uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Sirman Sukhmani Bhatia

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Biochemistry)

GRADE / DEGREE

Department of Biochemistry, Microbiology and Immunology

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Genetic Association of the APOE/APOC1/APOC4 Locus with Coronary Artery Disease

TITRE DE LA THÈSE / TITLE OF THESIS

Ruth McPherson

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Daniel Sparks

Frédérique Tesson

Gary W. Slater

Le Doyen de la Faculté des etudes supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Genetic Association of the APOE/APOC1/APOC4 Locus with Coronary Artery Disease

By

Simran Sukhmani Bhatia

Thesis Submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biochemistry, Microbiology and Immunology

Faculty of Medicine, University of Ottawa

Ottawa, Ontario, Canada

Your file Votre référence
ISBN: 978-0-494-73792-7
Our file Notre référence
ISBN: 978-0-494-73792-7

ABSTRACT

Genetic Association of the *APOE/APOC1/APOC4*
Locus with Coronary Artery Disease
By Simran Sukhmani Bhatia

Genome-wide association studies identified a 5' *APOC1* single nucleotide polymorphism (SNP), rs4420638 (minor allele frequency (MAF)=0.18) at 19q13.2, with a CAD risk odds ratio (OR) of 1.17 (1.08-1.28) in linkage disequilibrium with apoE4 risk SNP rs429358 (r^2=0.70; OR=1.06 (0.99-1.13), MAF=0.15). Differing OR and MAF led to the hypothesis that rs4420638 risk is partially independent to rs429358. Additional SNPs not in HapMap were genotyped by sequencing, however no strong linkage existed. Genotypically associated traits include serum apoE, as determined by ELISA (rs4420638 AA: 25.5±4.8; AB: 44.8±3.1; BB: 69.9±3.7ug/ml; p=7.27E-07, rs429358 AA: 14.8±3.4; AB: 40.1±3.1; BB: 67.0±3.4ug/ml; p=1.73E-08) and LDL (linear regression rs4420638 p=0.0007; rs429358 p=0.003) but not apoC1. Haplotype analysis indicated that rs429358 risk allele confers a greater CAD risk than rs4420638: rs4420638 risk allele alone is 6% in cases and 4% in controls; rs429358 risk allele is 3% in cases and 0.3% in controls. Likelihood ratio test confirms this conclusion.

ACKNOWLEDGEMENTS

I would first like to thank my supervisor, Dr Ruth McPherson, for the opportunity to complete my Masters of Science in her laboratory. Secondly, I'd like to acknowledge all the individuals and patients that volunteered for this study. Without their donation and support for the Ottawa Heart Institute, this genetics study would not have been possible. I am very grateful to Robert Davies – statistician from the Cardiovascular Research Methods Centre – who not only worked tirelessly with me on the statistics but also had the patience to answer every question I had regarding numbers. I would also like to thank the technicians in the Atherogenomics Laboratory and the John & Jennifer Ruddy Canadian Cardiovascular Genetics Centre for their work on the Affymetrix gene chip procedures. I am grateful to my committee members, Drs Alexandre Stewart and Ross Milne who provided their expertise to this project and helped review this manuscript. Of course, I would like to thank the members of my laboratory, Dr Olga Jarinova, Paulina Lau, Thet Naing, Adrianna Douvris and Adam Turner for making my time in the lab enjoyable.

Last but not least, I would like to give a heartfelt thank you to every friend, coworker and colleague I have had the pleasure of knowing during my Masters. It is because of them that I will look back on my time at the University of Ottawa Heart Institute with great fondness. Not only do I leave here with knowledge and research expertise, but also with genuine, rewarding and life-long friendships.

DEDICATION

This thesis is dedicated to my father, Baljit Singh Bhatia, whose unconditional love, support and knowledge helped me achieve every academic goal I have set out to accomplish. It is with his memory that I will continue to aspire towards higher achievements in all aspects of my life.

TABLE OF CONTENTS

Abstract ... ii
Acknowledgements ... iii
Dedication ... iii
Table of Contents ... iv
List of Figures ... vi
List of Tables ... vii
List of Abbreviations ... viii
1. Introduction ... 1
1.1 - Lipid Metabolism ... 1
1.1.1 Exogenous Pathway ... 3
1.1.2 Endogenous Pathway ... 6
1.1.3 Reverse Cholesterol Transport ... 7
1.1.4 Atherosclerosis ... 11
1.1.5 Role of ApoE in Lipid Metabolism ... 13
1.1.6 Role of ApoC1 and ApoC4 in Lipid Metabolism ... 14
1.2 - Genetics of Coronary Artery Disease ... 16
1.2.1 Genome-Wide Association Studies ... 18
1.3 - Effects of Genetic Variation in APOE on Apolipoprotein E Structure and Function ... 22
1.3.1 ApoE Protein Structure ... 24
1.3.2 Three Isoforms of ApoE ... 24
1.3.3 ApoE and CAD ... 27
1.3.4 ApoE in Alzheimer's Disease ... 27
1.4 - Hypothesis and Objectives ... 31
2. Materials and Methods ... 34
2.1 - Subjects ... 34
2.2 - Genotyping by Affymetrix® Genome-Wide Human SNP Array ... 34
2.3 - Genotyping by Sequencing ... 37
2.4 - PCR and Sequencing Reaction Set-up ... 40
2.5 - ApoE and ApoC1 ELISAs ... 42
2.6 - Statistical Analysis ... 43
3. Results ... 46
3.1 – Minor allele frequencies and odds ratios of genotyped APOE/C1/C4 SNPs in the Ottawa Heart Population ... 46
3.2 – Linkage Disequilibrium analysis for rs429358 and rs4420638 ... 51
3.3 – Haplotype analysis of ApoE SNPs and rs4420638 and Likelihood Ratio Test ... 53
3.4 – Only apoE, not apoC1, serum protein concentration is genotypically associated ... 56
3.5 – Plasma LDL concentration is associated with rs4420638 and apoE SNPs ... 61
3.6 – Odds ratios of the six apoE genotypes in the Ottawa Heart population ... 64

4. Discussion **67**

4.1 - MAF of risk SNPs in the OHS differ from other study populations 73
4.2 - ORs of rs429358, rs7412 and rs4420638 are more extreme than other study populations 76
4.3 - Minor Allele Frequencies of rs4420638 and rs429358 differ 79
4.4 - ApoE, but not apoC1 serum concentrations are associated with rs429358, rs4420638 and rs7412 genotypes 80
4.5 - LDL blood lipid levels are genotypically associated 83

5. Conclusion **86**

6. References **88**

Appendix 1: Detailed Competition ELISA Protocol Used to Determine Serum ApoE and ApoC1 Concentrations **96**

Appendix 2: Sample ApoE and ApoC1 Serum Concentration Calculations Example from Plate 14 ..... **100**

Appendix 3: Curriculum Vitae **105**

LIST OF FIGURES

Figure 1.1: Causes of Death of White Males and Females in USA in 2005 2

Figure 1.2: Schematic diagram representing the four main pathways involved in human lipid homeostasis 4

Figure 1.3: Schematic diagram of the uptake of exogenous lipids and the role of apoE in this process. 5

Figure 1.4: Reverse cholesterol transport. 9

Figure 1.5: Summary of the movement of lipoproteins in the endogenous, exogenous and reverse cholesterol transport pathways 10

Figure 1.6: Progression of atherosclerosis in coronary arteries 12

Figure 1.7: Schematic diagram of region 19q13.2 with all known polymorphisms 23

Figure 1.8: Ribbon model of unlipidated apolipoprotein E 25

Figure 1.9: Crystal structure of unlipidated apoE4 versus apoE3 26

Figure 1.10: Ribbon structure of apoE3 versus apoE2. 28

Figure 2.1: Schematic overview of genotyping by Affymetrix® Genome-wide Human SNP Array. 36

Figure 2.2: Schematic diagram of PCR amplicons used to sequence each SNP. 38

Figure 2.3: Example sequence result of amplicons after analysis by DNA Sequencing Analysis Software v5.2 (ABI). 41

Figure 2.4: Examples of standard curves obtained from Plate 14 44

Figure 3.1: Schematic map of APOE-APOC1/C4 region 19q13.2 identifying relative positions of all SNPs genotyped. 50

Figure 3.2: Linkage Disequilibrium Plot calculated from control samples (n=2145) of all SNPs genotyped using Haploview 4.2 software. 54

Figure 3.3: Bar graphs of concentration of fasting (a) LDL-cholesterol, (b) HDL-cholesterol, (c) triglycerides (TG), and (d) total cholesterol (TC) relative to the six apoE genotypes from control samples. 65

Figure 3.4: Odds ratios of six apoE genotypes (E2/E2, E2/E3, E2/E4, E3/E4 and E4/E4). 66

LIST OF TABLES

Table 1.1: Major loci for coronary disease identified and replicated in GWA studies (genome-wide significance p<5E-8). 21

Table 1.2: ApoE isoforms: structure and function (Hatters, Peters-Libeu, and Weisgraber 2006). 29

Table 1.3: Potential SNPs associated with lipid traits in the APOE/C1/C4 region. 33

Table 2.1: Table of physical characteristics of Cases and Controls. 35

Table 2.2: Four PCR amplicons were designed to include all seven SNPs for sequencing. 39

Table 3.1: Physical characteristics of CAD cases and healthy controls from the Ottawa Heart Study. Mean ± SD. 47

Table 3.2: Genotype frequencies of individuals genotyped in the Ottawa Heart Study cases and controls. 48

Table 3.3: Linkage disequilibrium r2 and D' values of each SNP genotyped relative to rs4420638 and rs429358 ε4 52

Table 3.4: Haplotype analysis using estimated 2-loci haplotype frequency prediction analysis. 55

Table 3.5: Likelihood Ratio Test of rs4420638 and rs429358 CAD risk models. 57

Table 3.6: Serum Apoliporotein E and C1 protein concentration in control samples as determined by ELISA method 59

Table 3.7: Association of serum Apoliporotein E and C1 concentration in control samples as determined by ELISA, with APOE genotype. 60

Table 3.8: Plasma concentration of fasting LDL and triglycerides (TG) in controls 62

Table 3.9: Concentration of fasting (a) LDL, (b) HDL, (c) triglycerides (TG), and (d) total cholesterol (TC) relative to the six apoE genotypes using control samples 63

Table 4.1: Comparison of the minor allele frequencies of rs4420638, rs429358 and rs7412 in the OHS with past studies 77

LIST OF ABBREVIATIONS

A	Adenine
Aβ	amyloid-beta
ABCA1	ATP-binding cassette transporter A1
ABI	Applied Biosystems
AD	Alzheimer's disease
Apo	Apolipoprotein
APP	amyloid precursor protein
Arg	Arginine
Asp	Aspartate
BACE	beta-site APP cleaving enzyme
BMI	Body mass index
BSA	bovine serum albumin
C	Cytosine
CAD	Coronary artery disease
CE	cholesteryl esters
CETP	Cholesterol ester transfer protein
CM	Chylomicrons
CNS	Central nervous system
CR	Chylomicron remnants
CSF	cerebrospinal fluid
Cys	Cysteine
DGI	Diabetes Genetics Inititative
DNA	Deoxyribonucleic assay
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
FC	free cholesterol
FUSION	Finland-United States Investigation of NIDDM Genetics
G	Guanine
Glu	Glutamate
GWAS	Genome-wide association study
HCR	Hepatic control region
HDL	High density lipoproteins
HL	Hepatic lipase
IDL	intermediate density lipoproteins
LCAT	lecithin-cholesterol acyltransferase
LD	Linkage disequilibrium
LDL	low-density lipoprotein
LDLr	low-density lipoprotein receptor
LPL	Lipoprotein lipase

LRP	low-density lipoprotein receptor-like protein
MAF	Minor allele frequency
ME	Multienhancer
MI	Myocardial infarction
nHDL	Nascent high density lipoprotein
NIA	National Institute on Aging
OHS	Ottawa Heart Study
OR	Odds ratio
oxLDL	oxidized low-density lipoprotein
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
RCT	Reverse cholesterol transport
RCT	Reverse cholesterol transport
RNA	Ribonucleic acid
RR	Ready reaction
SCAP	SREBP cleavage-activating protein
SNP	Single nucleotide polymorphism
SRB1	scavenger receptor B1
SRE	sterol regulatory elements
SREBP	sterol regulatory element binding protein
T	Thymine
TC	Total cholesterol
TG	Triglycerides
VLDL	very-low density lipoproteins

1. INTRODUCTION

One in two men and one in three women in North America will develop coronary artery disease (CAD) in their lifetime [1]. Although CAD most commonly manifests after the age of 60 in men and 65 in women, the development of arterial plaques begins at an early age [2,3]. CAD can lead to angina, heart attack, myocardial infarction (MI), and stroke. As a result, CAD is the leading cause of death in North America (**Figure 1.1**) [4]. Early identification of subjects at risk would allow for early treatment and positive life-style changes to deter the detrimental outcomes.

Major conventional risk factors of CAD include smoking, diabetes, hypertension and dyslipidemia. However, 10-20% of CAD patients have none of these (Khot et al. 2003). Many studies indicate that family history is also a major CAD risk factor [6-8]. Genetic factors, acting through conventional and unknown risk factors, are believed to account for half of CAD risk [9]. This thesis investigates the effect of single nucleotide polymorphisms (SNPs) at the *APOE/C1/C4* locus on plasma lipids and CAD risk.

1.1 - Lipid Metabolism

Like most all processes in the body, balancing lipid catabolism and anabolism requires multiple integrated processes. For lipid homeostasis the main pathways include, the exogenous (dietary) pathway, the endogenous pathway, reverse

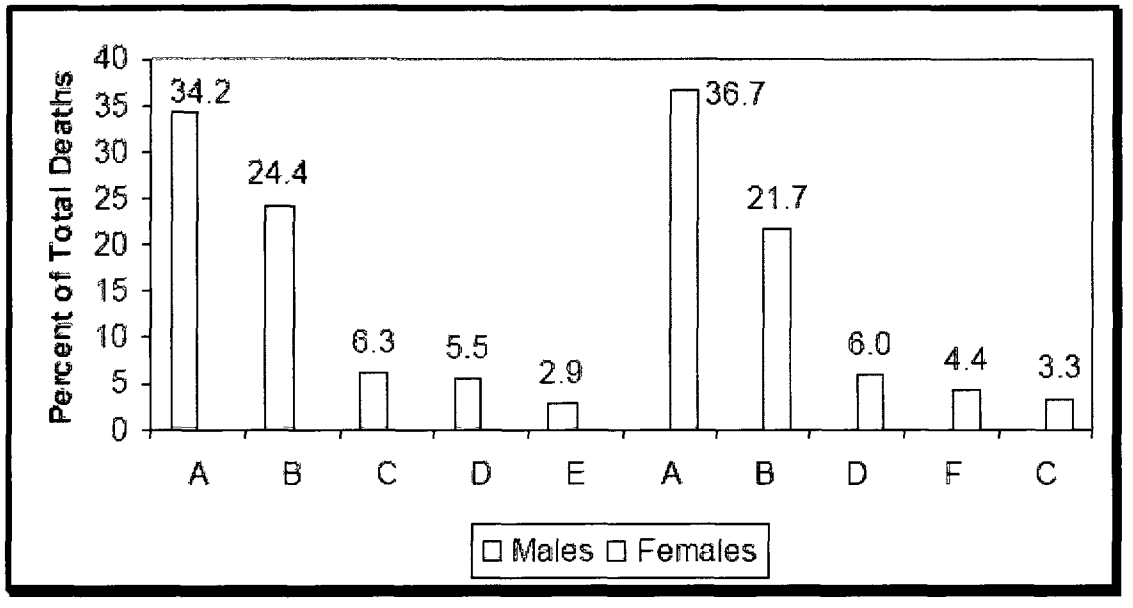

Percent of Total Deaths
40
35
30
25
20
15
10
5
0
34.2
24.4
6.3
5.5
2.9
36.7
21.7
6.0
4.4
3.3
A
B
C
D
E
A
B
D
F
C
Males
Females

Figure 1.1: Causes of Death of White Males and Females in USA in 2005.

Over one-third of all deaths in the United States are cardiovascular related. A: Total cardiovascular disease; B: cancer; C: accidents; D: chronic lower respiratory disease; E; diabetes mellitus; F: Alzheimer's disease. [American Heart Association 2009 Statistical Fact Sheet] [4]

cholesterol transport and excretion in the form of bile (**Figure 1.2**). All of these pathways lead to and from the liver – the main organ which maintains lipid homeostasis. A detailed diagram of the main processes involving lipid homeostasis can be seen in **Figure 1.5**.

1.1.1 Exogenous Pathway

In order to transport insoluble lipids from the intestine to the liver, triglycerides (TG) and cholesterol are packaged in chylomicrons (CM) in the epithelial cells of intestines. Chylomicrons are low-density lipid and cholesterol droplets associated with apolipoproteins – namely apoAI, apoB48, apoCII and apoE [10(pp491, 537)]. In this form, lipids can be transported to muscle tissues that require TGs for energy or hepatocytes and adipocytes for storage. At the vascular endothelium of adipose and muscle and other tissues, TGs are broken down to free fatty acids and monoglycerol by anchored lipoprotein lipase (LPL) [11(p541)]. Free fatty acids are available for oxidation, storage as cellular triglyceride or returned to the liver. This process continues until more than 80% of the original TG content is removed from CMs [11(p543)]. However, nearly all of the original cholesterol content is retained in the chylomicron remnants (CRs). The remnants are taken up by hepatocytes via receptor-mediated endocytosis. Lipoprotein-bound apoE serves as the main ligand for the LDL receptor (LDLr) and LDLr-like protein (LRP) present on hepatocytes (Kypreos et al. 2001). **Figure 1.3** shows the steps where apoE plays a role in the uptake of exogenous lipids.

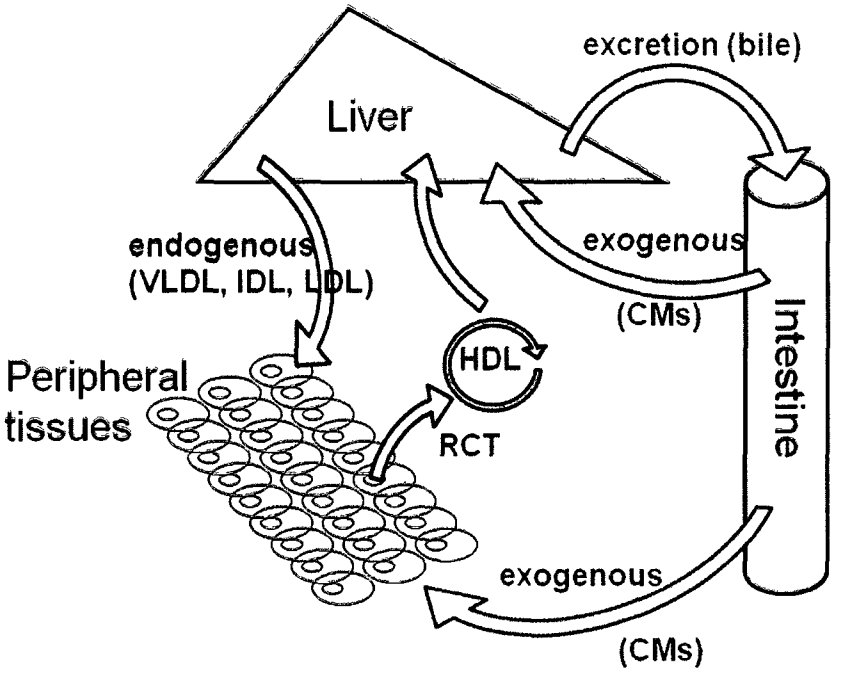

excretion (bile)
Liver
endogenous
(VLDL, IDL, LDL)
exogenous
(CMs)
Intestine
HDL
Peripheral
tissues
RCT
exogenous
(CMs)

Figure 1.2: Schematic diagram representing the four main pathways involved in human lipid homeostasis.

1) endogenous pathway (red) involving lipids excreted from the liver in the form of very-low density lipoproteins (VLDL), intermediate density lipoproteins (IDL), low-density lipoprotein (LDL) to peripheral tissues; 2) the exogenous pathway (green) linking lipids excreted from the intestine in the form of chylomicrons (CMs) to peripheral tissues and the liver; 3) reverse cholesterol transport (RCT) (purple) which transports excess cholesterol from peripheral tissues via high density lipoproteins (HDL) to the liver for storage or excretion; 4) excretion (blue) from the liver back to the intestine in the form of bile. In the intestine the excess bile and cholesterol can be excreted from the body or reabsorbed into the bloodstream via the exogenous pathway.

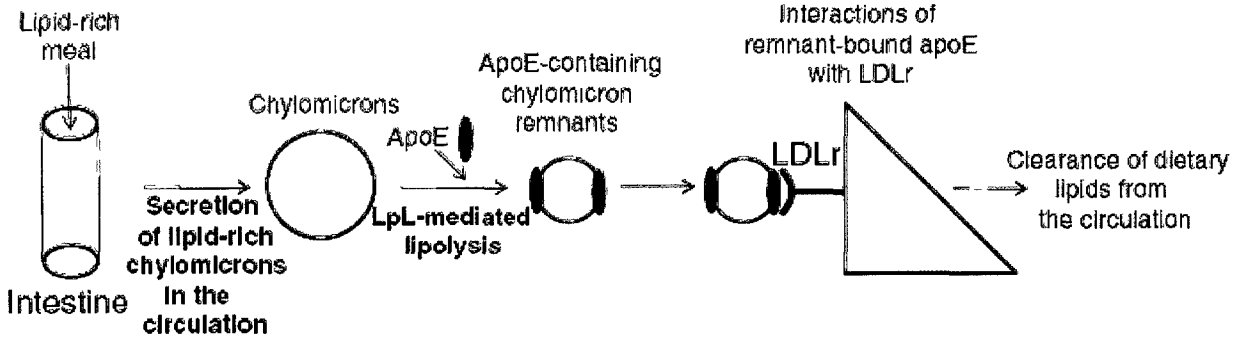

Lipid-rich meal
Intestine
Secretion of lipid-rich chylomicrons in the circulation
Chylomicrons
ApoE
LpL-mediated lipolysis
ApoE-containing chylomicron remnants
Interactions of remnant-bound apoE with LDLr
LDLr
Clearance of dietary lipids from the circulation

Figure 1.3: Schematic diagram of the uptake of exogenous lipids and the role of apoE in this process.

Lipids are packaged from the intestine in the form of chylomicrons and secreted into the bloodstream. As lipids are broken down by LPL on peripheral tissues, chylomicron remnants acquire apoE from the bloodstream. Lipid-bound apoE interacts with members of the LDL receptor family, including the LDLr, VLDLr and LRP, on hepatocytes or peripheral tissue cells for their endocytic removal from the bloodstream. [13]

Figure taken from Kypreos et al., (2009) Mechanisms of obesity and related pathologies: Role of apolipoprotein E in the development of obesity. *FEBS J* 276(20):5720-8, with permission. License agreement number 2416800726245

1.1.2 Endogenous Pathway

The liver synthesizes and secretes very-low-density lipoproteins (VLDL). These lipoproteins are similar in composition to CMs; however, they lack apoAI and instead of apoB48, VLDL contains full-length apoB100. ApoB100 serves as a ligand for the LDL receptor (LDLr), unlike apoB48. Similarly to CMs, TGs present in VLDL undergo lipolysis by LPL to deliver fatty acids to peripheral tissues. Remnants of VLDL – known as intermediate-density lipoproteins (IDL) – can be taken up by hepatocytes via the LDLr, VLDLr and LRP. A majority of IDL molecules, however, undergo further modification to form low-density lipoproteins (LDL) [11(p543)]. The conversion of IDLs to LDLs includes the complete removal of TGs, any remaining apoCII and apoE. The result is LDL, which consists of cholesterol, cholesteryl esters (CE) and apoB100.

Like IDL, LDL molecules can also be taken up by hepatocytes via the LDLr. However, 30-40% of the body's LDLrs are present on extrahepatic tissues [14(p558)]. The main purpose of LDL is to transport cholesterol to these tissues. The uptake of LDL is regulated by cholesterol homeostasis; as a cell requires more cholesterol, LDLr expression is upregulated resulting in increased endocytosis of LDL molecules. The method by which LDLr expression is upregulated is controlled by sterol regulatory elements (SRE). These are 10-base pair non-palindromic sequences that, when bound by SRE-binding protein (SREBP), initiate gene transcription in cholesterol depleted cells. Low levels of cholesterol trigger the eventual release of SREBP.

Inactive pre-SREBP is a dormant transmembrane protein in the endoplasmic reticulum (ER). Also present in the ER membrane is SREBP cleavage-activating protein (SCAP). The regulatory domain of pre-SREBP interacts with SCAP and, when cholesterol levels are low, SCAP ushers pre-SREBP to the Golgi. It is in the Golgi where pre-SREBP is cleaved to release the mature soluble transcription factor. It can then translocate to the nucleus where it binds SRE sequences on promoters of genes coding for proteins involved in lipid homeostasis – including the *LDLR* to enhance their transcription. [15(pp409-10)]

However, when cholesterol levels are optimal or high in the cell, cholesterol binds to SCAP causing it to conformationally change. SCAP can then bind to Insig, another transmembrane protein that remains anchored in the ER. As a result, the SCAP-SREBP complex also remains anchored in the ER preventing the upregulation of cholesterol synthesis proteins like the LDLr [15(p411)]

1.1.3 Reverse Cholesterol Transport

Reverse cholesterol transport (RCT) is the transport of excess cellular cholesterol from peripheral tissues back to the liver by high-density lipoproteins (HDL). Most peripheral cells are unable to catabolize cholesterol and maintenance of optimal cholesterol concentrations is vital for proper cellular function.

Cholesterol can be removed from peripheral cells by efflux to lipid-poor extracellular nascent HDL (nHDL). The ATP-binding cassette transporter A1 (ABCA1) facilitates efflux of cholesterol from sub-endothelial macrophages in arterial plaques to

nHDL. The simplest method of efflux is via gradient diffusion where cholesterol moves from a high concentration in cells to extracellular cholesterol acceptors, namely nHDL. SRB1 and ABCG1 also mediate cholesterol efflux to lipidated HDL particles rather than lipid poor apoA1. [16]

Cholesterol present on nHDL is converted to CE by lecithin-cholesterol acyltransferase (LCAT). HDL particles grow in size as more CE is converted resulting in particles of differing sizes, namely HDL2 and HDL3. Those enriched with triglycerides are processed to denser and smaller particles by hepatic lipase (HL) which is similar to LPL but resides on hepatocytes and adrenal glands. Cholesterol ester transfer protein (CETP) mediates the transfer of HDL-CE and VLDL and IDL-TG. Since VLDL and IDL are precursors to LDL, CETP indirectly mediates the homeostasis between HDL and LDL. Mature HDL delivers cholesterol and CE to hepatocytes via the scavenger receptor B1 (SR-B1), which facilitates the uptake of HDL and subsequent CE into hepatocytes [17], or by transferring CE to apoB lipoproteins, namely VLDL or LDL, both of which are taken up by the liver through LDLr-mediated endocytosis. In the liver, cholesterol may be esterified and stored in lipid droplets, repackaged into lipoproteins or converted to bile acids and secreted into the intestine in the form of bile. The main steps of reverse cholesterol transport are summarized in **Figure 1.4.** [18]

The most important lipoprotein in RCT is HDL. The unique feature of HDL is, unlike most other lipoproteins, which are secreted into plasma in lipid-rich form, pre-HDL secreted by the liver is usually unlipidated apoA1 [11(p544)]. The primary

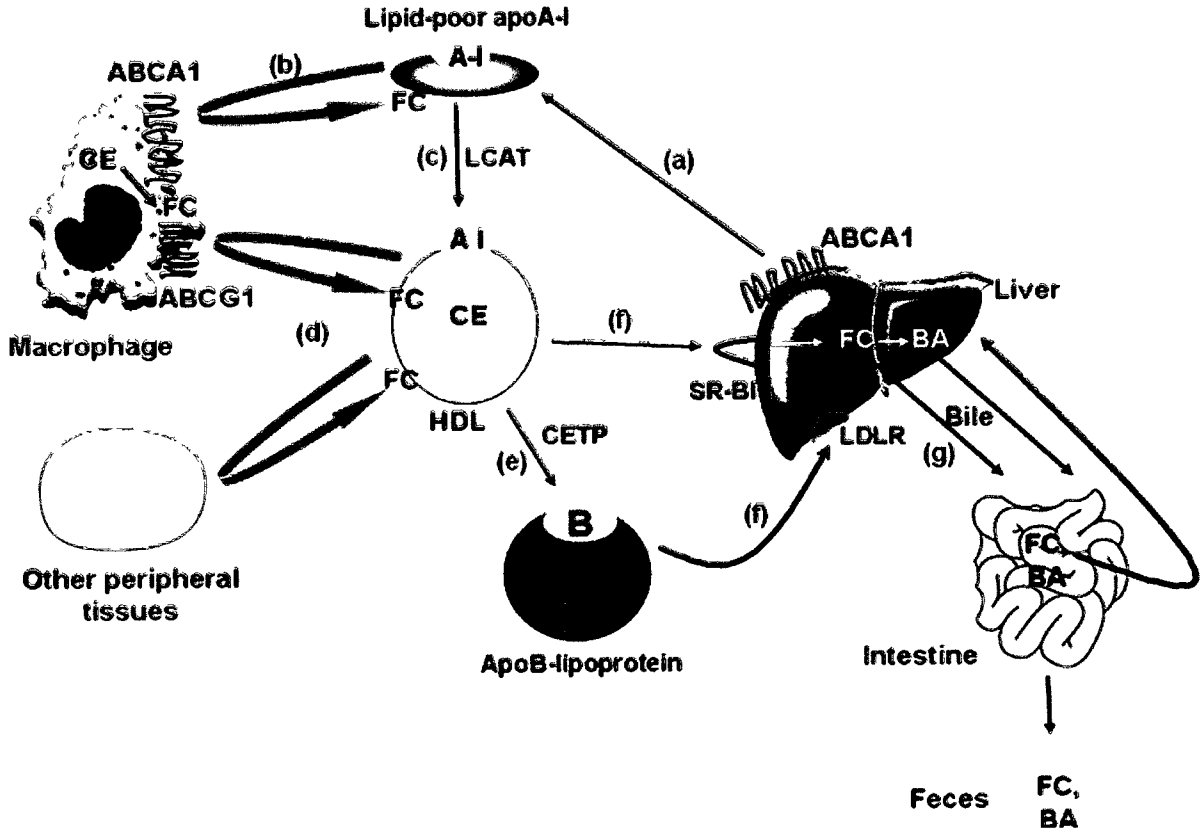

Lipid-poor apoA-I
ABCA1
(b)
A-I
FC
CE
FC
(c) LCAT
(a)
A I
FC
CE
ABCG1
Macrophage
(d)
FC
HDL
(f)
ABCA1
Liver
FC → BA
SR-BI
LDLR
Bile
(g)
CETP
(e)
B
(f)
FC
BA
Other peripheral tissues
ApoB-lipoprotein
Intestine
Feces
FC,
BA

Figure 1.4: Reverse cholesterol transport.

(a) nascent HDL, or lipid-poor apoA1, is excreted by the liver and (b) acquires free cholesterol (FC) from lipid-laden macrophages via the ubiquitous ATP-binding cassette transporter A1 (ABCA1). Cholesterol esters (CE) in macrophages are converted to FC to be taken up by extracellular nascent HDL. (c) FC is converted to CE in HDL by lecithin-cholesterol acyltransferase (LCAT) to give rise to larger mature HDL particles. (d) Mature HDL particles continue to take up FC from peripheral tissues and is mediated by ABCG1. (e) Cholesteryl ester transfer protein (CETP) mediates the exchange of HDL-CE and apoB-containing-lipoprotein-triglycerides (TG) (f) Accumulated cholesterol in HDL and apoB-containing-lipoproteins are taken up by the liver via the LDL receptor (LDLr) and scavenger receptor protein B1 (SRB1). (g) In the liver, CE and FC are broken down to bile acids and excreted to the intestine as bile. From the intestine, residual FC can be re-absorbed into the bloodstream or excreted in the feces [18]

Figure taken from Rader et al., (2009) The role of reverse cholesterol transport in animals and humans and relationship to atherosclerosis. *J. Lipid Res* 50suppl: S189-194, with permission.

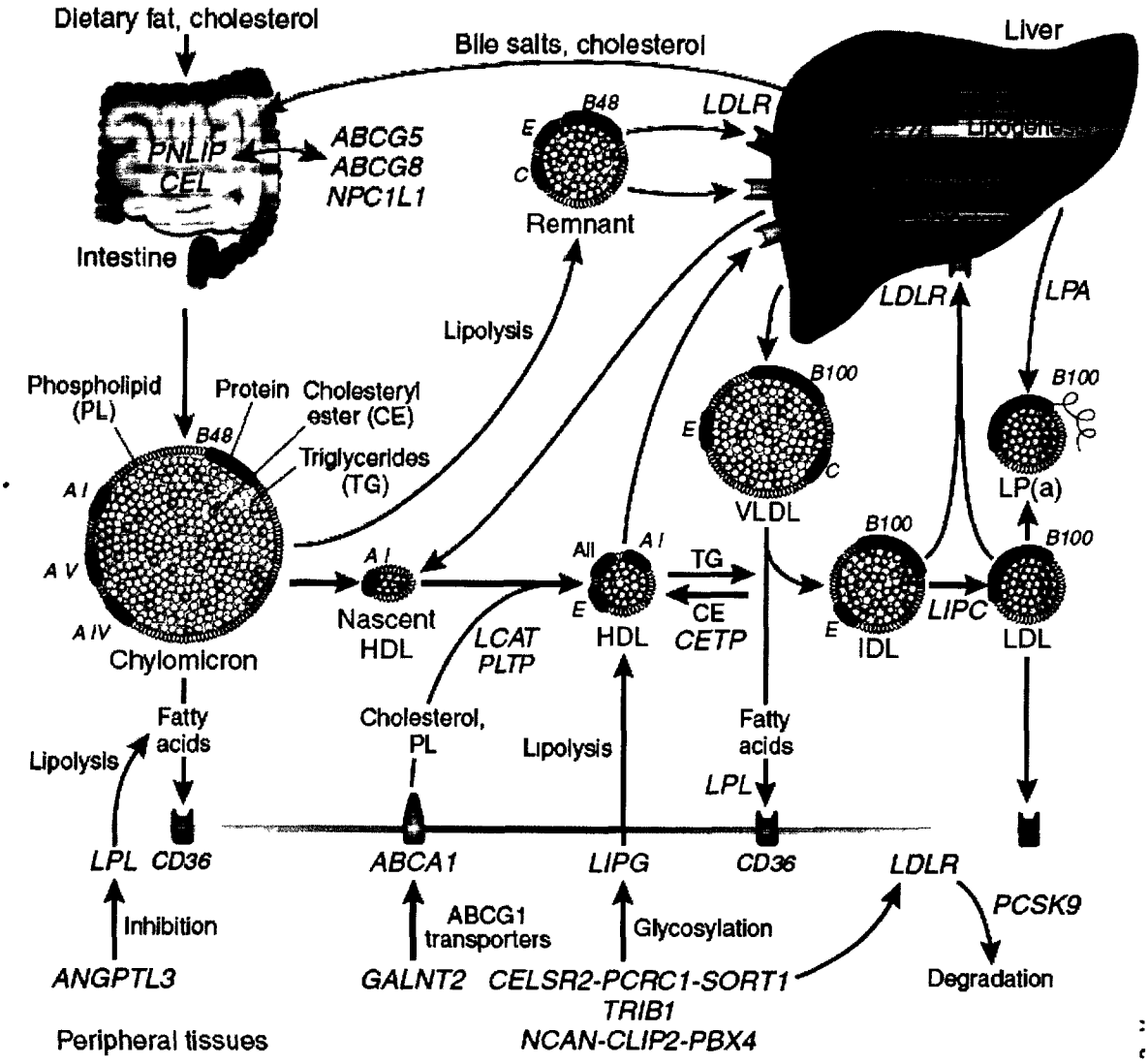

Dietary fat, cholesterol
Bile salts, cholesterol
Liver
PNLIP
CEL
Intestine
ABCG5
ABCG8
NPC1L1
B48
E
C
Remnant
LDLR
Lipoprotein
Lipolysis
LDLR
LPA
Phospholipid (PL)
Protein
Cholesteryl ester (CE)
Triglycerides (TG)
B48
A I
A V
A IV
Chylomicron
B100
E
C
VLDL
B100
LP(a)
A I
Nascent HDL
All
A I
E
HDL
TG
CE
CETP
B100
E
IDL
LIPC
B100
LDL
LCAT
PLTP
Fatty acids
Lipolysis
Cholesterol, PL
Lipolysis
Fatty acids
LPL
LPL
CD36
ABCA1
LIPG
CD36
LDLR
PCSK9
Inhibition
ABCG1 transporters
Glycosylation
Degradation
ANGPTL3
GALNT2
CELSR2-PCRC1-SORT1
TRIB1
NCAN-CLIP2-PBX4
Peripheral tissues

Figure 1.5: Summary of the movement of lipoproteins in the endogenous, exogenous and reverse cholesterol transport pathways.

The exogenous pathway begins in the intestine where lipids are secreted to the bloodstream packaged as chylomicrons (CMs). CMs transport triglycerides (TG) to peripheral tissues where they are lipolyzed to free fatty acids by lipoprotein lipase (LPL). CM remnants (CRs) acquire apoE from the bloodstream which interacts with the LDL receptor (LDLr) and LDL receptor-like protein (LRP) triggering receptor-mediated endocytosis. The endogenous pathway begins with the secretion of very-low-density lipoprotein (VLDL) by hepatocytes. VLDL delivers TGs to peripheral tissues similarly to CMs. As triglycerides are lipolyzed by LPL, VLDLs convert to intermediate-density lipoproteins (IDL) and low-density lipoproteins (LDL), both of which can be endocytosed by the LDLr on hepatocytes. LDL primarily delivers cholesterol to peripheral tissues. Reverse cholesterol transport begins with the excretion of nascent high-density lipoprotein (nHDL) from the liver. Excess cholesterol from peripheral tissues is shuttled to nHDL via ATP-binding cassette transporter A1 (ABCA1). Cholesterol in nHDL is converted to cholesterol esters (CE) by lecithin-cholesterol acyltransferase (LCAT) giving rise to mature HDL. HDL can exchange CE for VLDL-TG and IDL-TG via the cholesterol ester transfer protein (CETP) which mediates the balance between HDL and LDL. Mature HDL delivers cholesterol to hepatocytes by receptor mediated-endocytosis. Excess cholesterol in the liver is excreted back to the intestine in the form of bile. Genes shown in blue are those previously identified in genome-wide association studies (GWAS) to be associated with lipoprotein concentration. Genes in red are loci recently identified within the last two years of GWA studies [19]

Figure taken from Lusis et al., (2008) A treasure trove for lipoprotein biology. *Nat. Genet* 40(2): 129-130, with permission. License agreement number 2416810867760

purpose of HDL is to accumulate cholesterol and lipids from peripheral tissues [11(p544)]. Mature HDL is the smallest, most dense lipoprotein and consists of free cholesterol, CEs, some TGs, apoA1 and apoE.

1.1.4 Atherosclerosis

Atherosclerosis is a complex disease, which involves multiple environmental and genetic factors. It is characterized by the accumulation of lipids in the subendothelial space (in the case of CAD, coronary arteries) leading initially to the development of fatty streaks (**Figure 1.6**). The source of these lipids is plasma-derived LDL. Plasma concentrations are thus the major lipid risk factor for the development of this disease. It is the interaction of apoB and endothelial cell proteoglycans that retain LDL in the intimal layer. Consequently, other apoB containing lipoproteins such as VLDL and IDL can also add to the fatty streak. [20,21]

After accumulation, LDL becomes modified by oxidation resulting in pro-inflammatory oxLDL. Modified LDL triggers the recruitment of monocytes and lymphocytes by prompting endothelial cells to produce pro-inflammatory factors. Macrophages take up modified LDL leading to the formation of foam cells. Foam cells eventually rupture leaving their fatty content to add to the accumulating lipid-rich lesion. Fatty streaks eventually become fibrous plaques with the accumulation of smooth muscle cells. Plaques with thin fibrous caps are vulnerable to rupture and can

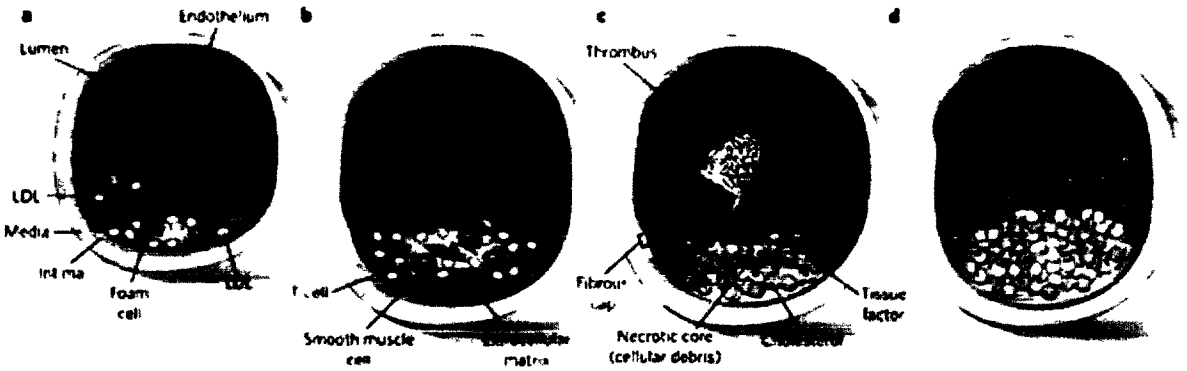

a
Endothelium
Lumen
LDL
Media
Intima
Foam cell
LDL
b
T cell
Smooth muscle cell
c
Thrombus
Fibrous cap
Tissue factor
Necrotic core (cellular debris)
d

Figure 1.6: Progression of atherosclerosis in coronary arteries.

(a) Formation of a fatty streak marks the initial stages leading to lethal atherogenic plaques; LDL particles from the bloodstream enter the sub-endothelial intima of coronary arteries stimulating the recruitment of macrophages. Unregulated endocytosis of lipoproteins by macrophages leads to the formation of foam cells (lipid-laden macrophages). Accumulation of foam cells forms fatty streaks. (b) Fatty streaks induce a chronic inflammatory state and recruit pro-inflammatory factors such as T-cells. Vascular smooth muscle cells excrete cellular-matrix components such as collagen leading to the formation of a fibrous cap (c). As foam cells rupture, cellular debris accumulate in the plaque's necrotic core. The fibrous cap of the lesion can be vulnerable to rupture causing a thrombus to form in the lumen. If the thrombus is large enough to stop blood flow, heart attack or myocardial infarction (MI) can result. (d) Alternatively, the plaque can continue to grow restricting the lumen size, thus, restricting blood flow and resulting in angina [22]

Figure taken from Rader and Daugherty, (2008) Translating molecular discoveries into new therapies for atherosclerosis. *Nature* 451(7181): 904-913, with permission. License agreement number 2416811082510

lead to thrombosis. It is critical if the thrombus blocks blood flow resulting in a myocardial infarction (MI). [23]

1.1.5 Role of ApoE in Lipid Metabolism

ApoE is involved at many levels of lipid homeostasis. Most apoE is synthesized in the liver, but adipose tissue, macrophages, brain and arterial wall cells also synthesize apoE to a lesser degree. A majority of this apolipoprotein is found in the plasma bound in lipoproteins, namely CR, VLDL and HDL. In the lipid-bound form, apoE is a high affinity ligand for the LDLr and LRP (Kypreos et al. 2001). When lipid-bound apoE binds to these receptors, the clearance of lipoproteins by hepatic endocytosis is triggered. This is true for VLDL and CR. For HDL, apoE functions as a receptor ligand in reverse cholesterol transport. (It is important to note that three apoE variants exist, E2, E3 and E4, each of which has different receptor and lipid binding affinities. This results in different physiological functions. See Section 1.3.2).

ApoE function is concentration dependant. Physiological levels of apoE function to promote VLDL and CR clearance, but increased apoE expression limits VLDL lipolysis. The high concentration of apoE competes out VLDL-bound apoCII, a cofactor for LPL. In this case, hypertriglyceridemia can result [24]. ApoE expression levels increase in the fasting state. High levels of apoE also promotes hepatocyte VLDL secretion and adipogenesis. In obesity, apoE levels decrease resulting in increased receptor-mediated endocytosis of lipoproteins by hepatocytes. Adipogenesis and adipocyte lipid uptake is impaired. [25,26]

ApoE is considered atheroprotective since, in HDL homeostasis, apoE facilitates cholesterol efflux. This is quite evident with macrophage-secreted apoE since it promotes cholesterol efflux from lipid-laden macrophages. In large HDL molecules, apoAI, the usual activator of LCAT, is not as efficient and, thus, apoE acts as the primary activator of LCAT. Although rare, humans with non-functional apoE have shown to have elevated plasma cholesterol and a disease known as familial apoE deficiency [27]. The apoE deficient mouse model ($apoE^{-/-}$) is highly susceptible to atherosclerosis since hepatocyte lipoprotein clearance is markedly impaired. A high fat diet markedly increases the development of atherosclerosis in $apoE^{-/-}$ mice. [28,29]

1.1.6 Role of ApoC1 and ApoC4 in Lipid Metabolism

Like apolipoprotein E, apoC1 and apoC4 are soluble plasma apolipoproteins primarily synthesized by the liver. Found in circulation, they are bound to CM, VLDL and HDL.

It has been shown *in vivo* that overexpression of apolipoprotein C1 causes hypercholesterolemia and even more pronounced hypertriglyceridemia [30,31]. This is primarily confined to VLDL lipids. ApoC1 inhibits the conversion of VLDL particles to LDL by interfering with lipoprotein lipase activity [31]. Additionally, VLDL clearance is hindered by apoC1, which inhibits VLDL binding to the LDLr and LRP. Thus, overexpression of apoC1 impairs VLDL metabolism and results in an increase in total cholesterol and triacylglycerides in VLDL particles. [32,33]

Not only does apoC1 affect VLDL metabolism, but recent studies have also shown apoC1 to have an effect on HDL. Overexpression of apoC1 is shown to stimulate lecithin: cholesterol acyltransferase (LCAT), which esterifies HDL-C [34]. It can also act as an inhibitor of both hepatic lipase (HL), which lipolyzes HDL triacylglycerides, and cholesteryl ester transfer protein (CETP), which exchanges HDL-CE for VLDL-TG . Consequently, HDL particle size and number increase. Clearance by hepatic endocytosis of these larger HDL particles may also be hindered by overexpressed apoC1, which was recently shown to inhibit HDL binding to SR-B1 on hepatocytes. [32]

An important variant has been identified in the promoter of *APOC1*; the *HpaI* polymorphism. This promoter variant is a CGTT insertion located 317 base pairs 5' of the *APOC1* transcription start site [35,36]. *HpaI* is associated with a 50% increase in apoC1 transcription *in vitro*, although since it is in strong linkage disequilibrium with the apoE SNPs, its independent physiological activity has been difficult to determine[36]. It has also been associated with Alzheimer's disease (AD) but this relationship may be due to *APOE* linkage[37].

ApoC4 is a relatively newly characterized apolipoprotein [38]. It is expressed in the liver and is not detected in human plasma. Although its function has not been fully elucidated, it is likely that apoC4 is involved in lipid homeostasis due to its structural similarity to other lipoproteins. Also, overexpression of human apoC4 results in increased plasma TG in transgenic mice which is associated with the development of atherosclerosis [39].

1.2 - Genetics of Coronary Artery Disease

Despite recent advances in understanding the pathophysiology of atherosclerosis and identifying its major risk factors of CAD, a major portion of heritable risk for this disease remains unexplained. A combination of environmental, genetic and epigenetic factors plays both additive and interactive roles in disease development. Thus, trends in CAD research have begun to focus more on the genetic contribution of the disease [40]. The heritability of CAD is estimated at 50% but this figure is much higher for early onset CAD as demonstrated in twin, adoption and family studies [7-9].

Monogenic disorders are studied using a traditional Mendelian approach. Although there are numerous monogenic diseases, a number of which result in premature CAD (e.g. familial hypercholesterolemia [41], progeria [42], homocysteinemia [43], etc) coronary artery disease itself is a common, polygenic disease. The fact that many, unrelated monogenic disorders can manifest atherosclerosis further emphasizes how complex and integrated CAD physiology really is. The genetic complexity of CAD is due to multiple factors: genes that contribute to the development of CAD itself, genes that contribute to the development of risk factors, the effect of lifestyle on these genes and their expression (diet, exercise, smoking, drugs), and the genetic interaction therein. The effect of each factor is minor, although additive. Thus, the identification of contributing factors poses some difficulty.

Genetic analyses must be carried out in large cohorts in order to identify common variants of small effect size. At the same time, CAD is a heterogeneous

disorder and phenotype, including age of onset, thus, the extent of coronary atherosclerosis or acute myocardial infarction need to be well defined. The common single nucleotide polymorphisms (SNPs) are the main focus of study for complex diseases such as CAD.

Over 23 million human SNPs have been assigned reference SNP (rs) numbers – 15 million of which have been validated with allele frequencies and additional genetic information in the National Centre for Biotechnology Information (NCBI) Entrez SNP database (dbSNP) (NCBI, National Library of Medicine). This is a considerable increase from the original 1.42 million SNPs first characterized by the International SNP Map Working Group in 2001 [44].

The International HapMap Project, which began in 2002, developed a database of human SNP variation to provide a tool to determine genetic factors associated with disease (International HapMap Consortium, 2003). Instead of sequencing entire genomes, as in the Human Genome Project [46], HapMap catalogs SNPs. This constitutes over 90% of genetic variation in the human population. Phase 1 of the International HapMap Project genotyped over 1 million SNPs with minor allele frequencies greater than 0.05 – so called "common" SNPs. The project included 269 individuals with ancestry from 4 geographic populations of Yoruba (Nigeria), Japan, China and North/East Europe (International HapMap Consortium, 2005). Phase 2 of the HapMap Project genotyped an additional 3.1 million SNPs from individuals in the same populations [48]. Due to the relative small sample size of HapMap, only SNPs with minor

allele frequencies over 0.05 can be detected. Detection of "rare" SNPs with allele frequencies less than 0.05 are more efficiently detected in larger sample populations.

Many SNPs do not result in amino acid changes but can alter expression patterns, transcription factor binding, or influence an individual's response to drugs. But despite the complexity of genetic studies, statistically significant regions in the genome have been replicated in multiple genome-wide association studies (GWAS) [49-51].

1.2.1 Genome-Wide Association Studies

Genome-wide association studies (GWAS) examine whole genomes of many individuals to determine the variance existing between each individual. These studies are usually conducted where two groups of individuals are compared – for example, diseased cases and healthy controls – where variation such as single nucleotide polymorphisms (SNPs) or copy number variations (CNVs) is screened. If certain SNP/CNVs are more prevalent in cases, the variance is said to be "associated" with the disease. These associations serve as markers for that region in the genome.

The purpose of GWA studies is to identify genetic variants that occur more frequently in diseased patients or have measureable phenotypes relative to matched controls. The completion of the Human Genome Project[52] and the International HapMap Project[47] have made these studies more technologically possible. Both projects have provided a catalogue of human genetic variation that allows for a systematic approach to scanning entire genomes. Single nucleotide polymorphisms are

the genetic variants measured. SNPs which occur more frequently in disease cases over healthy controls are "associated" with the disease and serve as markers identifying regions in the genome that may contribute to the phenotype.

Prior to GWA studies, candidate gene analysis was a method used to identify genes that increased disease risk. Candidate gene analysis is similar to GWAS however, instead of entire genome scans, a low number of plausibly associated genes are studied for specific variation. Although more efficient, this approach can lead to bias and cannot identify new genetic regions and pathways that may contribute to disease. For non-Mendelian diseases, such as CAD, multiple non-related genetic regions are cumulative in increasing disease risk.

The introduction of gene chips and large-scale sequencing procedures has advanced high-throughput genotyping technology. Gene chips or DNA microarrays can assay hundreds of thousands of SNP regions at once. In this study, the Affymetrix Gene Chip array was used (Affymetrix, Santa Clara, CA, USA). The analysis began with Human SNP Array 5.0 which genotyped 500, 000 SNPs and was upgraded to the 6.0 Array which can scan 1 million SNPs.

The SNPs which are identified as being associated with a disease can be linked to others that are causing the phenotypic effect, i.e. the "functional" SNPs. In genetics, this is due to linkage disequilibrium (LD). LD refers to the degree with which two alleles or loci are genetically inherited together. SNPs which are in strong LD have associated inheritance patterns and genotyping one can give you information about the other. Two values are used to measure LD: r-squared (r^2) and D-prime (D').

in value from 0.0 to 1.0 where 1.0 is "perfect" LD and both loci are always inherited together. Although both measure a degree of linkage, r^2 measures how interchangeable two SNPs are while D' measures the degree of recombination that has occurred between or around two SNPs. If the r^2 value is high, the minor allele of one SNP is always inherited with the minor allele of the other, thus, both SNPs would have similar minor allele frequencies. If the D' value of two SNPs is high, they are inherited together, however their allele frequencies may differ. D' values are always higher than r^2 values for the same two SNPs.

A number of GWAS have been conducted for CAD. A list of major loci that are associated with CAD risk as determined by GWA studies can be seen in **Table 1.1**. One of the first most robust regions to be linked with coronary artery disease risk is the 9p21.3 region[53,54]. Following the initial populations of Northern European descent, 9p21.3 was replicated in the Han Chinese[55,56], Japanese[57,58], Korean[59], Italian[60], Irish[61] and Indian[62] populations among others. To date, 9p21.3 remains to be the most replicated and robust locus associated with CAD risk. Despite its robustness, deciphering the molecular function of 9p21.3 has been challenging. No protein-coding genes fall within the risk region. The closest genes, *CDKN-2A* and *-2B,* are multiple kilobases downstream of the risk region. Recently, our group demonstrated that the risk region encompasses a functional enhancer which controls the expression of *ANRIL,* an anti-sense non-coding RNA molecule involved in the expression of cellular proliferation genes [63].

Band	SNP	Gene(s) in region	n	Risk Allele Freq (risk allele)	OR (95% CI)	p-value
1p13.3	rs599839*	*PSCR1*	83,873	0.78 (A)	1.11 (1.08; 1.15)	$2.89 \cdot 10^{-10}$
1p32.3	rs11206510***	*PCSK9*	102,352	0.82 (T)	1.08 (1.05; 1.11)	$9.10 \cdot 10^{-08}$
1q41	rs17465637****	*MIA3*	25,197	0.74 (C)	1.14 (1.09; 1.20)	$1.36 \cdot 10^{-08}$
2q33.1	rs6725887*	*WDR12*	77,954	0.15 (C)	1.14 (1.09; 1.19)	$1.12 \cdot 10^{-09}$
3q22.3	rs2306374*	*MRAS*	77,843	0.18 (C)	1.12 (1.07; 1.16)	$3.34 \cdot 10^{-08}$
6p24.1	rs12526453*	*PHACTR1*	83,050	0.67 (C)	1.10 (1.06; 1.13)	$1.15 \cdot 10^{-09}$
6q25.3	rs3798220**	*LPA*	32,584	0.02 (C)	1.54 (1.36; 1.74)	$9.62 \cdot 10^{-12}$
9p21.3	rs4977574*	*CDKN2A/B, ANRIL*	84,256	0.46 (G)	1.29 (1.23; 1.36)	$1.35 \cdot 10^{-22}$
10q11.21	rs1746048***	*CXCL12*	136,416	0.87 (C)	1.09 (1.07; 1.13)	$2.12 \cdot 10^{-10}$
12q24.12	rs3184504*	*SH2B3*	67,746	0.44 (T)	1.07 (1.04;1.10)	$6.35 \cdot 10^{-06}$
19p13.2	rs1122608*	*LDLR*	49,693	0.77 (G)	1.14 (1.09; 1.18)	$9.73 \cdot 10^{-10}$
21q22.11	rs9982601*	*MRPS6*	46,230	0.15 (T)	1.18 (1.12; 1.24)	$4.22 \cdot 10^{-10}$

Table 1.1: Major loci for coronary disease identified and replicated in GWA studies (genome-wide significance p<5E-8).

A number of loci and genes have been replicated in genome-wide association studies for CAD risk and lipid traits. In addition to 9p21 – the most robust loci implicated in CAD risk – and the *APOE/C1/C4* loci, many SNPs and genes repeatedly emerge as being statistically associated. A few have been listed here from two major GWAS meta-analyses conducted in 2009 and 2008. It is to be noted that genes in addition to those playing a role in lipid homeostasis are also associated.

Another locus strongly associated with CAD risk is the 19q13.2 *APOE-APOC1/C4* region that has been replicated in multiple GWAS (see **Figure 1.7**) [64,49,50]. Specifically, SNP rs4420638, which falls 3' of the *APOC1* gene, has a reported odds ratio (OR) of 1.17 (1.08-1.28) for CAD risk[50]. Its minor allele frequency (MAF) is 0.18. Although this is the only SNP of this region present on the Affymetrix gene chip, there are many additional SNPs of interest at this locus. Two of particular interest are rs429358 and rs7412, which lie in the protein-coding region of *APOE*. These SNPs are non-synonymous and give rise to the three isoforms of apoE protein: apoE2 ('protective' allele), apoE3 and apoE4 (risk allele). (See Section 1.3.2 for the specifics on each isoform.) A recent meta-analysis reported that each copy of apoE4 isoform increased CAD risk with and odds ratio (OR) of 1.06 whereas carriers of the apoE2 isoform conferred protection with an OR of 0.8 [65]. HapMap reports r^2 between rs429358, the apoE4 determining SNP, and rs4420638 as 0.70. There is no allele linkage between rs7412, the apoE2 determining SNP, and either risk SNPs ($r^2 \approx 0.0$). The MAF of rs429358 is 0.15 and rs7412 is 0.07.

1.3 - Effects of Genetic Variation in APOE on Apolipoprotein E Structure and Function

Three apoE isoforms exist due to two SNPs. SNP rs429358 alters amino acid 112 and rs7412 alters residue 158. The three apoE isoforms, E2, E3, and E4 contain cysteines or arginines at either position. The three alleles give rise to six different genotypes in the human population: three homozygote, ε3/ε3, ε2/ε2, ε4/ε4 and three heterozygote, ε2/ε3, ε4/ε3 and ε2/ε4.

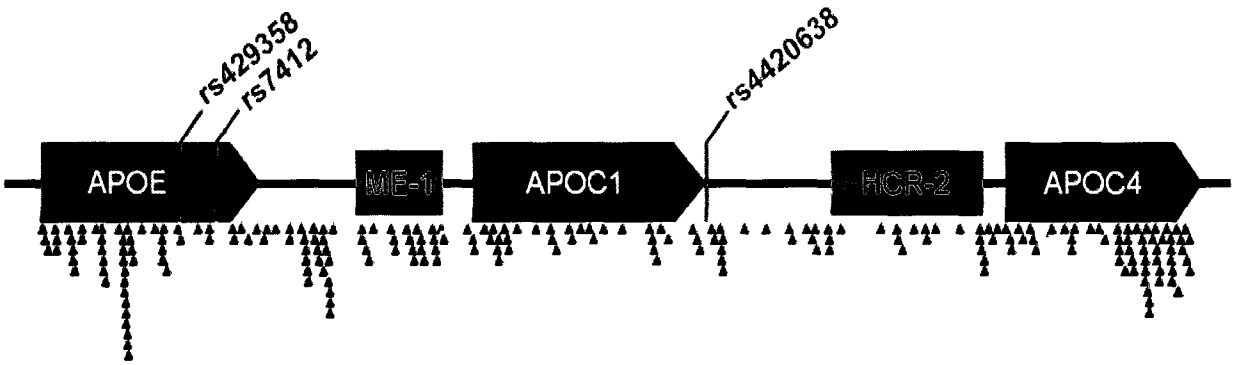

rs429358
rs7412
rs4420638
APOE
ME-1
APOC1
HCR-2
APOC4

Figure 1.7: Schematic diagram of region 19q13.2 with all known polymorphisms.

There are over 225 SNPs in this 39kb locus, as indicated by the small blue triangles. Only 23 of these have been characterized in one or more populations and are included in HapMap genotype data. This includes rs4420638, rs429358 and rs7412. The relative positions of rs429358 (chromosomal position: 45411941), rs7412 (45412079) and rs4420638 (45422946) are indicated in the diagram. The only SNP in this region on the Affymetrix gene chip is rs4420638. Blue arrows: transcriptional orientation of apolipoprotein genes. Purple boxes: enhancer regions. ME-1 enhancer region controls apolipoprotein expression in the central nervous system and could play a role in Alzheimer's disease. HCR-2 enhancer region controls expression in hepatocytes and could affect CAD physiology.

1.3.1 ApoE Protein Structure

Apolipoprotein E is a 299 amino acid protein. Its N-terminal domain is a 4-helix amphipathic bundle while its C-terminal is free-structure α-helices (**Figure 1.8**). The 4-helix bundle is the receptor-binding region while the α-helices at the C-terminal bind lipids. Thus, only in the lipid-bound state does the C-terminal have tertiary structure. Although each domain serves an independent purpose, interactions between the domains occur; a lipid-bound C-terminal is required for the N-terminal to bind LDL receptors with high affinity. Hence, lipid-free apoE cannot bind receptors efficiently. [66]

1.3.2 Three Isoforms of ApoE

Both *APOE* SNPs are in the receptor-binding region. ApoE3 is the common isoform and has an arginine at residue 158 and a cysteine at 112. ApoE4 has arginines at both residues. In E4, the charged, bulky arginine at position 112 interacts with a glutamate residue within the C-terminal lipid-binding region. This domain-interaction hinders the 'free' C-terminal α-helices (see **Figure 1.9**). As a result, apoE4 does not bind lipoproteins as efficiently as E3 and is found bound to larger VLDL and CR molecules rather than HDL. LDL receptor binding affinity of apoE4 is not affected. Since apoE4 binds more VLDL versus HDL, VLDL conversion to LDL is increased resulting in higher LDL-cholesterol (LDL-C) levels. [66]

ApoE2 has cysteines at both residues. The lipid-binding domain is not affected as in apoE4, thus apoE2 efficiently binds VLDL, CR and HDL though prefers smaller HDL molecules. However, the cysteine at residue 158 cannot form a salt-bridge with an

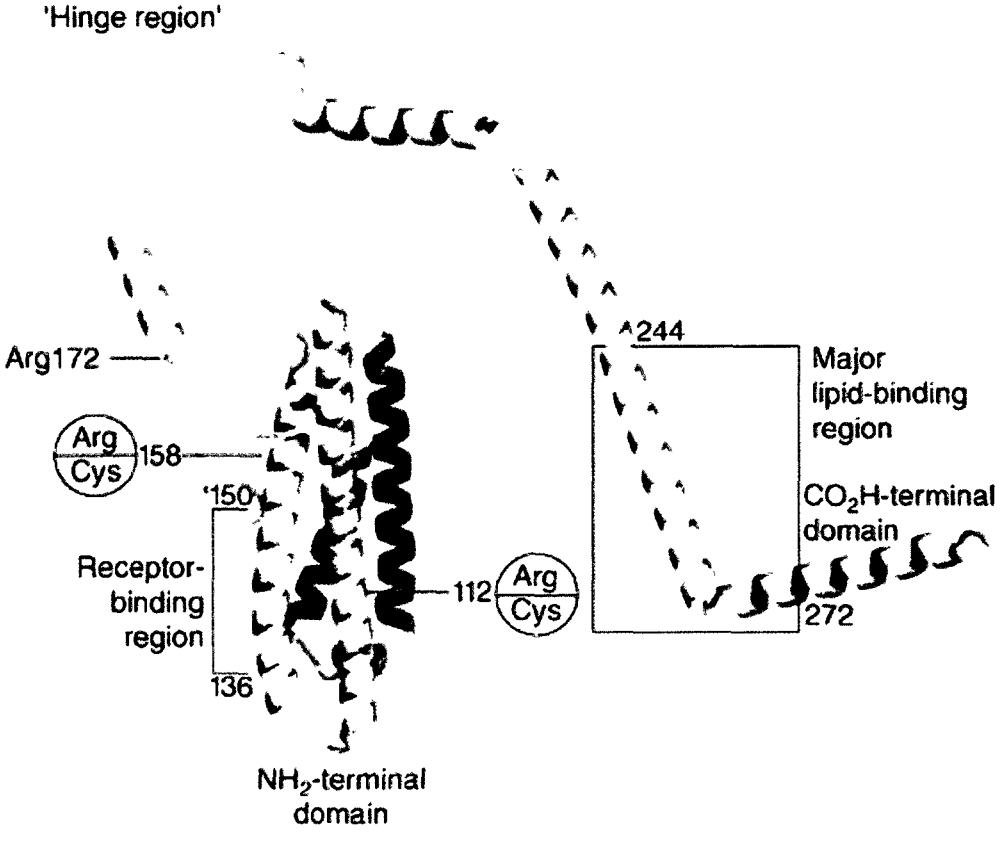

'Hinge region'
Arg172
Arg
Cys
158
150
Receptor-
binding
region
136
112
Arg
Cys
244
Major
lipid-binding
region
CO_2H-terminal
domain
272
NH_2-terminal
domain

Figure 1.8: Ribbon model of unlipidated apolipoprotein E.

N-terminal receptor binding region (coloured 4-alpha-helix bundle) is obtained from X-ray crystallography while the linker 'hinge region' and C-terminal lipid-binding region was based on structure prediction models due to its free-structure nature. SNP rs429358 alters amino acid position 158 while rs7412 alters position 112 to either arginine or cysteine. [66]

Figure taken from Hatters et al., (2006) Apolipoprotein E structure: insights into function. *Trends Biochem. Sci* 31(8): 445-454, with permission. License agreement number 24168120089745

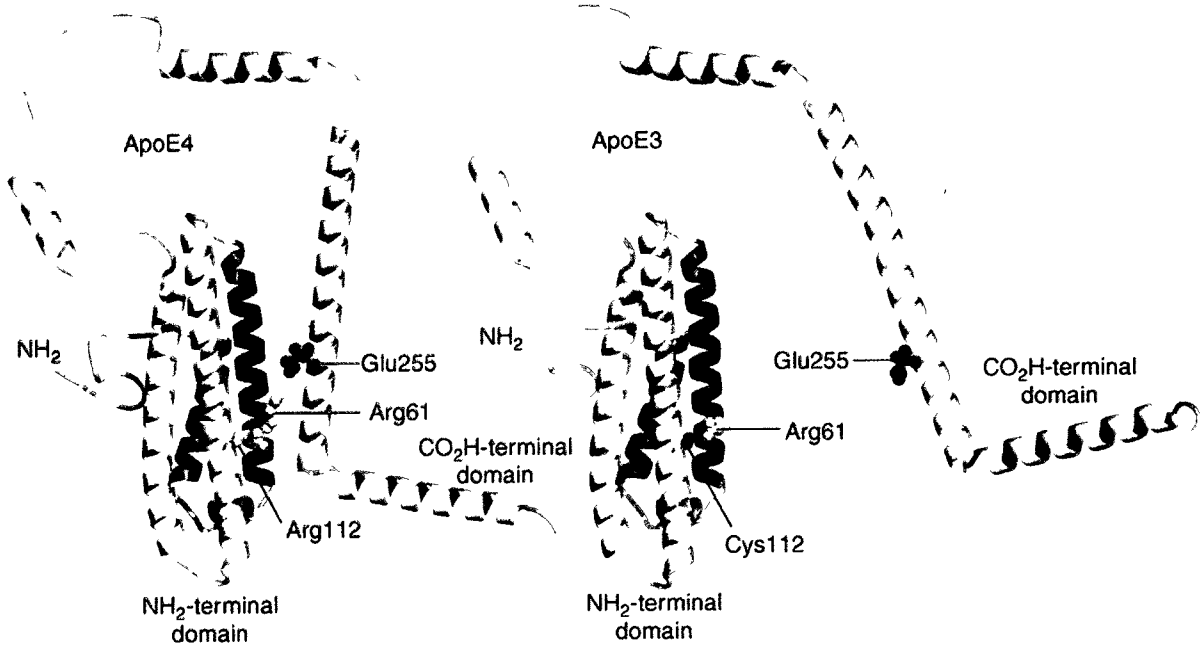

ApoE4
NH2
Glu255
Arg61
CO2H-terminal domain
Arg112
NH2-terminal domain
ApoE3
NH2
Glu255
CO2H-terminal domain
Arg61
Cys112
NH2-terminal domain

Figure 1.9: Crystal structure of unlipidated apoE4 versus apoE3.

Wild-type apoE3 contains a cysteine at position 112 (cys112). This cysteine affects the positioning of side-chain arginie at position 61 (arg61) which is oriented facing into the 4-helix bundle in apoE3. In apoE4, position 112 is an arginine (arg112) which disrupts the salt-brides orienting arg61. As a result, arg61 is positioned outward and can interact with glutamate 255 (glu255) on the free lipid-binding arm. This constricts the lipid-binding region resulting in apoE4's preferential binding to large lipoproteins VLDL and IDL. The receptor-binding region remains unaffected in apoE4 [66].

Figure taken from Hatters et al., (2006) Apolipoprotein E structure: insights into function. *Trends Biochem. Sci* 31(8): 445-454, with permission. License agreement number 24168120089745

aspartate in the receptor-binding region like wild-type arginine. As a result, the aspartate interacts with another arginine – one that is required for LDLr binding (**Figure 1.10**) [66]. Because of this, apoE2 has weak LDLr binding affinity – roughly 1% the affinity of E3 and E4. Thus, the uptake of apoE2 lipoproteins by hepatic LDLr is impaired. When hepatic LDLr expression is reduced by a cholesterol rich diet or hypothyroidism in subjects who are homozygous for the ε2 allele, VLDL and chylomicron remnants accumulate in the plasma sometimes resulting in a rare lipid disorder known as Type III dysbetalipoproteinemia. However, in general, ε2 carriers have a 20% lower risk of developing CAD, likely due to lower LDL-C levels[65].

1.3.3 ApoE and CAD

Studies have shown that individuals carrying the ε4 allele have a greater risk of developing cardiovascular disease relative to the more common ε3/3 individuals. The OR for each ε4 allele is 1.06 (0.99-1.13) as reported in a recent meta-analysis. The ε2 allele is 'protective' for CAD and, from the same meta-analysis, was determined to have an odds ratio of 0.80 (0.70-0.90) [65]. **Table 1.2** summarizes the structure and function of each apoE isoform.

1.3.4 ApoE in Alzheimer's Disease

After the liver, apoE has the highest expression level in the brain [67]. ApoE is the dominant apolipoprotein synthesized in the central nervous system (CNS). It is found on HDL-like particles, which are the only lipoproteins present in the CNS, or more specifically, in the cerebrospinal fluid (CSF). These particles are important in the

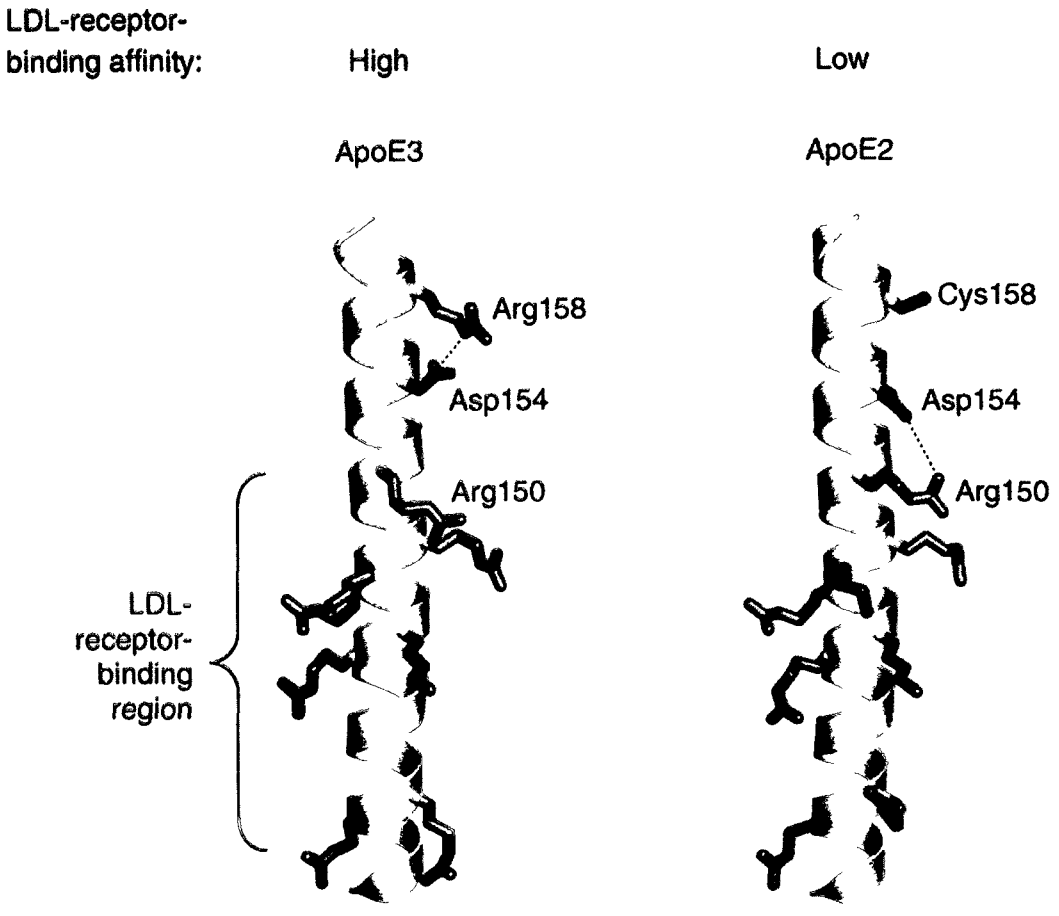

LDL-receptor-binding affinity:
High
Low
ApoE3
ApoE2
Arg158
Asp154
Arg150
Cys158
Asp154
Arg150
LDL-receptor-binding region

Figure 1.10: Ribbon structure of apoE3 versus apoE2.

Receptor binding region indicating amino acid change resulting from SNP rs7412 (See **Figure 1.9** for full ribbon structure for complete apoE structure). Wild-type apoE3 has an arginine at position 158 (Arg158), which forms a salt bridge with aspartate 154 (Asp154). When position 158 is a cysteine (cys158) as in apoE2, Asp154 must form a salt-bridge with arginine at position 150 (Arg150). This is a vital arginine required for optimal apoE receptor binding. As a result apoE2 has 1% the receptor binding affinity as apoE3 or apoE4. The lipid-binding region remains unaffected in apoE2. [66]

Figure taken from Hatters et al., (2006) Apolipoprotein E structure: insights into function. *Trends Biochem. Sci* 31(8): 445-454, with permission. License agreement number 24168120089745

Isoform	Freq	Amino Acid Residue		Functional Differences		Relationship to LDL-C levels	CAD Risk (OR)	Associated Disorders
		112	158	LDLr binding	Lipoprotein binding			
ApoE2	0.07	Cys	Cys	Low	HDL	Decreased	0.80	Type III hyperbetalipoproteinemia (rare)
ApoE3	0.78	Cys	Arg	High	HDL	Neutral	1.00	None
ApoE4	0.15	Arg	Arg	High	VLDL, IDL	Increased	1.06	Atherosclerosis, Alzheimers

Table 1.2: ApoE isoforms: structure and function (Hatters, Peters-Libeu, and Weisgraber 2006).

movement and delivery of cholesterol, which is required to maintain the proper function of synapses. Due to the blood-brain barrier, all cholesterol required by the brain is synthesized in situ. [68]

In addition to CAD, the *APOE* locus is a major genetic risk factor for late-onset Alzheimer's disease (AD). In the Caucasian population, heterozygous carriers of the ε4 allele are 3-fold more likely to develop AD and the risk increases 15-fold for homozygous carriers [69]. Similar to CAD, ε2 is protective against Alzheimer's disease [69]. GWAS have found the *APOE* region to be the strongest genetic risk factor for AD – the risk SNP being the only *APOE* SNP represented on the commonly used gene chips, rs4420638 [70]. The allele specific OR for AD for rs4420638 is 4.0 [71].

Numerous studies have attempted to decipher the mechanism by which apoE4 promotes AD, but the complete story remains elusive. Pathophysiologically, AD is characterized by neuronal, brain and synaptic atrophy. The formation of extracellular plaques composed of amyloid-β (Aβ) peptide and intracellular aggregates of phosphorylated Tau protein initiates and mediates the progression of the disease [72]. Aβ is cleaved from amyloid precursor protein (APP) by the enzyme β-site APP cleaving enzyme (BACE1) .

Some studies indicated that apoE may be an Aβ-binding protein that could interfere with peptide clearance [73]. The ε4 allele correlated with increased Aβ density in the brain of individuals at an early age [74]. However, other findings were contradictory [75,76]. *In vitro* studies demonstrated that lipidated apoE2 and E3 formed stronger complexes with Aβ-peptide than apoE4 [77]. Because of this inverse correlation

it was hypothesized that apoE2 and E3 bound and cleared Aβ more readily than E4. However, due to differing preparations and conditions used to study Aβ aggregation, understanding the role of apoE4 in AD requires further investigation.

1.4 - Hypothesis and Objectives

At the *APOE/C1/C4* locus, rs4420638 risk is strongly associated with plasma lipids and risk for both CAD and AD. This association has been attributed to its linkage to the apoE4 SNP, rs429358.

However, rs4420638 has a reported OR of 1.17 (1.08-1.28) for CAD risk[50] while a recent meta-analysis reported apoE4 isoform with and odds ratio (OR) of 1.06 [65]. HapMap reports r^2 between rs429358, and rs4420638 as 0.70. Additionally, in the Ottawa Heart Study (OHS), it was noted that the odds ratio of rs4420638 for CAD was considerably higher than that reported for the apoE4 determining SNP, rs429358. The MAF and r^2 values do not indicate near-perfect linkage. The rs4420638 SNP is also 5' to *APOC1* – an apolipoprotein which has an additional and important role in lipid metabolism.

Our original hypothesis was that rs4420638 alters CAD risk in part independently of the apoE4 and E2 SNPs. In addition, we sought to determine the effects of rs4420638 on plasma concentrations of apoE and apoC1.

The objectives of this project were to address the following:

1. Determine the effect of rs4420638 on plasma lipoprotein concentrations relative to apoE SNPs, rs429358 (ε4) and rs7412 (ε2). *Specifically, is rs4420638 a stronger determinant of plasma concentrations of triglycerides and LDL-C as compared to the apoE SNPs?*

2. Determine the odds ratio of rs4420638 for CAD relative to apoE SNPs ε4 and ε2. *Specifically, is rs4420638 more strongly associated with CAD risk as compared to the apoE SNPs?*

3. Analyze linkage between rs4420638 and apoE SNPs, lipoprotein association and CHD risk of potential SNPs indicated in **Table 1.3**.

4. Determine if genotypes of any or all of rs4420638 and *APOE* SNPs correlated with apoE and/or apoC1 serum protein expression by an ELISA assay

SNP	Position	Associated Trait	p-value
rs769449	50103781, ApoE intron	LDL-C	1.77×10^{-5}
rs769450	50102284, ApoE intron	HDL-C	0.023
rs445925	50107480, 3' ApoE	LDL-C	4.3×10^{-11}
rs483082	50108018, 3' ApoE, ME-1	LDL-C	0.084
rs3513657	50131003, HCR-2	LDL-C	4×10^{-4}

Table 1.3: Potential SNPs associated with lipid traits in the APOE/C1/C4 region.

These SNPs were genotyped in the same population as rs4420638 and APOE SNPs to determine linkage association.

2. MATERIALS AND METHODS

2.1 - Subjects

Controls. Healthy subjects with no symptoms, personal or family history of coronary artery disease were recruited. These included European white males over the age of 65 and women over 70 years as part of the Ottawa Heart Study (OHS). All subjects resided in the Ottawa region. For serum analysis of apoE and apoC1 by ELISA analysis, control subjects with no diabetic history and not on any lipid-altering medications were chosen. Written informed consent was obtained from all participants.

CHD Cases. Because early-onset CAD has a strong genetic component, cases were selected on the basis of severe premature coronary artery disease as determined by angiograms, angina, or coronary surgeries including stent or bypass surgery. Subjects were European white men younger than 60-years and women younger than 65-years. The physical characteristics of cases and controls can be seen in **Table 2.1**.

2.2 - Genotyping by Affymetrix® Genome-Wide Human SNP Array

DNA samples of cases and controls were genotyped by Affymetrix Gene Chip 500K or 6.0 arrays. The Affymetrix Genome-Wide Human SNP Nsp/Sty 6.0 Array (Affymetrix, CA) kit was used. A schematic diagram adapted from the kit protocol can be seen in **Figure 2.1**. The optimal protocol was followed as described in the

Trait	Case (n = 1083)	Control (n = 2145)
Age (y)	48.2 ± 7.2	75.0 ± 5.2
Gender - male	82.0%	51.6%
Non-smoker (ever)	31.6%	52.7%
Former smoker	56.8%	45.0%
Current smoker	11.5%	2.3%
Hypertension - yes	44.2%	36.4%
Average BMI (kg/m^2)	28.8 ± 4.6	26.1 ± 4.0
Triglycerides (mmol/L)	2.6 ± 2.2	1.4 ± 0.9
Total cholesterol (mmol/L)	6.1 ± 1.2	5.7 ± 1.0
LDL (mmol/L)	3.9 ± 1.0	3.6 ± 0.9
HDL (mmol/L)	1.1 ± 0.4	1.5 ± 0.4

Table 2.1: Table of physical characteristics of Cases and Controls.

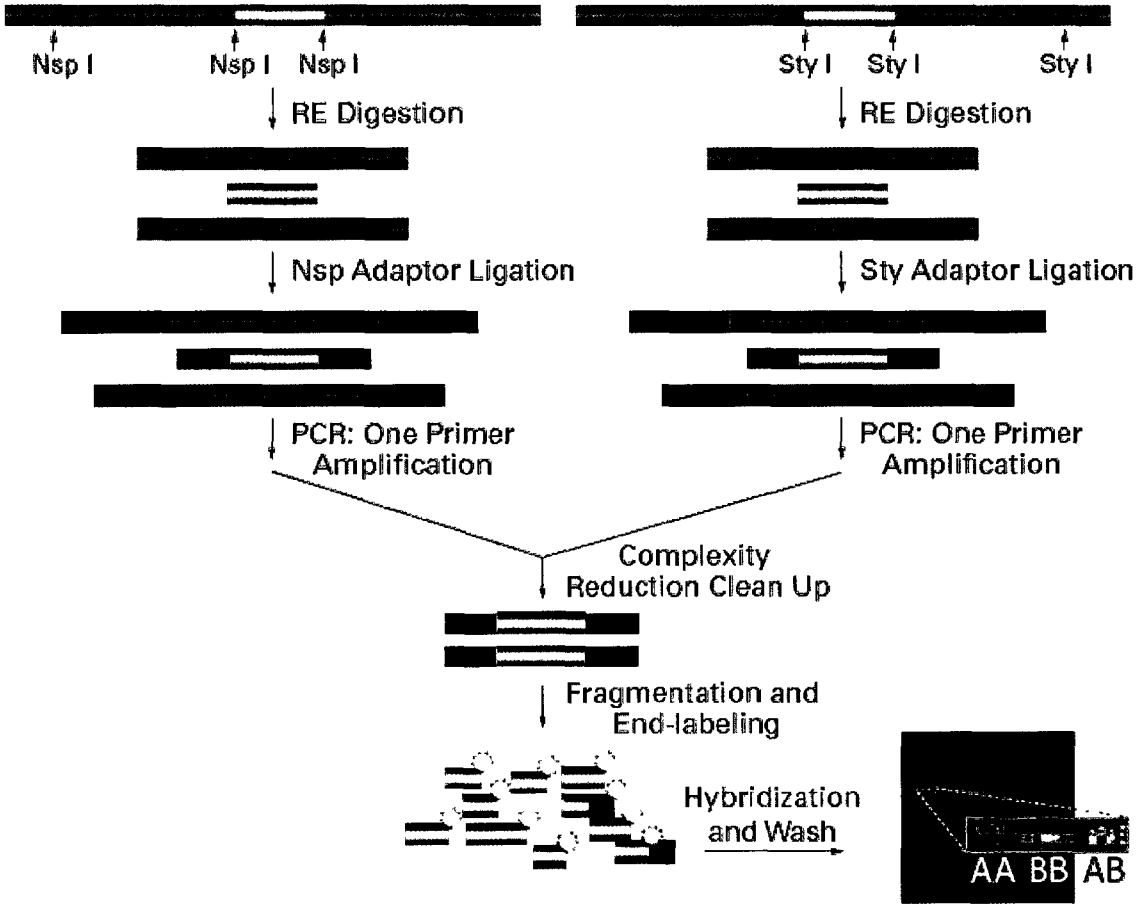

Nsp I
Nsp I
Nsp I
RE Digestion
Nsp Adaptor Ligation
PCR: One Primer Amplification
Sty I
Sty I
Sty I
RE Digestion
Sty Adaptor Ligation
PCR: One Primer Amplification
Complexity Reduction Clean Up
Fragmentation and End-labeling
Hybridization and Wash
AA BB AB

Figure 2.1: Schematic overview of genotyping by Affymetrix® Genome-wide Human SNP Array.

Two aliquots of each DNA sample are digested by NspI and StyI restriction enzymes. Adaptors that contain the primer recognition sequence are ligated to DNA fragments. A generic primer that recognizes the adaptor sequence is used to amplify the DNA fragments. Both reactions are combined and purified by a procedure using polystyrene beads. The purified DNA amplicons are fragmented, labeled and hybridized to gene chips. The array is subsequently analyzed by a genotyping calling algorithm called Birdseed. [78]

user guide. DNA samples were digested by two restriction enzymes NspI and StyI in separate reactions. Adaptors specific to NspI and StyI were ligated to digested DNA fragments followed by amplification of the fragments. The two reactions were combined and purified using polystyrene beads. Amplified DNA was fragmented, labeled and hybridized to gene chips. Subsequent washing and staining of gene chip arrays was run on Affymetrix GeneChip® Fluidics Station 250. The genotyping console uses a genotyping calling algorithm called Birdseed. All Affymetrix gene chip array procedures were conducted by technicians in the Ottawa Heart Institute Atherogenomics Laboratory as well as the John & Jennifer Ruddy Canadian Cardiovascular Genetics Centre.

2.3 - Genotyping by Sequencing

Seven SNPs were genotyped in the *APOE/C* locus by sequencing 4 polymerase chain reaction (PCR) fragments of 280-850 base pairs (bp) (**Figure 2.2**). Each PCR reaction contained 100-500ng of purified DNA sample, 0.4uM each PCR primer (Table 2) in 1x PCR buffer with 1x GC-rich solution and 0.2mM each dNTPs (Roche, FastStart Taq DNA Polymerase, dNTPack) in a 25ul reaction. PCR reactions were run on Eppendorf Mastercycler or Applied Biosystems (ABI) GeneAmp PCR System 9700 machines (95° for 3min, 95° for 30sec, annealing temp for 30sec, 72° for 1min30sec, 72° for 5min) for a total of 45 cycles at the respective annealing temperatures specified in **Table 2.2**.

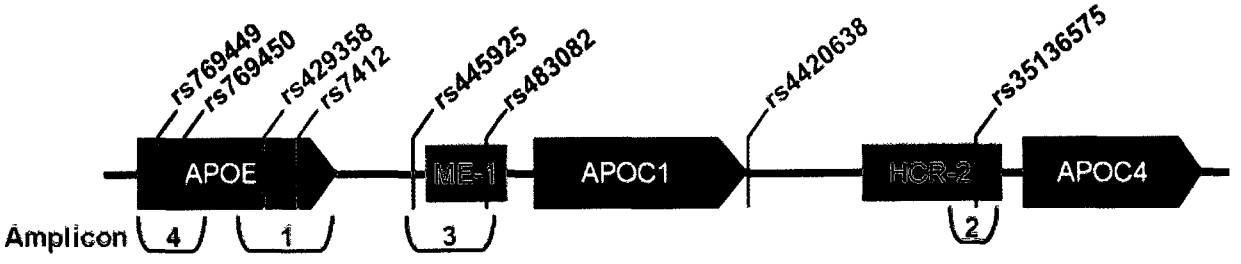

rs769449
rs769450
rs429358
rs7412
rs445925
rs483082
rs4420638
rs35136575
APOE
ME-1
APOC1
HCR-2
APOC4
Amplicon
4
1
3
2

Figure 2.2: Schematic diagram of PCR amplicons used to sequence each SNP.

Amplicon 1 included rs429358 and rs7412 and was 838 base pairs. Amplicon 2 was the smallest at 285 base pairs and included rs35156575 only. Amplicon 3 included rs445925 and rs483082 and was 697 base pairs. Amplicon 4 sequenced rs769449 and rs769450 and was 695 base pairs long. The only SNP on the whole-genome Affymetrix gene chip array was rs4420638 and thus, did not need to be manually sequenced for genotyping.

Amplicon	SNPS	PCR Primers	Sequencing Primer	Anneal Temp	PCR Length
1	rs429358 rs7412	Fwd: 5'GCCTCCTAGCTCCTTCTTCG Rev: 5'GCTGCATGTCTTCCACCAG	5'CTTCTCTCCCTCTTGGGTCTCTCT	52 °C	838 bp
2	rs35136575	Fwd: 5'TTTGGGGCTTGGTGACTTAG Rev: 5'TCCAAGAGATCAAGGGGATG	5'TTTGGGGCTTGGTGACTTAG	60.8 °C	285 bp
3	rs445925 rs483082	Fwd 5'GGAGTCGTATGGGACAGGAG Rev 5'TGGTTGTGAAAAGGACAGCA	5'GGAGTCGTATGGGACAGGAG	63.5 °C	697 bp
4	rs769449 rs769450	Fwd: 5'AATCACGGCAGGAAGATGAA Rev:5'TGTATTGAGTGTCCTTGTGTG	5'AATCACGGCAGGAAGATGAA	58.1 °C	695 bp

Table 2.2: Four PCR amplicons were designed to include all seven SNPs for sequencing.

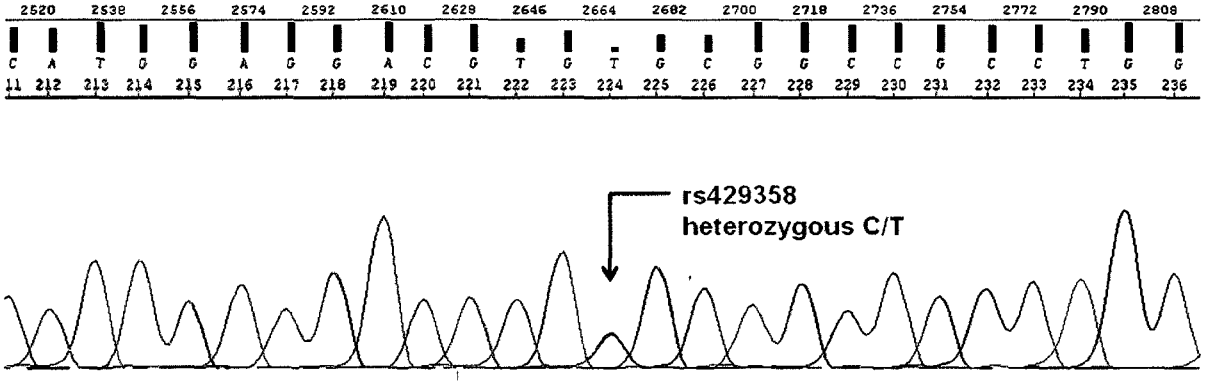
The PCR primers used to amplify each fragment are listed. Optimal sequencing primers were used for subsequent sequencing analysis. Only one PCR amplicon required a nested sequencing primer for optimal sequencing. All other products could be sequenced using the forward PCR primer as the sequencing primer. Optimal annealing temperatures were determined for each PCR primer pair. Fragment length refers to the length of the PCR product amplified for sequencing.

The PCR amplified reaction was used directly for subsequent sequencing. Sequencing was performed according to BigDye® Terminator v3.1 Sequencing Kit (ABI Prism®). Each 20ul reaction contained 1.5ul of PCR reaction, 0.16uM of respective sequencing primer (**Table 2.2**), 0.25x Ready Reaction (RR) mix and 0.625x BigDye Sequencing Buffer. Sequencing PCR was preformed on Eppendorf or ABI machines (96° for 30sec, 50° for 15sec, 60° for 4min for 25 cycles). Sequencing reactions were purified using Montage 96-well Sequencing Reaction Cleanup plates (Millipore). Reactions were analyzed on 48-channel 3730 DNA Analyzer (ABI) followed by sequencing analysis on DNA Sequencing Analysis Software v5.2 (ABI). Examples of analyzed sequences are shown in **Figure 2.3**.

2.4 - PCR and Sequencing Reaction Set-up

PCR reactions, sequencing reactions and sequence reaction clean-up steps were fully automated. The Hamilton Microlab STARlet pipettor was used to set up sample reactions in 96-well plates. This fully-automated equipment ensures accurate and precise uptake of reagent volumes. Pipetting of DNA and mastermix volumes were consistent for each sample genotyped. Venus One software technology was used to program each step in the PCR, sequencing and sequence reaction clean-up protocols.

(a)


2520 2538 2556 2574 2592 2610 2628 2646 2664 2682 2700 2718 2736 2754 2772 2790 2808
C A T D G A G G A C G T G T G C G G C C G C C T G G
11 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236
rs429358
heterozygous C/T

(b)

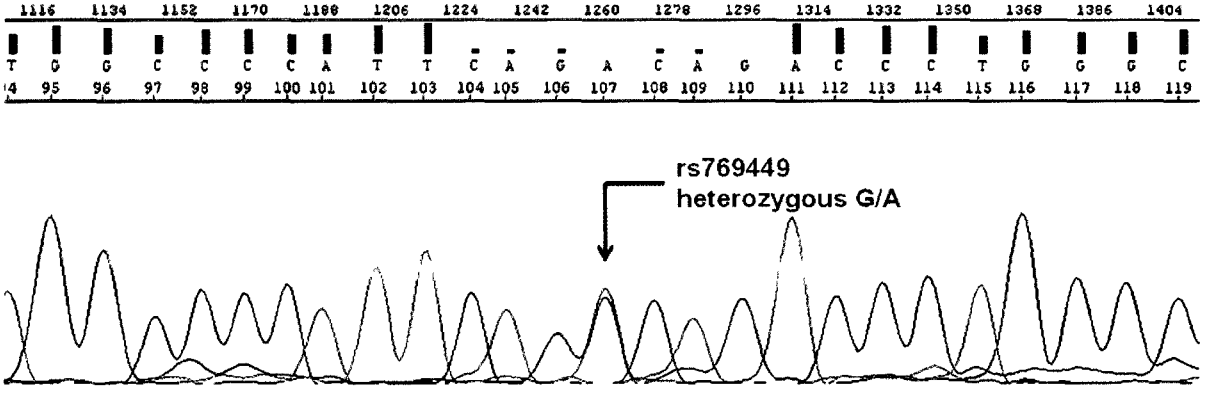

1116 1134 1152 1170 1188 1206 1224 1242 1260 1278 1296 1314 1332 1350 1368 1386 1404
T G G C C C C A T T C A G A C A G A C C C T G G G C
4 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119
rs769449
heterozygous G/A

Figure 2.3: Example sequence result of amplicons after analysis by DNA Sequencing Analysis Software v5.2 (ABI).

Sample sequence outputs of Amplicon 1 and Amplicon 2 show two heterozygous genotypes. Base pair signals were visually analyzed to determine each SNP's genotype since analysis software could not call a heterozygous genotype correctly. Bars above the called base indicate the quality of the sequence sample called where blue is accuate and red bars indicate a potential inaccuracy in base calling. Each colored peak represents one base where blue is cytosine (C), green is adenine (A), red is thymine (T) and black is guanine (G). (a) Amplicon 1: base pairs 212 to 236 are shown which include heterozygous C/T rs429358. (b) Amplicon 4: base pairs 95 to 119 are shown including heterozygous G/A rs769449.

2.5 - ApoE and ApoC1 ELISAs

ApoE and apoC1 concentrations in control serum samples were determined using competitive ELISA technique. Each serum sample was measured four times for each apolipoprotein; two 1/20 dilutions, and two 1/40 dilutions. Each apoliporotein ELISA was conducted concurrently for every serum sample. Phosphate-buffered saline (PBS) at pH 8.0 was used throughout. Commercially purified, human, lipid-free apolipoprotein (apoE and apoC1: Biodesign International) was coated overnight on 96-well plates holding high-binding polystyrene Microtiter® strips (Thermo Labsystems) at 4°C. A volume of 100ul of 0.5ug/ml of apolipoprotein in PBS-0.1% sodium azide was used to ensure saturated coating of each well. Concurrently, two dilutions (1/20 and 1/40) of human serum were incubated with 1/12800 of goat polyclonal, biotin-conjugated apoE and apoC1 antibody (apoE: Biodesign International; apoC1: Abcam Inc.) in PBS-0.1% sodium azide-0.5% bovine serum albumin (BSA)-0.05% Tween-20 (sample buffer) at 4°C overnight. Before use, the coated plates were washed with PBS containing 0.5% Tween-20 and blocked for 1 hr with 0.5% BSA in PBS at room temperature with shaking. To compete out the bound antibodies incubating with human serum, 100ul of each antibody-serum solution was added to each well and incubated for 1hr at 37°C. The plates were subsequently washed, and 1:30000 avidine-alkaline phosphatase (Sigma) in sample buffer was added to each well to bind to the biotin. Room temperature, ready-mixed alkaline phosphatase yellow-liquid substrate (Sigma) was added to each well (100ul) and incubated for 20-40 min. After incubation,

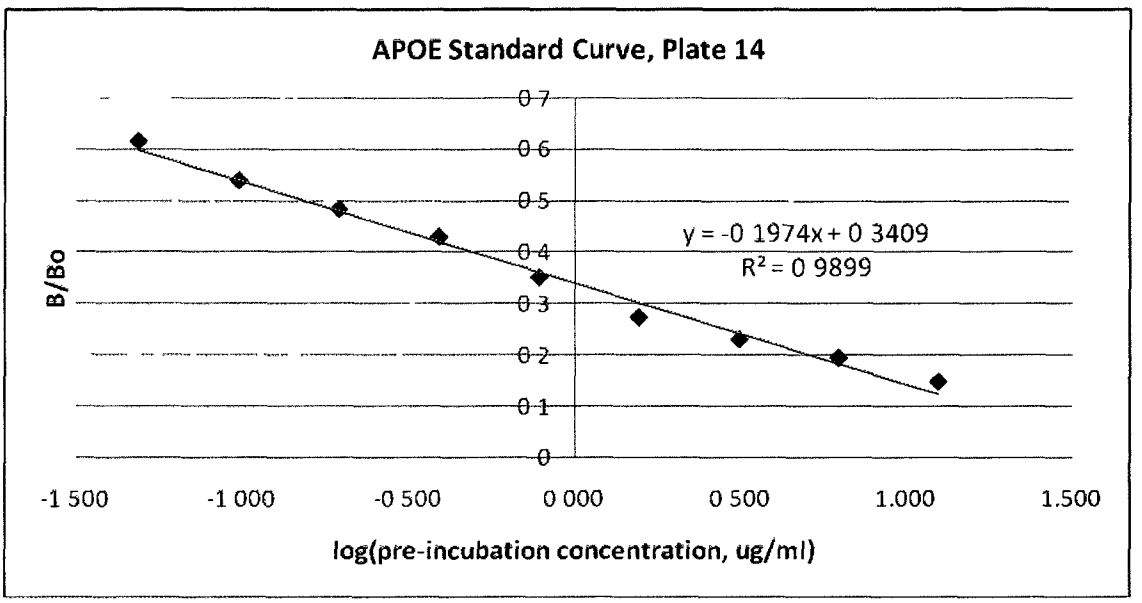
plates were read at 405nm on Microplate Reader Model 3550 (BioRad). A complete detailed protocol can be found in **Appendix 1: Detailed Competition ELISA Protocol Used to Determine Serum ApoE and ApoC1 Concentrations.**

The numerical readings of each well corresponded to the amount of free antibody present in the overnight serum-antibody incubation. Thus, the readings were inversely correlated with the amount of apolipoprotein present in each serum sample. Standard samples of pure apolipoprotein were serially diluted from 25.0 ug/ml to 0.0 ug/ml and were present in duplicate on each plate. Sample standard curves for apoE and apoC1 are shown in **Figure 2.4.** Additionally, two wells were left uncoated as a negative control to ensure complete apolipoprotein coating. Complete calculations used to determine each serum sample apoE and apoC1 concentrations can be found in **Appendix 2: Sample ApoE and ApoC1 Serum Concentration Calculations from Plate 14.**

2.6 - Statistical Analysis

Minor allele frequencies of each SNP in the OHS population were obtained by direct counting of genotyped samples in cases and controls. Allelic p-values were generated using a simple univariate linear regression model (**Table 3.4**). No physical or biological traits were adjusted for. The Bonferroni correction was applied to account for multiple testing. Linkage disequilibrium r^2 and D' values were calculated using R Programing (**Table 3.5**). Haploview 4.2 software was used to construct the linkage disequilibrium plot for controls (**Figure 3.2**). Similar D' values were obtained from both analyses. Haplotype frequencies were calculated using an estimated 2-locus

a)


APOE Standard Curve, Plate 14
0 7
0 6
0 5
0 4
0 3
0 2
0 1
0
B/Bo
y = -0 1974x + 0 3409
R² = 0 9899
-1 500
-1 000
-0 500
0 000
0 500
1.000
1.500
log(pre-incubation concentration, ug/ml)

b)

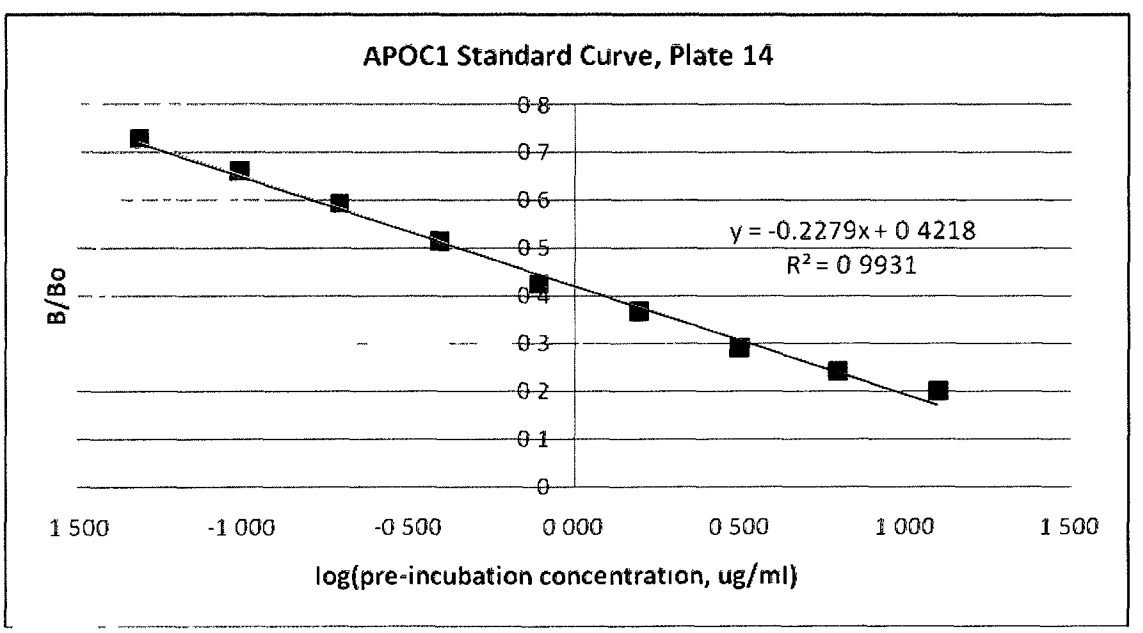

APOC1 Standard Curve, Plate 14
0 8
0 7
0 6
0 5
0 4
0 3
0 2
0 1
0
B/Bo
y = -0.2279x + 0 4218
R² = 0 9931
1 500
-1 000
-0 500
0 000
0 500
1 000
1 500
log(pre-incubation concentration, ug/ml)

Figure 2.4: Examples of standard curves obtained from Plate 14.

(a) APOE and (b) APOC1 ELISA assays. Each ELISA assay contained duplicates of nine standard samples of pure apoliporotein serially diluted from 25.0 ng to 0.0 ng (25.0, 12.5, 6.25, 3.13, 1.56, 0.78, 0.39, 0.20, 0.10, 0.05, 0.0 ug/ml). The absorbance readings at 405nm of each standard was averaged ("B") and normalized by dividing by the 0.0 ug/ml reading ("Bo"). The y-axis plots the normalized average absorbance ("B/Bo") of each serially diluted standard. These values are plotted against the log of each standard concentration (x-axis). Linear regression was used to obtain standard curve equations. The linear regression equations were used to determine the concentration of APOE and APOC1 in serum samples. All plates had standard curves with $R^2 > 0.98$.

haplotype frequency analysis (**Table 3.6 a, b, c**). For apoE and apoC1 ELISA data, a simple univariate linear regression model was applied to each genotype (**Table 3.7**, **Table 3.8**). ELISA data was not corrected for physical or biological characteristics, however, serum samples from non-diabetic individuals that were not on lipid-altering drugs were used. The Bonferroni correction was applied to account for multiple testing. Simple linear regression was also applied to plasma lipid analysis, LDL, HDL, TG, TC (**Table 3.9**, **Table 3.10**). Odds ratios were calculated using standard calculations programmed in R. Odds ratios and 95% confidence intervals for SNP and apoE genotypes were calculated using R programming. All statistical calculations were programmed in R with the help of statistician Robert Davies of the Ottawa Heart Institute.

3. RESULTS

3.1 - Minor allele frequencies and odds ratios of genotyped APOE/C1/C4 SNPs in the Ottawa Heart Population

To address the aforementioned aims, both ApoE SNPs rs429358 (ε4) and rs7412 (ε2) were genotyped by sequencing in the Ottawa Heart Population of elderly controls and early CAD cases. Healthy controls showed no symptoms or family history of coronary artery disease. Men over the age of 65 and women over 70-years of age were recruited as controls for this study. Early CAD cases were determined by angiograms, or had undergone stent or bypass surgeries. Cases included men younger than 60-years and women younger than 65-years of age. Physical characteristics of cases and controls are summarized in **Table 3.1**.

A total of over 1300 controls and over 1000 cases were genotyped for rs429358 and rs7412. The minor allele frequency of these SNPs was 0.12 for rs429358 and 0.08 for rs7412 in controls. The frequencies in cases was increased for the ε4 allele at 0.18 (p = 2.44E-9) and reduced for the ε2 allele at 0.05 (p = 0.00018). The odds ratio (OR) at a 95% confidence interval (CI) for rs429358 was highest at 1.64 (1.40, 1.93) while rs7412 was 0.63 (0.50, 0.80). (**Table 3.2**)

Affymetrix gene chip data was used to genotype rs4420638. The same caliber of cases and controls were used (cases n=926, controls n=2063). This SNP had a minor allele frequency of 0.15 in controls and 0.22 in cases (p = 5.52E-9). The OR for

Trait	Case (n = 1083)	Control (n = 2145)
Age (y)	48.2 ± 7.2	75.0 ± 5.2
Gender - male	82.0%	51.6%
Non-smoker (ever)	31.6%	52.7%
Former smoker	56.8%	45.0%
Current smoker	11.5%	2.3%
Hypertension - yes	44.2%	36.4%
Average BMI (kg/m^2)	28.8 ± 4.6	26.1 ± 4.0
Triglycerides (mmol/L)	2.6 ± 2.2	1.4 ± 0.9
Total cholesterol (mmol/L)	6.1 ± 1.2	5.7 ± 1.0
LDL (mmol/L)	3.9 ± 1.0	3.6 ± 0.9
HDL (mmol/L)	1.1 ± 0.4	1.5 ± 0.4

Table 3.1: Physical characteristics of CAD cases and healthy controls from the Ottawa Heart Study. Mean ± SD.

SNP	Chromosomal Position	Sample Size	AA	AB	BB	MAF	Allelic p-value	Odds Ratio (95% CI)
rs4420638	**45422946**	**cases (n = 926)**	**37**	**325**	**564**	**0.22**	**5.52e-9****	***1.52 (1.32,1.75)***
		controls (n = 2063)	**51**	**531**	**1481**	**0.15**		
rs429358 ε4	**45411941**	**cases (n = 1017)**	**29**	**312**	**676**	**0.18**	**2.44e-9****	***1.64 (1.40,1.93)***
		controls (n = 1361)	**23**	**279**	**1059**	**0.12**		
rs7412 ε2	**45412079**	**cases (n = 1018)**	**1**	**104**	**913**	**0.05**	**0.00018***	***0.63 (0.50,0.80)***
		controls (n = 1337)	**13**	**189**	**1135**	**0.08**		
rs769449 intronic	45410002	cases (n = 456)	9	112	335	0.14	0.007	*1.37 (1.09,1.71)*
		controls (n = 1154)	17	215	922	0.11		
rs769450 intronic	45410444	cases (n = 415)	60	197	158	0.38	0.15	0.89 (0.76,1.04)
		controls (n = 1110)	206	501	403	0.41		
rs445925	45415640	cases (n = 481)	0	80	401	0.08	0.05	0.77 (0.59,1.00)
		controls (n = 1152)	15	213	924	0.10		
rs483082 enhancer ME-1	45416178	cases (n = 482)	25	182	275	0.24	0.08	1.18 (0.98,1.41)
		controls (n = 1043)	46	351	646	0.21		
rs35136575 enhancer HCR-2	45439163	cases (n = 488)	51	185	252	0.29	0.014	*1.23 (1.04,1.44)*
		controls (n = 1210)	80	450	680	0.25		

Table 3.2: Genotype frequencies of individuals genotyped in the Ottawa Heart Study cases and controls.

Statistical difference determined for minor allele frequencies (MAF) of each SNP in cases relative to controls using a simple univariate linear regression model. Allelic p-values were not corrected for physical or biological traits. Odds ratios determined at a 95% confidence interval (CI). Statistically significant SNPs, as determined by allelic p-values, in bold. Bonferroni significance at p-value ≤ 0.006. SNPs with significant ORs as determined by 95% CI written in italics. A = minor allele; B = major allele.

rs4420638 was 1.52 (1.32, 1.75) calculated at a 95% confidence interval. (**Table 3.2**)

The possibility exists that rs4420638 may serve as a flag for more than the ε4 SNP it is in linkage with. In order to probe this, additional SNPs associated with CAD and/or lipids within the *APOE/C1/C4* region were genotyped to determine if linkage exists with either rs4420638 or ε4. Additional SNPs within this region have been identified as being associated with lipids as summarized in the Introduction in **Table 1.1.2** and depicted in **Figure 3.1**. These five SNPs have not been genotypically characterized in any other population. One of particular interest, rs35136575, lies within a downstream hepatic control region (HCR-2) which regulates transcription of all three apolipoproteins within this cohort [79,80]. ApoE intronic SNPs rs769449 and rs769450 and rs445925 and rs483082, both 5' to *APOC1*, have also shown strong association with lipids. Additionally, rs483082 lies within a multiple enhancer site 1 (ME1) which controls apoE expression in the central nervous system. Since apoE is a major apolipoprotein in the central nervous system, and apoE4 is a strong risk factor for Alzheimer's disease, altered apoE expression may play role in the onset of this disease [81]. Other than rs4420638, no additional apoE SNPs are present on the Affymetix gene chip. Thus, all additional SNPs were genotyped by sequencing in the same cases and controls.

Of the five, two SNPs had lower MAF in cases relative to controls, however neither reached strong statistical significance; rs769450 has a 0.38 MAF in cases and 0.41 MAF in controls (p = 0.15) with an OR of 0.89 (0.76, 1.04) and rs445925 is

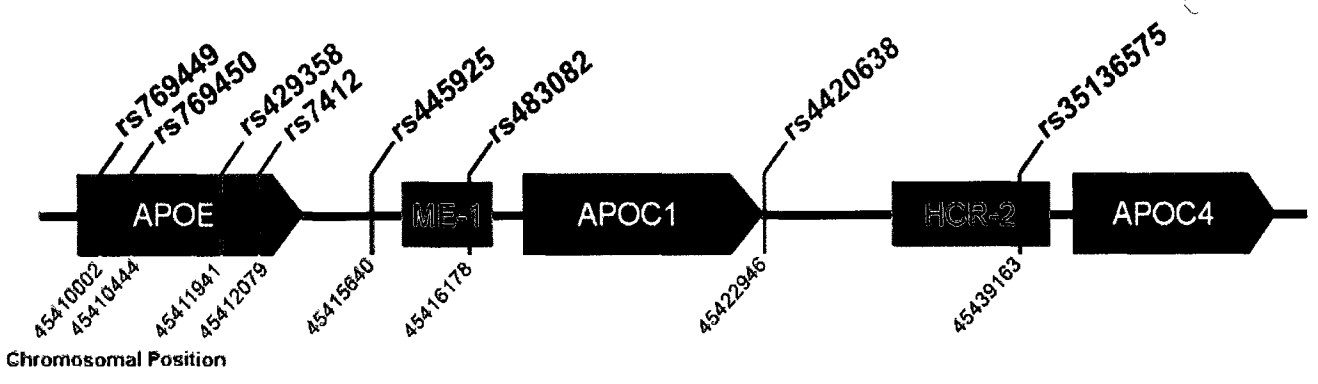

rs769449
rs769450
rs429358
rs7412
rs445925
rs483082
rs4420638
rs35136575
APOE
ME-1
APOC1
HCR-2
APOC4
45410002
45410444
45411941
45412079
45415640
45416178
45422946
45439163
Chromosomal Position

Figure 3.1: Schematic map of APOE-APOC1/C4 region 19q13.2 identifying relative positions of all SNPs genotyped.

Orange: *APOE* SNPs rs429358 and rs7412. Red: Affymetrix gene chip SNP rs4420638. Green: additional SNPs associated with lipids. Four PCR amplicons were designed for genotyping by sequencing; 1) including rs429358 and rs7412; 2) rs35136575; 3) rs445925 and rs483082; 4) rs769449 and rs769450. Blue arrows: transcriptional orientation of apolipoprotein genes. Purple boxes: enhancer regions. ME-1 enhancer region controls apolipoprotein expression in the central nervous system and could play a role in Alzheimer's disease. HCR-2 enhancer region controls expression in hepatocytes and could affect CAD physiology. 29kb between rs769449 and rs3513657.

0.08 MAF in cases and 0.10 MAF in controls (p = 0.05) and an OR of 0.77 (0.59, 1.00). The remaining three SNPs had increased MAF in cases. Intronic apoE SNP rs769449 had an odds ratio of 1.37 (1.09, 1.71) and a MAF of 0.14 in cases and 0.11 in controls but did not reach statistical significance (p = 0.006). Both SNPs in enhancer regions had increased MAF in cases; rs483082 had a MAF of 0.24 in cases and 0.21 in controls (p = 0.08) with and OR of 1.18 (0.98, 1.41). SNP rs35136575 had an OR of 1.23 (1.04, 1.44) with a MAF of 0.29 in cases and 0.25 in controls (p = 0.01). Only two of the five additional SNPs, rs769449 and rs35156575, had a significant OR value at 95% CI. (**Table 3.2**)

Of the apoE SNPs genotyped, those with statistically significant MAF differences between cases and controls were rs4420638, rs429358 and rs7412. The minor allele frequencies and odds ratios of all SNPs are summarized in **Table 3.2.**

3.2 - Linkage Disequilibrium analysis for rs429358 and rs4420638

The linkage disequilibrium (LD) between all SNPs was assessed by calculating r^2 and D' values using R Program 2.10.1 and Haploview 4.2 software. Linkage analysis of all SNPs with rs4420638 and rs429358 are summarized in **Table 3.3**. Of the seven SNPs, rs4420638, rs429358, rs7412 and both *APOE* intronic SNPs had high D' values of over 0.9. All additional SNPs showed weak linkage as assessed by D' and r^2. The highest r^2 linkage is between rs4420638 and rs429358 at 0.69. ApoE rs429358 and intronic *APOE* rs769449 SNPs also have mild linkage (r^2=0.67). No other SNPs showed significantly strong linkage to rs429358 or

SNP	rs4420638		rs429358 ε4	
	r^2	D'	r^2	D'
rs4420638	--	--	0.69	**0.97**
rs429358 ε4	0.69	**0.97**	--	--
rs7412 ε2	0.02	**1.00**	0.01	**1.00**
rs769449	0.47	0.88	0.67	0.90
rs769450	0.02	0.37	0.09	0.91
rs445925	0.00	0.09	0.00	0.03
rs483082	0.30	0.63	0.41	0.87
rs3513657	0.00	0.13	0.00	0.06

Table 3.3: Linkage disequilibrium r2 and D' values of each SNP genotyped relative to rs4420638 and rs429358 ε4.

The highest r^2 was found between rs4420638 and rs429358 at 0.69. Only apoE SNPs and rs4420638 had perfect D' linkage (in bold). Control samples used to determine linkage. (See **Figure 3.2** for D' linkage plot of all SNPs)

rs4420638. The lowest linkage was observed with enhancer SNP rs35136575 with both rs4420638 and rs429358 having r^2 and D' values less than 0.10. A linkage disequilibrium plot of all eight SNPs can be seen in **Figure 3.2**. This plot assesses the LD association between all eight SNPs using control genotypes.

3.3 - Haplotype analysis of ApoE SNPs and rs4420638 and Likelihood Ratio Test

A haplotype analysis was conducted on the three linked SNPs rs429358, rs7412 and rs4420638. This was to assess which paired SNP showed the strongest association with CAD. Analysis of SNPs rs4420638 and rs429358 indicated that the risk allele of rs429358 contributed to disease more strongly than rs4420638. The frequency of individuals bearing risk alleles for both SNPs was 15% for cases and 11% for controls. Frequency of individuals carrying rs429358 risk and rs4420638 reference alleles was 3% in cases and 0.3% in controls – a 10-fold difference between cases and controls. For rs4420638, 6% of cases and 4% of controls carry the rs4420638 risk allele alone – a 1.5-fold difference. Thus, since a much higher population of rs429358 risk carriers alone is present in cases, rs4293598 confers a greater CAD risk than does the risk allele of rs4420638 (see **Table 3.4a**).

Compared to the "protective" rs7412 SNP, rs4420638 risk allele had an increased frequency in cases than controls. In cases, 21% and in controls, 16% had risk alleles at both rs4420638 and rs7412 SNPs. In the case of rs7412 the "risk" allele is the major allele while the "reference" allele is the minor allele due to the protective nature

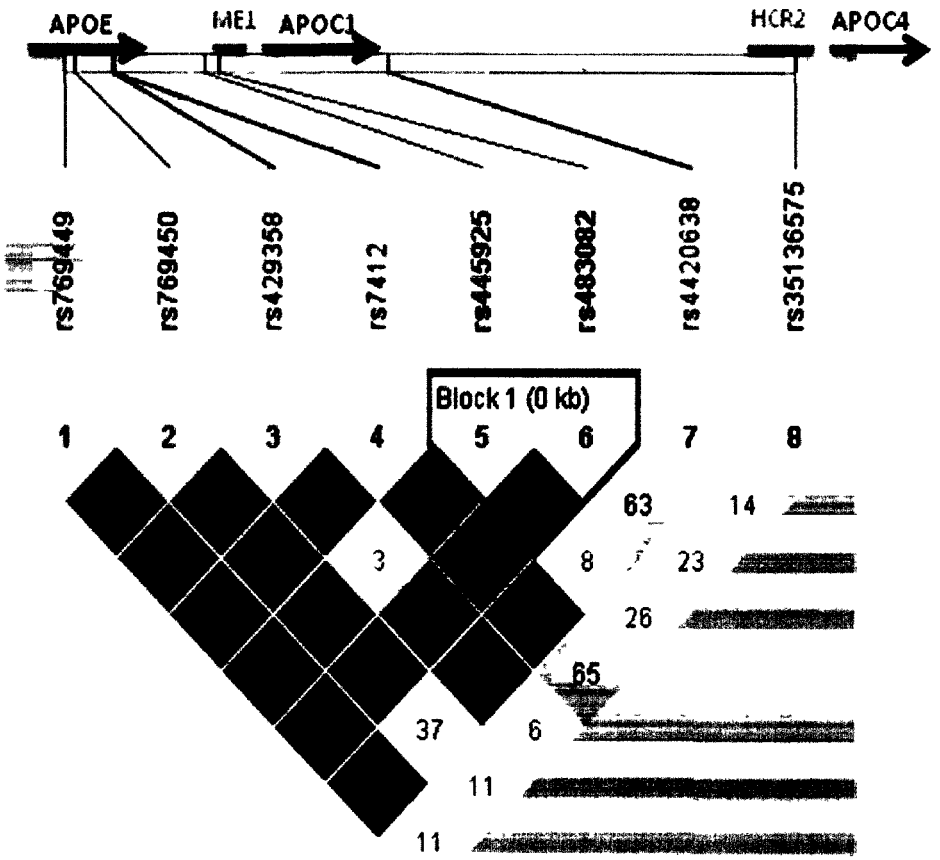

APOE
APOC1
HCR2
APOC4
rs769449
rs769450
rs429358
rs7412
rs445925
rs483082
rs4420638
rs35136575
Block 1 (0 kb)
1
2
3
4
5
6
7
8
63
14
3
8
23
26
65
37
6
11
11

Figure 3.2: Linkage Disequilibrium Plot calculated from control samples (n=2145) of all SNPs genotyped using Haploview 4.2 software.

Colour ranges from red (perfect linkage) to white (no linkage). Numbers in boxes represent D' values as a percentage ranging from 0 to 100. Red boxes with no numbers indicate perfect linkage (D' = 100). SNPs in green are rs429358 ε4, rs7412 ε2 and rs4420638. Black arrows represent transcriptional orientation of genes and blue bars are enhancer regions. Due to the tight *LD* between rs442925 and rs483082 Haploview 4.2 predicts these SNPs are in a haplotype block.

a)

rs4420638	rs429358 ε4	% in cases	% in controls
A (risk)	A (risk)	15	11
A	B	**6**	**4**
B	A	**3**	**0.3**
B	B	76	84

b)

rs4420638	rs7412 ε2	% in cases	% in controls
A (risk)	A	0.7	0
A	B (risk)	21	16
B	A	5	8
B	B	74	77

c)

rs429358 ε4	rs7412 ε2	% in cases	% in controls
A (risk)	A	0	0
A	B (risk)	18	12
B	A	5	8
B	B	77	80

Table 3.4: Haplotype analysis using estimated 2-loci haplotype frequency prediction analysis.

Haplotypes analyzed include (a) rs4420638 and rs429358; (b) rs4420638 and rs7412 and; (c) rs429358 and rs7412. Cases=859; controls=1244. Only samples with all three SNPs genotyped were analyzed. Relative percentage of individuals carrying risk/reference alleles of two SNPs in cases and controls are compared. Values that determine the functional difference between rs4420638 and rs429358 are in bold. There are 10-fold more cases than controls carrying rs429358 risk allele alone, while a 1.5-fold difference exists between cases and controls with only rs4420638 risk allele. Note that since rs7412 is a "protective" allele, the major allele (B) confers "risk."

of this SNP. Thus, there is a small percentage of cases (5%) and controls (8%) that carry both reference alleles at rs4420638 and rs7412 simply because rs7412 reference (minor) allele frequency is low (see Table 4). A lower percentage of cases, 74%, relative to controls, 77%, had risk alleles at rs7412 only. (**Table 3.4b**)

A similar pattern in seen in the haplotype analysis between rs7412 and rs429358. A higher percentage of cases (18%) have risk alleles at both SNPs than controls (12%). Similarly, 80% of controls and 77% of cases have the risk allele at rs7412 only. This may imply that the rs7412 risk allele alone does not contribute to CAD development as readily as risk alleles of rs4420638 or rs429358. Also, a small percentage of cases (5%) and controls (8%) carry reference alleles at both rs7412 and rs429358. Although the frequencies are smaller, there are still a higher percentage of controls than cases. (**Table 3.4c**)

Additionally, the likelihood ratio test reinforced this trend. There is no significant difference between the one variable CAD risk model using only rs429358 data and the two variable model using rs429358 and rs4420638. However, adding rs429358 data to a rs4420638 CAD risk model provides a significantly better assessment of CAD risk (**Table 3.5**).

3.4 – Only apoE, not apoC1, serum protein concentration is genotypically associated

Enzyme-linked immunosorbant assay (ELISA) was used to determine serum apoE and apoC1 protein concentration of genotyped controls. Serum apoE, but not

One Variable Model	Two Variable Model	p-value
rs429358	rs4420638 + rs429358	0.7
rs4420638	rs4420638 + rs429358	0.0005*

Table 3.5: Likelihood Ratio Test of rs4420638 and rs429358 CAD risk models.

There is a statistically significant difference between the one variable rs4420638 CAD risk model and the two variable model including rs429358 data (p=0.0005). When rs4420638 data is added to the single variable rs429358 CAD risk model, there is no significant difference in the fit of the model (p=0.7). Likelihood ratio test is based on samples with all three SNPs (rs4420638, rs429358 and rs7412) genotyped: 859 cases, 1244 controls.

apoC1 expression was associated with rs429358 and rs4420638. Minor alleles for both SNPs showed lower apoE expression while the homozygous major allele had more than double the ApoE expressed. For rs4420638 AA: 25.5 ± 4.8, AB: 44.8 ± 3.1, BB: 69.9 ± 3.7 ug/ml (mean ± SEM; n = 215; p = 7.27E-7). For rs429358 AA: 14.8 ± 3.4, AB: 40.1 ± 3.1, BB: 67.0 ± 3.4 ug/ml (mean ± SEM; n = 215; p = 1.73E-8). The opposite trend was seen for rs7412; the homozygous minor allele had nearly double the apoE serum concentration than the homozygous major allele; AA: 87.2 ± 9.7, AB: 71.5 ± 8.7, BB: 53.8 ± 2.7 (mean ± SEM; n = 215; p = 0.002). rs35136575 also associates with apoE serum levels similarly to rs429358 and rs4420638 (AA: 31.0 ± 4.6, AB: 52.5 ± 4.1, BB: 61.3 ± 3.7 ng/ml, mean ± SEM, p=0.005). ApoE and ApoC1 ELISA data are summarized in **Table 3.6**.

Serum protein apoE and apoC1 was also analyzed for the six apoE genotypes E2/E2, E2/E3, E2/E4, E3/E4 and E4/E4. Only apoE concentrations were associated with the genotypes while apoC1 did not correlate. ApoE concentration decreased from homozygous E2 to homozygous E4; E2/E2: 87.4 ± 8.2; E2/E3: 74.0 ± 9.0; E2/E4: 35.4 ± 2.6; E3/E3: 63.8 ± 3.6; E3/E4: 40.0 ± 3.2; E4/E4: 14.5 ± 3.0 ug/ml (mean ± SEM; n = 221; p = 1.11E-7). ApoC1 concentrations were relatively similar across genotypes; E2/E2: 132 ± 16.9; E2/E3: 124 ± 7.2; E2/E4: 120 ± 31.4; E3/E3: 141 ± 5.5; E3/E4: 141 ± 9.4; E4/E4: 140 ± 20.1 ug/ml (mean ± SEM; n = 221; p = 0.198). **Table 3.7** summarizes the *APOE* genotype correlation with apoE and apoC1 concentrations.

SNP	Genotype	ApoE (ug/ml)	p-value	ApoC1 (ug/ml)	p-value
rs4420638	AA (n=13)	25.5 ± 4.8	7.27E-07**	145.0 ± 14.4	0.22
	AB (n=66)	44.8 ± 3.1		145.0 ± 8.5	
	BB (n=136)	69.9 ± 3.7		134.0 ± 4.7	
rs429358 ε4	AA (n=8)	14.8 ± 3.4	1.73E-08**	142.0 ± 22.7	0.68
	AB (n=59)	40.1 ± 3.1		140.0 ± 9.2	
	BB (n=148)	67.0 ± 3.4		136.0 ± 4.5	
rs7412 ε2	AA (n=6)	87.2 ± 9.7	0.002*	143.0 ± 14.7	0.30
	AB (n=35)	71.5 ± 8.7		123.0 ± 7.1	
	BB (n=174)	53.8 ± 2.7		140.0 ± 4.7	

Table 3.6: Serum Apoliporotein E and C1 protein concentration in control samples as determined by ELISA method.

ApoE serum protein concentration is associated with rs4420638, rs35136575 and apoE SNP genotypes while apoC1 concentration shows no genotypic association. Simple linear regression model was used to determine p-values. Bonferroni corrected significance at p-value ≤ 0.006. Protein concentration calculated as mean ± SEM. A = minor allele; B = major allele.

ApoE Genotype	n	ApoE (ug/ml)	ApoE p-value	ApoC1 (ug/ml)	ApoC1 p-value
E2/E2	7	87.4 ± 8.2	1.11E-07**	132.0 ± 16.9	0.198
E2/E3	33	74.0 ± 9.0		124.0 ± 7.2	
E2/E4	3	35.4 ± 2.6		120.0 ± 31.4	
E3/E3	112	63.8 ± 3.6		141.0 ± 5.5	
E3/E4	57	40.0 ± 3.2		141.0 ± 9.4	
E4/E4	9	14.5 ± 3.0		140.0 ± 20.1	

Table 3.7: Association of serum Apoliporotein E and C1 concentration in control samples as determined by ELISA, with APOE genotype.

ApoE serum protein concentration is associated with *APOE* genotype while apoC1 concentration shows no association. Simple linear regression model used to determine p-values. Bonferroni corrected significance at $p \leq 0.025$. Protein concentration measured as mean ± SEM.

3.5 – Plasma LDL concentration is associated with rs4420638 and apoE SNPs

Plasma lipids were found to be associated with both apoE SNPs and rs4420638. Specifically LDL was associated with rs4420638, rs429358 and rs7412. rs35136575 did not reach statistical significance. For both rs4420638 and rs429358 LDL concentration decreased from homozygous minor allele to homozygous major allele; rs4420638 AA: 3.96 ± 0.19; AB: 3.74 ± 0.05; BB: 3.58 ± 0.03 mmol/L (mean ± SEM, n = 1210, p = 0.0007) rs429358 AA: 3.81 ± 0.23; AB: 3.78 ± 0.05; BB: 3.59 ± 0.03 mmol/L (mean ± SEM, n = 1210, p = 0.003). The opposite trend was seen for rs7412 where the homozygous minor allele had lower LDL than homozygous major alleles; AA: 3.29 ± 0.50; AB: 3.07 ± 0.06; BB: 3.73 ± 0.03 mmol/L (mean ± SEM, n = 1210, p = 1.78E-18). Triglycerides (TG) did not show a statistically significant trend with genotype. **Table 3.8** summarizes the lipid association data for the two apoE SNPs and rs4420638 for LDL and TG.

Plasma lipids were also analyzed with the six apoE genotypes E2/E2, E2/E3, E2/E4, E3/E4, and E4/E4. **Table 3.9a, b, c** and **d** summarize the association of the apoE genotypes with LDL, HDL, TG and total cholesterol. Association was seen between LDL (p = 3.67E-21) and total cholesterol (p = 1.03E-15), however HDL (p = 0.56), and TG (p = 0.44) were not statistically significant. Of the six genotypes, E2/E2 had the lowest LDL concentration with an increasing trend in concentration to homozygous E4; E2/E2: 3.29 ± 0.46; E2/E3: 3.03 ± 0.06; E2/E4: 3.75 ± 0.18; E3/E3: 3.72 ± 0.03; E3/E4: 3.82 ± 0.06 and E4/E4: 3.96 ± 0.97 mmol/L (mean ± SEM, n = 1290, p = 3.67E-21).

SNP	Genotype	LDL (mmol/L)	p-value	TG (mmol/L)	p-value (Additive)	p-value (Recessive)
rs4420638	AA (n=29)	3.96 ± 0.19	0.0007*	1.70 ± 0.19	0.01	0.002*
	AB (n=317)	3.74 ± 0.05		1.35 ± 0.04		
	BB (n=864)	3.58 ± 0.03		1.30 ± 0.02		
rs429358 ε4	AA (n=18)	3.81 ± 0.23	0.003*	1.66 ± 0.21	0.35	0.04
	AB (n=246)	3.78 ± 0.05		1.31 ± 0.04		
	BB (n=946)	3.59 ± 0.03		1.32 ± 0.02		
rs7412 ε2	AA (n=12)	3.29 ± 0.50	1.78e-18**	2.03 ± 0.35	0.07	0.0002*
	AB (n=171)	3.07 ± 0.06		1.32 ± 0.05		
	BB (n=1027)	3.73 ± 0.03		1.31 ± 0.02		

Table 3.8: Plasma concentration of fasting LDL and triglycerides (TG) in controls.

LDL and TG were measured at time of blood collection by technicians in the Atherogenomics Laboratory. Only LDL is correlated with rs4420638 and apoE SNP genotypes. Simple linear regression model used to determine additive p-values. Recessive test was also used for TG measurement to determine homozygous minor allele affect. Bonferroni corrected significance at $p \leq 0.006$. Lipids measured as mean ± SEM. A=minor allele; B=major allele.

a) LDL

ApoE Genotype	n	LDL (mmol/L)	p-value
E2/E2	13	3.29 ± 0.46	3.67e-21**
E2/E3	166	3.03 ± 0.06	
E2/E4	18	3.75 ± 0.18	
E3/E3	824	3.72 ± 0.03	
E3/E4	247	3.82 ± 0.06	
E4/E4	22	3.96 ± 0.21	

b)HDL

ApoE Genotype	n	HDL (mmol/L)	p-value
E2/E2	13	1.36 ± 0.12	0.48
E2/E3	166	1.45 ± 0.04	
E2/E4	18	1.42 ± 0.08	
E3/E3	824	1.48 ± 0.02	
E3/E4	247	1.48 ± 0.03	
E4/E4	22	1.37 ± 0.07	

c)TG

ApoE Genotype	n	TG (mmol/L)	p-value
E2/E2	13	2.08 ± 0.33	0.57
E2/E3	166	1.33 ± 0.05	
E2/E4	18	1.30 ± 0.13	
E3/E3	824	1.31 ± 0.02	
E3/E4	247	1.34 ± 0.05	
E4/E4	22	1.63 ± 0.17	

d) TC

ApoE Genotype	n	TC (mmol/L)	p-value
E2/E2	13	5.59 ± 0.51	2.04E-16**
E2/E3	166	5.09 ± 0.08	
E2/E4	18	5.76 ± 0.17	
E3/E3	824	5.79 ± 0.03	
E3/E4	247	5.91 ± 0.07	
E4/E4	22	6.07 ± 0.20	

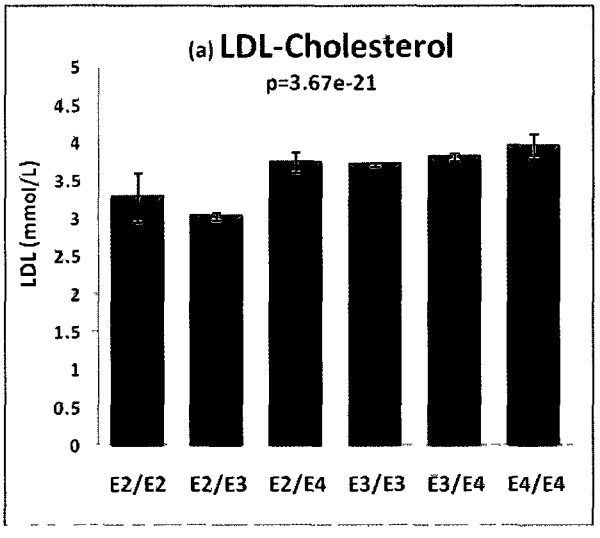
Table 3.9: Concentration of fasting (a) LDL, (b) HDL, (c) triglycerides (TG), and (d) total cholesterol (TC) relative to the six apoE genotypes using control samples.

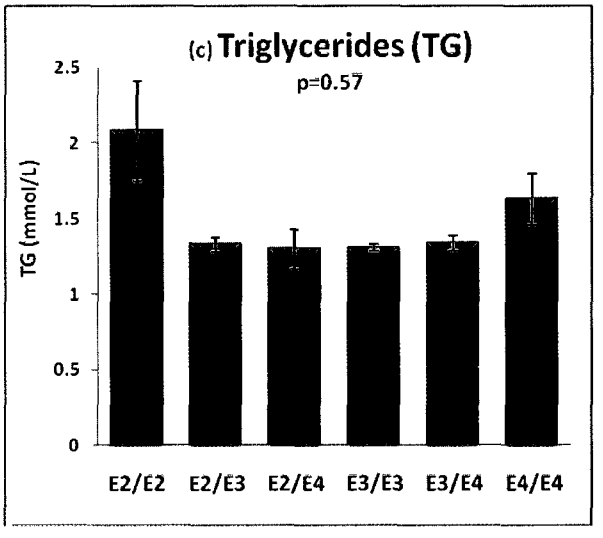
All lipid concentrations were determined at time of blood collection by technicians in the Atherogenomics Laboratory. Strong LDL and TC correlation exists with apoE genotypes. TG and HDL show no correlation. Simple linear regression model used to determine p-values. Bonferroni corrected significance at $p \leq 0.0125$. (See **Figure 3.3** for graphical representation of data).

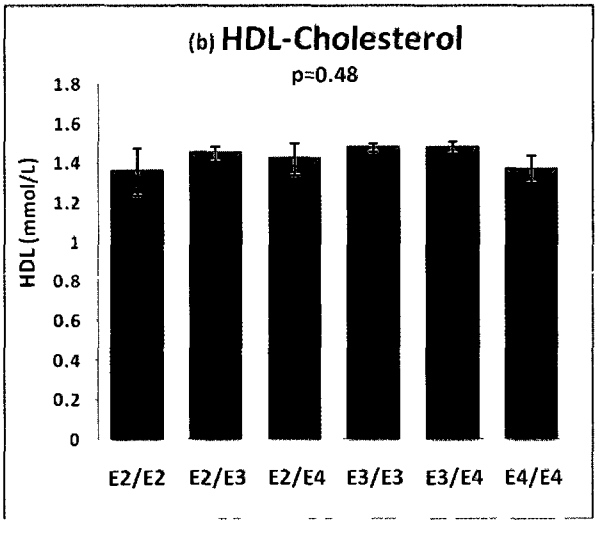
Total cholesterol showed a similar trend: E2/E2: 5.59 ± 0.51; E2/E3: 5.13 ± 0.08; E2/E4: 5.76 ± 0.17; E3/E3: 5.81 ± 0.03; E3/E4: 5.95 ± 0.07; E4/E4 6.07 ± 0.20 mmol/L (mean ± SEM, n = 1290, p = 2.04E-16). Bar graphs depicting each trend are seen in **Figure 3.3a, b, c** and **d**.

3.6 - Odds ratios of the six apoE genotypes in the Ottawa Heart population

Odds ratios for the six apoE genotypes were determined as seen in **Figure 3.4**. Relative to E3/E3 which was given an OR of 1.00, E2/E2 had the lowest OR at 0.11 (0.01, 0.84) followed by E2/E3 at 0.69 (0.52, 0.91). E2/E4 had a higher OR than E3/E3 at 1.83 (0.98, 3.42), but with high variance. E3/E4 had a OR of 1.63 (1.33, 1.99) followed by E4/E4 with 1.80 (1.03, 3.15) OR. All odds ratios were determined with 95% confidence intervals.


(a) LDL-Cholesterol
p=3.67e-21
LDL (mmol/L)
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
E2/E2
E2/E3
E2/E4
E3/E3
E3/E4
E4/E4


(c) Triglycerides (TG)
p=0.57
TG (mmol/L)
2.5
2
1.5
1
0.5
0
E2/E2
E2/E3
E2/E4
E3/E3
E3/E4
E4/E4


(b) HDL-Cholesterol
p=0.48
HDL (mmol/L)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
E2/E2
E2/E3
E2/E4
E3/E3
E3/E4
E4/E4

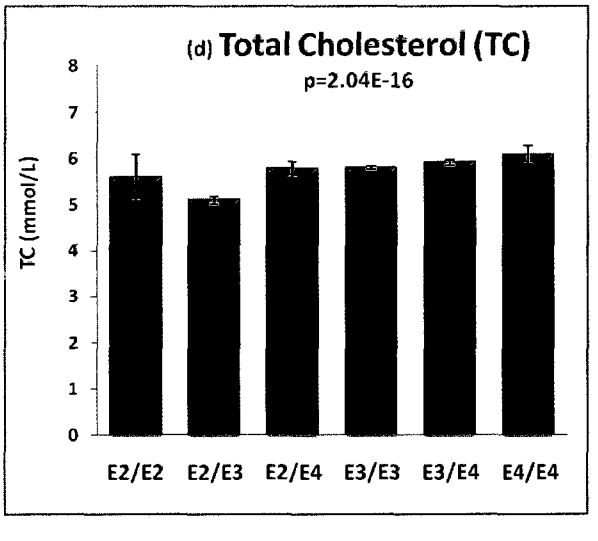

(d) Total Cholesterol (TC)
p=2.04E-16
TC (mmol/L)
8
7
6
5
4
3
2
1
0
E2/E2
E2/E3
E2/E4
E3/E3
E3/E4
E4/E4

Figure 3.3: Bar graphs of concentration of fasting (a) LDL-cholesterol, (b) HDL-cholesterol, (c) triglycerides (TG), and (d) total cholesterol (TC) relative to the six apoE genotypes from control samples.

All lipid concentrations were determined at time of blood collection. Strong LDL and TC correlation exists with apoE genotypes. TG and HDL show no correlation. However, TG concentrations are increased for E2/E2 and E4/E4 genotypes relative to E3/E3. (see **Table 3.9** for numerical value)

Genotype	n cases	n controls	OR	95% CI
E2/E2	1	13	0.11	(0.01,0.84)
E2/E3	81	169	0.69	(0.52,0.91)
E2/E4	23	18	1.83	(0.98,3.42)
E3/E3	592	847	1.00	-
E3/E4	289	254	1.63	(1.33,1.99)
E4/E4	29	23	1.80	(1.03,3.15)

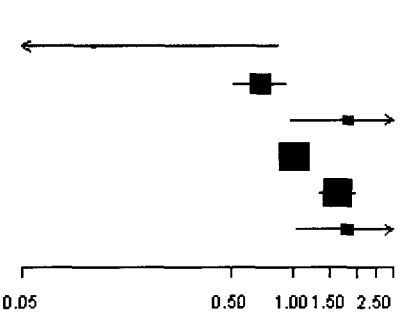


Figure 3.4: Odds ratios of six apoE genotypes (E2/E2, E2/E3, E2/E4, E3/E4 and E4/E4).

Values are relative to wild-type E3/E3 genotype which is given a relative odds ratio of 1.00. Box size is inversely proportional to sample variance. E2 carriers are protective for CAD, as indicated by OR <1.00 while E4 carriers have increased CAD risk as indicated by OR>1.00.

4. DISCUSSION

The *APOE/C1/C4* gene cluster on chromosome 19q13.2 has been identified as being a major locus for CAD risk. Numerous genome-wide association studies (GWAS) have identified a SNP within this region to be associated with CAD risk and plasma lipid traits [50,82,49,64]. The one SNP within this chromosomal region that is present on Affymetrix and Illumina high density gene chip arrays used in most GWA studies, is rs4420638. This SNP lies 5' to *APOC1* and is in mild linkage disequilibrium (LD) with *APOE* SNPs rs429358 and rs7412 ($r^2 = 0.70$).

The functionality of two *APOE* SNPs has been elucidated. Both rs429358 and rs7412 are non-synonymous SNPs which determine the three common apoE isoforms, apoE2, apoE3 and apoE4. Relative to E3, E2 carriers have lower LDL-C and total blood cholesterol, and are protected against CAD with an OR of 0.80 (0.70-0.90) as reported in a recent meta-analysis [65,66]. E4 carriers have higher LDL-C and total blood cholesterol, and an OR for CAD of 1.06 (0.99-1.13) [65,66]. The amino acid change conferred by each of rs7412 and rs429358 leads to a functional modification in either the apoE lipid-binding region or the receptor-binding region. ApoE2 has a 1% LDL-receptor-binding ability relative to apoE3 while apoE4 readily associates with large lipoproteins VLDL and IDL, rather than HDL [66]. This in turn, not only, affects the endocytic clearance of HDL-C, but also, apoE4 bound VLDL readily converts to LDL thus increasing LDL-C levels. Due to impaired receptor binding ability, ApoE2-bound-

VLDL cannot convert to LDL as readily as apoE4 and apoE3 containing VLDL. Also, apoE2 preferentially binds HDL particles [66].

In a recent meta-analysis of GWAS for CAD, rs4420638 was found to have an OR of 1.17 (1.08-1.28) for CAD, seemingly higher than that reported for apoE4 [50]. Although mild linkage exists between rs4420638 and rs429358 (r^2 = 0.70), the increased OR, differing MAFs (0.18 for rs4420638 versus 0.15 for rs429358) and stronger association of rs4420638 with blood lipid levels, leads us to hypothesize that the effect of rs4420638 on CAD risk is partially independent of linkage with rs429358.

The association of the *APOE/C1/C4* region to CAD risk could be due to more than the apoE SNPs resulting in three apoE isoforms. Studies have shown that altered expression of the apoliporoteins in this region alters lipid homeostasis. Overexpression of apoC1 was recently shown to inhibit HDL binding to hepatic SR-B1 [32]. This hinders HDL clearance. Additionally, overexpression of apoC1 stimulates LCAT which is responsible for HDL cholesterol esterification [34]. Increased apoC1 levels can also impair HL and CETP activity, thus, leading to larger and more numerous HDL particles [33]. Overexpression of apoE – irrespective of isoform – leads to impaired VLDL clearance since apoE competes out LPL cofactor apoCII and VLDL cannot be lipolysed as efficiently [24]. This naturally occurs in a fasting-state to increase lipoprotein blood levels [25]. Thus, altered expression levels of both apoE and apoC1 could be implicated in CAD risk.

For these reasons, further investigation of the *APOE/C1/C4* region may uncover additional molecular mechanisms contributing to increased CAD risk as flagged by rs4420638.

To address this, rs4420638, rs429358 and rs7412 were genotyped in the Ottawa Heart Study (OHS) population. Additionally, five SNPs within this region that were associated with LDL or HDL were also genotyped in the same cases and controls to determine linkage disequilibrium patterns. The five SNPs chosen had not previously been genotyped in any other large CAD case-control population.

We found that three of these five additional SNPs had an increased MAF in cases while the remaining two had a higher frequency in controls. However, none had statistically significant allelic p-values. It is possible that these SNPs may also play a role in CAD risk, however their association is not nearly as strong as the SNPs already being investigated in this region. The three original SNPs rs429358, rs7412 and rs4420638 were strongly statistically significant, and showed altered MAF in cases versus controls. As expected, both risk SNPs had a higher frequency in cases while rs7412 had a higher frequency in controls; rs4420638 was 0.22 in cases and 0.15 in controls (p=4.41E-8) and rs429358 was 0.18 in cases and 0.12 in controls (p=1.96E-8) while rs7412 had a 0.05 MAF in cases and 0.08 MAF in controls (p=0.0014). (See **Table 3.4**)

The five additional SNPs were chosen because of their association with blood lipid levels specifically HDL-C and LDL-C (see **Table 1.2**). They were genotyped in the OHS population to determine linkage to rs429358, rs7412 or rs4420638 SNPs. If

strong linkage were to exist between any of these SNPs and rs4420638, the functionality of rs4420638 could be further explored with the linked SNP. However, little or no linkage was seen between the additional five SNPs and the three original apoE SNPs, rs429358, rs7412 and rs4420638. The highest linkage existed between rs429358 and rs4420638 at an r^2 of 0.69 and D' of 0.97. All three of the original SNPs had D' values near 1.0 which indicates that they are inherited as a haplotype and may be functionally associated.

The next highest linkage was between rs769449 and rs429358 with an r^2 of 0.67 and D' of 0.90. Also, the neighboring rs769450 SNP had a D' of 0.91 with rs429358. This linkage is not unexpected since both rs769 SNPs and rs429358 are within the *APOE* gene. However, relative to rs4420638, the rs769 SNPs are likely having a smaller effect on apoE CAD risk since the r^2 between rs4420638 and rs429358 is higher at 0.69. Thus, since there exists a stronger linked SNP in this region (ie rs4420638), it is likely that the linkage between the rs769 SNPs and rs429358 is negligible in this case. Essentially, the five additional SNPs genotyped do not provide further information regarding rs4420638 CAD association. Thus, it can be concluded that the functionality of rs429358 and rs4420638 is not due to linkage to other SNPs in this locus (specifically those that have been genotyped in this study).

However, it can be noted that although, HCR-2 enhancer SNP rs35136575 has the lowest linkage with rs4420638 and rs429358, its OR is significant in a univariate analysis – that is, at a 95% CI, the SNP has an increased OR for CAD risk: OR = 1.23 (1.04, 1.44). Since no linkage exists with rs4420638 or rs429358, this SNP could confer

an independent effect on CAD risk. Future analysis of this enhancer region would be of interest to determine if rs35136575 alters apolipoprotein expression in hepatocytes, leading to increased CAD risk.

Because no other SNPs showed high LD with rs4420638 and the three main SNPs have a D' near 1.0 indicating a perfectly linked haplotype, the three SNPs, rs429358, rs7412 and rs4420638 are treated as a haplotype. Haplotype analysis evaluates the frequencies of the alleles of a combination of SNPs in cases versus controls. The ultimate haplotype combination is that of rs4420638 and rs429358 (see **Table 3.7a**). This analysis indicates which risk SNP is having a larger effect on CAD risk independent of the other. As expected, there are a higher proportion of cases than controls that have risk alleles at both SNPs. The frequency of individuals bearing risk alleles at both SNPs is 15% in cases and 11% in controls. When assessing the risk alleles of each SNP independently, there is a larger difference between the percentages of cases than controls with rs429358 risk allele alone than those with only the rs4420638 risk allele. The proportion of cases with only rs429358 risk allele was 3% while 0.3% of controls carry this risk allele alone. This is a 10-fold proportional difference. The proportion of individuals with a risk allele at only rs4420638 was 6% in cases and 4% in controls – a 1.5-fold proportional difference. This is indicative that rs429358 is the 'driving factor' for CAD risk in this region. A much higher population of cases exists when the risk allele of rs429358 is present without the rs4420638 risk allele.

The fact that, by sheer number, there are more cases with rs4420638 risk allele (6%) than rs429358 risk allele (3%) can be explained by their MAFs. Since rs4420638 has a MAF of 0.22 and rs429358 has a MAF of 0.18 in OHS cases, it is not unexpected to find a higher population of cases with the rs4420638 risk allele alone. This also indicates that rs4420638 is, in fact, associated with increased CAD risk, but not to the same degree that rs429358 does. From this haplotype analysis it can be concluded that rs429358 has a stronger effect on CAD risk than rs4420638. It cannot be determined conclusively whether or not the rs429358 effect is entirely independent of rs4420638. However, rs4420638 does not appear to have as strong an effect on CAD risk as rs429358.

The haplotype analyses for rs4420638 and rs429358 with rs7412 were as expected (see **Tables 3.7b, 3.6c**). When analyzing rs4420638 and rs7412, no controls, but 0.7% of cases had the rs4420638 risk allele and rs7412 reference allele. 21% of cases and 16% of controls had risk alleles at both SNPs. This trend is lost when rs4420638 is a reference allele; more controls (77%) than cases (74%) have only the rs7412 risk allele and again both SNPs are reference alleles (8% controls; 5% cases). When rs7412 is paired with rs429358 in a haplotype, a similar trend results. There are an increased number of cases (18%) than controls (12%) when both are risk alleles. When rs429358 is a reference allele – regardless of rs7412 – there is a higher percentage of controls (5% cases, 8% controls when both are reference alleles, and 77% cases, 80% controls when rs7412 is risk allele alone).

This confirms that rs7412 is not enforcing CAD risk, since risk is increased only when either rs4420638 or rs429358 also have the risk allele present. When pairing rs7412 risk with rs4420638 risk allele, the proportion of cases is higher than cases that have rs7412 and rs429358 risk alleles. Again, this can be explained by the differing MAF of the risk SNPs. Since rs4420638 has an increased MAF of 0.22 and rs429358 is 0.18 in the Ottawa Heart cases (**Table 3.4**), it is expected that this allele would be present in more cases than rs429358 when analyzing haplotypes. The increase may not necessarily imply that one SNP is functionally more dominant than the other.

4.1 - MAF of risk SNPs in the OHS differ from other study populations

There is a considerable difference between the MAF in the Ottawa Heart Study cases and controls (**Table 3.4**). In addition to the divergence within our population, the MAF of OHS controls is lower than controls used in other genome studies, including the general population sampled for the HapMap Project. According to HapMap, the MAF of rs4420638 is 0.18 in the general Caucasian population sampled for their study. The MAF of the OHS controls is lower at 0.15. Similarly, rs429358 has a reported MAF of 0.15 and rs7412 0.07 [66] however, here we report a MAF of 0.12 for rs429358 and 0.08 for rs7412 in controls. Many factors can explain this differentiation.

First, the populations differ phenotypically. The controls used in the OHS underwent a strict selection process; individuals that lacked any symptoms and physical signs of atherosclerosis were enrolled. In addition, since atherosclerosis is a late-onset disease, non-symptomatic men over 65 and women over 70 were recruited. This eliminated the possibility of incorporating younger controls that may develop

CAD at a later age. The individuals used for HapMap are a general sample of less than 100 Caucasian individuals of differing ages that may potentially be at any stage of disease. Since the HapMap samples were not phenotypically defined, individuals with some form of atherosclerosis were likely genotyped. In this case, if individuals with atherosclerosis were part of the population, there is a greater likelihood that their genotype would include SNPs associated with increased CAD risk. Thus, being risk SNPs, the frequencies of rs429358 and rs4420638 would increase if individuals with disease are not screened out.

Similarly, in Hatters' review, the population used was not defined and thus, it is likely that individuals with disease were genotyped [66]. The frequency of risk SNP rs429358 would increase in controls that are not phenotypically defined. A CAD "protective" SNP (i.e. rs7412) would have a higher frequency in a sample population that removed CAD patients, hence the differentiation between our controls and Hatters'. In other words, by ensuring our controls did not contain individuals with CAD, we indirectly selected against CAD risk SNPs and selected for CAD "protective" SNPs.

Additional GWA studies, including a number of recent meta-anlayses, show similar trends. A meta-analysis studying gene biomarkers for CAD by Wallace et al in 2008 used individuals with hypertension – a major CAD risk factor – in their study. Their reported MAF of rs4420638 was 0.19. It is highly likely that individuals with hypertension may have atherosclerosis. A GWA study on LDL-C by Sandhu et al, in 2008 reported a rs4420638 MAF of 0.18. Additional studies are compared with the OHS MAFs in **Table 4.1**.

Secondly, the method by which our control volunteers were recruited for the OHS may have inadvertently resulted in bias. Controls were recruited by advertising and consenting healthy, elderly participants actively volunteered to be included in this study. Blood samples for DNA analysis were collected at the Ottawa Heart Institute rather than the subject's home or long-term care facility. As a result, only control subjects that were cognitively sound were able to participate in this study. This is important to note since in addition to atherosclerosis, apoE is also genetically associated with Alzheimer's disease (repeat refs). Individuals showing any signs of Alzheimer's were not able to provide informed consent and were eliminated from the control population. Consequently, the risk SNPs for Alzheimer's were unintentionally selected against. In this case, this would include rs429358 and rs4420638, explaining why the MAF of each of these SNPs is lower than expected in the OHS controls. In contrast, since rs7412 is protective for Alzheimer's, this also explains its increased MAF in the OHS controls (0.08) relative to the general human population analyzed by Hatters (0.07). Thus, not only did our controls lack signs of CAD, but they were also not Alzheimer's patients. This stringency strongly selected against CAD and AD risk SNPs while selecting for "protective" SNPs.

A case-control study in 2006 by Zuo et al analyzed the genetic association of *APOE* in Alzheimer's patients. Their reported MAF of rs429358 was 0.36 in cases and 0.14 in controls while rs7412 had a frequency of 0.02 in cases and 0.06 in controls. Although they recruited healthy controls with no Alzheimer's disease, symptoms of atherosclerosis were not screened, which may explain the differing MAFs of their

controls and the OHS controls. Additional studies are compared with the OHS MAFs in **Table 4.1**.

4.2 - ORs of rs429358, rs7412 and rs4420638 are more extreme than other study populations

The odds ratios for CAD for the risk or protective alleles of the above SNPs in the Ottawa Heart Study population were proportionately higher or lower than reported by previous studies. In our population, the OR for rs7412 was 0.63 (0.50, 0.80) and 1.64 (1.40, 1.93) for rs429358. Previously reported ORs are 0.80 (0.70, 0.90) for rs7412 and 1.06 (0.99, 1.13) for rs429358 [65]. Similarly, rs4420638 had a previously reported OR of 1.17 (1.08-1.28) [50] while here we report 1.52 (1.32, 1.75).

This, again, can be explained by the extreme stringency with which the OHS controls were recruited. Our controls are healthy, older, cognitively able, consenting volunteers. Bennet et al's meta-analysis comprised of 82 studies totaling over 86, 000 "disease-free participants" [65]. The healthy controls were "people without known coronary or other diseases or clinical lipid abnormalities" [65]. However, the age of these participants was not reported. CAD and Alzheimer's are late-onset diseases and young participants that are healthy at the time of recruitment may manifest symptoms later on in life.

SNP	Study	MAF	Phenotype
rs4420638	**Ottawa Heart Study Controls**	**0.15**	**Healthy men over 65 and women over 70-years with no symptoms of CAD**
	HapMap	0.18	Not phenotypically defined
	Wallace et al. 2008	0.19	Meta-analysis including hypertensive individuals from British Genetics of Hypertension Study (BRIGHT), and Welcome Trust Case Control Consortium (WTCCC)
	Sandhu et al. 2008	0.18	Meta-analysis of five study populations of European individuals aged 35-79. Two sets included obese and dyslipidemic cases.
	Kathiresan et al. 2009	0.16	Meta-analysis of seven GWA studies including the Diabetes Genetics Initiative, Framingham Heart Study and others
	Willer et al. 2008	0.18	Meta-analysis of three GWA studies including diabetes case-control studies
	Barber et al. 2010	0.19	Statin GWAS using three study populations of CAD patients, healthy individuals and healthy volunteers on statin therapy
rs429358	**Ottawa Heart Study Controls**	**0.12**	**Healthy men over 65 and women over 70-years with no symptoms of CAD**
	Hatters et al. 2006	0.15	Not phenotypically defined
	Thompson et al. 2009	0.16	Statin therapy GWAS using patients with clinically evident CAD
	Zuo et al. 2006	0.36 (Case) 0.14 (Controls)	Case: European-American Alzheimer's patients Controls: No Alzheimer's
	Schmidt et al. 2010	0.15	Genomic study on motor neuron disease. Healthy controls with no neurological disorders.
rs7412	**Ottawa Heart Study Controls**	**0.08**	**Healthy men over 65 and women over 70-years with no symptoms of CAD**
	Hatters et al. 2006	0.07	Not phenotypically defined
	Thompson et al. 2009	0.06	Statin therapy GWAS using patients with clinically evident CAD
	Zuo et al. 2006	0.02 (Case) 0.06 (Controls)	Case: European-American Alzheimer's patients Controls: No Alzheimer's
	Schmidt et al. 2010	0.06	Genomic study on motor neuron disease. Healthy controls with no neurological disorders.

Table 4.1: Comparison of the minor allele frequencies of rs4420638, rs429358 and rs7412 in the OHS with past studies.

The minor allele frequencies of rs4420638 and rs429358 are lower in the OHS relative to other GWA studies and meta-analyses. The OHS controls are selected based on stringent restrictions including, age, past CAD history and current health, including cognitive ability. Other studies do not employ as stringent restrictions when selecting control subjects. The allele frequency of rs7412 is higher in our controls relative to others since our selection stringency strongly selected against CAD and AD risk SNPs and selected for "protective" SNPs.

Similarly, Willer et al's meta-analysis included samples from three studies: FUSION (Finland-United States Investigation of NIDDM Genetics) study, SardiNIA Study on Aging, and the Diabetes Genetics Inititative (DGI) study. Both FUSION and DGI studies identify genetic markers associated with increased type 2 diabetes risk [83,84]. The goal of the SardiNIA study – supported by the US National Institute on Aging (NIA) – is to identify genes associated with aging, specifically cardiovascular-related diseases [85]. Again, the population controls used in these studies underwent a different screening process than those of the OHS.

Because of the OHS "extreme" controls, a large difference between MAFs of cases and controls results. Extreme differences in MAFs results in extreme odds ratios, as seen in the SNPs rs429358, rs7412 and rs4420638. This explains why our ORs are considerably more extreme than previously reported values.

Comparing ORs calculated from the same sample population is more analogous. The numerical difference between previously reported ORs of rs4420638 and rs429358 indicated a possible functional difference between the SNPs. The meta-analyses by Willer et al (2008) calculated an OR of 1.17 (1.08-1.28) for rs4420638 and Bennet et al (2007) calculated 1.06 (0.99, 1.13) for rs429358. However, these reported ORs were not obtained from the same sample population. The OHS calculates a lower OR for rs4420638 (1.52 (1.32, 1.75)) than rs429358 (1.64 (1.40, 1.93)). Comparing these values is more analogous and likely corresponds to the true trend of CAD risk association of these SNPs. In other words, although both SNPs associate with increased

CAD risk, rs429358 is a stronger representative of the degree of risk at the 19q13.2 region than rs4420638.

4.3 - Minor Allele Frequencies of rs4420638 and rs429358 differ

Despite our finding that rs429358 has a stronger effect on CAD risk, rs4420638 has a higher MAF in the Ottawa Heart Population – in cases as well as controls (**Table 3.4**). If the association of rs4420638 with CAD risk is namely due to its linkage to rs429358, why do their MAFs differ considerably? An explanation for this can be provided by the theory proposed by Mahley & Rall (1999) and Finch & Sapolsky (1999) regarding the evolutionary emergence of the three apoE isoforms.

Mahley & Rall propose that ε4 is the ancestral mammalian allele despite ε3 being the more common allele in humans. Their theory stems from the observation that most all mammals – including primates – have only one ε4-like apoE allele. Animal apoE has an arginine, like apoE4, at position 112, instead of a cysteine like apoE3. However, in spite of arg112, animal apoE functions much like human apoE3. This because animal apoE lacks an arginine at position 61. As previously explained (see **Figure 1.9**), arg112 of apoE4 causes arg61 to interact with glutamate-255 causing a domain-domain interaction. This interaction hinders apoE4's lipid-binding domain and, thus, apoE4 preferentially binds larger lipoproteins. However, animal apoE lacks arg61 (most are threonine residues) and thus, operates much like human apoE3. [86,87]

In summary, they propose that human apoE4 gave rise to apoE3 which gave rise to apoE2. Both of these transformations involve arginine → cysteine mutations at

positions 112 then 158. Both codons fall in high-mutation CpG sites and involve C → T transitions. Over 35% of all mutations occur at CpG sites and over 90% of those mutations are transitions, namely, C → T transitions [88]. This is because methylated cytosine can undergo spontaneous deamination to thymine [88]. Thus, a C → T mutation is more readily explained than a T → C mutation, which would be required for E4 to emerge from E3 [86].

The perfect D' linkage between the two apoE SNPs (rs429358, rs7412) and rs4420638 as well as the haplotype analysis indicates that, although rs429358 has a stronger effect on CAD risk, it may not be independent of rs4420638. According to Mahley & Rall the evolutionary trend is ε4 → ε3, that is, ε4 is being 'lost'. Since it is inherited together with rs4420638, this would imply that, as ε4 is being selected against, rs4420638 would also become selected against. However, rs429358 has a stronger detrimental effect on atherosclerosis and Alzheimer's disease, thus, selection against ε4 may occur more readily. Additionally, ε4 falls in a CpG site with a high mutation rate while rs4420638 does not. Thus, the rate of mutation at ε4 is increased relative to rs4420638 which results in the lower risk allele frequency at rs429358 than rs4420638, as demonstrated in the OHS MAFs of these two SNPs.

4.4 - ApoE, but not apoC1 serum concentrations are associated with rs429358, rs4420638 and rs7412 genotypes

Apoliporotein E and C1 levels were measured in serum using ELISA methods. Serum samples from controls that were not on lipid-lowering medication were chosen.

The concentration of apolipoprotein in serum may not necessarily coincide with gene expression or mRNA levels. It is a measurement of the protein in serum at the time of blood sampling. This measurement gives insight into lipid homeostasis and the functionality of each of these apolipoproteins.

Since rs429358 and rs7412 affect the functionality of apoE, it is not unexpected to find these SNPs correlate with apoE concentration in serum. In other words, the genotype (i.e. functionality) of apoE has an effect on lipid homeostasis. Additionally, due to the linkage between rs429358 and rs4420638, it is not unexpected that rs4420638 is correlated to apoE serum concentration.

The minor alleles of rs4420638 and rs429358 correlate with lower apoE in serum. Rs429358 minor allele homozygotes: 14.8 ± 3.4 ug/ml; heterozygotes 40.1 ± 3.1 ug/ml whereas major allele homozygotes have the highest apoE serum concentration: 67.0 ± 3.4 ug/ml (p = 1.73E-08; mean ± SEM). Rs4420638 genotypes follow a similar trend: AA: 25.5 ± 4.8 ug/ml; AB: 44.8 ± 3.1 ug/ml; BB: 69.9 ± 3.7 ug/ml (p = 7.27E-07; mean ± SEM). Due to the "protective" nature of rs7412, this SNP follows an opposite trend: AA: 87.2 ± 9.7 ug/ml; AB: 71.5 ± 8.7 ug/ml; BB: 53.8 ± 2.7 ug/ml (p = 0.002; mean ± SEM). Enhancer SNP rs35136575 also associates with apoE serum levels (p=0.005) with a trend corresponding to risk SNPs. This enhancer region may alter apoE mRNA expression levels in hepatocytes leading to altered apoE levels in serum. Future gene expression assays, such as luciferase assays, or real-time PCR, would further elucidate the role rs35136575 has on apoE expression.

Naturally, the six apoE genotypes also show a similar trend. ApoE4 carriers had the lowest apoE serum concentrations: E4/E4 14.5 ± 3.0 ug/ml, while apoE2 carriers had the highest: E2/E2 87.4 ± 8.2 ug/ml; ($p = 1.11E-07$; mean ± SEM).

ApoE serves as a ligand for LDLr and other LDL receptor family members. It is a constituent of VLDL and CRs, which primarily transport lipids from the liver and intestine to peripheral tissues, and HDL, which transports lipids and cholesterol from peripheral tissues back to the liver. The functionality of apoE is tightly dependant on the apoE genotype. ApoE4 has a hindered lipid binding region and as a result, has a higher affinity for larger lipoproteins as VLDL and CRs. LDLr binding affinity is not impaired and, as a result, VLDL and CRs are readily lipolyzed and taken up by endocytosis. As VLDL and CR molecules are readily removed from plasma, this would result in a lower concentration of apoE as well. ApoE2 has impaired receptor-binding affinity and has a higher affinity for smaller lipoproteins as HDL. As a result, larger lipoproteins, namely, VLDL and CRs are not readily endocytosed by hepatocytes and remain in blood plasma for a longer period of time.[11,66,14]. This results in increased apoE serum concentration. This trend is evident in the serum control samples analyzed by ELISA in the OHS.

Unlike apoE, apoC1 serum concentrations did not associate with genotypes of any of the above SNPs ($p > 0.20$). Studies have shown that overexpression of apoC1 lead to hypercholesterolemia and hypertriglyceridemia by interfering with LPL activity. This primarily inhibits the conversion of VLDL to LDL. Additionally, overexpressed apoC1 inhibits VLDL binding to LDLr and LRP which also leads to

hypertriglyceridemia.[30,32,31]. Thus, the functionality of apoC1 is concentration dependent. It is important to address whether apoC1 expression and concentration levels are associated with genotype. Also, because rs4420638 lies 3' to apoC1, polyadenylation or additional 3' posttranscriptional or posttranslational modification factors could be affected by this SNP.

However, serum concentrations of apoC1 did not show a trend with rs429358, rs7412 or rs4420638. Thus, these genotypes do not affect the functionality of apoC1. However, since serum concentration does not measure mRNA levels, it cannot be concluded that these SNPs do not affect transcription of these apolipoproteins.

4.5 - LDL blood lipid levels are genotypically associated

Plasma LDL levels associate with rs4420638, rs429358 and rs7412 genotypes. Minor alleles of both rs429358 ($p = 0.0007$) and rs4420638 ($p = 0.003$) have higher LDL levels than major alleles (see **Table 3.9**). Homozygous minor allele carriers of rs7412 have lower LDL levels than major alleles (AA: 3.29 ± 0.50 mmol/L; AB: 3.07 ± 0.06 mmol/L; BB: 3.73 ± 0.03 mmol/L; $p = 1.78e\text{-}18$; mean ± SEM). LDL also shows a similar trend with apoE genotypes: E2/E2 have the lowest plasma LDL concentration at 3.29 ± 0.46 mmol/L and increases to E4/E4 carriers at 3.96 ± 0.21 mmol/L ($p = 3.67e\text{-}21$).

ApoE functionality directly correlates with lipid homeostasis. It is not unexpected to find apoE4 carriers with higher LDL levels than apoE2 carriers. ApoE4 has a higher affinity for VLDL molecules. This increases their interaction with

receptors and can readily be lipolysed to LDL molecules. Since apoE is not a constituent of LDL, the endocytic uptake of LDL is not increased and LDL clearance is normal. The conversion of VLDL to LDL is favoured with apoE4. ApoE2 has the opposite effect; apoE2 has impaired receptor binding and preferentially binds HDL molecules. Since less apoE2 binds VLDL and apoE2 receptor binding is inefficient, less VLDL is converted to LDL. This results in lower LDL plasma concentration. It is interesting to note that the p-value for rs7412 LDL association is extreme at 1.78e-18. There is no doubt that E2 has an extreme effect on LDL plasma concentration. The association is statistically stronger than E4 association. A reason for this extreme association could be due to other physiological processes that must compensate for the impaired apoE2 functionality causing plasma LDL concentration to remain extremely low. In order to counteract the impaired apoE2 binding to the LDLr, hepatocytes may upregulate membrane LDLr molecules. Thus, the inefficient apoE2 would have an increased chance of binding to a receptor. As a result, the few VLDL molecules that have bound apoE2 readily get processed to LDL. The primary ligand for LDLr is apoB100 – the only apoliporotein on LDL. Since apoB100 binds normally to LDLr, regardless of apoE genotype, the few molecules of LDL that exist, are readily taken up by LDLr mediated endocytosis. This would cause LDL to remain at an extremely low plasma concentration. This is true for rs7412 minor allele carriers.

Although there is no linear relationship between apoE genotype and triglyceride (TG) levels, the bar graph (**Figure 3.3c**) indicates that E2/E2 and E4/E4 carriers have higher TG levels than E3/E3 carriers. This is because both carriers have increased plasma lipoproteins. ApoE2 carriers would have higher VLDL and CR

molecules due to its impaired LDLr binding affinity. ApoE4 carriers have increased LDL molecules since apoE4 associates preferentially with VLDL and permits the conversion of VLDL to LDL more readily. Thus, both apoE4 and apoE2 carriers have increased plasma lipoprotein molecules that contain TGs. ApoE3 carriers function normally and no "step" in lipid homeostasis is hindered or facilitated.

5. CONCLUSION

Although rs4420638 associates with increased CAD risk, both the haplotype analysis of the apoE SNPs with rs4420638 and the likelihood ratio tests, indicate that rs429358 confers a stronger effect on CAD risk than rs4420638. Due to the mild linkage existing between these SNPs, determining if rs4420638 has a physiologically independent role in CAD risk is inconclusive. The relationship of rs4420638 to CAD risk can be explained by this linkage with rs429358 and is not due to the linkage with other SNPs associated with lipids in this region. Both SNPs similarly associate with apoE serum levels but not with apoC1 levels. Additionally both of these SNPs associate with total cholesterol levels and LDL-C.

The minor allele frequency (MAF) of rs429358 and rs4420638 was lower in the OHS controls relative to other populations since these controls were selected based on good health and age (65+ years), thus are unlikely to harbor these genetic variants associated with CAD and Alzheimer's disease. The odds ratios of these SNPs were more extreme in the OHS than other studies for similar reasons.

The association of rs4420638 is determined to be largely due to its linkage with rs429358, however the MAF between these SNPs is not identical. This can be understood by considering the theory proposed by Mahley & Rall (1999) and Finch & Sapolsky (1999) regarding the evolutionary emergence of the three apoE isoforms. If apoE4 is the ancestral isoform – as their theory proposes – since the E4 determining SNP, rs429358, falls in a high mutation CpG site, and is more detrimental than

rs4420638, it is likely that this SNP was more readily selected against relative to rs4420638. Rs4420638 is not present in a high mutation site. ApoE4 rs429358 may have been more readily selected against, hence its MAF is lower than rs4420638.

Future directions in further deciphering this region's role in CAD risk could include promoter or expression analysis associated with enhancer SNP rs35136575. This SNP has no linkage with rs4420638 or rs429358 but indicates increased CAD risk (OR=1.23 (1.04, 1.44)). It associates with apoE serum levels and, since it falls in a hepatic control region 2 (HCR-2), may alter hepatic mRNA expression of these apolipoproteins.

6. REFERENCES

1. Lloyd-Jones D, Adams RJ, Brown TM, et al. Executive summary: heart disease and stroke statistics--2010 update: a report from the American Heart Association. *Circulation.* 2010;121(7):948-954.

2. Strong JP, Malcom GT, Newman WP, Oalmann MC. Early lesions of atherosclerosis in childhood and youth: natural history and risk factors. *J Am Coll Nutr.* 1992;11 Suppl:51S-54S.

3. Willeit J, Kiechl S. Biology of arterial atheroma. *Cerebrovasc. Dis.* 2000;10 Suppl 5:1-8.

4. Lloyd-Jones D, Adams R, Carnethon M, et al. Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation.* 2009;119(3):480-486.

5. Khot UN, Khot MB, Bajzer CT, et al. Prevalence of conventional risk factors in patients with coronary heart disease. *JAMA.* 2003;290(7):898-904.

6. Rose G. FAMILIAL PATTERNS IN ISCHAEMIC HEART DISEASE. *Br J Prev Soc Med.* 1964;18:75-80.

7. Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N. Engl. J. Med.* 1994;330(15):1041-1046.

8. Zdravkovic S, Wienke A, Pedersen NL, et al. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J. Intern. Med.* 2002;252(3):247-254.

9. Chan L, Boerwinkle E. Gene-Environment Interactions and Gene Therapy in Atherosclerosis. *Cardiol Rev.* 1994;2(3):130-137.

10. Vance D, Vance J, Jonas A, Phillips MC. Lipoprotein Structure. In: *Biochemistry of Lipids, Lipoproteins and Membranes.* 5th ed. Amsterdam ; Boston : Elsevier, 2008; 2008:485-506.

11. Vance D, Vance J, Fielding CJ, Fielding PE. Dynamics of lipoprotein transport in the circulatory system. In: *Biochemistry of Lipids, Lipoproteins and Membranes.* 5th ed. Amsterdam ; Boston : Elsevier, 2008; 2008:533-553.

12. Kypreos KE, Teusink B, Van Dijk KW, Havekes LM, Zannis VI. Analysis of the structure and function relationship of the human apolipoprotein E in vivo, using

adenovirus-mediated gene transfer. *FASEB J.* 2001;15(9):1598-1600.

13. Kypreos KE, Karagiannides I, Fotiadou EH, et al. Mechanisms of obesity and related pathologies: role of apolipoprotein E in the development of obesity. *FEBS J.* 2009;276(20):5720-5728.

14. Vance D, Vance J, Schneider WJ. Lipoprotein Receptors. In: *Biochemistry of Lipids, Lipoproteins and Membranes.* 5th ed. Amsterdam ; Boston : Elsevier, 2008; 2008:555-578.

15. Vance D, Vance J, Liscum L. Cholesterol Biosynthesis. In: *Biochemistry of Lipids, Lipoproteins and Membranes.* 5th ed. Amsterdam ; Boston : Elsevier, 2008; 2008:399-421.

16. Kennedy MA, Barrera GC, Nakamura K, et al. ABCG1 has a critical role in mediating cholesterol efflux to HDL and preventing cellular lipid accumulation. *Cell Metab.* 2005;1(2):121-131.

17. Tall A, Silver D. Cholesterol Transport. *Making good cholesterol even better.* 2002. Available at: http://www.cumc.columbia.edu/news/in-vivo/Vol1_Iss5_mar11_02/cholesterol.html.

18. Rader DJ, Alexander ET, Weibel GL, Billheimer J, Rothblat GH. The role of reverse cholesterol transport in animals and humans and relationship to atherosclerosis. *J. Lipid Res.* 2009;50 Suppl:S189-194.

19. Lusis AJ, Pajukanta P. A treasure trove for lipoprotein biology. *Nat. Genet.* 2008;40(2):129-130.

20. Stary HC, Blankenhorn DH, Chandler AB, et al. A definition of the intima of human arteries and of its atherosclerosis-prone regions. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Arterioscler. Thromb.* 1992;12(1):120-134.

21. Stary H, Chandler A, Glagov S, et al. A definition of initial, fatty streak, and intermediate lesions of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Arterioscler Thromb Vasc Biol.* 1994. Available at: http://atvb.ahajournals.org.proxy.bib.uottawa.ca/cgi/content/abstract/14/5/840.

22. Rader DJ, Daugherty A. Translating molecular discoveries into new therapies for atherosclerosis. *Nature.* 2008;451(7181):904-913.

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

24. Huang Y, Ji Z, Brecht WJ, et al. Overexpression of Apolipoprotein E3 in Transgenic Rabbits Causes Combined Hyperlipidemia by Stimulating Hepatic VLDL Production and Impairing VLDL Lipolysis. *Arterioscler Thromb Vasc Biol.* 1999;19(12):2952-2959.

25. Huang ZH, Luque RM, Kineman RD, Mazzone T. Nutritional regulation of adipose tissue apolipoprotein E expression. *Am J Physiol Endocrinol Metab.* 2007;293(1):E203-209.

26. Chiba T, Nakazawa T, Yui K, Kaneko E, Shimokado K. VLDL induces adipocyte differentiation in ApoE-dependent manner. *Arterioscler. Thromb. Vasc. Biol.* 2003;23(8):1423-1429.

27. Schaefer EJ, Gregg RE, Ghiselli G, et al. Familial apolipoprotein E deficiency. *J. Clin. Invest.* 1986;78(5):1206-1219.

28. Ishibashi S, Brown MS, Goldstein JL, et al. Hypercholesterolemia in low density lipoprotein receptor knockout mice and its reversal by adenovirus-mediated gene delivery. *J. Clin. Invest.* 1993;92(2):883-893.

29. Pendse AA, Arbones-Mainar JM, Johnson LA, Altenburg MK, Maeda N. Apolipoprotein E knock-out and knock-in mice: atherosclerosis, metabolic syndrome, and beyond. *J. Lipid Res.* 2009;50 Suppl:S178-182.

30. van der Hoogt CC, Berbée JFP, Espirito Santo SMS, et al. Apolipoprotein CI causes hypertriglyceridemia independent of the very-low-density lipoprotein receptor and apolipoprotein CIII in mice. *Biochim. Biophys. Acta.* 2006;1761(2):213-220.

31. Berbée JFP, van der Hoogt CC, Sundararaman D, Havekes LM, Rensen PCN. Severe hypertriglyceridemia in human APOC1 transgenic mice is caused by apoC-I-induced inhibition of LPL. *J. Lipid Res.* 2005;46(2):297-306.

32. de Haan W, Out R, Berbée JFP, et al. Apolipoprotein CI inhibits scavenger receptor BI and increases plasma HDL levels in vivo. *Biochem. Biophys. Res. Commun.* 2008;377(4):1294-1298.

33. Jong MC, Hofker MH, Havekes LM. Role of ApoCs in lipoprotein metabolism: functional differences between ApoC1, ApoC2, and ApoC3. *Arterioscler. Thromb. Vasc. Biol.* 1999;19(3):472-484.

34. Jong MC, Voshol PJ, Muurling M, et al. Protection from obesity and insulin resistance in mice overexpressing human apolipoprotein C1. *Diabetes.* 2001;50(12):2779-2785.

35. Smit M, van der Kooij-Meijs E, Frants RR, Havekes L, Klasen EC. Apolipoprotein gene cluster on chromosome 19. Definite localization of the APOC2 gene and the polymorphic Hpa I site associated with type III hyperlipoproteinemia. *Hum. Genet.* 1988;78(1):90-93.

36. Xu Y, Berglund L, Ramakrishnan R, et al. A common Hpa I RFLP of apolipoprotein C-I increases gene transcription and exhibits an ethnically distinct pattern of linkage disequilibrium with the alleles of apolipoprotein E. *J. Lipid Res.* 1999;40(1):50-58.

37. Tycko B, Lee JH, Ciappa A, et al. APOE and APOC1 promoter polymorphisms and the risk of Alzheimer disease in African American and Caribbean Hispanic individuals. *Arch. Neurol.* 2004;61(9):1434-1439.

38. 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.

39. Allan CM, Taylor JM. Expression of a novel human apolipoprotein (apoC-IV) causes hypertriglyceridemia in transgenic mice. *J. Lipid Res.* 1996;37(7):1510-1518.

40. Shah SH. Gene polymorphisms and susceptibility to coronary artery disease. *Pediatr Blood Cancer.* 2007;48(7):738-741.

41. Jensen J, Blankenhorn DH, Kornerup V. Coronary disease in familial hypercholesterolemia. *Circulation.* 1967;36(1):77-82.

42. Baker PB, Baba N, Boesel CP. Cardiovascular abnormalities in progeria. Case report and review of the literature. *Arch. Pathol. Lab. Med.* 1981;105(7):384-386.

43. McCully KS, Ragsdale BD. Production of arteriosclerosis by homocysteinemia. *Am. J. Pathol.* 1970;61(1):1-11.

44. Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature.* 2001;409(6822):928-933.

45. The International HapMap Project. *Nature.* 2003;426(6968):789-796.

46. McPherson JD, Marra M, Hillier L, et al. A physical map of the human genome. *Nature.* 2001;409(6822):934-941.

47. A haplotype map of the human genome. *Nature.* 2005;437(7063):1299-1320.

48. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449(7164):851-861.

49. Wallace C, Newhouse SJ, Braund P, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am. J. Hum. Genet.* 2008;82(1):139-149.

50. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* 2008;40(2):161-169.

51. Barber MJ, Mangravite LM, Hyde CL, et al. Genome-wide association of lipid-lowering response to statins in combined study populations. *PLoS ONE.* 2010;5(3):e9763.

52. Dausset J, Cann H, Cohen D, et al. Centre d'Etude du polymorphisme humain (CEPH): Collaborative genetic mapping of the human genome. *Genomics.* 1990;6(3):575-577.

53. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science.* 2007;316(5830):1488-1491.

54. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science.* 2007;316(5830):1491-1493.

55. Zhou L, Zhang X, He M, et al. Associations between single nucleotide polymorphisms on chromosome 9p21 and risk of coronary heart disease in Chinese Han population. *Arterioscler. Thromb. Vasc. Biol.* 2008;28(11):2085-2089.

56. Ding H, Xu Y, Wang X, et al. 9p21 is a shared susceptibility locus strongly for coronary artery disease and weakly for ischemic stroke in Chinese Han population. *Circ Cardiovasc Genet.* 2009;2(4):338-346.

57. Hinohara K, Nakajima T, Takahashi M, et al. Replication of the association between a chromosome 9p21 polymorphism and coronary artery disease in Japanese and Korean populations. *J. Hum. Genet.* 2008;53(4):357-359.

58. Hiura Y, Fukushima Y, Yuno M, et al. Validation of the association of genetic variants on chromosome 9p21 and 1q41 with myocardial infarction in a Japanese population. *Circ. J.* 2008;72(8):1213-1217.

59. Shen G, Li L, Rao S, et al. Four SNPs on chromosome 9p21 in a South Korean population implicate a genetic locus that confers high cross-race risk for development of coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.* 2008;28(2):360-365.

60. Shen G, Rao S, Martinelli N, et al. Association between four SNPs on chromosome 9p21 and myocardial infarction is replicated in an Italian population. *J. Hum. Genet.* 2008;53(2):144-150.

61. Meng W, Hughes AE, Patterson CC, et al. Chromosome 9p21.3 is associated with early-onset coronary heart disease in the Irish population. *Dis. Markers.* 2008;25(2):81-85.

62. Maitra A, Dash D, John S, et al. A common variant in chromosome 9p21 associated with coronary artery disease in Asian Indians. *J. Genet.* 2009;88(1):113-118.

63. Jarinova O, Stewart AFR, Roberts R, et al. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler. Thromb. Vasc. Biol.* 2009;29(10):1671-1677.

64. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* 2009;41(1):56-65.

65. Bennet AM, Di Angelantonio E, Ye Z, et al. Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA.* 2007;298(11):1300-1311.

66. Hatters DM, Peters-Libeu CA, Weisgraber KH. Apolipoprotein E structure: insights into function. *Trends Biochem. Sci.* 2006;31(8):445-454.

67. Elshourbagy NA, Liao WS, Mahley RW, Taylor JM. Apolipoprotein E mRNA is abundant in the brain and adrenals, as well as in the liver, and is present in other peripheral tissues of rats and marmosets. *Proc. Natl. Acad. Sci. U.S.A.* 1985;82(1):203-207.

68. Pfrieger FW. Cholesterol homeostasis and function in neurons of the central nervous system. *Cell. Mol. Life Sci.* 2003;60(6):1158-1171.

69. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA.* 1997;278(16):1349-1356.

70. Li H, Wetten S, Li L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch. Neurol.* 2008;65(1):45-53.

71. Coon KD, Myers AJ, Craig DW, et al. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psychiatry.* 2007;68(4):613-618.

72. Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science.* 2002;297(5580):353-356.

73. Wisniewski T, Frangione B. Apolipoprotein E: a pathological chaperone protein in patients with cerebral and systemic amyloid. *Neurosci. Lett.* 1992;135(2):235-238.

74. Rebeck GW, Harr SD, Strickland DK, Hyman BT. Multiple, diverse senile plaque-associated proteins are ligands of an apolipoprotein E receptor, the alpha 2-macroglobulin receptor/low-density-lipoprotein receptor-related protein. *Ann. Neurol.* 1995;37(2):211-217.

75. Benjamin R, Leake A, Ince PG, et al. Effects of apolipoprotein E genotype on cortical neuropathology in senile dementia of the Lewy body and Alzheimer's disease. *Neurodegeneration.* 1995;4(4):443-448.

76. Heinonen O, Lehtovirta M, Soininen H, et al. Alzheimer pathology of patients carrying apolipoprotein E epsilon 4 allele. *Neurobiol. Aging.* 1995;16(4):505-513.

77. Aleshkov S, Abraham CR, Zannis VI. Interaction of nascent ApoE2, ApoE3, and ApoE4 isoforms expressed in mammalian cells with amyloid peptide beta (1-40). Relevance to Alzheimer's disease. *Biochemistry.* 1997;36(34):10571-10580.

78. Affymetrix Inc. Affymetrix Genome-Wide Human SNP Array 6.0 Data Sheet. 2007.

79. 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/C-II gene cluster. *J. Biol. Chem.* 1997;272(46):29113-29119.

80. Klos K, Shimmin L, Ballantyne C, et al. APOE/C1/C4/C2 hepatic control region polymorphism influences plasma apoE and LDL cholesterol levels. *Hum. Mol. Genet.* 2008;17(13):2039-2046.

81. Grehan S, Tse E, Taylor JM. Two distal downstream enhancers direct expression of the human apolipoprotein E gene to astrocytes in the brain. *J. Neurosci.* 2001;21(3):812-822.

82. Sandhu MS, Waterworth DM, Debenham SL, et al. LDL-cholesterol concentrations: a genome-wide association study. *Lancet.* 2008;371(9611):483-491.

83. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008;40(5):638-645.

84. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT LU, Saxena R, Voight BF, et al. Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. *Science.* 2007;316(5829):1331-1336.

85. Pilia G, Chen W, Scuteri A, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* 2006;2(8):e132.

86. Mahley RW, Rall SC. Is epsilon4 the ancestral human apoE allele? *Neurobiol. Aging.* 1999;20(4):429-430.

87. Finch CE, Sapolsky RM. The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms[small star, filled]. *Neurobiology of Aging.* 1999;20(4):407-428.

88. Cooper DN, Youssoufian H. The CpG dinucleotide and human genetic disease. *Hum. Genet.* 1988;78(2):151-155.

APPENDIX 1

Detailed Competition ELISA Protocol Used to Determine Serum ApoE and ApoC1 Concentrations

Competition ELISA

One full 96-well plate per Apolipoprotein measured

Day 1:

1) **Coat** the plate with purified human apolipoprotein (BioDesign): 100ul of 0.5ug/ml in each well overnight at 4°.
    a. ApoE (0.6mg/ml): 7.1ul in 9700ul of coating buffer (PBS/NaN3) (9692.9ul)
    b. ApoC1 (1.25mg/ml): 3.9ul in 9700ul of coating buffer (PBS/NaN3) (9696.1ul)

2) **Incubate** equal parts of 1/12800 diluted biotin conjugated antibody with sample (standards and serum) overnight at 4° (total volume should be at least 200ul to do in duplicate). Two dilutions of serum used: 1/20 and 1/40. Standards use purified human apolipoprotein ranging from 25ug/ml to 0.0 mg/ml in ½ dilution increments. (See Plate Layout for more detail.)
    To make 1/12800 diluted antibody
    a. 2ul antibody in 200ul sample buffer (198ul) – 1/100
    b. 62.5ul soln (a) in 8ml sample buffer (14.387ml) – 1/12800

Day 2:

3) **Wash** 3 x 200ul PBS-T (necessary to get rid of excess apolipoprotein in soln)

4) **Block** 1hr, shaking at room temperature: PBS/BSA/NaN3, 200ul per well.

5) **Incubate** at 37°C for 1hr (strict!) with 100ul of incubated samples with diluted antibody (from Step 2). Empty blocking solution thoroughly (do not wash) prior to adding samples.

6) **Wash** with 3 x 200ul PBS-T.

7) **Dilute** avidin-alkaline phosphatase to 1:30 000 (if doing two plates, ie, ApoE and ApoC1 at the same time, double these volumes)
    a. 5ul of enzyme to 495ul of sample buffer (1:100)
    b. 40ul of soln (a) to 1160ul sample buffer (1:3000)
    c. 1050ul of soln (b) to 9450ul sample buffer (1:30 000)

8) **Add** 100ul of diluted aa-phosphatase to each well. Incubate for 1hr at room temperature, shaking (take out ready-mixed liquid substrate to ensure it's at room temperature for use)

9) **Wash** with 3 x 200ul PBS-T

10) **Add** 100ul of room temperature ready-mixed liquid substrate. Read at 405nm when colour appears.

BUFFERS

COATING BUFFER – PBS/NaN3

Reagents	10ml (1 plate)	20ml (2 plates)	100ml	250ml
10x PBS	1 ml	2ml	10 ml	25 ml
2% NaN3	100 ul	200 ul	1 ml	2.5 ml
ddH2O	8.9 ml	17.8 ml	89 ml	222.5 ml

WASH – PBS-T

Reagents	250ml (1 plate)	500ml (2 plates)	1 L	2 L
10x PBS	25 ml	50ml	100 ml	200 ml
T-20	1.25 ml	2.5 ml	5 ml	10 ml
ddH2O	223.75 ml	447.5 ml	895 ml	1790 ml

BLOCKING BUFFER – PBS/BSA/NaN3

Reagents	20ml (1 plate)	40ml (2 plates)	100ml	500ml
10x PBS	2 ml	4 ml	10 ml	50 ml
2% NaN3	0.2 ml	0.4 ml	1 ml	5 ml
BSA	0.1 g	0.2 g	0.5 g	2.5 g
ddH2O	17.8 ml	35.6 ml	89 ml	445 ml

SAMPLE BUFFER – PBS/BSA/T-20/NaN3

Reagents	100ml (1 plate)	200ml (2 plates)	500ml	1L
10x PBS	10 ml	20 ml	50 ml	100 ml
2% NaN3	1 ml	2 ml	5 ml	10 ml
BSA	0.5 g	1 g	2.5 g	5 g
T-20	50ul	100ul	250ul	500ul
ddH2O	88.95 ml	177.9 ml	444.75 ml	889.5 ml

Plate Schematic (96-Well plate)

One plate can hold 11 standards in duplicate (rows A – B), two wells with no apolipoprotein coat as a controls (A12, B12), and 18 serum samples at 1/20 and 1/40 dilutions, each in duplicate (rows C – H).

	1	2	3	4	5	6	7	8	9	10	11	12
A	25.0	12.5	6.25	3.13	1.56	0.78	0.39	0.2	0.1	0.05	0.0	No coat
B	25.0	12.5	6.25	3.13	1.56	0.78	0.39	0.2	0.1	0.05	0.0	No coat
C	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40
D	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40
E	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40
F	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40
G	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40
H	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40	1/20	1/40

Standards (ug/ml)
Un-coated well control
Serum sample
Serum sample

APPENDIX 2

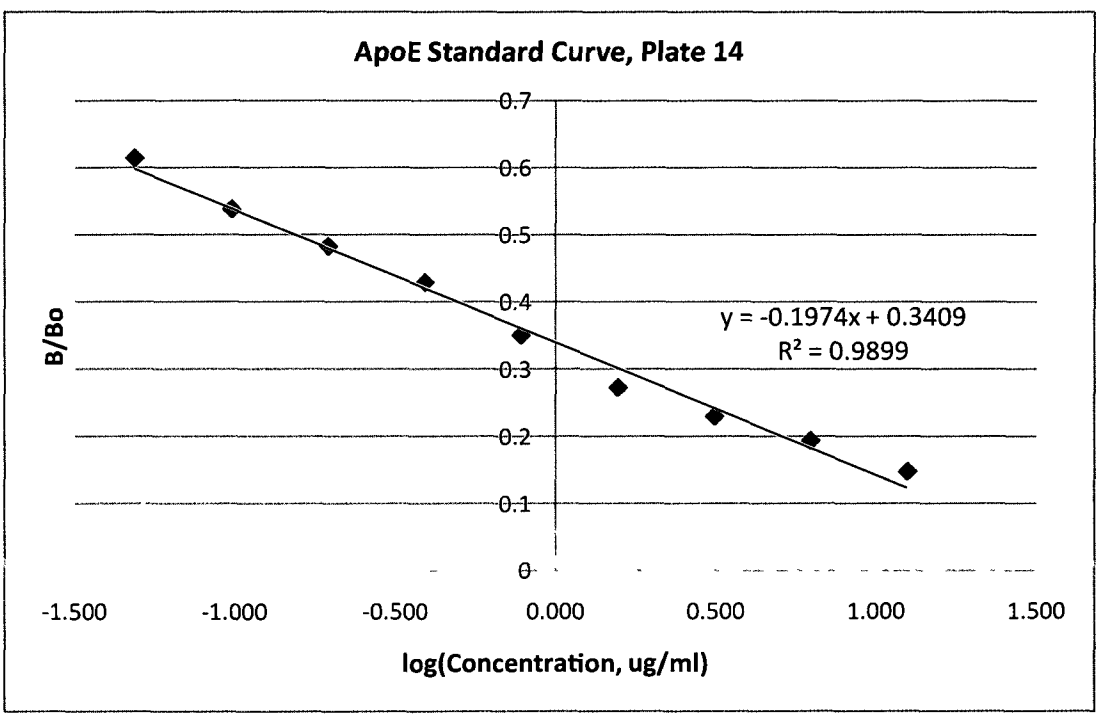
Sample ApoE and ApoC1 Serum Concentration Calculations

Example from Plate 14

ApoE Standards (ug/ml)

Absorbance (Abs) readings at 405 nm

Concentration	log(Conc)	*Abs 1*	*Abs2*	Mean Abs (B)	B/Bo
25.00	**1.398**	*0.169*	*0.158*	0.1635	**0.150207**
12.50	**1.097**	*0.168*	*0.153*	0.1605	**0.147451**
6.25	**0.796**	*0.213*	*0.209*	0.211	**0.193845**
3.13	**0.495**	*0.257*	*0.243*	0.25	**0.229674**
1.56	**0.194**	*0.302*	*0.291*	0.2965	**0.272393**
0.78	**-0.107**	*0.386*	*0.377*	0.3815	**0.350482**
0.39	**-0.408**	*0.485*	*0.449*	0.467	**0.429031**
0.20	**-0.709**	*0.549*	*0.502*	0.5255	**0.482774**
0.10	**-1.010**	*0.618*	*0.554*	0.586	**0.538356**
0.05	**-1.311**	*0.706*	*0.632*	0.669	**0.614607**
Bo = 0.0	--	*1.064*	*1.113*	1.0885	**1**



Standard Curve Linear Regression Equation

y = mx + b

y = -0.1974x + 0.3409

m= -0.1974

b= 0.3409

Serum Samples

y = mx + b

y = -0.1974x + 0.3409

m= -0.1974

b= 0.3409

$$x = \frac{(Abs/Bo) - b}{m} = \log(Conc)$$

Examples of three Serum samples from Plate 14

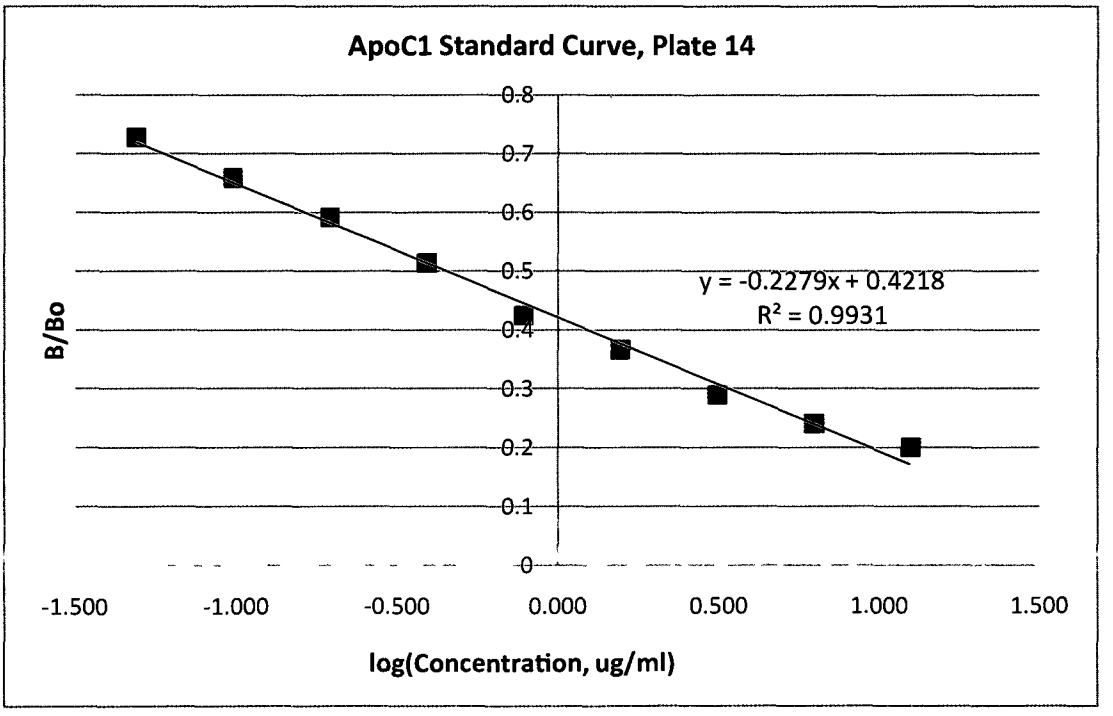
Sample #	Dilution	*Abs*	Abs/Bo = y	x=log(Conc)	Conc	Conc*Diln	Avg Conc
CONT-648	1/20	*0.415*	0.38125861	-0.20445093	0.624524	12.490478	**20.41683**
	1/20	*0.4*	0.36747818	-0.13464124	0.73343	14.668603	
	1/40	*0.429*	0.39412035	-0.26960663	0.537518	21.500738	
	1/40	*0.389*	0.35737253	-0.08344747	0.825187	33.007491	
CONT-683	1/20	0.237	0.21773082	0.62395733	4.206853	84.13706	**103.2362**
	1/20	0.211	0.19384474	0.74496079	5.558541	111.17081	
	1/40	0.298	0.27377124	0.34006462	2.188087	87.523486	
	1/40	0.261	0.23977951	0.51226184	3.252834	130.11334	
CONT-824	1/20	0.371	0.34083601	0.00032415	1.000747	20.014933	**23.39521**
	1/20	0.334	0.30684428	0.17252137	1.487721	29.754412	
	1/40	0.462	0.4244373	-0.42318794	0.377409	15.096353	
	1/40	0.402	0.36931557	-0.1439492	0.717878	28.71513	

Mean ApoE Concentrations were determined for each genotype

ApoC1 Standards (ug/ml)

Absorbance (Abs) readings at 405 nm

Concentration	log(Conc)	*Abs 1*	*Abs2*	Mean Abs (B)	B/Bo
25.00	**1.398**	*0.28*	*0.254*	0.267	**0.209658**
12.50	**1.097**	*0.266*	*0.245*	0.2555	**0.200628**
6.25	**0.796**	*0.312*	*0.302*	0.307	**0.241068**
3.13	**0.495**	*0.38*	*0.358*	0.369	**0.289753**
1.56	**0.194**	*0.492*	*0.442*	0.467	**0.366706**
0.78	**-0.107**	*0.545*	*0.537*	0.541	**0.424814**
0.39	**-0.408**	*0.682*	*0.627*	0.6545	**0.513938**
0.20	**-0.709**	*0.768*	*0.74*	0.754	**0.592069**
0.10	**-1.010**	*0.846*	*0.833*	0.8395	**0.659207**
0.05	**-1.311**	*0.929*	*0.926*	0.9275	**0.728308**
Bo = 0.0	--	*1.287*	*1.26*	1.2735	**1**



Standard Curve Linear Regression Equation

y = mx + b

y = -0.1974x + 0.3409

m= -0.2279

b= 0.4218

Serum Samples

y = mx + b

y = -0.2279x + 0.4218

m= -0.2279

b= 0.4218

$$x = \frac{(Abs/Bo) - b}{m} = \textbf{log(Conc)}$$

Examples of three Serum samples from Plate 14

Sample #	Dilution	*Abs*	Abs/Bo = y	x=log(Conc)	Conc	Conc*Diln	Avg Conc
CONT-648	1/20	*0.372*	0.29210836	0.56907256	3.707427	74.148532	**120.2474**
	1/20	*0.339*	0.26619552	0.68277523	4.816984	96.339687	
	1/40	*0.381*	0.2991755	0.53806274	3.451936	138.07744	
	1/40	*0.353*	0.27718885	0.63453774	4.3106	172.42401	
CONT-683	1/20	*0.302*	0.23714174	0.81026005	6.460409	129.20819	**160.3291**
	1/20	*0.252*	0.19787986	0.98253682	9.605873	192.11745	
	1/40	*0.375*	0.29446408	0.55873596	3.620228	144.80913	
	1/40	*0.351*	0.27561837	0.64142881	4.379543	175.18173	
CONT-824	1/20	*0.269*	0.2112289	0.92396272	8.393879	167.87758	**155.1441**
	1/20	*0.299*	0.23478602	0.82059665	6.616018	132.32035	
	1/40	*0.355*	0.27875932	0.62764667	4.242742	169.7097	
	1/40	*0.37*	0.29053789	0.57596363	3.766723	150.6689	

Mean ApoC1 Concentrations were determined for each genotype

APPENDIX 3

Curriculum Vitae

SIMRAN S. BHATIA

University of Ottawa Heart Institute

• EDUCATION

Masters of Science in Biochemistry 2008-Present
University of Ottawa, Ottawa, ON, Canada

Bachelor of Science (Hons.) in Biochemistry (Co-op) 2004-2007
University of Victoria, Victoria, BC, Canada

• PUBLICATIONS

Weiss, T. C.; Zhai G.G.; **Bhatia, S. S.**; Romaniuk, P. J. *An RNA Aptamer With High Affinity and Broad Specificity for Zinc Finger Proteins.* (2010) Biochemistry. 49(12): 2732-40. PMID: 20175561

Bhatia, S. S.; Weiss, T. C.; Romaniuk, P. J. *Contribution of Individual Amino Acids to the 5S RNA Binding Activity of the* Xenopus *Zinc Finger Protein p43.* (2008) Biochemistry. 47(32); 8398-8405. PMID: 18636752

• CONFERENCES

American Society of Human Genetics (ASHG) 59th Annual Meeting 2009
Honolulu, Hawaii, USA Oct 20-24, 2009
Poster Presented

University of Ottawa Heart Institute Research Day
University of Ottawa Heart Institute, Ottawa ON April 22, 2009
Seminar Presented (2nd Place Award in Research Category)

SCOLAR (Stroke, Cardiovascular, Obesity, Lipids, Atheroslerosis Research) Program
University of Alberta, Edmonton, Alberta May 5-16, 2008

• AWARDS AND SCHOLARSHIPS

Ontario Graduate Scholarship (OGS), 2009-2010
University of Ottawa Excellence Scholarship, 2009-2010
NSERC-Undergraduate Student Research Award, 2007, 2006
Deans' Honour List, University of Victoria, 2007
University of Victoria President's Entrance Scholarship, 2003
Khalsa Credit Union 'Award of Academics' Scholarship, 2003
Esquimalt High School Graduation Scholarship, 2003

• RESEARCH ASSISTANTSHIPS

University of Victoria, Victoria, BC
Supervisor: Dr Paul J Romaniuk
Research Assistant Jan-Aug 2006; May-Aug 2007
Research Report: RNA-Protein Interactions Between Zinc Finger Protein p43 and 5S RNA

- *NSERC-Undergraduate Student Research Award funded*

University of Alberta, Edmonton, AB
National High Field Nuclear Magnetic Resonance Centre (NANUC)
Research Assistant May-August 2005
Research Report: Metabolic Analysis of Urine from Healthy Human Subjects Using ^{1}H NMR Spectrometry