EFFECTS OF COMMON AND RARE GENETIC VARIANTS OF APOLIPOPROTEIN C4 ON HDL-CHOLESTEROL LEVELS

by

Zaheda Hassan Radwan

B.A.M.S., King Faisal University, Saudi Arabia, 2005

Submitted to the Graduate Faculty of

Department of Human Genetics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This Thesis was presented

by

Zaheda Hassan Radwan

It was defended on

December 06, 2010

and approved by

Thesis Advisor:
M. Ilyas Kamboh, Ph.D.
Professor and Department Chair
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
F. Yesim Demirci, M.D.
Assistant Professor
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Clareann H. Bunker, Ph.D., M.P.H.
Associate Professor
Department of Epidemiology
Graduate School of Public Health
University Of Pittsburgh

EFFECTS OF COMMON AND RARE GENETIC VARIANTS OF APOLIPOPROTEIN C4 ON HDL-CHOLESTEROL LEVELS

Zaheda Hassan Radwan, M.S.

University of Pittsburgh, 2010

Coronary heart disease (CHD) is a major public health problem in western countries as it continues to be a leading cause of premature mortality and morbidity. Several risk factors contribute to CHD risk, including dyslipidemia with low high density lipoprotein cholesterol (HDL-C) and high low density lipoprotein cholesterol (LDL-C). Meta analysis of genome wide linkage analysis in families with diverse ethnicity has revealed a strong linkage with different lipid traits on chromosome 19q13.2. There are several candidate genes present under this linkage region, including *APOE/C1/C4/C2* gene cluster. With the exception of *APOE*, other genes in this cluster have not been extensively evaluated in relation to lipid profile. Therefore, identifying *APOC4* genetic variants that modulate HDL-C level is a great public health importance. In this study we focused on the *APOC4* gene and hypothesized that rare and common variants in this gene could affect plasma lipid levels. Integration of common variants common disease (CVCD) and rare variants common disease (RVCD) hypotheses has been conducted in a limited number of studies. The aim of this study was to identify both common and rare variants in *APOC4* by sequencing individuals having extreme low and high HDL-C levels from U.S. non-Hispanic Whites (NHWs) and African Blacks, and to examine their effects on HDL-C and correlated lipid levels. In the sequencing analysis, a total of 65 variants were identified in NHWs and African Blacks. Of these 26 were present in NHWs and 51 in Blacks. Among NHWs, 31% of the low HDL-C group had rare or less common variants versus 10% of the high HDL-C group. On the

other hand, reverse trend was observed in the Black sample (46% of the low HDL-C group versus 54% of the high HDL-C). Screening of these observed rare and common variants in the complete NHWs and Blacks dataset would provide more information about their association with plasma HDL-C and correlated lipid traits.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First I would like to thank Dr. M. Ilyas Kamboh for giving me the opportunity to be involved in this exciting research. I have learned so much from this project and I appreciate that you allowing me part of your lab. I would like to thank Dr. F. Yesim Demirci for her countless effort in guidance and instructions that she has provided during my project. I really appreciate your patience in guiding and mentoring me over the last two years. I would also like to thank Amy Dressen for her assistance on the statistical analysis of this project.

Thank you to my coworkers and fellow students in Dr. Kamboh's lab who I have enjoyed working with over the last few years. I would like to thank, Dr.Irfan and Fahad Waqar in helping me with my training, Dr.Kahalifa and Yuee Wang for their support and answering my questions. Thank you Kate for helping me with the sequencing analysis. Thank you Jatinder so much for being a second mentor in clarifying all the complicated stuffs and I would like to thank Delik for her friendship and support over my studies.

I would like to thank my first family. Thank you very much my mom, dad, sisters and brothers for loves and support. Thank you my mom for all the efforts that you have spent in supporting my educational pursuit and your prayers. Thank you dad for your support, inspirations and motivations. Thanks sister Fatima and Sukainah for your love and support in taking care of my kids. Congratulation Jay for your graduation and wish you the best in your

graduate studies. Thank you my little brothers, Mohd and Safi, I know I was missed out over the last three years in supporting your education. Safi wish you the best in your studies and Mohd hopefully you graduate shortly.

Now it is time to thank my small family. Thank you my husband for your love, support and patience. Thank you Noor and Ali my lovely kids. I know I was missed out so much in raising you up in your early childhood.

Finally, before I forget I would like to thank my government especially King Abdullah in providing me scholarship and giving me this opportunity to pursue my graduate studies in one of the top ten ranked US universities, University of Pittsburgh. Also, I would like to thank Saudi Arabian Cultural Mission (SACM) employees for their countless efforts in advising and cooperation, especially my academic advisor: Mr. Aldehaiman thank you so much for being advisor and consultant at the same time.

1.0 IMPORTANCE AND SIGNIFICANCE

1.1 CARDIOVASCULAR DISEASE

1.1.1 Significance

Coronary heart disease (CHD) continues to be a leading cause of premature mortality and morbidity in western countries. According to the American Heart Association Statistics in 2006, cardiovascular disease (CVD) accounted for 56% of all deaths in the US (American Heart Association, 2009; Tsompanidi et al., 2010). National Center of Health Statistics (NCHS) data in 2006 estimated that 33% of premature deaths in US countries were attributed to CHD. Most strikingly, more than 81 million American adults live with various types of CVD and the estimated direct and indirect health care expenditures for 2006 were $165.4 billion (American Heart Association, 2009).

In addition to being a leading cause of premature mortality, CHD is a major cause of morbidities. CHD is a medical condition in which coronary arteries that carry blood to the heart become hardened and narrowed or blocked the condition being called, atherosclerosis. Atherosclerosis is the major manifestation of CHD and is associated with different kinds of diseases such as, cardiac ischemia (lack of blood supply), necrosis, chest pain and myocardial

infarction. Taking all together, CHD is a major cause of co-morbidities and premature mortality in western countries and thus is a major public health problem.

1.1.2 Cardiovascular Disease Risk Factors

Several risk factors contribute to CHD, including cigarette smoking, abnormal blood lipid levels, hypertension, diabetes, abdominal obesity, lack of physical activity, low daily fruit and vegetable consumption, alcohol overconsumption and psychosocial index (American Heart Association, 2009).

Numerous epidemiological studies and long-term outcomes trials confirm the association between lipid level and CHD risk (Chilton, 2004). It is estimated that lowering cholesterol level reduces the CHD risk by 32% (Ravnskov, 1992). Clinical trials suggested that lowering fat dietary consumption reduces the CHD risk (Mozaffarian et al., 2010). By looking only at cholesterol level as a major CHD risk factor, each lipoprotein particle has a distinct role in determining atherosclerosis risk in which high density lipoprotein cholesterol (HDL-C) has a protective property against atherosclerosis, while low density lipoprotein (LDL-C) increases the atherosclerosis risk.

1.2 LIPID AND CARDIOVASCULAR DISEASE

Atherosclerosis is considered to be a major manifestation of CHD and there is a complicated mechanism underling the pathophysiological development of atherosclerosis. Atherosclerosis is mainly arisen due to lipid accumulation along the coronary artery walls, which results in

blocking the blood flow that leads to several co-morbidities. It has been well established that abnormal plasma lipid profile modulates the risk of CHD in which different lipoprotein particles play a complicated role in developing atherosclerosis (discussed below in the lipoprotein and lipid metabolism section).

The opposite role of HDL-C and LDL-C in determining the CHD risk is well established in which high HDL-C is a protective factor, while increased LDL-C is a risk factor (Gordon et al., 1977). Interestingly 30-50% of patients with CHD have low HDL-C level (<40 mg/dl) (Sharrett et al., 2001). The inverse relationship between HDL-C level and CHD risk is well established. Each 1mg/dl increase in HDL-C decreases the risk of CHD by 2% in men and 3% in women (Gordon et al., 1989). On the other hand, 40% of CHD risk increases with each 38mg/dl increase in LDL-C level (Sharrett et al., 2001). In parallel, with each 10% reduction in LDL-C, the risk of CHD is decreased by 10% (Breil et al., 2009).

CHD prevention approaches are aiming to modulate the cholesterol level either by lowering LDL-C or increasing HDL-C level or combining both approaches together (Ferns and Keti, 2008). Evidence from clinical trial study reveals the effectiveness of using the first approach (reducing LDL-C level) either by using LDL-C lowering drugs or changing the life style associates with reduction in the CHD risk (Sharrett et al., 2001). There is, however, controversy in using the second approach (increasing HDL-C) as CHD prevention measure. On one hand, majority of individuals with CHD have low HDL-C level and high TG level in spite of low LDL-C (Rubins et al., 1995), suggesting that HDL-C is independent predictor of CHD risk and increasing HDL-C might be more effective in CHD prevention. On the other hand, meta analysis of 108 controlled trials using lipid modulating medication shows non-significant association between high HDL-C level and reduction of CHD risk, suggesting that HDL-C is not

an independent protective agent (Briel et al., 2009). Thus, whether lowering LDL-C or raising HDL-C level works independently is still a controversial issue. Thus, combining both approaches (lowering LDL-C and increasing HDL-C) would be the most effective preventive measure.

1.3 LIPOPROTEIN AND LIPID METABOLISM

1.3.1 Lipoprotein Particles

Since lipid particles are water insoluble molecules (hydrophobic), they require special carriers to be transported in the circulation. Small fatty acids are transported through serum albumins while large lipids molecules such as cholesterol, phospholipids, and triglycerides need specialized carriers like lipoprotein particles. Lipoprotein particles are composed of two layers outer-layer (hydrophilic) and inner-layer (hydrophobic). The outer layer is composed of phospholipids, unesterified cholesterol and apolipoproteins, while the inner layer's contents are cholesterol esters and triglycerides (Miles, 2003). The apolipoprotein component functions as modulators that facilitate internalizing the lipid through interaction with specific lipid receptors.

Lipoproteins are classified into six different particles based on their densities and electrophoresis mobility (Mahely et al., 1984). The lipoprotein density is a factor of lipid to protein ratio with highest protein contents having the highest density and the slowest mobility. Lipoprotein particles are classified into six different particles: chylomicron, chylomicron remnants, very low density lipoprotein (VLDL), low density lipoprotein (LDL), intermediate density lipoprotein (IDL) and high density lipoprotein (HDL) (Corella and Ordovas, 2005; Mahely et al., 1984; King, 2010).

The key element of lipoprotein particles is apolipoprotein component that determines the lipoproteins specificity through interaction with specific lipid receptors. Apolipoproteins posses a unique feature that enables lipoproteins to bind to lipid molecules and this feature is called amphipathic. All apolipoproteins have α- amphipathic helixes (both hydrophobic and hydrophilic ends) with a complex organization (Miles, 2003). The hydrophobic end had inward facing, while the hydrophilic end has an outward facing to interacts with the polar heads of phospholipids in order to facilitate internalizing the lipid particles (Miles, 2003).

Apolipoproteins play essential roles in lipid transport and lipid metabolism by acting as ligands for lipid receptors or cofactor for some lipolytic enzymes such as, lipoprotein lipase (LPL), hepatic lipase (HL), cholesteryl ester transfer protein (CETP), or lecithin-cholesterol acyltransferase (LCAT) (Mahely et al., 1984). Apolipoproteins are key elements in regulating lipid level through transporting lipid to the liver, and maintaining vital function through distributing lipid to the cells and utilizing lipid in maintaining the integrity of cell membrane and synthesizing steroid hormones (Mahely et al., 1984). Additional function of apolipoproteins is to maintain the integrity of lipoprotein structure (Mahely et al., 1984). Apolipoproteins modulate their affinities during lipid metabolism and become exchangeable among different lipoprotein particles. There are different classes of apolipoproteins (apo) such as; apoAI, apoAII, apoAIV, apoB48, apoB100, apoCI, apoCII, apoCIII, apoCIV, apoE, apoD, apoF, apoJ, and apoM (Mahely et al., 1984). Since apolipoprotein CIV (apoCIV) belongs to apolipoproteins family, it is more likely involved in lipid metabolism and this is a candidate gene that might modulate HDL-C level.

1.3.2 Lipid Metabolism

The lipid metabolism process is depicted in Figure 1. Lipid metabolism process begins with absorption of the dietary fatty acids in the intestine in Triglycerides (TG) form. The small fatty acids are transported through serum albumins to be taken directly either by metabolically active tissues such as, cardiac and skeletal muscle for energy or by adipose tissues for storage, while large lipid molecules require specialized lipoprotein particles. There are two major pathways of lipid metabolism: exogenous (dietary intake) and endogenous pathway (de novo lipid synthesis).

The exogenous pathway begins once the dietary lipids are absorbed into the intestine and packaged by chylomicron particle to be transported into the circulation (Fitzgerald, 2010). Chylomicron particles are TG-rich with apoB48, apoAI, apoAII and apoAIV as major apolipoproteins (King, 2010). In the circulation, chylomicron undergoes several modifications, under lipoprotein lipase (LPL) action; its TG contents are hydrolyzed into glycerol and three free fatty acids. The free fatty acids are taken up either by muscle tissues for energy or adipose tissues for storage. Consequently, chylomicrons are transferred into smaller particles called chylomicron remnants (Corella and Ordovas, 2005; Mahely et al., 1984; King, 2010). Apolipoproteins are exchanged among lipoprotein particles with more than 90% of apolipoproteins remaining free in the circulation. After hydrolyzing the TG content of chylomicron, apoAI and apoAII are transferred to HDL particle resulting in chylomicron remnants with apoB48 as a major apolipoprotein. Chylomicron remnants acquire apoE from HDL particle, which plays a crucial role in TG clearance through hepatic uptake of chylomicron remnants by LDL receptors (LDL-R) and LDL related protein 1 (LRP1) (Corella and Ordovas, 2005; Mahely et al., 1984; Miles, 2003; King, 2010). After hepatic uptake of chylomicron remnants, the endogenous pathway begins.

In endogenous pathway, VLDL particles are synthesized in the liver by using cholesterol ester (CE) and phospholipids (PL) residues of chyomicron remnants and acquired apoB100, apoE and apoCs as apolipoproteins. Afterward in the circulation, VLDL undergoes similar pathway as chylomicron. VLDL is hydrolyzed by LPL and transferred into VLDL remnants (Corella and Ordovas, 2005; Mahely et al., 1984; Miles, 2003; King, 2010). Small portion of VLDL remnants is catabolized in the hepatic tissue through receptor-mediated mechanism involves remnants receptors, LDL-R and LRP1, while the majority of VLDL remnants transferred into IDL and eventually to LDL. As mentioned before apolipoproteins change their lipoproteins affinity and migrate to another lipoprotein during the lipid metabolism process (Miles, 2003). After LPL-mediated TG lipolysis of VLDL particles, VLDL loses apoCs and apoE that are acquired by HDL with retention of apoB100, the major apolipoprotein component of LDL. There are three major known fates for LDL-C: 1) hepatic clearance, 2) uptaken by extrahepatic tissues or 3) infiltrates into arterial walls and involved in atherosclerosis. Since up to 70% of LDL-R are present in the liver, most LDL is uptaken by the liver and part of the remainder LDL is taken up by extrahepatic tissues for cellular membrane maintenance and steroid hormone synthesis (Grundy, 1983). Hepatic uptake of LDL cholesterol depends on the numbers of LDL-R expressed on the hepatic cells. The magnitude of intracellular cholesterol level down regulates the number of LDL-R (Grundy, 1983). Thus, there is negative feedback in which high intracellular cholesterol cuts down the number of LDL-R, which results in reducing the hepatic clearance of LDL-C and increasing the cholesterol level in the plasma. Eventually, presence of LDL-C at high level in the plasma results in penetration the arterial walls and accumulation of LDL-C on the intima, which initiates the atherosclerosis.

The atherogenic property of LDL-C is well documented and several randomized control trials and epidemiological studies confirm the direct relationship between LDL-C and CHD. High LDL-C is present in approximately 40-50% of patients with CHD (Griffin, 1999). The physical feature of LDL with a small molecular size enables LDL-C particle to penetrate the arterial walls (Grundy, 1983). Accumulation of LDL on the arterial walls is known as proatherogenic event. Under normal condition, no LDL-C is transferred to the arterial walls because majority of LDL-C is transported to the liver or into HDL-C. In individuals with high LDL-C, LDL-C accumulates on the arterial walls and once it is internalized, it gets oxidized (Adames et al., 2000). Oxidation process leads to the activation of endothelial region and triggers monocytes attraction. Monocytes modify into macrophages that engulf the oxidized LDL and transfer into foam cells. Then, inflammatory process takes place that results in severe damage and initiates the pathogenesis of atherosclerosis (Adames et al., 2000; Ross, 1999; Life Extension Foundation, 2010).

Even though LDL is a major proatherogenic factor, other TG-rich lipoproteins (TGRL) have similar role in initiation atheroscerosis. Animal studies show association between high dietary intake of cholesterol and low hepatic uptake in rabbit and this association could be involved in atherosclerosis (Grundy, 1983). However, chylomicron remnants present at low concentration in humans due to maximum level of hepatic clearance, any defect in the hepatic uptake mechanism could result in accumulation of chylomicron remnants, which might lead to atherosclerosis (Grundy, 1983).

Numerous *in vitro* studies showed the potential contribution of TGRL in proatherosclerosis mechanism. Several hypotheses presume TGRL involvement in atherosclerosis. Findings from cell culture study confirm the cytotoxic effects of TGRL on

endothelial cells (Spiedel et al., 1990). Furthermore, another *in vitro* study concluded that the end product of TGRL hydrolysis (free fatty acid) increases endothelial cells permeability and eventually increases LDL retention in the intima (Hennig et al., 1985). The relatively small molecular size of VLDL, VLDL remnants and LDL enable them to penetrate and being trapped into arterial intima (Hyson et al., 2003). Findings from various *in vitro* studies confirm the involvement of TGRL and its products in impairing the endothelial cells integrity and initiating atherosclerosis (Hyson et al., 2003). Another study concluded a direct correlation between triglyceride and LDL-C, which supports TGRL involvement in atherosclerosis (Krauss, 1991).

Furthermore, TGRL promotes atherosclerosis not only by increasing LDL-C level but also by impairing the efficiency of RCT (reverse cholesterol transport) function of HDL particle. Patients with Tangier's syndrome have low HDL-C and high VLDL due to mutation in ATP-binding cassette sub-family A (ABCA1) gene (Corella and Ordovas, 2005). The possible mechanism underlying low HDL-C in Tangier's disease is that High level of TG stimulates CETP to transfer TG from VLDL to LDL and to HDL that results in inefficient HDL particles. TG-rich HDL has an inefficient RCT function which accelerates the onset of atherosclerosis (Corella and Ordovas, 2005).

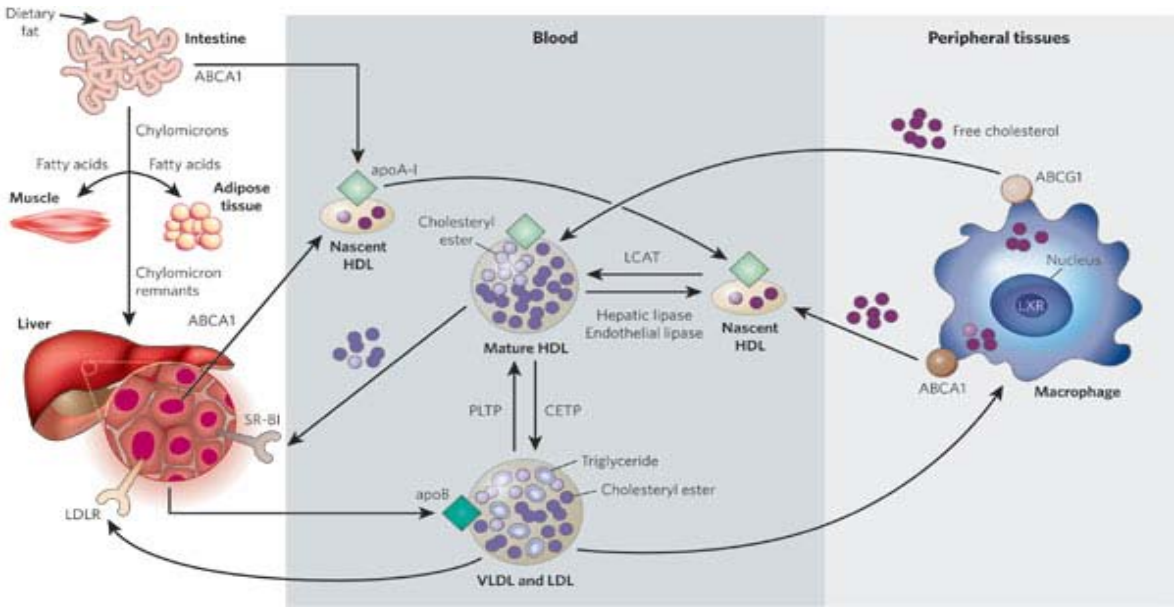

Figure 1 Lipoprotein Roles in Lipid Metabolism and Atherosclerosis (with permission from Rader and Daugherty, 2008)

1.4 HIGH DENSITY LIPOPROTEIN CHOLESTEROL (HDL-C)

HDL lipoprotein has the highest density with 45-55% apolipoproteins, 22-36% phospholipids, 15-20% esterified cholesterol, 3-5% unesterified cholesterol and approximately 5% triglycerides. ApoAI and apoAII are the major HDL apolipoproteins with a crucial function in HDL metabolism. HDL also contains other apolipoproteins such as, apoCs and apoE (Tsompanidi et al., 2010).

There are two main sources of HDL; de novo synthesis in the hepatic cells and intestine, or generated from other lipoprotins mainly LDL. HDL-C metabolism as illustrated by Fitzgerald (2010) in abcam website shows that HDL-C is generated primarily from other lipoproteins after

transferring the cholesterol and phospholipids into apoAI and forming pre B-HDL particle. ATP-binding cassette (ABCA1) is the major catalyst of this process. Furthermore, LCAT enzyme plays important role in esterifying the cholesterol content of the nascent HDL and transferring the nascent HDL (disc-shaped) to a mature HDL (spherical-shaped) (Corella and Ordovas, 2005; Fitzgerald, 2010).

The cholesterol content of HDL is taken up by the hepatic tissue either directly through scavenger receptors B1 (SR-B1) without internalizing HDL particle or indirectly through transferring cholesterol to other lipoprotein particles (VLDL and LDL) through LDL-receptor or LRP1-mediated hepatic uptake (King, 2010)). HDL-C is transferred to VLDL and LDL under action of HDL-related enzymes such as CETP (King, 2010).

1.4.1 Antiatherogenic Feature of HDL-C

Epidemiological studies confirm the inverse association between HDL-C and CHD. There are several hypotheses that explain the potential antiatherogenic property of HDL-C. Involvement of HDL in several mechanisms explains its atheroprotective function, including reverse cholesterol transport (RCT), antioxidant function and anti inflammatory function.

One of the hypotheses is involvement of HDL in RCT. In RCT mechanism, HDL-C transports the cholesterol from extrahepatic tissue, including macrophages in the atherosclerotic lesion site to the liver for catabolism. Thus, RCT mechanism prevents accumulation of cholesterol in the extrahepatic tissue and arterial walls and in turn inhibits atherosclerosis initiation event (Tsompanidi et al., 2010; Assman et al., 1996). It is hypothesized that LCAT enzyme, that is activated by apoAI enhances the reverse cholesterol transport mechanism

through esterification of free cholestrol. Then, the esterfied cholesterol gets internalized into HDL lipid core and eventually catabolized in the liver, which results in alteration of the HDL particle configuration from disc-like to a spherical-shape particle. So, LCAT is involved in modifying a nascent HDL into mature HDL-C (Kris-Etherton and Etherton, 1982).

In addition to the RCT mechanism, HDL has antioxidant property that provides direct or indirect protection of LDL from oxidation. This antioxidant property is mainly due to presence of apoAI and antioxidant enzymes. *In vitro* studies shows that apoAI inhibits LDL oxidation either by making LDL molecule resistant to lipoxygenase enzyme which involves in fatty acid oxidation, or by removing oxidation prone molecules from LDL and make it resistant to be oxidized. Additional molecule participates into antioxidant function is antioxidant enzymes such as, paraoxonase (PON) and acetyl-hydrolase platelet activation factor (PAFAH), which inhibit LDL oxidation (Corella and Ordovas, 2005; Tsompanidi et al., 2010).

Additionally, HDL has antiinflammatory feature that provides protection from inflammation. Several *in vitro* studies show that HDL inhibits expression of pro inflammatory adhesion molecules (such as C-reactive protein and prostaglandins from monocytes) and stimulates expression of transforming growth factor beta 2 (TGFβ2). Additional *in vitro* studies concluded that involvement of HDL in stimulating endothelial nitric oxide synthase (eNOS), which functions as vascular relaxant (Tsompanidi et al., 2010). Furthermore, HDL has a major role in clearing away all the atherogenic agents. Phenotypic characteristic of Tangier's syndrome with high TG and low HDL-C explains the major role of HDL in maintaining TGRL homeostasis. Taking all these findings together explain the antiatherogenic role of HDL (Corella and Ordovas, 2005).

Since elevated level of HDL and lower level of LDL predicts low risk of CHD, most therapeutic and prevention approaches targeting increasing HDL cholesterol level to minimize CHD risk.

1.4.2 Genetics of HDL-Cholesterol

The well-established inverse association between HDL-C and CHD risk provides compelling rational in identifying the genetic basis of HDL-C. Several family and twin studies have confirmed the genetic basis of HDL-C with an estimated heritability of 40-60% (Qasim and Rader, 2006; Lusis et al., 2004). To a lesser extent additional factors explain the remaining variation in HDL-C level, including age, gender, obesity, physical activity, diet, smoking, alcohol consumption and other metabolic disorder such as diabetes mellitus and liver diseases. The variation in HDL-C is a complex trait in which multifactorial genetic-environmental interaction takes place in modulating HDL-C level.

Several linkage and association studies have been conducted over the last 30 years in order to determine the genetic basis of HDL-C level. Several monogenic susceptible loci that influence the high and low level of HDL-C have been identified. Familial hypoalphalipoproteinemia (FHA) is the most common monogenic disorder with extremely low level of HDL-C and apoAI. However, the monogenic susceptible loci, which have been identified, explain only 1% of HDL-C heritability (Miller and Zhan, 2004). Since HDL-C is a complex trait, there is a complicated mechanism influencing the variation of HDL-C level. There are two major approaches used to identify susceptible loci for complex diseases, including candidate gene studies and genome wide studies.

Candidate gene study is a hypothesis-based approach, which can be classified further into association studies and resequencing association studies (Weisglass-Volkov and Pajukanta, 2010). Association studies compare the allele or genotype frequency in cases and control. Genes for association studies are chosen based on their location (under the linkage peak) or based on their biological role in the disease pathogenesis. Thus, we can classify candidate gene association studies into positional or biological candidate genes. The major difference between candidate genes association studies and resequencing association studies is the resolution (ability to detect common or rare variants). In candidate gene association studies, only common genetic variants with MAF≥ 5% are genotyped, which is known as linkage disequilibrium based studi es (LD-based studies). Only tag SNPs (the representative SNPs that are in LD with other proxies) that could give us information about their proxies are genotyped. In contrast to candidate gene association studies, the resequencing approach is LD-based free association studies in which the entire gene is sequenced to overcome the candidate gene association studies limitations in detecting rare variants with MAF<5%. Resequencing enables researchers to detect both common and rare variants. Although resequencing association studies have better resolution in identifying rare variants, it is an expensive and laborious approach (Weisglass-Volkov and Pajukanta, 2010).

Many candidate gene association studies have been conducted over the past years (Weisglass-Volkov and Pajukanta, 2010). Numerous genes involved in lipid metabolism (includes: apolipoproteins, enzymes, receptors, lipid transfer protein, transporters and transcription factors) have been tested and yielded inconsistent results. The major reason behind inconsistency among various studies is insufficient sample size, which complicates identifying common SNPs with small to moderate effect sizes (Weisglass-Volkov and Pajukanta, 2010). Several susceptible loci have been identified using the candidate genes approach; *CETP*, *LIPC*,

LPL, *LCAT*, *ABCA1*, *APOA1*, *APOC3*, *APOA5*, *APOE*, *SR-B1* and *PON1* (Weisglass-Volkov and Pajukanta, 2010; Sviridov and Nestel, 2007; Klos and Kullo, 2007; Boes et al., 2009; Pollex and Hegele, 2007). Several candidate genes studies emphasized the genetic effect of some apolipoproteins such as *APOA2, APOA4, APOE* and *APOB* (Boes et al., 2009) but little attention has been paid to *APOC4,* which we have investigated in this study.

The second category of genetic studies is genome wide study (GWS) that is also further classified into association or linkage studies. In GWS, common variants with MAF $\geq$ 5% from the whole genome are genotyped under hypothesis-free settings. The major difference between genome wide association studies (GWAS) and genome wide linkage studies (GWLS) is the target population. GWAS are a population-based approach, while GWLS studies are family-based approach (Weisglass-Volkov and Pajukanta, 2010).

Linkage studies test the co-segregation of chromosomal regions among family members. Linkage studies are useful in localizing the causal variant of rare monogenic disorders but it is not applicable for complex diseases (Rust et al., 1998). Association studies are more powerful in detecting the genetic basis of complex diseases than linkage analysis in having better resolution and using fewer number of markers than linkage analysis (Weisglass-Volkov and Pajukanta, 2010). Linkage with HDL-C has been reported on different chromosomes, including 9p (Arya et al., 2002), and chromosome 7 (Adeyemo et al., 2005) in Mexican American families. In Utah families, 11q23 was linked with HDL-C, while 8q and 15 were linked to HDL-C in San Antonio Hispanic families (Kort et al., 2000). Recent genome wide linkage study revealed significant linkage peak at chromosome 19p13 with a LOD score of 1.14 in population of Asian ancestry (Park, 2008). Meta analysis of Linkage studies on families with diverse ethnicity revealed a broad linkage peak on chromosome 19 (19p13-19q13.24) with different lipid traits, which

suggest that complex genetic interaction among different genes in this linkage region might have direct impact on modulating lipid profile (Malhotra et al., 2007). Most interestingly, *APOE/C1/C4/C2* genes lie under this linkage region.

GWAS opens a great opportunity for researchers to identify novel susceptible loci with no prior knowledge of their biological involvement in the etiology of complex disease pathogenesis. GWAS is a hypothesis free approach in which the common variants with MAF ≥ 5% in the genome are genotyped. Although GWAS have discovered numerous susceptible loci influence HDL-C level variation, they explain only 10% of HDL-C heritability (Kathiresan et al., 2009). There are numerous assumptions explaining the GWAS missing heritability. One assumption is that the most common genetic variants have a small to moderate effect size that could not pass the GWAS significance level (p-value $<5x10^{-8}$). Another assumption is that rare variants with a large effect size might account for a greater portion of HDL-C heritability and GWAS miss identifying rare variants because it is based on common variants common disease hypothesis (CVCD). Thus, integration of common variants common diseases (CVCD) and multiple rare variants common diseases (MRVCD) hypotheses would enable researchers to identify both common and rare variants.

To date, GWAS have discovered around ~ 40 loci influencing lipid level and about 16 loci for HDL-C in European populations (Willer et al., 2008; Kathiresan et al., 2009; Aulchenko et al., 2009 and Sabatti et al., 2009). HDL-C susceptible loci are within or near these genes: *ABCA1,* angiopoietin-like-4 *(ANGPTLA), APOA1/C3/A4/A5* genes cluster, *APOB, CETP,* free fatty acids desaturase (*FADS (1/2/3)),* N- acetylgalactosaminyltransferase 2 *(GALNT2),* hepatic nuclear factor *(HNF4A), LCAT, LIPC, LIPG, LPL,* MAP kinase activating death domain-folate hydrolase-nuclear oxysterol receptors (*MADD-FOLHI-NRIH3),* methylmalonic aciduria (*MVK-*

MMAB), phospholipid transfer protein *(PLTP)* and tetratricopeptide repeat domain 39B *(TTC39B)* (Kathiresan et al., 2009).

Susceptible loci influence HDL-C according to their role in HDL-C metabolism. According to Rader and Maugeals (2000), some genes with known role in accelerating reverse cholesterol transport (RCT) function (such as; *ABCA1, LCAT, PLTP* and *LPL*) result in increasing HDL-C, while other genes that participating in hepatic uptake of HDL-C and catabolism such as (*SR-B1, HL* and endothelial lipase) result in decreasing the HDL-C level.

1.5 *APOC4* GENE

1.5.1 Gene Structure

APOC4 was identified and characterized by Allan et al (1995) within the *APOE/C1/C4/C2* gene cluster on chromosome 19q13.2. *APOE/C1/C4/C2* is a 48 kb gene cluster in which *APOC1* is located either 4.3 or 5.3 kb downstream from *APOE* gene with the same transcriptional orientation (Davison et al., 1986; Myklibost and Rogene, 1986; Lauer et al., 1988; Li et al., 1988). There is one copy of *APOC1* gene that is located 7.5 kb from *APOC1* with no mRNA product in any tissue called, pseudo *APOC1* gene. *APOC4* is located 555 bp upstream from *APOC2* with the same transcriptional orientation (Allan et al., 1995).

APOC4 gene spans a 3.3 kb and composed of three exons and two introns (Figure 2). *APOC4* sequence encodes 127 amino acids with 25 residues in the signal peptides (Allan et al., 1995). The first exon encodes most of the signal peptides while the second one encodes the C-

terminal region. Each of the second and third exon has amphipathic helical structure, lipid-binding domain (Zhang et al., 1996). Exon 1 encodes for amino acid 1-25, exon 2 encodes for amino acid 26-73 and exon 3 encodes for the remaining amino acid (Kotite et al., 2003).

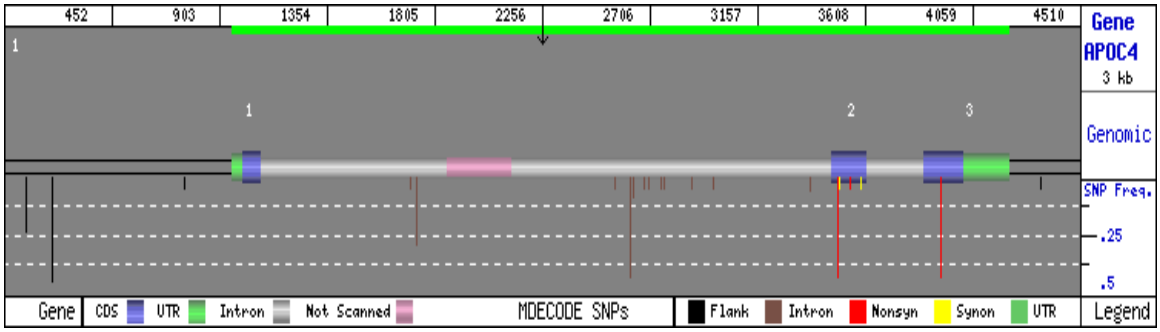

Figure 2 *APOC4* Gene Structure with Three Exons and Two Introns (adopted from SeattleSNPs database)

The expression of all four genes in this cluster, *APOE/C1/C4/C2,* is regulated by two (600 bp) hepatic control elements (HCE-I and HCE-II) (Allan et al., 1995; Simonet et al., 1993). These two hepatic control elements work independently; presence of one HCE is enough for sufficient expression of each gene in this cluster (Shachter et al., 1993; Simonet et al., 1993; Allan et al., 1995; Allan et al., 1997).

APOC4 is expressed only in the liver with undetectable level in the plasma. RNase protection analysis, confirmed the lower expression of *APOC4* in the liver with approximately 100-fold lower than the expression of *APOC2* gene (Allan et al., 1995). In comparison with the mice *Apoc4* gene sequence, the human *APOC4* promoter lacks the typical TATA box motifs. So, presence of transcription factor binding sites (Sp1) is necessary for gene expression. At least

three Sp1 are required for binding with the transcription factor to initiate transcription process (Allan et al., 1995). Moreover, human *APOC4* gene contains five Sp1 like-motif sequences with GGAGGG instead of GGCGGG (with C>A substitution). It has been found that (C>A) substitution in Sp1 like motif associates with three-fold reduction in the binding affinity of Sp1 motif to the transcription factor, which partially explains the lower expression rate of *APOC4* mRNA in the liver (Allan et al., 1995).

Zhang et al. (1996) reported the lower expression rate of *APOC4* in human liver compared to other species such as rabbits. It has been found that the Rabbit *Apoc4* gene is similar to human *APOC4* gene in lacking a typical TATA box in the 5'flanking region with one exception (Kotite et al., 2003). Rabbit *Apoc4* gene contains a purine-rich sequence GGGACAG(G/A) repeated nine times in tandem within the 5'UTR with two additional repeats in the 5' flanking region, whereas human and mouse *APOC4* gene contains only one copy of this purine-rich sequence (Kotite et al., 2003). In human *APOC4*, the purine-rich sequence is located at position 527-533bp in the 5'flanking region according to reference sequence of *APOC4* in SeattleSNPs database. Since purine-rich sequence is involved in transcription of numerous genes, the higher number of purine-rich sequence repeats in rabbit explains the higher expression level of *Apoc4* gene in rabbit than in human (Kotite et al., 2003).

There are several hypotheses explaining the low expression rate of *APOC4* mRNA in human liver compared to the other apolipoproteins. Lacking the TATA box motif, weak Sp1 motifs and presence of only one copy of the purine-rich sequence, as described above, explain part of the lower expression rate of *APOC4* gene in human liver compared to the rabbit.

SDS-PAGE analysis revealed presence of four different isoforms of *APOC4* with 12.5-19 kDa. This high molecular mass is a result of complex N-linked oligosccharides with variable

sialylations (Allan et al., 1996). ApoCIV protein structure resembles other apolipoproteins structure in having two amphipathic a-helical structures (lipid binding domain). Segrest et al (1994), localized the lipid binding domain in the exchangeable apoplipoproteins by using LOCATE program. They found that apolipoproteins contain a special amphipathic feature that facilitates lipid-binding function. Class A1 helical structure is located within the second exon between residues (47-74) in which the positively charged proteins arranged as polar-nonpolar interface and negatively charged proteins present at the center of polar face. The second helical structure is class Y, which is located within the thirds exon between (95-116) residues. In the Y class helical domain, the positively charged proteins present at the center of polar interface (Segrest et al., 1994; Allan et al., 1995). ApoCIV unique structure suggests its involvement in lipid metabolism.

1.5.2 Biological Function of ApoCIV

In vitro studies have shown several regulatory effects of apoCs on different enzymes and receptors involved in lipid metabolism (Jong et al., 1999). Conservation of *APOC4* genomic and protein structure among different species implies the significant role of apoCIV in human (Allan et al., 1995). ApoCIV protein structure with two amphipathic helical domains suggests its involvement in lipid metabolism (Allan at al., 1995). Moreover, Kim et al. (2008) reported that overexpression of *APOC4* in patients with hepatitis C virus infection (HCV) and those patients manifest liver steatosis, suggesting that overexpression of *APOC4* interferes with TG metabolism which results in TG accumulation and liver steatosis (Kim et al., 2008).

ApoCs are present on VLDL and HDL lipoproteins. In postprandial condition, apoCs are transfered from HDL to chylomicron and VLDL (Jong et al., 1999). In normotriglyceridemic

condition, more than 80% of apoCIV is present on VLDL (representing 0.7% of apo VLDL) and most of the remainder is on HDL, while apoCI and apoCII distribute equally on both HDL and VLDL (Kashyap et al., 1977; Schonfeld et al., 1979). Overexpression of *APOC4* in transgenic mice causes hypertrigylceridemia suggests a direct association between apoCIV and TG metabolism (Allan et al., 1996). This phenotype resembles the resulting phenotype of overexpression other apoCs (apoC1, C2 and C3) (Aalto-Setala et al., 1992; deSilva et al., 1994; Shachter et al., 1994; Simonet et al., 1991). The similar hypertrigylceridemic phenotype of expression all apoCs suggests that all human apoCs have a common underlying mechanism in trigylceride metabolism (Allan et al., 1996). There are two assumptions behind accumulation of TG-rich lipoprotein (VLDL) in *APOC4* transgenic mice: 1) impairing hepatic uptake of VLDL through interfering with apoE mediated cellular uptake, and 2) impairing the TG lipolysis mechanism through interfering with LPL or HL activity (Allan et al., 1996).

Numerous evidences from several studies support the first assumption behind VLDL accumulation, impairing the hepatic uptake of VLDL. In transgenic mice, overexpression of *APOC2* and *APOC3* is associated with elevated TG level with minimum changes on total cholesterol level due to interference with hepatic uptake of TG-rich lipoprotein (VLDL) (Aalto-Setala et al., 1992; de Silva et al., 1994; Shachter et al., 1994). Furthermore, in vitro studies have reported that *APOC1* and *APOC2* overexpression impairs VLDL binding to LRP1 receptors through displacing apoE (Swaney and Weisgraber, 1994; Weisgraber et al., 1990). Similar findings from other studies confirm that expression of apoCs interferes with apoE-mediated hepatic uptake of VLDL and LDL through impairing the interaction between apoE and LDL receptor (Sehayek and Eisenberg, 1991). In parallel, the findings of hypertrigylceridemic phenotype in *APOC4* transgenic mice without change in LPL and hepatic lipase activity,

supports the first assumption behind VLDL accumulation (Allan et al., 1996). Most probably, the potential underlying mechanism behind VLDL accumulation is related to impairing apoE-mediated hepatic uptake of VLDL through interaction with LDL-R or LRP1 receptors without disturbing triglyceride lipolysis mechanism (Allan et al., 1996).

Conservation of *APOC4* coding region sequences among different species suggests the significant function of apoCIV. It is assumed that expression of *APOC4* is associated with the developmental stage and disease status, suggesting that the expression of *APOC4* is low under normal condition and could be increased under certain condition such as acute illness (Allan et al., 1996). Overexpression of *APOC4* in patients with HCV infection supports the possibility of increasing *APOC4* expression under acute illness (Kim et al., 2008). Additional evidence from in vitro study suggests the potential role of apoCIV in lipid metabolism and atherosclerosis. Expression of *ApoE/C1/C4/C2* genes cluster in lipid-loaded macrophages, suggests their potential antiatherogenic role in cholesterol efflux in a manner similar to the RCT mechanism (Mak et al., 2002).

1.5.3 Genetic of *APOC4*

According to Chip bioinformatics database, about 180 SNPs (single nucleotide polymorphisms) have been identified in the *APOC4* gene. Although several studies have been conducted to identify genetic variants in the entire *APOE/C1/C4/C2* genes cluster, a little attention has been paid to examine the extent of genetic variants in the *APOC4* gene individually.

Genome wide linkage scans in the Strong Heart Family Study (SHFS) identified strong linkage signal with LDL-C on chromosome 19q13.41 with LOD score of 4.3 (North et al., 2006). Several candidate genes are present under this linkage region, including *APOE/C1/C4/C2* gene cluster. Most interestingly, no significant association was found between *APOE* genetic variants and LDL-C in SHFS, suggesting that other genetic variants in nearby genes could be responsible for the linkage signal (North et al., 2006). In parallel, meta-analysis of linkage studies performed on families with type II diabetes mellitus revealed strong linkage with multiple lipid traits at this locus on ch19q13.13-13.43 in populations with diverse ethnic history, including total cholesterol and LDL-C (Malhorta et al., 2007). In the same study they found broad linkage peak on chromosome19 (19p12-q13.13), which could explain the complexity of the genetic of lipid in which multiple genes interact with each other and modulate lipoproteins concentration (Malhorta et al., 2007).

Moreover, GWAS in 2008 confirmed association signal of (rs4420638, *APOC1*) at *APOE/C1/C4/C2* genes cluster with LDL-C ($p=1x10^{-60}$) (Kathiresan et al., 2008). Several hypotheses explain the difficulties of identifying functional alleles in GWAS studies. GWAS is unable to detect rare variants with modest or high effect size, and common variants with small to modest effect size. It is possible that there are multiple rare variants in *APOC4* that are associated with lipid profile but they were not identified in GWAS. For this reason, we are integrating CVCD and MRVCD hypotheses to detect both common and rare variants in the *APOC4* gene.

To our knowledge, there is only one previous study aimed to detect genetic variants in the *APOC4* gene. Kamboh et al. (2000) identified five SNPs in the coding regions and their exon-intron boundaries of the *APOC4*. Two SNPs were at nt.968 and nt.979 in non-coding part of the

first exon, two SNPs were at codon 36 and 52 in the second exon, and one was at codon 96 in exon 3. Significant association was found between Pro36Leu and Leu96Arg and elevated TG level in NHW women (p= 0.03 and 0.08, respectively). Another association was found between SNP at nt.979 and HDL-C level in NHW men and Lp (a) levels in NHW women with p= 0.05. These *APOC4* genetic variants explained only 2% of the TG variation (Kamboh et al., 2000).

1.6 SPECIFIC AIMS

The objective of this study was to examine the role of complete genetic variants in the *APOC4* gene in relation to plasma lipid profile in two racial groups comprising non-Hispanic whites (NHWs) from the US and African Blacks from Nigeria. Since *APOC4* is a candidate gene, we tested both CVCD and MRVCD hypotheses in order to identify both common and rare variants. The objectives of this study were fulfilled by carrying out the following aims:

Aim 1:

Resequence the entire *APOC4* gene in individuals with extreme HDL-C levels falling in the upper fifth percentile (47 whites, 48 Blacks) and lower fifth percentile (48 Whites, 47 Blacks) to detect both rare (MAF < 5%) and common (MAF ≥ 5%) variants.

Aim 2:

Examine the distribution of common and rare variants identified in Aim 1 between the high and low HDL-C groups.

Aim 3:

Genotype *APOC4* common tag SNPs in the entire samples of NHW (n=623) and Black (n=788), and to examine their association with plasma lipid profile.

2.0 SUBJECTS AND METHODS

2.1 SUBJECTS

2.1.1 Study Population

We used two well-characterized epidemiological samples, including non-Hispanic whites (NHWs) (n=623) and African Blacks (n=788). NHW samples were collected as part of the San Luis valley Diabetes Study (SLVDS). SLVDS was established as geographical based case-control study of non-insulin dependant diabetes mellitus and cardiovascular disease in Alamosa and Conejos counties of south Colorado (Hamman et al., 1989). The demographic and health data and other confounding factors such as dietary intake, smoking, alcohol consumption and physical activity were collected from all participants. The basic characteristics of this study are described in Rewers et al. (1993) and Hamman et al. (1989). Samples of African Black were recruited in Benin City, Nigeria as part of study on CHD risk factors in Blacks. Participants' demographic and health information were collected and more information can be found in Bunker et al. (1995, 1996). The demographic characteristics of both populations are summarized in Table 1.

Esterase-oxidase method was used to measure total cholesterol level. Total HDL-C and LDL-C level were determined enzymatically after dextran sulfate magnesium precipitation

(Harris et al., 1998; Richmond, 1973), and trigylceride level was determined enzymatically by using Stavropoulous and Crouch procedure (Stavropoulous and Crouch, 1974). The DNA used for sequencing and TaqMan genotyping was extracted from clot sample in Blacks and from Buffy coat in NHWs using standard DNA extraction procedures.

Table 1 Demographic Characteristic of Study Populations

Variable	NHWs (n=623)		African Blacks (n=788)	
	Men	Women	Men	Women
Sample size (n)	295	328	495	293
Age (Yrs)	52.9 ± 0.6	52.4 ± 0.6	42.5 ± 0.4	38.7 ± 0.4
BMI (kg/m^2)	26.2 ± 0.2	24.8 ± 0.2	22.0 ± 0.2	24.3 ± 0.3
LDL-C (mg/dl)	139.8 ± 2.0	134.7 ± 2.0	104.7 ± 0.1	117.3 ± 0.1
HDL-C (mg/dl)	43.9 ± 0.6	56.3 ± 0.7	45.9 ± 0.6	50.6 ± 0.7
Triglycerides (mg/dl)	147.6 ± 4.1	128.2 ± 2.9	77.8 ± 1.7	62.8 ± 1.4
Total cholesterol (mg/dl)	213.7 ± 2.2	217.7 ± 2.1	167.7 ± 1.6	181.3 ± 2.2

2.1.2 Resequencing Samples

Ninety-five NHWs (47 individuals having high HDL-C and 48 having low HDL-C) and ninety-five African Blacks (48 having high HDL-C and 47 having low HDL-C) were selected for *APOC4* gene resequencing to detect both rare and common variants. The characteristics of subjects selected for resequencing are summarized in Table 2.

Table 2 Characteristics of *APOC4* Resequencing Samples

	NHWs (n=95)			African Blacks (n=95)		
	(high HDL-C) (n=47)	(low HDL-C) (n=48)	P-value	(high HDL-C) (n=48)	(low HDL-C) (n=47)	P-value
Sex (M/F)	24/23	24/24	0.92	24/24	23/24	0.92
Age (Yrs)	55.45 ± 9.8	53.03 ± 10.54	0.25	41.26 ± 8.72	40.87 ± 7.16	0.8
BMI ($kg/m^{2)}$	23.17 ± 3.17	27.35 ± 3.90	<0.001	22.06 ± 4.71	23.91 ± 5.51	0.08
TOTAL-C (mg/dl)	227.34 ± 51.76	208.81 ± 44.65	0.06	201 ± 39.68	141.68 ±31.03	<0.001
LDL-C (mg/dl)	126.84 ± 46.95	136.95 ± 41.28	0.28	112.55 ± 39.75	95.04 ± 28.28	0.02
TG (mg/dl)	114.09 ± 60.88	240.21±153.22	<0.001	61.98 ± 19.85	95.79 ± 73.21	0.003
HDL-C (mg/dl)	77.68 ± 13.32	31.81 ± 4.37	<0.001	76.05 ± 7.53	25.51 ± 5.66	<0.001

2.2 METHODS AND MATERIAL

2.2.1 PCR Amplification and Sequencing

Publicly available information at SeattleSNPs database (http://pga.mbt.washington.edu/) was used to order M13 tagged primers, which generated seven overlapping fragments. These primary Seattle primers cover genomic region of 4,510 bp that includes 956 bp in the 5'flanking region, all 3 exons (614 bp), 2 introns (2,644 bp), and only 297 bp in 3'flanking region. To extend the amplified sequences on the 3'flanking region, we adopted additional sequences (576 bp) located at (4511-5086) bp from CHIP database to amplify 873 bp in the 3'flanking region and we designed PCR primer to amplify this region. For some fragments there are multiple primers sets (fragment 1, 4, 7), Table 3 lists the primer sets used in this study for PCR amplification. Because of having difficulties in amplifying fragment 4, we designed PCR primers to fill out the gap between the third and fourth fragment. We had sequence gap within fragment 3 in both samples (NHWs and African Blacks) and in fragment 7 in the African Blacks samples, and for this reason we designed internal sequencing primers to fill these gaps. Primer3 software version 0.4 (http://frodo.wi.mit.edu/primer3/) was used to design PCR and sequencing primers. The primers' sequences are summarized on Table 3.

Table 3 Polymerase Chain Reaction (PCR) Primers

Fragment	Fragment size	Forward Primer's Sequence	Reverse PCR Primer's Sequence	Internal Sequencing Primers *
1	(711)bp	5'-tgtaaaacgacggccagtGAGACGGAGTCTTGCTCTTTCGC-3'	3'-caggaaacagctatgaccTCAGTTTCCTCCTCCATAAAGTG-5'	
2	(807)bp	5'-tgtaaaacgacggccagtATTACAGGCACGCATCACTACTT-3'	3'-caggaaacagctatgaccCACACAGATGATCCCAGTTTGTA-5'	
3	(800)bp	5'-tgtaaaacgacggccagtGTGAGAAGAAGTGGGTGGAGG-3'	3'-caggaaacagctatgaccGTCAGGATATGGAGACCATCCTG-5'	5'-GCATGTCTGTGTGCATATGTGT-3' 3'-GTGGCGTGTGTCTGTAGTCC-5'
4*	(320)bp	5'-CCAGGATGGTCTCCATATCC-3'	3'-GTGCCTGGCCCTGTATTAAA-5'	
5	(848)bp	5'-tgtaaacgacggccagtATTCTAGATCAGCATTATCCAGTA-3'	3'-caggaaacagctatgaccTCTTCACCTTGTGTCAGTAGTCC-5'	
6	(1010)bp	5'-tgtaaaacgacggccagtGTCCACAGAGGTAGCTCAGACAG-3'	3'-caggaaacagctatgaccGCTCCTCTCTGTGACCTAGGAGT-5'	
7	(986)bp	5'-tgtaaaacgacggccagtAAAGCTAAAGATGAGTCGCTGG-3'	3'-caggaaacagctatgaccGGAGTCAGAGCTTGTAGGAGACA-5'	5'-CAGAGAACACCTGGGGAGAG-3' 3'-AAGAGATCTCGCTGTGTTGC-5'
8*	(739)bp	5'-CTTCAAGGCGTGTCAGTTTC-3'	3'-ATTTGTGGAGTGTGGTGGTG-5'	

Lowercase represent M13-tag primers, while uppercase represent primers sequence, * represents the primers we designed.

These amplicons generated eight overlapping fragments that covered a total genomic region of 5,086 bp, including 956 bp in the 5'flanking region, 614 bp of exons, 2,644 bp of introns, and 876 bp in the 3' flanking region. The entire gene was sequenced from both directions. There were two small gaps 33 bp between (1842-1875 bp) and 16 bp between (2147-2163). The polymerase chain reaction (PCR) reaction and cycling condition are summarized in Table 4. As part of optimization process, some modifications have been made either by increasing $MgCl_2$, changing the annealing temperature, or using DMSO for CG-rich regions during the PCR reaction.

Table 4 PCR Reaction and Cycling Condition

<table>
<tr><th colspan="3">PCR Reaction Conditions
(Reaction volume 25 µL)</th><th>PCR Cycle</th></tr>
<tr><td colspan="3">Genomic DNA (1ng/dl) 3.0 µL</td><td rowspan="8">1. 95˚C for 5 minutes.
2. 95˚C for 45 sec.
3. 58-60˚C for 45 sec
4. 72˚C for 1 min.
Repeats 2-4 for 40 cycles
5. 72˚C for 10 min.
6. Cool down to 4˚C</td></tr>
<tr><td rowspan="7">Master Mix (22µL)</td><td>d.H2O</td><td>11.75-13.25 µL</td></tr>
<tr><td>10x BufferGold</td><td>2.5 µL</td></tr>
<tr><td>$MgCl_2$ (25 mM)</td><td>1.5-3.0 µL</td></tr>
<tr><td>dNTPs (1.25 mM)</td><td>3.8 µL</td></tr>
<tr><td>Forward Primer (20 mM)</td><td>0.4 µL</td></tr>
<tr><td>Reverse Primer (20mM)</td><td>0.4 µL</td></tr>
<tr><td>Tag Polymerase enzyme Enzyme (5U/µL)</td><td>0.15 µL</td></tr>
</table>

After PCR, gel electrophoresis was performed to confirm DNA amplification prior to DNA sequencing. Invitrogen TM E-Gel® 96 2% with SYBR® Safe precaste gels (Invitrogen corporation, Carlsbad, CA) was used for large scale PCR, while regular 2% agarose gel was used for small scale PCR for samples that failed in the initial DNA amplification. The amplified

samples were sent to a commercial lab for automated fluorescence-based cycle sequencing and capillary electrophoresis on ABI 3730x1DNA Analyzers (Genomic Services of Beckman Coulter Genomics, Denver, MA). Variant Reporter version 1.0 (Applied Biosystem, Foster City, CA) and Sequencher version 4.8 (Gene Codes Corporation, Ann Arbor, MI), were used to analyze sequencing data.

2.2.2 TaqMan Genotyping of *APOC4* Common Variants

A total of seven TaqMan premade SNP genotyping assays were ordered to genotype 4 tag SNPs in NHWs and 7 tag SNPs in Blacks. Genotyping data of Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) and Yoruban in Ibadan, Nigeria (YRI) populations were obtained from publicly available HapMap database by using HapMap release #27 (www.hapmap.org) that covered the same *APOC4* region used in resequencing. Tag SNPs for both populations were selected by running Tagger option in Haploview program (Broad institute of MIT and Harvard, 2010) at r^2 cutoff 0.7 and MAF cutoff 0.049. There were 2 tag SNPs bins in CEU versus 7 tag SNPs bins in YRI. TaqMan assays for 6 out of 7 tag SNPs bins were available and 3 out of these 7 TaqMan assays were corresponding to 1 tag SNP bin in CEU, and the remaining tag SNP bin in CEU (rs5158) was genotyped also in Blacks. Thus, we genotyped 4 tag SNPs in NHWs and 7 tag SNPs in blacks. Information related to tag SNPs and TaqMan SNP genotyping assays is summarized in Table 5 and Table 6, respectively.

Table 5 Tag SNPs and Captured Alleles in CEU and YRI

CEU			YRI		
Bin no.	**Allele captured**	**Genotyped SNP**	**Bin no.**	**Allele captured**	**Genotyped SNP**
1	**rs5167 (3927)** **rs5157 (2623)** rs2288911 (4746) rs2288912 (4661) **rs1132899 (3498)**	rs5167 rs5157 rs1132899	1	rs12709884 (4154) rs5159 (2971) **rs12721104 (3380)**	rs12721104
2	**rs5158 (2640)**	rs5158	2	rs2288911 (4746) rs2288912 (4661)	------
			3	**rs5167 (3927)**	rs5167
			4	**rs5155 (2559)**	rs5155
			5	**rs5157 (2623)**	rs5157
			6	**rs10425530 (4157)**	rs10425530
			7	**rs1132899 (3498)**	rs1132899

Bolded SNPs represent the genotyped SNPs by TaqMan.

Table 6 TaqMan SNPs Genotyping Assays

SNP reference ID	Position on NCBI	Assay type	Functional Location	Assay ID	Population
rs1132899	45448036	Premed	missense	C__1841831_10	NHWs & Blacks
rs5158	45447178	Premade	intron	C__11466146_10	NHWs & Blacks
rs5157	45447161	Premade	intron	C__1841830_10	NHWs & Blacks
rs5155	45447097	Premade	intron	C__26681856_10	Blacks
rs10425530	45448695	Premade	3'UTR	C__29686817_10	Blacks
rs12721104	45447918	Premade	intron	C__33605659_10	Blacks
rs5167	45448465	Premade	missense	C__1341833_20	NHWs & Blacks

TaqMan genotyping involves DNA amplification and end-point fluorescence reading using the ABI Prism 7900HT instrument. The TaqMan genotyping Master Mix and Assay were added to dried whole genome amplified DNA in 384-well plate. For TaqMan genotyping we followed ABI manufactures protocol (Applied Biosystem, 2007) with some modifications on the DNA concentration and the number of cycle that are illustrated on Table 7.

Table 7 TaqMan SNPs Genotyping Condition

TaqMan Reaction total volume (5μL)		PCR condition
d.H2O	2.43 μL	1. 95˚ for 10 min. 2. 95˚ for 45 sec. 3. 60˚ for 1 min. -repeat 2-3 49x
Master Mix	2.5 μL	
Assay	0.06 μL	

2.3 STATISTICAL ANALYSIS

Allele and genotype frequencies were determined by using direct counting. Concordance of the genotype distribution to Hardy-Weinberg Equilibrium (HWE) was tested using chi-squared (X^2) test for each variant. Haploview software was used to analyze the variant allele frequencies, their distributions among low and high HDL-C groups, and their LD patterns. For the resequencing samples, X^2 test was used to compare the allele frequencies between low and high HDL-C groups. For those SNPs that were genotyped in the entire samples, linear regression was performed to test the effects of genotypes on the means of four lipid traits (total cholesterol,

HDL-C, LDL-C, and TG). To minimize the effect of non-normality, natural log transformation was used to transform HDL-C and TG values in NHWs. In Blacks, log transformation was used to transform total cholesterol and TG, and the square root transformation used to transform HDL-C and LDL-C. The significant covariates were identified using stepwise regression in both directions. In NHWs, covariates included in the final model were gender, age, BMI (weight [kg]/height $[m]^2$), and smoking. Among Blacks: gender, age, waist, Jobmin (minutes walking or bicycling to work each day (min)), and Staff (staff level (junior/senior)) covariates were included in the analysis. The additive and dominant models were used for data analysis. The R statistical software package (version 2.3.1, http://www.r- project.org) and Statistical Analysis Software (SAS) were used to perform all computations. A *p*-value of less than 0.05 under one of these models was considered as suggestive evidence of association.

3.0 RESULTS

3.1 *APOC4* RESEQUENCING

Resequencing of the *APOC4* gene in 190 NHW chromosomes (94 in high HDL-C group and 96 in low HDL-C group) and 190 African Black chromosomes (96 in high HDL-C group and 94 in low HDL-C group) revealed a total of 65 variants, including one dinucleotide microsatellite. Table 8 summarizes *APOC4* variants identified in this study in NHWs and Blacks. Of these observed 65 variants, one was a dinucleotide microsatellite, 4 were insertions or deletions (indels), and the remaining 60 variants were single nucleotide substitutions (SNPs). Of the 65 variants identified, 15 were located in the 5'flanking region, 1 was in the 5'UTR, 29 were in introns, 2 were in splice sites, 7 were in exons, 2 were in the 3' UTR and 9 were in the 3' flanking region. Twenty-six variants were present in NHWs, 51 were observed in African Blacks and 13 were present in both (Figure 3). Five of the seven exonic variants were non-synonymous resulting in amino acid changes, including two new (3969A>C [Lys110Tyr] and 4012G>A [Lys124Arg]) observed in Blacks only at low frequency (0.011 and 0.005, respectively). While 26 of the observed variants have previously been reported in databases (SeattleSNPs or CHIP), we observed 39 new variants in this study, including the dinucleotide microsatellite repeats at position 4929-4959 bp in both populations. For all samples in both populations there was GC instead of CG at position 1722-1723 bp, which is different from the reference sequence.

Table 8 *APOC4* Variants Identified in NHWs and Blacks

APOC4 variants Position (bp)	Base Change	Ref SNP ID	Location	Amino acid change	NHWs		Blacks	
					MAF%	Call rate%	MAF %	Call rate%
65	C>T		5' flanking region				0.007	77.9
92_94	del3	rs12721101	5' flanking region		0.259	87.4	0.262	86.3
108	G>A		5' flanking region		0.259	87.4	0.262	86.3
116	A>G		5'flanking region		0.007	78.9		
150_151	ins114		5' flanking region		0.256	88.4	0.283	94.7
204	A>G	rs4803773	5' flanking region		0.235	52.6	0.233	94.7
233	C>T		5' flanking region				0.006	93.7
245	G>T		5' flanking region				0.006	94.7
368	A>G		5' flanking region				0.011	93.7
438	G>A		5' flanking region				0.006	94.7
489	C>T		5' flanking region				0.017	93.7
636	C>T		5' flanking region		0.005	97.9		
637	G>T	rs73558107	5' flanking region				0.043	98.9
757	C>A	rs12721105	5' flanking region				0.054	97.9
870	G>A		5'flanking region		0.005	97.9		
968	A>G		5' UTR		0.038	97.9		
1088	T>G		splice site				0.006	91.6
1130	T>C		intron 1				0.005	97.9
1150	A>G		intron 1		0.005	100		
1192	G>A		intron 1				0.006	91.6
1229	G>C		intron 1		0.011	100		
1325_1327	del3		intron 1		0.147	89.5	0.047	89.5
1430_1431	Ins1		intron 1				0.033	94.7
1702	G>A	rs12721102	intron 1				0.006	91.6
1719	C>A		intron 1				0.006	91.6
1733	C>T	rs1271111	intron 1		0.263	82.1	0.244	90.5
1823	C>G		intron 1		0.187	78.9	0.203	77.9
2063	C>G		intron 1		0.005	98.9		
2099	G>T		intron 1				0.011	98.9
2467	C>T		intron 1				0.017	92.6
2557	C>A		intron 1		0.006	91.6		
2559	C>T	rs5155	intron 1				0.088	95.8
2607	G>A	rs5156	intron 1				0.016	96.8
2623	C>T	rs5157	intron 1		0.477	90.5	0.163	96.8
2640	C>T	rs5158	intron 1		0.153	92.6	0.033	95.8
2641	G>A		intron 1				0.005	97.9
2678	G>C		intron 1				0.005	96.8

Table 8 (Continued)

2683	G>A	rs12721109	intron 1		0.011	92.6		
2767	G>T	rs12721107	intron 1				0.017	93.7
2971	A>G	rs5159	intron 1				0.15	94.7
3213	T>C	rs28616151	intron 1				0.056	93.7
3348	G>A		intron 1				0.006	86.3
3363	G>A		intron 1				0.043	85.3
3380	G>A	rs12721104	intron 1				0.159	86.3
3498	C>T	rs1132899	exon 2 (non-synonymous)	Pro36Leu	0.476	87.4	0.223	87.4
3502	C>T	rs10423683	exon 2 (Synonymous)	Ser37Ser			0.06	87.4
3546	G>A	rs12691089	exon 2 (non-synonymous)	Gly52Asp	0.006	88.4		
3592	C>T	rs12691090	exon 2 (synonymous)	Asp67Asp			0.027	96.8
3700	G>A		intron 2				0.005	97.9
3792	G>A	rs5165	intron 2				0.017	94.7
3847	T>C		splice site		0.011	96.8		
3927	T>G	rs5167	exon 3 (non-synonymous)	Leu96Arg	0.339	94.7	0.472	94.7
3969	A>C		exon 3 (non-synonymous)	Lys110Tyr			0.011	97.9
4012	G>A		exon 3 (non-synonymous)	Lys124Arg			0.005	95.8
4154	G>A	rs12709884	3' UTR				0.115	95.8
4157	G>A	rs10425530	3' UTR				0.1	94.7
4533	C>T		3' flanking region				0.011	92.6
4579	G>A		3' flanking region				0.052	91.6
4628	G>A	rs12721063	3' flanking region		0.011	94.7		
4661	C>G	rs2288912	3' flanking region		0.494	91.6	0.244	88.4
4746	T>G	rs2288911	3' flanking region		0.500	93.7	0.224	89.5
4844	G>A		3' flanking region				0.107	88.4
4895	G>A		3' flanking region		0.006	91.6		
4912	G>C		3' flanking region				0.006	88.4
4929-4959	dinucleotide repeats		3'flanking region		-----	-----	-----	-----

Variants with refSNPs ID represent previously reported variants and those present in both populations are highlighted in yellow.

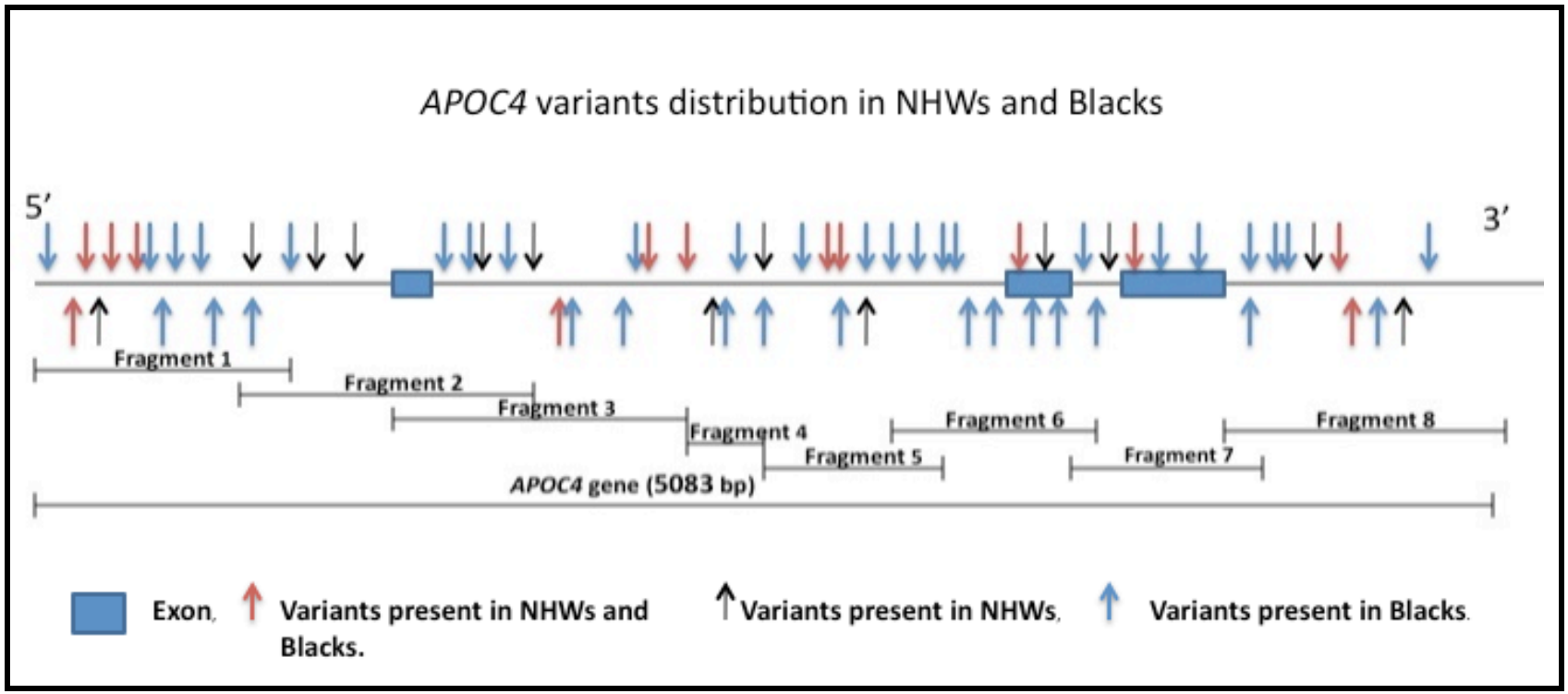

Figure 3 Distribution of APOC4 Genetic Variants

3.1.1 Non-Hispanic Whites

A total of 26 variants were identified in NHWs plus one dinucleotide microsatellite. Twelve of 26 variants were previously reported in databases and 14 variants were new. Since the dinucleotide microsatellite was not fully analyzed, it is not included in the total number of ethnic-specific variants. Only 8 of these variants had MAF<1%, 5 variants had MAF 1-5% and 13 variants had MAF≥ 5%. Among 26 variants identified in this study, 7 were located in the 5'flanking region, 1 was in the 5'UTR, 10 were in introns, 1 was in the splice site, 3 variants were in exons, and 4 in the 3'flanking region. Three of 26 variants were indels (two small deletions and one large insertion) with the number of bases affected range was between 3-114 bp. One deletion (92_94del3) was in the 5'flanking region, one insertion (150_151ins114) was in the 5'flanking region and one deletion (1325_1327del3) was in intron 1. There was one variant in the splice site, 3847 T>C. The 3 variants in the coding region were non-synonymous SNPs resulted in amino acid changes: 3498 C>T/ rs1132899 (Pro36Leu)/ MAF=0.476 was in exon 2, 3546G>A/ rs12691089 (Gly52Asp)/ MAF=0.006 was in exon 2, and 3927T>G/ rs5167 (Leu96Arg)/ MAF=0.339 was located in exon 3.

3.1.2 African Blacks

A total of 51 variants were identified in African Black plus one dinucleotide microsatellite, including 4 indels. Twenty-three of the 51 identified variants were previously reported in databases, and 28 were new. Since the dinucleotide microsatellite was not fully analyzed, it is not included in the total number of ethnic-specific variants. Fifteen variants had MAF<1%, 15 had MAF 1-5%, and 21 had MAF≥ 5%. Of the variants identified in this study, 12 were located in the 5'flanking region, 24 were located in introns, one was in the splice site, 6 were in exons, 2 were in the 3'UTR, and 6 were in the 3'flanking region. Four out of 51 variants were indels (2 deletions and 2 insertions) with number of bases affected range was between 1-114 bp. One deletion and one insertion were located in the 5' flanking region (92_94del3, 150_151ins114), and additional one deletion and one insertion were located in intron 1 (1325_1327del3, 1430_1431ins1). There was one variant in the splice site 1088 T>G. Among the 6 variants in the coding region 4 variants were non-synonymous, including two unreported SNPs. The two known non-synonymous variants were: 3498 C>T/ rs1132899 (Pro36Leu)/ MAF=0.223 was in exon 2 and 3927 T>G/ rs5167 (Leu96Arg) / MAF=0.472 was in exon 3. The two new non-synonymous variants were observed only in Blacks: 3969A>C (Lys110Tyr) / MAF=0.011 was in exon 3 and 4012 G>A (Lys124Arg)/ MAF=0.005was in exon 3.

3.1.3 *APOC4* Annotated Sequence

Variants identified in the *APOC4* gene are depicted in a colored FASTA representation of *APOC4* annotated reference sequence adopted from SeattleSNPs database (http://pga.gs.washington.edu/) with additional sequence (from 4,511-5,086) from CHIP database (dbSNPs build: 130) (http://snpper.chip.org/) in Figure 4. The annotated reference sequence was modified by including the identified variants in this study with population-specific variants representation. The dinucleotide microsatellite was not included in this annotated sequence. Variants identified in both populations (NHWs and African Blacks) (n=13) are depicted in **RED** font. **Blue** font used for variants identified in African Blacks population only (n=38) and **Black** font used for variants identified only in NHWs population (n=13). All variants that identified previously in database but not in our study had MAF<0.02 and depicted with **Dark green** font. Yellow highlight is used for deletion (n=2) and blue highlight is used for bases flank the insertion (n=2) with the population-specific color font. Variants identified in Seattle and in our study include ref SNPs ID in addition to its position according to annotated Seattle reference sequence and allele change. The color code used in SeattleSNPs for the reference sequence is as follows: light grey for flanking regions and introns, green for UTR, dark blue for exons, purple for repeat regions, and light blue for regions not scanned in SeattleSNPs database.

```
GAGACGGAGT CTTGCTCTTT CGCCCAGGCT GGACTGCAGT GGCGCGATCT 50  | REPEAT
CGGCTCAATG CAAGCTCCAC CTCCCAGGTT CACGCCATTC TCCTGCCTCA 100 | p.65[C>T]|
p.92_94/rs12721101[delCCT]
GCCTCCCGAG TAGCTAGGAC TACAGGCGCC TGCCACCACG CCTGGCTAAT 150 | p.108[G>A]| p.116[A>G]
TTTTCATATT TTTAGTAGAG ATGGGGTTTC ACCGTGTTAG CCAAGATGGT 200 | P.150_151[ins114]
CTCAATCTCC TGACCTCGTG ATCCGCCTGC CTCGGCCTCC CAAAGTGTTG 250 | p.204/rs4803773[A>G]|
p.233[C>T]| p.245/[G>T]
GGGTTACAGG CATGAGCCAC CGCGCCTGGC CAACAGCAAT GATCTTTGAG 300 | REPEAT
CACCTATATT GCCAGTCTCC ACGGTAAGAG CTTTCTTCAT TTTTTGTTTT 350 | REPEAT
GTTTTGTTTC AAGACAGAGT CTTGCTCTGT CACCCAGGCT GGAGTGCAGT 400 | p.368[A>G]
GGTGTGATCG CGGCTCACTG CAGCCTTCAC TTCCCGGGTT CAAGCCATTC 450 | p.438[G>A]
TCCTGCCTCA GCCTCCCAAG TAGCTGGGAT TACAGGCACG CATCACTACT 500 | p.489[C>T]
TCTGGCTAAT TTTTGTATTT TTAGTAGGGA CAGGGTTTTT CACCATGTTG 550
GCCAGGTTGG TCTCAAACTC CTGGCCTCAT ATGATCTGCC CACCTCGGCC 600
TCCCAAAGTG CTGGGATTAC AGGCGTGAGC CACTGCGCCT TTCTTTGTAT 650 | REPEAT| p.636[C>T]|
p.637/rs73558107[G>T]
TTGTTCAAGT AATATACTGA AATATGTACT GTGCCTCCCA CTTTATGGAG 700
GAGGAAACTG AGGCCAGCAA ATGAGGCTGT CATGGGAGGT GGAGACAGGA 750
TTTGAACCTG CCTCAGTGCA GGAGGCTCAA GAGCCTCTGT CTTCTCTCAG 800 | p.757/rs12721105[C>A]
GGCACTGTGT GGGAGGGTGA GAAGGAGGGA GGCCCACAGA GGCATGACCT 850
CTGATTGCCA CTGTCACCTG GGCCCTGCTC TCTGAAGTCT CTGCCAAGCG 900 | p.870[G>A]
GGGAGGTGGC CGGGGGAGGG CCCTGCTCTG TGCAGCCTCC CCTCCCCCGG 950
CCCGCAGAGT TGAGCACAGA GGGACAGAGG CACGGAACCC CCAGAAATGT 1000 | Exon 1 | UTR| p.968[A>G]
                                                    M   1
CCCTCCTCAG AAACAGGCTC CAGGCCCTGC CTGCCCTGTG CCTCTGCGTG 1050
S  L  L  R   N  R  L   Q  A  L   P  A  L  C   L  C  V  18
CTGGTCCTGG CCTGCATTGG GGGTGAGAAG AAGTGGGTGG AGGGATGTGG 1100 | p.1088[T>G]
 L  V  L   A  C  I  G                                  25
GGCCCACACC TGGTGGGTGT GAGTGTGGCT GTGTGTCCTG TGGCTCTGTA 1150 | p.1130[T>C]| P.1150[A>G]
GCCACGTGAG ACATGAGTAC GGAGTGTGTG CGTTTCATGG CGTGCGTATG 1200 | p.1192[G>A]
CATGTGCGTG TCGGGGAGTG TGTGTGTGGG TGGCTGAGAG TGAAGTGTGA 1250 | p.1229[G>C]
ATGTCACATT GGTACAAACT GGGATCATCT GTGTGTGTGC ACGTGCGTGC 1300
GTGGAAGTGG GAGTATGCAG TCGTGGTAAA AAAGTGCATG TCTGTGTGCA 1350 | p.1325_1327[delGGT]
TATGTGTATT TGTGTGCACC TGTCTCTCTG TGGGGTATGT GTGTGCAAAA 1400
TATTTGAGTG TGTGGACATG TGTGAGGGGG TGAGTGTGTG CTGGTGTGTA 1450 | p.1430_1431[insG]
CGTCTGTGTT TTGCATATGC ATTTTTTTTT TTTTTTTGA GACGGAGTCT 1500 | REPEAT
CACTCTGTCA CCCAGGCTGG AGTGCAGTGG TAGCAGTGGT GCGATCTTGG 1550
CTCACTGCAT CATCCGCCTA CCCGTTTCAA GGGATTCTCC TGCCTCAGTC 1600
TTCAGAGTAT TTGGGACTAC AGACACACGC CACCATGCCT GGCTAATTTT 1650 | REPEAT
TTTTTTTTGA GACGGAGTCT CGCTCTGTTA CCCAGGCTGG AGTGCAGTGG 1700
CGTGATCTTG GCTCACTGCA AGCTCCGCCT CCCGGGTTCA CGCCATTCTC 1750 | p.1702/rs12721102[G>A]|
p.1719[C>A]| p.1733/rs12721111[C>T]
CTGCCTCAGC CTCCCGAGTA GCTGGGACTA CAGGAGCCCA CCACCACGCC 1800
TGGCTAATTT TTTGTATTTT TACTAGAGAC GGGGTTTCGC CGTGTTAGCC 1850 | p.1823[C>G]
AGGATGGTCT CCATATCCTG ACCTCGTGAT CCGCCTGCCT CGGCCTTCCA 1900 | NOT SCANNED
AAGTGCTAGG ATTATAGGCG TGAGCCACTG CGCCTGGCCA ATGCCTGGCT 1950 | REPEAT
AATTTTTTTA TATTTTTGGT AGAGACAGGG TTTTGCCATG TTGCCCAGGC 2000
TGGTCTTGAA ATCCTGACCT CAGGTGATCC GCCCGCCTTG GCCTCCCAAA 2050
GTGCTGGGAT TACAGGCATG AGCCACCACG CCCGGCCATG TACTTTATGT 2100 | p.2063[C>G]| p.2099[G>T]
TAAAATGGGA TCATATTCTA GATCAGCATT ATCCAGTAGA AATTTAAATT 2150
TTTAATACAG GGCCAGGCAC GGTGGCTCAT GCCTGTAATC CCAGCACTTT 2200 | REPEAT
CGGAGGCCGA GGCGGGTGGA TCGCAAGGTC AGGAGATTTG AGATCATCCT 2250
GGCTAACAGA TGGGTAAAAA CCCATCTCTA CTAAAAATAC AAAAAATTAG 2300
CCATGCATGG TGGCATGCGC CTGTAGTCCC AGCTACTCGG GAGGCTGAGG 2350
CCGGAGAATC ACTTGAACCC GGGAGGCAGA GGTTGCAGTG AGCCGAGATC 2400
GCGCCACTGC ATTCCAACCT GGGTGACAGA GCGAGACTCC GTCTGAAAAA 2450
AAAAAAAAAT TTAACACGTA TGTAGACAAT GTGCAAGGCA CCATTCCATG 2500 | p.2467[C>T]| REPEAT
TGCATCGTAT GTAGTAACTC TTAATTCTCA CGATAACCCT GAGGTAGATA 2550
TTATTACCCC GTTCTACAAA AGGAGAAACA GTCCTGGGGA GACAGGATAA 2600 | p.2557[C>A]|
p.2559/rs5155[C>T]
```

```
GTCACCGGCC AAGGCACACA GCCAGCTACA TGTGGCCCCC GCGTGACGGC 2650 | p.2607[G>A]/rs5156|
p.2623/rs5157[C>T] | p.2640/rs5158[C>T]| p.2641[G>A]
TGGTCTCTGT AGGCGAGGCT TTGTCCAGAT GCGTGGGTAG AAGGTCTGGC 2700 | p.2678[G>C]|
p.2683/rs12721109[G>A]
CCGGAAAGAG GAACTGACAG CAAGGCTAAG CCAATGTCTG CCCCTGGGGG 2750 | p.2703/[G>T]
CAGAAAGTCA CCTCCTGCTC TCCCTCCACT GTCCACAGAG GTAGCTCAGA 2800 | p.2753[G>A] |
p.2767/rs12721107[G>T]
CAGGGTGGGG GTCACAGGAG AACGAAGGGA GAAGGGGGTA GTTCCTGGGC 2850
AGCAAAATCA GGTGGTGAAG GGAGGCATCA GAGGATGGCA ATTAGAGAGG 2900 | p.2886[T>C]
CCATTAGAGG GGAACCACAG GCAGACAGGG TGACAGGAGG GACTACTGAC 2950
ACAAGGTGAA GAGATGGCCC AGCCGGACGG GGTGGCTCAC ATCTGTAATC 3000 | REPEAT |
p.2971/rs5159[A>G]
CCAGCATTTT GGGAGCCCGA GGTGGGTGGA TCACTTGAGG TCAGGAGTTC 3050
GAGGCCCCAA CATGGCAAAA CCCCATCTCT TCTAAAAATA CAAAAATTAG 3100
CCGGGCATGA TGGCAGATGC CTGTAATCCC TGCTACTCGG GAGGCTGAGG 3150
CAGGAAAATT GCCTGAATCC AGGAGGTGGA GGTTGCAATG AGACGAGATC 3200
ATGACACTGC ACTCCACCCT GGGCAACAGA GCAAGAGACT GACTCTGTCT 3250 | p.3213/rs28616151[T>C]
CATAAAAAAA AAGAAAAAAG AAAAAAAAAA AGAGATGGCT GATGGTTAAA 3300
GAGGGGTTAG CGGTCAGGGG ACACATAAGG GTAAAGGCAG GAGGCAAGAG 3350 | p.3348[G>A]
GACTGGCAGG GGGCTGCCCC TGGGCCACCG GGAGCGACAC AGGATGAGCA 3400 | p.3363[G>A]|
p.3380/rs12721104[G>A]
TGGAGGGAAA GGGAGAAGGG GATTCTAGGG TCCCAGCCTA CCCAAGTTGC 3450
CCTCTGGTTC CACCTAGCAT GCCAGCCAGA GGCCCAGGAA GGAACCCCGA 3500 | Exon 2 |
p.3498/rs1132899[C>T]
                  A   C  Q  P  E   A  Q  E   G  T  P   36
GCCCCCCACC AAAGCTAAAG ATGAGTCGCT GGAGCCTGGT GAGGGGCAGG 3550 | p.3502/rs10423683[C>T]|
p.3546/rs12691089[G>A]
S  P  P  P   K  L  K   M  S  R   W  S  L  V   R  G  R  53
ATGAAGGAGC TGCTGGAGAC AGTGGTGAAC AGGACCAGAG ACGGGTGGCA 3600 | p.3592/rs12691090[C>T]
 M  K  E   L  L  E  T   V  V  N   R  T  R   D  G  W  Q 70
ATGGTTCTGG TGAGGGTGTG CTGGGCTGGG TGGTGGGAGG GGACTCCTGG 3650 | REPEAT
  W  F  W                                              73
GTCTGAGGGA GGAGGGGCTG GGGCCTGGAC CCCTGAGTCT CAGGGAGGAG 3700 | p.3700[G>A]
GAAAGGGTGG GAGTGGGGCT GTGACCCCTA GGTCTGGGAG GAGTGGAGGG 3750
TTAGAGCTGA GAGCAGGAAC TCCTAGGTCA CAGAGAGGAG CGGATAAATG 3800 | p.3792/rs5165[G>A]
GGGCAGAGAA CACCTGGGGA GAGCTGGGGC CTCCACTGTG ATGTCCTCTC 3850 | p.3847[T>C]
TCCTGTAGGA GCCCGAGCAC CTTCCGGGGC TTCATGCAGA CCTACTATGA 3900 | Exon 3
          S  P  S  T   F  R  G   F  M  Q   T  Y  Y  D 87
CGACCACCTG AGGGACCTGG GTCCGCTCAC CAAGGCCTGG TTCCTCGAAT 3950 | p.3927/rs5167[T>G]
  D  H  L   R  D  L   G  P  L  T   K  A  W   F  L  E   103
CCAAAGACAG CCTCTTGAAG AAGACCCACA GCCTGTGCCC CAGGCTTGTC 4000 | p.3969/A>C]
S  K  D  S   L  L  K   K  T  H   S  L  C  P   R  L  V  120
TGTGGGGACA AGGACCAGGG TTAAAATGTT CATAAAAGCC AGGTGTGGTT 4050 | UTR | REPEAT|
p.4012[G>A]
 C  G  D   K  D  Q  G                                  127
GTGGCGGGTG CCTGTAGTCC CAGCTACTCA GGAGGCTGAG GTAGGATGAT 4100
GGCTTGAGCC CAGGAGTTCG AGACCAGCCT GGGCAACACA GCGAGATCTC 4150
TTGGGGGTAA AACAAAAAGA AAAAAAAAAG TTCATACTTC TCCAATAAAT 4200 | p.4154/rs12709884[G>A]|
p.4157/rs10425530[G>A]
AAAGTCTCAC CTGTGTCCCT GTCTGGATCC TTCCCCAGTG TGGCCAGAAA 4250
AAAACCCACC CCACTGCCTC CCAGGAATCA ATGAGTAGAA GAGGTGACAC 4300
CTGATGGGGA AGGAAGAGTA GGGAGGTCGG GAAGGGTATC AAGGAATAAC 4350 | p.4346[A>T]
ACCCTATTGT GGGCTTGCGG AGAATGGGGG ACTTCAAGGC GTGTCAGTTT 4400
CAGGAGGGTG AGGGCAGGAG CGTGGGTGGA GTCAGCAGGT CCCCATGATG 4450
GCCCTCACTG AGAGCTTCGC CCTTGTCTCC TACAAGCTCT GACTCCATTC 4500
CCAGTGGGCA CCCAGCACCT CCAACCCCTC CACAGCCCCC AACCCAGCCT 4550 | p.4533[C>T]
CTGTCGGAGG CGAATTCTCA GAGTGAGGGT TCCCTGTCAC TTGAGAGAAG 4600 | p.4579[G>A]
GTTCCCTGTG ACGTGACCTT GGGGGACGTC ATTGCCCTTT CTGTCCCCAC 4650 | p.4628/rs12721063[G>A]
CCACCCCCTC CCGCAGTTCTG TTGGCCAGGA CTTTGGCCTA GACAAAGGAT 4700 | p.4661/rs2288912[C>G]
GGGGGTTGTG GCTGTGGAGC GGAAGTGGGT CTCAACCACT ATAAATCCTC 4750 | p.4746/rs2288911[T>G]
TCTGTGCCCG TCCGGAGCTG GTGAGGACAG CCTGCCAGAG TCTGGTAAGA 4800
AAGGGACTCA GGGTGCGGGG ACAGGGGGGC GTCAGCAGGG AGAGGGCAAA 4850 | p.4844[G>A]
GATCGATAAA GCAGGAATTT TAAGAGGCAC AATATTAGAA GCCCGTGTTG 4900 | p.4895[G>A]
```

```
GAACCATGAC TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG 4950 | p.4912[G>C]
TGTGAGAGAG AGAGAGAGGG AGATGGAGTC TCGCTATGTA GCCCAGGCTA 5000
GACTCAAACT CCTGGGCTCA AGCAATCCTC CTGCCTCAGC CTCCCCAGTA 5050
GCTGGGACTA CAGGTGCACC ACCACACTCC ACAAAT                5086
```

Figure 4 *APOC4* Annotated Sequences

3.2 DISTRIBUTION OF *APOC4* VARIANTS IN HIGH AND LOW HDL-C GROUPS

3.2.1 Non-Hispanic Whites

Of the 26 variants identified in this study in NHWs, 13 had MAF<5%, and 13 had MAF≥ 5%. Among the fourteen variants that are identified in our study but they were not previously reported in other databases (Seattle and CHIP), the MAF range was 0.005-0.259. The 4 common variants that have not been reported previously are: (108G>A/MAF=0.259, 5'flanking region), (150_151ins114/MAF=0.256, 5'flanking region), (1325_1327del3/MAF=0.147, intron 1), and (1823C>G/MAF=0.187, intron 1). The108C>G and 150_151ins114 variants were in complete LD with previously reported 92_94del3. None of the thirteen common variants showed significant difference between low and high HDL-C group. Table 9 shows the allele distribution of common variants among low and high HDL-C groups. Among relatively uncommon or rare variants (MAF<5%), while 8 variants were present only in low HDL-C group, only 2 variants were present in high HDL-C group. Three variants were present in both low and high HDL-C groups. Table 10 shows the distribution of rare or relatively uncommon variants among low and high HDL-C groups. The number of individuals having a minimum of one rare variant was higher in the low HDL-C group (15/48, 31.3%) compared to the high HDL-C group (5/47, 10.6%). Of the exonic variants, one non-synonymous variant located in exon 2 (rs12691089 [Gly52Asp]/ MAF=0.011) was present only in the low HDL-C group. Rare variants genotypes include individuals with less than 100% genotype call rate, which might underestimate the magnitude in the difference of allele distribution among low and high HDL-C group.

Table 9 Distribution of *APOC4* Common Variants in High and Low HDL-C Group in NHWs

APOC4 variant Position	Base Change	Ref SNP ID	Location	amino acid change	all MAF %	Frequency in high HDL-C	Frequency in low HDL-C	P value
92_94	del3	rs12721101	5' flanking region		0.259	0.262	0.256	0.926
108	G>A		5'flanking region		0.259	0.262	0.256	0.926
150_151	ins114		5'flanking region		0.256	0.262	0.250	0.860
204	A>G	rs4803773	5'flanking region		0.235	0.212	0.261	0.565
1325_1327	del3		Intron 1		0.147	0.138	0.154	0.740
1733	C>T	rs12721111	Intron 1		0.263	0.276	0.250	0.709
1823	C>G		Intron 1		0.187	0.191	0.183	0.897
2623	C>T	rs5157	Intron 1		0.477	0.436	0.489	0.329
2640	C>T	rs5158	Intron 1		0.153	0.159	0.149	0.860
3498	C>T	rs1132899	exon 2 (non-synonymous)	Pro36Leu	0.476	0.423	0.477	0.200
3927	T>G	rs5167	exon 3	Leu96Arg	0.339	0.372	0.309	0.368
4661	C>G	rs2288912	3' flanking region		0.494	0.489	0.478	0.657
4746	T>G	rs2288911	3' flanking region		0.5	0.476	0.478	0.544

Table 10 Distribution of *APOC4* Relatively Uncommon and Rare Variants in Low and High HDL-C Groups in NHWs

APOC4 variant Position	Base Change	Ref SNP ID	Location	All MAF %	MAF% high HDL-C	MAF% low HDL-C
116	A>G		5'flanking region	0.007	0.000	0.013
636	C>T		5' flanking region	0.005	0.000	0.011
870	G>A		5'flanking region	0.005	0.011	0.000
968	A>G		5' UTR	0.038	0.011	0.062
1150	A>G		intron 1	0.005	0.011	0.000
1229	G>C		intron 1	0.011	0.000	0.021
2063	C>G		intron 1	0.005	0.000	0.011
2557	C>A		intron 1	0.006	0.000	0.011
2683	G>A	rs12721109	intron 1	0.011	0.012	0.011
3546	G>A	rs12691089	exon 2 (non-synonymous)	0.006	0.000	0.011
3847	T>C		splice site	0.011	0.000	0.021
4895	G>A		3' flanking region	0.006	0.000	0.011
4628	G>A	rs12721063	3' flanking region	0.011	0.012	0.011

Yellow highlighted variants represent variants that were uniquely present in low HDL-C group, while green highlighted variants represent variants that were uniquely present in high HDL-C group.

3.2.2 African Blacks

Out of 51 variants identified in Blacks, 21 had MAF ≥ 5%, and 30 had MAF<5%. Among the 28 variants that were identified in our study but they were not previously reported in other databases (Seattle and CHIP), the MAF range was 0.005-0.283. In addition to 108G>A, 150_151ins114 and 1823C>G, two common variants were located in the 3'flanking region, (4579G>A/MAF=0.052 and 4844G>A/MAF=0.107) were not identified previously. No statistically significant difference was found in the allele frequencies of common variants between low and high HDL-C groups. Two variants 150_151ins114 and 4579G>A showed borderline associations with respective p-value 0.094, 0.081. Table 11 shows the distribution of common variants in low and high HDL-C groups. Of the 30 relatively uncommon and rare variants (MAF<5%), 10 were present in the low HDL-C group and 7 were present in the high HDL-C group. Table 12 shows the distribution of rare and relatively uncommon variants in low and high HDL-C groups. The number of individuals having a minimum of one rare or less common variant was quite comparable in both groups with a little bit higher in the high HDL-C group (26/48, 54%) than in the low HDL-C group (22/47, 46%). Of the exonic variants, one non-synonymous variant located in exon 3 4012G>A (Lys124Arg) was present only in the low HDL-C group.

Table 11 Distribution of *APOC4* Common Variants in Low and High HDL-C Groups in Blacks

APOC4 variants Position	Base Change	Ref SNP ID	Location	Amino acid change	all MAF %	Frequency in high HDL-C group	Frequency in low HDL-C group	p-value
92_94	Del3	rs12721101	5' flanking region		0.262	0.209	0.321	0.106
108	G>A		5' flanking region		0.262	0.209	0.321	0.106
150_151	Ins114		5' flanking region		0.283	0.228	0.341	0.094
204	A>G	rs4803773	5' flanking region		0.233	0.250	0.216	0.589
757	C>A	rs12721105	5' flanking region		0.054	0.064	0.043	0.538
1733	C>T	rs1271111	intron 1		0.244	0.228	0.262	0.602
1823	C>G		intron 1		0.203	0.238	0.156	0.220
2559	C>T	rs5155	intron 1		0.088	0.067	0.109	0.317
2623	C>T	rs5157	intron 1		0.163	0.163	0.163	1.000
2971	A>G	rs5159	intron 1		0.150	0.152	0.148	0.933
3213	T>C	rs28616151	intron 1		0.056	0.064	0.048	0.639
3380	G>A	rs12721104	intron 1		0.159	0.151	0.167	0.786
3498	C>T	rs1132899	exon 2 (non-synonymous)	Pro36Leu	0.223	0.227	0.218	0.886
3502	C>T	rs10423683	exon 2 (Synonymous)	Ser37Ser	0.060	0.068	0.051	0.648
3927	T>G	rs5167	exon 3 (non-synonymous0	Leu96Arg	0.472	0.554	0.500	0.465
4154	G>A	rs12709884	3' UTR		0.115	0.106	0.125	0.694
4157	G>A	rs10425530	3' UTR		0.100	0.106	0.093	0.765
4579	G>A		3' flanking region		0.052	0.023	0.081	0.081
4661	C>G	rs2288912	3' flanking region		0.244	0.267	0.220	0.470
4746	T>G	rs2288911	3' flanking region		0.224	0.250	0.198	0.413
4844	G>A		3' flanking region		0.107	0.110	0.105	0.915

Yellow highlighted variants have marginally significant p-value

Table 12 Distribution of *APOC4* Relatively Uncommon and Rare Variants in Low and High HDl-C Groups in Blacks

APOC4 variants	Base Change	Ref SNP ID	Function	all MAF %	MAF% low HDL-C group	MAF% high HDL-C group
65	C>T		5' flanking region	0.007	0.015	0.000
233	C>T		5' flanking region	0.006	0.011	0.000
245	G>T		5' flanking region	0.006	0.011	0.000
368	A>G		5' flanking region	0.011	0.011	0.011
438	G>A		5' flanking region	0.006	0.011	0.000
489	C>T		5' flanking region	0.017	0.023	0.011
637	G>T	rs73558107	5' flanking region	0.043	0.043	0.042
1088	T>G		intron 1	0.006	0.012	0.000
1130	T>C		intron 1	0.005	0.000	0.011
1192	G>A		intron 1	0.006	0.000	0.011
1325_1327	del3		intron 1	0.047	0.023	0.073
1430_1431	ins1		intron 1	0.033	0.033	0.033
1702	G>A	rs12721102	intron 1	0.006	0.000	0.011
1719	C>A		intron 1	0.006	0.000	0.011
2099	G>T		intron 1	0.011	0.011	0.010
2467	C>T		intron 1	0.017	0.011	0.023
2607	G>A	rs5156	intron 1	0.016	0.011	0.021
2640	C>T	rs5158	intron 1	0.033	0.022	0.043
2641	G>A		intron 1	0.005	0.011	0.000
2678	G>C		intron 1	0.005	0.011	0.000
2767	G>T	rs12721107	intron 1	0.017	0.000	0.035
3348	G>A		intron 1	0.006	0.000	0.012
3363	G>A		intron 1	0.043	0.038	0.048
3592	C>T	rs12691090	exon 2 (non-synonymous)	0.027	0.011	0.042
3700	G>A		intron 2	0.005	0.000	0.010
3792	G>A	rs5165	intron 2	0.017	0.012	0.000
3969	A>C		exon 3 (non-synonymous)	0.011	0.011	0.011
4012	G>A		exon 3 (non-synonymous)	0.005	0.011	0.000
4533	C>T		3' flanking region	0.011	0.011	0.011
4912	G>C		3' flanking region	0.006	0.120	0.000

Yellow highlighted variants represent variants that were uniquely present in low HDL-C group, while green highlighted variants represent variants that were uniquely present in high HDL-C group.

3.3 LD AND TAGGER ANALYSIS OF *APOC4* VARIANTS

3.3.1 Non-Hispanic Whites

LD and tagger analysis (by using Haploview software) of 13 common variants with MAF≥ 0.049 and r^2 cutoff 0.9 identified 6 tag-SNPs Bins in NHWs (Table 13). Figure 5 shows LD plot of common variants identified in our study in NHWs.

Table 13 Tagger Results of Common Variants Identified in NHWs by Using Haploview

Bin	Test	Alleles Captured	Genotyped SNPs
1	4746	4746, 204,**3498** , **2623**, 4661	3498 (rs1132899) 2623 (rs5157)
2	108	92_94del3, 108, 150_151ins114	--------
3	1325_1327del3	1325_1327del3, **2640**	2640 (rs5158)
4	1733	1733	--------
5	3927	**3927**	3927(rs5167)
6	1823	1823	---------

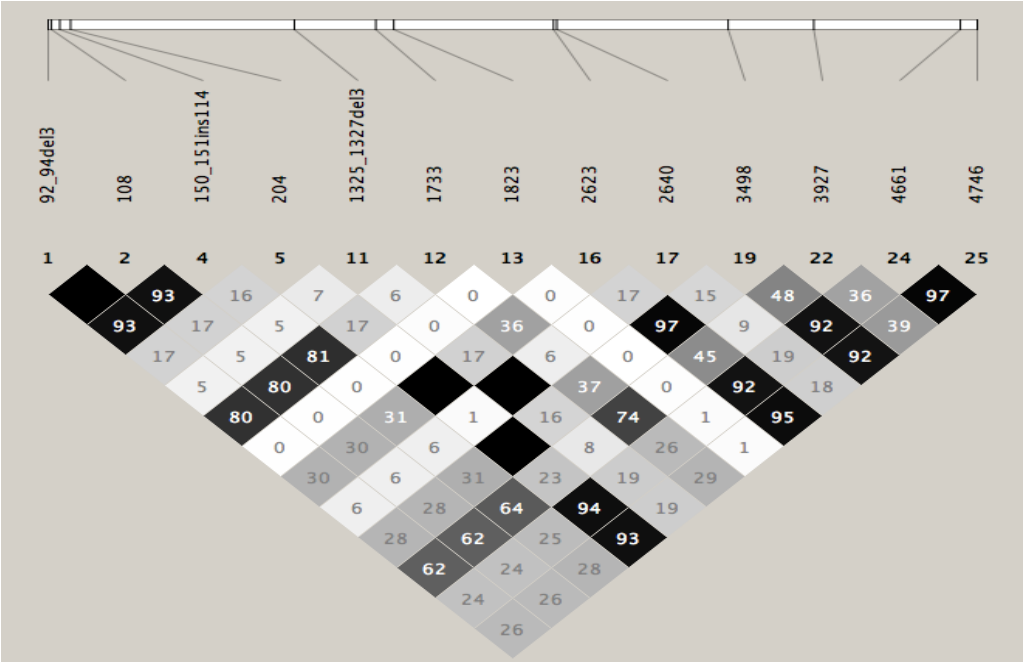

Figure 5 LD Plot of Common Variants Identified in NHWs

3.3.2 African Blacks

LD and tagger function from haploview analysis by using r^2 cutoff 0.9 of 21 common variants identified in our study in Blacks yielded 15 Bins of tag-SNPs (Table 14). Figure 6 illustrates the LD pattern of common variants in Black.

Table 14 Tagger Results of Common Variants Identified in Blacks

Bin	Test	Alleles Captured	Genotyped SNPs
1	92_94del3	92_94del3,150_151ins114,108	----------
2	3502	3502,3213,757	----------
3	3380	2971,**3380**	3380(rs12721104)
4	204	204,**3498**	3498 (rs1132899)
5	4579	4579	----------
6	4154	4154	----------
7	4746	4746	----------
8	1823	1823	----------
9	2559	**2559**	2559 (rs5155)
10	4157	**4157**	4157(rs10425530)
11	4661	4661	----------
12	2623	**2623**	2623(rs5157)
13	4844	4844	----------
14	3927	**3927**	3927(rs5167)
15	1733	1733	----------

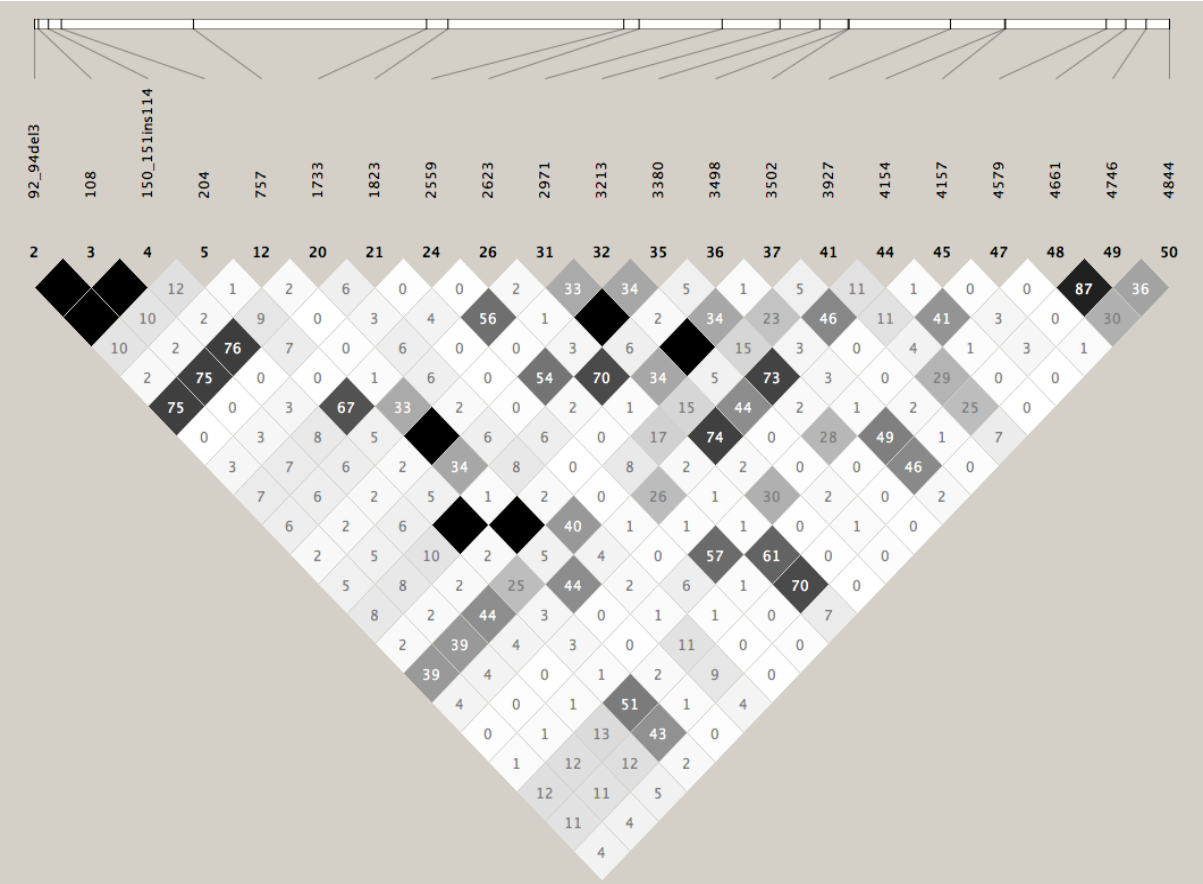

Figure 6 LD Plot of Common Variants Identified in Blacks

3.4 GENOTYPING COMMON VARIANTS IN THE TOTAL NHW AND BLACK SAMPLES

We genotyped seven common variants in Blacks and four common variants in NHWs using premade TaqMan SNPs genotyping assays. Genotyping features for all genotyped SNPs are summarized in Table 15.

Table 15 TaqMan SNPs Genotyping Features

Ref SNP ID	*Position	location	Hapmap MAF%		NHWs		African Blacks	
			CEU	YRI	Call Rate	HWE p-value	Call Rate	HWE p-value
rs1132899	3498	missense	(T) 0.392	(T) 0.291	0.9919	0.655	0.986	0.699
rs5158	2640	intron	(T) 0.092	(T) 0.025	0.9919	0.919	0.979	1.000
rs5157	2623	intron	(C) 0.448	(T) 0.203	0.9967	0.824	0.97	0.772
rs5155	2559	intron	(T) 0.000	(T) 0.067	-----	------	0.9809	1.000
rs10425530	4157	3'UTR	(A) 0.000	(A) 0.100	-----	------	0.973	0.775
rs5167	3927	intron	(G) 0.328	(G) 0.404	0.9919	0.266	0.9809	0.549
rs12721104	3380	missense	(A) 0.000	(A) 0.086	-----	------	0.9733	1.000

* Positions according to Seattle ref sequence.

None of the variants deviated from HWE (p-value >0.05) in either sample. LD pattern of the common variants screened in the total NHW and Black samples were repeated which yielded similar pattern of LD as seen in subset of the sequenced individuals. Figure 7 and Figure 8 show LD pattern of common variants screened in the entire NHW and Black samples, respectively.

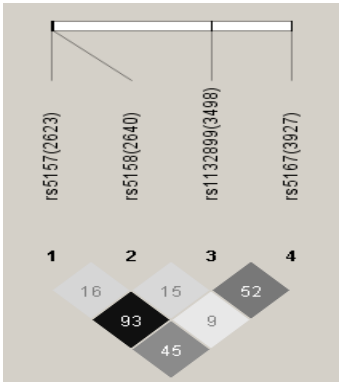

Figure 7 LD Plot of Common Variants Screened in the Total NHW Sample

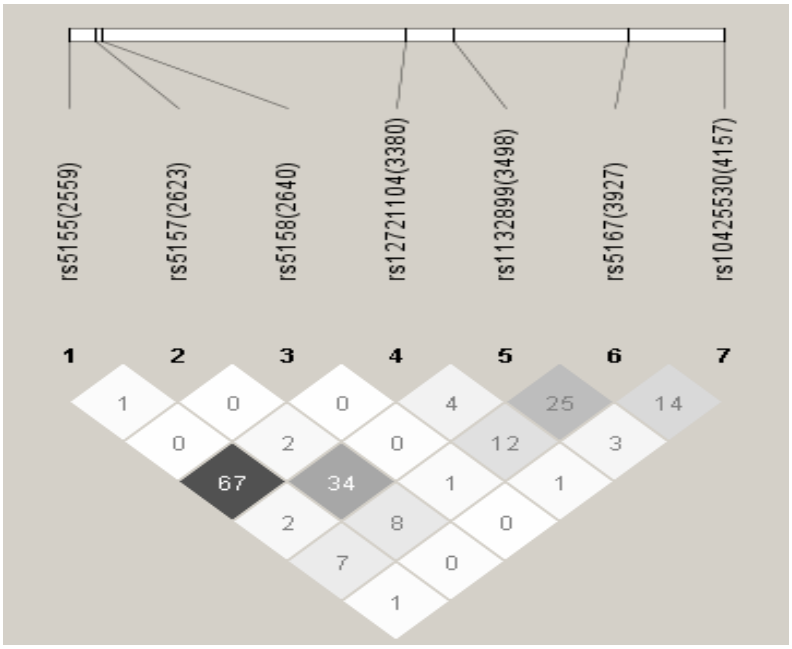

Figure 8 LD Plot of Common Variants Screened in the Total Black Sample

3.4.1 Association Analysis of the Common Variants Genotyped in the Total NHW and Black Samples

Table 16 and Table 17 show the genotype count, covariate adjusted mean (for each genotype) for four lipid traits (cholesterol, HDL-C, LDL-C and TG), and adjusted p-value for each variant genotyped in the total NHW and Black samples, respectively. The variation in individual number is a factor of genotype data and lipid profile availability. Both additive and dominant models were used in the association analysis. No statistically significant association was found between Tag-SNPs genotyped and lipid traits in NHWs. We checked the variants that had significant association in African Blacks: rs5158 and rs5167. Although no significant p-values were observed for rs5158 (intron 1, MAF=0.153) with TG (p=0.148) and for rs5167 (exon 3, MAF=0.339) with LDL-C (p=0.127), trend in lipid trait was observed with minor allele heterozygosity. In the Black samples, significant association was observed between rs5158 (intron 1, MAF=0.033) and TG (p=0.030). Borderline associations were between rs5167 (exon 3, MAF=0.472) and LDL-C (p=0.0615), rs1132899 (exon 2, MAF=0.233) and total Cholesterol (p=0.091), and rs10425530 (3'UTR, MAF=0.1) and HDL-C (p=0.099).

Table 16 Association Analysis of Common SNPs Screened in the Total NHW Sample

rs5158				
Lipid Profile	Model	Genotype (count)	mean ± SE	P-value
Cholesterol	Additive	CC (456)	218.75± 1.98	0.167
		CT (146)	212.61± 3.34	
		TT (11)	217.93±12.07	
	Dominant	CC (456)	218.76± 1.97	0.121
		CT/TT (157)	212.98± 3.22	
*HDL-C	Additive	CC (456)	50.18± 0.60	0.159
		CT (149)	50.54± 1.01	
		TT (11)	59.52± 3.69	
	Dominant	CC (456)	50.19± 0.61	0.423
		CT/TT (160)	51.16± 0.96	
LDL-C	Additive	CC (451)	140.08± 1.84	0.224
		CT(147)	136.12± 3.08	
		TT(11)	133.67± 11.15	
	Dominant	CC(451)	140.08± 1.84	0.229
		CT/TT (158)	135.94± 2.97	
*TG	Additive	CC (454)	144.64± 3.30	0.148
		CT(148)	136.15± 5.54	
		TT(11)	124.40± 20.14	
	Dominant	CC (454)	144.63± 3.3	0.183
		CT/TT (159)	135.33± 5.34	
rs1132899				
Lipid Profile	Model	Genotype (count)	mean ± SE	P-value
Cholesterol	Additive	CC (158)	218.45± 3.23	0.997
		CT (313)	216.17± 2.34	
		TT (142)	218.58± 3.44	
	Dominant	CC (158)	218.44± 3.32	0.680
		CT/TT (455)	216.92± 1.97	
*HDL-C	Additive	CC (159)	51.33± 0.98	0.538
		CT (315)	49.84± 0.71	
		TT (142)	50.37± 1.05	
	Dominant	CC (159)	51.33± 0.98	0.277
		CT/TT (457)	50.0± 0.60	
LDL-C	Additive	CC (158)	139.25± 2.99	0.989
		CT (311)	139.09± 2.17	
		TT (140)	139.20± 3.21	
	Dominant	CC (158)	139.25± 2.99	0.971
		CT/TT (451)	139.12± 1.83	
*TG	Additive	CC (159)	139.61± 5.33	0.290
		CT (313)	141.31± 3.87	
		TT (141)	147.16± 5.71	
	Dominant	CC (159)	139.58± 5.32	0.466
		CT/TT (454)	143.29± 3.26	

Table 16 (Continued)

rs5157				
Lipid Profile	Model	Genotype (count)	mean ± SE	P-value
Cholesterol	Additive	CC (153)	216.45± 3.27	0.903
		CT (312)	217.77± 2.35	
		TT (151)	217.00± 3.33	
	Dominant	CC (153)	216.45± 3.27	0.774
		CT/TT (463)	217.52± 1.95	
*HDL-C	Additive	CC (154)	51.14± 1.00	0.595
		CT (314)	50.06± 0.72	
		TT (151)	50.23± 1.02	
	Dominant	CC (154)	51.14± 1.00	0.414
		CT/TT (465)	50.11± 0.60	
LDL-C	Additive	CC (153)	137.55± 3.02	0.903
		CT (310)	140.32± 2.18	
		TT (149)	138.02± 3.10	
	Dominant	CC (153)	137.56± 3.02	0.559
		CT/TT (459)	139.58± 1.82	
*TG	Additive	CC (154)	138.98± 5.43	0.322
		CT (312)	142.11± 3.91	
		TT (150)	145.74± 5.56	
	Dominant	CC (154)	138.97± 5.43	0.390
		CT/TT (462)	143.29± 3.26	
rs5167				
Lipid Profile	Model	Genotype (count)	mean ± SE	P-value
Cholesterol	Additive	TT (245)	214.52± 2.60	0.123
		TG (296)	218.45± 2.44	
		GG (72)	221.92± 4.74	
	Dominant	TT (245)	214.53± 2.60	0.162
		TG/GG (368)	219.16± 2.18	
*HDL-C	Additive	TT (247)	50.74± 0.79	0.770
		TG (297)	49.74± 0.74	
		GG (72)	51.10± 1.44	
	Dominant	TT (247)	50.75± 0.79	0.500
		TG/GG (369)	50.02± 0.66	
LDL-C	Additive	TT (244)	136.23± 2.41	0.127
		TG (293)	140.77± 2.27	
		GG (72)	142.20± 4.38	
	Dominant	TT (244)	136.24± 2.41	0.117
		TG/GG (365)	141.07± 2.03	
*TG	Additive	TT (245)	142.50± 4.43	0.976
		TG (296)	141.92± 4.06	
		GG (72)	144.30± 7.91	
	Dominant	TT (245)	142.51± 4.33	0.964
		TG/GG (368)	142.41± 3.64	

* Log transformed value.

Table 17 Association Analysis of Common SNPs Screened in the Total Black Sample

rs1132899				
Lipid profile	Model	Genotype(count)	mean ± SE	P-value
*Cholesterol	Additive	CC (430)	172.89± 1.80	0.389
		CT (280)	178.62± 2.25	
		TT (39)	166.37± 5.87	
	Dominant	CC (430)	172.92± 1.80	0.092
		CT/TT (319)	177.15± 2.12	
**HDL-C	Additive	CC (427)	48.75± 0.61	0.656
		CT (279)	48.85± 0.76	
		TT (39)	46.91± 1.99	
	Dominant	CC (427)	48.76± 0.61	0.907
		CT/TT (318)	48.62± 0.72	
**LDL-C	Additive	CC (432)	109.98± 1.60	0.402
		CT (281)	115.13± 1.99	
		TT (39)	105.15± 5.22	
	Dominant	CC (432)	110.00± 1.60	0.106
		CT/TT (320)	113.94± 1.88	
*TG	Additive	CC (437)	69.29± 1.59	0.987
		CT(281)	69.64± 1.99	
		TT (40)	70.31± 5.15	
	Dominant	CC (437)	69.29± 1.59	0.918
		CT/TT (321)	69.72± 1.87	
rs5158				
Lipid profile	Model	Genotype (count)	mean ± SE	P-value
*Cholesterol	Additive	CC (712)	174.63± 1.45	0.955
		CT (32)	173.76± 6.46	
		TT (0)		
	Dominant	CC (712)	174.63± 1.45	0.955
		CT/TT (32)	173.76± 6.46	
**HDL-C	Additive	CC (708)	48.68± 0.49	0.564
		CT (32)	47.64± 2.19	
		TT (0)		
	Dominant	CC (708)	48.68± 0.49	0.564
		CT/TT (32)	47.64± 2.19	
**LDL-C	Additive	CC (716)	111.47±1.28	0.450
		CT (31)	115.41±5.84	
		TT (0)		
	Dominant	CC (716)	111.47± 1.28	0.450
		CT/TT (31)	115.41± 5.84	
*TG	Additive	CC (721)	69.88± 1.27	0.031
		CT (32)	57.65± 5.70	
		TT (0)		
	Dominant	CC (721)	69.88± 1.27	0.031
		CT/TT (32)	57.65± 5.70	

Table 17 (Continued)

rs5157				
Lipid profile	Model	Genotype (count)	mean ± SE	P-value
*Cholesterol	Additive	CC (503)	173.50± 1.70	0.324
		CT (215)	178.00± 2.57	
		TT (20)	167.71± 8.24	
	Dominant	CC (503)	173.51± 1.70	0.155
		CT/TT (235)	177.11± 2.46	
**HDL-C	Additive	CC (500)	48.62± 0.57	0.765
		CT (214)	49.36± 0.86	
		TT (20)	43.49± 2.77	
	Dominant	CC (500)	48.63± 0.57	0.766
		CT/TT (234)	48.85± 0.83	
**LDL-C	Additive	CC (503)	110.59± 1.51	0.301
		CT (218)	114.31± 2.27	
		TT (20)	109.12± 7.32	
	Dominant	CC (503)	110.59± 1.51	0.194
		CT/TT (238)	113.87± 2.17	
*TG	Additive	CC (509)	69.00± 1.48	0.795
		CT (217)	69.65± 2.25	
		TT (21)	73.23± 7.06	
	Dominant	CC (509)	69.00± 1.48	0.884
		CT/TT (238)	69.97± 2.15	
rs5155				
Lipid profile	Model	Genotype (count)	mean ± SE	P-value
*Cholesterol	Additive	CC (608)	175.22± 1.53	0.153
		CT (131)	172.20± 3.19	
		TT (6)	148.85± 14.73	
	Dominant	CC (608)	175.21± 1.53	0.267
		CT/TT (137)	171.15± 3.12	
**HDL-C	Additive	CC (604)	48.78± 0.52	0.384
		CT (131)	48.64± 1.09	
		TT (6)	38.80± 5.02	
	Dominant	CC (604)	48.78± 0.52	0.650
		CT/TT (137)	48.20± 1.06	
**LDL-C	Additive	CC (613)	111.97± 1.37	0.240
		CT (129)	109.61± 2.87	
		TT (6)	95.56± 13.17	
	Dominant	CC (613)	111.96± 1.37	0.327
		CT/TT (135)	108.97± 2.81	
*TG	Additive	CC (616)	69.80± 1.37	0.534
		CT (132)	68.87± 2.87	
		TT (6)	72.10± 13.27	
	Dominant	CC (616)	69.80± 1.37	0.495
		CT/TT (138)	69.01± 2.80	

Table 17 (Continued)

rs10425530				
Lipid profile	Model	Genotype (count)	mean ± SE	P-value
*Cholesterol	Additive	GG (582)	174.82± 1.59	0.582
		GA (149)	176.07± 3.05	
		AA (8)	152.37± 13.13	
	Dominant	GG (582)	174.81± 1.60	0.854
		GA/AA (157)	174.92± 2.99	
**HDL-C	Additive	GG (581)	48.24± 0.54	0.187
		GA (146)	50.34± 1.04	
		AA (8)	45.01± 4.44	
	Dominant	GG (581)	48.24± 0.54	0.099
		GA/AA (154)	50.07± 1.02	
**LDL-C	Additive	GG (584)	112.12± 1.42	0.490
		GA (150)	112.04± 2.71	
		AA (8)	94.23± 11.69	
	Dominant	GG (584)	112.12± 1.42	0.700
		GA/AA (158)	111.17± 2.65	
*TG	Additive	GG (590)	69.68± 1.40	
		GA (151)	68.56± 2.68	0.594
		AA (8)	67.19± 11.61	
	Dominant	GG (590)	69.68± 1.40	0.665
		GA/AA (159)	68.50± 2.62	
rs12721104				
Lipid profile	Model	Genotype (count)	mean ± SE	P-value
*Cholesterol	Additive	GG (553)	174.64± 1.63	0.833
		GA (173)	175.22± 2.81	
		AA (13)	168.34± 10.10	
	Dominant	GG (553)	174.63± 1.63	0.964
		GA/AA (186)	174.72± 2.71	
**HDL-C	Additive	GG (549)	48.75± 0.55	0.357
		GA (173)	48.66± 0.94	
		AA (13)	42.96± 3.39	
	Dominant	GG (549)	48.73± 0.55	0.639
		GA/AA (186)	48.25± 0.91	
**LDL-C	Additive	GG (557)	111.58± 1.45	0.980
		GA (172)	112.15± 2.52	
		AA (13)	109.50± 9.02	
	Dominant	GG (557)	111.57± 1.45	0.934
		GA/AA (185)	111.96± 2.42	
*TG	Additive	GG (560)	68.73± 1.44	0.187
		GA (175)	71.43± 1.49	
		AA (13)	79.72± 8.99	
	Dominant	GG (560)	68.75± 1.44	0.225
		GA/AA (188)	72.02± 2.40	

Table 17 (Continued)

rs5167				
Lipid profile	Model	Genotype (count)	Mean ± SE	P-value
*Cholesterol	Additive	TT (219)	177.28± 2.52	0.117
		TG (361)	174.96± 1.99	
		GG (165)	171.93±2.88	
	Dominant	TT (219)	177.28± 2.52	0.195
		TG/GG (526)	173.99± 1.66	
**HDL-C	Additive	TT (219)	48.21± 0.85	0.665
		TG (359)	48.85± 0.68	
		GG (163)	48.70± 0.98	
	Dominant	TT (219)	48.21± 0.85	0.533
		TG/GG (522)	48.81± 0.56	
**LDL-C	Additive	TT (220)	114.85± 2.24	0.062
		TG (363)	112.06± 1.77	
		GG (165)	108.47± 2.57	
	Dominant	TT (220)	114.85± 2.24	0.137
		TG/GG (528)	110.92± 1.48	
*TG	Additive	TT (222)	67.84± 2.20	0.248
		TG (367)	69.40± 1.74	
		GG (166)	70.71± 2.54	
	Dominant	TT (222)	67.84± 2.20	0.329
		TG/GG (533)	69.81± 1.46	

* Cholesterol and TG values are log transformed.
** HDL and LDL values are square root transformed.

4.0 DISCUSSION

The *APOC4* gene belongs to the apolipoprotein family, which plays an essential role in lipid metabolism, and is located within the *APOE/C1/C4/C2* gene cluster on chromosome 19q13.2. ApoCIV is like other apolipoproteins, has two α-helical amphipathic structures that enable apolipoproteins binding to lipid molecules. The presence of apoCIV on VLDL and HDL lipoprotein particles strongly suggests its involvement in VLDL and HDL metabolisms.

Although the exact function of apoCIV in lipid metabolism is not well known, overexpression of *APOC4* in transgenic mice yields hypertrigylceridemic phenotype, suggesting its important role in TG metabolism. Meta analysis of Linkage studies on families with diverse ethnicity revealed a broad linkage peak on chromosome 19 (19p13-19q13.24) with different lipid traits, which suggest that complex genetic interaction among different genes in this linkage region might have direct impact on modulating lipid profile (Malhotra et al., 2007). Since *APOC4* is part of the *APOE/C1/C4/C2* gene cluster in the linked region on chromosome 19, genetic variations in the *APOC4* gene alone or in conjunction with other candidate genes in the region might affect plasma lipid profile.

Extensive genetic association studies have been performed on the other apolipoprotein genes in this gene cluster, especially on *APOE*. In order to fully evaluate the genetic role of these genes in lipid metabolism, it is important to examine all apolipoprotein genes in this gene cluster. To our knowledge, only one previous study from our group sequenced the coding and intron-

exon boundaries of the *APOC4* gene in randomly selected 50 NHWs individuals and identified five point mutations (Kamboh et al., 2000). However, since individuals were not selected with regards to their lipid profile and the entire gene was not sequenced, the previous study fell short of identifying all possible variants in the *APOC4* gene.

To our knowledge, this is the first study that aimed to resequence the entire *APOC4* gene with its flanking sequence in selected individuals with extreme lipid profile in order to identify potential all common and rare variants in *APOC4* that are associated with lipid level.

4.1 COMPARISON OF OUR STUDY RESULTS WITH PUBLICLY AVAILABLE DATABASES

We compared our results to publicly available databases (SeattleSNPs and CHIP). SeattleSNPs sequenced a total of 4,510 bp genomic region of *APOC4,* including 956 bp in the 5'flanking region and 297 bp in the 3'flanking region in 48 African Americans (24 from Jackson, Mississippi and 24 from Coreill, New Jersey), and 48 individuals of European Descent (24 from Rochester, Minnesota and 24 from North Karelia, Finland). All individuals were selected randomly irrespective of their lipid level. In this study, we sequenced a total of 5,083 bp genomic region of *APOC4*, including all exons and introns plus 956 bp in the 5'flanking region, and 876 bp in the 3'flanking region in a total of 95 NHWs and 95 Blacks having extreme low and high HDL-C levels.

We noticed difference in our results compared to publicly available databases (SeattleSNPs and CHIP), which are more likely due to our larger sample size, selection criteria

or our larger sequencing genomic region. Furthermore, we used African sample compared to African Americans used in SeattleSNPs that have considerable white admixture. For variants that were identified in our study but not reported in databases, could be due to our larger sample size and selection criteria.

A total of 30 variants have been previously reported in publicly available databases (SeattleSNPs or CHIP) compared to a total of 65 variants identified in this study. In public databases (SeattleSNPs and CHIP) 15 variants have been reported in Europeans with 6 and 9 variants having MAF<5%, and MAF ≥ 5%, respectively. By comparison, a total of 26 variants plus a dinucleotide microsatellite were identified in this study in NHWs with 13 variants having MAF<5%, and 13 variants having MAF≥ 5%. A total of 25 variants have been reported in databases in African American with 10 and 15 variants having MAF < 5%, and MAF≥5%, respectively. By comparison, we found 51 variants plus a dinucleotide microsatellite in Blacks; 30 variants were observed at MAF<5%, and 21 variants were observed at MAF≥ 5%. We found all the reported common variants in African Americans and Europeans in our study, except 3 reported rare variants in Europeans (2703G>T/ MAF=0.025, 2753G>A/ MAF=0.02, and 3532A>G/ MAF=0.008) and 4 reported rare variants in African Americans, (2683G>A/ MAF=0.02, 2886T>C/ MAF=0.02, 3532A>G/ MAF=0.017, and 4346A>T/ MAF=0.02). We identified 39 new variants in this study, 14 of them were present in NHWs and 28 in African Blacks.

Five variants affecting the coding region were reported in databases (3 non-synonymous, and 2 synonymous), including 3498C>T (Pro36Leu) in exon 2, 3502C>T (Ser37Ser) in exon 2, 3546G>A (Gly52Asp) in exon 2, 3592C>T (Asp67Asp) in exon 2, and 3927T>G (Leu96Arg) in Exon 3. Of the coding variants, 3502C>T (Ser37Ser) and 3592C>T (Asp67Asp) were present in

African Americans only, 3546G>A (Gly52Asp) in European only, and 3498C>T (Pro36Leu) and 3927T>G (Leu96Arg) in both populations. In this study, we identified all the 5 reported exonic variants plus two additional non-synonymous variants among African Blacks, including 3969A>C (Lys110Tyr) in exon 3 and 4012G>A (Lys124Arg) in exon 3.

Three main factors may explain our ability in identifying more variants in this study than in previous studies: 1) our sequencing sample size was larger, 2) individuals were selected based on their extreme low or high HDL-C levels, and 3) our larger sequencing genomic region. In comparison to public databases, 5 unique variants were identified in Europeans, and 15 unique variants in African Americans. We found 13 variants unique to NHWs, and 38 variants unique to Blacks. Similar to publicly available databases, we found more unique variants in Blacks than in NHWs.

In our study, we found 4 indels (2 deletions, and 2 insertions), including 92_94del3, 150_151ins114, 1325_1327del3, and 1430_1431ins1 in which the first three indels were present in both samples (NHW and African Black), and the last insertion was present only in Blacks. By comparison, the databases reported only one deletion (92_94del3), which was common in both European and African American populations. The largest insertion, 150_151ins114 was LINE-1 (long interspersed nuclear element-1), which was in complete LD with previously reported 92_94del3. Another variant, 108G>A, in addition to 150_151ins114 was in strong LD with 92_94del3 and both of them were not reported in databases.

4.2 DISTRIBUTION OF *APOC4* VARIANTS IN HIGH AND LOW HDL-CHOLESTEROL GROUPS

The distribution of unique rare or less common variants (MAF<5%) was similar in both samples (NHW and African Black) in which more unique rare variants were present in the low HDL-C group than in the high HDL-C group (8 vs. 2 in NHWs and 10 vs. 7 in African Blacks). Accumulation of rare variants in the low HDL-C group more likely has damaging effect (Cohen et al., 2004). Although the distribution of unique rare variants in low and high HDL-C groups was similar in both NHW and Black samples, the percentage of individuals with rare or less common variants in the low HDL-C group was higher in NHWs and lower in Blacks compared to the high HDL-C group.

In NHWs, 31.3% of individuals in the low HDL-C group (15/48) had rare or less common variants versus 10.6% in the high HDL-C group (5/47). In NHWs, rare variants could have a protective feature against fatal illness that enable individuals with extremely low HDL-C survive. Conversely, in Blacks we observed more individuals with rare or less common variants in the high HDL-C group (26/48) than in the low HDL-C group (22/47) (54% vs. 46%). The difference in the rare or less common variants distribution among NHWs and Blacks could explain the variation in HDL-C level in which Blacks have higher HDL-C level than Whites (American Heart Association, 2002). Furthermore, different LD pattern in both samples (NHW and African Black) could explain the variation in rare or less common variants distribution among low and high HDL-C groups in NHWs and African Blacks.

4.3 COMPARISON OF OUR STUDY RESULTS WITH PUBLISHED LITERATURE

One recent prospective study was conducted to test the association between common variants within *APOE/C1/C4/C2* gene cluster and lipid traits (Ken-Dror et al., 2010). A total of 9 common SNPs within this cluster, including *APOE/* rs405509, *APOE/* rs439401, *APOC1*/ rs4420638, *APOC1*/ rs4800770, *APOC1*/ rs7259004, *APOC4*/ rs12691089, *APOC4*/ rs5167, *APOC2*/ rs5127, *APOC2*/ rs10413089 were genotyped in 2,767 middle-aged men from the Second Northwick Park Heart Study (NPHSII). Only two *APOC4* common exonic SNPs were included in the genotyped SNPs; rs12691089 (Gly52Asp) in exon 2, and rs5167 (Leu96Arg) in exon 3. rs5167 was strongly associated with high HDL-C and apoAI levels ($p<0.001$), and LDL-C ($p= 0.03$) (Ken-Dror et al., 2010).

By comparison, we did not observe striking association with HDL-C or LDL-C in NHWs but a borderline association was observed with LDL-C ($p=0.127$). Similarly, a borderline association was observed between rs5167 and LDL-C ($p=0.061$) in Blacks. Although we did not find an association between this variant and HDL-C in NHWs and Blacks ($p=0.770$ and 0.532, respectively), in the sequencing analysis the minor allele frequency was higher in the high HDL-C group than in the low HDL-C group in both NHWs and African Blacks (0.372 vs. 0.309 in NHWs and 0.554 vs. 0.500 in Blacks), which was consistent with the previous results. Our small sample size might explain our inability in detecting the small effect size of some common variants or the modest effect size of some rare variants. Although there was no significant

association between rs12691089 (Gly52Asp) and HDL-C levels in Ken-Dror et al. (2010), we observed this SNP in the low HDL-C group only in NHWs in this study.

Previously, Kamboh et al. (2000) reported association of *APOC4* exonic variants: 3498T>C/ rs1132899 (Pro36Leu) and 3927T>G/ rs5167 (Leu96Arg) and high TG level in NHW women (p=0.03 and 0.08, respectively). Another association was found in men between 979G>A and HDL-C (p=0.06) (Kamboh et al., 2000). Although we did not find statistically significant association between these genetic variants, we observed a trend of association of 3498T>C (rs1132899) and 3927T>G (rs5167) and TG in NHWs which was consistent with the previous findings. Although we used the same NHW population with similar sample size (n=623 in this study vs. n=592 in Kamboh et al. (2000)), we did not observe significant association of rs1132899 and rs5167 with TG. One of the factors that might explain our limitation in detecting the reported associations is our association analysis method. Due to gender-specific effect, we did not expect to find this association in the entire sample since Joint gender analysis was performed in this study. So, stratifying based on gender could be more powerful in detecting gender-specific effect.

4.4 CONCLUSIONS AND FURTHER DIRECTIONS

Coronary heart disease continues to be a leading cause of premature mortality and co-morbidity in western countries and dyslipidemia with high LDL-C and low HDL-C as major risk factors for CHD. Little attention has been paid to the role of *APOC4* genetic variants in relation to plasma lipid levels, although it is an important biological and positional candidate gene that is part of the *APOE/C1/C4/C2* gene cluster on chromosome 19q13.2. Even though there is no known major role of apoCIV in HDL metabolism, this study was undertaken to catalogue *APOC4* common and rare variants and to assess their effects on lipid traits.

Our findings were consistent with previous results. For the reported associations, even though we did not obtain statistically significant associations, which could be due to our small sequencing sample size, we did observe trend in the corresponding lipid trait in the same direction. We observed borderline associations in some variants with different lipid traits that were unreported previously and need to be confirmed by genotyping in the total samples. More unique rare variants were observed in the low HDL-C group than the high HDL-C group in both samples (NHWs and Blacks). Further studies in larger sample size are warranted in order to confirm the previously reported association signals and borderline association observed in this study. Furthermore, it is important to examine the distribution of unique rare variants in the total sample, and to detect more rare variants with modest effect size or common variants with small effect size. Since *APOC4* is part of the *APOE/C1/C4/C2* gene cluster, a joint analysis of all these genes would provide more information about their role in lipid metabolism.

BIBLIOGRAPHY

Aalto-Setala K, Fisher EA, Chen X, Chajek-Shaul T, Hayek T, Zechner R, Walsh A, Ramakrishnan R, Ginsberg HN, Breslow JL. Mechanism of hypertriglyceridemia in human apolipoprotein (apo) CIII transgenic mice. Diminished very low density lipoprotein fractional catabolic rate associated with increased apoCIII and reduced apoE on the particles. *J Clin Invest.* 1992;90(5):1889-900.

Adames MR, Kinlay S, Blake G, Orford JL, Ganz P, Selwyn AP. Pathophysiology of atherosclerosis: development, regression, restenosis. *Curr Ather Rep*. 2000;2(3):251-8

Adeyemo AA, Johnson T, Acheampong J, Oli J, Okafor G, Amoah A, Owusu S, Agyenim-Boateng K, Eghan BA Jr, Abbiyesuku F, Fasanmade O, Rufus T, Doumatey A, Chen G, Zhou J, Chen Y, Furbert-Harris P, Dunston G, Collins F, Rotimi C. A genome wide quantitative trait linkage analysis for serum lipids in type 2 diabetes in an African population. *Atherosclerosis*. 2005;181(2):389-97.

Allan CM, Walker D, Taylor JM. Evolutionary duplication of a hepatic control region in the human apolipoprotein E locus: identification of a second region that confers high level and liver-specific expression of the human apolipoprotein E gene in transgenic mice. *J Biol Chem*. 1995;270(44):26278-81.

Allan CM, Taylor JM. Expression of a novel human apolipoprotein (apoC-IV) causes hypertriglyceridemia in transgenic mice. *J Lipid Res*. 1996;37(7):1510-8.

Allan CM, Walker D, Segrest JP, Taylor JM. Identification and characterization of a new human gene (APOC4) in the apolipoprotein E, C-I, and C-II gene locus. *Genomics*. 1995;28(2):291-300.

Allan CM, Taylor S, Taylor JM. Two hepatic enhancers, HCR.1 and HCR.2, coordinate the liver expression of the entire human apolipoprotein E/C-I/C-IV/CII gene cluster. *J Biol Chem*. 1997;272(46):29113-9.

Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid I, Pramstaller P, Penninx B, Janssens, Wilson A J, Spector T, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet*. 2009;41(1):47-55.

American Heart Association. High blood Cholesterol and Other Lipids. 2002.

American Heart Association. Heart Diseases and Stroke Statistics. 2009.

Applied Biosystems. TaqMan® Genotyping Master Mix Protocol. 2007.

Arya R, Duggirala R, Almasy L, Rainwater DL, Mahaney MC, Cole S, Dyer TD, Williams K, Leach RJ, Hixson JE, MacCluer JW, O'Connell P, Stern MP, Blangero J. Linkage of high-density lipoprotein-cholesterol concentrations to a locus on chromosome 9p in Mexican Americans. *Nat Genet*. 2002;30:102-5.

Assmann G, Schulte H, von Eckardstein A, Huang Y. High-density lipoprotein cholesterol as a predictor of coronary heart disease risk. The PROCAM experience and pathophysiological implications for reverse cholesterol transport. *Atherosclerosis.* 1996;124(S):11-20.

Boes E, Coassin S, Kollerits B, Heid IM, Kronenberg F. Genetic-epidemiological evidence on genes associated with HDL cholesterol levels: a systematic in-depth review. *Exp. Gerontol*. 2009;44(3):136-60.

Briel M, Ferreira-Gonzalez I, You JJ, Karanicolas PJ, Akl EA, Wu P, Blechacz B, Bassler D, Wei X, Sharman A, Whitt I, Alves da Silva S, Khalid Z, Nordmann AJ, Zhou Q, Walter SD, Vale N, Bhatnagar N, O'Regan C, Mills EJ, Bucher HC, Montori VM, Guyatt GH. Association between change in high density lipoprotein cholesterol and cardiovascular mortality and morbidity: systematic review and meta regression analysis. *BMJ*. 2009;338:b92.

Bunker CH, Ukoli FA, Matthews KA, Kriska AM, Huston SL, Kuller LH. Weight threshold and blood pressure in a lean black population. *Hypertension.* 1995;26(4):616-23.

Bunker GH, Ukoli FA, Okoro FI, Olumu AB, Kriska AM, Huston SL, Markovic N, Kuller LH. Correlates of serum lipids in a lean black population. *Atherosclerosis.* 1996;123(1):215-25.

Chilton RJ. Pathophysiology of Coronary Heart Disease: A Brief Review. *JAOA.* 2004;104(9): 5-8.

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple Rare Alleles Contribute to Low Plasma Levels of HDL Cholesterol. *Science*. 2004;305(5685):869-72.

Corella D, and Ordovas JM. Single nucleotide polymorphisms that influence lipid metabolism: Interaction with dietary factors. *Annu Rev Nutr*. 2005;25:341–90.

Davison PJ, Norton P, Wallis SC, Gill L, Cook M, Williamson R, Humphries SR. There are two gene sequences for human apolipoprotein C1 (APO C1) on chromosome 19, one of which is 4 kb from the gene for apoE. *Biochem Biophys Res Commun*. 1986;136(3):876-84.

DeSilva HV, Lauer SJ, Wang J, Simonet WS, Weisgraber KH, Mahley RW, Taylor JM. Overexpression of human apolipoprotein C-111 in transgenic mice results in an accumulation of apolipoprotein B48 remnants that is corrected by excess apolipoprotein E. *J Biol Chem.* 1994;269(3):2324-35.

Ferns G, Keti V. HDL cholesterol modulation and Its Impacts on the management of cardiovascular risk. *Ann Clin Biochem*. 2008;45(pt 2):122-8.

Fitzgerald M. Lipid metabolism pathway. Abcam plc. 1998-2010. http://www.abcam.com/index.html?pageconfig=resource&rid=11512

Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am J Med*. 1977;62(5):707-14.

Gordon D, Probstfield J, Garrison RJ, Neaton JD, Castelli WP, Knoke JD, Jacobs DR, Bangdiwala Sh, Tyroler HA. High-Density Lipoprotein Cholesterol and cardiovascular disease four prospective american studies. *Circulation*. 1989;79(1):8-15.

Griffin BA. Lipoprotein atherogenicity: an overview of current mechanisms. *Pro Nutr Soc*. 1999; 58(1):163-9.

Grundy SM. Absorption and Metabolism of Dietary Cholestrol. *Ann Rev Nutr*. 1983;3:71-96.

Hamman RF, Marshall JA, Baxter J, Kahn LR, Mayer EJ, Orleans M, Murphy JR, Lezotte DC. Methods and prevalence of non-insulin dependent diabetes mellitus in a biethnic Colorado population. The San Luis Valley Diabetes Study. *Am J Epidemiol.* 1989; 129(2):295-311.

Harris MR, Bunker CH, Hamman RF, Sanghera DK, Aston CE, Kamboh MI. Racial differences in the distribution of a low density lipoprotein receptor-related protein (LRP) polymorphism and its association with serum lipoprotein, lipid and apolipoprotein levels. *Atherosclerosis.* 1998;137(1):187-95.

Hennig B, Shasby DM, Spector AA. Exposure to fatty acid increases human low density lipoprotein transfer across cultured endothelial monolayers. *Circ Res*. 1985;57(5):776-80.

Hyson D. Rutledge JC, Berglund L. Postprandial lipemia and cardiovascular disease. *Curr Atheros Rep*. 2003;5(6):437-44.

Jong MC, Hofker MH, Havekes LM. Role of apoCs in lipoprotein metabolism. *Arteriovascler thromb vasc Biol.* 1999;19(3):472-84.

Kamboh MI, Aston CE, Hamman RF. DNA sequence variation in human apolipoprotein C4 gene and its effect on plasma lipid profile. *Atherosclerosis*. 2000;152(1):193-201.

Kashyap MS, Srivastava LS, Chen CY, Perisutti G, Campbell M, Lutmer RF, Glueck CJ. Radioimmunoassay of human apolipoprotein CII. A study in normal and hypertriglyceridemia subjects. *J Clin Invest.* 1977;60(1):171-80.

Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M. Six new loci associated with blood low-density lipoprotein, high-density lipoprotein. *Nat Gent*. 2008; 40(2):189-97.

Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet*. 2009;41(1):56–65.

Ken-Dror G, Talmud PJ, Humphries SE, Drenos F. APOE/C1/C4/C2 gene cluster genotypes, haplotypes and lipid levels in prospective coronary heart disease risk among UK healthy men. *Mol Med*. 2010;16(9-10):389-99.

Kim E, Li K, Lieu C, Tong S, Kawai S, Fukutomi T, Zhou Y, Wands J, Li J. Expression of apolipoprotein C-IV is regulated by Ku antigen/peroxisome proliferator-activated receptor gamma complex and correlates with liver steatosis. *J Hepatol*. 2008;49(5):787-98.

King MD. Intestinal uptake of Lipids [internet]. 2010 [cited Oct 20, 2010]. Available from: http://themedicalbiochemistrypage.org/lipoproteins.html

Klos KL, Kullo IJ. Genetic determinants of HDL: monogenic disorders and contributions to variation. *Curr Opin Cardiol*. 2007;22(4):344-51.

Kotite L, Zhang LH, Yu Z, Burlingame AL, Havel RJ. Human apoC-IV: isolation, characterization, and immunochemical quantification in plasma and plasma lipoproteins. *JLR*. 2003;44(7):1387-94.

Kort EN, Ballinger DG, Ding W, Hunt SC, Bowen BR, Abkevich V, Bulka K, Campbell B, Capener C, Gutin A, Harshman K, McDermott M, Thorne T, Wang H, Wardell B, Wong J, Hopkins PN, Skolnick M, Samuels M. Evidence of linkage of familial hypoalphalipoproteinemia to a novel locus on chromosome 11q23. *Am J Hum Genet*. 2000;66(6):1845-56.

Krauss R. Heterogeneity of plasma low density lipoproteins and atherosclerosis risk. *Curr Opin Lipidol*. 1991;5(5):339-49.

Kris-Etherton P, Etherton T. The Role of Lipoprotein in lipid metabolism of meat animals. *J Anim Sci.* 1982;55(4):804-17.

Lauer S, Walker D, Elshourbagy NA, Reardon CA, Levy-Wilson B, Taylor JM. Two copies of the human apolipoprotein C-I gene are linked closely to the apolipoprotein E gene. *J Biol Chem*. 1988;263(15):7277-86.

Life Extension Foundation [Internet] .1995-2010 [cited Oct 18, 2010]. Available from: http://www.lef.org/protocols/heart_circulatory/coronary_artery_disease_atherosclerosis_01.htm

Li WH, Tanimura M, Luo CC, Datta S, Chan L. The apolipoprotein multigene family: biosynthesis, structure, structure-function relationships, and evolution. *J Lipid Res*. 1988;29(3):245-71.

Lusis AJ, Mar R, Pajukanta P. Genetics of atherosclerosis. *Annu Rev Genomics Hum Genet.* 2004;5:189-218.

Mahely RW, Innerarity TL, Rall SC, Weisgraber KH. Plasma lipoproteins: apolipoprotein structure and Function. *J Lip Res*. 1984;25(12):1277-94.

Mak PA, Laffitte BA, Desrumaux C, Joseph SB, Curtiss LK, Mangelsdorf DJ, Tontonoz P, Edwards PA. Regulated expression of the apolipoprotein E/C-I/C-IV/C-II gene cluster in murine and human macrophage. *JBC*. 2002;277(35):31900-8.

Malhotra A, Elbein SC, Ng MC, Duggirala R, Arya R, Imperatore G, Adeyemo A, Pollin TI, Hsueh WC, Chan JC, Rotimi C, Hanson RL, Hasstedt SJ, Wolford JK. Meta-analysis of genome-wide linkage studies of quantitative lipid traits in families ascertained for type 2 diabetes. *Diabetes*. 2007;56(3):890-6.

Miles B. Review of lipoproteins [Internet]. 2003 [cited Oct 30, 2010]. Available from: http://www.tamu.edu/classes/bmiles/lectures/Lipid%20Transport.pdf

Miller M, Zhan M. Genetic determinants of low high- density lipoprotein cholesterol. *Curr Opin Cardiol*. 2004;19(4):380-4.

Mozaffarian D, Micha R, Wallace S. Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: A systematic review and meta-analysis of randomized controlled trials. *PLoS Med*. 2010;7(3):e1000252.

Myklebost O, Rogne S. The gene for human apolipoprotein C1 is located 4.3 kilobases away from the apolipoprotein E gene on chromosome 19. *Hum Genet*. 1986;73(4):286-9.

North KE, Goring HH, Cole SA, Diego VP, Almasy L, Laston S, Cantu T, Howard BV, Lee ET, Best LG, Fabsitz RR, MacCluer JW. Linkage analysis of LDL Cholesterol in American Indian population: The strong heart family study. *J Lip Res*. 2006;47(1):59-66.

Park H. Kim JI, Cho SI, Sung J, Kim HL, Ju YS, Bayasgalan G, Lee MK, Seo JS. Genome-wide linkage study for plasma HDL cholesterol level in an isolated population of Mongolia. *Genomic inform*. 2008; 6(1):8-13.

Peschle C, Mavilio F, Care A, Migliaccio G, Migliaccio AR, Salvo G, Samoggia P, Petti S, Guerriero R, Marinucci M, Lazzaro D, Russo G, Mastrober-ardino G. Haemoglobin switching in human embryos: asynchrony of zeta----alpha and epsilon----gamma-globin switches in primitive and definite erythropoietic lineage. *Nature*. 1985;313(5999):235-8.

Pollex RL, Hegele RA. Genetic determinants of plasma lipoproteins. *Nat Clin Pract Cardiovasc. Med*. 2007;4(11):600-9.

Qasim A, Rader DJ. Human genetics of variation in high-density lipoprotein cholesterol. *Curr Atheroscler Rep*. 2006;8(3):198-205.

Rader D, Maugeals C. Genes influencing HDL metabolism: new perspectives and implications for atherosclerosis prevention. *Mol Med Today*. 2000;6(4):170-5.

Rader DJ, Daugherty A. Translating molecular discoveries into new therapies for atherosclerosis. *Nature*. 2008; 451(7181):904-13.

Ravnskov U. Cholestrol lowering trials in coronary heart disease: frequency of citation and outcomes. *BMJ*. 1992;305(6844):15-9.

Rewers M, Shetterly SM, Hoag S, Baxter J, Marshall J, Hamman RF. Is the risk of coronary heart disease lower in Hispanics than in non-Hispanic whites? The San Luis Valley Diabetes Study. *Ethnicity Dis.* 1993;3(1):44-54.

Richmond W. Preparation and properties of a cholesterol oxidase from Nocardia sp. and its application to the enzymatic assay of total cholesterol in serum. *Clin Chem.* 1973;19(12): 1350-6.

Ross R. Atherosclerosis: an inflammatory disease. *N Engl J Med*. 1999;340(2):115-226.

Rubins HB, Robins SJ, Collins D, Iranmanesh A, Wilt TJ, Mann D, Mayo-Smith M, Faas FH, Elam MB, Rutan GH, et al. Distribution of lipids in 8500 men with coronary Artery disease. *Am J Cardiol*. 1995;75(17):1196-201.

Rust S, Walter M, Funke H, von Eckardstein A, Cullen P, Kroes HY, Hordijk R, Geisel J, Kastelein J, Molhuizen HO, Schreiner M, Mischke A, Hahmann HW, Assmann G. Assignment of Tangier disease to chromosome 9q31 by a graphical linkage exclusion strategy. *Nat Genet*. 1998;20(1):96–8.

Sabatti C, Service SK, Hartikainen A, Pouta A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varilo T, Kaakinen M, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet*. 2009;41(1):35–46.

Segrest JP, Jones MK, Mishra VK, Anantharamaiah GM, Garber DW. ApoB-100 has a pentapartite structure composed of three amphipathic alpha-helical domains alternating with two amphipathic beta-strand domains detection by the computer program LOCATE. *Arterioscler Thromb Vasc Biol.* 1994;14(10):1674-85.

Sehayek E, Eisenberg S. Mechanisms of inhibition by apolipoprotein C of apolipoprotein E dependent cellular metabolism of human triglyceride-rich lipoproteins through the low density lipoprotein receptor pathway. *J Rio. Chem.* 1991;266(27):18259-67.

Schonfeld G, George PK, Miller J, Reilly P, Witztum J. Apolipoprotein C–II and C–III levels in hyperlipoproteinemia. *Metabolism.* 1979;28(10):1001-10.

Shachter NS, Zhu Y, Walsh A, Breslow JL, Smith JD. Localization of a liver-specific enhancer in the apolipoprotein E/C-I/C-II gene locus. *J Lipid Res*. 1993;34(10):1699-1707.

Shachter NS, Hayek T, Leff T, Smith JD, Rosenberg DW,Walsh A, Ramakrishnan R,Goldberg IJ,Ginsberg HN, and Breslow JL. Overexpression of apolipoprotein CII causes hypertriglyceridemia in transgenic mice. *J. Clin. Znuat*. 1994;93(4):1683-90.

Sharrett AR, Ballantyne CM, Coady SA, Heiss G, Sorlie PD, Catellier D, Patsch W. Atherosclerosis risk in communities study group. Coronary heart disease prediction from lipoprotein cholesterol levels, triglycerides, lipoprotein(a), apolipoproteins A-I and B, and HDL density subfractions – The Atherosclerosis Risk in Communities (ARIC) Study. *Circulation*. 2001;104(10):1108-13.

Simonet WS, Bucay N, Lauer SJ, Taylor JM. A far-downstream hepatocyte-specific control region directs expression of the linked human apolipoprotein E and C-I genes in transgenic mice. *J Biol Chem*. 1993;268(11):8221-9.

Simonet, WS, Bucay N, Pitas RE, Lauer SJ, Taylor JM. Multiple tissue-specific elements control the apolipoprotein E/C-I gene locus in transgenic mice. *J Biol Chem.*1991;266(14):8651-4.

Spiedel M, Booyse FM, Abrams A, Moore MA, Chung BH. Lipolyzed hypertriglyceridemic serum and triglyceride-rich lipoprotein cause lipid accumulation in and are cytotoxic to cultured human endothelial cells. High density lipoproteins inhibit this cytotoxicity. *Thromb Res*. 1990;58(3):251-64.

Stavropolous WS, Crouch RD. A new colorimetric procedure for the determination of serum triglycerides. *Clin Chem.* 1974;20:957.

Sviridov D, Nestel PJ. Genetic factors affecting HDL levels, structure, metabolism and function. *Curr Opin Lipidol*. 2007;18(2):157-63.

Swaney JB, Weisgraber KH. Effect of apolipoproteinC-I peptides on the apolipoproteinE content and receptor-binding properties of beta-migrating very low density 1ipoproteins. *J Lipid Res.* 1994;35(1):134-42.

Tsompanidi EM, Brinkmeier MS, Fotiadou EH, Giakoumi SM, Kypreos KE. HDL biogenesis and functions: Role of HDL quality and quantity in atherosclerosis. *Atherosclerosis.* 2010;208(1):3-9.

Weisgraber KH, Mahley RW, Kowal RC, Herz J, Goldstein JL, Brown MS. Apolipoprotein C-I modulates the interaction of apolipoprotein E with beta-migrating very low density lipoproteins (beta-VLDL) and inhibits binding of beta-VLDL to low density lipoprotein receptor-related protein. *J Biol Chem.*1990; 265(36):22453-9.

Weissglas-Volkov D, Pajukanta P. Genetic causes of high and low serum HDL-Cholesterol. *J Lipid Res*. 2010;51(8):2032-57.

Willer CJ, Sanna S, Jackson A, Scuteri A, Bonnycastle L, Clarke R, Heath S, Timpson N, Najjar S, Stringham H, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet*. 2008;40(2):161-9.

Zhang LH, Kotite L, Havel RJ. Identification, characterization, cloning and expression of apolipoprotein C-IV, a novel sialoglycoprotein of rabbit plasma lipoproteins. *J Biol Chem*. 1996;271(3):1776-83.