Hind Elias Muallem

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Pathology and Laboratory Medicine.

Chapel Hill
2008

Approved by:

Advisor: Nobuyo Maeda

Reader: Kari North

Reader: Kirk wilhemsen

Reader: Bernard Weissman

Reader: Margaret-Gully

ABSTRACT

Hind Muallem: Association of Genetic Variations in the 3'UTR of the Human Low Density Lipoprotein Receptor with the Lipid Parameters in the Atherosclerosis Risk in Communities (ARIC) Study

(Under the direction of Nobuyo Maeda)

The low-density lipoprotein receptor (LDLR) plays a pivotal role in cholesterol homeostasis. However, the role of genetic variations in the 3' UTR of the *LDLR* in relation to plasma cholesterol has been largely understudied. Six SNPs, G44243A, G44332A, C44506G, G44695A, C44857T and A44964G, within the 5' region of the 3'UTR fall into three common haplotypes, *GGCGCA*, A*GCACG*, and *GGCGTA*, occurring at frequencies of 0.45, 0.31 and 0.17 respectively in Caucasians (n=29) and 0.13, 0.13 and 0.38 respectively in African Americans (n=32), with three other haplotypes occurring at lesser frequencies. Genotyping of two "haplotype tagging" SNPs, C44857T and A44964G, in the Atherosclerosis Risk in Community (ARIC) study population showed that in Caucasians, but not in African Americans, the inferred *TA* haplotype had a significant LDL-cholesterol lowering effect. The adjusted LDL-cholesterol levels in the *TA/TA* diplotypes were lower by 6.10 mg/dl in men ($P<0.001$) and by 4.63 mg/dl in women ($P<0.01$) than in individuals with other diplotypes. Caucasian men homozygous for *CA,* in contrast, showed significantly

higher LDL-cholesterol (P<0.04), lower HDL-cholesterol (P<0.02) and higher LDL/HDL ratios (P<0.001). To identify if these haplotypes are causative, we studied the expression of a reporter gene in a tissue culture based system carrying a 3'UTR that includes the 1kb nucleotide sequences corresponding to the *AGCACG*, *GGCGTA* and *GGCGCA* haplotypes. The three haplotypes of the 3'UTR sequences had no difference in their effect on the expression. Therefore, the association in Caucasians with plasma lipid profiles indicative of higher cardiovascular risk is not caused by polymorphisms in the 3'UTR of the LDLR gene, but with other polymorphisms that are in linkage disequilibrium with them.

ACKNOWLEDGEMENTS

My special and sincere thanks go to Nobuyo Maeda and Olivier Smithies. They are an inspiration and light that shines in the field of research and science. I consider myself very lucky to have had the chance to spend five years in their lab and under their guidance.

I would like to thank my committee members for their guidance, support, and advice in the progress of my project: Kari North, Kirk Wilhemsen, Bernard Weissman, and Margaret-Gully.

Furthermore, I owe much thanks and gratitude to my collaborators: Gerardo Heiss, Eric Boerwinkle,' Megan Grove,' Mary Wojczynski, and Xia Li.'

I am highly indebted to all Smities Maeda's Laboratory members for each and every one of them helped me in a different way: Lance Johnoson, Kumar Pandya, Ray Givens, Ray Fox, Avani Pendse, Jose Arbones-Mainer, Micheal Altenberg, Jennifer Wilder, John Hagaman, Jenny Langenbach. Also special thanks go to Masao Kakoki, Hyung-Suk kim, Sylvia Hiller, John Cowhig, Xianwen Yi, Seigo Hatada, Tomoko Hatada, and Sarah Karkanawi.

Last but not least, I am immensely thankful to Bill Coleman, our chairman Dr. Jennet Charles, Frank Church, and Nadia Maalouf, for their support to all the graduate students in the department and Dorothy Poteat for her administrative assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Familial Hypercholesterolemia	FH
Familial Defective Apolipoprotein B-100	FDB
Autosomal Recessive Hypercholesterolemia	ARH
Low density lipoprotein cholesterol	LDL-C
Coronary heart disease	CHD
Low density lipoprotein	LDL
Low density lipoprotein gene	*LDLR*
Apolipoprotein	APO
Proprotein convertase subtilisin–like kexin type 9	PCSK9
Steroid regulatory element	SRE
Steroid regulatory element binding protein	SREBP
Steroid regulatory element binding protein cleavage activating protein	SCAP
Total Cholesterol	TC
Green fluorescent protein	GFP
High density lipoprotein	HDL
Triglyceride	TG
Very low-density lipoprotein	VLDL
Low density lipoprotein	LDL
Intermediate density lipoprotein	IDL
Sterol sensing domain	SSD

Site -1 protease	S1P
Site -2 protease	S2P
Atherosclerosis Risk in Communities Study	ARIC
Restriction fragment length polymorphisms	RFLP
Single nucleotide polymorphism	SNP
Linkage disequilibrium	LD

Chapter I

Introduction

Hypercholesterolemia, Atherosclerosis, and Coronary Heart Diseases

Hypercholesterolemia is a major factor in the development of atherosclerosis. Atherosclerosis leading to coronary heart disease (CHD) is the cause of more than 25% of the deaths in the United States. Every year around 1.2 million Americans are diagnosed with CHD. Hypercholesterolemia is clearly a complex and multigenic condition. Four main monogenic diseases have been identified.

1 - Familial Hypercholesterolemia (FH)

2 - Familial Defective Apolipoprotein B-100 (FDB)

3 - Autosomal Recessive Hypercholesterolemia (ARH)

4 - Autosomal Dominant Hypercholesterolemia due to mutations in *PCSK9* gene

Familial Hypercholesterolemia (FH)

Familial hypercholesterolemia (FH) is an autosomal dominant disease that is characterized by a hallmark of elevated low density lipoprotein cholesterol (LDL-C) and premature CHD. FH is caused by mutations in the low density lipoprotein receptor gene (*LDLR*) (Hobbs et al. 1992; Hobbs et al. 1988; Levy et al. 1997; Pereira et al. 1995). FH heterozygote occurs 1 in 500 individuals. Plasma LDL-C in these patients is increased 2 fold (300-500 mg/dl). FH homozygote is rare, occuring 1 in million. Plasma LDL-C in FH homozygotes is increased 3 to 6 fold (600-1200 mg/dl).

LDLR is a cell surface receptor and together with its ligands ApoE / ApoB, plays a pivotal role in cholesterol homeostasis. LDLR is synthesized as a glycoprotein precursor of 120 KD, which is changed to a mature glycoprotein of 160 KD due to the addition of a 40 KD protein (Tolleshaug et al. 1982). It is located on chromosome 19p13.1-13.3. It encodes

839 amino acids, and has multiple copies of Alu sequences of repetitive DNA (Yamamoto et al. 1984). The LDLR is a member of the supergene families, since 13 out of the 18 exons of the gene encode proteins that are homologous to other genes, such as C9 component of complement, epidermal growth factor precursor, and blood clotting system factor XI and X (Sudhof et al. 1985).

The LDLR is made up of five domains. The first domain is the ligand binding region for ApoE and ApoB-100, which consists of seven repeats rich in cysteine. The second domain is the epidermal growth factor-precursor homology domain which is required for the dissociation of the ligand in the acidic medium of the endosome. The third domain is the O-linked glycan region. The fourth domain is the transmembrane domain that anchor the LDLR in the lipid bilayer, and the fifth domain is the cytosolic region which is involved in endocytosis and internalization of the LDL-LDLR complex (Gent and Braakman 2004; Sudhof et al. 1985) (Figure 1.1).

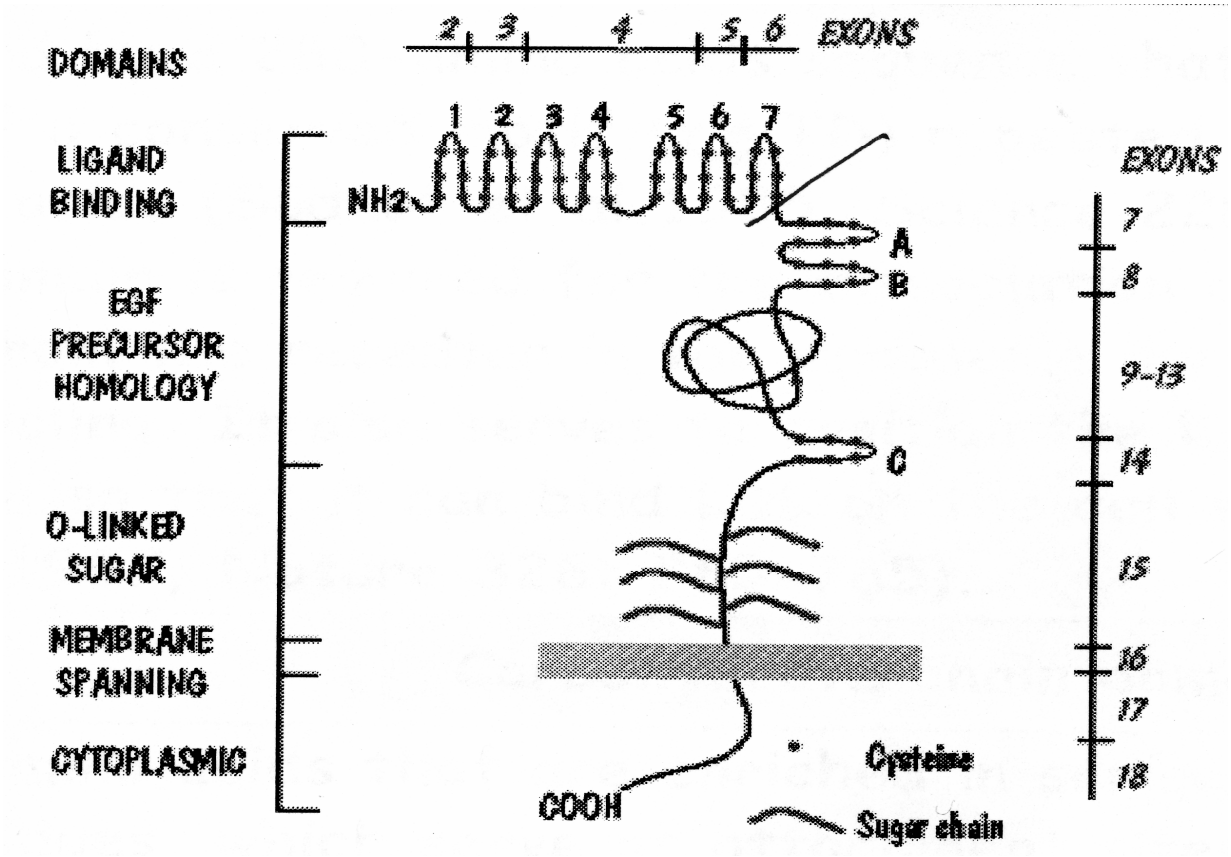

Figure 1.1. Structure of the human low density lipoprotein receptor. On the right are the exons encoding the protein and on the left are the domains of the mature protein.

Mutations in the LDLR lead to an elevated level of total cholesterol and LDL-C (Hobbs et al. 1992). There are two classified groups of LDLR mutations: 1) Receptor negative or null mutations, and 2) receptor defective mutations. Within these groups there are five different classes of LDLR mutations (Austin et al. 2004; Gent and Braakman 2004).

Class 1: This is called a receptor negative mutation. The LDLR synthesis is disrupted due to different mutations leading to a complete absence of functional protein.

Class 2: Normally the LDLR is synthesized as 120 KD in the endoplasmic reticulum (ER) and then migrates to the Golgi where the O and N sugars are added to change it to 160 KD protein. In a class 2 mutation the protein does not leave the ER and is not transported to the Golgi. As a result it is degraded before it is transported to the cell surface.

Class 3: The LDLR is synthesized and transported to the cell surface, but does not properly bind the low density lipoprotein particle (LDL). This is caused by deletions or mutations in the ligand binding domains.

Class 4: The surface receptor is not clustered in coated pits, and as a result can not internalize the bound LDL particle into the cell. This is caused by mutations in the cytoplasmic domain.

Class 5: Mutations occur in the epidermal growth factor homology domain which hinders the dissociation of the receptor and ligand in the endosome in an acidic environment. This decreases the recycling of the receptor to the cell surface.

Class I is referred to as a receptor negative mutation while the other 4 classes are referred to as receptor defective mutations.

All of the above class mutations cause an impaired function of the LDLR which leads to FH and to abnormal clearance of cholesterol and LDL-C from the circulation, and hence an increased risk of atherosclerosis.

Familial Defective Apolipoprotein B-100

Apolipoprotein B-100 (ApoB-100) is synthesized in the liver; it is made up of 4536 amino acids, and weighs 550 KD. ApoB-100 plays an important role in LDL clearance. It is the only protein constituent of LDL and is the ligand that is recognized by the LDLR. Genetic disorders of ApoB-100 cause familial defective apolipoprotein B-100 (FDB), a disorder which leads to hypercholesterolemia and an increased risk of CHD (Innerarity et al. 1990). The first identified mutation in ApoB-100 was a single base substitution in exon 26 at position 3500. This single base substitution (G/A) CGG to CAG, leads to the amino acid substitution of glutamine for arginine (R3500Q) (Innerarity et al. 1990; Jacobsen et al. 2002; Soria et al. 1989). The frequency of the mutation is 1/500 in Caucasians. The mutation reduces the affinity of LDL to LDLR by approximately 95 %, resulting in hypercholesterolemia (Tybjaerg-Hansen et al. 1998).

Along with the role of ApoB-100 in mediating the clearance of LDL, it is also required for the synthesis, assembly and secretion of the hepatic triglyceride rich lipoprotein particle, very low density lipoprotein (VLDL), into the circulation. A full length ApoB-100 is needed for normal secretion of VLDL. Mutations in the coding region of the gene result in a premature termination codon (Burnett et al. 2003; Burnett et al. 2007; Linton et al. 1992). Most of these nonesense or frame shift mutations cause a truncated ApoB-100. Truncation could vary in size from B-2 to B-89. Short truncations of B-27/B29 are not

detectable in plasma because they are not secreted into the circulation (Lancellotti et al. 2004; Schonfeld 2003). Hence, these mutations are cause of familial hypobetalipoproteinemia (FHBL). FHBL is a genetic disorder that is characterized by low levels of plasma cholesterol, LDL-C and ApoB.

For example a nucleotide change (G/T) in the 5'splice donor site in intron 5 leads to improper splicing and results in the absence of plasma ApoB-100. Another nucleotide change (C/T) in exon 10 at residue 412 causes a nonsense mutation leading to a truncated ApoB-100 (Huang et al. 1991). A novel missense mutation (R463W) was discovered in a Lebanese Christian (Burnett et al. 2003). The homozygotes of this mutation have very low levels of ApoB-100 that is barely detectable in plasma.

Also, a new mutation, (R3480P), was discovered in 2005. Heterozygotes of this mutation have a lower binding affinity to LDLR yet, they have hypobetalipoproteinemia. This is mainly due to the reduced amount of VLDL that is converted to LDL (Benn et al. 2005).

Autosomal Recessive Hypercholesterolemia (ARH)

The *ARH* gene encodes a putative adaptor protein called ARH. Its main function is to link the LDLR to the clathrin protein of the coated pit for endocytosis. ARH is made up of 308 amino acids and four domains, an N-terminal domain, of unknown function, a phosphotyrosin binding domain (PTB) that binds the (FDNPVY) sequence in the LDLR cytoplasmic tail for internalization, a clathrin box sequence that binds with high affinity to clathrin, and a C- terminal domain that binds to the adaptor protein -2 (AP-2). Therefore, ARH mediates the interaction of the three proteins, LDLR, clathrin, and AP-2 to promote the clustering and endocytosis of the LDL-LDLR complex (Garuti et al. 2005; He et al.

2002). ARH is tissue specific since it is required for LDLR function in the liver but not in fibroblast.

Mutations in the *ARH* gene cause autosomal recessive hypercholesterolemia, (ARH) a disorder characterized by impaired internalization of LDLR in the liver. Clinically, it is similar to familial hypercholesterolemia which is caused by mutations in *LDLR* gene. ARH is a very rare disorder worldwide. Around 50 recognized affected ARH individuals are mainly of Sardinian or Middle Eastern origin (Arca et al. 2002). The two mutations that were found in the Sardinian population are a nonsense mutation at G65A, which led to a stop codon at W22X, and an insertion at 432insA which also led to a truncated non functional protein (Garcia et al. 2001). In affected Lebanese individuals the prevalent mutations are a nonsense mutation at C406T, resulting in a stop codon at Q136X, and another missense mutation at C605A resulting in a stop codon at P202H (Garcia et al. 2001; Soutar et al. 2003).

Autosomal Dominant Hypercholesterolemia due to mutations in *PCSK9*

In the past few years, a newly discovered serine protease called proprotein convertase subtilisin–like kexin type 9 (PCSK9) was shown to play an important role in LDL metabolism. PCSK9 is a glycoprotein that encodes 692 amino acids. It has N-terminal prodomain, a catalytic domain and a C-terminal domain (Benjannet et al. 2004; Naureckiene et al. 2003; Seidah et al. 2003). In 2003, Abifadel et al reported that some missense mutations in the gene cause autosomal dominant hypercholesterolemia (Abifadel et al. 2003). Since then, more research was done to elucidate the mechanism by which PCSK9 plays a role in lipoprotein metabolism. Some studies have shown that over expression of

PCSK9 in the liver of mice has no effect on the mRNA of LDLR, but, reduces LDLR by degrading it post transcriptionally (Fasano et al. 2007; Horton et al. 2007; Maxwell and Breslow 2004; Maxwell et al. 2005). The actual site of action of PCSK9 has not yet been identified. However, Helen Hobbs and her group showed that stable over expression of PCSK9 has no effect on either synthesis of LDLR or trafficking of LDLR out of the endoplasmic reticulum in HepG2 cells (Horton et al. 2007). Also, inhibition of the proteosome does not affect LDLR reduction by PCSK9, but disassembly of the Golgi complex did inhibit the degradation of LDLR (Horton et al. 2007). They also showed that the addition of purified PCSK9 to the medium in HepG2 cells decreased the number of LDLR on the cell surface, and that both proteins are co-immunoprecipitated. This suggests that the two proteins are physically attached and that PCSK9 functions extracellularly. In addition, the ARH that is required for the internalization of LDLR is also required for the degradation of LDLR by PCSK9. Hence, the absence of ARH inhibits the degradation of LDLR by PCSK9 (Lagace et al. 2006).

Based on the findings mentioned above, a possible mechanism in which PCSK9 causes LDLR degradation is through binding to LDLR on the cell surface in a catalytically inactive state. In the endosome, the pH is acidic and PCSK9 becomes active and degrades the LDLR. Another possible mechanism is that PCSK9 inhibits normal recycling of LDLR after internalization, and instead directs LDLR to the lysozome where it is degraded rather than recycled to the cell surface.

Having shown the importance of PCSK9 in lipoprotein metabolism, we would expect that mutations in the gene also affect plasma levels of LDL-C. Many researchers have shown that missense and nonsense mutations in the gene that cause loss of PCSK9

function are associated with hypocholesterolemia (Abifadel et al. 2003; Chen et al. 2005). Other missense mutations that cause gain of function are associated with hypercholesterolemia. Loss of function mutations are more common than gain of function mutations (Fasano et al. 2007; Kotowski et al. 2006; Zhao et al. 2006). Examples of gain of function mutations are S127R in the prodomain, F216L(Abifadel et al. 2003), D374Y (Leren 2004; Timms et al. 2004) N425S in the catalytic domain, and R496 in the C terminal domain. Some known loss of function mutations are R46L, ΔR97 (Zhao et al. 2006), G106R (Berge et al. 2006), Y142X (Cohen et al. 2005) in the prodomain, L253F in the catalytic domain, A443T (Kotowski et al. 2006) and C679X (Benjannet et al. 2006) in the C terminal domain. There are also other missense mutations and in-frame deletions that cause low levels of LDL-C (Berge et al. 2006; Cameron et al. 2006; Kotowski et al. 2006).

Hypocholesterolemia that is caused by loss of function mutations is not associated with diseases, as it is benign, unlike abetalipoproteinemia and homozygous hypobetalipoproteinemia, both of which are associated with hepatic steatosis, malnutrition and fat-soluble vitamine deficieny. Hence, PCSK9 is considered a good therapeutic target for lowering LDL-C and prevention of coronary heart disease.

Lipoprotein Metabolism

Lipoprotein particles are made up of apoproteins and lipids, and facilitate the transport of lipids in the aqueous medium of the plasma. They are spherical in shape with a hydrophobic core of cholesterol esters and triglycerides, surrounded by a polar layer of phospholipids and hydrophilic proteins. There are six classes of lipoproteins: chylomicrons, chylomicron remanents, very low density lipoprotein (VLDL), intermediate density lipoprotein (IDL), low density lipoprotein (LDL) and high density lipoprotein (HDL) (Figure 1.2) http://images.google.com.

Chylomicrons are assembled in the intestinal mucosa and they transport dietary cholesterol and triglycerides from the intestine to the liver and other tissues. They carry apoB-48 and apoE on their surface. Chylomicron remnants deliver dietary cholesterol, apoB-48 and apoE to the liver via the chylomicron remanent receptor. VLDL transport endogenous triglycerides and cholesterol to extra hepatic tissues, mainly muscle and adipose tissue, as a source of energy or for storage. They have apoE, apoC, and apoB-100. IDL are formed when triglycerides are removed from VLDL. They are either taken by the liver or are converted to LDL. LDL is the main carrier of cholesterol to the tissues in humans. It has only apoB-100. HDL transports cholesterol from the tissues back into the liver for excretion in bile. They have apoE, apoC, and apoA.

Cholesterol plays an important role in body functions. It is an important component of cell membranes as well as a precursor of steroid hormones, estrogen, androgen, mineralcorticoids and glucocorticoids. Estrogen and androgen are produced by the adrenal cortex and the gonads; they are responsible for the function of female and male secondary sex organs. Mineralcorticoids and glucocorticoids are both formed in the adrenal cortex.

Mineralcorticoids maintain water and salt balance, while glucocorticoids affect fat, gluconeogenesis, glycogen, and protein degradation.

Cholesterol is also a precursor of bile acids which are synthesized in the liver. They are stored in the gallbladder and then secreted into the small intestine to render dietary lipids soluble for absorption.

Since lipids are insoluble in plasma, they are transported via lipoprotein through three major pathways (Figure 1.2).

1.-..Exogenous pathway

2.-..Endogenous pathway

3.-..Reverse cholesterol pathway.

Exogenous pathway

Dietary lipids are insoluble in the aqueous environment of the small intestine. They become soluble through the action of bile acids, which are secreted from the liver. Once solubilized they become accessible to pancreatic lipase which hydrolyzes dietary triglyceride (TG) into free fatty acids (FFA) , and mono and diacyl glycerol (Dawson and Rudel 1999). These products diffuse into the intestinal epithelial cells where they are resynthesised as TG. The entrocytes package the cholesterol, TG, and phospholipids into chylomicrons which enter the circulation via the lymphatic system. Before entering the circulation, apolipoprotein B-48 (ApoB-48) apolipoprotein A (ApoA) A-I, II and IV are the predominant apolipoproteins on the chylomicrons. ApoB-48 is synthesized only by the small intestine and is generated by a stop codon at postion 2153 of the ApoB mRNA which results in 48 % of the full length of ApoB-100. In the blood stream they acquire apolipoprotein E (ApoE), Apolipoprotein C (ApoC) CI, CII and CIII from other

lipoproteins, mainly high density lipoproteins (HDL) (Tall 1990). In the capillaries of muscle and adipose tissue, lipoprotein lipase hydrolyzes the TG in chylomicrons into FFA and glycerol. The FFA is used as a source of energy for muscles and peripheral tissues, or is stored in adipose tissues in the form of TG. The chylomicron remnants, i.e. cholesterol and apoproteins, are taken up by the liver by the chylomicron remanent receptor called low density lipoprotein receptor –related protein (LRP) (Mahley and Ji 1999)

Endogenous pathway

While fasting, the liver becomes the main source of plasma lipoproteins. Excess dietary lipids are stored in the liver in the form of triglycerides. The lipids are packaged into a VLDL particle that has ApoB-100 and is secreted into the plasma through the space of Disse. Like chylomicrons, the lipoprotein lipase converts VLDL to intermediate density lipoprotein (IDL). The IDL are removed by the liver via the LDLR. As the IDL lose more TG, they are converted into smaller, denser, cholesterol- ester rich LDL particles. LDL is the main plasma carrier of cholesterol to tissues. They do not contain ApoE on their surface and can be removed solely by the ApoB -100 ligand. ApoB -100 has a lower affinity for the LDLR than ApoE, which leads to higher plasma half life of circulating LDL. The LDL is taken up by the tissue through LDLR by endocytosis. The maximum uptake is in the liver, which accounts for 75%, and in the adrenals and adipose tissues (Goldstein and Brown 1984; Rudel et al. 1986). Inside the peripheral cell the apoproteins are degraded and cholesterol esters are hydrolyzed to give free cholesterol which is incorporated into the plasma membrane and used for the synthesis of steroids. Excess cholesterol in the liver is excreted in the form of bile acids. When LDL circulates for a long time in circulation, it is

oxidized and is taken up the macrophage scavenger receptor, which can lead to the formation of foam cells in the vessel wall (Quinn et al. 1987).

Reverse Cholesterol pathway

HDL is another class of lipoproteins. HDL is synthesized in the liver and small intestine, has ApoCI, ApoCII and ApoA1 on its surface, and plays an important role in reverse cholesterol transport. It removes extra cholesterol from the tissues and returns them to the liver by three different pathways. One pathway is the transfer of cholesterol to LDL and VLDL via the cholesterol ester transfer protein (CETP). The second pathway is through the accumulation of ApoE on the HDL particle and then clearance by direct binding to the LDLR through ApoE. Thirdly, the cholesterol is taken up directly by the scavenger receptor (SR-B1) without the uptake of particles (Acton et al. 1996; Steinberg 1996).

Taken together we conclude that cholesterol carried by high non-HDL particles in circulation is associated with high risk for CHD while, cholesterol carried by HDL particles is associated with low risk of CHD.

Lipoprotein Metabolism

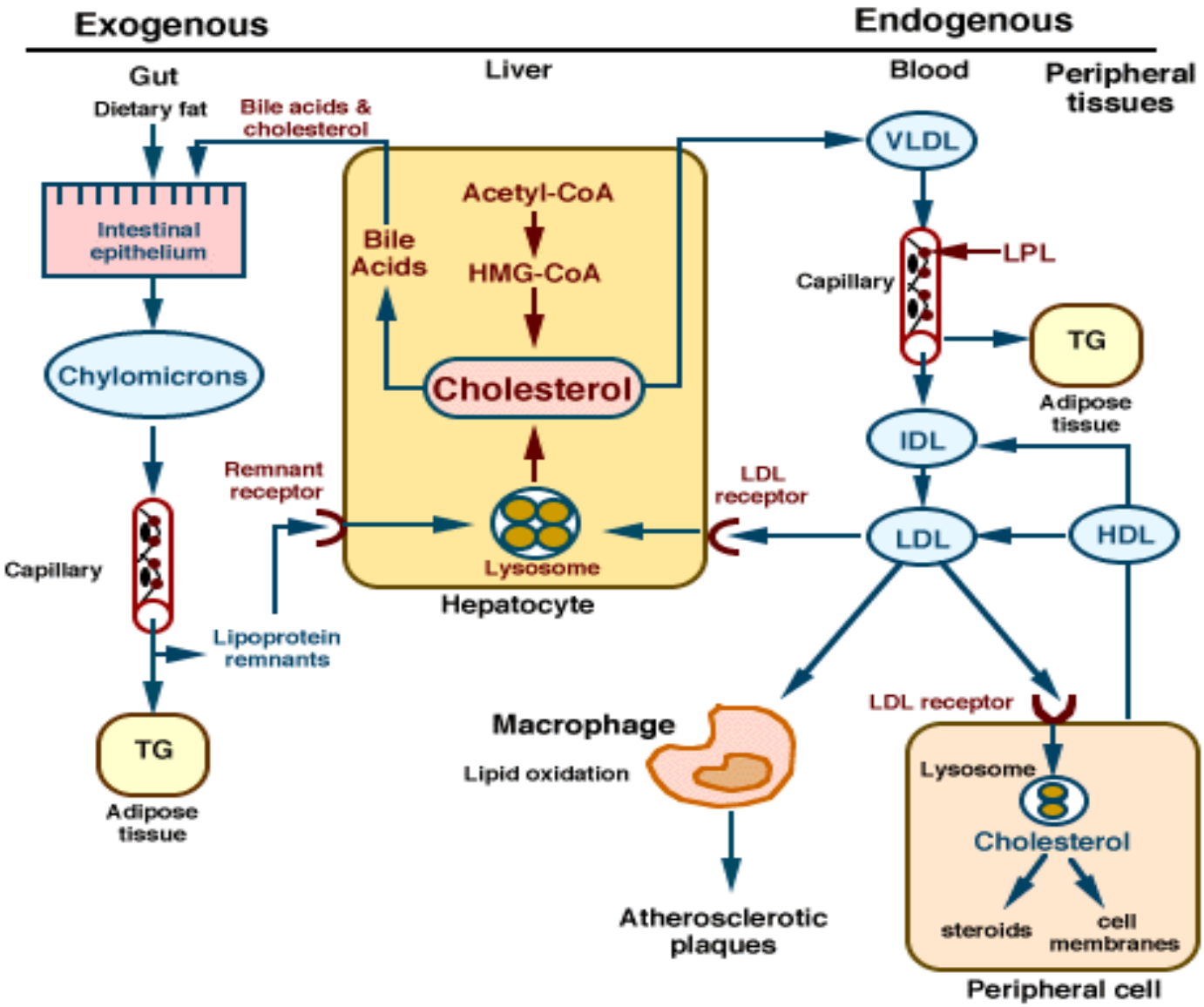

Figure 1.2. Human Lipid Metabolism. The lipid metabolism is divided into three pathways, the exogenous pathway (left) which transport dietary lipids, the endogenous pathway (right) which transports lipids from the liver to the circulation and the reverse cholesterol transport (lower right).

Regulation of the *LDLR* gene

Two regions are known to regulate transcription of the *LDLR* gene, the Promoter region and downstream of the gene (Figure 1.3).

Transcriptional elements in the promoter region lies within 177 base pair which consist of two TATA boxes and three GC rich imperfect repeats of16 bp in length each (Sudhof et al. 1987). Repeat 1 and 3 contain binding sites for SP1 transcription factor and they are responsible for the basal level of *LDLR* expression, repeat 3 contains a sterol regulatory element binding site (SRE-1) where sterol regulatory binding protein binds and enhances transcription (Briggs et al. 1993).

The transcription of the gene is regulated by the intracellular level of cholesterol (Smith et al. 1990). When the intracellular level of cholesterol is low, *LDLR* transcription is increased, and its amount on the cell surface is also increased, to actively bind and internalize LDL-C. On the other hand, when the cholesterol level within cells is high, transcription is diminished and the uptake of LDL-C is suppressed (Dawson et al. 1988). This regulation depends on sterol regulatory element binding protein (SREBP) which binds to (SRE-1) in repeat 3 (Briggs et al. 1993). Nuclear localization of SREBP is regulated by another protein called steroid regulatory element binding protein cleavage activating protein (SCAP). SCAP has a sterol sensing domain (SSD), and when intracellular sterol content is low, SCAP activates site -1 protease (SIP) which cleaves SREBP at one site. The cleaved product is then cleaved by site-2 protease, and the active N-terminal domain of SREBP migrates to the nucleus to start transcription (Hua et al. 1993; Wang et al. 1994). When intracellular cholesterol is high, SCAP inactivates SIP and hence inhibits the cleavage of the N-terminal domain (Dawson et al. 1988).

In addition to the enhancer activity in the promoter of *LDLR*, Wang et al identified a DNA sequence highly conserved among primates approximately 13 kb downstream of the gene. This PS2 sequence contains SRE binding motif TCACCCCAC which binds SREBP-2. The PS2 sequence enhances luciferase gene expression driven by the proximal human *LDLR* promoter. The activity of the test vector containing *LDLR* promoter/SREBP-2/PS2 was 5 fold higher than the vector containing *LDLR* promoter (Wang et al. 2006).

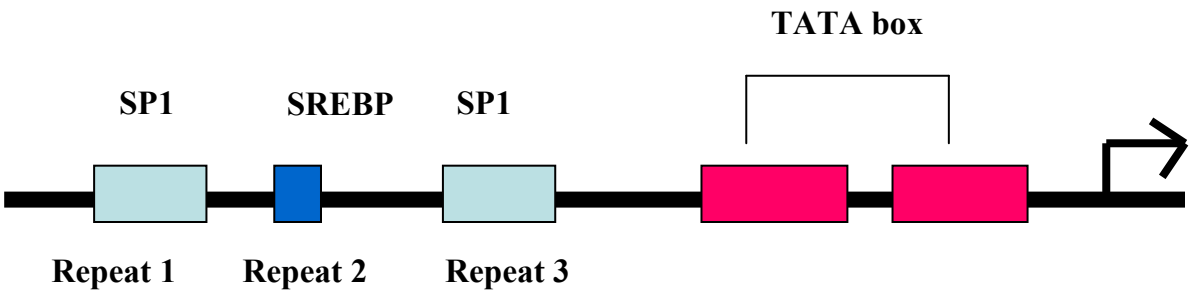

Figure 1.3. The sequences that regulate the transcription of the *LDLR* gene are localized in 177 bp upstream of the initiation codon (AUG), mainly two TATA boxes and three direct repeats. Repeat 1 and repeat 3 bind Sp1 which governs the basal level of transcription, while repeat 2 binds SREBP that drives cholesterol mediated regulation of the transcription**.**

Polymorphisms in the human *LDLR* and plasma lipid parameters

While mutations that render the *LDLR* gene defective cause familial hypercholesterolemia (FH), an inherited autosomal dominant disease with hallmarks of elevated plasma LDL-C levels and premature CHD, most individuals who suffer from CHD do not have FH and their plasma lipoprotein levels are within the reference range, there is unequivocal evidence that lowering serum cholesterol reduces the risk of CHD events (van Aalst-Cohen et al. 2004). Because of its pivotal role in lipid metabolism, polymorphisms rendering LDLR coding sequence changes, as well as quantitative change of *LDLR* expression are a candidate risk factor for CHD, justifying association studies of common genetic variations in the gene.

Several studies have shown associations between genetic variations in *LDLR* and plasma levels of LDL-C (Ahn et al. 1994; Klausen et al. 1993; Poledne et al. 1993; Taylor R 1988). For example, an association between high plasma LDL-C levels and the absence of a *Pvu* II restriction sites in intron 15 has been reported in normolipidemic Norwegian and German populations (Pedersen and Berg 1988; Pedersen JC 1988; Schuster et al. 1990). In addition the presence of *AvaII* site in exon 13 has been reported previously to be associated with increased LDL-C levels in Italian population, Hispanics and non-Hispanics and Brazilian whites.(Humphries et al. 1991; Salazar et al. 2000). Recently a new study has shown that a SNP in exon 12 (rs 688) is associated with increased total and LDL-C in women (Zhu et al. 2007).

To date, there is no completed study on the association of polymorphisms in the 3'UTR regulatory region of the gene. Towards this goal, is my chapter two, I studied the

association of three common polymorphisms with various lipid parameters, and the effect of these SNPs on gene expression.

Atherosclerosis risk and *APOE* isoforms in humans

Apo E, a glycoprotein of 299 amino acids, plays a pivotal role in the clearance of plasma chylomicron, chylomicron remnants, VLDL, IDL and HDL via its interaction with the LDLR and LDLR like proteins (LRP) .(Corbo and Scacchi 1999; Corbo RM 1999; Hussain et al. 1999; Melman et al. 2002a). It is mainly synthesized in the liver, and to a lesser extent in the kidney, brain, adrenals and monocytes. In humans, there are di-allelic single nucleotide polymorphisms located in exon 3 at positions T3937C and C4075T which together yield the three isoforms, *APOE*2*, *APOE*3* and *APOE*4*, with gene frequencies of 0.08, 0.77, and 0.15 respectively (Table 1) (Banares et al. 2005; Boerwinkle and Utermann 1988b; - Weisgraber 1994; Weisgraber 1994) (Table 1.1).

Table 1.1. SNP position and amino acid changes in human *APO gene*

SNP postion	AA position	AA WT	AA mutant	Mutation
3937	112	Cystein	Arginine	TGC CGC
4075	158	Arginine	Cystein	CGC TGC

AA, amino acid; WT, wild type

These isoforms differ in the amino acids at position 112 and 158. The most frequent isoform, *APOE*3*, has cysteine and arginine at these positions, while *APOE*2* has two cysteines and *APOE*4* has two arginines (Lund-Katz et al. 2001). The genotype frequency distribution follows the same pattern in the population with E3/E3 as the highest followed by E3/E4, E2/E3, E2/E4, E4/E4 and E2/E2.

The isoforms have different binding affinities for the LDLR. ApoE4 has a higher binding affinity than apoE3, while apoE2 binds very poorly (Lund-Katz et al. 2001; Saito et al. 2003; Saito H 2003). Despite its poor receptor binding however, individuals with apoE2 have a reduced level of cholesterol and a reduced risk of atherosclerosis, except for 5%–10 % of apoE2 homozygotes who have a very high level of plasma cholesterol and exhibit Type III hyperlipoproteinemia (Kypreos et al. 2003.). In contrast, individuals with apoE4 have slightly increased plasma cholesterol levels and an increased risk of atherosclerosis (Banares et al. 2005; Malloy I 2004). This paradox in binding affinity and atherosclerosis risk has been explained by Davingon's hypothesis, which states that the high binding affinity of apoE4 for the receptor results in rapid clearance of chylomicron and VLDL remnants, causing intracellular increase of cholesterol, as a consequence, there is a down regulation of the *LDLR*, while the low binding affinity of apoE2 results in slow clearance of chylomicron and VLDL remnants , hence low intracellular level of cholesterol, and therefore an up regulation of the *LDLR* (Davignon et al. 1988). According to Sing and Davignon, 8.3 % of the total variance for LDL-C is caused by the apoE isoforms (Sing and Davignon 1985). To compare the risk attributed by the different *APOE* isoforms, and compared to the homozygous *APOE*3* genotype as a reference genotype, one allele of *APOE*2* lowers cholesterol level by 10-14 mg/d, while one allele of *APOE*4* raises the level by 8-10 mg/dl

(Hallman et al. 1991). This difference is shown in most populations, regardless of the differences in the genetic background and life style such as diet and exercise.

Functional gene-gene interaction between *Apo E* isoforms and *LDLR* polymorphisms: A determinant of lipid level variations

Mice with humanized genes for lipoprotein metabolism

To test Davignon's hypothesis that up regulation or down regulation of *LDLR* in individuals with *APOE*2* or *APOE*4*, respectively, is the cause of the isoform-dependent risk for atherosclerosis, Malloy et al replaced the coding sequence of the mouse *Apoe* gene with that coding for the three human isoforms, *APOE*2,* or *APO*E3,* or *APOE*4*. Surprisingly, they found that the phenotypes of mice having the human apoE proteins are not the same as in humans. Rather, they found that the three types are what would be expected simply from the biochemical behavior of the proteins. Thus, the mice with human, apoE3 and apoE4, both of which bind to the receptor well, are normolipidemic and did not develp atherosclerosis on a western-type diet, while all the mice with human apoE2, which does not bind to the receptor well, exhibit type III hyperlipoproteinemia and develop atherosclerosis (Malloy I 2004).

Unexpectedly, and contrary to the current hypothesis, when Malloy et al increased expression of the LDLR in these mice, the mice with human apoE4 (4h), developed severe atherosclerosis due to an accumulation of cholesterol in the plasma when fed a western-type diet (Malloy I 2004). This trend was not seen in mice with apoE3 (3h). The same increase of LDLR, on the other hand, completely ameliorated the hyperlipidemia and atherosclerosis of the apoE*2*/E2 mice (2h) (Knouff et al. 2001). This finding led Malloy and his group to

propose a new "ApoE Trapping" hypothesis: namely, that the LDLR can trap apoE4 and reduce its availability for transfer to newly formed lipoproteins leading to an increase in the plasma cholesterol concentration (Michael K. Altenburg et al. 2004).

The overall consequence is that increased *LDLR* expression has harmful effects in mice with apoE4. In contrast, apoE2 with lower affinity for the LDLR virtually escapes trapping by the LDLR while, increased LDLR helps clearance of apoE2 enriched lipoprotein particles.

Therefore, the normal variations in the basal level of *LDLR* expression cause a difference in relative ratio of apoE/LDLR in humans as illustrated in the grey area of Figure 1.4, and this difference determines the atherosclerosis risk in apoE isoform dependent fashion.

My chapter 3 tests whether LDLR polymorphisms and human apoE polymorphisms interact or not in determining plasma lipoprotein profiles using ARIC population.

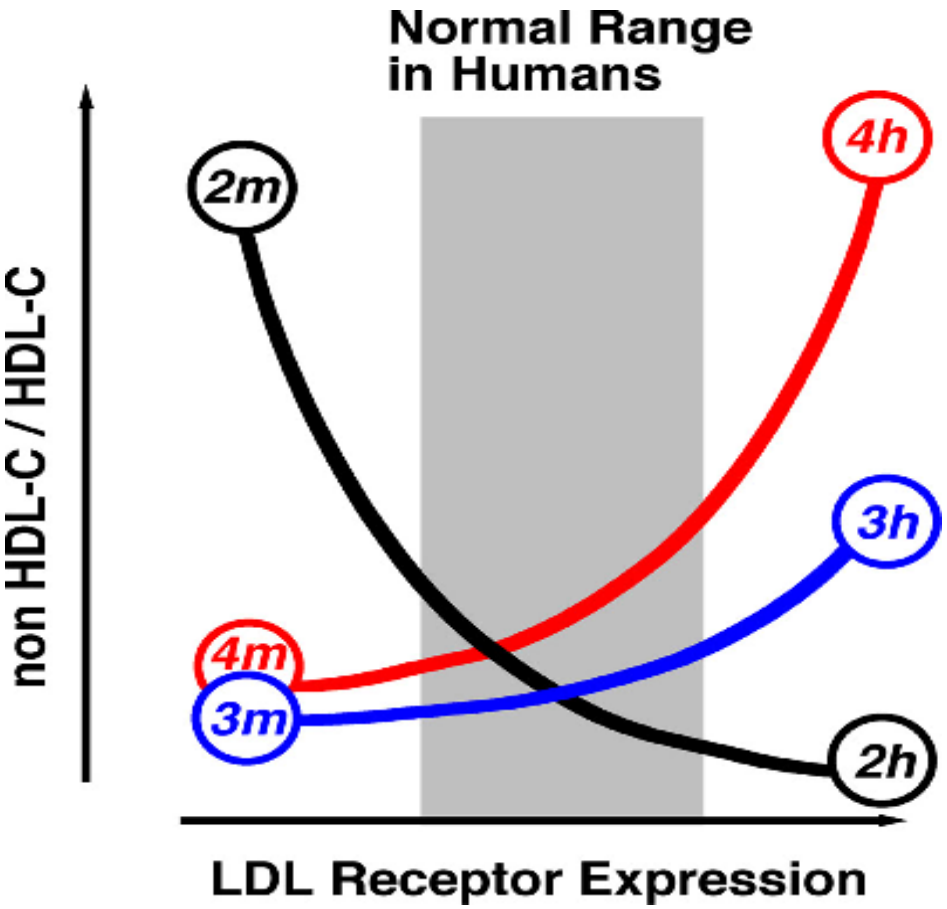

Figure 1.4. Interaction of *APOE* isoforms and *LDLR* expression in Atherosclerosis Risk
On the left hand side of the figure, 2m,3m and 4m represent mice that have mouse *Ldlr* and human apoE isoforms. We see that at a normal level of expression of LDLR, the mice expressing human apoE2 (2m) have a higher risk of developing atherosclerosis than the mice expressing human apoE4 (4m) as measured by the ratio of non HDL-C/HDL-C. The relative order of the risk for atherosclerosis in mice with the human apoE isoforms (2m>, 4m>, 3m) is shown in the left side of Figure 1
On the right hand side of the figure, 2h, 3h and 4h represent mice that carry the human apoE isoforms as well as one copy of the human *LDLR*. When the mouse *Ldlr* is replaced with the human *LDLR,* the expression is increased 2.5 times. The relationship between risk and *APOE* genotype is changed in the mice with a high level of expression of *LDLR*; 2h mice have a lower risk of atherosclerosis than the 4h mice. The relative order of the atherosclerosis risk of the mice with the human receptor (2h <, 3h<, 4h) is shown in the right side of the figure.
Since risk in humans with different apoE isoforms are much less, we expect that normal ranges of *LDLR* expression in humans are somewhere in the middle as shown in the grey box.

Enhancer in intron I of the human *LDLR* gene

Enhancers, by definition, are the DNA sequences where activator proteins bind to increase significantly the rate of transcription. The exact location of an enhancer on a gene varies, and it may be located thousands of nucleotides away from the promoter of a gene. Also, enhancers activate transcription independent of orientation, i.e. whether they are in the forward or reverse direction. Once activator proteins bind to the enhancer, they recruit histone acetylase transferase and ATP-dependent chromatin remodeling complex, both of which aid in the assembly of transcription factors at the promoter and thereby stimulate transcription initiation. In some genes, enhancers are located in introns. For example, enhancers have been found in intron I of *HPRT* gene (Reid et al. 1990), β actin (Frederickson 1989), adenosine deaminase (Aronow et al. 1989) and others (Banerji et al. 1983; Basler et al. 1989; Chung and Perry 1989; Oshima et al. 1990; Reddy and Reddy 1989; Rippe et al. 1989). The transcriptional ability of the DNA sequence 5' to the *LDLR* gene has been well characterized by Sudhof et al (Sudhof et al. 1987), who found that the 1774 bp promoter sequence is sufficient for the control of *LDLR* gene expression. This region contains a TATA box, but not CAAT box. It also contains three short direct sequence repeats where steroid regulatory element binding protein (SREBP) binds. There appears to be no enhancer activitiy in at least 6.5 Kb of the 5'upsream of the promoter region, as measured by reporter assay (Sudhof et al. 1987). In addition to the enhancer activity in the promoter of *LDLR*, a primate-specific enhancer element PS2 was found downstream of the gene as mentioned earlier in the introduction (Wang et al. 2006).

Research Project

Based on the important role that the LDLR plays in the development of hypercholesterolemia and CHD, my specific aims were:

1 - To identify six common *LDLR* sequence polymorphisms in the 3'UTR of the gene and to investigate the association of the conferred common haplotypes with the lipid parameters in the ARIC Study. Determine the effect of these polymorphisms on a reporter gene expression in cultures cells.

2 - Analyze gene-gene interaction between *LDLR* diplotypes and *Apo*E* isoforms in determining plasma lipid profile and cardiovascular risk using multiple regression analysis in ARIC study.

3 - Test enhancer activity in intron I of the human *LDLR* gene and identify the transcription factors that bind to specific sequences in the intron. Then study the biological effect of the SNPs that are located in this region.

Chapter II

Common Genetic Variations in the 3'UTR of the Human LDL-Receptor Gene and their Associations with Plasma Lipid Levels in the Atherosclerosis Risk in Communities (ARIC) Study

Hind Muallem[1], Kari E North[2], Masao Kakoki[1], Mary K. Wojczynski[2], Xia Li[2], Megan Grove[3], Eric Boerwinkle[3], Kirk C. Wilhelmsen[4], Gerardo Heiss[2], and Nobuyo Maeda[1, *]

Abstract

The low-density lipoprotein receptor (LDLR) plays a pivotal role in cholesterol homeostasis. However, the role of genetic variations in the 3' UTR of the *LDLR* in relation to plasma cholesterol has been largely understudied. Six SNPs, G44243A, G44332A, C44506G, G44695A, C44857T and A44964G, within the 5' region of the 3'UTR fall into three common haplotypes, *GGCGCA*, A*GCACG*, and *GGCGTA*, occurring at frequencies of 0.45, 0.31 and 0.17 respectively in Caucasians (n=29) and 0.13, 0.13 and 0.38 respectively in African Americans (n=32), with three other haplotypes occurring at lesser frequencies. Genotyping of two "haplotype tagging" SNPs, C44857T and A44964G, in the Atherosclerosis Risk in Community (ARIC) study population showed that in Caucasians, but not in African Americans, the inferred *TA* haplotype had a significant LDL-cholesterol lowering effect. The adjusted LDL-cholesterol levels in the *TA/TA* diplotypes were lower by 6.10 mg/dl in men ($P<0.001$) and by 4.63 mg/dl in women ($P<0.01$) than in individuals with other diplotypes. Caucasian men homozygous for *CA,* in contrast, showed significantly higher LDL-cholesterol ($P<0.04$), lower HDL-cholesterol ($P<0.02$) and higher LDL/HDL ratios ($P<0.001$). To identify if these haplotypes are causative, we studied the expression of a reporter gene in a tissue culture based system carrying a 3'UTR that includes the 1kb nucleotide sequences corresponding to the *AGCACG*, *GGCGTA* and *GGCGCA* haplotypes. The three haplotypes of the 3'UTR sequences had no difference in their effect on the expression. Therefore, the association in Caucasians with plasma lipid profiles indicative of higher cardiovascular risk is not caused by polymorphisms in the 3'UTR of the LDLR gene, but with other polymorphisms that are in linkage disequilibrium with them.

Introduction

Low-density lipoprotein receptor (LDLR) is a cell surface receptor that mediates the uptake of LDL particles from the circulation via receptor-mediated endocytosis (Goldstein et al., 1995; Hobbs et al., 1990). Mutations that render the *LDLR* gene defective cause familial hypercholesterolemia (FH), an inherited autosomal dominant disease with hallmarks of elevated plasma LDL-cholesterol (LDL-C) levels and premature coronary heart disease (CHD). Although most individuals who suffer from CHD do not have FH and their plasma lipoprotein levels are within the reference range, there is unequivocal evidence that lowering serum cholesterol reduces the risk of CHD events (Gotto, 1994). Thus, because of its pivotal role in lipid metabolism, a quantitative change of *LDLR* expression is plausibly associated with risk for CHD, and warrants association studies of common genetic variations in non-coding regions of the *LDLR* gene that may influence its expression at transcriptional or post-transcriptional mRNA levels (Sudohof et al., 1987; Wilson et al., 1998). In addition, LDLR level is also determined by translation as well as protein degradation. For example, genetic variations in the *PCSK9* gene coding for proprotein convertase subtilisin/kexine Type 9, affect PSCK in protein degradation pathway, and increased PSCK activity reduces hepatic LDLR protein in the liver and leads to higher levels of plasma LDL cholesterol (Cohen et al., 2006; Kotowski et al., 2006).

Several studies have shown associations between genetic variations in the *LDLR* gene and plasma levels of LDL (Taylor et al., 1988; Poledne et al., 1993; Klausen et al., 1993; Ahn et al., 1994). For example, an association between high plasma LDL-C levels and the absence of a *Pvu* II restriction site in intron 15 has been reported in normolipidemic Norwegian and German populations (Pedersen et al., 1988; Schuster et al., 1990). However,

to our knowledge no studies have been reported that attempt to identify the causative polymorphisms in the *LDLR* that functionally impact on plasma LDL-C. Towards this goal, we report on the common allelic variations in the 3' untranslated region (UTR) of the *LDLR* gene in humans, their functional impact on gene expression, and their association with variation in plasma lipid levels.

The 3'UTR sequences have long been recognized as a regulatory region that is important for the appropriate expression of many genes (Conne et al., 2000; Chen et al., 2006). It contains cis-elements that affect the stability of transcripts, and/or influence the translation by interacting with the 5' sequence of the transcripts. For example, a stretch of AU-rich element (ARE) containing an AUUUA penta-nucleotide motif serves as a recognition site for specific proteins that increase the rate of poly A tail shortening and mRNA degradation. The 3'UTR of the human *LDLR* gene includes three regions each containing an ARE, and Wilson et al. (1998) demonstrated that each element shortened the half-life of the reporter gene transcript by approximately 30% when introduced into the 3'UTR of the β globin gene in a tissue culture based system. Furthermore, we have previously shown that the removal of two of the three AREs in the 3'UTR of the *LDLR* gene leads to more stable transcripts and increased steady state levels of LDLR mRNA in mice (Knouff et al., 2001). Therefore, common SNPs within or near these cis-elements may affect the binding of proteins to ARE or other control elements and hence alter gene expression in general human populations.

We describe here the identification of common genetic variations in the 3'UTR of the human *LDLR* gene. We also demonstrate that these common variants have significant

association with plasma lipoprotein profiles in Caucasians enrolled in the Atherosclerosis Risk in Communities (ARIC) study.

Materials and Methods

Identification of common polymorphisms in the mRNA LDLR 3'UTR

DNA was isolated from blood samples or fibroblasts from unrelated individuals, five Caucasians and nine African Americans (Maeda et al., 1983; Maeda, 1991). The use of DNA in the current study was approved by the institutional review board of the University of North Carolina. One kb DNA containing the 5' of the 3'UTR of the human *LDLR* was PCR amplified from the genomic DNA using 1:1 mixture of Tgo DNA polymerase (Roche Applied Science, Penzberg Germany) and Taq DNA polymerase (New England Biolabs, Beverly, MA). The forward and reverse primers, 5'-TTTACGCGTACGTGGCGTGAACATCTGC-3' and 5'-TTTGGCGCGCCTCACGATCACGGCTTTGAAG-3', contained additional nucleotides to generate *Mlu*I and *Asc*I restriction sites at the 5' and 3' ends of the PCR products, respectively. The PCR product was run on 0.8 % agarose gel to confirm the correct size, purified using QIAquick gel extraction kit (Qiagen Inc, Valencia, CA), and sequenced with each of the PCR primers. Nucleotide sequences were compared to the *LDLR* gene sequence in the GenBank database (NC_000019). PCR products from some individuals were cloned in pBluescript II vector (Stratagene, La Jolla, CA) to verify each haplotype seqeunce. The *Mlu*I/*Asc*I fragments with the three common haplotypes were isolated from these plasmids and cloned into the *Mlu*I site of the *Hprt-hrGFPn* 3'UTR test plasmid vector below.

ARIC study population and genotyping

The ARIC study is a prospective epidemiologic study whose main target is to investigate atherosclerotic vascular diseases and their risk factors (investigators 1989). Blood samples were drawn for the analysis of lipids and lipoproteins and DNA was extracted for genotyping (Brown et al., 1993). The present study includes 10,602 Caucasians and 3,497 African Americans. Fasting plasma lipid data at the first cohort examination visit (1987-1989) was used for this analysis.

A total of 19,986 wells including negative and positive controls, non-blinded replicates, and blinded duplicates of 5% of the study samples were genotyped for C44857T and A44964G using TaqMan SNP allelic discrimination with ABI 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). Primers and probes designed by Applied Biosystems were used (a reference number 4331183: Assay ID# C_2804304_10 / dbSNP# rs1433099 / Public location: Chr 19, Base: 11,103,658, and Assay ID#C_2804305_10 / dbSNP# rs2738466: Base: 11,103,765). A 5 µl reaction mixture was used for PCR amplification that contains 3 ng of genomic DNA in 2.5 µl TaqMan Universal PCR master mix (Ranade et al., 2001). Allelic fluorescence was measured using the 7900HT and analyzed with SDS version 2.0, software (Perkin Elmer/ Applied Biosystems, Foster city, CA). Replicate quality control samples were 100% concordance.

Statistical analysis

All analyses were stratified by race and gender. Haplotypes were estimated manually and by using the phase reconstruction method (PHASE 2.1), which assigned the most probable haplotype pair to each individual (Stephens, 2001; Stephens, 2003).

Deviations from Hardy-Weinberg equilibrium expectations were tested using a goodness-of-fit test. Multiple regression analysis was conducted utilizing SAS Version 8.1 making no assumptions about the inheritance pattern, applying a general model (two degree of freedom test, SAS Institute, Cary, NC). Multivariable linear regression analysis was conducted to compare the mean estimated values of total cholesterol, LDL-C, HDL-C and LDL/HDL ratio in groups of individuals with different *LDLR* SNPs and diplotypes. Analyses were carried out stratified by race and gender either adjusting for age, or fully adjusting for other well established lipid related covariates including age, field center, current drinker, current smoker, former smoker, use of medicine to lower cholesterol, use of lipid-lowering medications, medications that secondarily affect cholesterol, BMI and waist-to-hip ratio for men. Additional adjustments for menopausal status and gonadal hormone use were made for women. The mean adjusted lipid values within race and sex and stratified by *LDLR* genotype were calculated using ANOVA implemented in SAS Version 8.1.

Reporter gene expression assay in mouse embryonic stem cells

The *Hprt-hr*GFPn 3'UTR test vector contains a reporter gene coding for humanized renilla Green Fluorescence Protein with a nuclear localizing sequence (*hrGFPn*, Stratagene, La Jolla,CA) driven by human β actin promoter that contained the first two non-coding exons (Kakoki et al., 2004). The vector also contains two regions of the mouse *Hprt* gene sequence that allows homologous recombination in embryonic stem cells. *Mlu*I/*Asc*I fragments of the *LDLR* 3'UTR were inserted into a *Mlu*I site between the stop codon of *hrGFPn* and a poly A addition sequence of the *Hprt-hr*GFPn 3'UTR test vector. 3'UTR test plasmid DNA (10 μg) was linearized with *Pme*I and electroporated into mouse embryonic

stem cells, E14TG2a, that lack promoter and exon 1 and 2 of the *Hprt* gene. Since homologous recombination restores the *Hpr*t gene, only correctly targeted cells survive in a medium containing 4.8 μg/ml thymidine, 16 μg/ml hypoxanthine and 0.175 μg/ml aminopterin (HAT). After selection for 10 days, drug resistant colonies were picked and expanded, and the fluorescence was measured with a Cytomation MoFlo cytometer having an argon laser for excitation at 488nm (Becton Dickinson, San Jose, CA). 10,000 events were collected for every analysis and the data was analyzed using summit 3.1 (DaKo cytomation, carpenteria, CA). RNA was extracted from cells with Trizol reagent, and the mRNA of *hrGFPn* was measured by quantitative reverse transcription-PCR with the ABI 7700 (Applied Biosystems, Foster City, CA;(Kim et al. 1990). Forward and reverse primers for hrGFPn amplification were 5'- ATCCTGGTGTACCGCCTGAA-3' and 5'-TGCTGGATGAAGTGGTACTC-3', respectively. The probe for hrGFPn detection was 5'-Fam-CAAGTCCTTGTAGATCTCCTGGAGC-TAMARA-3'. The β actin gene expression was used as an internal standard for each reaction.

Results

SNPs and haplotypes of the *LDLR* 3'UTR sequence

We amplified 1.0 kb of DNA corresponding to the 5' part of 3'UTR of the human *LDLR* gene by PCR in genomic DNA isolated from five unrelated Caucasians and nine unrelated African Americans. Nucleotide sequence analyses of the PCR products confirmed polymorphisms at positions 44243, 44332, 44506, 44695, 44857, and 44964 within this region (Figure 2.1; SNP annotations: rs14158, rs3826810, rs2738464, rs2738465, rs1433099, and rs2738466, respectively). One of the five Caucasians, C5, was homozygous for *GGCGCA* at the six SNP sites matching with the reference sequence (NC_000019). We assigned this haplotype as Type I. Individuals C1, C2, and C3 were heterozygous for the Type I and Type II carrying *AGCACG* at the six SNP sites. An individual C4 was homozygous for *GGCGTA*, which we assigned as Type III. These haplotypes were cloned and confirmed via direct sequencing. Three additional haplotypes were found in African American individuals. A haplotype with nucleotide of *GGGGTA* (Type IV) was in individuals AA3 and AA4, who were both heterozygous for this and a Type III. Type V *GGGGCA* was seen in AA5 in combination with a Type III. The G44332A polymorphism was rare with A being present in only one individual, AA1, who carry a sequence with *GACACA* (Type VI). Again we established the haplotypes IV, V and VI by cloning of the individual DNA fragments followed by sequencing. Haplotypes of the samples studied are summarized in Table 2.1.

The six haplotypes also account for the majority of nucleotide polymorphisms of the 24 European Americans and 23 African Americans genotyped by Hinds *et al* (2005) and the PARC Study in Entrez SNP, (NCBI

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=snp). Thus the haplotypes of 20 of the 24 European Americans are easily assigned into Types I, II or III. The remaining three are heterozygous for Types II and IV, and one is heterozygous for Types III and VI. Similarly, the haplotypes of 23 African Americans are readily assignable with haplotypes I through VI, except for two who are either heterozygous for Type I and IV or for Types III and V. Considering all available genotype data (Table 2.2), Types I, II and III are the most common haplotypes, occurring at frequencies of 0.45, 0.31 and 0.17 respectively in Caucasians (n=29) and 0.13, 0.13 and 0.38 respectively in African Americans (n=32). The Type IV haplotype occurs at a frequency of 0.19 in African Americans. The three common sequences Type I, II and III together account for 93% of the 3'UTR sequence of the *LDLR* gene in Caucasians, while they account for 64% of that in African Americans.

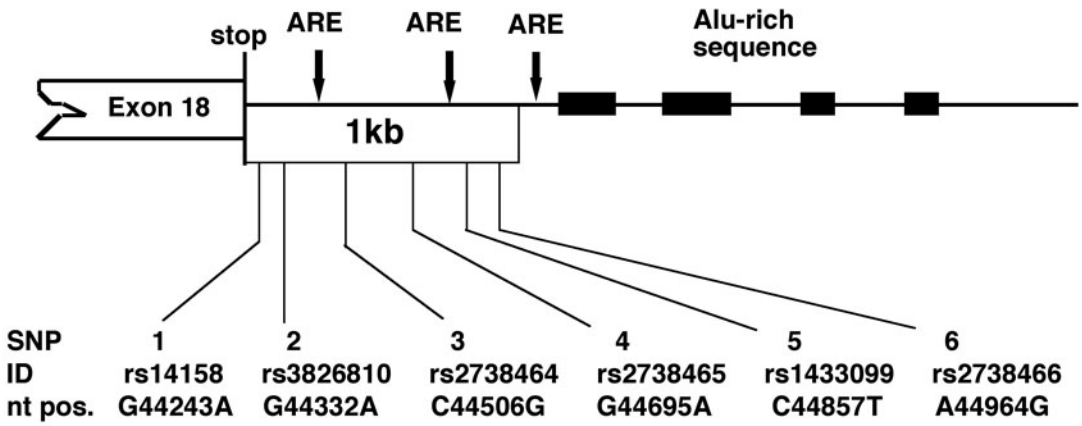

Figure 2.1. 3'UTR of the human *LDLR* gene. A white box following the stop codon within the exon 18 is the 1kb region studied in the present work. Positions of three ARE sequences are indicated by arrows. Black boxes represent the Alu sequences in the remaining 2.5kb of the 3'UTR. Locations of SNP sites and major and minor nucleotides are shown with the Gene bank dnSNP identification (ID) and nucleotide positions according to NC_000019 (nt. pos).

Table 2.1 SNPs in the 3'UTR of *LDLR* seen in individuals (AA=African American, C=Caucasian)

ID	44243	44332	44506	44695	44857	44964	Haplotypes (Predicted)
AA1	*G/G*	*G/A*	*C/C*	*G/A*	*C/T*	*A/A*	III / VI*
AA2	*G/A*	*G/G*	*C/C*	*G/A*	*C/T*	*A/G*	II / III
AA3	*G/G*	*G/G*	*C/G*	*G/G*	*T/T*	*A/A*	III / IV*
AA4	*G/G*	*G/G*	*C/G*	*G/G*	*T/T*	*A/A*	III / IV
AA5	*G/G*	*G/G*	*C/G*	*G/G*	*C/T*	*A/A*	III / V*
AA6	*G/G*	*G/G*	*C/C*	*G/G*	*T/T*	*A/A*	III / III
AA7	*G/A*	*G/G*	*C/C*	*G/A*	*C/T*	*A/G*	II / III
AA8	*G/A*	*G/G*	*C/C*	*G/A*	*C/C*	*A/G*	I / II
AA9	*G/G*	*G/G*	*C/C*	*G/G*	*T/T*	*A/A*	III / III
C1	*G/A*	*G/G*	*C/C*	*G/A*	*C/C*	*A/G*	I / II*
C2	*G/A*	*G/G*	*C/C*	*G/A*	*C/C*	*A/G*	I / II
C3	*G/A*	*G/G*	*C/C*	*G/A*	*C/C*	*A/G*	I / II
C4	*G/G*	*G/G*	*C/C*	*G/G*	*T/T*	*A/A*	III / III*
C5	*G/G*	*G/G*	*C/C*	*G/G*	*C/C*	*A/A*	I / I*

*Established by cloning and sequencing.

Genotype assignment and frequencies in ARIC subjects

We searched in NCBI genome databank the homologous 3'UTR sequence of the *LDLR* gene in other primates. The nucleotides at the six SNP sites are *GGCGTA* in Colobus (AC150433), *GGCGTG* in Baboons (AC140974), *GGCGTG* in Rhesus (CN642890), and *GGCGCG* in Chimpanzees (NW_119754). While the nucleotides at the six SNP sites of the most common haplotype (*GGCGTA*) in African Americans were shared with those in the Colobus monkey, other combinations including the most common haplotype (*GGCGCA*) were not seen in these primate species. The numbers of sequence entry for each of the different primate species available from the NCBI databank were limited. However, when multiple sequences were available, we did not find intra-species differences at these sites, although we were not able to find whether these sequences were all from a single individual or from multiple different individuals.

Since parallel occurrence of the same nucleotide substitutions during the primate evolution can be considered rare, the polymorphisms at positions 44857 and 44964 are likely of ancient origins. In addition, the combinations of these two SNPs can predict the haplotypes, at least in the majority of Caucasians. Consequently, we chose to genotype these two SNPs (rs1533099 and rs2738466 respectively) in the large ARIC study. The distributions of C44857T and A44964G diplotypes were in Hardy-Weinberg equilibrium for both races ($p>0.05$). The relative frequencies of C44857T and A44964G were significantly different between African Americans and Caucasians (Table 2.3). The frequency of C at 44857 was higher in Caucasians (0.73) compared to the African Americans (0.46). The Caucasians had a lower frequency of A at position 44964 than did the African Americans (0.74 vs 0.83). Allelic frequencies of C44857T, A44964G and their combinations in the

ARIC subjects and in the individuals used to estimate the frequencies of 3'UTR hapotypes (Table 2.2 and 2.3) were not significantly different (Fisher's exact test).

Table 2.2 Frequencies of the 3'UTR haplotypes of the human *LDLR* gene

Haplotype		Caucasians			African Americans		
		This study n=10	Perlegen n=48	Total %	This study n=18	Perlegen n=46*	Total %
I	*GGCGCA*	5	21	45%	1	7	13%
II	*AGCACG*	3	15	31%	3	5	13%
III	*GGCGTA*	2	8	17%	10	14	38%
IV	*GGGGTA*	0	3	5%	2	10	19%
V	*GGGGCA*	0	0	0%	1	5	9%
VI	*GACACA*	0	1	2%	1	5	9%

- Two individuals were heterozygous for either Type I/IV or III/V. Chromosome numbers.were counted.

Three combinations of these two SNPs C44857-A44964 (*CA*: corresponding to haplotypes I, V and VI), C44857-G44964 (*CG*: corresponding to haplotype II) and T44857-A44964 (*TA*: corresponding to haplotypes III and IV) predicted the majority of the haplotypes in the ARIC population (Table 2.2). However, two of 10,602 Caucasians were genotyped as carrying *T/T* at 44857 and *G/G* at 44964; twelve individuals were *C/T* at 44857 but *G/G* at 44964. This suggests that *TG* is a minor haplotype present at a low frequency in this population. Thus, while likelihood-based haplotype inference assigned the 1,465 individuals who are doubly heterozygous for C44857T and A44964G as *CG/TA*, a fraction of them (less than 20) may be *CA/TG*. Of 3,497 African Americans, the inferred haplotype combinations of *TG/TG*, *CG/TG* and *TA/TG* were seen 2, 4, and 43 times respectively. Thus the *TG* haplotype is present at a higher frequency in African Americans than in Caucasians, and approximately 5-10% of those being assigned as *CG/TA* may be *CA/TG*. Individuals with *TG/TG*, *CG/TG* and *TA/TG* were grouped as rare.

The inferred diplotype frequencies differed significantly between Caucasians and African Americans in this study. The Caucasian group had a higher frequency than African Americans for *CA/CA* (22.18% vs. 9.09%), *CA/CG* (25.31% vs 9.78%) and *CG/CG* (6.21% vs 2.60%), while the African American group had higher frequencies than Caucasians in *CA/TA* (31.20% vs 24.7%), *CG/TA* (18.19% vs 13.82%) and *TA/TA* (27.80% vs 7.67%).

Table 2.3. Frequencies of C44857T and A44964G, and of combined alleles in ARIC study population

	Caucasian		African American	
	Men N=4,967	Women N=5,635	Men N=1,316	Women N=2,181
		C44857T		
C/C	2,628 (52.9%)	3,065 (54.4%)	296 (22.5%)	455 (20.8%)
C/T	1,932 (39.0%)	2,155 (38.2%)	661 (50.2%)	1,069 (49.0%)
T/T	407 (8.2%)	415 (7.4%)	359 (27.3%)	657 (30.1%)
		A44964G		
A/A	2,749 (55.3%)	3,035 (53.8%)	881 (66.9%)	1,500 (68.8%)
A/G	1,890 (38.1%)	2,265 (40.2%)	406 (30.8%)	615 (28.2%)
G/G	328 (6.6%)	335 (5.9%)	29 (2.2%)	66 (3.0%)
		Alleles (C44857T - A44964G)		
C-T	46.8%	47.6%	31.7%	28.9%
C-G	25.6%	26.0%	17.6%	16.6%
T-A	27.5%	26.4%	51.7%	53.9%
T-G	0.05%	0.08%	0.6%	0.7%

Inferred haplotypes of *LDLR* 3'UTR and lipid profiles in ARIC subjects.

Of the 15,792 total subjects who agreed to participate in the ARIC study, 1,693 were excluded because their lipid and lipoprotein data at their first visit (baseline) were not available or because their plasma total cholesterol levels were four standard deviations from the mean, or their informed consent precluded the use of their DNA. The study population characteristics are presented in Supplementary Table 2.7, subdivided by race and gender.

Multiple regression analysis was conducted to compare the mean estimated values of total cholesterol, LDL-C, HDL-C and LDL/HDL ratio in groups of individuals with different *LDLR* genotypes (Table 2 4). Fully adjusted mean LDL-C levels were 7 mg/dl and 4 mg/dl higher in *C/C* (44857) homozygous Caucasian males and females than *T/T* (44857) homozygous (P=0.0016 and P<0.06 respectively). LDL-C levels in African American females were 3 mg/dl higher in *C/C* homozygous compared to *T/T* homozygotes (P=0.43). In contrast, no effects of C44857T on LDL-C levels were detected in African American males. The A44965G polymorphism had no effects on the lipid profiles in either population. Analysis carried out adjusting only for age showed no differences in mean values and significance from fully adjusted analysis (not shown).

Table 2.4. Adjusted plasma lipid parameters in ARIC subjects according to genotypes of C44857T and A44965G

SNP	C44857T			P*	A44965G			P*
	C/C	C/T	T/T		A/A	A/G	G/G	
C, Men	N=2,628	N=1,932	N=4,07		N=2,749	N=1,890	N=328	
LDL/HDL[a]	3.54±0.02	3.46±0.03	3.34±0.06	*0.005*	3.50±0.02	3.49±0.03	3.49±0.07	0.93
HDL[b]	42±1	43±1	42±1	0.080	43±1	43±1	42±1	0.36
LDL[c]	141±1	140±1	134±2	*0.003*	140±1	139±1	140±1	0.77
TC[d]	211±1	211±1	205±2	*0.006*	211±1	210±1	210±2	0.88
C, Women	N=3,065	N=2,155	N=415		N=3,035	N=2,265	N=335	
LDL/HDL[a]	2.61±0.02	2.58±0.02	2.51±0.05	0.17	2.59±0.02	2.57±0.02	2.67±0.06	0.31
HDL[b]	57±1	58±1	57±1	0.61	58±1	58±1	58±1	0.73
LDL[c]	135±1	136±1	131±2	0.07	136±1	134±1	138±2	0.22
TC[d]	218±1	218±1	214±2	0.15	218±1	217±1	221±2	0.36
AA, Men	N=296	N=661	N=359		N=881	N=406	N=29	
LDL/HDL[a]	3.01±0.07	2.9911±0.05	3.0915±0.07	0.47	3.07±0.04	2.92±0.06	2.97±1.00	0.13
HDL[b]	51±1	51±1	49±1	0.07	50±1	51±1	51±3	0.26
LDL[c]	137±2	138±2	137±2	0.91	139±1	137±2	126±8	0.19

TC[d]	212±3	212±2	209±2	0.61	212±2	210±2	205±8	0.66
AA,Women	N=455	N=1069	N=657		N=1500	N=615	N=66	
LDL/HDL[a]	2.66±0.05	2.53±0.04	2.62±0.05	0.08	2.59±0.03	2.58±0.05	2.57±0.14	0.99
HDL[b]	58 ±1	58±1	57 ±1	0.43	57±1	58 ±1	61±2.	0.27
LDL[c]	141±2.	135±1	138±2	*0.04*	137±1	137±2	140±5	0.80
TC[d]	221±2	214±1	216±2	*0.03*	215±1	217±2	222±6	0.35

We next compared the mean value of each lipoprotein variable of the individuals with inferred C44857T - A44965G diplotypes against the mean of the given variable of all other individuals who did not have the inferred diplotype (Table 2.5). The most striking associations between the diplotypes and plasma cholesterol levels were observed in Caucasian men. Caucasian men with *TA/TA* had 6.06 mg/dl lower total cholesterol (*P*= 0.002), 6.10 mg/dl lower LDL-C (*P*= 0.001) and lower LDL/HDL ratio (*P*= 0.01) than those with other diplotypes. Conversely, men with the *CA/CA* had 2.46 mg/dl higher LDL-C (*P*=0.04) and 0.91 mg/dl lower HDL-C (*P*= 0.02) than those with other diplotypes respectively. The estimated mean of LDL/HDL ratio was significantly higher in men with the *CA/CA* (*P*=0.001). Caucasian women with the *TA/TA* diplotypes demonstrated a similar pattern of results with 4.07 mg/dl lower total cholesterol (*P*= 0.04), and 4.63 mg/dl lower LDL-C (*P*=0.01) when compared to women with all other diplotypes. The cholesterol raising effects of *CA/CA* diplotypes were not observed in the Caucasian female ARIC participants. Inferred diplotypes were not associated with plasma lipid profiles in African American men except that *TA/TA* diplotypes had significantly lower (2.12 mg/dl, *P*=0.03) HDL-C levels when compared to all other African American men. No diplotype effects were noted in the African American women. We also compared the mean value of each lipoprotein variable of the individuals with inferred C44857T- A44965G diplotypes against the mean of the most common diplotype of each group (CA/CG in Caucasians and CA/TA in African Americans) and results were very similar. (Supplementary Table 2.8).

Table 2.5. Diplotype frequencies and estimated differences in lipoprotein cholesterol levels in individuals with a given diplotype compared to those with other diplotypes.

Diplotypes	Frequency	TC[a]		LDL-C[b]		HDL-C[c]		LDL/HDL[d]	
	N (%)	Estimate	(*P**)	Estimate	(*P**)	Estimate	(*P**)	Estimate	(*P**)
Caucasian Men									
CA/CA	1,104 (22.2%)	1.85	(0.15)	*2.46*	*(0.04)*	*-0.91*	*(0.02)*	*0.14*	*(0.001)*
CA/CG	1,196 (24.1%)	0.29	(0.82)	0.15	(0.89)	0.28	(0.46)	0.007	(0.87)
CA/TA	1,243 (25.0%)	1.37	(0.27)	1.06	(0.36)	*0.99*	*(0.009)*	-0.04	(0.27)
CG/CG	328 (6.6%)	-0.49	(0.82)	-0.13	(0.94)	-0.93	(0.15)	-0.003	(0.96)
CG/TA	689 (13.9%)	-1.08	(0.48)	-1.46	(0.32)	-0.05	(0.90)	-0.03	(0.53)
TA/TA	402 (8.1%)	*-6.06*	*(0.002)*	*-6.10*	*(0.001)*	-0.15	(0.80)	*-0.16*	*(0.01)*
Caucasian Women									
CA/CA	1,248 (22.2%)	0.53	(0.67)	0.56	(0.64)	-0.15	(0.74)	0.01	(0.67)
CA/CG	1,487 (26.4%)	-0.50	(0.67)	-0.45	(0.69)	-0.32	(0.47)	0.02	(0.51)
CA/TA	1,376 (24.4%)	1.48	(0.23)	2.08	(0.07)	-0.20	(0.66)	0.03	(0.37)
CG/CG	330 (5.9%)	2.96	(0.19)	2.80	(0.20)	0.15	(0.85)	0.07	(0.25)
CG/TA	776 (13.8%)	-1.28	(0.41)	-2.10	(0.15)	*1.22*	*(0.03)*	*-0.10*	*(0.02)*
TA/TA	411 (7.3%)	*-4.07*	*(0.04)*	*-4.63*	*(0.01)*	-0.14	(0.85)	-0.09	(0.10)
African American Men									
CA/CA	124 (9.4%)	4.64	(0.26)	4.90	(0.21)	-0.57	(0.70)	0.14	(0.21)
CA/CG	145 (11.0%)	-1.36	(0.73)	-3.52	(0.34)	*2.99*	*(0.03)*	-0.16	(0.13)
CA/TA	413 (31.4%)	2.31	(0.38)	1.35	(0.59)	0.54	(0.57)	0.001	(0.99)
CG/ CG	27 (2.1%)	-5.09	(0.55)	-11.24	(0.17)	-0.98	(0.75)	-0.02	(0.94)

CG/TA	246 (18.7%)	-0.67 (0.83)	0.02 (0.99)	0.54 (0.63)	-0.10 (0.27)
TA/TA	344 (26.2%)	-2.56 (0.36)	-0.18 (0.94)	*-2.12 (0.03)*	0.10 (0.19)
African American Women					
CA/CA	194 (8.9%)	0.94 (0.78)	1.74 (0.59)	-0.68 (0.60)	0.05 (0.52)
CA/CG	197 (9.0%)	*8.11 (0.01)*	*7.04 (0.03)*	0.62 (0.62)	0.13 (0.16)
CA/TA	678 (31.1%)	-3.26 (0.12)	-2.82 (0.16)	0.03 (0.97)	-0.07 (0.18)
CG/CG	64 (2.9%)	7.21 (0.21)	3.87 (0.48)	2.61 (0.22)	0.02 (0.91)
CG/TA	390 (17.9%)	-2.18 (0.39)	-3.07 (0.21)	0.94 (0.32)	-0.08 (0.22)
TA/TA	628 (28.8%)	0.33 (0.87)	1.16 (0.57)	-0.85 (0.29)	0.06 (0.29)

**P* value is from a test comparing the mean adjusted estimate of the lipid value in individuals with a given diplotype compared to all others who do not have the diplotype of interest. The p value from an overall test comparing the unadjusted estimate of the lipid value in individuals with a given diplotype compared to all others who do not have the diplotype of interest are similar to the adjusted (not shown).

[a,] Values were adjusted for age, field center, current smoker, use of medicine to lower cholesterol. In addition, values were adjusted for former smoker and waist to hip ratio for men, and for medicine that secondarily affects cholesterol, menopausal status, current hormone usage, BMI, and waist-to-hip ratio for women.

[b,] Values were adjusted for age, field center, current smoker, use of medicine to lower cholesterol. In addition, values were adjusted for former smoker and leisure physical activity index for men, and for menopausal status, current hormone usage, interaction between menopause and hormone usage, and waist-to-hip ratio for women.

[c,] Values were adjusted for age, current smoker, use of medicine to lower cholesterol, and sport physical activity index. In addition, values were adjusted for field center and former smoke for men, and for current drinker, menopausal status, current hormone usage, interaction between menopause and hormone usage, BMI and waist-to-hip ratio for women.

[d,] Values were adjusted to age, current smoker, current drinker, use of medicine to lower cholesterol, use of medicine that secondarily affects cholesterol, BMI, and waist-to-hip ratio. In addition, values were adjusted for field center for men, and for former smoker, menopausal status, current hormone usage, interaction between menopause and hormone usage for women.

Effects of *LDLR* 3'UTR sequences on gene expression

To test whether the common sequences Type I (*GGCGCA*), Type II (*AGCACG*), and Type III (*GGCGTA*) in the 3'UTR of *LDLR* affect the level of gene expression, we used a *Hprt-hr*GFPn-3'UTR test system (Kakoki et al., 2004). Since Type I, type II, and type III did not show any associations with lipid parametrs in African Americans, we then studied the expression of additional reporter constructs containing type IV (GGGGTA) and type V (GGGGCA) sequences which are more common in African Americans. The five plasmid constructs, each carrying the 3'UTR sequences of Type I, II, III, IV and V immediately 3' to the stop codon of the *hrGFPn* reporter gene, were introduced into mouse embryonic stem (ES) cells with a deletion in the *Hprt* gene. Homologous recombination between the reporter constructs with ES cell genome inserts a single copy of the reporter gene into the *Hprt* locus and simultaneously restores the *Hprt* gene function and allows the cells to survive in a medium containing thymidine, hypoxanthine and aminopterin (Figure 2.2). The experiment was repeated four times and the average of the median intensity of GFP Fluorescence of 10,000 cells collected by flow cytometer per experiment was compared. The results demonstrate that the *hrGFPn* mRNA levels measured by quantitative reverse transcription-PCR showing that mRNA levels in the cells with Type I, II, III, IV,and V sequences was not statistically significant. Also, the GFP fluorescent intensity correlated well with the *hrGFPn* mRNA levels. These data demonstrate that the common polymorphisms in the 3'UTR of the human *LDLR* gene do not affect expression of the reporter gene in a tissue culture system (Table 2.6).

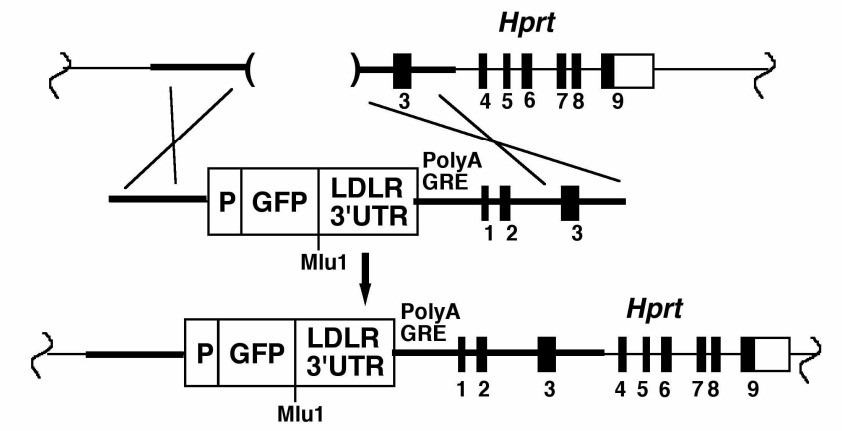

Figure 2.2. Effects of the 3'UTR haplotypes I, II III IV and V on reporter gene expression, 3'UTR test system. The target *Hprt* locus (top line) in the mouse ES cells, E14TG2a, has a deletion of DNA including promoter and exons 1 and 2 as indicated by the parenthesis. Black and white boxes indicate coding and non-coding exons 3 through 9. The reporter gene construct (middle line) contains a human β actin promoter, the *hrGFPn* coding sequence, and 5' and 3' regions of homology arms (thick horizontal lines). Each of the 1kb DNA from the human *LDLR* 3'UTR was inserted into the *Mlu*I site just 3' to the stop codon. Homologous recombination between the two arms of homologies (large Xs) restores the *Hprt* gene and inserts a single copy of the reporter gene (bottom line).

Table 2. 6. Effects of the 3'UTR haplotypes I, II, III, VI and V on reporter gene expression

Haplotypes		Reporter Gene Expression	
		Fluorescence	mRNA
I	GGCGCA	100±13	100±6
II	AGCACG	106±13	133±22
III	GGCGTA	102±14	139±15
IV	GGGGTA	108±13	121±13
V	GGGGCA	109±13	133±17

Haplotypes are nuclerotide positions 44243, 44332, 44506, 44695, 44857, and 44964 (rs14158, rs3826810, rs2738464, rs2738465, rs1433099, and rs2738466). *hr*GFPn fluorescence activity. The fluorescence of 10,000 cells was measured in each experiment and the activity is expressed relative to the mean fluorescence of Type I haplotype as 100%. *hr*GFP mRNA levels. RNA was isolated from HAT resistant cells and the amount of *hr*GFP mRNA relative to β actin was determined by RT-PCR and expressed relative to the mean of type I haplotype as 100 %.

Discussion

In this study we demonstrated that, in a tissue culture based reporter gene assay, common polymorphisms in the 3'UTR sequence of the human *LDLR* gene do not influence the gene expression at the mRNA level. We also demonstrated that the common haplotypes in the 3'UTR contribute to inter-individual variations in plasma total cholesterol, LDL-C, HDL-C, and LDL/HDL ratio in Caucasians.

Since the LDLR plays an important role in plasma lipoprotein homeostasis, genetic variations in the *LDLR* gene could contribute to inter-individual differences in lipoprotein cholesterol levels in the general population. In testing this hypothesis, we focused on common variations in the 3'UTR because it is considered a regulatory region that is important for the appropriate expression of many genes by affecting the mRNA degradation and stability. Within the 5' portion of the 3'UTR of the human *LDLR* gene there are two stretches of DNA that are conserved between rodents and humans in the *LDLR* gene, and each of these stretches contains AUUUA motifs (underlined in Supplementary Figure 1). Although none of the SNP sites overlap directly with the AUUUA sequence, two of the SNP sites, 44332 and 44857, are located within these homology blocks. As seen in table 6, of the five most common polymorphisms in the 3'UTR of the human *LDLR* gene, the expression of the reporter gene in a tissue culture system was not statistically different among the five types.

Our analysis also detected an association between 3'UTR polymorphisms and variation in the lipoprotein cholesterol levels in the general population. Although the proportions of the sample variance explained by the two SNPs are small (see Supplemental Table 2.2 for comparison of the size of the effects with some other factors), our data

suggests that some of the genetic variations in the 3'UTR are important determinants of coronary heart disease risk in Caucasians in the ARIC study. Thus Caucasian men homozygous for the *CA/CA* diplotype had a significantly higher LDL-C and significant lower HDL-C levels. As a result, they also had a 4.5% higher LDL/HDL ratio, an established risk factor for CHD (Panagiotakos et al., 2003). Conversely, Caucasian men that were homozygous for the *TA/TA* diplotype had significantly lower total cholesterol and LDL-C, and a 5.3% lower LDL/HDL ratio. The effects of these *LDLR* polymorphisms in Caucasian women were generally smaller than those detected in men. Particularly noteworthy is the absence of the LDL-C increasing effect of the *CA/CA* genotype in women, while the LDL-C lowering effect of *TA/TA* genotype was still significant.

To our knowledge, there have been no previous reports of any positive or negative associations between *LDLR* polymorphisms and plasma lipid profiles in African American population. In our study, we did not detect any association between the 3'UTR diplotypes and the lipid profiles in the African American ARIC population. Importantly, however, we note that our present study has limited haplotype assignments for African American individuals. Genotyping at the two SNP sites can reasonably identify the most common 3'UTR haplotypes, *GGCGCA*, *AGCACG* and *GGCGTA*, which account for 93% of the 3'UTR sequences of the *LDLR* gene in Caucasians. In contrast, these three haplotypes account for only 64% of the haplotypes in African Americans. In addition, the two SNP sites that we chose to genotype did not distinguish *GGGGCA* and *GACACA* from *GGCGCA*, or *GGGGTA* from *GGCGTA*. Thus, only 43% of the *CA* haplotype is *GGCGCA* in African Americans, and 66% of the *TA* haplotype is *GGCGTA*. This limitation in haplotype assignment is likely to have reduced the power of detecting association between the

diplotypes and plasma lipids in African Americans. The haplotype *GGGGTA* is more prevalent in African Americans (19%) than in Caucasians (5%), and the haplotype *GGGGCA* is only seen in African Americans (9%). These two haplotypes share the *G* allele at 44506. The effect of the G allele at 44506 on the expression of the *LDLR* gene in *vitro* was not different from other types.

Another important consideration is the effect of a long-range linkage disequilibrium observed in this region of chromosome in Caucasians. Previous studies (Knoblauch et al., 2004) estimated the presence of 22 haplotypes from 7 SNPs in 32 kb of the DNA, spanning intron 1 to the 3'UTR among a total of 1054 normolipidemic subjects from 218 families in the German population: three major haplotypes accounting for 71 % of the chromosomes. Their finding that the long range *LDLR h*aplotypes contribute to the genetic variance of LDL (5%), HDL (6%) and LDL/HDL ratio (2%) is thus consistent with the significant association of the 3'UTR haplotypes with these parameters in Caucasians we observed in the current study. In addition, the *Ava*II site in exon 13 and *Pvu*II site in intron 15, the presence and absence respectively of which have been reported previously to be associated with increased LDL-C levels in an Italian population, Hispanic and non-Hispanics, and Brazilian whites, are also in linkage disequilibrium with the 3'UTR polymorphisms (Ahn et al., 1994; Schuster et al; 1990, Knoblauch et al., 2004; Salazar et al., 2000; Humphries et al., 1991). This long range linkage disequilibrium demonstrated that the real causative polymorphism(s) which affect the plasma lipid parameters are not the 3'UTR sequence per se but others that are in linkage disequilibrium with them.

The primary finding of this study is that the *GGCGCA* haplotype is associated with higher levels of total and LDL-C in Caucasians, whereas the *GGCGTA* haplotype is associated with lower levels of cholesterol levels.

In conclusion, our findings show that common genetic variation in the 3'UTR of the *LDLR* are significantly associated with plasma lipid levels in normolipidic Caucasian subjects. Further studies are necessary to elucidate the mechanism underlying this association.

Acknowledgment

The study was supported by a grant HL 42630 to NM. The authors thank Ms Sylvia Hiller and Annette Staton for help in tissue culture, Dr. Larry Arnolds for flowcytometric analyses, and Dr. Hyung-Suk Kim for quantitative mRNA analyses. DNA sequencing was carried out by the Sequencing Core Facility of the University of North Carolina. The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021 and N01-HC-55022. The authors thank the staff and participants of the ARIC study for their important contributions.

Supplemental Materials

ARIC Population : The present study includes 10,602 Caucasians and 3,497 African Americans. The study population characteristics presented in Supplementary Table 1, show that current alcohol consumption was reported more frequently in Caucasians than in African Americans, while more male African American participants reported that they were current smokers in comparison to Caucasian males. A larger proportion of African American women reported that they had reached menopausal status in comparison to Caucasian women, whereas more Caucasian women reported both previous and current gonadal hormone use in comparison to African American women. Mean total cholesterol, and HDL-C concentrations in women of both races were higher than those of the men, but Caucasian men had lower HDL-C values in comparison to African American men. There were no significant differences in the LDL-C levels between African American men and women, but Caucasian men had slightly higher LDL-C concentrations in comparison to Caucasian women. The analyses were carried out stratified by race and gender either adjusting for age, or fully adjusting for age, field center, current alcohol consumption, current smoking, former smoking, use of lipid-lowering medications, use of medications that secondarily affect cholesterol, body mass index (BMI) and waist-to-hip ratio for men. Additional adjustments for menopausal status and gonadal hormone use were made for women.

Supplementary **Table2.7.** Basic characteristics of ARIC population.

	Caucasian		African American	
	Men	Women	Men	Women
Drinking status				
Current drinker	3,419 (69.0%)	3,411 (60.6%)	652 (49.9%)	453 (21.0%)
Former drinker	1,033 (20.9)	805 (14.3%)	359 (27.5%)	419 (19.4)
Current smoker	1,182 (23.8)	1,393 (24.7%)	494 (37.6%)	527 (24.2%)
Former smoker	2,397 (48.3)	1,388 (24.7%)	441 (33.5%)	378 (17.4%)
Cholesterol lowering medication				
Primary	183 (3.7)	184 (3.3%)	17 (1.3%)	32 (1.5%)
Secondary	1,039 (21.1%)	1,312 (23.4%)	331 (25.7%)	764 (35.4%)
Menopausal status				
Menopausal	N/A	3,452 (61.5%)	N/A	1,395 (64.4%)
Hormonal use				
Current hormone usage	N/A	1,165 (21.3%)	N/A	302 (14.6%)
Menopause and current hormone usage	N/A	777 (14.2%)	N/A	239 (11.6%)
Centers				
Forsyth County, NC	1,483 (29.8%)	1,737 (30.8%)	162 (12.3%)	238 (10.9%)
Washington, MD and Jackson, MS	1,756 (35.4%)	1,916 (34.0%)	1,154 (87.7%)	1,943 (89.1%)
Minneapolis, MN	1,728 (34.8%)	1,982 (36.2%)	0	0
Continuous Covariates				
Age	55±5.7	54±5.7	54±6.0	53±5.7
BMI	27.43±4.00	26.61±5.49	27.63±4.83	30.72±6.50
Waist to hip ratio	0.97±0.05	0.89±0.08	0.94±0.06	0.90±0.08
Sport physical activity index	2.67±0.82	2.42±0.77	2.26±0.74	2.10±0.65
Leisure physical activity index	2.42±0.52	2.50±0.54	2.07±0.57	2.07±0.58
Lipid values				
TC (mg/dl)	211±38	218±41	211±44	217±44
LDL-C (mg/dl)	140±35	135±40	138±41	137±42
HDLC (mg/dl)	43±12	58±17	50±17	58±17
LDL/HDL	3.49±1	2.60±1	3.02±1	2.61±1

Values are mean±SD. N/A, not applicable. BMI, body mass index; TC, total cholesterol.

Supplementary **Table 2.8.** Diplotype frequencies and the mean adjusted estimate of cholesterol in individuals with a given diplotype compared to the most common diplotype.

Diplotypes	Frequency N (%)	TC[a] Estimate (*P**)	LDL-C [b] Estimate (*P**)	HDL-C[c] Estimate (*P**)	LDL/HDL[d] Estimate (*P**)
		Caucasian Men			
CA/CA	1,104 (22.2%)	0.85 (0.59)	1.64 (0.27)	-0.98 (0.06)	0.09 (0.09)
CA/CG	1,196 (24.1%)	1.0 (ref)	1.0 (ref)	1.0 (ref)	1.0 (ref)
CA/TA	1,243 (25.0%)	0.13 (0.93)	0.25 (0.86)	*0.98 (0.049)*	-0.09 (0.08)
CG/CG	328 (6.6%)	-1.12 (0.64)	-0.47 (0.83)	-1.09 (0.15)	-0.01 (0.86)
CG/TA	689 (13.9%)	-1.66 (0.37)	-1.51 (0.38)	- 0.09 (0.88)	-0.06(0.36)
TA/TA	402 (8.1%)	*-6.48 (0.003)*	*-6.14 (0.003)*	-0.58 (0.42)	*-0.15 (0.047)*
		Caucasian Women			
CA/CA	1,248 (22.2%)	0.10 (0.95)	0.37 (0.80)	0.05 (0.94)	-0.01 (0.79)
CA/CG	1,487 (26.4%)	1.0 (ref)	1.0 (ref)	1.0 (ref)	1.0 (ref)
CA/TA	1,376 (24.4%)	1.24 (0.41)	1.62 (0.26)	0.51 (0.42)	0.02 (0.68)
CG/CG	330 (5.9%)	1.65 (0.50)	1.68 (0.47)	0.62 (0.55)	0.01 (0.92)
CG/TA	776 (13.8%)	-1.55 (0.38)	-2.25 (0.18)	*1.96 (0.01)*	*-0.13 (0.02)*
TA/TA	411 (7.3%)	-3.90 (0.08)	*-4.68 (0.03)*	0.74 (0.43)	*-0.14 (0.04)*
		African American Men			
CA/CA	124 (9.4%)	2.62 (0.56)	3.28 (0.45)	-0.74 (0.67)	0.12 (0.36)
CA/CG	145 (11.0%)	-1.61 (0.71)	-1.91 (0.64)	3.25 (0.05)	-0.16 (0.22)
CA/TA	413 (31.4%)	1.0 (ref)	1.0 (ref)	1.0 (ref)	1.0 (ref)
CG/CG	27 (2.1%)	-6.26 (0.47)	-10.99 (0.19)	-0.97 (0.77)	-0.03 (0.91)
CG/TA	246 (18.7%)	-1.10 (0.76)	0.55 (0.87)	0.54 (0.70)	-0.06 (0.55)
TA/TA	344 (26.2%)	-2.76 (0.39)	-0.51 (0.87)	-1.83 (0.14)	0.09 (0.41)
		African American Women			
CA/CA	194 (8.9%)	4.42 (0.22)	4.20 (0.23)	-0.48 (0.74)	0.12 (0.21)
CA/CG	197 (9.0%)	*11.45 (0.001)*	*9.16 (0.01)*	0.13 (0.93)	*0.19(0.04)*
CA/TA	678 (31.1%)	1.0 (ref)	1.0 (ref)	1.0 (ref)	1.0 (ref)
CG/CG	64 (2.9%)	11.12 (0.05)	8.33 (0.13)	2.60 (0.25)	0.12 (0.44)
CG/TA	390 (17.9%)	2.41 (0.39)	0.77 (0.78)	0.50 (0.66)	0.01(0.88)
TA/TA	628 (28.8%)	3.59 (0.14)	3.47 (0.14)	-1.10 (0.26)	0.12 (0.07)

***P* value is from a test comparing the mean adjusted estimate of the lipid value in individuals with a given diplotype compared to the reference (individuals with *CA/CG* in Caucasians and with *CA/TA* in African Americans)

[a,] Values were adjusted for age, field center, current smoker, use of medicine to lower cholesterol. In addition, values were adjusted for former smoker and waist to hip ratio for men, and for medicine that secondarily affects cholesterol, menopausal status, current hormone usage, BMI, and waist-to-hip ratio for women.

[b,] Values were adjusted for age, field center, current smoker, use of medicine to lower cholesterol. In addition, values were adjusted for former smoker and leisure physical activity index for men, and for menopausal status, current hormone usage, interaction between menopause and hormone usage, and waist-to-hip ratio for women.

[c,] Values were adjusted for age, current smoker, use of medicine to lower cholesterol, and sport physical activity index. In addition, values were adjusted for field center and former smoke for men, and for current drinker, menopausal status, current hormone usage, interaction between menopause and hormone usage, BMI and waist-to-hip ratio for women.
[d,] Values were adjusted to age, current smoker, current drinker, use of medicine to lower cholesterol, use of medicine that secondarily affects cholesterol, BMI, and waist-to-hip ratio. In addition, values were adjusted for field center for men, and for former smoker, menopausal status, current hormone usage, interaction between menopause and hormone usage for women.

Supplementary **Table 2.9.** Estimate of the proportion of the sample variance explained by C44857T, A44965G, age, smoking and total covariates

	C44857T	A44965G	Age	Smoking	Total Covariates
C, Men					
LDL/HDL ratio	0.28%	0.001%	0.002%	1.22%	8.38%
HDL (mg/dl)	0.21%	0.05%	0.22%	0.69%	13.3%
LDL (mg/dl)	0.27%	0.01%	0.17%	0.27%	1.44%
TC (mg/dl)	0.23%	0.002%	0.3%	0.35%	1.96%
C, Women					
LDL/HDL ratio	0.09%	0.02%	1.33%	2.35%	22.3%
HDL (mg/dl)	0.07%	0.01%	0.001%	1.79%	25.9%
LDL (mg/dl)	0.08%	0.04%	4.12%	0.55%	11.6%
TC (mg/dl)	0.05%	0.03%	6.14%	0.12%	10.5%
AA, Men					
LDL/HDL ratio	0.12%	0.31%	0.002%	0.82%	12.2%
HDL (mg/dl)	0.43%	0.39%	0.02%	0.14%	16.2%
LDL (mg/dl)	0.01%	0.20%	0.10%	3.18%	3.66%
TC (mg/dl)	0.07%	0.05%	0.09%	2.31.%	2.79%
AA,Women					
LDL/HDL ratio	0.34%	0.01%	1.51%	0.79%	5.3%
HDL (mg/dl)	0.17%	0.10%	0.07%	1.33%	15.3%
LDL (mg/dl)	0.4%	0.11%	2.63%	0.001%	6.3%
TC (mg/dl)	0.46%	0.25%	3.01%	0.04%	5.3%

Supplementary Figure **2.3.** Nucleotide sequence of the 5' portion of *LDLR* 3'UTR

44156 agacagatgg tcagtctgga ggatgacgtg gcgTGAacat ctgcctggag tcccgcccct
44216 gcccagaacc cttcctgaga cctcgccGgc cttgttttat tcaaagacag agaagaccaa
44276 agcattgcct gccagagctt tgttttatat atttattcat ctgggaggca gaacaGgctt
44336 cggacagtgc ccatgcaatg gcttgggttg ggattttggt ttcttccttt cctgctgaag
44396 gataagagaa acaggcccgg ggggaccagg atgacacctc catttctctc caggaagttt
44456 tgagtttctc tccaccgtga cacaatcctc aaacatggaa gatgaaaggg Caggggatgt
44516 caggcccaga gaagcaagtg gctttcaaca cacaacagca gatggcacca acgggacccc
44576 ctggccctgc ctcatccacc aatctctaag ccaaacccct aaactcagga gtcaacgtgt
44636 ttacctcttc tatgcaagcc ttgctagaca gccaggttag cctttgccct gtcacccccG
44696 aatcatgacc cacccagtgt ctttcgaggt gggtttgtac cttccttaag ccaggaaagg
44756 gattcatggc gtcggaaatg atctggctga atccgtggtg gcaccgagac caaactcatt
44816 caccaaatga tgccacttcc cagaggcaga gcctgagtca cCggtcaccc ttaatattta
44876 ttaagtgcct gagacacccg gttaccttgg ccgtgaggac acgtggcctg cacccaggtg
44936 tggctgtcag gacaccagcc tggtgcccAt cctcccgacc cctacccact tccattcccg
44996 tggtctcctt gcactttctc agttcagagt tgtacactgt gtacatttgg catttgtgtt
45056 attattttgc actgttttct gtcgtgtgtg ttgggatggg atcccaggcc agggaaagcc
45116 cgtgtcaatg aatgccgggg acagagaggg gcaggttgac cgggacttca aagccgtgat
45176 cgtgaatatc gagaactgcc attgtcgtct ttatgtccgc ccacctagtg cttccacttc

The stop codon (TGA) is capitalized. Nucleotide positions are numbered according to the *LDLR* gene sequence in the GenBank database (NC_000019). Major nucleotides at the six SNP sites are shown in blue capital letters. Underlines indicate the regions highly conserved between humans and rodents. ARE sequences are shown in green letters, a stem loop is shown in pink letters, and the positions corresponding to the PCR primers used to amplify the 1kb DNA are shown in red letters.

Chapter III

Lack of Evidence for Gene-Gene Interaction Between the Human Low Density Lipoprotein Receptor and Apolipoprotein E Genotypes in The Atherosclerosis Risk in Communities (ARIC) study.

Hind Muallem[1], Kari E North[2], Masao Kakoki[1], Xia Li[2],

Megan Grove[3], Eric Boerwinkle[3], Kirk C. Wilhelmsen[4], Gerardo Heiss[2],

and Nobuyo Maeda[1, *]

Abstract

The low density lipoprotein receptor (LDLR) with its ligand apolipoprotein E (apoE) plays an essential role in cholesterol homeostasis. Genetic variations in the human low density lipoprotein gene (*LDLR*) and *APOE* isoforms have been shown to associate with the levels of plasma low density lipoprotein cholesterol (LDL-C). However, the gene-gene interaction between *LDLR* and *APOE* in determining the risk of atherosclerosis in humans has not been interrogated

A nucleotide change C44857T in the 5' region of the 3'UTR of the human *LDLR* gene was genotyped in the Atherosclerosis Risk in Communities (ARIC) study. Genotyping two SNPs in exon 4 of the human *APOE* gene, T3937C, and C4075T, in ARIC population resulted in three isoforms, *APOE*2, APOE*3,* and *APOE*4*. Our results showed that there is no interaction between between *LDLR* and *APOE* genotypes in ARIC population. Thus *LDLR* genotype effect and *APOE* genotype effect are independent.

Introduction

LDLR with its ligands apoE and apoB play a pivotal role in cholesterol homeostasis (Goldstein 1995; Hobbs et al. 1990). Mutations in the *LDLR* gene which lead to a decreased expression of *LDLR* cause familial hypercholesterolemia (FH), a phenomena characterized by an accumulation of cholesterol in the circulation and premature coronary heart disease (CHD). Moreover, single nucleotide polymorphisms in the human *LDLR* gene cause a quantitative change of *LDLR* expression that are associated with either an increased or a decreased level of LDL-C and an increased or a decreased risk of atherosclerosis.

Hypercholesterolemia is a complex and multifactorial disease, and the effect of one genetic variation may be interact with the effect of other genetic variations in other genes as well as gene environment interaction (Wang and Zhao 2003). Association studies in the *LDLR* gene and the risk of CHD have been well studied, but the interaction of *LDLR* gene with other genes, mainly those that are involved in lipid metabolism, has been under estimated.

ApoE is a glycoprotein that plays a key role in the clearance of plasma chylomicrons, very low density lipoproteins (VLDL), and high density lipoproteins (HDL) via its interaction with the low density lipoprotein receptor (LDLR) (Corbo and Scacchi 1999; Hussain et al. 1999; Melman et al. 2002b). In the population there are three common isoforms, *APOE*2*, *APOE*3* and *APOE*4*, with allele frequencies of 0.08, 0.77, and 0.15 respectively (Banares et al. 2005; Boerwinkle and Utermann 1988a; Weisgraber 1990) These isoforms differ in the amino acids at position 112 and 158. The most frequent isoform, *APOE*3*, has cysteine and arginine at these positions, *APOE*2* has two cysteines and *APOE*4* has two arginines (Lund-Katz et al. 2001; Weisgraber 1990).

The isoforms have different binding affinities for LDLR; apoE4 has a higher binding affinity than apoE3, while apoE2 binds very poorly (Lund-Katz et al. 2001; Saito et al. 2003; Saito H 2003). Despite its poor receptor binding however, individuals with apoE2 have a reduced level of cholesterol and a reduced risk of atherosclerosis, except for 5%–10 % of apoE2 homozygous who develop Type III hyperlipoproteinemia exhibiting, very high levels of plasma cholesterol and triglycerides (TG).(Kypreos et al. 2003.). In contrast, individuals with apoE4 have significantly increased plasma LDL-C levels and an increased risk of atherosclerosis compared to those without (Banares et al. 2005; Boerwinkle et al. 1987; Bouthillier et al. 1983; Malloy I 2004)

This paradox in binding affinity and atherosclerosis risk has been attributed to the high affinity of apoE4 for the receptor resulting in rapid clearance of chylomicrons and VLDL remnants causing a down regulation of the *LDLR*, while the low affinity of apoE2 results in slow clearance of chylomicrons and VLDL remnants and therefore, an up regulation of the *LDLR* (Davignon et al. 1988). The increase in *LDLR* expression ameliorated the hyperlipidemia and atherosclerosis of the mice that were homozygous for the human *APOE*2* isoforms.(Knouff et al. 2001).

On the other hand, Malloy et al showed that, mice fed a western-type diet, and carry the human *APOE*4,* with an increased expression of *LDLR*, developed severe atherosclerosis due to the accumulation of cholesterol in the plasma (Malloy I 2004). In contrast, there was no harmful effect on mice carrying human *APOE*3*. This finding led Malloy and his group to postulate an "ApoE Trapping" hypothesis which states that, apo E4 which has a high affinity for the receptor is trapped by it, and hence render it unavailable for the newly formed lipoproteins resulting in an increase in cholesterol levels in plasma.

Above observations in mice suggest that there may be gene-gene interaction between *LDLR* and *APOE* in humans affecting atherosclerosis risk.

A previous study by Muallem et al showed that the mean LDL-C in C/C rs1433099 (C44857) homozygous Caucasians male and female was 7.0 and 4.0 mg/dl higher than T/T (44857T) homozygous. Also in C/C homozygous African American female LDL-C was 3 0 mg/dl higher compared to T/T homozygous.(Muallem et al. 2007). Since there were no functional differences between these 3'UTR sequences, we concluded that C (44857) is linked to variations in the *LDLR* gene and neighboring genes which causes plasma lipoprotein profiles associated with high risk for CHD.

In the present paper, we studied the interaction between a nucleotide change rs1433099, C44857T in the 3'UTR of the human *LDLR* gene and the *APOE* isoforms and their effect on lipid levels in the Atherosclerosis Risk in Communities study.

Material and Methods

ARIC study population

The ARIC study is a prospective epidemiologic study whose main target is to investigate atherosclerotic vascular diseases and their risk factors (The ARIC (investigators 1989). Blood samples were drawn for the analysis of lipids and lipoproteins and DNA was extracted for genotyping (Brown et al. 1993). The present study includes 13,838 Caucasians and 3,481 African Americans. Fasting plasma lipid data at the first cohort examination visit (1987-1989) was used for this analysis. The genotyping strategy for the *LDLR* and *APOE* polymorphisms have been previously described (Muallem et al. 2007; Sturgeon et al. 2005).

Statistical analysis

All analyses were stratified by race. Deviations from Hardy-Weinberg equilibrium expectations were tested using a goodness-of-fit test. Multiple linear regression analysis was conducted to evaluate the mean estimated values of TC, LDL-C, HDL-C and LDL/HDL ratio in groups of individuals with different *LDLR* and *APOE* genotypes. Individuals with the *LDLR C44875T* C/T genotype and *APOE* E3 isoform were considered the referent group. The Model evaluated interaction effect between *LDLR C44875T* genotypes and *APOE* genotypes on these lipid traits, on a multiplicative scale, using multiple partial F test for overall effect of interaction terms, with individuals carrying *LDLR C44875T* C/T and *APOE* E3 genotype as a reference group. All analyses were performed separately in Caucasians and African Americans, using SAS software, version 9.1.3.

Results

For the purpose of this analysis, *LDLR* and *APO E* genotypes were grouped into three categories: *LDLR* C/C, C/T and T/T, C/T the reference group, and *APOE*2* (2/2, 2/3, 2/4), *APOE*3* (3/3) the reference group and *APOE*4* (3/4 and 4/4). As shown in table 1 the frequency distribution of the combination of *LDLR* genotypes, C/C, T/T and *APOE* genotypes, E2, E4 are different in African Americans and Caucasians. Frequencies for *LDLR* C/C: *APOE*2*, and *LDLR* C/C: *APOE*4* are higher in African Americans than in Caucasians (10.1%, 6.1%, vs 1.9%, 1.2%), while Caucasians have higher frequencies for *LDLR* T/T: *APOE*2* and *LDLR* T/T: *APOE*4* compared to African Americans (13.6%, 8.5% vs. 7.2%, 4.0%) Table 3.1.

As for the interaction term, no evidence of significant interaction was seen between *LDLR* and *APOE* genotypes among Caucasians and African Americans in ARIC population. Tables 3.2-3.9.

Table 3.1. Frequency of *LDLR* and *APOE* interaction stratified by race.

Caucasians N=13,838	*APOE*	*LDLR*		
		CC	CT	TT
	E2	193 (1.9%)		1406 (13.6%)
	E3		7752 (74.9%)	
	E4	126 (1.2)		880 (8.5%)
African Americans N= 3,481				
	E2	350 (10%)		
	E3		2528 (72.6%)	
	E4	212 (6 %)		139 (4%)

CT and E3 are the reference groups

Table 3.2. Interaction effect between *C44875T* polymorphisms in *LDLR* and *APOE* polymorphisms on plasma lipoproteins in Caucasians

APOE	Total cholesterol*		
	LDLR C44875T		
	C/C	C/T	T/T
ϵ2	-1.56 ± 2.3 (P=0.498)	-11.28 ± 1.76 (P<0.001)	-1.84 ± 1.92 (P0.339)
ϵ3	1.01 ± 1.05 (P=0.338)	0	-6.16 ± 1.95 (P=0.002)
ϵ4	2.44 ± 4.28 (P=0.569)	6.87 ± 1.47 (P<0.001)	2.7 ± 3.63 (P=0.458)
*P**	*P* = 0.6191		

* Unit of measurement is mg/dl

Table 3 3. Interaction effect between *C44875T* polymorphisms in *LDLR* and *APOE* polymorphisms on plasma lipoproteins in Caucasians

APOE	LDL-C*		
	LDLR C44875T		
	C/C	C/T	T/T
ϵ2	-3.87 ± 2.16 (P=0.074)	-13.61 ± 1.66 (P<0.001)	-2.65 ± 1.81 (P=0.143)
ϵ3	1.78 ± 0.98 (P=0.071)	0	-6.04 ± 1.83 (P=0.001)
ϵ4	0.23 ± 4.04 (P=0.954)	7 ± 1.38 (P<0.001)	2.35 ± 3.42 (P=0.492)
*P**	*P* = 0.2094		

* Unit of measurement is mg/dl

Table 3 4. Interaction effect between *C44875T* polymorphisms in *LDLR* and *APOE* polymorphisms on plasma lipoproteins in Caucasians

APOE	HDL*		
	LDLR C44875T		
	C/C	C/T	T/T
ϵ2	1.35 ± 0.88 (P=0.128)	0.02 ± 0.68 (P=0.972)	0.25 ± 0.74 (P=0.738)
ϵ3	-1.27 ± 0.4 (P=0.002)	0	-0.68 ± 0.75 (P=0.361)
ϵ4	0.36 ± 1.64 (P=0.826)	-1.34 ± 0.56 (P=0.018)	-0.57 ± 1.39 (P=0.683)
*P**	*P* = 0.6216		

* Unit of measurement is mg/dl

Table 3 5. Interaction effect between *C44875T* polymorphisms in *LDLR* and *APOE* polymorphisms on plasma lipoproteins in Caucasians

APOE	LDL/HDL		
	LDLR C44875T		
	C/C	C/T	T/T
ϵ2	-0.21 ± 0.07 (P=0.005)	-0.25 ± 0.06 (P<0.001)	-0.06 ± 0.06 (P=0.298)
ϵ3	0.13 ± 0.03 (P<0.001)	0	-0.08 ± 0.06 (P=0.21)
ϵ4	-0.11 ± 0.14 (P=0.423)	0.22 ± 0.05 (P<0.001)	0.06 ± 0.12 (P=0.614)
*P**	*P* = 0.0638		

* Unit of measurement is mg/dl

Table 3 6. Interaction effect between *C44875T* polymorphisms in *LDLR* and *APOE* polymorphisms on plasma lipoproteins in African Americans

APOE	Total cholesterol*		
	LDLR C44875T		
	C/C	C/T	T/T
ϵ2	1.45 ± 5.23 (P=0.782)	-13.04 ± 2.86 (P<0.001)	0.59 ± 4.3 (P=0.892)
ϵ3	4.32 ± 2.79 (P=0.121)	0	1.34 ± 2.61 (P=0.608)
ϵ4	1.57 ± 4.62 (P=0.734)	8.63 ± 2.38 (P<0.001)	-1.72 ± 3.93 (P=0.662)
*P**	*P* =0.9653		

* Unit of measurement is mg/dl

Table 3 7. Interaction effect between *C44875T* polymorphisms in *LDLR* and *APOE* polymorphisms on plasma lipoproteins in African Americans

APOE	LDL-C*		
	LDLR C44875T		
	C/C	C/T	T/T
ϵ2	2.58 ± 4.98 (P=0.605)	-15.39 ± 2.71 (P<0.001)	3.65 ± 4.09 (P=0.373)
ϵ3	2.11 ± 2.66 (P=0.427)	0	1.62 ± 2.48 (P=0.514)
ϵ4	0.08 ± 4.4 (P=0.985)	8.16 ± 2.26 (P<0.001)	1.07 ± 3.74 (P=0.775)
*P**	*P* = 0.9278		

* Unit of measurement is mg/dl

Table 3 8. Interaction effect between *C44875T* polymorphisms in *LDLR* and *APOE* polymorphisms on plasma lipoproteins in African Americans

APOE	HDL*		
	LDLR C44875T		
	C/C	C/T	T/T
ϵ2	-2.1 ± 2.08 (P=0.313)	1.27 ± 1.14 (P=0.264)	-2.05 ± 1.71 (P=0.232)
ϵ3	1.46 ± 1.11 (P=0.188)	0	-2.06 ± 1.04 (P=0.047)
ϵ4	1.79 ± 1.84 (P=0.33)	-1.36 ± 0.95 (P=0.151)	-0.2 ± 1.56 (P=0.9)
*P**	*P* = 0.424		

* Unit of measurement is mg/dl

Table 3.9. Interaction effect between *C44875T* polymorphisms in *LDLR* and *APOE* polymorphisms on plasma lipoproteins in African Americans

APOE	LDL/HDL		
	LDLR C44875T		
	C/C	C/T	T/T
ϵ2	0.08 ± 0.15 (P=0.584)	-0.3 ± 0.08 (P<0.001)	0.15 ± 0.12 (P=0.2)
ϵ3	0.02 ± 0.08 (P=0.769)	0	0.12 ± 0.07 (P=0.113)
ϵ4	-0.11 ± 0.13 (P=0.376)	0.2 ± 0.07 (P=0.003)	0.09 ± 0.11 (P=0.431)
*P**	*P* = 0.4442		

* Unit of measurement is mg/dl

*Values were adjusted for age, Weld center, current drinker, current smoker, use of medicine to lower cholesterol, use of medicine that secondarily affect cholesterol, BMI and waist to hip ratio. In addition, values were adjusted for formal smokers, menopausal status and current hormone usage for women.

***P* value is from the multiple partial F test for overall effect of interaction between C44857T polymorphism in *LDLR* and *APOE* polymorphisms on total cholesterol, LDL-C, HDL-C and LDL/HDL.

Discussion

The effect of gene-gene interaction between *LDLR* and *APOE* on lipid profile was studied by Pederson and Berg in 1990 and Ahn et al in 1994 (Ahn et al. 1994; Pedersen and Berg 1989, 1990). Pederson and Berg studied the interaction between *Pvu* II polymorphism in intron 15 of the human *LDLR* and *APOE* genotypes in 200 unrelated Norwegian subjects. They found that individuals having the *APOE4* and carrying the *Pvu*.II positive allele had lower total cholesterol than those individuals having *APOE4* and carrying the *Pvu* II negative allele linked to TT. Hence, the increasing effect of *APOE4* on cholesterol was diminished by the presence of *Pvu* II positive allele in this study. On the other hand, Ahn et al studied the interaction of two common polymorhisms, *Ava* II (exon 13) and *Nco* I (exon 18) with *APOE* genotypes in 385 normolipidic Hispanic and 543 non-Hispanic whites, from Colorado. Unlike Pederson and Berge, Ann et al did not find an interaction between *LDLR* and *APOE* genotypes. The limitation of the two studies is the small sample size. Our aim in this paper is to study the interaction with a large sample size that has enough power to detect gene-gene interaction.

Gene-gene interaction between *LDLR* and *APOE* was detected in Alzheimer's disease (AD) (Cheng et al. 2005). In this study authors studied two nucleotide changes in exon 8 rs11669576 (A24441G) and exon 13 rs5925 (C33078T) in the human *LDLR* gene for functional interaction with *APO*E4* isoform in determining the risk of AD. They found that individuals with *APOE*4* and who are homozygous for *LDLR* G/G and C/C at the same two SNP sites have a nine fold increase in the risk for AD compared to individuals with other combination of *LDLR* genotypes.

Malloy et al showed that in mice, there is an interaction between *LDLR* and *APOE* genotypes governed by the level of expression of *LDLR*, *APOE* genotypes and western type diet as mentioned in the introduction. Mice carrying the human *APOE4* and having an increased expression of *LDLR* developed sever atherosclerosis. In contrast, the increased expression of *LDLR* in mice having *APOE2* were protected from the hyperlipidemia and atherosclerosis.

Our analysis using the ARIC population showed that there is no statistically significant interaction between *LDLR* genotypes and *APOE* isoforms, suggesting that the *APOE* genotype effect and *LDLR* genotype effect on plasma lipid levels are independent.

This lack of interaction has to be interpreted with caution. Even though the strength of this study is the large sample size of both Caucasians (13,838) and African Americans (3,481), when the samples are divided among *LDLR* genotypes and *APOE* isoforms, sample size becomes small in some groups. This reduces statistical power and might limit the identification of gene-gene interaction (Moore and Williams 2002).

Another consideration is the fact that the C44857T SNP studied here is not a causative SNP, but it is in linkage disequilibrium with other causative SNPs in the gene. Since we do not know the nature of the causative SNPs, and the extent of linkage disequilibrium with the C44857T, the small effect associated with this SNP may not be sufficient to reveal the interaction with the strong effects of apoE polymorphisms.

It is, therefore, necessary to identify functional *LDLR* gene polymorphisms that influence the basal level of *LDLR* expression to test the interaction between the *LDLR* expression and apoE polymorphisms observed in mice.

Chapter IV

Enhancer activity in the first intron of the human low density lipoprotein receptor gene.

Hind Muallem[1] and Nobuyo Maeda[1]

Abstract

The low density lipoprotein receptor plays an important role in cholesterol homeostasis. Enhancer activities that affect their transcriptional control within the promoter region and that located 13 kb downstream of the human low density lipoprotein (*LDLR*) gene have been well characterized, but, regulatory elements residing within the gene have not been studied. We have identified transcriptional enhancer elements within a 100 bp fragment in the first intron of *LDLR* gene. The 100 bp fragment enhances the luciferase activity of a reporter gene when inserted into pGL3-promoter vector by 3.0 fold. The enhancement is orientation independent. However, neither the fragment containing 5' half of the 100 bp, or that containing only 3' half of the 100 bp were able to confer the enhancer activity. A sequence motif in the 5' half is similar to the Sp1 binding site. Supershift with Sp1 polyclonal antibody suggests that Sp1 is necessary but not sufficient for the enhancing activity. A sequence motif AGCTATCCCC in the 3' half is a potential GATA protein binding site. However, co-transfection with GATA expression vectors did not increase the enhancing activity suggesting that the GATA proteins are not involved in controlling transcription. Thus other factors synergistically enhance the transcription of the *LDLR* gene.

Introduction

The low density lipoprotein receptor (LDLR) is a cell surface receptor that plays a pivotal role in lipoprotein metabolism. Mutations in the *LDLR* gene cause familial hypercholesterolemia which is characterized by an increase in low density lipoprotein cholesterol (LDL-C) and premature coronary heart disease (CHD). Transcription of the *LDLR* gene is regulated by intracellular level of cholesterol (Smith et al. 1990). When intracellular level of cholesterol is low, transcription of *LDLR* is increased and its amount on the cell surface is also increased to actively bind and internalize LDL-C. On the other hand, when cholesterol level within cells is high, transcription is diminished and the uptake of LDL-C is suppressed (Dawson et al. 1988). The basal and cholesterol mediated transcription of *LDLR* depend on elements within 177 base pairs in the promoter region which contains two TATA boxes and three GC rich imperfect repeats of 16 bp each (Sudhof et al. 1987). Repeat 1 and 3 contain binding sites for Sp1 transcription factor that are responsible for the basal level of *LDLR* expression. Repeat 3 contains a sterol regulatory element binding site (SRE-1), where sterol regulatory element binding protein (SREBP) binds and enhances transcription (Briggs et al. 1993). Nuclear localization of SREBP is regulated by another protein called steroid regulatory element binding protein cleavage activating protein (SCAP). SCAP has a sterol sensing domain (SSD), and when cytoplasmic intracellular sterol content is low, SCAP activates site -1 protease (SIP) which cleaves SREBP at one site. The cleaved product is then cleaved by site-2 protease (S2P) and the active N-terminal domain of SREBP migrates to the nucleus to initiate transcription (Hua et al. 1993; Wang et al. 1994).

In addition to the enhancer activity in the promoter of *LDLR*, Wang et al identified a DNA sequence highly conserved among primates approximately 13 kb downstream of the gene. This PS2 sequence contains SRE binding motif TCACCCCAC which binds SREBP-2. The PS2 sequence enhances luciferase gene expression driven by the proximal human *LDLR* promoter. The activity of the test vector containing *LDLR* promoter/SREBP-2/PS2 was 5 fold higher than the vector containing *LDLR* promoter

The 5'region of the *LDLR* gene is rich in G/C nucleotides. The G/C rich region extends well beyond exon I into the 5'portion of intron I. Many genes with a similar sequence feature are known to contain enhancer sequences in their first intron. For example, the human β –enolase gene, Adolase β gene, adiponectin gene, *HPRT* gene, β actin, smooth muscle actin, and adenosine deaminase gene (Aronow et al. 1989; DeLong and Smith 2005; Frederickson et al. 1989; Gregori et al. 1998; K Satyamoorthy et al. 1977; Mack and Owens 1999; Qiao et al. 2005; Reid et al. 1990). We therefore hypothesized that the G/C rich region of intron I of the human *LDLR* gene may contain transcriptional enhancer activity. In the present paper we identified an enhancer activity in a 100 bp fragment containing a binding site for the transcription factor Sp1. Insertion of the 100 bp into a pGL3-promoter vector led to a 3.0 fold increase in luciferase activity in human embryonic kidney 293 cells.

Material and Methods

Plasmid Construction

A phage clone containing approximately 10kb of DNA spanning part of the promoter, exon I and two third of intron I of the human *LDLR* gene was obtained by screening the lamda phage library (Kim et al. 1990). The phage DNA was digested with *Eco*R1 restriction enzyme to generate three fragments, Fa1, 2448 bp (859-3307), Fa2, 1668 bp (3308-4976) and F a3, 4872 bp (4977-9849). The nucleotide positions are according to GenBank accession number AY324609 as shown in Figure 4.1.

Each fragment was cloned into *Eco*R1 site in pBluescript II KS (-) plasmid vector (Stratagene). The correctness of the sequence was verified by sequencing using specific forward and reverse pBluescript II KS (-) plasmid primers, M13F3 and M13R3. PGL3- promoter vector (Promega) contains a multiple cloning sites, followed by SV40 promoter, cDNA of firefly luciferase gene (driven by SV40 promoter) and ampicilline resistant gene β –lactamase. The *Xho*1 site within the cloning site was replaced with *Eco*R1 site into which these *Eco*R1 fragments were inserted. pR-Tk vector containing cDNA of Renilla luciferase (Promega) was used to control for transfection efficiency.

Amplification of fragments Fa5, Fa8 and Fa11

DNA fragment, Fa5, that contains the first 823 bp (2489-3312) of intron I, was generated by PCR in a 25 µl reaction mixture using Phusion High-Fidelity DNA Polymerase kit (BioLabs, New England). The forward and reverse primers, 5'-GAA TTCGTAAGGCTTGCTCCAGGCGC-3' and 5'-GGTGGCCTGTTCGGACTACA-3', introduced *Eco*R1 sites at both the 5' and 3' ends of the PCR product. Human genomic

DNA (Maeda 1991) was denatured at 98ºC for 30 seconds, followed by annealing /elongation at 72ºC for 1.0 minute and 30 cycles. Fa8, a 100 bp (3730-3830), spanning a *Bgl*2 site within Fa2 fragment was generated using Fast Start High Fidelity PCR System (Roche, Indianapolis, IN, USA). The forward and reverse primers were, 5'-GAATTCTTGAAACGTTGCCTCAGAAT-3' and 5'-GAATTCACCCCAGGGGGTCGGTCCCG-3'. The 25µl PCR reaction containing human genomic DNA was denatured at 95ºC for 2.0 minutes followed by annealing /elongation at 72ºC for 1.0 minutes and 35 cycles. Fa11, a 245 bp (3211–3455) 5' of the *Bgl*2 site was generated using Fast Start High Fidelity PCR System using the forward and reverse primers: 5- GAATTCATCCTGAGTTGGGTGCAGTC -3, AND 5'-GAATTCACAAGGTGAGGCTCAGACTT-3'. Amplification conditions are the same as Fa8. The PCR products were run on 0.8% agarose gel, purified using QIA quick gel extraction kit 250 (Qiagen Inc, Valencia, CA), digested with *Eco*R1 and sequenced with the PCR primers to verify the sequences. The purified PCR was cloned into the pGL3-promoter and used to test enhancer activities.

Two clones that contained Fa9, (1169 bp, 3807-4976) and Fa10, (778 bp, 4198-4976) were obtained by nested deletions with exonuclease III of the clone containing Fa2 in the forward direction using Erase a base system (Promega) according to the manufacturer's protocol.

Transient Transfection and Luciferase Activity

Human embryonic kidney cells, 293A, were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 2mM L-glutamine, and

penicillin - streptomycin. Cells were grown to 90% confluency in 6mm plates and transiently transfected with a mixture containing 10 µl Lipofectamine 2000 (Invitrogen,Carlsbad, CA) and approximately 0.8 µg DNA composed of 20:1 molar ratio of a pGL3-promoter test vector to pRL-TK control vector (Promega, Madison, WI). Cells were incubated in above medium at 37ºC and 5 % CO2 for 24 hours. Cells were harvested and lysed in 300 µl of DMEM medium for luciferase assay. Luciferase activity was determined by using the Dual- Glo luciferase Reporter Assay System (Promega). For Firefly luciferase activity, 75 µl of the lysate was added to 75 µl of Dual-Glo luciferase reagent, mixed on a shaker for half an hour and measured using Orion Microplate Luminometer. Subsequently, Renilla luciferase activity (control vector) was measured as above, 75 µl of Dual-Glo.Stop& Glo reagent was added to150 µl of the above mixture. The relative activity of the test vector was determined by dividing firefly luciferase activity by Renilla luciferase activity. Each test vector was assayed in triplicate of transfection per experiment and the experiment was repeated at least twice. Co-transfection with the GATA's expression vectors (kindly provided by Dr. T Kodoma at Tokyo U Japan) was carried at 1:1 molar ratio (approximately 0.7µg of each vector DNA) with the test plasmid that contained Fa8 fragment

Elecrtophoretic Mobility shift Assay

A100 bp DNA fragment of Fa8 was cut from agarose, and labeled with [ά- P^{32}] dCTP using Klenow fragment of polymerase enzyme (Roche, Indianapolis IN) and 6bp random oligonucleotides. Nuclear protein was extracted from 293 cells following the manufacturer's protocol (Active Motif, Carlsbad, CA). Protein concentration was

determined by BCA protein Assay Kit, (Pierce, Boston, MA). For electrophoretic mobility shift assay (EMSA), 7.0 µg of nuclear extract was incubated at room temperature for 30 minutes with 20 µl reaction mixture containing gel shift binding buffer (5 mM MgCl2, 2.5 mM EDTA, 2.5 mM DTT, 250 mM NaCl, 50 mM Tris-HCl, (pH 7.5) and 0.25 µg/µl poly (dI–dC) in 20 % glycerol (Promega). 25,000 cpm of labeled Fa8 was added to the mixture and incubation proceeded for another half an hour. The whole reaction mixture was loaded on a 4% non-denaturing polyacrylamide gel and electrophoresed for one hour and a half at 170 volts in 0.25 x TEB buffer (22mM Tris-HCl, 22mM boric acid and 0.5mM EDTA). Gels were dried and the image was analyzed by autoradiography

For competition assay, 200 molar excess of cold competitive or non competitive DNA fragments were added to the above EMSA reaction mixture. For supershift assay, 4 µg of antibodies against human Sp1, GATA1, GATA2, GATA3, GATA4 and GATA6 (Santa Cruz Biotechnology, CA) were added to the reaction mixture containing labeled DNA, binding buffer and nuclear protein (Mack and Owens 1999).

Results

Identification of a DNA sequence with a transcriptional enhancer activity in intron I of the human *LDLR* gene

To identify the regulatory sequence (s) in intron I, a phage clone containing part of the promoter, exon I and two third of intron I was digested with *Eco*R1 to produce Fa1, Fa2 and Fa3 as shown in Figure 4.1.

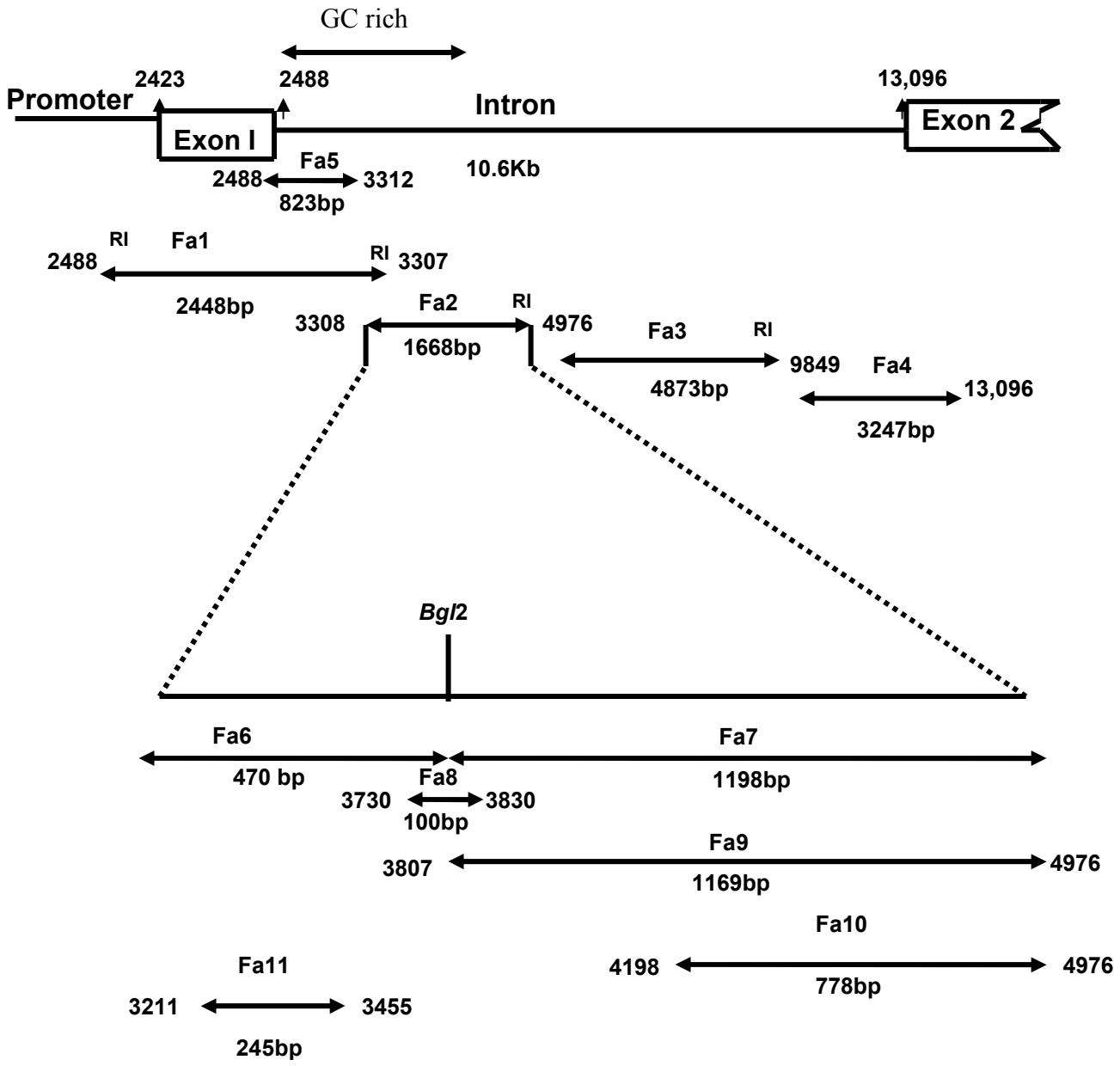

Figure 4.1. A schematic representation of the 5'portion of the human *LDLR* and DNA fragments studied. Fa1, Fa2, and Fa3, were generated by digestion of a phage clone with *Eco*R1. Nucleotide position was indicated according to Accession number AY324609. Double arrowed lines indicate fragments used to generate test vectors. Fa4 was not examined.

DNA fragments, Fa1, Fa2, and Fa3, were cloned separately into pGL3-promoter vector upstream of the luciferase gene. To test their enhancer activities, test plasmid DNAs were co- transected into 293 cells with pRL-TK plasmid as a transfection control. Test vector containing fragment Fa1 gave a slightly higher luciferase activity than the pGL3 vector (1.6±0.12 vs 1.0±0.24, $p<0.01$) (Figure 4.2). This activity was probably due to the promoter region, since a test vector with PCR fragment, Fa5, which contains the beginning of intron I but not promoter and exon I showed a luciferase activity that was not statistically different from the control vector pGL3 (Figure 4.3). In a marked contrast, test vector with Fa2, showed 3.7 fold higher activity than the pGL3 vector (3.7±0.180 vs 1.0±0.024, $p<0.001$) (Figure 4.2). The luciferase activity of the lysates of the cells transfected with the test vector containing Fa3 was not statistically significant from the pGL3 vector.

The enhancer element, by definition, should enhance transcription of nearby genes in an orientation independent fashion. To test whether enhancement by the Fa2 fragment is orientation independent, the 1.6 kb *Eco*R1 fragment was inserted into PGL3 vector in a reverse orientation, Ra2. The luciferase activity in the cells transfected with Ra2 test vector was also significantly enhanced and indistinguishable from the activity in the cells transfected with Fa2. Thus 1.6 kb *Eco*R1 fragment in intron I of the *LDLR* gene contains an enhancer element (s).

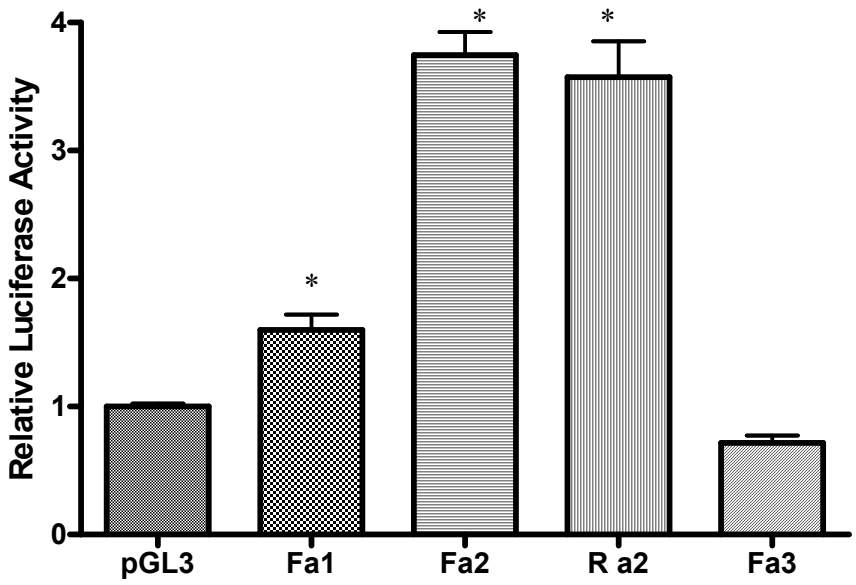

Figure 4.2. Analysis of enhancer activity of the DNA fragments, Fa1, Fa2, Ra2, (Fa2 in reverse direction) and Fa3, from intron I of *LDLR*. The results were expressed as relative to the luciferase activity of cells transfected with a control vector PGL3 as 1. Transfcetion was carried in triplicate in each experiment, and the experiment was repeated three times. Bars indicate means ±SE from these experiments. * indicates the significant difference from the activity of PGL3, $p<0.01$.

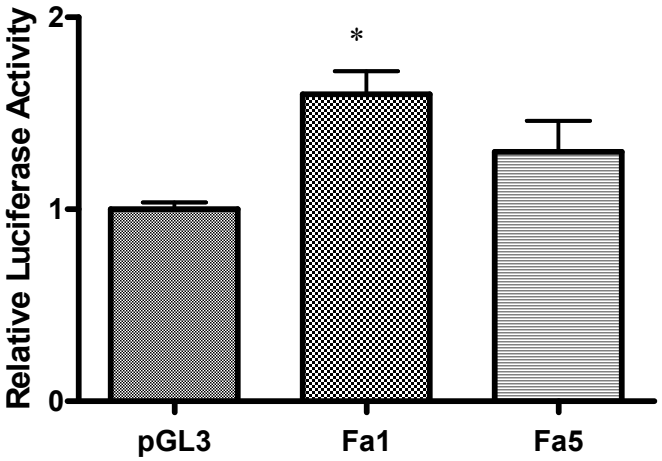

Figure 4.3. Analysis of enhancer activity in the first part of intron I. Fa5 is Fa1 excluding the promoter and exon I. Transfcetion was carried in triplicate in each experiment, and the experiment was repeated three times. Bars indicate means ±SE from these experiments. * $P<0.01$ compared to PGL3.

To further localize the position of elements responsible for the enhancer activity within Fa2, we digested the fragment into two, Fa6, (470 bp, 3308-3777) and Fa7, (1198, 3778-4976) by cutting it at the unique *Bgl*2 restriction site (Figure 4.1), and cloned them into *Eco*R1/*Bgl*2 sites of the PGL3 vector. Transfection into 293 cells revealed that neither Fa6 nor Fa9 conferred enhancer activity (Figure 4.4), Luciferase activity in the cells transfected with Fa6 test vector was the same as with the control PGL3 activity; while the activity of the Fa7 vector was significantly reduced (0.32±0.04 vs1.00±0.04, $p<0.01$). We tested two independent clones of the test vectors, and the results were the same. Thus cutting at the *Bgl*2 site led to the loss of enhancer activity of Fa2.

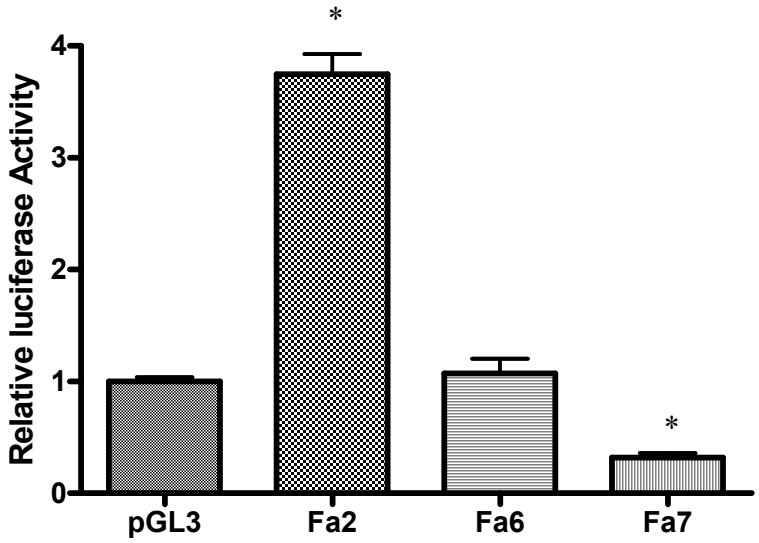

Figure 4.4. Loss of enhancer activity of Fa2 by cutting at the *Bgl*2. Fa6 is 470 bp of the 5'portion and Fa7 is 1198bp of the 3'portion of fragment Fa2. Transfcetion was carried out in triplicate in each experiment, and the experiment was repeated three times. Bars indicate means ±SE from these experiments. * $P<0.01$ compared to PGL3.

We also tested the activity of vectors containing sequences Fa9, (1169 bp), and Fa10 (778 bp), that are 3'parts of the Fa7 (Figure 4.1). The luciferase activity of cells transfected with Fa10 was slightly higher than that of cells transfected with PGL3 (1.2±0.1 vs 1.0 ±0.04 $p<0.0\ 1$) (Figure 4.5), but Fa9 did not show any enhancing activity. The suppression seen in Fa7 was absent in either Fa9 or Fa10, suggesting that if a negative enhancer is present with Fa7, it is likely located within 29 bp of its 5' portion (3778-3807) of Fa8.

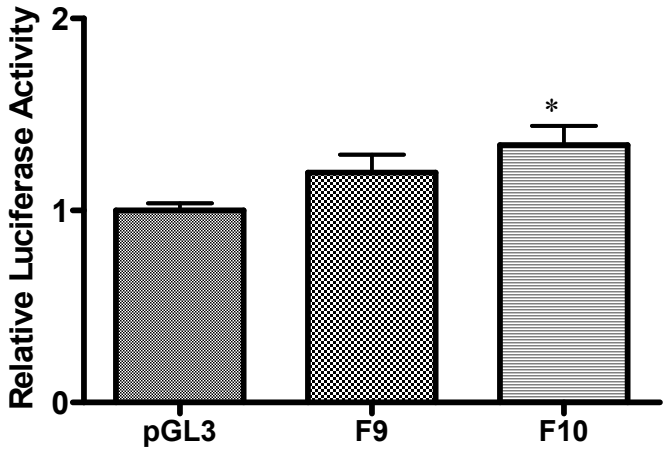

Figure 4.5. To locate the base pairs responsible for the decreased activity in Fa7. Fa9 is 1169 bp and F10 is 778 bp of the 3'portion of *Bgl*2 cutting site. Transfcetion was carried in triplicate in each experiment, and the experiment was repeated three times. Bars indicate means ±SE from these experiments. * $P<0.01$ compared to PGL3.

Enhancer activity lies within 100 bp (Fa8) of Fa2 flanking a *Bgl*2 site

The loss of activity by cutting at the *Bgl*2 site led us to hypothesize that the region important for the enhancing activity of the Fa2 locates near this site. To test this, we made a 100 bp DNA fragment, Fa8, containing 50 bp on either side of the *Bgl*2 site by PCR. As shown in Figure 4.6, Fa8 was sufficient for 3.7 fold increase of luciferase activity over the empty vector as there is no statistical difference between the relative luciferase activities of Fa8 containing vector compared to the vector of Fa2 inserted in both directions. We concluded from this experiment that the enhancer element (s) localizes within the 100 base pairs of DNA flanking the *Bgl*2 site.

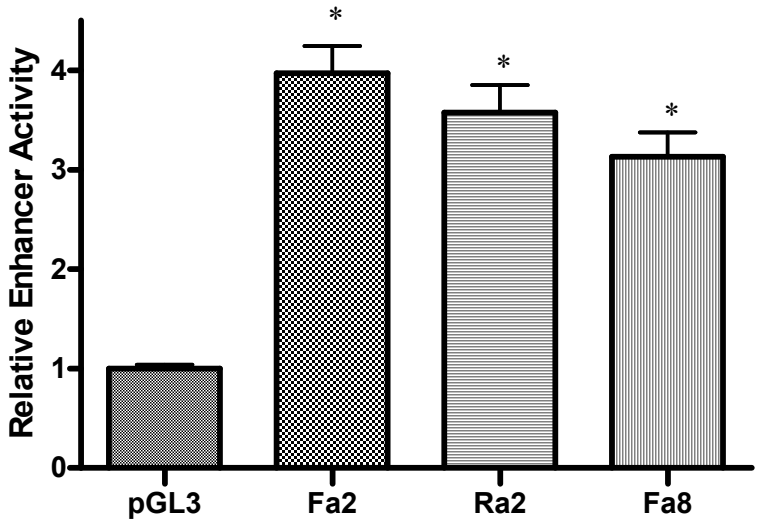

Figure 4.6. Enhancer activity is located in 100 bp within Fa2. Transfcetion was carried in triplicate in each experiment, and the experiment was repeated three times. Bars indicate means ±SE from these experiments. * P<0.01 compared to PGL3.

We first examined whether the 100 bp, Fa8, sequence contains any known transcription factor binding sites using programs available in websites: (http://tfbind.hgc.jp and http://www.cbrc.jp/htbin/nph-tfsearch). The search revealed among others three potential sites highly homologous to consensus binding sites for SREBP, GATA and Sp1 as shown in Figure 4.7.

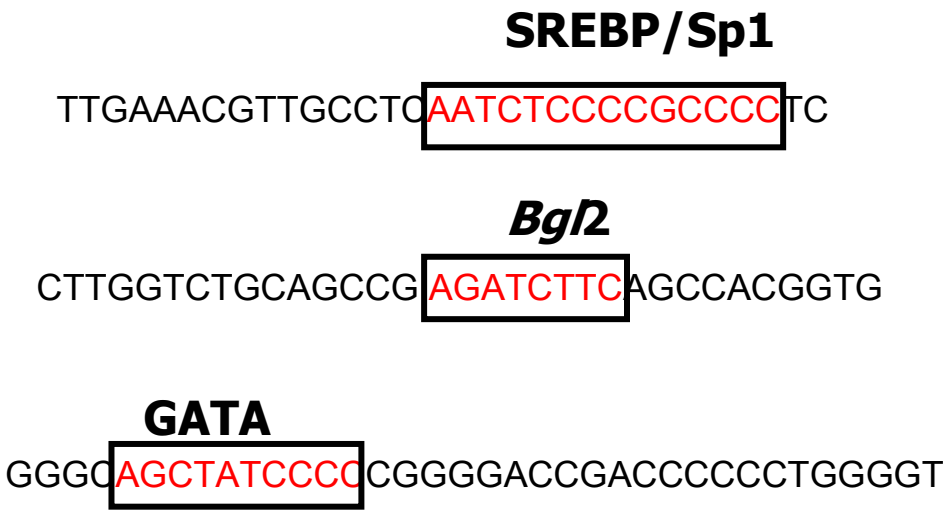

Figure 4.7. DNA sequence of Fa8 Potential transcription binding sites

SREBP is not the transcription factor that binds to Fa8

Since SREBP is known to regulate the *LDLR* gene expression, we next tested whether the enhancer activity is influenced by the cholesterol content in the culture medium. To do so, 293 cells were cultured in a lipoprotein free medium for 24 hrs, prior to transfection, and continued in culture after transfection in either a lipoprotein free medium – cholesterol (induction), or in a medium containing cholesterol and OH cholesterol (Suppression). The effectiveness of the system was tested using a test vector which contains Fa1 that has a known SREBP site in the promoter region. As Figure 4.8 A shows, the luciferase activity from the Fa1 test vector transfected cells was significantly induced in lipoprotein free medium and suppressed in cholesterol containing medium. In contrast, the activity of the cells transfected with Fa8 test vector was not influenced by the amount of cholesterol in the medium. We concluded from these results, that SREBP is not likely to be involved in the enhancing activity of Fa8.

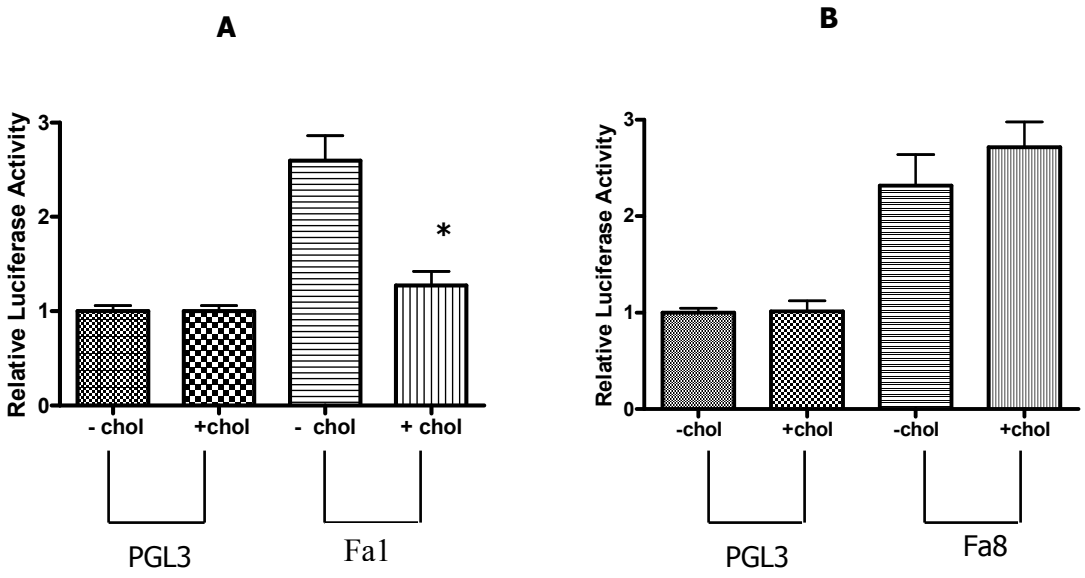

Figure 4.8. SREBP effects on enhancer activities of Fa1 (A) and Fa8 (B). Transfcetion was carried in triplicate in each experiment, and the experiment was repeated three times. Bars indicate means ±SE from these experiments. chol=cholesterol. * $P<0.01$ between suppression and repression media.

Effects of GATA transcription factors

To identify if GATA transcription factors bind to fragment Fa8, we tested the activity of the test vector containing Fa8 alone compared to the same vector co-transfected with GATA1-6 expression vectors. The result of this experiment was in conclusive since as seen in Figure 4.9 the activity of Fa8 was statistically different from the activity of GATA1-6 ($p<0.001$).

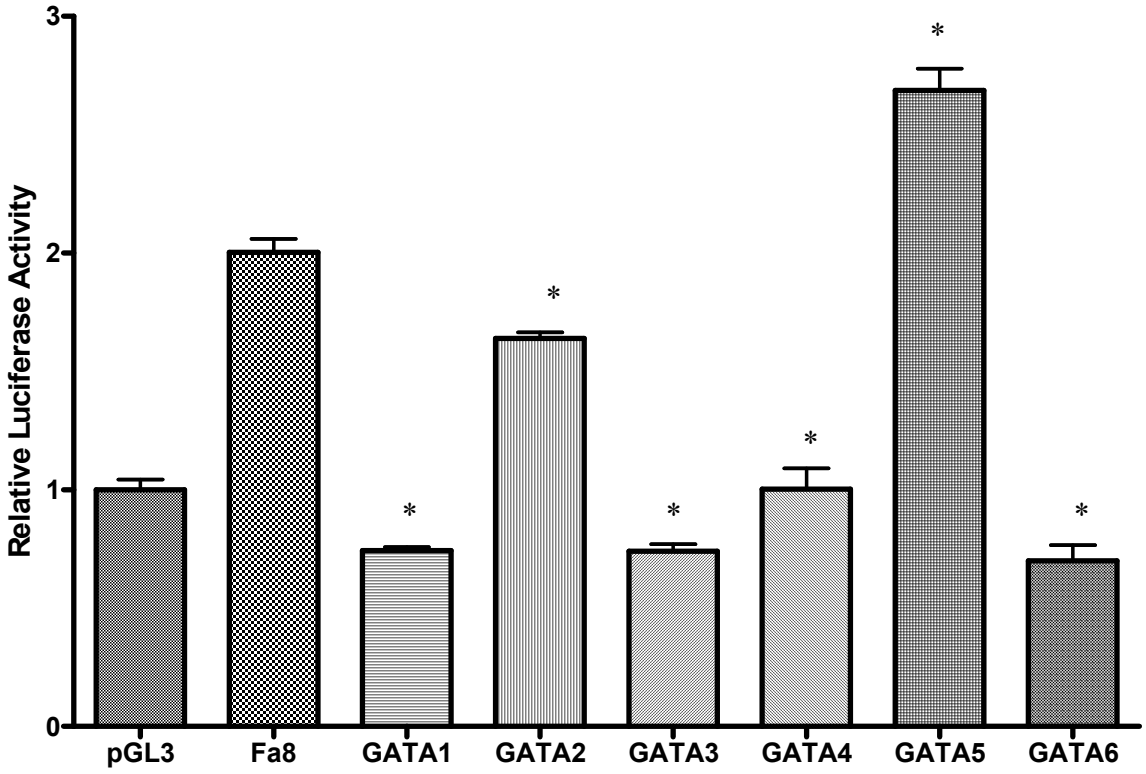

Figure 4.9. To test if GATAs transcription factors bind to Fa8. Transfcetion was carried in triplicate, and the averages of the three plates from a single experiment are shown. Bars indicate means ±SE from one experiment. * P<0.001 compared to Fa8.

To check if GATAs expression vectors work in a dose dependent manner, we repeated the same experiment as above with GATA1, GATA5 and GATA6 with decreasing doses (0.38, 0.25 and 0.1 μg). The test vector containing Fa8 was co-transfected with 0.38, 0.25 and 0.1 μg of GATA1, GATA5 and GATA6. The activity of the test vector containing Fa8 alone was not statistically different from the activity of the same vector co-transfected with GATA1, GATA5 and GATA6 with the three doses (Figure 4.10).

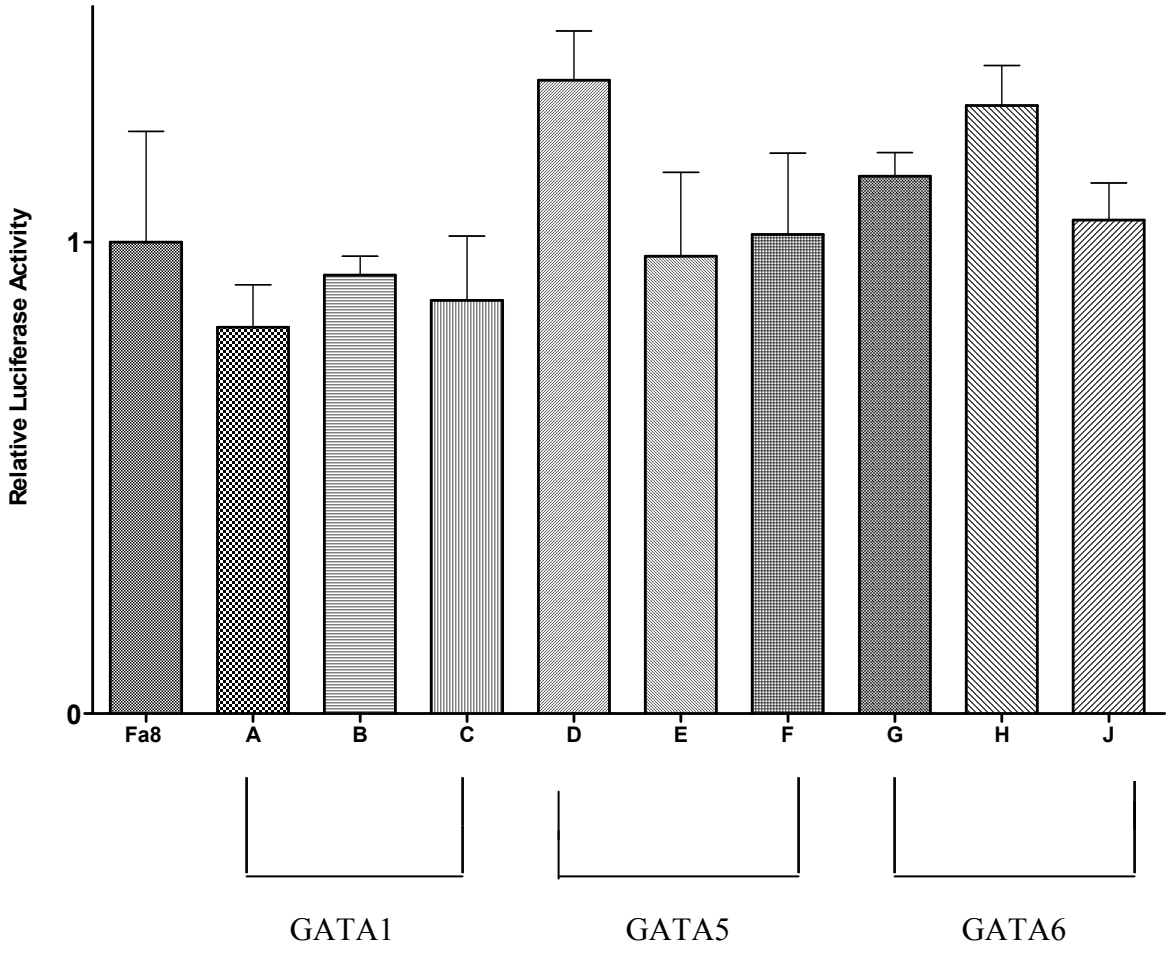

Figure 4.10. Co-transfection Fa8 with GATA1, GATA5 and GATA6 in a dose dependent manner. Transfcetion was carried in triplicate in the experiment and the experiment was done once. Bars indicate means ±SE from one experiment.
***** P<0.001 compared to Fa8.
A, B and C = Co-transfection Fa8 with 0.38, 0.25 and 0.1 µg of GATA1
D, E, and F = Co-transfection Fa8 with 0.38, 0.25 and 0.1 µg of GATA5
G, H, and J = Co-transfection Fa8 with 0.38, 0.25 and 0.1 µg of GATA6

We next repeated the same experiment as above with GATA5 alone in a dose dependent manner but in the increasing order. The test vector containing Fa8 was co-transfected with 1.5 and 2.0 µg of GATA5. The activity of the test vector containing Fa8 alone was not statistically different from the activity of the same vector co-transfected with GATA5 at the two doses (Figure 4.11).

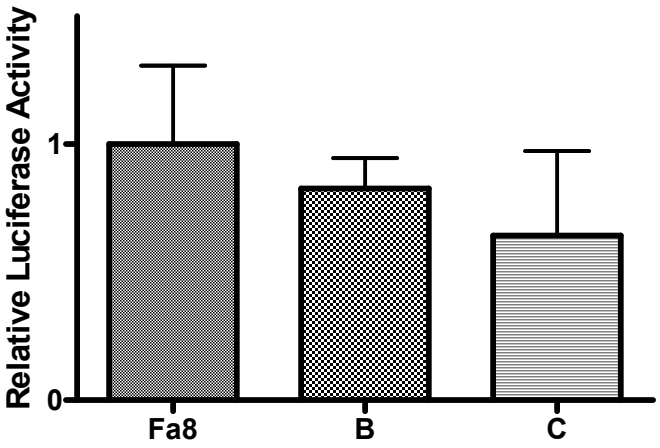

Figure 4.11. Co-transfection Fa8 with GATA5 in a dose dependent manner. Transfcetion was carried in triplicate in the experiment, and the experiment was done once. Bars indicate means ±SE from one experiment. * P<0.001 compared to Fa8.
A = Co-transfection Fa8 with 1.5 μg of GATA5
D = Co-transfection Fa8 with 2.0 μg of GATA5

All of the above results indicate that co-transfection of GATA,s do not influence the enhancer activity of fragment Fa8. The dose dependent experiments were repeated with the empty vector and the results were the same. There was no statistical difference in the activity of the PGL3 compared to the activity of the empty vector co-transfected with GATAs.

SP1 is likely involved in the enhancing activity of Fa8

In order to identify the potential transcription factor(s) that bind to the enhancer element(s) that localize within the Fa8 fragment, a band shift assay with 293 cells nuclear extract was performed. As seen in Figure 4.12 A, when the labeled 100bp Fa8 fragment was incubated with the nuclear extract, a complex was formed as a discrete band that migrated slower than the free or unbound DNA on the gel (lane 2). Incubation with 200 molar excess of cold Fa8 fragment decreased the intensity of the band indicating that the binding is specific (lane 3), while incubation with 200 molar excess of oligonucleotides containing AP2 sequence (promega) showed no competition (lane 4). Moreover, incubation with 200 molar excess of Fa11 located upstream of Fa8 that contains multiple potential Sp1 binding sites also showed competition (Figure 4.12 B, lane 3). Incubation with 200 molar excess of DNA sequence 5'-ATTCGATCGGGGCGGGGCGAGC-3' containing SP1 binding site (Promega) also showed competition (Figure 4.12 B, lane 4).

To verify that Sp1 binds to the Fa8 sequence, we performed a super shift assay with polyclonal antibody against Sp1. As shown in lane5 of Figure 4.12 A, the addition of Sp1 antibody resulted in a super shift of the nuclear extract probe complex, indicating that the shift band contained Sp1.

To test if GATA transcription factors bind to Fa8 sequence, we performed super shift assay with antibodies against GATA1 GATA 3 and GATA 4 as seen in Figure 4.12 C (lane 3, 4 and 5), the nuclear extract probe complex was not shifted. Addition of antibodies against GATA2 and GATA6 also did not result in super shift (Figure 4.12 D lane 3 and lane 4).

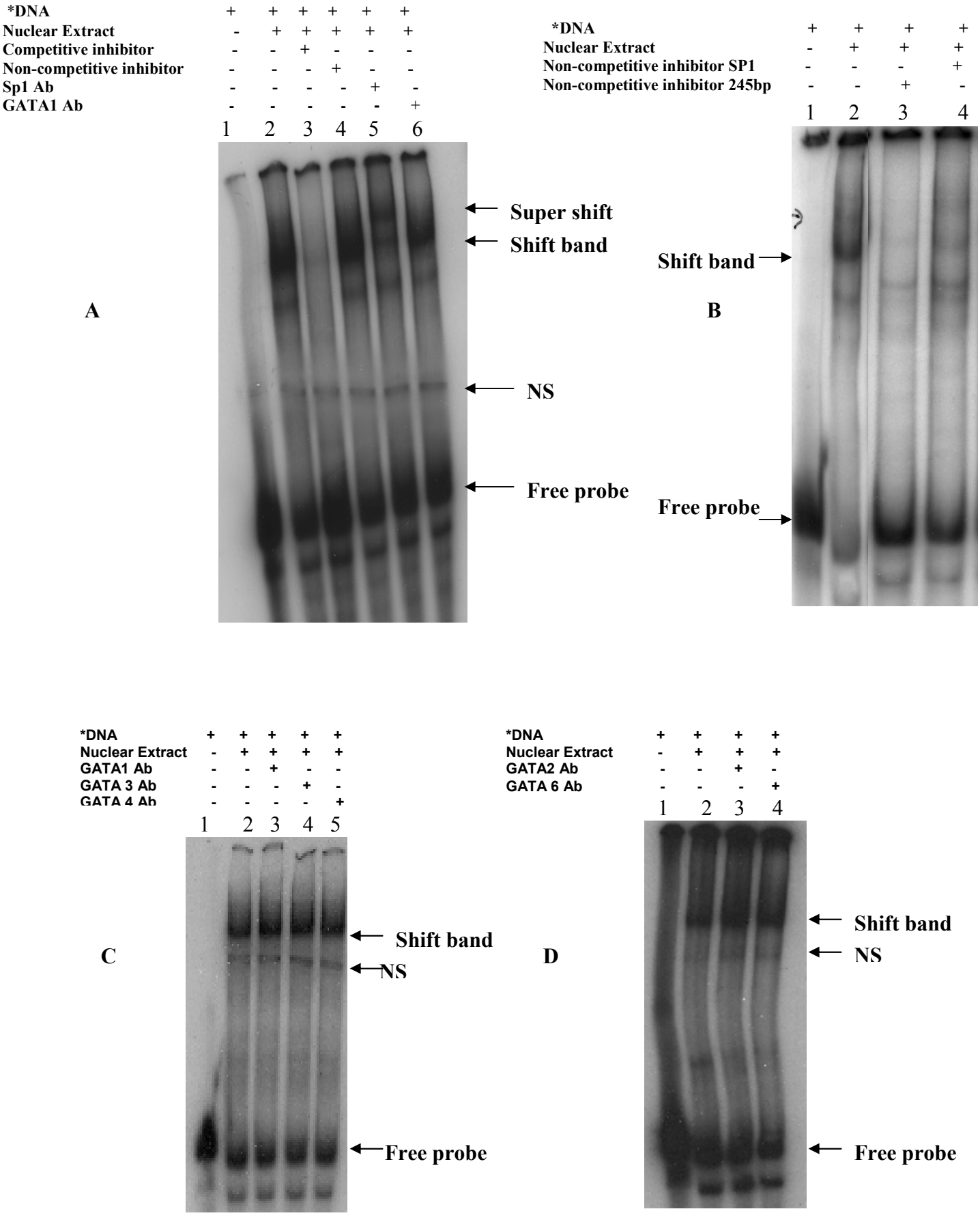
*DNA + + + + + +
Nuclear Extract - + + + + +
Competitive inhibitor - - + - - -
Non-competitive inhibitor - - - + - -
Sp1 Ab - - - - + -
GATA1 Ab - - - - - +
1 2 3 4 5 6
Super shift
Shift band
NS
Free probe
A
*DNA + + + +
Nuclear Extract - + + +
Non-competitive inhibitor SP1 - - - +
Non-competitive inhibitor 245bp - - + -
1 2 3 4
Shift band
Free probe
B
*DNA + + + + +
Nuclear Extract - + + + +
GATA1 Ab - - + - -
GATA 3 Ab - - - + -
GATA 4 Ab - - - - +
1 2 3 4 5
Shift band
NS
Free probe
C
*DNA + + + +
Nuclear Extract - + + +
GATA2 Ab - - + -
GATA 6 Ab - - - +
1 2 3 4
Shift band
NS
Free probe
D

Figure 4.12. DNA mobility shift assay. (A) 7.0 µg of nuclear extract of 293 cells were incubated at room temperature for 30 minutes with 20 µl reaction mixture containing gel shift binding buffer as mentioned in material and methods. 25,000 cpm of labeled probe was added to the mixture and incubation proceeded for another half an hour. The samples were loaded on a 4% non-denaturing polyacrylamide gel and the DNA: protein complex was analyzed by autoradiography (lane 2). For competition assay, 200 molar excess of cold competitive inhibitor (lane 3) and non competitive inhibitor (lane 4) were added to the above reaction before the addition of the labeled probe. Super shift was observed upon the addition of 4 µg of Sp1 antibody in the mixture before the addition of labeled probe (lane 5). No super shift band was seen with GATA1 antibody (Lane 6). (B) Incubation with 200 molar excess of Fa11 showed competition (lane 3). Incubation with 200 molar excess of SP1 sequence 5'-ATTCGATCGGGGCGGGGCGAGC-3' also showed competition (lane 4). (C) 4 µg of GATA1 (lane 3), GATA3 (lane 4), and GATA4 (lane 5) antibodies were assayed for super shift as for SPI. (D) GATA2 (lane 3) and GATA 6 (lane 4) antibodies were assayed for super shift as above.

Discussion

In the present study we have identified in intron I of the human *LDLR* gene a DNA sequence that enhances transcription of a reporter gene. Previous studies have shown enhancer activity within the promoter which is governed by three direct repeats. Repeat 1 and 3 bind Sp1 transcription factor to drive the basal level of transcription, while repeat 2 binds the sterol regulatory element binding protein (SREBP) to the sterol regulatory element (SRE) to drive the cholesterol mediated transcription. Recently, Wang et el also showed that there is an enhancer activity approximately 13 kb downstream of the *LDLR* gene.

Losing enhancer activity by cutting at the *Bgl*2 site within the Fa8 led us to speculate that another transcription factor is also necessary for the full activity. But, when the transcription factor binding sites were physically separated by cutting at *Bgl*2 site, there was a complete loss of function of the enhancer. This finding led us to speculate that the presence of at least two transcription factors one binding to 5' and the other binding to 3' to the *Bgl*2 sites are necessary for the enhancing activity.

The result of the supershift with the Sp1 polyclonal antibody suggests that Sp1-binding is the likely factor that is necessary but not sufficient for this activity. Sp1 is ubiquitously expressed and beside its role in recruiting the basal transcription machinery, it also works with other transcription factors to control specific gene expression (Gyrd-Hansen et al. 2002; Martino et al. 2001; Qin et al. 1999). For example, Qin et al showed that Sp1 and Ap2 are both important for the costitutive matrix metallproteinase-2-gene (MMP-2) expression in astroglioma cells. Mutations in Sp1 A site reduceded MMP-2 transcription by 41-70 % compared to the wild type, while mutation in Ap2 caused 36-60 % reduction in transcription compared to the wild type. Double mutations at binding sites Sp1A and Ap2

decreased transcription to 74-94%, which means that both Sp1 and Ap2 are functionally important for transcription of metallproteinase-2-gene expression.

Sp1 is also well known to interact with other transcription factors including GATAs. Both Sp1 and GATA proteins have zinc finger domain which facilitate the interaction between them through their finger region. However, our experiments showed that the luciferase activity in the cells transfected with Fa8 test vector was not statistically significant from the activity of Fa8 test vector co-transfected with GATAs expression vectors. Hence, this implies that the Sp1 binding partner is a different transcription factor and not the GATAs transcription factors. Hence, further work is necessary to identify the other transcription factor (s).

In the present study we verified the presence of an enhancer within intron 1 of the human *LDLR* gene and identified the transcription factor Sp1 that bind to specific sequences. Regardless of the precise mechanism of the transcriptional activity enhancement, we pointed out that there are several SNP sites near to this 100 bp sequence. It would also be important to verify whether the nearby polymorphisms have any effects on the enhancer activity.

Chapter V

Variations in the human low density lipoprotein receptor

In this chapter I will summarize the well studied genetic variations i.e. the single nucleotide polymorphisms and the haplotypes that have been identified in the human low density lipoprotein receptor gene (*LDLR*), and have been associated with serum level of total cholesterol (TC) and low density lipoprotein cholesterol (LDL-C) in normalipidic individuals. As I have mentioned earlier, the *LDLR* gene plays an important role in cholesterol homeostasis, hence any identification of genetic variations that affect the level of TC and LDL-C is important in the human population. While the individual SNPs may not be causative, they are in linkage disequilibrium with other causative SNPs in the gene or neighboring genes. Beside the fact that some of these variations in the *LDLR* are associated with familial hypercholesterolemia, others contribute to the variations in TC and LDL-C in normal individuals.

Restriction fragment length polymorphisms

Early studies used restriction fragment length polymorphisms (RFLPs) to detect genetic variations in the human *LDLR* gene. For example, Schuster et al in 1990 analyzed in 324 normocholesterolemic individuals in Germany between the presence and absence of sites for the restriction enzymes *Stu* I (exon 8) *Pvu* II, *Apa*LI (intron 15) and *Nco* I (exon 18) and their association with plasma lipid levels. They found that the presence of *Pvu* II was significantly associated with lower LDL-C by 3.0 mg/dl. Of the sixteen possible RFLP haplotypes, nine were detected in this population and only four were common. As summarized in Table 5.1, the carries of haplotype 3 had lower LDL-C than the sample mean (average excess) by

by, - 9.49 mg/dl. Carriers of Haplotypes 2 and 4 were associated with a very small reduction of LDL-C by, - 0.15 mg/dl and – 0.41 mg/dl. While carriers of haplotype 1 showed a modest increase of LDL-C by 1.44 mg/dl (Schuster et al. 1990; Templeton 1987).

Table 5.1. Deduced haplotypes of four RFLPs in subjects from Germany.

Haplotype	*Stu* I	*Apa*LI	*Pvu* II	*Nco* I	Frequency	Average Excess
	Exon 8	Intron 15	Intron 15	Exon 18	%	mg/dl*
1	+	+	-	+	27	+1.44
2	+	-	-	-	25	-0.15
3	+	-	+	+	20	-9.49
4	+	+	-	-	12	-0.41

+ The presence of the cutting site, - the absence of the cutting site.
*mg/dl LDL-C. Four common haplotypes out of sixteen total haplotypes.

Smilarly, in 1991 Humphries et al studied four RFLPs detected by the enzymes, *Ava* II, (exon 13), *Pvu* II (intron 15) and *Apa*LI (3'UTR) in 289 normolipidic individuals in Italy (table 5.2). They also found that the presence of the *Pvu* II site in intron 15 had an average excess of LDL-C by - 9.0 mg/dl. Of the nine RFLP haplotypes identified, four were common. Haplotype 1 that contains the *Pvu* II site was associated with an average excess of LDL-C by - 11.5 mg/dl. Haplotypes 2-4 were associated with an average excess LDL-C of + 8.35 mg/dl, + 4.55 mg/dl, and + 3.05 mg/dl respectively as seen in Table 5.2.

Clearly these two studies concur and demonstrate that the major effect is defined by the presence or absence of *Pvu* II site in intron 15, and other sites are less important.

Table 5.2. Common haplotypes in normolipidic individuals from Italy.

Haplotype	*Ava* II	*Apa*LI 5'	*Pvu* II	*Apa*LI 3'	Frequency	Average Excess
	Exon 13	Intron 15	Intron 15	3'UTR	%	mg/dl*
1	-	-	+	+	27	- 11.5
2	-	+	-	+	6.2	+ 8.35
3	+	+	-	+	43.5	+ 4.55
4	-	-	-	-	17.9	+ 3.05

+ presence of a cutting site, - absence of a cutting site
*Apa*L1 5', cutting site in intron 15, and *Apa*L1 3', cutting site in the 3'UTR.
* mg/dl of LDL-C. Four common haplotypes out of nine total haplotypes.

Ava II in exon 13 and *Nco* I in exon 18 were also studied by Ahn and his group in 1994 in 385 normolipidic Hispanic and 543 non-Hispanic whites, from Colorado. The *Ava* II positive allele was associated with significantly increased TC and LDL-C by 6.52 mg/dl and 4.95 mg/dl in non-Hispanic women and 3.69 mg/dl and 2.29 mg/dl in Hispanic women. The *Ava* II negative allele decreased significantly TC and LDL-C by 4.84 mg/dl and 3.68 mg/dl in non-Hispanic women; similar reduction was seen in Hispanic women. *Nco* I positive allele was also associated with significantly increased TC and LDL-C by 3.66 mg/dl and 3.07 mg/dl but only in Hispanic women. The *Nco* I negative allele decreased significantly TC and LDL-C by 8.99 mg/dl and 7.55 mg/dl in Hispanic women (Ahn et al. 1994). There was no effect of these two SNPs in men. Although both SNPs are in exons, amino acid sequence of LDL receptor is not altered by either nucleotide change.

Finally, not all studies have shown that these sites are important. In 1998 Taylor et al studied four RFLPs generated by the restriction sites, *Stu* I (exon 8), *Pvu* II, *Apa*LI (intron 15) and *Nco* I (exon 18) in a sample of 484 normolipidic individuals in London Authors found that there was no statistical difference in the mean level of serum TC and LDL-C among all the genotypes (Taylor et al. 1988). The subjects were Caucasian males and females. The discrepancy in the effect of *Pvu* II positive allele among Taylor's population and Humphries and Schusters's population may be attributed to the fact that different populations may have different gene-gene interaction as well as gene environment interaction, hence masking the effect of *Pvu* II on LDL-C in this population.

Single nucleotide polymorphisms in *LDLR* gene

Studies have been shifted in recent years to the analysis of SNPs in the *LDLR* gene. For example, Boright et al in 1998 studied the association of a polymorphism rs688 (C1773T) in exon 12 to LDL-C levels in a normolipidic sample of 719 Alberta Hutterites. They found that the average effect (from the sample mean) of 1773T allele on LDL-C was + 3.9 mg/dl, while the average effect of C1773 allele LDL-C was - 6.2mg/dl. The influence of the two alleles is small but significant. This nucleotide change within exon 12 does not change the amino acid, hence authors concluded that it is in linkage disequilibrium with other functional variants (Boright et al. 1998). However, as being described below, this polymorphism may also have a functional significance.

Knoblauch et al in 2002 studied the effect of common SNP Haplotypes of the *LDLR* gene in 732 individuals in Germany. They showed that SNPs in *LDLR*: rs5930 (G26462A Exon 10), rs5925 (A33078G Exon 13), rs5927 (C36138T.Exon 15), fall into 7 haplotypes: GTG, GTA, GCG, GCA, ATG, ATA, and ACG. The only haplotype that was associated with lipid profiles was GTG, the presence of which was associated with a significant decrease of TC by 0.32 mg/dl and LDL-C by 0.41 mg/dl in 102 individulas (Knoblauch et al. 2002).

Authors extended their study to include extra SNPs, rs2228671 (C13109T Exon 2), rs885765 (A18896C Intron 4), rs5930 (G26462A Exon 10), rs5925 (A33078G Exon 13), rs5927 (C36138T Exon 15), *Pvu* II intron 15, C/T and rs1433099 (C44857T 3'UTR). Twenty two haplotyes were conferred in 1054 subjects in Germany. Authors concluded (2004) that the contribution of the *LDLR* haplotypes to the genetic variance of LDL-C was 5 %, HDL-C 6% and LDL/HDL 2% (Knoblauch et al. 2004). None of the above

polymorphisms have been reported to have functional influence, and therefore are thought to be in linkage disequilibrium with a causative polymorphism (s).

In 2007 we studied the haplotypes of four SNPs within the 5' region of the 3'UTR ,rs14158 (G44243A), rs2738465 (G44695A), rs1433099 (C44857T) rs2738466 (A44964G). We further sought association of the three common haplotypes, *GGCGCA*, A*GCACG*, and *GGCGTA*, with total plasma cholesterol and LDL-C among subjects in the Atherosclerosis Risk in Communities (ARIC) study. The mean LDL-C levels were 7 mg/dl and 4 mg/dl higher in *C/C* (44857) homozygous Caucasian males and females than *T/T*homozygotes. LDL-C levels in African American females were 3 mg/dl higher in *C/C* homozygous compared to *T/T* homozygotes. In contrast, no effects of C44857T on LDL-C levels were detected in African American males. The A44965G polymorphism had no effects on the lipid profiles in either population. When haplotypes were analyzed, we found that Caucasian men with *TA/TA* haplotypes had 6.06 mg/dl lower TC and, 6.10 mg/dl lower LDL-C. In contrast, Caucasian men with the *CA/CA* haplotype had 2.46 mg/dl higher LDL-C. Caucasian women with the *TA/TA* haplotype had 4.07 mg/dl lower TC, and 4.63 mg/dl lower LDL-C. There was no effect of these haplotypes on African American men and women (Muallem et al. 2007). The contribution of the 3'UTR haplotypes in Caucasians to the genetic variance of LDL-C, HDL and LDL/HDL is consistent with Knoblauch et al's finding. The importance of our work is that it is the first study to investigate a large number of normolipidic subjects of both Gaucasians and African Americans.

A potential polymorphism associated with alteration in the *LDLR* expression was recently reported by Zhu and his group in 2007. These authors studied the function of a SNP rs688 (C1773T) in exon 12, which is common in many populations as European

Caucasians, African populations and some Asian populations. It has a frequency of approximately 60% in Caucasians This SNP was first identified by leitersdorf and Hobbs in 1988 (Leitersdorf and Hobbs 1988) and was studied for association by Boright (Boright et al. 1998). Zhu et al found by examining *LDLR* expression in the liver biopsy samples that the 1773T allele decreased significantly the splicing efficiency of exon 12 in females. Cells in females homozygous for 1773T allele had a decreased splicing efficiency by about 8.6% compared to females homozygous for the 1773 C allele. The 1773T allele leads to an increased amount of LDLR transcripts lacking exon 12 that results in shifting of the reading frame and premature terminational protein. In this study 1773T allele was found to be associated with 10 % increase in TC and LDL-C in premenopausal women. On the other hand, the decrease in splicing efficiency was not significant in males and hence there was no association of this SNP with LDL-C among men (Zhu et al. 2007). The lack of association between rs688 and LDL-C in males could be explained by the gender difference in exon 12 splicing efficiency between males and premenopausal females. This difference could be due to the modulating effect of estrogen on the SRp40, the predicted splicing factor that binds to exon splicing enhncers (ESE). The C1773T does not alter the amino acid sequence, but is within the exon enhancer splicing region. This is the first potential functional common SNP reported in the human *LDLR* beside the functional work that was intensively done on the promoter portion of the gene by Goldstein, Hobbs and their group.

Linkage disequilibrium in *LDLR* gene

Having found association between SNPs in the *LDLR* gene and the lipid profile, and based on the fact that only one SNP rs688 in exon 12 is functional, the question remains as to whether all other SNPs associated with plasma LDL-C are in linkage disequilibrium with the functional rs 688. Humphries and his group found that *Ava* II, *Apa*LI, (5' and 3') and *Pvu* II are in linkage disequilibrium however, the main effect on LDL-C is the presence of *Pvu* II allele. The effect of *Pvu* II positive allele is not due to *Pvu* II per se, because *pvu* II is located in intron 15, hence it requires that it should be in linkage disequilibrium with another functional polymorphism

In addition, Ahn and his group showed that *Ava* II and *Nco* I are in linkage disequilibrium in both Hispanic and non-Hispanic populations and they both increase the level of TC and LDL-C in the two populations. The presence of both polymorphisms in a coding sequence resulting in synonymous amino acids is suggestive that they are not functional. Rather they are in linkage disequilibrium with other functional SNPs else where in the gene. Using Perlegen Genome Browser to detect linkage disequlibrium blocks in human *LDLR* gene in Caucasians, showed that AvaII in exon 13 is in linkage disequilibrium with rs688 (C1773T) in exon 12 due to their physical proximity. Zhu et al showed that this rs688 is functional because females homozygous for T/T have decreased exon 12 splicing efficiency by 8.6% compared to females homozygous for C/C. Hence, the increasing effect of AvaII and NcoI positive alleles on LDL-C may be explained by the linkage disequilibrium of these two SNPs with the functional rs668 that has the similar effect on LDL-C. Moreover, the 3 SNPs did not show any effect in men.

The effect of the four SNPs in the 3'UTR region of the gene, three of which are in linkage disequilibrium, (rs14158, rs2738456 and rs2738466) on LDL-C studied by Muallem et al is not due to the 3'UTR sequence per se but others that are in linkage disequilibrum with them. Since it is possible that the effect is due to rs688, we searched HapMap phased haplotypes of the 2 SNPs and we found that approximately 86 % of the individuals having rs1433099 44857T in the 3'UTR in Caucasians, also have rs688 1773C in exon 12 indicating that 44857T and 1773C are completely linked and both have a lowering effect on LDL-C. On the other hand, we found that approximately 60 % of individuals having C44857 also have C1773 and the other 40 % have 1773T indicating that linkage is low. Hence, the effect of C44857 on LDL-C is dependent on the linkage with either 1773T that decreases LDLR or 1773C that increases LDLR.

On the other hand, rs688 is not studied in African Americans, implying that C44857T may be in linkage disequilibrium with other SNP in the gene.

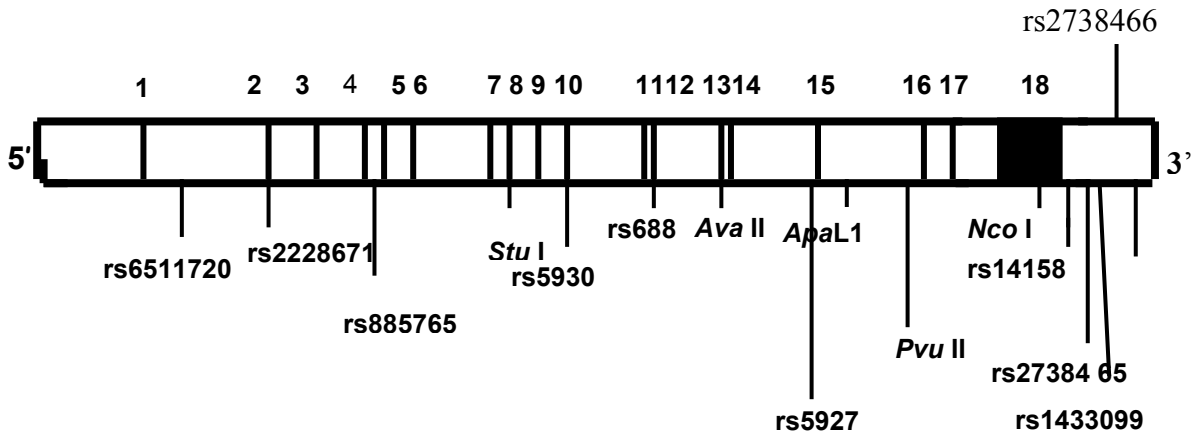

Figure 5.1. Map of the polymorphisms mentioned in this chapter and their location in the gene.

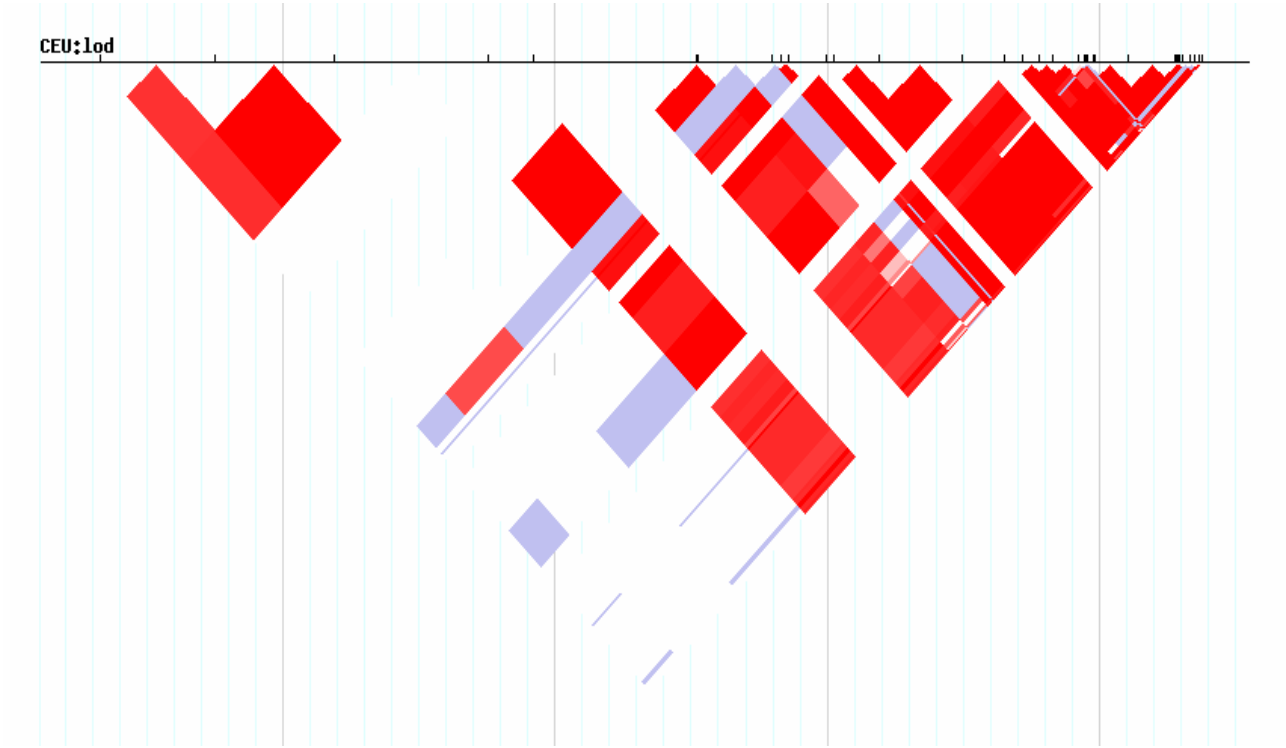

Figure 5.2. Linkage disequilibrium in the human *LDLR* gene copied from HapMap Genome Browser (http://www.hapmap.org). LD plot is from CEU population from HapMap. Darker regions implicate higher LD calculated by pairwise r2-values between markers.

Association of *LDLR* SNPs and coronary heart disease

Even though coronary heart disease is caused by familial hypercholesterolemia (FH), an inherited autosomal dominant disease with hallmarks of elevated plasma LDL-C levels as a result of mutations in the *LDLR* gene, most individuals who suffer from CHD do not have FH and their plasma lipoprotein levels are within the reference range.

For example, Salazar et al in 2000 studied the association of the *Pvu* II polymorphism with difference in lipid concentration in 228 white subjects with high and low risk of coronary heart disease in Brazil. Consistent with other observations, the presence of *Pvu* II was associated with decreased TC and LDL-C levels in the Brazilian population. The *Pvu* II positive allele frequency in individuals with high risk group was 0.25 compared to low risk group. In addition, the *Pvu* II positive allele was also found at a higher frequency among elderly people over the age of 65 suggesting that the allele may be associated with longer survival

In addition, in our study we found that 22 % Caucasian men homozygous for the *CA/CA* diplotype had a significantly higher LDL-C and significant lower HDL-C levels. As a result, they also had a 4.5% higher LDL/HDL ratio, an established risk factor for CHD (Panagiotakos et al., 2003). Conversely, 8 % Caucasian men that were homozygous for the *TA/TA* diplotype had significantly lower total cholesterol and LDL-C, and a 5.3% lower LDL/HDL ratio.

Recently, Willer et al studied the effect of some genetic variants in several genes affecting lipid metabolism including *LDLR*. They found that a SNP rs6511720 (G4504T) in intron I of *LDLR* increased LDL-C by 9.17 mg/dl. This SNP was approximately 1.2% more common among coronary artery disease (CAD) cases than in the control group (Willer et al.

2008). Hence, SNPs in *LDLR* gene are a good genetic marker to predict the risk of CHD in the population.

In summary, the *LDLR* gene is one locus with multiple alleles. Association studies of single nucleotide polymorphisms in *LDLR* with lipid parameters have been well studied. Extra work is required to investigate more functional polymorphisms beside the rs688 present in exon 12.

References

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

Acton S, Rigotti A, Landschulz KT, Xu S, Hobbs HH, Krieger M (1996) Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. Science 271: 518-20

Ahn YI, Kamboh MI, Aston CE, Ferrell RE, Hamman RF (1994) Role of common genetic polymorphisms in the LDL receptor gene in affecting plasma cholesterol levels in the general population. Arterioscler Thromb 14: 663-70

Arca M, Zuliani G, Wilund K, Campagna F, Fellin R, Bertolini S, Calandra S, Ricci G, Glorioso N, Maioli M, Pintus P, Carru C, Cossu F, Cohen J, Hobbs HH (2002) Autosomal recessive hypercholesterolaemia in Sardinia, Italy, and mutations in ARH: a clinical and molecular genetic analysis. Lancet 359: 841-7

Aronow B, Lattier D, Silbiger R, Dusing M, Hutton J, Jones G, Stock J, McNeish J, Potter S, Witte D, et al. (1989) Evidence for a complex regulatory array in the first intron of the human adenosine deaminase gene. Genes Dev 3: 1384-400

Austin MA, Hutter CM, Zimmern RL, Humphries SE (2004) Familial hypercholesterolemia and coronary heart disease: a HuGE association review. Am J Epidemiol 160: 421-9

Banares VG, Peterson G, Aguilar D, Gulayin R, Sisu E, Wyszynski DF, Pivetta OH, Tavella MJ (2005) Association between the APOE*4 allele and atherosclerosis is age dependent among Argentine males. Hum Biol 77: 247-56

Banerji J, Olson L, Schaffner W (1983) A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33: 729-40

Basler K, Siegrist P, Hafen E (1989) The spatial and temporal expression pattern of sevenless is exclusively controlled by gene-internal elements. Embo J 8: 2381-6

Benjannet S, Rhainds D, Essalmani R, Mayne J, Wickham L, Jin W, Asselin MC, Hamelin J, Varret M, Allard D, Trillard M, Abifadel M, Tebon A, Attie AD, Rader DJ, Boileau C, Brissette L, Chretien M, Prat A, Seidah NG (2004) NARC-1/PCSK9 and its natural mutants: zymogen cleavage and effects on the low density lipoprotein (LDL) receptor and LDL cholesterol. J Biol Chem 279: 48865-75

Benjannet S, Rhainds D, Hamelin J, Nassoury N, Seidah NG (2006) The proprotein convertase (PC) PCSK9 is inactivated by furin and/or PC5/6A: functional consequences of natural mutations and post-translational modifications. J Biol Chem 281: 30561-72

Benn M, Nordestgaard BG, Jensen JS, Nilausen K, Meinertz H, Tybjaerg-Hansen A (2005) Mutation in apolipoprotein B associated with hypobetalipoproteinemia despite decreased binding to the low density lipoprotein receptor. J Biol Chem 280: 21052-60

Berge KE, Ose L, Leren TP (2006) Missense mutations in the PCSK9 gene are associated with hypocholesterolemia and possibly increased response to statin therapy. Arterioscler Thromb Vasc Biol 26: 1094-100

Boerwinkle E, Visvikis S, Welsh D, Steinmetz J, Hanash SM, Sing CF (1987) The use of measured genotype information in the analysis of quantitative phenotypes in man. II. The role of the apolipoprotein E polymorphism in determining levels, variability, and covariability of cholesterol, betalipoprotein, and triglycerides in a sample of unrelated individuals. Am J Med Genet 27: 567-82

Boerwinkle E, Utermann G (1988) Simultaneous effects of the apolipoprotein E polymorphism on apolipoprotein E, apolipoprotein B, and cholesterol metabolism. Am J Hum Genet 42: 104-12

Bouthillier D, Sing CF, Davignon J (1983) Apolipoprotein E phenotyping with a single gel method: application to the study of informative matings. J Lipid Res 24: 1060-9

Brown SA, Hutchinson R, Morrisett J, Boerwinkle E, Davis CE, Gotto AM, Jr., Patsch W (1993) Plasma lipid, lipoprotein cholesterol, and apoprotein distributions in selected US communities.

Boright AP, Connelly PW, Brunt JH, Morgan K, Hegele RA (1998) Association and linkage of LDLR gene variation with variation in plasma low density lipoprotein cholesterol. J Hum Genet 43: 153-9

Briggs MR, Yokoyama C, Wang X, Brown MS, Goldstein JL (1993) Nuclear protein that binds sterol regulatory element of low density lipoprotein receptor promoter. I. Identification of the protein and delineation of its target nucleotide sequence. J Biol Chem 268: 14490-6

Brown SA, Hutchinson R, Morrisett J, Boerwinkle E, Davis CE, Gotto AM Jr, Patsch W (1993) Plasma lipid, lipoprotein cholesterol, and apoprotein distributions in selected US communities. The Atherosclerosis Risk in Communities (ARIC) Study. Arterioscler Thromb 13:1139-1158.

Burnett JR, Shan J, Miskie BA, Whitfield AJ, Yuan J, Tran K, McKnight CJ, Hegele RA, Yao Z (2003) A novel nontruncating APOB gene mutation, R463W, causes familial hypobetalipoproteinemia. J Biol Chem 278: 13442-52

Burnett JR, Zhong S, Jiang ZG, Hooper AJ, Fisher EA, McLeod RS, Zhao Y, Barrett PH, Hegele RA, van Bockxmeer FM, Zhang H, Vance DE, McKnight CJ, Yao Z (2007) Missense mutations in APOB within the betaalpha1 domain of human APOB-100 result in

impaired secretion of ApoB and ApoB-containing lipoproteins in familial hypobetalipoproteinemia. J Biol Chem 282: 24270-83

Cameron J, Holla OL, Ranheim T, Kulseth MA, Berge KE, Leren TP (2006) Effect of mutations in the PCSK9 gene on the cell surface LDL receptors. Hum Mol Genet 15: 1551-8

Chen SN, Ballantyne CM, Gotto AM, Jr., Tan Y, Willerson JT, Marian AJ (2005) A common PCSK9 haplotype, encompassing the E670G coding single nucleotide polymorphism, is a novel genetic marker for plasma low-density lipoprotein cholesterol levels and severity of coronary atherosclerosis. J Am Coll Cardiol 45: 1611-9

Chen JM, Ferec C, Cooper DN (2006) A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. Hum Genet 120:1-21.

Conne B, Stutz A, Vandli JD (2000) The 3' untranslated region of messenger RNA: a molecular 'hotspot' for pathology. Nature Med 6:637-641.

Cheng D, Huang R, Lanham IS, Cathcart HM, Howard M, Corder EH, Poduslo SE (2005) Functional interaction between APOE4 and LDL receptor isoforms in Alzheimer's disease. J Med Genet 42: 129-31

Chung S, Perry RP (1989) Importance of introns for expression of mouse ribosomal protein gene rpL32. Mol Cell Biol 9: 2075-82

Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet 37: 161-5

Corbo RM, Scacchi R (1999) Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? Ann Hum Genet 63: 301-10

Davignon J, Gregg RE, Sing CF (1988) Apolipoprotein E polymorphism and atherosclerosis. Arteriosclerosis 8: 1-21

Dawson PA, Hofmann SL, van der Westhuyzen DR, Sudhof TC, Brown MS, Goldstein JL (1988) Sterol-dependent repression of low density lipoprotein receptor promoter mediated by 16-base pair sequence adjacent to binding site for transcription factor Sp1. J Biol Chem 263: 3372-9

Dawson PA, Rudel LL (1999) Intestinal cholesterol absorption. Curr Opin Lipidol 10: 315-20

DeLong CJ, Smith WL (2005) An intronic enhancer regulates cyclooxygenase-1 gene expression. Biochem Biophys Res Commun 338: 53-61

Frederickson RM, et al., (1989) 5'flanking and first intron sequences of the human beta - actin gene required for efficient promotor activity. Nucleic Acids Res 17(1): 253-70

Fasano T, Cefalu AB, Di Leo E, Noto D, Pollaccia D, Bocchi L, Valenti V, Bonardi R, Guardamagna O, Averna M, Tarugi P (2007) A novel loss of function mutation of PCSK9 gene in white subjects with low-plasma low-density lipoprotein cholesterol. Arterioscler Thromb Vasc Biol 27: 677-81

Garcia CK, Wilund K, Arca M, Zuliani G, Fellin R, Maioli M, Calandra S, Bertolini S, Cossu F, Grishin N, Barnes R, Cohen JC, Hobbs HH (2001) Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. Science 292: 1394-8

Garuti R, Jones C, Li WP, Michaely P, Herz J, Gerard RD, Cohen JC, Hobbs HH (2005) The modular adaptor protein autosomal recessive hypercholesterolemia (ARH) promotes low density lipoprotein receptor clustering into clathrin-coated pits. J Biol Chem 280: 40996-1004

Gent J, Braakman I (2004) Low-density lipoprotein receptor structure and folding. Cell Mol Life Sci 61: 2461-70

Goldstein JL, Brown MS (1984) Progress in understanding the LDL receptor and HMG-CoA reductase, two membrane proteins that regulate the plasma cholesterol. J Lipid Res 25: 1450-61

Goldstein J, Hobbs H, Brown M (1995) Familial Hypercholesterolemia. In Scriver CR, Beaudet AL, Sly WS, Valle D (Eds) The Metabolic and Molecular basis of inherited disease. 7th ed. McGraw-Hill, New York, NY USA, Vol II, pp 1981-2030.

Gotto AM Jr (1994) Heart disease in the assessment and treatment of hypercholesterolemia: coronary artery disease and other atherosclerotic disease, family history, and left ventricular hypertrophy. Am J Med 96:9S-18S.

Gregori C, Porteu A, Lopez S, Kahn A, Pichard AL (1998) Characterization of the aldolase B intronic enhancer. J Biol Chem 273: 25237-43

Gyrd-Hansen M, Krag TO, Rosmarin AG, Khurana TS (2002) Sp1 and the ets-related transcription factor complex GABP alpha/beta functionally cooperate to activate the utrophin promoter. J Neurol Sci 197: 27-35

Hallman DM, Boerwinkle E, Saha N, Sandholzer C, Menzel HJ, Csazar A, Utermann G (1991) The apolipoprotein E polymorphism: a comparison of allele frequencies and effects in nine populations. Am J Hum Genet 49: 338-49

He G, Gupta S, Yi M, Michaely P, Hobbs HH, Cohen JC (2002) ARH is a modular adaptor protein that interacts with the LDL receptor, clathrin, and AP-2. J Biol Chem 277: 44044-9

Hobbs HH, Brown MS, Goldstein JL (1992) Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. Hum Mutat 1: 445-66

Hobbs HH, Russell DW, Brown MS, Goldstein JL (1990) The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. Ann Rev Genet 24:133-170.

Hobbs HH, Leitersdorf E, Goldstein JL, Brown MS, Russell DW (1988) Multiple crm-mutations in familial hypercholesterolemia. Evidence for 13 alleles, including four deletions. J Clin Invest 81: 909-17

Horton JD, Cohen JC, Hobbs HH (2007) Molecular biology of PCSK9: its role in LDL metabolism. Trends Biochem Sci 32: 71-7

Hua X, Yokoyama C, Wu J, Briggs MR, Brown MS, Goldstein JL, Wang X (1993) SREBP-2, a second basic-helix-loop-helix-leucine zipper protein that stimulates transcription by binding to a sterol regulatory element. Proc Natl Acad Sci U S A 90: 11603-7

Huang LS, Kayden H, Sokol RJ, Breslow JL (1991) ApoB gene nonsense and splicing mutations in a compound heterozygote for familial hypobetalipoproteinemia. J Lipid Res 32: 1341-8

Humphries S, Coviello DA, Masturzo P, Balestreri R, Orecchini G, Bertolini S (1991) Variation in the low density lipoprotein receptor gene is associated with differences in plasma low density lipoprotein cholesterol levels in young and old normal individuals from Italy. Arterioscler Thromb 11: 509-16

Hussain MM, Strickland DK, Bakillah A (1999) The mammalian low-density lipoprotein receptor family. Annu Rev Nutr 19: 141-72

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. Science 307:1072-1079.

Innerarity TL, Mahley RW, Weisgraber KH, Bersot TP, Krauss RM, Vega GL, Grundy SM, Friedl W, Davignon J, McCarthy BJ (1990) Familial defective apolipoprotein B-100: a mutation of apolipoprotein B that causes hypercholesterolemia. J Lipid Res 31: 1337-49

Jacobsen N, Fenger M, Bentzen J, Rasmussen SL, Jakobsen MH, Fenstholt J, Skouv J (2002) Genotyping of the apolipoprotein B R3500Q mutation using immobilized locked nucleic acid capture probes. Clin Chem 48: 657-60

Kakoki M, Tsai YS, Kim HS, Hatada S, Ciavatta DJ, Takahashi N, Arnold LW, Maeda N, Smithies O (2004) Altering the expression in mice of genes by modifying their 3' regions. Dev Cell 6:597-606

Knoblauch H, Bauerfeind A, Krahenbuhl C, Daury A, Rohde K, Bejanin S, Essioux L, Schuster H, Luft FC, Reich JG (2002) Common haplotypes in five genes influence genetic

variance of LDL and HDL cholesterol in the general population. Hum Mol Genet 11: 1477-85

Knoblauch H, Bauerfeind A, Toliat MR, Becker C, Luganskaja T, Gunther UP, Rohde K, Schuster H, Junghans C, Luft FC, Nurnberg P, Reich JG (2004) Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. Hum Mol Genet 13: 993-1004

K Satyamoorthy, S J Samulewicz, L D Thornburg, A Basu, Howe aCC (1977) Identification of an intronic enhancer that nullifies upstream repression of SPARC gene expression. Nucleic Acids Research 25: 3169-3174

Kim HS, Smithies O, Maeda N (1990) A physical map of the human salivary proline-rich protein gene cluster covers over 700 kbp of DNA. Genomics 6: 260-7

Klausen IC, Hansen PS, Gerdes LU, Rudiger N, Gregersen N, Faergeman O (1993) A PvuII polymorphism of the low density lipoprotein receptor gene is not associated with plasma concentrations of low density lipoproteins including LP(a). Hum Genet 91: 193-5

Knouff C, Malloy S, Wilder J, Altenburg MK, Maeda N (2001) Doubling expression of the low density lipoprotein receptor by truncation of the 3'-untranslated region sequence ameliorates type iii hyperlipoproteinemia in mice expressing the human apoe2 isoform. J Biol Chem 276: 3856-62

Kotowski IK, Pertsemlidis A, Luke A, Cooper RS, Vega GL, Cohen JC, Hobbs HH (2006) A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. Am J Hum Genet 78: 410-22

Kypreos K, Li X, Dijk Kv, Havekes L, Zannis V (2003.) Molecular mechanisms of type III hyperlipoproteinemia: The contribution of the carboxy-terminal domain of ApoE can account for the dyslipidemia that is associated with the E2/E2 phenotype. Biochemistry. 42(33): 9841-53

Lagace TA, Curtis DE, Garuti R, McNutt MC, Park SW, Prather HB, Anderson NN, Ho YK, Hammer RE, Horton JD (2006) Secreted PCSK9 decreases the number of LDL receptors in hepatocytes and in livers of parabiotic mice. J Clin Invest 116: 2995-3005

Lancellotti S, Di Leo E, Penacchioni JY, Balli F, Viola L, Bertolini S, Calandra S, Tarugi P (2004) Hypobetalipoproteinemia with an apparently recessive inheritance due to a "de novo" mutation of apolipoprotein B. Biochim Biophys Acta 1688: 61-7

Leren TP (2004) Mutations in the PCSK9 gene in Norwegian subjects with autosomal dominant hypercholesterolemia. Clin Genet 65: 419-22

Levy E, Minnich A, Cacan SL, Thibault L, Giroux LM, Davignon J, Lambert M (1997) Association of an exon 3 mutation (Trp66-->Gly) of the LDL receptor with variable

expression of familial hypercholesterolemia in a French Canadian family. Biochem Mol Med 60: 59-69

Linton MF, Pierotti V, Young SG (1992) Reading-frame restoration with an apolipoprotein B gene frameshift mutation. Proc Natl Acad Sci U S A 89: 11431-5

Lund-Katz S, Wehrli S, Zaiou M, Newhouse Y, Weisgraber KH, Phillips MC (2001) Effects of polymorphism on the microenvironment of the LDL receptor-binding region of human apoE. J Lipid Res 42: 894-901

Leitersdorf E, Hobbs HH (1988) Human LDL receptor gene: HincII polymorphism detected by gene amplification. Nucleic Acids Res 16: 7215

Maeda N, Bliska JB, Smithies O (1983) Recombination and balanced chromosome polymorphism suggested by DNA sequences 5' to the human delta-globin gene. Proc Natl Acad Sci USA 80:5012-5016.

Maeda N (1991) DNA polymorphisms in the controlling region of the human haptoglobin genes: a molecular explanation for the haptoglobin 2-1 modified phenotype. Am J Hum Genet 49:158-166.

Mack CP, Owens GK (1999) Regulation of smooth muscle alpha-actin expression in vivo is dependent on CArG elements within the 5' and first intron promoter regions. Circ Res 84: 852-61

Martino A, Holmes JHt, Lord JD, Moon JJ, Nelson BH (2001) Stat5 and Sp1 regulate transcription of the cyclin D2 gene in response to IL-2. J Immunol 166: 1723-9

Mahley RW, Ji ZS (1999) Remnant lipoprotein metabolism: key pathways involving cell-surface heparan sulfate proteoglycans and apolipoprotein E. J Lipid Res 40: 1-16

Malloy I AM, Knouff C, Lanningham-Foster L, Parks S, Maeda N (2004) harmful effect of increased LDLR expression in mice with human Apo E*4 but not Apo E*3. Arterioscler Throm Vasc Biol. January 2004:91-97.: 91-97

Maxwell KN, Breslow JL (2004) Adenoviral-mediated expression of Pcsk9 in mice results in a low-density lipoprotein receptor knockout phenotype. Proc Natl Acad Sci U S A 101: 7100-5

Maxwell KN, Fisher EA, Breslow JL (2005) Overexpression of PCSK9 accelerates the degradation of the LDLR in a post-endoplasmic reticulum compartment. Proc Natl Acad Sci U S A 102: 2069-74

Melman L, Geuze H, Li Y, McCormick L, Kerkhof PV, Strous G, Schwartz A, Bu G (2002) Proteasome regulates the delivery of LDL receptor-related protein into the degradation pathway. Mol Biol Cell Sep;13(9): 3325-35

Michael K. Altenburg, Sudi IM, Chris Knouff, John SP, Maeda N (2004) ApoE4 trapping by the low density lipoprotein receptor. ELSEVIER 1262: 388-391

Moore JH, Williams SM (2002) New strategies for identifying gene-gene interactions in hypertension. Ann Med 34: 88-95

Muallem H, North KE, Kakoki M, Wojczynski MK, Li X, Grove M, Boerwinkle E, Wilhelmsen KC, Heiss G, Maeda N (2007) Quantitative effects of common genetic variations in the 3'UTR of the human LDL-receptor gene and their associations with plasma lipid levels in the Atherosclerosis Risk in Communities study. Hum Genet 121: 421-31

Naureckiene S, Ma L, Sreekumar K, Purandare U, Lo CF, Huang Y, Chiang LW, Grenier JM, Ozenberger BA, Jacobsen JS, Kennedy JD, DiStefano PS, Wood A, Bingham B (2003) Functional characterization of Narc 1, a novel proteinase related to proteinase K. Arch Biochem Biophys 420: 55-67

Oshima RG, Abrams L, Kulesh D (1990) Activation of an intron enhancer within the keratin 18 gene by expression of c-fos and c-jun in undifferentiated F9 embryonal carcinoma cells. Genes Dev 4: 835-48

Panagiotakos DB, Pitsavos C, Skoumas J, Chrysohoou C, Toutouza M, Stefanadis CI,Toutouzas PK (2003) Importance of LDL/HDL cholesterol ratio as a predictor for coronary heart disease events in patients with heterozygous familial hypercholesterolaemia: a 15-year follow-up (1987-2002). Curr Med Res Opin 19:89-
94.

Pedersen JC, Berg K (1989) Interaction between low density lipoprotein receptor (LDLR) and apolipoprotein E (apoE) alleles contributes to normal variation in lipid level. Clin Genet 35: 331-7

Pedersen JC, Berg K (1990) Gene-gene interaction between the low density lipoprotein receptor and apolipoprotein E loci affects lipid levels. Clin Genet 38: 287-94

Pedersen J, Berg K (1988) Normal DNA polymorphism at the low density lipoprotein receptor (LDLR) locus associated with serum cholesterol level. Clin Genet. 34(5): 306-12

Pereira E, Ferreira R, Hermelin B, Thomas G, Bernard C, Bertrand V, Nassiff H, Mendez del Castillo D, Bereziat G, Benlian P (1995) Recurrent and novel LDL receptor gene mutations causing heterozygous familial hypercholesterolemia in La Habana. Hum Genet 96: 319-22

Poledne R, Pisa Z, Berg K (1993) Normal genetic variation at the low density lipoprotein receptor (LDLR) locus influences cholesterol levels in children. Clin Genet 43: 122-6

Poledne R, Pisa Z, Berg K (1993) Normal genetic variation at the low density lipoprotein receptor (LDLR) locus influences cholesterol levels in children. Clin Genet 43: 122-6

Quinn MT, Parthasarathy S, Fong LG, Steinberg D (1987) Oxidatively modified low density lipoproteins: a potential role in recruitment and retention of monocyte/macrophages during atherogenesis. Proc Natl Acad Sci U S A 84: 2995-8

Qiao L, Maclean PS, Schaack J, Orlicky DJ, Darimont C, Pagliassotti M, Friedman JE, Shao J (2005) C/EBPalpha regulates human adiponectin gene transcription through an intronic enhancer. Diabetes 54: 1744-54

Qin H, Sun Y, Benveniste EN (1999) The transcription factors Sp1, Sp3, and AP-2 are required for constitutive matrix metalloproteinase-2 gene expression in astroglioma cells. J Biol Chem 274: 29130-7

Ranade K, Chang MS, Ting CT, Pei D, Hsiao CF, Olivier M, Pesich R, Hebert J, Chen YD, Dzau VJ, Curb D, Olshen R, Risch N, Cox DR, Botstein D (2001) High-throughput genotyping with single nucleotide polymorphisms. Genome Res 11:1262-1268

Reid LH, Gregg RG, Smithies O, Koller BH (1990) Regulatory elements in the introns of the human HPRT gene are necessary for its expression in embryonic stem cells. Proc Natl Acad Sci U S A 87: 4299-303

Reddy CD, Reddy EP (1989) Differential binding of nuclear factors to the intron 1 sequences containing the transcriptional pause site correlates with c-myb expression. Proc Natl Acad Sci U S A 86: 7326-30

Reid LH, Gregg RG, Smithies O, Koller BH (1990) Regulatory elements in the introns of the human HPRT gene are necessary for its expression in embryonic stem cells. Proc Natl Acad Sci U S A 87: 4299-303

Rippe RA, Lorenzen SI, Brenner DA, Breindl M (1989) Regulatory elements in the 5'-flanking region and the first intron contribute to transcriptional control of the mouse alpha 1 type I collagen gene. Mol Cell Biol 9: 2224-7

Rudel LL, Parks JS, Johnson FL, Babiak J (1986) Low density lipoproteins in atherosclerosis. J Lipid Res 27: 465-74

Saito H, Dhanasekaran P, Baldwin F, Weisgraber KH, Phillips MC, Lund-Katz S (2003) Effects of polymorphism on the lipid interaction of human apolipoprotein E. J Biol Chem 278: 40723-9

Salazar LA, Cavalli SA, Hirata MH, Diament J, Forti N, Giannini SD, Nakandakare ER, Bertolami MC, Hirata RD (2000) Polymorphisms of the low-density lipoprotein receptor gene in Brazilian individuals with heterozygous familial hypercholesterolemia. Braz J Med Biol Res 33: 1301-4

Salazar LA, Hirata MH, Forti N, Diament j, Giannini SD, Jaqueline IS, Hirata RD (2000) *PvuII* intron 15 polymorphism at the LDL receptor gene is associated with differences in serum lipid concentrations in subjects with low and high risk for coronary artery disease from Brazil. Clinica Chimi Acta 293:75-88.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978-989.

Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162-1169.

Smith JR, Osborne TF, Goldstein JL, Brown MS (1990) Identification of nucleotides responsible for enhancer activity of sterol regulatory element in low density lipoprotein receptor gene. J Biol Chem 265: 2306-10

Sturgeon JD, Folsom AR, Bray MS, Boerwinkle E, Ballantyne CM (2005) Apolipoprotein E genotype and incident ischemic stroke: the Atherosclerosis Risk in Communities Study. Stroke 36: 2484-6

Schonfeld G (2003) Familial hypobetalipoproteinemia: a review. J Lipid Res 44: 878-83

Schuster H, Humphries S, Rauh G, Held C, Keller C, Wolfram G, Zollner N (1990) Association of DNA-haplotypes in the human LDL-receptor gene with normal serum cholesterol levels. Clin Genet 38: 401-9

Seidah NG, Benjannet S, Wickham L, Marcinkiewicz J, Jasmin SB, Stifani S, Basak A, Prat A, Chretien M (2003) The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): liver regeneration and neuronal differentiation. Proc Natl Acad Sci U S A 100: 928-33

Sing C, Davignon J (1985) Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. Am J Hum Genet 37: 268–85

Smith JR, Osborne TF, Goldstein JL, Brown MS (1990) Identification of nucleotides responsible for enhancer activity of sterol regulatory element in low density lipoprotein receptor gene. J Biol Chem 265: 2306-10

Soria LF, Ludwig EH, Clarke HR, Vega GL, Grundy SM, McCarthy BJ (1989) Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. Proc Natl Acad Sci U S A 86: 587-91

Soutar AK, Naoumova RP, Traub LM (2003) Genetics, clinical phenotype, and molecular cell biology of autosomal recessive hypercholesterolemia. Arterioscler Thromb Vasc Biol 23: 1963-70

Steinberg D (1996) A docking receptor for HDL cholesterol esters. Science 271: 460-1

Sudhof TC, Goldstein JL, Brown MS, Russell DW (1985) The LDL receptor gene: a mosaic of exons shared with different proteins. Science 228: 815-22

Sudhof TC, Van der Westhuyzen DR, Goldstein JL, Brown MS, Russell DW (1987) Three direct repeats and a TATA-like sequence are required for regulated expression of the human low density lipoprotein receptor gene. J Biol Chem 262: 10773-9

Tall AR (1990) Plasma high density lipoproteins. Metabolism and relationship to atherogenesis. J Clin Invest 86: 379-84

Taylor R JM, Seed M, Humphries S. (1988) Four DNA polymorphisms in the LDL receptor gene: their genetic relationship and use in the study of variation at the LDL receptor locus. J Med Genet. 1988 Oct;25(10):653-9. . J Med Genet. 25(10): 653-9.

Timms KM, Wagner S, Samuels ME, Forbey K, Goldfine H, Jammulapati S, Skolnick MH, Hopkins PN, Hunt SC, Shattuck DM (2004) A mutation in PCSK9 causing autosomal-dominant hypercholesterolemia in a Utah pedigree. Hum Genet 114: 349-53

Tolleshaug H, Goldstein JL, Schneider WJ, Brown MS (1982) Posttranslational processing of the LDL receptor and its genetic disruption in familial hypercholesterolemia. Cell 30: 715-24

Tybjaerg-Hansen A, Steffensen R, Meinertz H, Schnohr P, Nordestgaard BG (1998) Association of mutations in the apolipoprotein B gene with hypercholesterolemia and the risk of ischemic heart disease. N Engl J Med 338: 1577-84

Templeton AR (1987) The general relationship between average effect and average excess. Genet Res 49: 69-70

The ARIC investigators (1989) The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol 129:687-702.

Van Aalst-Cohen ES, Jansen AC, De Jongh S, De Sauvage Nolting PR, Kastelein JJ (2004) Clinical, diagnostic, and therapeutic aspects of familial hypercholesterolemia. Semin Vasc Med 4: 31-41

Wang QF, Prabhakar S, Wang Q, Moses AM, Chanan S, Brown M, Eisen MB, Cheng JF, Rubin EM, Boffelli D (2006) Primate-specific evolution of an LDLR enhancer. Genome Biol 7: R68

Wang X, Sato R, Brown MS, Hua X, Goldstein JL (1994) SREBP-1, a membrane-bound transcription factor released by sterol-regulated proteolysis. Cell 77: 53-62

Wang S, Zhao H (2003) Sample size needed to detect gene-gene interactions using association designs. Am J Epidemiol 158: 899-914

Weisgraber KH (1994) Apolipoprotein E : structure-function relationship. .Adv protein Chem.: 249-302

Weisgraber KH (1990) Apolipoprotein E distribution among human plasma lipoproteins: role of the cysteine-arginine interchange at residue 112. J Lipid Res 31: 1503-1

Wilson GM, Vasa MZ, Deeley RG (1998) Stabilization and cytoskeletal-association of LDL receptor mRNA are mediated by distinct domains in its 3' untranslated region. J Lipid Res 39:1025-1032.

Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, Strait J, Duren WL, Maschio A, Busonero F, Mulas A, Albai G, Swift AJ, Morken MA, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen WM, Li Y, Scott LJ, Scheet PA, Sundvall J, Watanabe RM, Nagaraja R, Ebrahim S, Lawlor DA, Ben-Shlomo Y, Davey-Smith G, Shuldiner AR, Collins R, Bergman RN, Uda M, Tuomilehto J, Cao A, Collins FS, Lakatta E, Lathrop GM, Boehnke M, Schlessinger D, Mohlke KL, Abecasis GR (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet 40: 161-9

Yamamoto T, Davis CG, Brown MS, Schneider WJ, Casey ML, Goldstein JL, Russell DW (1984) The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. Cell 39: 27-38

Zhao Z, Tuakli-Wosornu Y, Lagace TA, Kinch L, Grishin NV, Horton JD, Cohen JC, Hobbs HH (2006) Molecular characterization of loss-of-function mutations in PCSK9 and identification of a compound heterozygote. Am J Hum Genet 79: 514-23

Zhu H, Tucker HM, Grear KE, Simpson JF, Manning AK, Cupples LA, Estus S (2007) A common polymorphism decreases low-density lipoprotein receptor exon 12 splicing efficiency and associates with increased cholesterol. Hum Mol Genet 16: 1765-72