Fibrinolytic Activity In Vitro and in Rat Carotid Artery. Circulation Research. 2001;89(12):1240-5.
175. Hajjar KA, Jacovina AT, Chacko J. An endothelial cell receptor for plasminogen/tissue plasminogen activator. I. Identity with annexin II. Journal of Biological Chemistry. 1994;269(33):21191-7.
176. Johnsson N, Marriott G, Weber K. p36, the major cytoplasmic substrate of src tyrosine protein kinase, binds to its p11 regulatory subunit via a short amino-terminal amphiphatic helix. The European Molecular Biology Organization. 1988;7(8):2435-42.
177. Filipenko NR, MacLeod TJ, Yoon C-S, Waisman DM. Annexin A2 Is a Novel RNA-binding Protein. Journal of Biological Chemistry. 2004;279(10):8723-31.
178. Kassam G, Choi K-S, Ghuman J, Kang H-M, Fitzpatrick SL, Zackson T, et al. The Role of Annexin II Tetramer in the Activation of Plasminogen. Journal of Biological Chemistry. 1998;273(8):4790-9.
179. Cesarman GM, Guevara CA, Hajjar KA. An endothelial cell receptor for plasminogen/tissue plasminogen activator (t-PA). II. Annexin II-mediated enhancement of t-PA-dependent plasminogen activation. Journal of Biological Chemistry. 1994;269(33):21198-203.
180. Hajjar KA, Mauri L, Jacovina AT, Zhong F, Mirza UA, Padovan JC, et al. Tissue Plasminogen Activator Binding to the Annexin II Tail Domain: Direct Modulation by Homocysteine. Journal of Biological Chemistry. 1998;273(16):9987-93.
181. Jost M, Thiel C, Weber K, Gerke V. Mapping of three unique Ca2+-binding sites in human annexin II. European Journal of Biochemistry. 1992;207(3):923-30.
182. Becker T, Weber K, Johnsson N. Protein-protein recognition via short amphiphilic helices; a mutational analysis of the binding site of annexin II for p11. The European Molecular Biology Organization Journal. 1990;9(13):4207-13.
183. Rety S, Sopkova J, Renouard M, Osterloh D, Gerke V, Tabaries S, et al. The crystal structure of a complex of p11 with the annexin II N-terminal peptide. Nature Structural & Molecular Biology. 1999;6(1):89-95.
184. Kube E, Becker T, Weber K, Gerke V. Protein-protein interaction studied by site-directed mutagenesis. Characterization of the annexin II-binding site on p11, a member of the S100 protein family. Journal of Biological Chemistry. 1992;267(20):14175-82.
185. Madureira PA, O'Connell PA, Surette AP, Miller VA, Waisman DM. The Biochemistry and Regulation of S100A10: A Multifunctional Plasminogen Receptor Involved in Oncogenesis. Journal of Biomedicine and Biotechnology. 2012;2012:353687.

186. Burger A, Berendes R, Liemann S, Benz J, Hofmann A, Göttig P, et al. The Crystal Structure and Ion Channel Activity of Human Annexin II, a Peripheral Membrane Protein. Journal of Molecular Biology. 1996;257(4):839-47.
187. Madureira PA, Surette AP, Phipps KD, S.Taboski MA, Miller VA, Waisman DM. The role of the annexin A2 heterotetramer in vascular fibrinolysis. Blood. 2011;118(18):4789-97.
188. Bharadwaj A, Bydoun M, Holloway R, Waisman D. Annexin A2 Heterotetramer: Structure and Function. International journal of molecular sciences. 2013;14(3):6259-305.
189. Zhang D-W, Garuti R, Tang W-J, Cohen JC, Hobbs HH. Structural requirements for PCSK9-mediated degradation of the low-density lipoprotein receptor. Proceedings of the National Academy of Sciences. 2008;105(35):13045-50.
190. Luna Saavedra YG, Day R, Seidah NG. The M2 Module of the Cys-His-rich Domain (CHRD) of PCSK9 Protein Is Needed for the Extracellular Low-density Lipoprotein Receptor (LDLR) Degradation Pathway. Journal of Biological Chemistry. 2012;287(52):43492-501.
191. Holla ØL, Cameron J, Tveten K, Strøm TB, Berge KE, Laerdahl JK, et al. Role of the C-terminal domain of PCSK9 in degradation of the LDL receptors. Journal of Lipid Research. 2011;52(10):1787-94.
192. Ly K, Luna Saavedra YG, Canuel M, Routhier S, Desjardins R, Hamelin J, et al. Annexin A2 Reduces PCSK9 Protein Levels via a Translational Mechanism and Interacts with the M1 and M2 Domains of PCSK9. Journal of Biological Chemistry. 2014;289(25):17732-46.
193. Sharma M, Ansari MT, Abou-Setta AM, Soares-Weiser K, Ooi TC, Sears M, et al. Systematic Review: Comparative Effectiveness and Harms of Combination Therapy and Monotherapy for Dyslipidemia. Annals of Internal Medicine. 2009;151(9):622-30.
194. Pandor A, Ara R, Tumur I, Wilkinson A, Paisley S, Duenas A, et al. Ezetimibe monotherapy for cholesterol lowering in 2,722 people: systematic review and meta-analysis of randomized controlled trials. Journal of internal medicine. 2009;265(5):568-80.
195. Insull Jr W. Clinical utility of bile acid sequestrants in the treatment of dyslipidemia: a scientific review. Southern medical journal. 2006;99(3):257-73.
196. Carlson LA. Nicotinic acid: the broad-spectrum lipid drug. A 50th anniversary review. Journal of Internal Medicine. 2005;258(2):94-114.
197. Brautbar A, Ballantyne CM. Pharmacological strategies for lowering LDL cholesterol: statins and beyond. Nature Reviews Cardiology. 2011;8(5):253-65.
198. Jones PH, Davidson MH, Kashyap ML, Kelly MT, Buttler SM, Setze CM, et al. Efficacy and safety of ABT-335 (fenofibric acid) in combination with rosuvastatin in patients with mixed dyslipidemia: A phase 3 study. Atherosclerosis. 2009;204(1):208-15.
199. Rynkiewicz A, Cybulska B, Banach M, Filipiak K, Guzik T, Idzior-Waluś B, et al. Management of familial heterozygous hypercholesterolemia: Position Paper of the Polish Lipid Expert Forum. Journal of Clinical Lipidology. 2013;7(3):217-21.
200. Cuchel M, Bruckert E, Ginsberg HN, Raal FJ, Santos RD, Hegele RA, et al. Homozygous familial hypercholesterolaemia: new insights and guidance for clinicians to improve detection and clinical management. A position paper from the Consensus Panel on Familial Hypercholesterolaemia of the European Atherosclerosis Society. European heart journal. 2014;35(32):2146-57.
201. Sjouke B, Kusters DM, Kindt I, Besseling J, Defesche JC, Sijbrands EJG, et al. Homozygous autosomal dominant hypercholesterolaemia in the Netherlands: prevalence, genotype–phenotype relationship, and clinical outcome. European heart journal. 2015;36(9):560.

202. Cuchel M, Meagher EA, du Toit TH, Blom DJ, Marais AD, Hegele RA, et al. Efficacy and Safety of a Microsomal Triglyceride Transfer Protein Inhibitor in Homozygous Familial Hypercholesterolemia. Lancet (London, England). 2013;381(9860):40-6.
203. Græsdal A, Bogsrud MP, Holven KB, Nenseter MS, Narverud I, Langslet G, et al. Apheresis in homozygous familial hypercholesterolemia: The results of a follow-up of all Norwegian patients with homozygous familial hypercholesterolemia. Journal of Clinical Lipidology. 2012;6(4):331-9.
204. Stein EA, Honarpour N, Wasserman SM, Xu F, Scott R, Raal FJ. Effect of the Proprotein Convertase Subtilisin/Kexin 9 Monoclonal Antibody, AMG 145, in Homozygous Familial Hypercholesterolemia. Circulation. 2013;128(19):2113-20.
205. Marais AD. Familial Hypercholesterolaemia. The Clinical Biochemist Reviews. 2004;25(1):49-68.
206. Moorjani S, Torres A, Gagn C, Brun D, Lupien P, Roy M, et al. Mutations of low-density-lipoprotein-receptor gene, variation in plasma cholesterol, and expression of coronary heart disease in homozygous familial hypercholesterolaemia. The Lancet. 1993;341(8856):1303-6.
207. Ito MK, Watts GF. Challenges in the Diagnosis and Treatment of Homozygous Familial Hypercholesterolemia. Drugs. 2015;75(15):1715-24.
208. Thompson GR, Catapano A, Saheb S, Atassi-Dumont M, Barbir M, Eriksson M, et al. Severe hypercholesterolaemia: therapeutic goals and eligibility criteria for LDL apheresis in Europe. Current Opinion in Lipidology. 2010;21(6):492-8.
209. Bambauer R, Schiel R, Latza R. Low-density Lipoprotein Apheresis: An Overview. Therapeutic Apheresis and Dialysis. 2003;7(4):382-90.
210. Armstrong VW, Schleef J, Thiery J, Muche R, Schuff-Werner P, Eisenhauer T, et al. Effect of HELP-LDL-apheresis on serum concentrations of human lipoprotein(a): kinetic analysis of the post-treatment return to baseline levels. European Journal of Clinical Investigation. 1989;19(3):235-40.
211. Schuff-Werner P, Gohlke H, Bartmann U, Baggio G, Corti MC, Dinsenbacher A, et al. The HELP-LDL-apheresis multicentre study, an angiographically assessed trial on the role of LDL-apheresis in the secondary prevention of coronary heart disease. II. Final evaluation of the effect of regular treatment on LDL-cholesterol plasma concentrations and the course of coronary heart disease*. European Journal of Clinical Investigation. 1994;24(11):724-32.
212. Yamamoto A, Kawaguchi A, Harada-Shiba M, Tsushima M, Kojima S-i. Apheresis Technology for Prevention and Regression of Atherosclerosis: An Overview. Therapeutic Apheresis. 1997;1(3):233-41.
213. Gordon BR, Kelsey SF, Bilheimer DW, Brown DC, Dau PC, Gotto AM, et al. Treatment of refractory familial hypercholesterolemia by low-density lipoprotein apheresis using an automated dextran sulfate cellulose adsorption system. The American Journal of Cardiology. 1992;70(11):1010-6.
214. Tatamia R, Inoue N, Itoh H, Kishino B, Koga N, Nakashima Y, et al. Regression of coronary atherosclerosis by combined LDL-apheresis and lipid-lowering drug therapy in patients with familial hypercholesterolemia: a multicenter study. Atherosclerosis. 1992;95(1):1-13.
215. Kitabatake A, Sato H, Hori M, Kamada T, Kubori S, Hoki N, et al. Coronary atherosclerosis reduced in patients with familial hypercholesterolemia after intensive cholesterol lowering with low-density lipoprotein-apheresis: 1-year follow-up study. The Osaka LDL-Apheresis Multicenter Trial Group. Clinical Therapeutics. 1994;16(3):416-28.
216. Nishiura S. L-CAPS Group. Regression of coronary artery disease in familial hypercholesterolemia: Japan LDL-apheresis Coronary Atherosclerosis Prospective Study (L-

CAPS), L-CAPS Group, Department of Cardiology, Yokohama Rosal Hospital. Therapeutic apheresis and dialysis. 2003;7.
217. Raal FJ, Santos RD, Blom DJ, Marais AD, Charng M-J, Cromwell WC, et al. Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous familial hypercholesterolaemia: a randomised, double-blind, placebo-controlled trial. The Lancet. 2010;375(9719):998-1006.
218. Santos RD, Duell PB, East C, Guyton JR, Moriarty PM, Chin W, et al. Long-term efficacy and safety of mipomersen in patients with familial hypercholesterolaemia: 2-year interim results of an open-label extension. European Heart Journal. 2015;36(9):566-75.
219. Crooke ST, Geary RS. Clinical pharmacological properties of mipomersen (Kynamro), a second generation antisense inhibitor of apolipoprotein B. British Journal of Clinical Pharmacology. 2013;76(2):269-76.
220. Sahebkar A, Watts GF. New LDL-Cholesterol Lowering Therapies: Pharmacology, Clinical Trials, and Relevance to Acute Coronary Syndromes. Clinical Therapeutics. 2013;35(8):1082-98.
221. Faiz F, Hooper AJ, van Bockxmeer FM. Molecular pathology of familial hypercholesterolemia, related dyslipidemias and therapies beyond the statins. Critical Reviews in Clinical Laboratory Sciences. 2012;49(1):1-17.
222. Hussain MM, Rava P, Walsh M, Rana M, Iqbal J. Multiple functions of microsomal triglyceride transfer protein. Nutrition & Metabolism. 2012;9:14-.
223. Shan L, Pang L, Zhang R, Murgolo NJ, Lan H, Hedrick JA. PCSK9 binds to multiple receptors and can be functionally inhibited by an EGF-A peptide. Biochemical and Biophysical Research Communications. 2008;375(1):69-73.
224. McNutt MC, Kwon HJ, Chen C, Chen JR, Horton JD, Lagace TA. Antagonism of Secreted PCSK9 Increases Low Density Lipoprotein Receptor Expression in HepG2 Cells. Journal of Biological Chemistry. 2009;284(16):10561-70.
225. Du F, Hui Y, Zhang M, Linton MF, Fazio S, Fan D. Novel Domain Interaction Regulates Secretion of Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9) Protein. Journal of Biological Chemistry. 2011;286(50):43054-61.
226. Graham MJ, Lemonidis KM, Whipple CP, Subramaniam A, Monia BP, Crooke ST, et al. Antisense inhibition of proprotein convertase subtilisin/kexin type 9 reduces serum LDL in hyperlipidemic mice. Journal of Lipid Research. 2007;48(4):763-7.
227. Gupta N, Fisker N, Asselin M-C, Lindholm M, Rosenbohm C, Ørum H, et al. A Locked Nucleic Acid Antisense Oligonucleotide (LNA) Silences PCSK9 and Enhances LDLR Expression In Vitro and In Vivo. PLoS ONE. 2010;5(5):e10682.
228. Frank-Kamenetsky M, Grefhorst A, Anderson NN, Racie TS, Bramlage B, Akinc A, et al. Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates. Proceedings of the National Academy of Sciences. 2008;105(33):11915-20.
229. Cameron J, Ranheim T, Kulseth MA, Leren TP, Berge KE. Berberine decreases PCSK9 expression in HepG2 cells. Atherosclerosis. 2008;201(2):266-73.
230. Kong W, Wei J, Abidi P, Lin M, Inaba S, Li C, et al. Berberine is a novel cholesterol-lowering drug working through a unique mechanism distinct from statins. Nature Medicine 2004;10(12):1344-51.
231. Lipovšek D. Adnectins: engineered target-binding protein therapeutics. Protein Engineering, Design and Selection. 2011;24(1-2):3-9.
232. Dias CS, Shaywitz AJ, Wasserman SM, Smith BP, Gao B, Stolman DS, et al. Effects of AMG 145 on Low-Density Lipoprotein Cholesterol Levels: Results From 2 Randomized, Double-Blind, Placebo-Controlled, Ascending-Dose Phase 1 Studies in Healthy Volunteers

and Hypercholesterolemic Subjects on Statins. Journal of the American College of Cardiology. 2012;60(19):1888-98.
233. Stein EA, Mellis S, Yancopoulos GD, Stahl N, Logan D, Smith WB, et al. Effect of a Monoclonal Antibody to PCSK9 on LDL Cholesterol. New England Journal of Medicine. 2012;366(12):1108-18.
234. Blom DJ, Hala T, Bolognese M, Lillestol MJ, Toth PD, Burgess L, et al. A 52-Week Placebo-Controlled Trial of Evolocumab in Hyperlipidemia. New England Journal of Medicine. 2014;370(19):1809-19.
235. Giugliano RP, Desai NR, Kohli P, Rogers WJ, Somaratne R, Huang F, et al. Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 in combination with a statin in patients with hypercholesterolaemia (LAPLACE-TIMI 57): a randomised, placebo-controlled, dose-ranging, phase 2 study. The Lancet.380(9858):2007-17.
236. Roth EM, Taskinen M-R, Ginsberg HN, Kastelein JJP, Colhoun HM, Robinson JG, et al. Monotherapy with the PCSK9 inhibitor alirocumab versus ezetimibe in patients with hypercholesterolemia: Results of a 24 week, double-blind, randomized Phase 3 trial. International Journal of Cardiology. 2014;176(1):55-61.
237. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. New England Journal of Medicine. 2017;376(18):1713-22.
238. Landlinger C, Pouwer MG, Juno C, van der Hoorn JWA, Pieterman EJ, Jukema JW, et al. The AT04A vaccine against proprotein convertase subtilisin/kexin type 9 reduces total cholesterol, vascular inflammation, and atherosclerosis in APOE*3Leiden.CETP mice. European Heart Journal. 2017;38(32):2499-507.
239. Ogura M. PCSK9 inhibition in the management of familial hypercholesterolemia. Journal of Cardiology. 2017.
240. Fitzgerald K, White S, Borodovsky A, Bettencourt BR, Strahs A, Clausen V, et al. A Highly Durable RNAi Therapeutic Inhibitor of PCSK9. New England Journal of Medicine. 2017;376(1):41-51.
241. Wang X, Raghavan A, Chen T, Qiao L, Zhang Y, Ding Q, et al. CRISPR-Cas9 Targeting of PCSK9 in Human Hepatocytes In Vivo—Brief ReportSignificance. Arteriosclerosis, Thrombosis, and Vascular Biology. 2016;36(5):783-6.
242. Sive JI, Göttgens B. Transcriptional network control of normal and leukaemic haematopoiesis. Experimental Cell Research. 2014(0).
243. Clapier CR, Cairns BR. The Biology of Chromatin Remodeling Complexes. Annual Review of Biochemistry. 2009;78(1):273-304.
244. Aalfs JD, Kingston RE. What does ‘chromatin remodeling’mean? Trends in Biochemical Sciences. 2000;25(11):548-55.
245. Locker J. Transcription factors: Gulf Professional Publishing; 2001.
246. Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science. 1989;245(4916):371-8.
247. Tjian R, Maniatis T. Transcriptional activation: a complex puzzle with few easy pieces. Cell. 1994;77(1):5-8.
248. Triezenberg SJ. Structure and function of transcriptional activation domains. Current Opinion in Genetics & Development. 1995;5(2):190-6.
249. Mulier MM, Gerster T, Schaffner W. Enhancer sequences and the regulation of gene transcription. European Journal of Biochemistry. 1988;176(3):485-95.

250. Amouyal M. The remote control of transcription, DNA looping and DNA compaction. Biochimie. 1991;73(10):1261-8.
251. Berk AJ. Activation of RNA polymerase II transcription. Current Opinion in Cell Biology. 1999;11(3):330-5.
252. Kadonaga JT. Regulation of RNA Polymerase II Transcription by Sequence-Specific DNA Binding Factors. Cell. 2004;116(2):247-57.
253. Schleif R. DNA looping. Annual Review of Biochemistry. 1992;61(1):199-223.
254. Li R, Knight JD, Jackson SP, Tjian R, Botchan MR. Direct interaction between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. Cell. 1991;65(3):493-505.
255. Wamstad JA, Wang X, Demuren OO, Boyer LA. Distal enhancers: new insights into heart development and disease. Trends in Cell Biology. 2014;24(5):294-302.
256. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010;466(7307):714-9.
257. Darlington GJ, Wang N, Hanson RW. C/EBPα: a critical regulator of genes governing integrative metabolic processes. Current Opinion in Genetics & Development. 1995;5(5):565-70.
258. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. Nature Reviews Genetics. 2004;5(7):522-31.
259. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. cell. 2004;116(2):281-97.
260. Carrington JC, Ambros V. Role of microRNAs in plant and animal development. Science. 2003;301(5631):336-8.
261. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215-33.
262. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific microRNAs from mouse. Current Biology. 2002;12(9):735-9.
263. Moore KJ, Rayner KJ, Suárez Y, Fernández-Hernando C. microRNAs and cholesterol metabolism. Trends in Endocrinology & Metabolism. 2010;21(12):699-706.
264. Yang Z, Cappello T, Wang L. Emerging role of microRNAs in lipid metabolism. Acta Pharmaceutica Sinica B. 2015;5(2):145-50.
265. Esau C, Davis S, Murray SF, Yu XX, Pandey SK, Pear M, et al. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. Cell metabolism. 2006;3(2):87-98.
266. Elmén J, Lindow M, Schütz S, Lawrence M, Petri A, Obad S, et al. LNA-mediated microRNA silencing in non-human primates. Nature. 2008;452(7189):896-9.
267. Iliopoulos D, Drosatos K, Hiyama Y, Goldberg IJ, Zannis VI. MicroRNA-370 controls the expression of MicroRNA-122 and Cpt1α and affects lipid metabolism. Journal of lipid research. 2010;51(6):1513-23.
268. Thomas LF, Saito T, Sætrom P. Inferring causative variants in microRNA target sites. Nucleic acids research. 2011:gkr414.
269. The International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-45.
270. The International HapMap Consortium. The International HapMap Project. Nature. 2003;426(6968):789-96.
271. The International HapMap Consortium. A haplotype map of the human genome. Nature. 2005;437(7063):1299-320.
272. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851-61.

273. The International HapMap3 Consortium. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467(7311):52-8.
274. The Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
275. The ENCODE Consortium. (ENCyclopedia Of DNA Elements) Project. Science. 2004;306(5696):636-40.
276. The Encode Project Consortium. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). PLoS Biology. 2011;9(4):e1001046.
277. The Encode Project Consortium. An Integrated Encyclopedia of DNA Elements in the Human Genome. Nature. 2012;489(7414):57-74.
278. Puckelwartz MJ, McNally EM. Genetic Profiling for Risk Reduction in Human Cardiovascular Disease. Genes. 2014;5(1):214-34.
279. Center C. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11:241-7.
280. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53.
281. Global Lipid Genetic Consortium's [Available from: http://csg.sph.umich.edu/.
282. The CARDIoGRAM plus C4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. Nature genetics. 2013;45(1):25-33.
283. Manolio TA. Bringing genome-wide association findings into clinical use. Nat Rev Genet. 2013;14(8):549-58.
284. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotech. 2012;30(11):1095-106.
285. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics. 2013;45(6):580-5.
286. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 2015;348(6235):648-60.
287. Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, et al. The international HapMap project. Nature. 2003;426(6968):789-96.
288. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome research. 2002;12(6):996-1006.
289. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. Nucleic Acids Res. 2014;42.
290. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research. 2012;40(D1):D930-D4.
291. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. 2005;21(13):2933-42.
292. Dweep H, Sticht C, Pandey P, Gretz N. miRWalk – Database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes. Journal of Biomedical Informatics. 2011;44(5):839-47.
293. Liu C, Zhang F, Li T, Lu M, Wang L, Yue W, et al. MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. BioMed central genomics. 2012;13(1):661.
294. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human microRNA targets. PLoS Biol. 2004;2(11):e363.
295. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010;7(4):248-9.

296. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protocols. 2009;4(8):1073-81.
297. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nature methods. 2014;11(4):361-2.
298. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Research. 2013;41(D1):D808-D15.
299. Miller G, Bauer K, Barzegar S, Cooper J, Rosenberg R. Increased activation of the haemostatic system in men at high risk of fatal coronary heart disease. Thrombosis and haemostasis. 1996;75(5):767-71.
300. Ken-Dror G, Humphries S, Kumari M, Kivimaki M, Drenos F. A genetic instrument for Mendelian randomization of fibrinogen. Eur J Epidemiol. 2012;27(4):267-79.
301. Dorn GW, Cresci S. Genome-wide association studies of coronary artery disease and heart failure: where are we going? Pharmacogenomics. 2009;10(2):213-23.
302. Cooper JA, Miller GJ, Bauer KA, Morrissey JH, Meade TW, Howarth DJ, et al. Comparison of Novel Hemostatic Factors and Conventional Risk Factors for Prediction of Coronary Heart Disease. Circulation. 2000;102(23):2816-22.
303. Miller G, Bauer K, Barzegar S, Foley A, Mitchell J, Cooper J, et al. The effects of quality and timing of venepuncture on markers of blood coagulation in healthy middle-aged men. Thrombosis and haemostasis. 1995;73(1):82-6.
304. Marmot M, Brunner E. Cohort Profile: The Whitehall II study. International Journal of Epidemiology. 2005;34(2):251-6.
305. Marmot M, Steptoe A. Whitehall II and ELSA: Integrating epidemiological and psychobiological approaches to the assessment of biological indicators. National Research Council. 2008.
306. Drenos F, Talmud PJ, Casas JP, Smeeth L, Palmen J, Humphries SE, et al. Integrated associations of genotypes with multiple blood biomarkers linked to coronary heart disease risk. Human Molecular Genetics. 2009;18(12):2305-16.
307. Shah T, Engmann J, Dale C, Shah S, White J, Giambartolomei C, et al. Population Genomics of Cardiometabolic Traits: Design of the University College London-London School of Hygiene and Tropical Medicine-Edinburgh-Bristol (UCLEB) Consortium. PloS one. 2013;8(8):e71345.
308. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature genetics. 2012;44(8):955-9.
309. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013;45(6):580-5.
310. Folkersen L, Hooft Fvt, Chernogubova E, Agardh HE, Hansson GK, Hedin U, et al. Association of Genetic Risk Variants With Expression of Proximal Genes Identifies Novel Susceptibility Genes for Cardiovascular Disease. Circulation: Cardiovascular Genetics. 2010;3(4):365-73.
311. Smith AJP, Ahmed F, Nair D, Whittall R, Wang D, Taylor A, et al. A functional mutation in the LDLR promoter (-139C>G) in a patient with familial hypercholesterolemia. European journal Humen Genetics. 2007;15(11):1186-9.
312. Smith AJP, Humphries SE. Characterization of DNA-Binding Proteins Using Multiplexed Competitor EMSA. Journal of Molecular Biology. 2009;385(3):714-7.
313. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research. 1988;16(3):1215.

314. Usifo E, Leigh SEA, Whittall RA, Lench N, Taylor A, Yeats C, et al. Low-Density Lipoprotein Receptor Gene Familial Hypercholesterolemia Variant Database: Update and Pathological Assessment. Annals of Human Genetics. 2012;76(5):387-401.
315. Leitersdorf E, Van der Westhuyzen D, Coetzee G, Hobbs H. Two common low density lipoprotein receptor gene mutations cause familial hypercholesterolemia in Afrikaners. Journal of Clinical Investigation. 1989;84(3):954.
316. Varghese MJ. Familial hypercholesterolemia: A review. Annals of Pediatric Cardiology. 2014;7(2):107-17.
317. Mollaki V, Progias P, Drogari E. Familial Hypercholesterolemia in Greek children and their families: Genotype-to-phenotype correlations and a reconsideration of LDLR mutation spectrum. Atherosclerosis. 2014;237(2):798-804.
318. Lombardi MP, Redeker EJ, Defesche JC, Kamerling SW, Trip MD, Mannens MM, et al. Molecular genetic testing for familial hypercholesterolemia: spectrum of LDL receptor gene mutations in The Netherlands. Clinical genetics. 2000;57(2):116-24.
319. Norsworthy PJ, Vandrovcova J, Thomas ER, Campbell A, Kerr SM, Biggs J, et al. Targeted genetic testing for familial hypercholesterolaemia using next generation sequencing: a population-based study. BMC Medical Genetics. 2014;15(1):70.
320. Kessling AM, Seed M, Taylor R, Wynn V, Humphries SE. Rising cholesterol levels in children with Familial Hypercholesterolaemia. Biomedicine & Pharmacotherapy. 1990;44(7):373-9.
321. Santos PCJL, Morgan AC, Jannes CE, Turolla L, Krieger JE, Santos RD, et al. Presence and type of low density lipoprotein receptor (LDLR) mutation influences the lipid profile and response to lipid-lowering therapy in Brazilian patients with heterozygous familial hypercholesterolemia. Atherosclerosis. 2014;233(1):206-10.
322. Santos PCJL, Pereira AC. Type of LDLR mutation and the pharmacogenetics of familial hypercholesterolemia treatment. Pharmacogenomics. 2015;16(15):1743-50.
323. Lind S, Rystedt E, Eriksson M, Wiklund O, Angelin B, Eggertsen G. Genetic characterization of Swedish patients with familial hypercholesterolemia: a heterogeneous pattern of mutations in the LDL receptor gene. Atherosclerosis. 2002;163(2):399-407.
324. Langenhoven E, Warnich L, Thiart R, Rubinsztein DC, van der Westhuyzen DR, Marais AD, et al. Two novel point mutations causing receptor-negative familial hypercholesterolemia in a South African Indian homozygote. Atherosclerosis. 1996;125(1):111-9.
325. Fouchier SW, Defesche JC, Umans-Eckenhausen MA, Kastelein JJ. The molecular basis of familial hypercholesterolemia in The Netherlands. Human Genetics. 2001;109(6):602-15.
326. Salazar LA, Hirata MH, Cavalli SA, Nakandakare ER, Forti N, Diament J, et al. Molecular basis of familial hypercholesterolemia in Brazil: Identification of seven novel LDLR gene mutations. Human Mutation. 2002;19(4):462-3.
327. Mozas P, Castillo S, Tejedor D, Reyes G, Alonso R, Franco M, et al. Molecular characterization of familial hypercholesterolemia in Spain: Identification of 39 novel and 77 recurrent mutations in LDLR. Human Mutation. 2004;24(2):187-.
328. Innerarity TL, Weisgraber KH, Arnold KS, Mahley RW, Krauss RM, Vega GL, et al. Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding. Proceedings of the National Academy of Sciences. 1987;84(19):6919-23.
329. Gallagher JJ, Myant NB. The affinity of low-density lipoproteins and of very-low-density lipoprotein remnants for the low-density lipoprotein receptor in homozygous familial defective apolipoprotein B-100. Atherosclerosis. 1995;115(2):263-72.
330. Myant NB, Forbes SA, Day INM, Gallagher J. Estimation of the Age of the Ancestral Arginine3500→ Glutamine Mutation in Human ApoB-100. Genomics. 1997;45(1):78-87.

331. Hatmi ZN, Tahvildari S, Gafarzadeh Motlag A, Sabouri Kashani A. Prevalence of coronary artery disease risk factors in Iran: a population based survey. BMC Cardiovascular Disorders. 2007;7:32-.
332. Hovsepian S, Kelishadi R, Djalalinia S, Farzadfar F, Naderimagham S, Qorbani M. Prevalence of dyslipidemia in Iranian children and adolescents: A systematic review. Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences. 2015;20(5):503.
333. Nemati MH. Coronary revascularization in a child with homozygous familial hypercholesterolemia. Interactive cardiovascular and thoracic surgery. 2010;10(1):131-2.
334. Nemati MH, Astaneh B, Joubeh A. Triple coronary artery bypass graft in a 10-year-old child with familial hypercholesterolemia. General thoracic and cardiovascular surgery. 2009;57(2):94-7.
335. Fard-Esfahani P, Zeinali C, Rouhi Dehboneh S, Taghikhani M, Khatami S. A Novel Mutation in exon 4 of the low density lipoprotein (LDL) receptor gene in an Iranian Familial Hypercholesterolemia patient. Iranian Biomedical Journal. 2005;9(3):139-42.
336. Fard-Esfahani P, Mohammadi-Torbati P, Khatami S, Zeinali S, Allahyari MTM. Familial Defective Apoliporrotein B 100: Frequency of R3500q Mutation of Apoliporotein B Gene in Iranian Hypercholesterolemic Patients. Acta Medica Iranica. 2005;43(3):193-6.
337. Farrokhi E, Shayesteh F, Asadi Mobarakeh S, Roghani Dehkordi F, Ghatreh Samani K, Hashemzadeh Chaleshtori M. Molecular Characterization of Iranian Patients with Possible Familial Hypercholesterolemia. Indian Journal of Clinical Biochemistry. 2011;26(3):244-8.
338. Farrokhi E, Shirani M, Modarresi M, Roghani F, Hashemzadeh M. The study of mutations of the 9 exons of LDLR gene in patients with familial hypercholesterolemia in Cheharmahal Bakhtiari province. Arak Medical University Journal. 2011;13(4):30-7.
339. Al-Ashwal A, Alnouri F, Sabbour H, Al-Mahfouz A, Al-Sayed N, Razzaghy-Azar M, et al. Identification and treatment of patients with homozygous familial hypercholesterolaemia: information and recommendations from a Middle East advisory panel. Current vascular pharmacology. 2015;13(6):759-70.
340. Saadat M, Ansari-Lari M, Farhud D. Short report consanguineous marriage in Iran. Annals of human biology. 2004;31(2):263-9.
341. Damgaard D, Larsen ML, Nissen PH, Jensen JM, Jensen HK, Soerensen VR, et al. The relationship of molecular genetic to clinical diagnosis of familial hypercholesterolemia in a Danish population. Atherosclerosis. 2005;180(1):155-60.
342. Hartog C, Fryer A, Upadhyaya M. Mutation analysis of iduronate-2-sulphatase gene in 24 patients with Hunter syndrome: Characterisation of 6 novel mutations. Human Mutation. 1999;14(1):87-.
343. Sözen MM, Whittall R, Öner C, Tokatlı A, Kalkanoğlu HS, Dursun A, et al. The molecular basis of familial hypercholesterolaemia in Turkish patients. Atherosclerosis. 2005;180(1):63-71.
344. Schmidt HHJ, Tietge UJF, Buettner J, Barg-Hock H, Offner G, Schweitzer S, et al. Liver transplantation in a subject with familial hypercholesterolemia carrying the homozygous p.W577R LDL-receptor gene mutation. Clinical Transplantation. 2008;22(2):180-4.
345. Nissen H, Lestavel S, Hansen TS, Luc G, Bruckert E, Clavey V. Mutation screening of the LDLR gene and ApoB gene in patients with a phenotype of familial hypercholesterolemia and normal values in a functional LDL receptor/apolipoprotein B assay. Clinical Genetics. 1998;54(1):79-82.
346. Nobe Y, Emi M, Katsumata H, Nakajima T, Hirayama T, L. Wu L, et al. Familial Hypercholesterolemia in Utah Kindred with Novel 2412-6 Ins G Mutations in Exon 17 of the LDL Receptor Gene. Japanese Heart Journal. 1999;40(4):435-41.

347. Ekström, Abrahamson, Wallmark, Florén, Nilsson E. Mutations in the low-density lipoprotein receptor gene in Swedish familial hypercholesterolaemia patients: clinical expression and treatment response. European Journal of Clinical Investigation. 1998;28(9):740-7.
348. Kuhrová V, Francová H, Zapletalová P, Freiberger T, Fajkusová L, Hrabincová E, et al. Spectrum of low density lipoprotein receptor mutations in Czech hypercholesterolemic patients. Human Mutation. 2002;19(1):80-.
349. Yu W, Nohara A, Higashikata T, Lu H, Inazu A, Mabuchi H. Molecular genetic analysis of familial hypercholesterolemia: spectrum and regional difference of LDL receptor gene mutations in Japanese population. Atherosclerosis. 2002;165(2):335-42.
350. Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(44):15942-7.
351. Ollila J, Lappalainen I, Vihinen M. Sequence specificity in CpG mutation hotspots. FEBS Letters. 1996;396(2):119-22.
352. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics. 2012;13(7):484-92.
353. Zhi Y, Huang Y, Li Z, Zhang R, Wang S. Hypermethylation in promoter area of LDLR gene in atherosclerosis patients. Journal of molecular cell biology. 2007;40(6):419-27.
354. Hattori H, Hirayama T, Nobe Y, Nagano M, Kujiraoka T, Egashira T, et al. Eight novel mutations and functional impairments of the LDL receptor in familial hypercholesterolemia in the north of Japan. Journal of Human Genetics. 2002;47(2):80-7.
355. Leitersdorf E, Tobin EJ, Davignon J, Hobbs HH. Common low-density lipoprotein receptor mutations in the French Canadian population. Journal of Clinical Investigation. 1990;85(4):1014-23.
356. Lamon-Fava S, Diffenderfer MR, Marcovina SM. Lipoprotein(a) metabolism. Current Opinion in Lipidology. 2014;25(3):189-93.
357. Durrington PN, Schofield JD, Siahmansur T, Soran H. Lipoprotein (a): gene genie. Current Opinion in Lipidology. 2014;25(4):289-96.
358. Bos S, Yayha R, van Lennep JER. Latest developments in the treatment of lipoprotein (a). Current Opinion in Lipidology. 2014;25(6):452-60.
359. Orsó E, Ahrens N, ć D, Schmitz G. Familial hypercholesterolemia and lipoprotein(a) hyperlipidemia as independent and combined cardiovascular risk factors. Atherosclerosis Supplements. 2009;10(5):74-8.
360. Jaeger BR, Richter Y, Nagel D, Heigl F, Vogt A, Roeseler E, et al. Longitudinal cohort study on the effectiveness of lipid apheresis treatment to reduce high lipoprotein(a) levels and prevent major adverse coronary events. Nature Clinical Practice Cardiovascular Medicine. 2009;6(3):229-39.
361. Thompson GR. Recommendations for the use of LDL apheresis. Atherosclerosis. 2008;198(2):247-55.
362. Schmitz G, Orsó E. Lipoprotein(a) hyperlipidemia as cardiovascular risk factor: pathophysiological aspects. Clinical Research in Cardiology journal Supplements. 2015;10(Suppl 1):21-5.
363. Andersen LH, Miserez AR, Ahmad Z, Andersen RL. Familial defective apolipoprotein B-100: A review. Journal of Clinical Lipidology. 2016;10(6):1297-302.
364. Natesampillai S, Fernandez-Zapico ME, Urrutia R, Veldhuis JD. A Novel Functional Interaction between the Sp1-like Protein KLF13 and SREBP-Sp1 Activation Complex Underlies Regulation of Low Density Lipoprotein Receptor Promoter Function. Journal of Biological Chemistry. 2006;281(6):3040-7.

365. Liu J, Ahlborn TE, Briggs MR, Kraemer FB. Identification of a Novel Sterol-independent Regulatory Element in the Human Low Density Lipoprotein Receptor Promoter. Journal of Biological Chemistry. 2000;275(7):5214-21.
366. Khamis A, Palmen J, Lench N, Taylor A, Badmus E, Leigh S, et al. Functional analysis of four LDLR 5′UTR and promoter variants in patients with familial hypercholesterolaemia. European Journal of Human Genetics. 2015;23(6):790-5.
367. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190-5.
368. Donnelly P. Progress and challenges in genome-wide association studies in humans. Nature. 2008;456(7223):728-31.
369. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature reviews genetics. 2008;9(5):356-69.
370. Bradley DT, Hughes AE, Badger SA, Jones GT, Harrison SC, Wright BJ, et al. A variant in LDLR is associated with abdominal aortic aneurysm. Circulation: Cardiovascular Genetics. 2013;6(5):498-504.
371. Trompet S, de Craen A, Postmus I, Ford I, Sattar N, Caslake M, et al. Replication of LDL GWAs hits in PROSPER/PHASE as validation for future (pharmaco) genetic analyses. BMC Medical Genetics. 2011;12(1):131.
372. Kathiresan S, Willer CJ, Peloso G, Demissie S, Musunuru K, Schadt E, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nature genetics. 2009;41(1):56-65.
373. Anand SS, Xie C, Paré G, Montpetit A, Rangarajan S, McQueen MJ, et al. Genetic variants associated with myocardial infarction risk factors in over 8000 individuals from five ethnic groups: The INTERHEART Genetics Study. Circulation Cardiovascular genetics. 2009;2(1):16-25.
374. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nature Genetics. 2008;40(2):189-97.
375. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nature genetics. 2008;40(2):161-9.
376. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nature genetics. 2008;41(1):47-55.
377. Miljkovic I, Yerges-Armstrong LM, Kuller LH, Kuipers AL, Wang X, Kammerer CM, et al. Association analysis of 33 lipoprotein candidate genes in multi-generational families of African ancestry. Journal of lipid research. 2010;51(7):1823-31.
378. Rose A. Intron-mediated regulation of gene expression. Nuclear pre-mRNA processing in plants: Springer; 2008. p. 277-90.
379. Calhoun VC, Stathopoulos A, Levine M. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. Proceedings of the National Academy of Sciences. 2002;99(14):9243-7.
380. Su W, Jackson S, Tjian R, Echols H. DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes & development. 1991;5(5):820-6.
381. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analyses identifies 13 new susceptibility loci for coronary artery disease. Nature genetics. 2011;43(4):333-8.

382. The Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet. 2011;43(4):339-44.
383. Fairoozy RH, White J, Palmen J, Kalea AZ, Humphries SE. Identification of the Functional Variant(s) that Explain the Low-Density Lipoprotein Receptor (LDLR) GWAS SNP rs6511720 Association with Lower LDL-C and Risk of CHD. PLOS ONE. 2016;11(12):e0167676.
384. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336-7.
385. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic acids research. 2015;44(D1):D877-D81.
386. Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. UCSC genome browser tutorial. Genomics. 2008;92(2):75-84.
387. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(50):21931-6.
388. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics. 2007;39(3):311-8.
389. Calo E, Wysocka J. Modification of enhancer chromatin: what, how and why? Molecular cell. 2013;49(5).
390. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee B-K, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome research. 2011;21(10):1757-67.
391. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473(7345):43-9.
392. Mittal V, Ma B, Hernandez N. SNAPc: a core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain. Genes & development. 1999;13(14):1807-21.
393. Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. Genes & development. 1996;10(21):2657.
394. Sadowski CL, Henry RW, Kobayashi R, Hernandez N. The SNAP45 subunit of the small nuclear RNA (snRNA) activating protein complex is required for RNA polymerase II and III snRNA gene transcription and interacts with the TATA box binding protein. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(9):4289-93.
395. Amsellem S, Briffaut D, Carrié A, Rabès J, Girardet J, Fredenrich A, et al. Intronic mutations outside of Alu-repeat-rich domains of the LDL receptor gene are a cause of familial hypercholesterolemia. Human Genetics. 2002;111(6):501-10.
396. Chai J, Tarnawski AS. Serum response factor: discovery, biochemistry, biological roles and implications for tissue injury healing. Journal of physiology and pharmacology 2002.
397. Chai J, Tarnawski A. Serum response factor: discovery, biochemistry, biological roles and implications for tissue injury healing. Journal of physiology and pharmacology: an official journal of the Polish Physiological Society. 2002;53(2).

398. Pak YK. Serum response element-like sequences of the human low density lipoprotein receptor promoter: possible regulation sites for sterol-independent transcriptional activation. Biochemistry and Molecular Biology International journal. 1996;38(1):31-6.
399. Selvaraj A, Prywes R. Expression profiling of serum inducible genes identifies a subset of SRF target genes that are MKL dependent. BMC Molecular Biology. 2004;5(1):1-15.
400. Smaczniak C, Immink RGH, Angenent GC, Kaufmann K. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development. 2012;139(17):3081-98.
401. Dalton S, Treisman R. Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element. Cell. 1992;68(3):597-612.
402. Price MA, Rogers AE, Treisman R. Comparative analysis of the ternary complex factors Elk-1, SAP-1a and SAP-2 (ERP/NET). The EMBO Journal. 1995;14(11):2589-601.
403. Wang D-Z, Chang PS, Wang Z, Sutherland L, Richardson JA, Small E, et al. Activation of Cardiac Gene Expression by Myocardin, a Transcriptional Cofactor for Serum Response Factor. Cell. 2001;105(7):851-62.
404. Selvaraj A, Prywes R. Megakaryoblastic leukemia-1/2, a transcriptional co-activator of serum response factor, is required for skeletal myogenic differentiation. The Journal of Biological Chemistry. 2003;278.
405. Sepulveda JL, Vlahopoulos S, Iyer D, Belaguli N, Schwartz RJ. Combinatorial expression of GATA4, Nkx2-5, and serum response factor directs early cardiac gene activity. The Journal of Biological Chemistry 2002;277.
406. Sun K, Battle MA, Misra RP, Duncan SA. Hepatocyte expression of Serum Response Factor is essential for liver function, hepatocyte proliferation and survival, and postnatal body growth in mice. Hepatology (Baltimore, Md). 2009;49(5):1645-54.
407. Chawla A, Repa JJ, Evans RM, Mangelsdorf DJ. Nuclear receptors and lipid physiology: opening the X-files. Science. 2001;294(5548):1866-70.
408. Glass CK, Rosenfeld MG. The coregulator exchange in transcriptional functions of nuclear receptors. Genes & development. 2000;14(2):121-41.
409. Kishimoto M, Fujiki R, Takezawa S, Sasaki Y, Nakamura T, Yamaoka K, et al. Nuclear Receptor Mediated Gene Regulation through Chromatin Remodeling and Histone Modifications. Endocrine Journal. 2006;53(2):157-72.
410. de Lera AR, Bourguet W, Altucci L, Gronemeyer H. Design of selective nuclear receptor modulators: RAR and RXR as a case study. Nature Reviews Drug Discovery. 2007;6:811.
411. Niederreither K, Subbarayan V, Dolle P, Chambon P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. Nature Genetics. 1999;21(4):444-8.
412. Clarke N, Germain P, Altucci L, Gronemeyer H. Mode of action of RAR–RXR heterodimers. Expert Reviews in Molecular Medicine. 2004;6(25).
413. Xu L, Glass CK, Rosenfeld MG. Coactivator and corepressor complexes in nuclear receptor function. Current Opinion in Genetics & Development. 1999;9(2):140-7.
414. Darnell JE. STATs and gene regulation. Science. 1997;277(5332):1630-5.
415. Ramana CV, Chatterjee-Kishore M, Nguyen H, Stark GR. Complex roles of Stat1 in regulating gene expression. Oncogene. 2000;19(21).
416. Planas AM, Berruezo M, Justicia C, Barrón S, Ferrer I. Stat3 Is Present in the Developing and Adult Rat Cerebellum and Participates in the Formation of Transcription Complexes Binding DNA at the sis-Inducible Element. Journal of Neurochemistry. 1997;68(4):1345-51.

417. Qin J-Z, Kamarashev J, Zhang C-L, Dummer R, Burg G, Dobbeling U. Constitutive and Interleukin-7- and Interleukin-15-Stimulated DNA Binding of STAT and Novel Factors in Cutaneous T Cell Lymphoma Cells. 2001;117(3):583-9.
418. Richard AJ, Stephens JM. The role of JAK–STAT signaling in adipose tissue function. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease. 2014;1842(3):431-9.
419. Xu D, Yin C, Wang S, Xiao Y. JAK-STAT in lipid metabolism of adipocytes. JAK-STAT. 2013;2(4).
420. Chinetti G, Fruchart JC, Staels B. Peroxisome proliferator-activated receptors (PPARs): Nuclear receptors at the crossroads between lipid metabolism and inflammation. Inflamm res. 2000;49(10):497-505.
421. Korzus E, Torchia J, Rose DW, Xu L, Kurokawa R, McInerney EM, et al. Transcription Factor-Specific Requirements for Coactivators and Their Acetyltransferase Functions. Science. 1998;279(5351):703-7.
422. Ouchi T, Lee SW, Ouchi M, Aaronson SA, Horvath CM. Collaboration of signal transducer and activator of transcription 1 (STAT1) and BRCA1 in differential regulation of IFN-γ target genes. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(10):5208-13.
423. Zhang JJ, Vinkemeier U, Gu W, Chakravarti D, Horvath CM, Darnell JE. Two contact regions between Stat1 and CBP/p300 in interferon γ signaling. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(26):15092-6.
424. Zhu M-h, John S, Berg M, Leonard WJ. Functional Association of Nmi with Stat5 and Stat1 in IL-2- and IFN γ-Mediated Signaling. Cell. 1999;96(1):121-30.
425. Müller H-P, Sogo J, Schaffner W. An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. Cell. 1989;58(4):767-77.
426. Ptashne M. Gene regulation by proteins acting nearby and at a distance. Nature. 1985;322(6081):697-701.
427. Ptashne M. How eukaryotic transcriptional activators work. Nature. 1988;335(6192):683-9.
428. Ptashne M, Gann A. Transcriptional activation by recruitment. Nature. 1997;386(6625):569-77.
429. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295(5558):1306-11.
430. Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nature protocols. 2007;2(7):1722-33.
431. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, et al. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. Genome research. 2006;16(10):1299-309.
432. Poirier S, Mayer G, Poupon V, McPherson PS, Desjardins R, Ly K, et al. Dissection of the Endogenous Cellular Pathways of PCSK9-induced Low Density Lipoprotein Receptor Degradation: Evidence for an Intracellular Route. The Journal of Biological Chemistry. 2009;284(42):28856-64.
433. Hedhli N, Falcone DJ, Huang B, Cesarman-Maus G, Kraemer R, Zhai H, et al. The Annexin A2/S100A10 System in Health and Disease: Emerging Paradigms. Journal of Biomedicine and Biotechnology. 2012;2012:406273.
434. Grewal T, Enrich C. Annexins — Modulators of EGF receptor signalling and trafficking. Cellular Signalling. 2009;21(6):847-58.
435. Grewal T, Wason SJ, Enrich C, Rentero C. Annexins–insights from knockout-mice. Biological Chemistry. 2016;397(10):1031–53.

436. Oliver S. Proteomics: Guilt-by-association goes global. Nature. 2000;403(6770):601-3.
437. Gillis J, Pavlidis P. "Guilt by Association" Is the Exception Rather Than the Rule in Gene Networks. PLoS Computational Biology. 2012;8(3):e1002444.
438. Gao F, Ihn HE, Medina MW, Krauss RM. A common polymorphism in the LDL receptor gene has multiple effects on LDL receptor function. Human Molecular Genetics. 2013;22(7):1424-31.
439. Fairoozy RH, Cooper J, White J, Giambartolomei C, Folkersen L, Wannamethee SG, et al. Identifying low density lipoprotein cholesterol associated variants in the Annexin A2 (ANXA2) gene. Atherosclerosis. 2017;261:60-8.
440. Slatkin M. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics. 2008;9(6):477-85.
441. McDowell I, Pai A, Guo C, Vockley CM, Brown CD, Reddy TE, et al. Many long intergenic non-coding RNAs distally regulate mRNA gene expression levels. bioRxiv. 2016;044719.
442. Griner NB, Young D, Chaudhary P, Mohamed AA, Huang W, Chen Y, et al. ERG Oncoprotein Inhibits ANXA2 Expression and Function in Prostate Cancer. Molecular cancer research : MCR. 2015;13(2):368-79.
443. Nikolaev LG, Akopov SB, Didych DA, Sverdlov ED. Vertebrate Protein CTCF and its Multiple Roles in a Large-Scale Regulation of Genome Activity. Current Genomics. 2009;10(5):294-302.
444. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell. 2007;128(6):1231-45.
445. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012;489(7414):91-100.
446. Tarailo-Graovac M, Chen N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Current Protocols in Bioinformatics: John Wiley & Sons, Inc.; 2009.
447. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nature Genetics. 2013;45(10):1238-43.
448. Cariou B, Le May C, Costet P. Clinical aspects of PCSK9. Atherosclerosis. 2011;216(2):258-65.
449. Cui Q, Ju X, Yang T, Zhang M, Tang W, Chen Q, et al. Serum PCSK9 is associated with multiple metabolic factors in a large Han Chinese population. Atherosclerosis. 2010;213(2):632-6.
450. Baass A, Dubuc G, Tremblay M, Delvin EE, O'Loughlin J, Levy E, et al. Plasma PCSK9 is associated with age, sex, and multiple metabolic markers in a population-based sample of children and adolescents. Clinical Chemistry. 2009;55(9):1637-45.
451. Ludwig MZ. Functional evolution of noncoding DNA. Current Opinion in Genetics & Development. 2002;12(6):634-9.
452. Arnone MI, Davidson EH. The hardwiring of development: organization and function of genomic regulatory systems. Development. 1997;124(10):1851-64.
453. Decker EL, Nehmann N, Kampen E, Eibel H, Zipfel PF, Skerka C. Early growth response proteins (EGR) and nuclear factors of activated T cells (NFAT) form heterodimers and regulate proinflammatory cytokine gene expression. Nucleic acids research. 2003;31(3):911-21.

454. Bresnick EH, Katsumura KR, Lee H-Y, Johnson KD, Perkins AS. Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. Nucleic acids research. 2012;40 (13):5819-31.
455. Ko LJ, Engel JD. DNA-binding specificities of the GATA transcription factor family. Molecular and cellular biology. 1993;13(7):4011-22.
456. Bulger M, Groudine M. Looping versus linking: toward a model for long-distance gene activation. Genes & development. 1999;13(19):2465-77.
457. Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. CTCF Tethers an Insulator to Subnuclear Sites, Suggesting Shared Insulator Mechanisms across Species. Molecular cell. 2004;13(2):291-8.
458. Liu Y, Bondarenko V, Ninfa A, Studitsky VM. DNA supercoiling allows enhancer action over a large distance. Proceedings of the National Academy of Sciences. 2001;98(26):14883-8.
459. Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. Nature protocols. 2007;2(8):1849-61.
460. Guan W, Cheng F, Wu H, Cao Q, Zhu X, Fan Y, et al. GATA binding protein 3 is correlated with leptin regulation of PPARγ1 in hepatic stellate cells. Journal of Cellular and Molecular Medicine. 2016:n/a-n/a.
461. Bonder MJ, Kasela S, Kals M, Tamm R, Lokk K, Barragan I, et al. Genetic and epigenetic regulation of gene expression in fetal and adult human livers. BMC genomics. 2014;15(1):860.
462. Lang G, Gombert WM, Gould HJ. A transcriptional regulatory element in the coding sequence of the human Bcl-2 gene. Immunology. 2005;114(1):25-36.
463. Pierce RA, Moore CH, Arikan MC. Positive transcriptional regulatory element located within exon 1 of elastin gene. American Journal of Physiology - Lung Cellular and Molecular Physiology. 2006;291(3):L391-L9.
464. Turley SD, Spady DK, Dietschy JM. Role of liver in the synthesis of cholesterol and the clearance of low density lipoproteins in the cynomolgus monkey. Journal of lipid research. 1995;36(1):67-79.
465. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics. 2014;46(3):310.
466. Xie D, Boyle AP, Wu L, Zhai J, Kawli T, Snyder M. Dynamic Trans-Acting Factor Co-localization in Human Cells. Cell. 2013;155(3):713-24.
467. Ambros V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. Cell. 2003;113(6):673-6.
468. Gamazon ER, Ziliak D, Im HK, LaCroix B, Park DS, Cox NJ, et al. Genetic architecture of microRNA expression: implications for the transcriptome and complex traits. The American Journal of Human Genetics. 2012;90(6):1046-63.
469. Min H, Yoon S. Got target?: computational methods for microRNA target prediction and their extension. Experimental & Molecular Medicine. 2010;42(4):233-44.
470. Erhard F, Haas J, Lieber D, Malterer G, Jaskiewicz L, Zavolan M, et al. Widespread context dependency of microRNA-mediated regulation. Genome research. 2014;24(6):906-19.
471. Hollås H, Aukrust I, Grimmer S, Strand E, Flatmark T, Vedeler A. Annexin A2 recognises a specific region in the 3′-UTR of its cognate messenger RNA. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2006;1763(11):1325-34.
472. Anni V, Hanne H, Ann Kari G, Aase MR. Multiple Roles of Annexin A2 in Post-Transcriptional Regulation of Gene Expressio. Current Protein & Peptide Science. 2012;13(4):401-12.

473. Zeng J, Xiong Y, Li G, Liu M, He T, Tang Y, et al. MiR-21 is overexpressed in response to high glucose and protects endothelial cells from apoptosis. Experimental and clinical endocrinology & diabetes: official journal, German Society of Endocrinology [and] German Diabetes Association. 2013;121(7):425-30.
474. Chen L, Li Y, Cui J, Ning J, Wang G, Qian G, et al. MiR-206 Controls the Phenotypic Modulation of Pulmonary Arterial Smooth Muscle Cells Induced by Serum from Rats with Hepatopulmonary Syndrome by Regulating the Target Gene, Annexin A2. Cellular Physiology and Biochemistry. 2014;34(5):1768-79.
475. Keklikoglou I, Hosaka K, Bender C, Bott A, Koerner C, Mitra D, et al. MicroRNA-206 functions as a pleiotropic modulator of cell proliferation, invasion and lymphangiogenesis in pancreatic adenocarcinoma by targeting ANXA2 and KRAS genes. Oncogene. 2015;34(37):4867-78.
476. Zhang J, Cheng J, Zeng Z, Wang Y, Li X, Xie Q, et al. Comprehensive profiling of novel microRNA-9 targets and a tumor suppressor role of microRNA-9 via targeting IGF2BP1 in hepatocellular carcinoma: Oncotarget; 2015.
477. Bronisz A, Wang Y, Nowicki MO, Peruzzi P, Ansari KI, Ogawa D, et al. Extracellular Vesicles Modulate the Glioblastoma Microenvironment via a Tumor Suppression Signaling Network Directed by miR-1. Cancer Research. 2014;74(3):738-50.
478. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic acids research. 2008;36(suppl 1):D149-D53.
479. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Research. 2009;19(1):92-105.
480. Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature. 2008;455(7209):58-63.
481. Lopez-Ramirez MA, Wu D, Pryce G, Simpson JE, Reijerkerk A, King-Robson J, et al. MicroRNA-155 negatively affects blood–brain barrier function during neuroinflammation. The FASEB Journal. 2014;28(6):2551-65.
482. Miller AM, Gilchrist DS, Nijjar J, Araldi E, Ramirez CM, Lavery CA, et al. MiR-155 has a protective role in the development of non-alcoholic hepatosteatosis in mice. 2013.
483. Ekimler S, Sahin K. Computational methods for microRNA target prediction. Genes. 2014;5(3):671-83.
484. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biology. 2010;11(8):R90-R.
485. Venkataraman S, Birks DK, Balakrishnan I, Alimova I, Harris PS, Patel PR, et al. MicroRNA 218 Acts as a Tumor Suppressor by Targeting Multiple Cancer Phenotype-associated Genes in Medulloblastoma. Journal of Biological Chemistry. 2013;288(3):1918-28.
486. Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. Cell. 2013;153(3):654-65.
487. Fujiwara T, Katsuda T, Hagiwara K, Kosaka N, Yoshioka Y, Takahashi R-U, et al. Clinical Relevance and Therapeutic Significance of MicroRNA-133a Expression Profiles and Functions in Malignant Osteosarcoma-Initiating Cells. STEM CELLS. 2014;32(4):959-73.
488. Fichtlscherer S, De Rosa S, Fox H, Schwietz T, Fischer A, Liebetrau C, et al. Circulating microRNAs in patients with coronary artery disease. Circulation research. 2010;107(5):677-84.
489. Wagner J, Riwanto M, Besler C, Knau A, Fichtlscherer S, Röxe T, et al. Characterization of Levels and Cellular Transfer of Circulating Lipoprotein-Bound MicroRNAs. Arteriosclerosis, thrombosis, and vascular biology. 2013;33(6):1392-400.

490. Chen T, Yan H, Li Z, Jing T, Zhu W, Ge J, et al. MicroRNA-155 regulates lipid uptake, adhesion/chemokine marker secretion and SCG2 expression in oxLDL-stimulated dendritic cells/macrophages. International journal of cardiology. 2011;147(3):446-7.
491. Nazari-Jahantigh M, Wei Y, Noels H, Akhtar S, Zhou Z, Koenen RR, et al. MicroRNA-155 promotes atherosclerosis by repressing Bcl6 in macrophages. The Journal of Clinical Investigation. 2012;122(11):4190-202.
492. Tian F-J, An L-N, Wang G-K, Zhu J-Q, Li Q, Zhang Y-Y, et al. Elevated microRNA-155 promotes foam cell formation by targeting HBP1 in atherogenesis. Cardiovascular research. 2014;103(1):100-10.
493. Kishore R, Verma SK, Mackie AR, Vaughan EE, Abramova TV, Aiko I, et al. Bone marrow progenitor cell therapy-mediated paracrine regulation of cardiac miRNA-155 modulates fibrotic response in diabetic hearts. PloS one. 2013;8(4):e60161.
494. Marais AD, Kim JB, Wasserman SM, Lambert G. PCSK9 inhibition in LDL cholesterol reduction: Genetics and therapeutic implications of very low plasma lipoprotein levels. Pharmacology & Therapeutics. 2015;145:58-66.
495. Dullaart RPF. PCSK9 Inhibition to Reduce Cardiovascular Events. New England Journal of Medicine. 2017;0(0).
496. Humphries SE, Whittall RA, Hubbart CS, Maplebeck S, Cooper JA, Soutar AK, et al. Genetic causes of familial hypercholesterolaemia in patients in the UK: relation to plasma lipid levels and coronary heart disease risk. Journal of Medical Genetics. 2006;43(12):943-9.
497. He Q-C, Hu Y-Y, Zhang Q-P, Tan L-L, Liu Y-H, Liu T, et al. A meta-analysis of three identified single nucleotide polymorphisms at 1p13.3 and 1q41 and their associations with lipid levels and coronary artery disease. The Kaohsiung Journal of Medical Sciences. 2017;33(1):1-10.
498. Welsh MJ, Smith AE. Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. Cell. 1993;73(7):1251-4.
499. The ENCODE Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.
500. Hsu Patrick D, Lander Eric S, Zhang F. Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell. 2014;157(6):1262-78.
501. Grundy SM, Cleeman JI, Merz CNB, Brewer HB, Clark LT, Hunninghake DB, et al. Implications of Recent Clinical Trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. Circulation. 2004;110(2):227.
502. Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. The Lancet. 2016;388(10059):2532-61.
503. Nordestgaard BG, Chapman MJ, Humphries SE, Ginsberg HN, Masana L, Descamps OS, et al. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease : Consensus Statement of the European Atherosclerosis Society. European Heart Journal. 2013;34(45):3478-90.
504. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, et al. Effect of Long-Term Exposure to Lower Low-Density Lipoprotein Cholesterol Beginning Early in Life on the Risk of Coronary Heart Disease. Journal of the American College of Cardiology. 2012;60(25):2631.
505. Wald DS, Bestwick JP, Morris JK, Whyte K, Jenkins L, Wald NJ. Child–Parent Familial Hypercholesterolemia Screening in Primary Care. New England Journal of Medicine. 2016;375(17):1628-37.
506. Futema M, Cooper JA, Charakida M, Boustred C, Sattar N, Deanfield J, et al. Screening for familial hypercholesterolaemia in childhood: Avon Longitudinal Study of Parents and Children (ALSPAC). Atherosclerosis. 2017;260:47-55.

Published paper from this thesis

1. PLOS ONE, 2016

PLOS | ONE

RESEARCH ARTICLE

Identification of the Functional Variant(s) that Explain the Low-Density Lipoprotein Receptor (*LDLR*) GWAS SNP rs6511720 Association with Lower LDL-C and Risk of CHD

Roaa Hani Fairoozy[1☯*], Jon White[2☯], Jutta Palmen[1], Anastasia Z. Kalea[1], Steve E. Humphries[1☯]

1 Centre for Cardiovascular Genetics, BHF Laboratories, Institute of Cardiovascular Science, University College London, London, United Kingdom, 2 University College London Genetics Institute, Department of Genetics, Environment and Evolution, London, United Kingdom

☯ These authors contributed equally to this work.
* roaa.fairoozy.12@ucl.ac.uk

OPEN ACCESS

Citation: Fairoozy RH, White J, Palmen J, Kalea AZ, Humphries SE (2016) Identification of the Functional Variant(s) that Explain the Low-Density Lipoprotein Receptor (*LDLR*) GWAS SNP rs6511720 Association with Lower LDL-C and Risk of CHD. PLoS ONE 11(12): e0167676. doi:10.1371/journal.pone.0167676

Editor: Nanette H Bishopric, University of Miami School of Medicine, UNITED STATES

Received: March 7, 2016

Accepted: November 20, 2016

Published: December 14, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: RHF is funded by King Abdullah Medical city in Holly city (KAMC), the Ministry of Health, Saudi Arabia. AZK is funded by a National Institute for Health Research, University College London Hospitals, Biomedical Research Centre Cardiometabolic Programme (BRC105CMSH/5982). SEH is a British Heart Foundation (BHF) Professor and he and JP are funded by BHF grant

Abstract

Background

The Low-Density Lipoprotein Receptor (*LDLR*) SNP rs6511720 (G>T), located in intron-1 of the gene, has been identified in genome-wide association studies (GWAS) as being associated with lower plasma levels of LDL-C and a lower risk of coronary heart disease (CHD). Whether or not rs6511720 is itself functional or a marker for a functional variant elsewhere in the gene is not known.

Methods

The association of *LDLR* SNP rs6511720 with incidence of CHD and levels of LDL-C was determined by reference to CARDIoGRAM, C4D and Global lipids genetics consortium (GLGC) data. SNP annotation databases were used to identify possible SNP function and prioritization. Luciferase reporter assays in the liver cell line Huh7 were used to measure the effect of variant genotype on gene expression. Electrophoretic Mobility Shift Assays (EMSAs) were used to identify the Transcription Factors (TFs) involved in gene expression regulation.

Results

The phenotype-genotype analysis showed that the rs6511720 minor allele is associated with lower level of LDL-C [beta = -0.2209, $p = 3.85 \times 10^{-262}$], and lower risk of CHD [log (OR) = 0.1155, $p = 1.04 \times 10^{-7}$]. Rs6511720 is in complete linkage. Rs6511720 is in complete linkage disequilibrium (LD) with three intron-1 SNPs (rs141787760, rs60173709, rs57217136). Luciferase reporter assays in Huh7 cells showed that the rare alleles of both rs6511720 and rs57217136 caused a significant increase in *LDLR* expression compared to the common alleles (+29% and +24%, respectively). Multiplex Competitor-EMSAs (MC-

2. Atherosclerosis, 2017

Contents lists available at ScienceDirect

Atherosclerosis

journal homepage: www.elsevier.com/locate/atherosclerosis

Identifying low density lipoprotein cholesterol associated variants in the *Annexin A2* (*ANXA2*) gene

Roaa Hani Fairoozy [a,*], Jackie Cooper [a], Jon White [b], Claudia Giambartolomei [c], Lasse Folkersen [d,e], S. Goya Wannamethee [f], Barbara J. Jefferis [f], Peter Whincup [g], Yoav Ben-Shlomo [h], Meena Kumari [i,j], Mika Kivimaki [j], Andrew Wong [k], Rebecca Hardy [k], Diana Kuh [k], Tom R. Gaunt [l], J.P. Casas [m], Stela McLachlan [n], Jackie F. Price [n], Aroon Hingorani [o], Anders Franco-Cereceda [p], Thomas Grewal [q], Anastasia Z. Kalea [a], Steve E. Humphries [a], on behalf of the UCLEB consortium

[a] Cardiovascular Genetics, BHF Laboratories, Institute of Cardiovascular Science, University College London, London, United Kingdom
[b] UCL Genetics Institute, Department of Genetics, Environment and Evolution, University College London, London, United Kingdom
[c] Division of Psychiatric Genomics, Department of Psychiatry and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA
[d] The Karolinska Institute, Stockholm, Sweden
[e] Department of Bioinformatics, Technical University of Denmark, Lyngby, Denmark
[f] UCL Department of Primary Care & Population Health, UCL Institute of Epidemiology, University College London, London, United Kingdom
[g] Population Health Research Institute, St George's University of London, Cranmer Terrace, London, United Kingdom
[h] School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom
[i] Institute for Social and Economic Research, University of Essex, Colchester, United Kingdom
[j] Department of Epidemiology & Public Health, UCL Institute of Epidemiology & Health Care, University College London, London, United Kingdom
[k] MRC Unit for Lifelong Health and Ageing, London, United Kingdom
[l] MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, United Kingdom
[m] Farr Institute of Health Informatics, University College London, London, United Kingdom
[n] Centre for Population Health Sciences, The University of Edinburgh, Edinburgh, United Kingdom
[o] Genetic Epidemiology Group, Institute of Cardiovascular Science, University College London, London, United Kingdom
[p] Cardiothoracic Surgery Unit, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
[q] Faculty of Pharmacy, University of Sydney, Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 27 January 2017
Received in revised form
24 March 2017
Accepted 12 April 2017
Available online 13 April 2017

Keywords:
Annexin A2
Proprotein convertase subtilisin/kexin type-9
Low-density lipoprotein cholesterol-receptor
Low-density lipoprotein cholesterol
Single nucleotide polymorphism
Coronary heart disease

ABSTRACT

Background and aims: Annexin-A2 (AnxA2) is an endogenous inhibitor of proprotein convertase subtilisin/kexin type-9 (PCSK9). The repeat-one (R1) domain of AnxA2 binds to PCSK9, blocking its ability to promote degradation of low-density lipoprotein cholesterol-receptors (LDL-R) and thereby regulate low-density lipoprotein cholesterol (LDL-C) levels. Here we identify variants in *ANXA2* influencing LDL-C levels and we determine the molecular mechanisms of their effects.

Results: The *ANXA2* single nucleotide polymorphism (SNP) genotype-phenotype association was examined using the Second-Northwick-Park Heart Study (NPHSII) (n~2700) and the UCL-LSHTM-Edinburgh-Bristol (UCLEB) consortium (n~14,600). The *ANXA2*-R1 domain coding-SNP rs17845226 (V98L) associated with LDL-C, homozygotes for the minor allele having ≈ 18.8% higher levels of LDL-C ($p = 0.004$), and higher risk of coronary heart disease (CHD) ($p = 0.04$). The SNP is in modest linkage disequilibrium ($r^2 > 0.5$) with two intergenic SNPs, rs17191344 and rs11633032. Both SNPs showed allele-specific protein binding, and the minor alleles caused significant reduction in reporter gene expression (≈ 18%, $p < 0.001$). In the expression quantitative trait loci (eQTL) study, minor allele homozygotes have significantly lower levels of *ANXA2*-mRNA expression ($p = 1.36 \times 10^{-05}$).

Abbreviations: AnxA2, annexin A2; PCSK9, proprotein convertase subtilisin/kexin type-9; LDLR, low-density lipoprotein cholesterol-receptor; LDL-C, low-density lipoprotein cholesterol; NPHSII, Second-Northwick-Park Heart Study; UCLEB, UCL-LSHTM-Edinburgh-Bristol consortium; CHRD, cysteine-histidine-rich domain of PCSK9; FAIRE, formaldehyde assisted isolation of regulatory elements; CTCF, CTC-binding factor.

* Corresponding author. Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, Rayne Building Rm 318, 5 University St, London WC1E 6JF, United Kingdom.
E-mail address: roaa.fairoozy.12@ucl.ac.uk (R.H. Fairoozy).

http://dx.doi.org/10.1016/j.atherosclerosis.2017.04.010

3. Scientific Reports, 2017

SCIENTIFIC REPORTS

OPEN

Received: 18 July 2017
Accepted: 15 November 2017
Published online: 06 December 2017

The Genetic Spectrum of Familial Hypercholesterolemia (FH) in the Iranian Population

R. H. Fairoozy[1,2], M. Futema[3], R. Vakili[4,5], M. R. Abbaszadegan[5], S. Hosseini[5], M. Aminzadeh[6], H. Zaeri[7], M. Mobini[8], S. E. Humphries[1] & A. Sahebkar[9,10]

Familial hypercholesterolemia (FH) is an autosomal dominant disorder associated with premature cardiovascular disease (CVD). Mutations in the *LDLR*, *APOB*, and *PCSK9* genes are known to cause FH. In this study, we analysed the genetic spectrum of the disease in subjects from the Iranian population with a clinical diagnosis of FH. Samples were collected from 16 children and family members from five different cities of Iran. Probands were screened for mutations in the *LDLR*, *APOB*, and *PCSK9* genes using next generation sequencing, with results confirmed by Sanger sequencing. The likely pathology of identified variants was examined using *in silico* tools. Of the probands, 14 had a clinical diagnosis of homozygous FH and two of heterozygous FH. No mutations were found in either *APOB* or *PCSK9*, but nine probands were homozygous for seven different *LDLR* mutations, with p.(Trp577Arg) occurring in three and p.Val806Glyfs*11 occurring in two patients. Two mutations were novel: p.(Leu479Gln) and p.(Glu668*). Seven probands with a clinical diagnosis of FH were mutation negative. This pilot study, integrating clinical and molecular-based techniques, begins to elucidate the FH heterogeneity and the mutation spectrum in the Iranian population. Such information is important for future disease management and cost savings.

Familial hypercholesterolemia (FH) is an autosomal dominant disorder resulting in elevated plasma low-density lipoprotein cholesterol (LDL-C). Heterozygous FH (HeFH) is common, with an estimated prevalence of at least 1/500[1] in most European populations, but being higher in some populations such as the Danes[2], the Afrikaners in South Africa[3], French Canadians[4] and the Dutch[5], with recent estimates in the UK suggesting around 1/270[6,7]. Plasma LDL-C in HeFH is two-to-three-fold higher than normal. Individuals with this form of FH are more likely to develop premature coronary heart disease (CHD) in the second or third decade of their lives[1]. The second form is homozygous FH (HoFH), which has an estimated prevalence of 1 in 1 million in most populations, however, some populations have reported higher prevalence such as 1:300,000 in the Netherlands[5]. Individuals with HoFH have six-to-eight-fold higher levels of LDL-C plasma than normal and develop CHD in the early stages of their lives, often dying before the age of 20[1]. FH is caused by a mutation in one of three genes: the low-density lipoprotein cholesterol receptor (*LDLR*), Apolipoprotein B gene (*APOB*), or a gain-of-function mutation in the gene for proprotein convertase subtilisin/kexin type-9 (*PCSK9*). Autosomal recessive hypercholesterolemia (ARH), a rare inheritance of FH, has been reported[8] and occurs when individual inherits two pathogenic variants in the low-density lipoprotein adaptor protein 1 (*LDLAP1*) gene. In the UK, the vast majority (93%) of identified mutations are found in the *LDLR* gene, a further 5% in *APOB*, and about 2% in the *PCSK9* gene[9,10].

[1]Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, London, United Kingdom. [2]Molecular Diagnostic Unit, Clinical Laboratory Department, King Abdullah Medical city in Makkah, Makkah, Saudi Arabia. [3]Centre for Cardiology in the Young, Institute of Cardiovascular Science, University College London, London, United Kingdom. [4]Department of Pediatric Endocrinology and Metabolism, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. [5]Medical Genetics Research Center, Mashhad University of Medical Sciences, Mashhad, Iran. [6]Diabetes Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran. [7]Neonatal and Children Health Research Centre, Golestan University of Medical Sciences, Gorgan, Iran. [8]School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. [9]Biotechnology Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran. [10]School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran. Correspondence and requests for materials should be addressed to S.E.H. (email: steve.humphries@ucl.ac.uk)