Associations between specific ApoE genetic variants and their interactions with environmental factors in relation to the lipid profile of black South Africans

L Meades

20687508

Dissertation submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in Nutrition at the Potchefstroom Campus of the North-West University

Supervisor:	Dr KR Conradie
Co-supervisor:	Dr GW Towers
Assistant promotor:	Dr C Nienaber-Rousseau

May 2014

It all starts here ™

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT ®

Assosiasies tussen spesifieke ApoE genetiese variante en hulle interaksie met omgewingsfaktore in verhouding tot die lipied profiel van swart Suid-Afrikaners

L Meades

20687508

Verhandeling voorgelê vir gedeeltelike nakoming van die vereistes vir die graad *Magister Scientiae* in Voeding aan die Noord-Wes Universiteit, Potchefstroom kampus

Studieleier: Dr KR Conradie

Medestudieleier: Dr GW Towers

Hulpstudieleier: Dr C Nienaber-Rousseau

Mei 2014

It all starts here ™

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT ®

ABSTRACT

Introduction: Cardiovascular disease (CVD) is the leading cause of global mortality and its prevalence is increasing among black South Africans in spite of their favourable lipid profile. Apolipoprotein E (*ApoE*) is a well-described risk factor for CVD and certain polymorphisms within this gene alter the lipid profile. The author hypothesised that there are population-specific effects within the *ApoE* gene that are responsible for the favourable lipid profile observed in black South Africans whose effects are being altered by environmental factors.

Objectives: The main aim of this study was to investigate the associations between specific *ApoE* single nucleotide polymorphisms (SNPs) and the lipid profile of a black South African population, taking into account certain environmental and phenotypic factors in order to explore the interaction effects between these variables.

Methods: Genotyping within this cross-sectional study (n=1 588), nested within the Prospective Urban and Rural Epidemiology (PURE) study, was achieved using Illumina's® GoldenGate Genotyping Assay with VeraCode® technology on the BeadXpress® platform (proprietary multiplex fluorescent hybridisation assays on a bead array substrate) or the Bio-Rad CFX Manager© (version 2.0). The Konelab20i™ auto analyser was used for quantitative determination of serum total cholesterol; high-density lipoprotein cholesterol (HDL-C) and triglyceride (TG) concentrations. Low-density lipoprotein cholesterol concentrations were estimated by the Friedewald equation.

Results: All SNPs adhered to the assumptions of the Hardy-Weinberg equilibrium, yet the frequency of the SNPs often differed from that reported in other ethnic groups. The well-reported rs429358 and rs7412 SNPs (as the constituent SNPs of the haplotype-genotypes) presented with the strongest associations with various components of the blood lipid profile in the black South African cohort under investigation. Two gene-environment (rs405509 and rs7412) interaction effects on TG remained significant after conducting *post hoc* tests. Two genotype-phenotype interaction effects between the rs7412 SNP and body mass index and gamma-glutamyl transferase on the HDL-C concentrations remained significant after conducting *post hoc* tests.

Conclusions: The variety of associations between these particular SNPs and the blood lipid profile determined in the present cohort strongly indicates that it is integral to any public health investigation into CVD development that these SNPs be investigated. This study further produced greater insight into the biological mechanisms underlying serum lipid and cholesterol concentrations in a black South African population. Therefore, from

these results it is evident that the lipid profile of black South Africans is most definitely influenced by not only genetic variations in the *ApoE* gene and certain environmental factors, but by the interaction between these factors as well. The present study is the largest study to date to investigate the effect of polymorphisms in the *ApoE* gene on the lipid profile of black South Africans.

Key words: *ApoE* gene, *ApoE* polymorphisms, BeadXpress®, black South African population, cardiovascular disease, diet, environment, interaction effect, lipid profile

OPSOMMING

Agtergrond: Kardiovaskulêre siekte (KVS) is 'n hoofoorsaak van sterftes wêreldwyd en die voorkoms daarvan is aan die toeneem onder swart Suid-Afrikaners ten spyte van hulle gunstige lipiedprofiel. Apolipoproteïen E (*ApoE*) is 'n bekende risikofaktor vir KVS en daar is al bewys dat sekere polimorfismes in hierdie geen die lipiedprofiel beïnvloed Die hipotese is dat daar bevolkingspesifieke effekte in die *ApoE*-gene is wat vir die gunstige lipiedprofiel, wat waargeneem word in swart Suid-Afrikaners, verantwoordelik is.

Doelwit: Die hoofdoel van hierdie studie was om die assosiasies tussen spesifieke *ApoE*-polimorfismes en die lipiedprofiel van 'n swart Suid-Afrikaanse bevolking te ondersoek, met inagneming van sekere omgewings- en fenotipiese faktore ten einde die interaksie-effek tussen hierdie veranderlikes te ondersoek.

Studieontwerp en -metodes: Genotipering in hierdie deursnee-studie (n = 1,588), binne die Prospektiewe Stedelike en Landelike Epidemiologiese (*PURE*) studie is met behulp van die Illumina® se GoldenGate Genotipering (gepatenteerde veelvoudige fluoresserende hibridisering analise op 'n kraletjie opstelling substraat) deur middel van die VeraCode-tegnologie® op die BeadXpress®-platform en die Bio Rad CFX-masjien© (weergawe 2.0) gedoen. Die serum totale cholesterol (TC), hoëdigtheid-lipoproteïencholesterol (HDL-C) en trigliseried- (TG) konsentrasies is kwantitatief bepaal deur middel van die Konelab20i ™-ontleder.. Laedigtheid-lipoproteïencholesterol-konsentrasies is beraam deur middel van die Friedewald-vergelyking.

Resultate: Al die polimorfismes het voldoen aan die aannames van die Hardy-Weinberg-ewewig, maar die verspreiding van die polimorfismes het verskil van dié wat gerapporteer is in ander etniese groepe. Die bekende rs429358- en rs7412-polimorfismes (wat die samestellende polimorfismes van die haplotipe-genotipes is) het die sterkste assosiasies met verskeie komponente van die bloedlipiedprofiel in die swart Suid-Afrikaanse groep getoon. Twee statisties betekenisvolle interaksie-effekte is tussen sekere polimorfismes (rs405509 en rs7412) en die omgewing, op TG gerapporteer. Verder is twee statisties betekenisvolle interaksie-effekte tussen die rs7412-polimorfisme en sekere fenotipes [(liggaamsmassaindeks (LMI) en gamma-glutamiel transferase), op die HDL-C konsentrasies gerapporteer.

Gevolgtrekking: Die verskeidenheid assosiasies wat tussen hierdie spesifieke polimorfismes en die bloedlipiedprofiel in die huidige etniese groep bepaal is, dui sterk daarop dat dit 'n integrale deel van enige openbaregesondheid-ondersoek oor KVS moet wees. Die identifisering van sleuteldeterminante van plasmalipoproteïenkonsentrasies

(verstedeliking, alkoholgebruik, sekere dieetfaktore en LMI) het verder tot die bestaande literatuur bygedra en ook 'n mate van insig in die biologiese meganismes wat onderliggend is aan die die serumlipiedkonsentrasies in 'n swart Suid-Afrikaanse bevolking gegee. Uit hierdie resultate is dit duidelik dat die lipiedprofiel van swart Suid-Afrikaners beslis nie net deur die genetiese variasies in die *ApoE*-geen beïnvloed word nie, maar ook deur sekere omgewingsfaktore, sowel as deur die wisselwerking tussen hierdie faktore. Dit blyk dat die huidige studie tot op hede die mees omvangryke studie is wat die effek van polimorfismes (in die *ApoE*-geen) op die lipiedprofiel van swart Suid-Afrikaners ondersoek het.

Sleutelwoorde: *ApoE*-geen, *ApoE*-polimorfismes, BeadXpress®, swart Suid-Afrikaanse bevolking, kardiovaskulêre siekte, dieet, omgewing, interaksie-effek, lipiedprofiel

TABLE OF CONTENTS

LIST OF ABBREVIATIONS i
LIST OF SYMBOLS UNITS iv
LIST OF FIGURES vi
LIST OF TABLES vii
LIST OF EQUATIONS viii
ACKNOWLEDGEMENTS ix

CHAPTER ONE 1
1.1 BACKGROUND AND MOTIVATION 1
1.2 RESEARCH AIMS AND OBJECTIVES 5
1.3 CHAPTER OUTLINE 6

CHAPTER TWO
2.1 CARDIOVASCULAR DISEASE 8
2.1.1 Conventional CVD risk factors 11
2.1.2 Ethnicity as a non-modifiable risk factor for CVD development 13
2.1.3 The blood lipid profile as a modifiable risk factor of CVD development 18
2.1.3.1 Non-genetic factors that might influence the blood lipid profile 21

CHAPTER THREE
3.1 JUSTIFICATION FOR INVESTIGATING THE ROLE OF THE *APOE* GENE IN CVD RISK DEVELOPMENT 33
3.2 THE STRUCTURAL, MECHANISTIC AND FUNCTIONAL ASPECTS OF THE *APOE* GENE 35
3.3 GENETIC VARIANTS IN THE *APOE* GENE 41
3.4 ADDITIONAL SINGLE NUCLEOTIDE POLYMORPHISMS IN THE *APOE* GENE 42
3.5 THE ANTHROPOLOGY OF THE *APOE* GENE 46
3.6 ASSOCIATIONS AND INTERACTIONS BETWEEN *APOE* GENETIC VARIANTS WITH OTHER ENVIRONMENTAL AND PHENOTYPIC TRAITS IN RELATION TO CVD 48

CHAPTER FOUR
4.1 PROSPECTIVE URBAN AND RURAL EPIDEMIOLOGICAL STUDY 51
4.1.1 Aim of the study 51
4.1.2 Participant selection 52
4.2 ETHICAL ASPECTS 52
4.3 GENETIC ANALYSIS 53
4.3.1 Deoxyribonucleic acid isolation 53
4.3.2 Sequencing of the *ApoE* gene 57
4.3.3 The identification and selection of SNPs within the *ApoE* gene 60
4.3.4 Process of SNP identification and validation using the BeadXpress® platform 62
4.3.5 Genotyping using the BeadXpress® T Reader 65
4.3.5.1 Genotyping of the *ApoE* polymorphism at rs7412 67
4.4 STATISTICAL ANALYSIS 68
4.4.1 Stratification criteria 68
4.4.2 Normality testing/distribution 69
4.4.3 Statistical testing 69

CHAPTER FIVE
5.1 DEMOGRAPHIC AND BIOCHEMICAL CHARACTERISTICS OF THE PURE POPULATION 72
5.2 GENETIC DATA COLLECTION 83
5.2.1 DNA isolation 83
5.2.2 Polymerase chain reaction amplification and automated sequencing 84
5.2.3 Identification of polymorphisms in the *ApoE* gene 85
5.2.3.1 Amplification of fragment one in the *ApoE* gene 85
5.2.3.2 Amplification of fragment two in the *ApoE* gene 87
5.2.3.3 Amplification of fragment three of the *ApoE* gene 89
5.2.3.4 Amplification of fragment four in the *ApoE* gene 91
5.2.3.5 Amplification of fragment five in the *ApoE* gene 93
5.2.3.6 Polymorphism in the *ApoE* gene, reported in literature 94
5.3 BEADXPRESS® ANALYSES 95
5.3.1 Designing custom GoldenGate® genotyping assays 95
5.3.2 Generation and analysis of the GoldenGate® genotyping data 98
5.4 REAL-TIME PCR ANALYSES 107
5.5 ALLELE FREQUENCY DETERMINATION 108
5.5.1 SNP rs1081101 110
5.5.2 SNP rs405509 115
5.5.3 SNP rs440446 120
5.5.4 SNP rs429358 123
5.5.5 SNP rs7412 130
5.5.6 *ApoE* haplotype-genotypes (the ε2/ ε3/ ε4 epsilons) 143
5.6 SUMMARY OF THE RESULTS 153

CHAPTER SIX 159

ADDENDUM A 163
ADDENDUM B 168
ADDENDUM C 169
ADDENDUM D 171
REFERENCE LIST 176

LIST OF ABBREVIATIONS

Abbreviations are listed in alphabetical order:

VIC	2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein
dNTP	2'-deoxynucleotide-5'-triphosphate
ddNTPs	2', 3'-ddideoxyribonucleotide triphosphates
FAM	6-caboxyfluorescien
AMDR	acceptable macronutrient distribution range
AMI	acute myocardial infarction
A	adenine
ATP	adenosine triphosphate
ABCA1	adenosine triphosphate binding cassette transporter A1
ASO	allele-specific oligonucleotides
ANOVA	analysis of variance
T_a	annealing temperatures
APOA1	apolipoprotein A1
Apo B	apolipoprotein B
Apo E	apolipoprotein E
Arg	arginine
Asp	aspartic acid
ADT	assay design tool
β-VLDL	beta-very low density lipoprotein
BP	blood pressure
BMI	body mass index
BRISK	Black Risk Study
CHO	carbohydrates
CVD	cardiovascular disease
CEN	Centre of Excellence for Nutrition
CBVD	cerebrovascular disease
CETP	cholesterol ester transfer protein
CLD	chronic lifestyle diseases
CHD	coronary heart disease
Chr	chromosome
CM	chylomicrons
(c)SNPs	coding single nucleotide polymorphisms
COL1A1	collagen, Type I, Alpha1
CHF	congestive heart failure
CAD	coronary artery disease
CHD	coronary heart disease
CRP	C-reactive protein
Cys	Cysteine
C	Cytosine
DNA	deoxyribonucleic acid
DRI	dietary reference intake
DBP	distolic blood pressure
DZ	dizygotic
ddH_2O	double distilled water
ESRD	end-stage renal disease
E1	ApoE 1 isoform
E2	ApoE 2 isoform
E3	ApoE 3 isoform
E3r	ApoE 3r isoform
E4	ApoE 4 isoform
EB	elution buffer
EDTA	ethylenediamine tetra-acetic acid
FVIII	factor VII
FH	familial hypercholesterolaemia
FBG	fibrinogen
GGT	gamma-glutamyl transferase

GWAS	genome wide association studies
gDNA	genomic DNA
G	guanine
HWE	Hardy-Weinberg equilibrium
HL	hepatic triglyceride lipase
HDL-C	high density lipoprotein cholesterol
HOMA	homeostatic model assessment
Hcy	homocysteine
HIV	human immunodeficiency virus
HLP	Hyperlipoproteinaemia
IL6	interleukin 6
IDL	intermediate density lipoprotein
IOM	International Organization for Migration
IHD	ischaemic heart disease
IS	ischaemic stroke
LCAT	lecithin-cholesterol acyltransferase
LVH	left ventricular hypertrophy
LD	linkage disequilibrium
LPL	lipoprotein lipase
LSO	locus-specific oligonucleotide
LDL-C	low density lipoprotein cholesterol
LDL-R	low density lipoprotein receptors
MRC	Medical Research Council
T_m	melting temperature
MTHFR	methylene-tetrahydrofolate reductase
MAF	minor allele frequency
MZ	monozygotic
MUFAs	mono unsaturated fatty acids
MI	myocardial infarction
NCEP	National Cholesterol Education Program
NRF	National Research Foundation
NO	nitric oxide
NCD	non-communicable diseases
NWU	North-West University
OPA	oligonucleotide pool assay
OGTT	oral glucose tolerance test
PMPs	paramagnetic particles
PON1	paraoxonase 1
$PAI\text{-}1_{act}$	plasminogen activator inhibitor type-1
PAD	peripheral arterial disease
PEG	polyethylene glycol
PCR	polymerase chain reaction
PUFAs	polyunsaturated fatty acids
PURE	Prospective Urban and Rural Epidemiology study
PK	proteinase K
QFFQ	quantitative food frequency questionnaire
REACH	Reduction of Atherothrombosis for Continued Health
Rs	reference sequence
SFA	saturated fatty acids
SMAC	Sequential Multiple Analyser Computer
SNPs	single nucleotide polymorphisms
SEV	standard elution volume
±SD	standard deviation
SBP	systolic blood pressure
T	thymine
t-PA	tissue plasminogen activator
TC	total cholesterol
TE	total energy
THUSA	Transition and health during urbanisation of South Africans
TFA	trans fatty acid
TRL	triacylglycerol-rich lipoprotein
TG	triglycerides
TARF	Turkish Adult Risk Factor
UK	United Kingdom

UCSC	University of California Santa Cruz
VLDL-C	very low density lipoprotein cholesterol
VLDL	very low density lipoproteins
WGHS	Women's Genome Health Study
WHO	World Health Organization
WHR	waist-to-hip-ratio

LIST OF SYMBOLS AND UNITS

Symbols and units are listed in alphabetical order:

AU	arbitrary unit
Bp	base pair
B	Beta
χ^2	chi-square
©	copyright
R	correlation coefficient
↓	decrease
°C	degrees Celsius
Δ	delta
=	equal
ΔG	Gibbs free energy
G	gram
G	gravitational force
>	greater than
≥	greater than or equal to
↑	increase
Kb	kilo-base pairs
kDA	kilo Dalton
$kg.m^{-2}$	kilograms per meter squared; unit of body mass index
kJ	kilojoules
L	litre
M	meter
µ	micro
$\mu g.mL^{-1}$	micro grams per milli liters
Mg	milligram
mL	millilitre
mm Hg	millimeter of mercury
$mmol.L^{-1}$	milli mole per litre
x *g*	multiplied by gravitational force
$ng.\mu L^{-1}$	nanogram per microlitre
-	negative; minus
N	number of; sample size
%	percentage
±	plus minus
+	positive
P	p-value, indicates statistical significance
®	registered trademark
<	smaller than
≤	smaller than or equal to
~	tilde
™	Trade mark
$U.L^{-1}$	units per litre

LIST OF FIGURES

Figure 1.1 A graphic illustration of the main aim of the present study 6
Figure 2.1 The continuous set of pathophysiological events involved in CVD development 9
Figure 2.2 Global mortality and burden of disease characteristics of CVD and their major risk factors 12
Figure 2.3 CVD characteristic distributions by region in the REACH registry 15
Figure 3.1 Candidate genes contributing to CVD development 30
Figure 3.2 The lipid transport process via the different lipoproteins as a group 35
Figure 3.3 Influence of amino acid position 158 on LDL-R-binding activity 39
Figure 4.1 Maxwell® 16 DNA Purification Cartridges 55
Figure 4.2 Print screen of the UCSC Genome browser used to obtain the rs numbers for identified SNPs 61
Figure 4.3 Steps involved in the custom GoldenGate genotyping assay design 62
Figure 4.4 Flowchart of the steps involved in ordering a custom OPA 63
Figure 4.5 Location of the identified and selected SNPs in the ApoE gene 64
Figure 5.1 Photographic representation of the successful amplification of fragment one of the *ApoE* gene at 56°C 86
Figure 5.2 Representative electropherogram for the SNPs rs405509 and rs440446 identified in fragment one of the *ApoE* gene 87
Figure 5.3 Photographic representation of the successful amplification of fragment two of the *ApoE* gene at 67.2°C 88
Figure 5.4 Representative electropherogram for the SNP rs769449 identified in fragment two within the ApoE gene 88
Figure 5.5 Photographic representation of the successful amplification of fragment three of the *ApoE* gene at 53°C 90
Figure 5.6 Representative electropherogram for the novel SNP identified in fragment three in the *ApoE* gene 91
Figure 5.7 Photographic representation of the successful amplification of fragment four of the ApoE gene at 54.3°C 92
Figure 5.8 Representative electropherogram for the SNP rs769452 identified in fragment four in the *ApoE* gene 92
Figure 5.9 Photographic representation of the successful amplification of fragment five of the *ApoE* gene at 61.4°C 93
Figure 5.10 Representative electropherogram for the SNPs rs429358 and rs7412 identified in fragment five in the *ApoE* gene 94
Figure 5.11 GoldenGate® Assay contamination control as displayed in the GenomeStudio® Genotyping Module Controls Dashboard for the current study 100
Figure 5.12 GoldenGate® Assay allele specific extension control displayed in the GenomeStudio® Genotyping Module Controls Dashboard for the current study 101
Figure 5.13 GoldenGate® Assay PCR uniformity control displayed in the GenomeStudio® Genotyping Module Controls Dashboard for the current study 102
Figure 5.14 GoldenGate® Assay extension gap control displayed in the GenomeStudio® Genotyping Module Controls Dashboard 102
Figure 5.15 GoldenGate® Assay First Hybridisation Control displayed in the GenomeStudio® Genotyping Module Controls Dashboard in this study 103
Figure 5.16 GoldenGate® Assay Second Hybridisation Control displayed in the GenomeStudio® Genotyping Module Controls Dashboard in this study 104
Figure 5.17 Scatter plot of 10% GC Score compared to call rates of the investigated SNPs 105
Figure 5.18 Genomestudio® shade call regions for SNP rs769449 as an example for low cluster separation scores 106
Figure 5.19 Genomestudio® shade call regions for SNP rs769452 107
Figure 5.20 Controls used in the genotype analyses of the rs7412 SNP via the Bio Rad CFX Manager© (version 2.0) 108
Figure 5.21 GenomeStudio® shade call regions for SNP rs1081101 in the PURE dataset 111

Figure 5.22 Scatterplot of PAI-1_{act} against triglyceride concentrations categorised by rs1081101 114
Figure 5.23 Genomestudio® shade call regions for SNP rs405509 115
Figure 5.24 Significant genotype-environment interaction effects on TG concentrations for the three genotype groups of rs405509 119
Figure 5.25 Genomestudio® shade call regions for SNP rs440446 120
Figure 5.26 Genomestudio® shade call regions for SNP rs429358 123
Figure 5.27 The interactive effect of dietary MUFA intake with rs429358 on triglyceride concentrations for each of the genotype groups 127
Figure 5.28 The interactive effect of dietary TFA intake with rs429358 on triglyceride concentrations for each of the genotype groups 128
Figure 5.29 The interactive effect of dietary SFA intake with rs429358 on triglyceride concentrations for each of the genotype groups 129
Figure 5.30 The interactive effect of PAI-1_{act} concentrations with rs429358 on triglyceride concentrations for each of the genotype groups 130
Figure 5.31 Bio-Rad CFX Manager© allelic discrimination for SNP rs7412 131
Figure 5.32 The interactive effect of locality with rs7412 on triglyceride concentrations for each of the genotype groups 134
Figure 5.33 The interactive effect of dietary MUFA intake with rs7412 on triglyceride concentrations for each of the genotype groups 136
Figure 5.34 The interactive effect of dietary TFA intake with rs7412 on triglyceride concentrations for each of the genotype groups 137
Figure 5.35 The interactive effect of dietary fibre intake with rs7412 on triglyceride concentrations for each of the genotype groups 138
Figure 5.36 The interactive effect of GGT concentrations with rs7412 on HDL-C concentrations for each of the genotype groups 140
Figure 5.37 The interactive effect of GGT concentrations with rs7412 on TG concentrations for each of the genotype groups 141
Figure 5.38 The interactive effect of BMI with rs7412 on HDL-C concentrations for each of the genotype groups 143
Figure 5.39 The interactive effect of dietary cholesterol intake with the *ApoE* haplotype-genotypes on HDL-C concentrations for each of the genotype groups 150
Figure 5.40 The interactive effect of BMI with the *ApoE* haplotype-genotypes on HDL-C concentrations for each of the genotype groups 152
Figure 5.41 The interactive effect of PAI-1_{act} with the *ApoE* haplotype-genotypes on TG concentrations for each of the genotype groups 153

LIST OF TABLES

Table 2.1	Categorisation of CVD risk factors	12
Table 2.2	Biochemical values of the lipid profile for CHD risk assessment in Caucasian individuals	19
Table 3.1	Associations between the (additional) genetic variants and the risk factor and incidence rate of CVD	44
Table 4.1	Primers and PCR protocols used for the sequencing of the *ApoE* gene, per PCR fragment (1 to 6)	58
Table 4.2	Summary of the characteristics of the final SNPs in the ApoE gene sent for genotyping by the BeadXpress® T-Reader	64
Table 4.3	Real-time PCR temperature conditions	68
Table 5.1	Continuous baseline characteristics per gender and locality, within the PURE study population	73
Table 5.2	Categorical baseline characteristics per gender and locality, in the PURE study population	81
Table 5.3	Validation scores for the identified SNPs within the *ApoE* gene investigated in the PURE study	96
Table 5.4	Summary of the characteristics of the SNPs in the *ApoE* gene that were genotyped in black South African subjects	98
Table 5.5	IllumiCode Sequence IDs of the control analyses undertaken in the BeadXpress® system, along with the description and expected outcome of each	99
Table 5.6	Spearman Rank Order correlations between identified variables and the lipid profile in the current PURE cohort	109
Table 5.7	Variables confounded for per lipid profile determinant in the present cohort	110
Table 5.8	Adherence to the assumptions of Hardy-Weinberg Equilibrium of the rs1081101 locus in the present PURE population using chi-square test analyses	112
Table 5.9	Associations of the lipid profile determinant concentrations with rs1081101 locus in the present PURE population	113
Table 5.10	Adherence to the assumptions of Hardy-Weinberg Equilibrium of the rs405509 locus in the present PURE population using chi-square test analyses	116
Table 5.11	Associations of the lipid profile determinant concentrations with the rs405509 locus in the present PURE population	118
Table 5.12	Adherence to the assumptions of Hardy-Weinberg Equilibrium of the rs440446 locus in the present PURE population using chi-square test analyses	121
Table 5.13	Associations of the lipid profile determinant concentrations with the rs440446 locus in the present PURE population	122
Table 5.14	Adherence to the assumptions of Hardy-Weinberg Equilibrium of the rs429358 locus within the present PURE population using chi-square test analyses	124
Table 5.15	Associations of the lipid profile determinant concentrations with rs429358 locus in the present PURE population	125
Table 5.16	Adherence to the assumptions of Hardy-Weinberg Equilibrium of the rs7412 locus in the present PURE population using chi-square test analyses	132
Table 5.17	Associations of the lipid profile determinant concentrations with the rs7412 locus in the present PURE population	133
Table 5.18	Adherence to the assumptions of Hardy-Weinberg Equilibrium of the various haplotype genotypes in the present PURE population using chi-square test analyses	144
Table 5.19	Associations of lipid profile determinant concentrations with haplotype genotypes in the present PURE population	147
Table 5.20	Summary of all the significant results observed	156

LIST OF EQUATIONS

Equation 4.1 Measurement of DNA purity 54
Equation 4.2 Measurement of DNA concentration 54
Equation 4.3 Measurement of DNA yield 55
Equation 4.4 Measurement of DNA absorbance 56

ACKNOWLEDGEMENTS

This study would not have been attainable without the indispensible assistance of a number of people and institutions. My highest level of appreciation goes to the following individuals and institutions:

Dr Karin Conradie (my supervisor) for her devotion, time and determination in training and mentoring me in the specific laboratory skills and knowledge needed to complete my studies. All that I know regarding a nutrigenetics laboratory, I have learnt from her, along with many valuable lessons on how to complete the experimental part of a study successfully. Dr Wayne Towers (my co-supervisor), for his exceptional efforts and dedication in developing and mentoring me to acquire the analytic skills needed to complete a dissertation. Thank you both for sharing your experience and expertise to enable me to develop and continue as an academic and researcher. Dr Cornelie Nienaber-Rousseau (my co-supervisor), not only for her enthusiasm and willingness always to assist and share her scientific and nutritional knack throughout the entire study, but for the inspiration she is as a professional and friend.

Dr Suria Ellis, for her assistance in the statistical analyses conducted for this study. Prof. Annamarie Kruger, Prof. Edelweiss Wentzel-Viljoen and each individual involved in the collection and analysis of the PURE 2005 baseline data. The NRF[1] for funding this study, North West University for the bursaries awarded for the duration of my study as well as the personnel of the Ferdinand Postma Library, with special thanks to Mrs Anneke Coetzee, for assistance with interlibrary loans. Mrs Barbara Bradley, for her language and editing skills.

My parents (for making it possible to further my studies) along with the rest of the family for encouraging me throughout my studies and for being my haven where I could recover my driving force. Thank you for your unconditional love and support. Hannes Meades, for acknowledging my aspiration to further my studies and to excel therein and keeping me motivated to the end. You are my best friend and I will always be grateful to you. Bianca Swanepoel, for her reassuring and uplifting personality and her willingness to share her thoughts and knowledge. Thank you for sharing this academic journey with me. I will

[1] Disclaimer: Any opinion, findings and conclusions or recommendations expressed in this material are those of the author and therefore, the NRF does not accept any liability in regard thereto.

forever cherish our friendship. Thank you all for your love and comfort throughout this study.

Most of all, the Lord who has not only given me the privilege to further my studies, but has held my hand throughout my life's journey, guiding me, encouraging me and most of all, developing me into what I can become:

"What you are is God's gift to you, what you become is your gift to God."

CHAPTER ONE

Introduction

This dissertation investigates specific factors that have been reported to contribute to cardiovascular disease (CVD) risk development. The introductory chapter defines the research problem this work aims to address and states the hypothesis. In addition, to enhance clarity, certain operational definitions for terms used in this work will be discussed in this chapter.

1.1 BACKGROUND AND MOTIVATION

The World Health Organisation (WHO) anticipates that CVD will soon be the leading cause of death globally, to such an extent that by 2030 it is predicted that almost 23.6 million people will die of these disorders. CVD can be defined as a set of disorders that affect the proper functioning of blood vessels as well as the heart (WHO, 2011). In the past, cardiology [referring to the study of the anatomy, normal functions and disorders of the heart (Harris *et al*., 2006)] focussed research initiatives on the myocardium (cells forming the bulk of the heart wall). However, it seems as if the myocardium plays only a small role and that the real problem pertains to the blood vessels (Libby, 2003). Moreover, it is important to distinguish between CVD and coronary heart disease (CHD), as these two terms are often used interchangeably. The latter, also known as coronary artery disease, (but referred to as CHD in the present dissertation for consistency) is an abnormal condition that may affect the arteries of the heart and produce various pathological effects *i.e.* reducing the oxygen and nutrient flow to the myocardium (Harris *et al*., 2006). However, CVD encompasses CHD and other disorders including ischaemic heart disease (IHD) and cerebrovascular disease (CBVD) among others.

It is important to be aware of the key drivers of this disease. Overweight and obesity, high blood pressure and cholesterol are all nutrition-related cardiovascular risk factors that are among the leading causes of CVD mortality and morbidity in the world (Ezatti *et al*., 2005; Yusuf *et al*., 2001a; Reddy & Yussuf, 1998). Most of the increased incidence of CVD can be attributed to increased exposure to the various risk factors, mainly lifestyle changes, *i.e.* dietary changes, increases in body weight, a decrease in physical activity and increasing tobacco use and alcohol consumption (Beaglehole, 2001; Walker & Sareli, 1997). In

addition, when the life expectancy of an individual increases, the period of exposure to CVD risk factors will also increase; therefore, the increase in global longevity could result in a higher prevalence of CVD (Mayosi *et al*., 2009; Reddy & Yussuf, 1998).

It has been reported that urbanisation is increasing dramatically in developing countries together with an increase in adverse lifestyle behaviours that increase the incidence of obesity and diet-related non-communicable diseases (NCDs) (Popkin, 2006; Yach *et al*., 2004; Shetty, 2002; Beaglehole, 2001). Furthermore, it has been reported that more than 80% of global CVD deaths (affecting men and women equally) occur in low- and middle-income countries (WHO, 2011). This emphasises the quandary that while the incidence of CVD is decreasing in developed countries, this incidence is increasing in developing countries (Mackay *et al*., 2004). For the purpose of this dissertation, South Africa will be regarded as a developing country when compared to first world countries. The reason for highlighting this is that if one compares South Africa to other African countries, South Africa is regarded as a developed country. However, South Africa contains areas that can be considered as developed and others that are regarded as developing (http://www.afesis.org.za/Sustainable-Settlements-Articles/beyond-the-rural-and-urban-development-frameworks-theorising-the-relationship, 10 October 2013).

In South Africa, NCDs are estimated to account for 29% of all deaths, of which CVD contributes as much as 11% (WHO, 2011). It has been well reported that the socio-economic background (psychosocial factor) of many black South African populations is changing rapidly from rural, traditional African lifestyles to westernised diets and lifestyle behaviours, typified by a high fat intake, a low carbohydrate and fibre intake and sedentary behaviour (Alberts *et al*., 2005; Vorster, 2002; Yusuf *et al*., 2001b). This phenomenon is known as the nutrition transition and its associations with chronic diseases has been investigated and well documented in South Africa. However, the true impact of this phenomenon is not recognised in all parts of Africa (Popkin, 2006; Vorster *et al*., 2005; Vorster, 2002; Vorster *et al*., 1999; Steyn *et al*., 1991).

Various dietary factors have been associated with CVD development and its incidence at community level, with trans fatty acid (TFA) and saturated fatty acid (SFA) intake being the strongest dietary factors influencing CVD risk (De Backer, 2008). TFA not only increases low-density lipoprotein cholesterol (LDL-C) concentrations, as does SFA, but decreases high-density lipoprotein cholesterol (HDL-C) as well, making it the most adverse dietary

factor affecting the lipid profile and ultimately CVD development (Brouwer *et al*., 2010; Chardigny *et al*., 2008; Mozaffarian & Willet, 2009; Micha & Mozaffarian, 2008).

Other dietary factors that have also been reported to play a role in CVD development are the fibre content of a diet, *i.e*. the inclusion of whole-grain products, fruit and vegetables, the amount of refined and processed sugars and the amount of salt consumed (De Backer, 2008). Unhealthy dietary habits are reported as primary CVD risk factors, which result in elevated total cholesterol (TC), LDL-C and triglyceride (TG) concentrations, as well as low HDL-C concentrations, which are reported to be secondary risk factors (Bersamin *et al*., 2008).

Furthermore, it is of relevance to investigate the underlying causes and factors that contribute to the development of CVD in different populations, as it has been well reported that CVD risk differs between ethnic groups (Holvoet *et al*., 2007; Forouhi & Sattar, 2006). Throughout this dissertation the term “ethnicity” will be used and not “nationality” or “race” to reflect the cultural traditions, environmental exposure and common history shared by a group of people along with certain genetic characteristics due to various reproductive patterns (Burchard *et al*., 2003). Since ethnicity encompasses genetic and cultural factors, it could influence the risk of a specific group of people in the development of CVD (Hernandez & Blazer, 2006) and should be taken into account in studies such as the work presented in this dissertation.

For a number of years most studies on CVD were conducted on Caucasian populations and a limited number of studies investigated other ethnic groups (Steyn *et al*., 2005; Yusuf *et al*., 2004). However, this gap was recognised and researchers started to conduct multi-ethnic studies such as the Black Risk (BRISK) study, the INTERHEART (global and Africa) studies and the Reduction of Atherothrombosis for Continued Health (REACH) registry (Bhatt *et al*., 2006; Steyn *et al*., 2005; Steyn *et al*., 1991). The prevalence of CHD among black South African populations has generally been low because of the reportedly favourable lipid profile of this population, *i.e*. presenting with relatively low LDL-C and high HDL-C concentrations (Alberts *et al*., 2005; Walker & Sareli, 1997). Yet, the incidence of cardiovascular events in the black South African population has increased over the last few years (Tibazarwa *et al*., 2009). This emphasises the importance of studying CVD risk factors (*i.e.* lipid profile) and mechanisms contributing to this epidemic (Masemola *et al*., 2007; Alberts *et al*., 2005).

Cardiovascular events also occur in individuals in whom the traditional risk factors for CVD are not present (Ridker *et al*., 2004; Greenland *et al*., 2003), thus highlighting the relevance and importance of considering CVD as a multi-factorial disease. A multi-factorial disease is defined as a disease that is the consequence of the interaction of multiple pathogenic mechanisms, including various genes, acting either alone or with one another, with or without the involvement of environmental factors (Harris *et al*., 2006; Gelehrter *et al*., 1998). In order to understand multi-factorial diseases such as CVD, it is necessary to analyse any clinically important interactions between major environmental or phenotypic factors and the genetic predisposition of an individual (Casas *et al*., 2006). A multi-factorial approach needs to be followed in order to understand the factors contributing to the individual lipid profile response to dietary interventions and the prescription of more successful and beneficial health solutions (Ordovas, 2006; Ordovas & Mooser, 2004). An example of such an approach is nutrigenetics, a research field investigating gene-diet interactions.

Since the major risk factors of CVD include the markers of the lipid profile, it is important to investigate factors that influence the lipid profile using a multi-factorial approach. The apolipoprotein E (*ApoE*) gene is the best-studied candidate gene in the lipid field; it is located on human chromosome 19q13.2 and encodes a 34 kilo Dalton (kDa) glycosylated polymorphic protein (Siest *et al.,* 1995; Davignon *et al*., 1988; Lusis *et al.,* 1986). This gene participates in the transport and metabolism of plasma cholesterol and TG (Mahley & Huang, 1999). Three common and frequently reported protein alleles exist (ε2, ε3, ε4) and each allele encodes a single isoform, respectively: ApoE-2, ApoE-3 and ApoE-4 (Frikke-Schmidt *et al*., 2000). These alleles are co-dominantly inherited and therefore six *ApoE* haplotype-genotypes commonly arise namely, ε2/ε2, ε2/ε3, ε3/ε3, ε3/ε4, ε4/ε4 and ε2/ε4 (Zannis *et al.,* 1982). Even though the literature refers to these genetic variants as the ε2, ε3, ε4 *alleles*, it is important to highlight that they result from the strong linkage disequilibrium (LD) between two common non-synonymous single nucleotide polymorphisms (SNPs) and, therefore, should in actual fact be referred to as *haplotypes*, as in the present study. However, in the present study, various other allelic variations, in the form of SNPs in the *ApoE* gene, were also identified and studied. These *ApoE* polymorphisms were selected based on sequence analysis of individuals from the black South African population in order to find a possible explanation for the ethnic variations that could predict future CVD events, owing to differences in allele frequencies among different ethnicities (Ioannidis *et al.*, 2004; Gerdes*,* 2003; Stengård *et al.,* 1998).

Little data are available on the relationship between *ApoE* polymorphisms and blood lipid concentrations in the black South African population (Masemola *et al*., 2007). These data are important to obtain as it is hypothesised that the genetic impact of *ApoE* polymorphisms on the plasma TC variance is one of the most powerful genetic components in the regulation of cholesterol concentrations at population level (Davignon *et al.*, 1988).

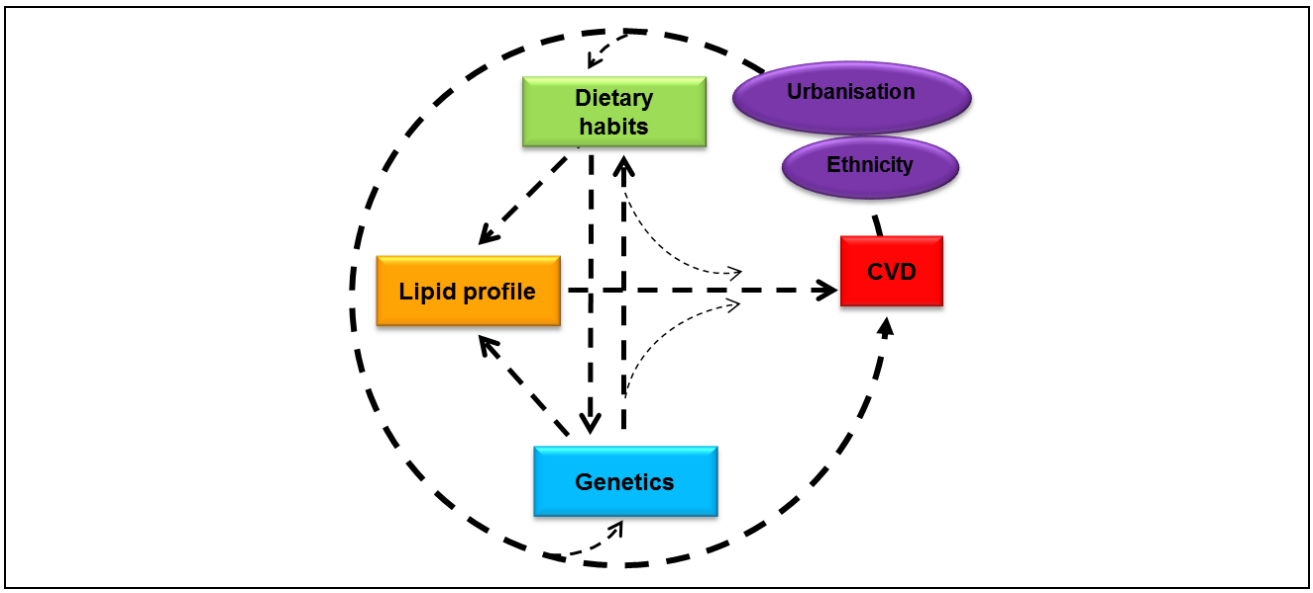
In the present study, which is nested within the South African arm of the Prospective Urban and Rural Epidemiology (PURE) study, approximately 2 000 samples were analysed at a genotypic, dietary, lifestyle and epidemiological , enabling the detection of associations between these polymorphisms and the lipid profile. The present study has the potential to fill a gap in the literature regarding a detailed genotype frequency description of the *ApoE* gene of a black South African cohort and the elucidation of any gene-environmental interactions that may affect the lipid profile and consequently CVD risk of the black South African population.

From all the above-mentioned, one can infer that it is valuable to investigate the predisposing effect of various lifestyle factors, in congruence with genetic aspects, involved in CVD development, which is currently emerging in the black South African population. In this way a multi-factorial disease, such as CVD, can be understood better and it can be determined whether these interactions play a causal role in the increased susceptibility to CVD in the black South African population.

1.2 RESEARCH AIMS AND OBJECTIVES

The aim of the larger PURE study is to investigate population-specific factors such as genetic, environmental and lifestyle factors, including nutrition, physical activity, tobacco use, alcohol consumption and epidemiological transition, and determining whether these factors are associated with an increased risk of NCDs, such as CVD (Teo *et al.*, 2009). A visual representation giving a holistic view of the main aim of the work represented in this dissertation is given in Figure 1.1. The main aim of the present dissertation was to investigate the effect of reported as well as novel SNPs within the *ApoE* gene on the lipid profile in black South Africans, from rural and urban areas, and to explore possible gene-environment interactions that may occur (Figure 1.1).

Figure 1.1 A graphic illustration of the main aim of the present study



ApoE = apolipoprotein E; CVD = cardiovascular disease; SNPs = single nucleotide polymorphisms

The specific objectives were to:

a. determine the genotype distributions of the various reported as well as novel alterations within the *ApoE* gene in a black South African population;
b. assess whether the identified polymorphisms in the *ApoE* gene are associated with identified risk biomarkers (*i.e*. the lipid profile) of CVD, in a black South African population;
c. explore the possible interaction effects of environmental factors (*i.e*. locality) with specific *ApoE* genetic variations on the lipid profile of black South Africans;
d. explore possible genotype-phenotype [*e.g*. genotype-body mass index (BMI)] interaction effects on the lipid profile of black South Africans, and
e. explore the possible interaction effects of specific dietary factors (*i.e*. dietary fat intake and alcohol consumption) with specific *ApoE* polymorphisms on the lipid profile of black South Africans.

1.3 CHAPTER OUTLINE

This dissertation is presented in chapter format. It was technically edited in the style required by the North-West University (NWU), and has been edited by a language editor. This introductory chapter is followed by two comprehensive exploratory literature reviews. Chapter two is a literature review investigating the conventional risk factors of CVD and elaborates on the contributing role of ethnicity (a black South African population) and environmental factors (*i.e*. dietary habits, epidemiological transition) on the lipid profile. The second literature review (Chapter three) is based on the molecular genetics of CVD,

focusing specifically on the genetic variants identified in the *ApoE* gene and their effect on the lipid profile. Furthermore, the *ApoE* gene's effect on CVD development is discussed in Chapter three, with specific reference to the polymorphisms identified in the present cohort.

Chapter four describes the PURE study design along with the methods used for participant enrolment. All experimental methods used in the present study are explained in detail in this chapter, as well as the anthropometric measurements, dietary intake analysis, various biochemical analyses and statistical analyses undertaken.

Chapter five presents the discussion of the results obtained and comparisons thereof with the available and relevant literature. Chapter five includes the baseline characteristics of the PURE study population. Furthermore, this chapter is divided into two main sections, the first pertaining to the population baseline characteristics and the second to the genetic characteristics. Each of the investigated SNPs is discussed separately, focusing on the implications of the significant associations and interactions observed. Chapter five concludes with a summary of the major findings.

The final chapter of this dissertation are the concluding chapter. The concluding chapter highlights novel observations made in the study, examines the limitations of this study and makes recommendations for future research.

CHAPTER TWO

The influential and interactive role of various risk factors on CVD development

The wealth of Africa lies in the diversity of the populations inhabiting this continent. South Africa is a country of ethnic diversity with people from all over Africa inhabiting it as a result of migrations from other countries (Schuster *et al*., 2010). This diverse population consists of a number of different ethnic groups that have their origins in Asia, Africa and Europe.

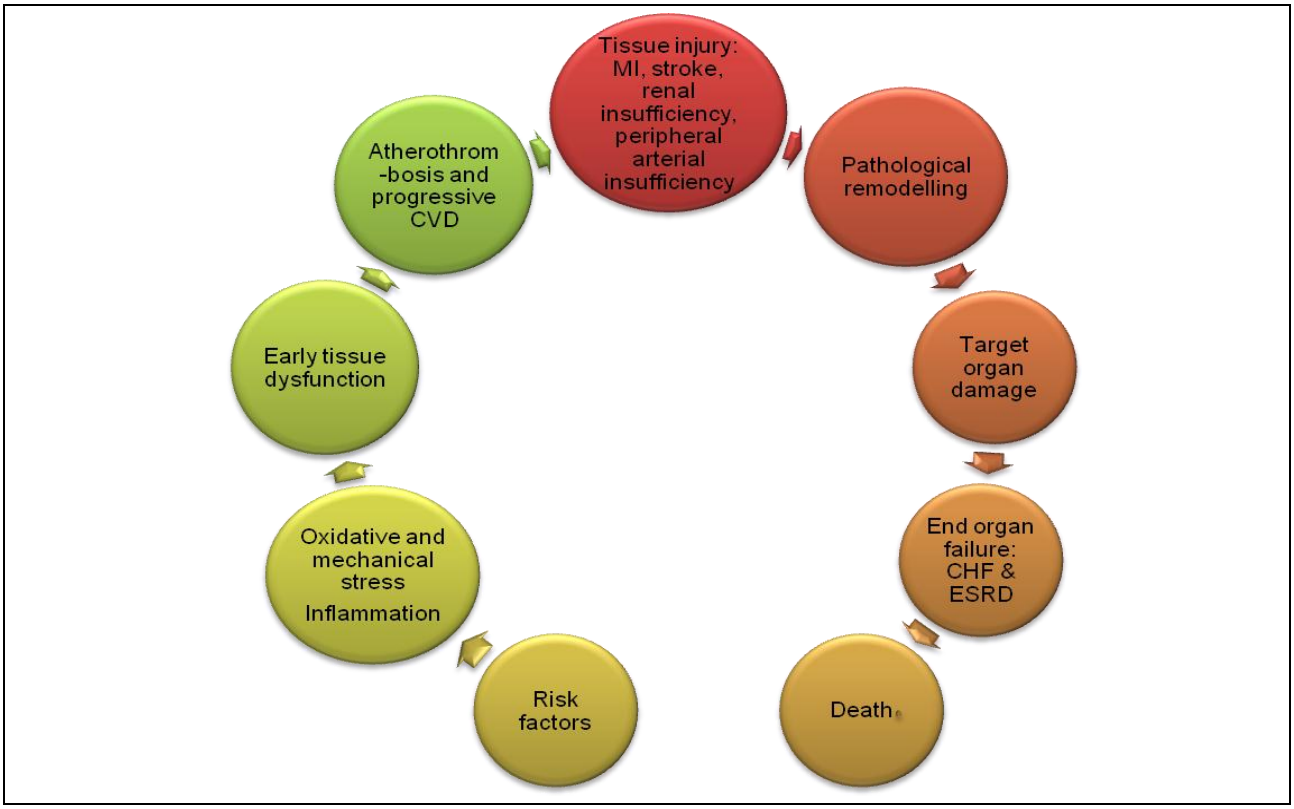
The socio-economic environment of many black South Africans, in particular, is changing rapidly from the more rural, traditional African lifestyles to more westernised lifestyles with the adoption of more imprudent dietary patterns (Steyn *et al*., 2005; Mackay *et al*., 2004; Walker & Sareli, 1997). Most industrialised countries follow Westernised lifestyles, yet the CVD incidence is decreasing in these countries in response to the implementation of prevention programs targeting the control of risk factors (Okrainec *et al*., 2004). Contrary to this, the CVD incidence as well as the incidence of many other chronic lifestyle diseases (CLD) are increasing in developing countries and can be expected to escalate in the future among black South African populations, as the movement from rural to urban areas increases (Masemola *et al.*, 2007; Mackay *et al.,* 2004; Steyn *et al*., 1991). Reasons other than urbanisation can also contribute to the increased CVD incidence among (black) South Africans e.g. the genetic makeup of (black) South Africans which is a research field that has not been extensively studied to date.

2.1 CARDIOVASCULAR DISEASE

CVD is a condition typified by abnormal or impaired functioning of the heart and/ or blood vessels (Harris *et al*., 2006). These disorders include: atherosclerosis, CBVD, congenital heart disease, deep vein thrombosis, peripheral arterial disease (PAD), pulmonary embolism, rheumatic heart disease and systemic venous hypertension (Harris *et al*., 2006; Forrester, 2004). Atherosclerosis is a major health concern in both developed and developing countries making it a global health burden. Furthermore, atherosclerosis is the devastating underlying condition in patients who develop CHD, myocardial infarction (MI), heart failure, PAD and stroke (Paré *et al*., 2007; Dzau *et al*., 2006).

CHD and stroke are within the top ten leading causes of mortality worldwide (WHO, 2011). This finding is responsible for the sense of urgency that has arisen, to investigate the reasons for the high incidence rate of CHD. Investigation into the pathophysiological CVD continuum over the last two decades has resulted in the determination that CVD incidence is initiated by specific risk factors, which in turn instigate the processes that result in tissue damage, leading to the pathogenicity of these disorders (Dzau *et al*., 2006). This is in accordance with the reality that CHD is also in fact one of the most evident consequences of the high prevalence of nutrition-related risk factors in the population globally (Hawkesworth *et al*., 2010). Figure 2.1 represents a continuum indicating the progression of CVD and a number of additional metabolic pathways, which form part of the pathophysiological CVD continuum, including oxidative stress, early endothelial tissue dysfunction, inflammatory progressions and vascular remodelling in the initiation and prolongation of atherosclerotic disease.

Figure 2.1 The continuous set of pathophysiological events involved in CVD development



CHF = congestive heart failure; CVD = cardiovascular disease; ESRD = end-stage renal disease; MI = myocardial infarction; adapted from Dzau *et al*. (2006)

The most important underlying pathological process for CVD is atherosclerosis. Growing evidence associates lipid metabolism with atherosclerosis development, as modified or oxidised LDL-C has been identified as the best recognised initiating event of atherosclerosis (Ordovas, 2009; De Caterina *et al*., 2006). The accumulation of

lipoproteins in the vessel wall (hypercholesteroleamia) is one of the consequences of a damaged endothelium (Schachter, 1997). In time, these lipoproteins can become modified by chemical acetylation, resulting in oxidised species of LDL-C that result in the attraction of monocytes and lymphocytes to the vessel wall, where monocytes subsequently convert into macrophages. The macrophages recognise and take up the oxidised LDL-C particles through scavenger receptors. This in turn gives rise to cholesterol-loaded "foam cells", which are the trademarks of early atherosclerotic lesions, termed "fatty streaks" (Lusis *et al*., 2003; Schachter, 1997). The foam cells eventually die and the presence of inflammatory cells triggers the migration of smooth muscle cells from the media to the intimae and in turn the proliferation of the smooth muscle cells in the intimae. These cells secrete collagen and form an apparently protective "fibrous cap" overlaying the lesion (Lusis, 2003; Schachter, 1997). MI is usually caused by rupture or erosion of such lesion(s), leading to the formation of a thrombus, which results in either the stenosis or occlusion of the vessel (Lusis, 2003; Vorster *et al*., 1998; Reilly & Cawley, 1996). Human lesions differ widely in composition and there are undoubtedly multiple processes that can result in clinically significant events.

In essence, CVD can be described as an adverse effect that arises in response to physiological processes not being in tolerable homeostasis, since the slightest disruption (initiated by risk factors) of the normal functions of endothelial cells can stimulate pathological vascular responses such as smooth muscle cell proliferation, vaso-constriction, inflammation and thrombosis (Dzau *et al*., 2006). The undesirable presence of these phenotypic traits could possibly be as a result of genetic variations, environmental exposure or an imbalance between these factors (Stephens & Humphries, 2003).

Over the years, many epidemiological and interventional studies have identified specific factors associated with increased risk of CVD, which resulted in establishing scientific guidelines and risk assessment models that can be applied to individuals and populations to reduce the risk of disease development (Ordovas, 2006). A well-used method in identifying individuals at high risk of CVD is to establish whether the traditional risk factors are present (Wilson, 2000). Almost 70% of all individuals at risk of CVD have multiple risk factors, which interact concurrently to increase the total risk of CVD, where the risk is about four times higher in the presence of one risk factor to 60 times higher in the presence of five risk factors (Wilson, 2000). However, even when one obtains detailed information on the environmental exposure and genetic makeup of an individual, the onset,

development and seriousness of the disease cannot be predicted with certainty owing to a lack of a comprehensive understanding of these multi-factorial diseases (Sing *et al*., 2003). It is in the interaction between these factors that the determinants of the phenotypic outcome of an individual possibly lies (Sing *et al*., 2003). It is important that the conventional CVD risk factors be investigated (Section 2.1.1), but it is even more important to be conscious of the interactive effects between these traditional factors and novel risk factors such as genetic susceptibility, which will be discussed in Chapter three.

2.1.1 Conventional CVD risk factors

According to Hopkins and Williams, 246 (risk) factors were associated with CVD in 1981 (Hopkins & Williams, 1981). A large number of clinical trials have been conducted in the last two decades with the aim of increasing the understanding of the causes and contributing factors of diseases, while facilitating the identification of new risk factors, which could possibly improve the methods for identifying persons who are in the early stages of CVD or at high risk of developing it (Ware, 2006). In 2004, the WHO stated that more than 300 risk factors for CVD had been identified (Mackay *et al*., 2004). Some of the risk factors, that have been identified and consequently added to the list of reported CVD risk factors, as they might assist in improving CVD risk prediction, include abnormal blood coagulation [markers of blood clotting *i.e.* fibrinogen and plasminogen activator inhibitor type 1 (PAI-1)], inflammation (inflammatory markers *i.e.* elevated CRP concentrations) and raised homocysteine (Hcy) concentrations (Mackay *et al*., 2004). The increase in the number of risk factors identified emphasises not only that awareness of CVD has increased, but also that scientific knowledge of CVD in terms of its aetiology, treatments and “novel” risk factors has expanded.

Some of these risk factors can be categorised as major modifiable, other modifiable and non-modifiable risk factors, as indicated in Table 2.1 (Mackay *et al*., 2004). Both non-modifiable risk factors (genetic predisposition to the disease, family history, age and gender) and ones that can be modified by lifestyle interventions (obesity, hypertension, hypercholesterolaemia, diabetes, alcohol consumption, atherogenic diet, lack of physical activity and tobacco use) enhance CVD incidence rates (Campbell *et al*., 2008; Mackay *et al*., 2004).

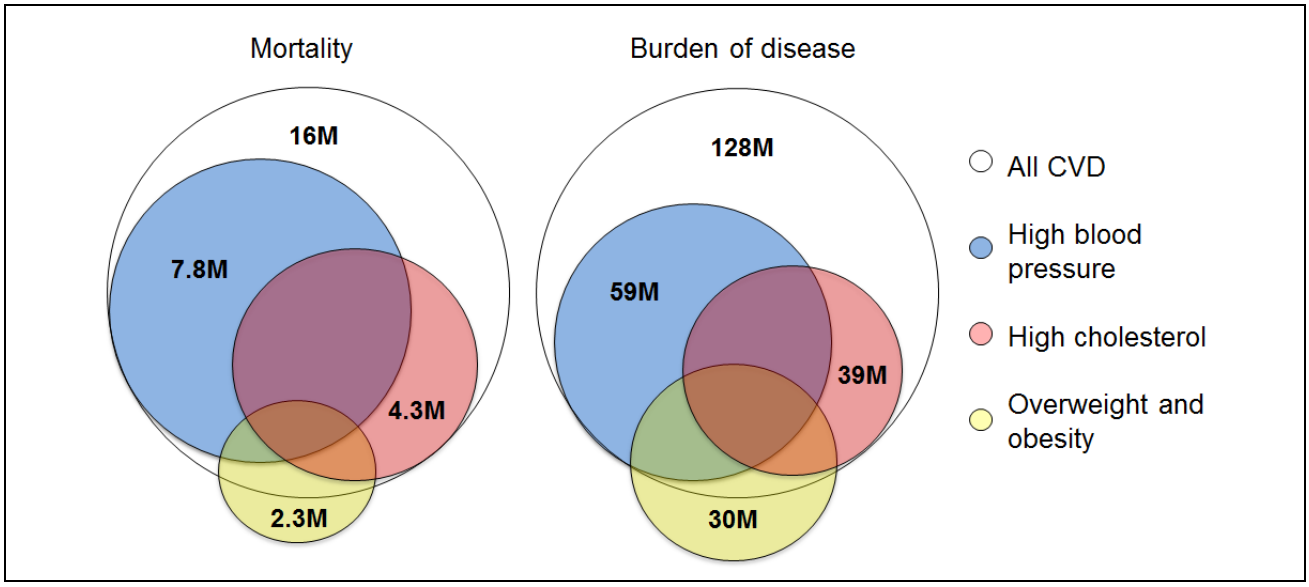
Table 2.1 Categorisation of CVD risk factors

Major modifiable risk factors	Other modifiable risk factors	Non-modifiable risk factors
Elevated blood pressure	Alcohol use	Genetics or family history
Abnormal blood lipids	Use of certain medicine	Ethnicity
Tobacco use	Elevated serum lipoprotein (a)	Gender
Physical inactivity	LVH	Advancing age
Obesity	Psychosocial stress	---
Unhealthy diets	Psychosocial factors (urbanisation, poverty *etc*.)	---
Diabetes mellitus	Mental illness	---

LVH = Left ventricular hypertrophy; adapted from Mackay *et al*. (2004)

The major risk factor for CVD is BMI, as it is an indicator of the risk of overweight or obesity, altered blood lipid determinants, diabetes and hypertension (Litwin, 2008; Ezzati *et al*., 2005; Cleeman *et al*., 2001) as seen in Figure 2.2. The total global mortality estimates have indicated that elevated blood lipid profile concentrations, tobacco use and hypertension are reported as the top three causes of death in developed countries and are also the major constituents of a high-risk profile for CVD (Ezzati *et al*., 2005; Ezzati *et al*., 2003; Ezzati *et al*., 2002).

Figure 2.2 Global mortality and burden of disease characteristics of CVD and their major risk factors



CVD = cardiovascular disease; M = million; adapted from Ezzati *et al*. (2005)

Over the years, the influential role of non-modifiable risk factors in CVD risk development has also been studied extensively. Advancing age is the most influential independent risk factor for CVD; the risk of stroke doubles each decade after the age of 55 (Ezatti *et al*., 2005; Mackay *et al.*, 2004), possibly as a result of the increased degree and duration of exposure to CVD risk factors. Furthermore, the absolute risk of CVD increases

with age among all individuals of a population and, therefore, relative risk is often used as a more sensitive metric in an elderly population as it is based on only the major risk factors (Grundy *et al*., 1999). However, with advancing age the impact of certain CVD risk factors, such as dyslipidaemia, diabetes and fibrinogen, is reduced (Navas-Nacher *et al*., 2001).

Many research projects have been conducted in order to understand the differences and similarities between females and males at the population level (*i.e.* behaviours, lifestyles and environment), the individual level and the cellular and molecular levels (Hernandez & Blazer, 2006). According to Mackay and co-workers (2004), higher CHD rates occur among men compared to premenopausal women. However, the risk increases significantly for women after the protective effect of oestrogen is lost (de Backer, 2008). Nonetheless, the risk of developing stroke has been reported to be similar for both genders (Mackay *et al.,* 2004).

Another non-modifiable risk factor of CVD that is of interest to the present study is ethnicity, since the occurrence of different CHD mortality and morbidity rates within and between populations has been observed (Dahlöf, 2010; Bhatt *et al*., 2006; Mutch *et al*., 2005). This is most likely due to the different populations that are under investigation in any particular study, being not only genetically diverse, but also being exposed to different contributing environmental risk factors (Anand *et al*., 2009; Sing *et al*., 2003). The following section will discuss the effect of ethnicity, as a non-modifiable risk factor on CVD development, in greater detail.

2.1.2 Ethnicity as a non-modifiable risk factor for CVD development

Most of the knowledge available on the risk factors of CVD has been collected from studies conducted mainly among European populations (Forouhi & Sattar, 2006; Steyn *et al*., 2005; Yusuf *et al*., 2004; Yusuf *et al*., 2001a). The global scientific field has become increasingly uncertain as to the extent to which the existing data available for CVD risk factors and prevention strategies are applicable to the rest of the populations across the globe and, therefore, the era of multi-ethnic studies has begun (Yusuf *et al*., 2004; Keil *et al*., 1993).

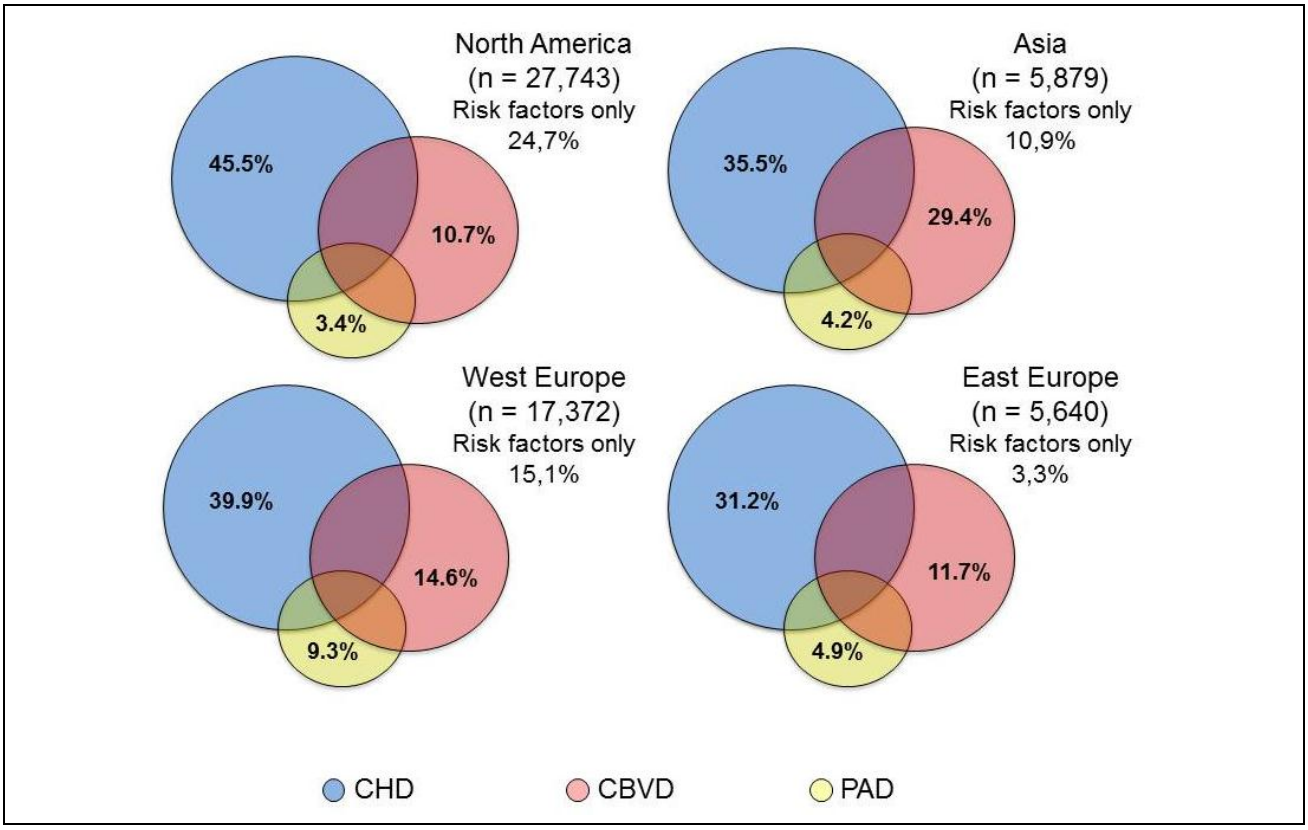
One of the first multi-ethnic studies to clarify whether the effects of risk factors do vary among different ethnic populations from 52 countries (located in Africa, Asia, Australia, Europe, the Middle East, North America and South America), was the INTERHEART

global study (Yusuf *et al*., 2004). The INTERHEART global study also investigated the importance of the associations between reported and easily measured risk factors (tobacco use, lipids, self-reported hypertension and diabetes, obesity, diet, physical activity, alcohol consumption and psychosocial factors) and CHD on the population attributable risks (PAR) for acute MI (AMI) (Yusuf *et al*., 2004). Yusuf and co-investigators (2004) reported that when the population was subdivided by ethnicity, the investigated risk factors accounted for a very high proportion of the PAR in every ethnic group: Europeans, 86%; Chinese, 90%; Latin Americans, 90%; South Asians, 92%; black Africans, 92% and Arabs, 93%, indicating consistency across a number of ethnic, gender and age groups. However, even though the risk factor incidence among the populations investigated by the INTERHEART study were similar, the CVD incidence among these ethnic groups differed, which suggests that other influential factors, such as genetics, determine the risk in populations or ethnicities (Cargill *et al*., 1999; Halushka *et al*., 1999; Wang *et al*., 1998; Li & Sadler, 1991). In addition, various other studies have reported a higher CVD burden among South Asians, African-Caribbeans [usually based in the United Kingdom, (UK)], African-Americans and Mexican Americans when compared to European Caucasians (Forouhi & Sattar, 2006; McKeigue *et al*., 1993). The higher prevalence of CVD can possibly be attributed to the fact that different ethnic groups are predisposed to developing CVD at different rates (Forouhi & Sattar, 2006), possibly because of different economic advances or genetic differences, as mentioned before (Thorogood *et al*., 2007; Cargill *et al*., 1999; Halushka *et al*., 1999; Wang *et al*., 1998; Li & Sadler, 1991). Although the problem of CVD may be pandemic, the aforementioned results signify the possibility that CHD prevention strategies can be based on similar strategies across multiple ethnic groups around the world (Bhatt *et al*., 2006; Yusuf *et al*., 2004).

The REACH registry is another multi-ethnic study, which was initiated in 2003 with the aim of investigating the impact of traditional and novel risk factors on the prevalence of cardiovascular ischaemic events among patients with (or with a high risk of) atherothrombosis (Ohman *et al*., 2006). The REACH registry collected global data from more than 44 countries across six regions (Asia, Australia, Europe, Latin America, the Middle East and North America) on atherosclerosis risk factors from 67 888 middle-aged patients (Ohman *et al*., 2006). The results reported confirmed the findings from the INTERHEART global study in that classic cardiovascular risk factors (hypertension, high cholesterol concentrations, diabetes, obesity and tobacco use) are consistent and common in various ethnic populations, even if they do tend to be undertreated and less controlled in many regions of the world (Bhatt *et al*., 2006). Furthermore, the REACH

registry determined differences in the distribution of CVD characteristics in the different categories of CVD, *i.e.* CHD, CBVD, and PAD, among different regions (ethnicities), as indicated in Figure 2.3 (Bhatt *et al*., 2006; Dahlöf, 2010).

Figure 2.3 CVD characteristic distributions by region in the REACH registry



CHD = coronary heart disease; CBVD = cerebrovascular disease; PAD = peripheral arterial disease; n = number of subjects; % = percentage; risk factors = ≥3 atherosclerosis risk factors; adapted from Dahlöf (2010)

From Figure 2.3 it is clear that the risk factor incidence is the lowest in the East European region (3.3%), yet the CHD incidence (31.2%) is comparable to that of the Asian region (35.5%), despite the risk factor incidence in the Asian region being almost threefold higher (10.9%). The comparison of PAD between these two regions results in similar observations as determined in the case of CHD. This suggests that the incidence of CHD and PAD is possibly a result of contributing factors other than the risk factor prevalence measured by the REACH study (Dahlöf, 2010). The CVD incidence for the Asian region is also almost threefold that of the East European region, suggesting that CVD can possibly be more strongly determined in the Asian population by the risk factor prevalence, since the clustering and the interaction of risk factors in individuals increase CVD risk (Dahlöf, 2010). Yet, when comparing these two regions with the highest risk factor prevalence for the Northern American and Western European regions, a totally different combination of disease prevalence *versus* risk factor prevalence is observed compared to the regions with the lower reported percentages (Dahlöf, 2010). Even though the risk

factor prevalence is 24.9% in the North American region, the CVD incidence is only 10.9%, whereas these statistics are 15.1% and 14.9% in the Western European region for risk factor and CHD rates, respectively (Dahlöf, 2010). Both the risk factor prevalence and CHD incidence are highest in the North American region; however, for the other regions the CHD incidence is also high, but the risk factor prevalence is lower (Dahlöf, 2010). The fact that the incidence of the disease differs between the four regions, irrespective of the risk factor prevalence, supports the previous suggestion that other factors are also at play in determining the incidence of the disease. It is possible that the risk within populations is mainly determined by environmental factors, while the risk among populations or ethnicities is generally determined by genetic components (Forouhi & Sattar, 2006).

Even though this registry is geographically diverse, this study did not include an important ethnic group (especially to the present study) *i.e*. the African population (Bhatt *et al*., 2006). Fortunately, other studies have investigated the CVD burden of disease prevalence among populations of African descent (Zoratti *et al*., 1998; Cappuccio, 1997; Walker & Sareli, 1997), which have been reported to present with lower rates of CHD regardless of their higher rates of insulin resistance, diabetes, hypertension and stroke, while maintaining a healthy lipid profile (Forouhi & Sattar, 2006; Zoratti *et al*., 1998). Since the present study is being conducted in a (black South) African population, it is critical to investigate the discoveries and progress that have been made in this ethnic group as well.

Over a decade ago, Walker and Sareli (1997) reported on numerous studies conducted before the 1970s, which investigated the prevalence and incidence of CVD among Africans. The conclusion was that the prevalence of CVD risk factors among black Africans was low and the occurrence of CVD (especially CHD) in this ethnic group was almost unheard of (Walker & Sareli, 1997; Steyn *et al*., 1991). However, a decade later scientists started reporting that the CVD mortality rate for Africans was almost that of the Caucasian populations (Loock *et al*., 2006; Steyn *et al*., 2005; Walker & Sareli, 1997; Keil *et al*., 1993; Steyn *et al*., 1991).

When comparing the incidence of CVD among different African populations, controversial results are reported. Zoratti (1998) reported a reduced incidence of CHD among the Afro-Caribbeans based in the UK; in contrast to this Forouhi and Sattar (2006) revealed that CVD was the leading cause of death among African-Americans. These controversial results support the aforementioned notion that genetic variations might be responsible for

the differences in the incidence of CVD between populations or ethnicities (Forouhi & Sattar, 2006).

The INTERHEART Africa study was conducted in nine sub-Saharan African countries where 578 AMI cases and 785 controls (also in an urban setting) were recruited from 1999 to 2003 (Steyn *et al*., 2005). The aim of the study was to determine the strength of the association between several CVD risk factors, not only individually, but in combination with AMI in the African population in general and in the three subgroups (European Africans (17%), coloured Africans (47%) and black Africans (36%)]. More than 80% of the participants were from South Africa (Steyn *et al*., 2005). The risk factors that presented the strongest association with AMI among the total African group were their previous family history of hypertension and diabetes, again emphasising the imperative role of genetics. A striking observation of this study is the low number of AMI cases among black Africans, mainly outside South Africa, which suggests that AMI in black Africans who live in less developed areas of sub-Saharan Africa may still be relatively rare.

Limited data are available on the current CVD morbidity and mortality rates in South Africa. According to Statistics South Africa (2010), diseases of the circulatory system have been increasing in the proportion of all deaths in the last few years. In 2010, diseases of the circulatory systems were reported to be the second most common cause of death at 15.2% (Statistics South Africa, 2010). Furthermore, diseases of the circulatory system such as other forms of heart disease, cerebrovascular diseases, ischaemic heart diseases and hypertensive diseases contributed a total of 18.4% and 30.0% of the deaths for individuals between the age of 50 and 64 years of age and 65 years and older, respectively and were considered among the ten leading causes of death for those age groups (Statistics South Africa, 2010).

Furthermore, differences in the main causes of mortality among the various ethnic groups in South Africa also exist (Statistics South Africa, 2005; Wyndham, 1979). The CVD mortality rate in South Africa is highest among Indians, followed by coloured people, with the lowest rates reported for Caucasians and black South Africans (Norman *et al*., 2007; Steyn, 2007). Before the 1990s, the CHD prevalence and mortality rates were already reported to be low among black South Africans (Alberts *et al*., 2005; Steyn *et al*., 2005). However, over the last few years, the incidence of cardiovascular cases in black South Africans has increased, emphasising the importance of studying CVD risk factors and

mechanisms contributing to this epidemic (Pieters *et al*., 2011a; Masemola *et al.*, 2007; Alberts *et al.*, 2005).

As early as 1991, researchers already recognised the scarcity of reliable IHD data for black South African populations (Steyn *et al*., 2005; Steyn *et al*., 1991). The BRISK study attempted to assess the overall IHD risk in a black South African population of the Cape Peninsula and therefore determined the prevalence of a combination of the major risk factors (Steyn *et al*., 1991). In general the percentage of males (30.8%) who presented with at least one risk factor was almost threefold that of females (12.5%) and almost half of both the males and females who presented with such a risk factor were of an older age (55-64 years). These results indicated that black South African urban men (from the Cape Peninsula) already had a risk profile for developing IHD.

More recently, Pieters *et al*. (2011a) conducted a case control study, investigating the risk factor profile of CHD in black South Africans and reported that the same risk factors (such as a higher prevalence of tobacco use, hypertension, insulin resistance, increased TC and LDL-C concentrations), which are typically seen in Caucasians, were prevalent at increased levels among the black South African CHD patients. Only modest differences were observed between patients and controls for all the risk factors, except for insulin resistance and small dense LDL-C, which were reported to be the main predictors of CHD development in the black South African population (Pieters *et al*., 2011a). The role of various other environmental risk factors affecting the blood lipid profile and in this way the risk of CVD development will be discussed in Section 2.1.3.

2.1.3 The blood lipid profile as a modifiable risk factor of CVD development

Lipoproteins, which are the major indicators of CVD risk, consist of lipids and proteins that are mainly produced in the liver and the intestine and are involved in the transport and redistribution of lipids in the body (Ordovas, 2009). A clear link has been established between blood lipid determinants and CVD development, with LDL-C being the major lipoprotein involved in plasma cholesterol transport (Lichtenstein *et al.*, 2006; Ordovas, 2006). This led to the National Cholesterol Education Program (NCEP) identifying LDL-C and high-density lipoproteins (HDL-C) as the primary preventative targets to avert CVD development (Lichtenstein *et al.*, 2006; Ordovas, 2006), since LDL-C concentrations are reported to be positively correlated with CVD risk development, whereas HDL-C

concentrations are reported to be negatively correlated with the risk of developing CVD (Cleeman *et al*., 2001).

The risk assessment of CHD requires obtaining an extensive lipoprotein profile (Table 2.2). These profile measurements include TC, LDL-C, HDL-C and blood TG such as very low-density lipoproteins (VLDL) and should be obtained every five years after the age of 20 years (Wilson *et al.*, 1998). The clinical standards for healthy cholesterol concentrations, as defined by the NCEP-Adult Treatment Panel, are depicted in Table 2.2.

Table 2.2 Biochemical values of the lipid profile for CHD risk assessment in Caucasian individuals

Designation	TC (mmol/L)	LDL-C (mmol/L)	HDL-C (mmol/L)	Fasting TG (mmol/L)
Desirable	< 5.17	< 2.59	≥ 1.56	< 1.69
Borderline Risk	5.17-6.18	3.36-4.11	1.03-1.53	1.69-2.25
High Risk	≥ 6.21	4.14-4.89	< 1.03	2.26-5.63

< less than; ≤ less or equal to; > more than; ≥ more than or equal to; HDL-C = high density lipoprotein cholesterol, LDL-C = low density lipoprotein cholesterol; mmol.L^{-1} = millimole per litre; TG = triglycerides; adapted from Wilson *et al.* (1998)

Achieving and maintaining lipid homeostasis are vital, since an increase in LDL-C concentration increases the risk of developing CVD (Cleeman *et al*., 2001). The coordinated action of a large number of factors, such as binding proteins, apolipoproteins, enzymes and receptors (involving hundreds of different genes), together assures the maintenance of lipid homeostasis (Ordovas, 2009). In the event that lipid homeostasis is not maintained, the chronic overload of plasma LDL-C is one of the mechanisms that might initiate the continuous set of pathophysiological events involved in CVD development (Section 2.1, as indicated in Figure 2.1).

It is important to be aware of the consequences of elevated LDL-C concentrations, since individuals are placed into risk categories based on LDL-C measurements and identification of accompanying risk determinants, including all the aforementioned factors (Campbell *et al*., 2008). Even though LDL-C has been the primary target for lipid-lowering strategies, elevated TG concentrations have also been suggested as an independent CVD risk factor (Onat *et al*., 2006; Austin *et al*., 1998). TG does not have atherogenic properties directly, but many TG-rich lipoproteins, such as chylomicron (CM) remnants, VLDL remnants and VLDL particles, independently promote atherogenesis (Talayero & Sacks, 2011). Therefore, in addition to the role of modified LDL-C particles in atherosclerosis development, vascular inflammation can be initiated by TG-rich lipoproteins *via* toll-like receptor 2 signalling (Libby *et al*., 2011; Talayero & Sacks, 2011). In order to prevent atherogenic inflammation from the accumulation of cytokines,

endothelial accumulation should be prevented by clearing the TG-rich lipoproteins from the plasma and increasing liver metabolism of these particles (Talayero & Sacks, 2011). Delayed clearance of these particles has been attributed to the inhibition of the binding of the ApoE and ApoB-100 proteins to the hepatic receptors, resulting in reduced ability of lipoproteins to be taken up by the liver (Talayero & Sacks, 2011).

Furthermore, HDL-C concentrations have been reported to correlate inversely with CVD risk. The mechanisms involved include complicated pathways of reverse cholesterol transport mediated by HDL-C particles, which transfer cholesterol from peripheral tissue to the liver for excretion *via* membrane-bound adenosine triphosphate (ATP) binding cassette transporters (Libby *et al*., 2011; Tall *et al*., 2008). Another mechanism employs cholesterol ester transfer protein (CETP) inhibitors, which increase HDL-C concentrations (Tall *et al*., 2008).

The TC:HDL-C ratio is another means of measurement to predict CHD risk, where a high TC value suggests elevated LDL-C, whereas a low HDL-C value often suggests abnormalities such as elevated VLDL concentrations (Wilson *et al.*, 1998). A TC:HDL-C ratio of 1:5 has been reported to be protective against CVD and since high percentages of African populations have been reported with ratios higher than this cut-off value, this might possibly explain the lower CHD prevalence in South Africa (Mollentze *et al*., 1995; Steyn *et al*., 1991). Thorogood and co-workers (2007) reported that the majority (95%) of the inhabitants of a rural community had this protective ratio, which supported similar reports by other researchers (Steyn *et al*., 1991). In contrast to this, a few studies investigating the lipid profile of rural black communities have reported a higher than expected incidence of hypercholesterolaemia among these individuals (Masemola *et al.,* 2007; Thorogood *et al*., 2007; Alberts *et al.,* 2005; Oosthuizen *et al*., 2002) when using cut-off values developed for Caucasian populations.

It is important to emphasise the role and necessity of integrating biological risk factors and mechanisms with social and behavioural risk factors in order to gain a better understanding of the determinants of disease development, which cause differences in health rates among populations (Stephens *et al.*, 2008; Hernandez & Blazer, 2006). The results reported by Thorogood *et al*. (2007) are an example of these aspects in a South African context. Thorogood *et al*. (2007) reported that the TC concentrations in a rural black South African community at the Agincourt study site (Bushbuckridge district in the Mpumalanga Province) were higher than those reported for a rural community in another

geographical region in South Africa (Kahn *et al*., 2007; Oosthuizen *et al*., 2002). This supports the notion that the implementation of the necessary lifestyle modifications can also help to maintain lipid homeostasis (Ordovas, 2009). The influential role of some of these non-genetic factors on the blood lipid profile will be discussed in the following section.

2.1.3.1 Non-genetic factors that might influence the blood lipid profile

The NCEP lays strong emphasis on the primary role that lifestyle treatment should play in lowering LDL-C (Lichtenstein *et al*., 2006). There are a few distinctive (lifestyle modifying) factors that can lower LDL-C and raise HDL-C concentrations, such as giving up smoking, weight control, physical activity, soluble dietary fibre intake, moderate alcohol consumption and the replacement of SFAs in the diet with monounsaturated fatty acids (MUFAs) or polyunsaturated fatty acids (PUFAs) (Campbell *et al*., 2008; de Backer, 2008).

The risk or benefit determination of a high fat diet, specifically a diet that is high in SFA and/or TFA, is mainly based on the lipid and lipoprotein responses, specifically on LDL-C concentrations, as they respond to dietary fatty acid changes (Kris-Etherton *et al*., 2007). According to the American Dietetic Association and the Dieticians of Canada report, for each percentage increase in energy from SFA in the diet, LDL-C increases by 1% and CHD risk is doubled. A dose-dependent effect has been reported for TFA, with a TFA intake range of 0.5% to 10% of total energy (TE) intake, resulting in elevated LDL-C concentrations (Lichtenstein *et al*., 2006). High TFA intake has been associated with an increase in the relative risk of CHD development, CHD death rate, risk of fatal and nonfatal MI and sudden death (Lichtenstein *et al*., 2006; Hunt *et al*., 2005).

Epidemiological studies have also reported a positive association between SFA intake and CHD incidence, while clinical trials have reported elevated LDL-C concentrations in response to SFA intake (Brown & Goldenstein, 2006; Lichtenstein *et al*., 2006). Dietary recommendations encourage the limitation of SFAs to less than 7% of TE intake, to lower dietary cholesterol intake to < 200 mg/day and to increase (viscous) fibre intake to 10-24 g/day and plant sterols to 2 g/day, as these adjustments enhance the lowering of LDL-C concentrations. The strong, continuous, graded, and independent positive association between the entire range of TC and LDL-C and CVD is exponential, indicating that a given absolute difference in TC or LDL-C from any point in the distribution is associated with a constant percentage difference in CHD risk (de Backer, 2008).

It has also been speculated that the epidemiological transition is one of the most recent contributing factors to CVD development (Pieters *et al*., 2011; Sliwa *et al*., 2008; Steyn *et al*., 2005; Oosthuizen *et al*., 2002; Vorster, 2002) and, therefore, a discussion of the epidemiological transition and urbanisation as a CVD risk factor globally and among black South Africans follows. The epidemiological transition has led to a pattern of risk factors that increase CVD risk, including indicators of socio-economic class (poverty), obesity, hypertension, elevated cholesterol concentrations, tobacco use and a sedentary lifestyle (Mayosi *et al*., 2009; Tibazarwa *et al*., 2009; de Backer, 2008; Schutte *et al*., 2006; Yusuf *et al*., 2001a).

Rural sub-Saharan Africa is at an early stage of the economic and health transition, consequently limited published data are available on the prevalence of CVD risk factors in these populations (Leeder, 2004). However, various risk factors for atherosclerotic vascular disease in both urban and rural communities of South Africa have emerged over the last several decades (Opie & Mayosi, 2005). Mortality rate projections made by Leeder, (2004) indicate that by 2040 the demographic changes in this population alone will result in a doubling of the number of cardiovascular deaths in South Africa, even if the risk factor profile or the mortality rates from CVD stay constant. It is important, however, to determine the manner in which the epidemiological transition and urbanisation of (South) Africans leads to a higher prevalence of certain types of CVD, such as stroke or CBVD, than that of other types, *i.e*. CHD (Reddy & Yusuf, 1998).

Data pertaining to the urbanisation of (black) South Africans over the years are scarce, inconsistent and/or difficult to obtain. However, the available reports indicate that urbanisation is well advanced in South Africa, with more than half of the population residing in urban areas (Armstrong *et al*., 2008; http://www.doh.gov.za/docs/reports/2003/sadhs2003/part1.pdf). The WHO has reported that urbanisation in South Africa has increased with 9% from 1990 to 2009 (WHO, 2010). The latest census conducted in South Africa in 2011 revealed that the population is estimated to have increased by more than 20% from 1996 to 2011, with almost 40% of the population residing in the largest metropolitan areas (Johannesburg and Cape Town), indicating that urbanisation is continuing at a rapid pace in South Africa (http://www.ihsglobalinsight.co.za/News/news.asp?id=1026: 7 January 2013).

Numerous studies have investigated the effect that urbanisation and other lifestyle behaviours have had on CVD development in South African populations (Tibazarwa *et al*., 2009; Sliwa *et al*., 2008; Masemola *et al*., 2007; Thorogood *et al*., 2007; Alberts *et al*., 2005; Steyn *et al*., 2005; Vorster, 2002; Oosthuizen *et al*., 2002; Yusuf *et al*., 2001a; Mollentze *et al*., 1995). The differences in CVD prevalence according to social class observed and documented in various populations can only to some extent be explained by the conventional CVD risk factors (de Backer, 2008). Studies of data from South Africa have similarly reported an increase in the prevalence of hypertension and obesity with increasing wealth. Furthermore, the exposure to some risk factors, such as moderate tobacco use, was inversely associated with wealth, restating concerns that risk factors for NCD might become highest in poor populations (Mayosi *et al*., 2009).

A study conducted by Alberts and co-workers (2005) determined the prevalence of CVD and associated risk factors in a rural black community in the Limpopo province, using the Framingham risk equation (Anderson *et al.*, 1991). It is important to note that the Framingham risk score has not been validated in the (black) South African population; this should be kept in mind when interpreting the results obtained. Alberts *et al.* (2005) reported that 20% of the women and approximately 33% of the men had a 20% or higher chance of suffering a CVD event in the next 10 years, despite living in poor socioeconomic conditions. Socio-economic status can be determined by income, occupation, education *etc.* and these factors have been associated with CVD risk factors *i.e.* blood pressure, BMI, tobacco use and cholesterol concentrations (Winkleby *et al.,* 1999). A study conducted on a black population in Cape Town in the Cape Peninsula, reported that even though urbanisation resulted in higher educational status and diets richer in micronutrients, the atherogenicity within these populations is also significantly higher (Steyn *et al*. 1991). This suggests that an improved socio-economic status does not necessarily lead to an improved nutritional status, but leads to a nutritional pattern that predisposes individuals to the development of atherosclerosis (Steyn *et al*., 1991).

It has been hypothesised that with urbanisation, traditional diets are replaced by a Western diet, which is typified by high fat intake and low carbohydrate and fibre intake (Alberts *et al*., 2005; Vorster, 2002; Yusuf *et al*., 2001a; Walker & Sareli, 1997). The nutrition transition is characterised by increased energy consumption, a decrease in the use of coarse and unprocessed cereals and an increase in the consumption of fats, especially from animal sources, which are high in TFA and SFA (Teo *et al*., 2009). The dietary distribution of urbanised black South Africans has changed drastically compared to their

rural counterparts, with an approximate 10% increase of the TE intake being allocated to fat intake, whereas carbohydrate intake has decreased by more than 5% (Bourne *et al*., 2002). Furthermore, the majority of the studies suggest that rural populations have already been developing significant concentrations of chronic disease risk factors and have a CVD risk profile similar to those present in urban areas of South Africa (Alberts *et al*., 2005). However, a study conducted by Steyn *et al*. (2001) on the adult population of Dikgale (situated in Limpopo, South Africa), reported that adults in rural settings still consumed a moderately traditional diet, which is low in fat and high in carbohydrates, yet low in dietary fibre. Steyn *et al.* (2001) concluded that the type of food items eaten in a rural context did not always contribute to a healthy meal plan, which emphasises that this shift towards a diet similar to that in developed countries is also noticeable and occurring in people from rural areas (Bourne *et al*., 2002).

The “Heart of Soweto” study is a large-scale survey and is one of the many studies with the intention of investigating the effect of urbanisation on heart disease risk among a black South African population in the urbanised town of Soweto (Tibazarwa *et al*., 2009). This study observed that heart failure was the most common primary diagnosis in this population, with moderate to severe systolic dysfunction evident in around half the cases. Black Africans were more likely to be diagnosed with heart failure than were others, but far less likely to be diagnosed with CHD, which is in agreement with similar reports (Pieters *et al*., 2011a). Contrary to the favourable lipid profile of the black African population, Tibazarwa and co-workers (2009) suggested that CVD risk factors are highly prevalent in this population because of the high-risk Westernised lifestyles that were adopted. Furthermore, this research group (Tibazarwa *et al*., 2009) reported that more than 75% of the “Heart of Soweto” study population had at least one measurable major risk factor for CVD *i.e*. >40% were obese, approximately 30% were hypertensive and >10% had elevated serum cholesterol concentrations (Tibazarwa *et al*., 2009; Sliwa *et al*., 2008). Even though such data allows researchers to establish whether epidemiological transition (in specific populations) has expanded the clinical CVD spectrum beyond the traditional CVD risk factors (Sliwa *et al*., 2008), the results reported by Tibazarwa *et al*. (2009) and Sliwa *et al*. (2008) are somewhat surprising, as the prevalence of the risk factors is not as high as one would expect it to be in a CVD-affected population (Yusuf *et al*., 2001b). These results are in accordance with those observed in other ethnicities from around the world, as discussed in Section 2.1. The frequency of each of these risk factors among those burdened with the disease was somewhat different to that observed in the “Heart of Soweto” study population. Even though the frequency of obesity (~23%) is reported to be

lower than that reported by Tibazarwa *et al*. (2009), the frequencies of high blood pressure (~46%) and elevated cholesterol concentrations (~31%) were higher (Tibazarwa *et al*., 2009; Ezzati *et al*., 2005). It is worthwhile to also draw local comparisons with similar studies as wel. Connor and co-workers (2008) reported on the prevalence of stroke risk factors in a South African general practice population. The aforementioned risk factors were common in all the population groups, although fluctuating in their distribution among the groups. The mean age of the participants was 50.7 years, with 67% of the participants having one risk factor and 40% having two or more risk factors. The most common risk factor in all of the ethnic groups was hypertension, but the highest prevalence of 59% was among the black patients, whereas dyslipidaemia had the lowest prevalence among the black patients (5%), but was the most prevalent factor among the Caucasian individuals (37%). These values are different from those reported by Tibazarwa and co-workers (2009), in that the hypertension frequencies among the "Heart of Soweto" study population were approximately half those of the general practice population, whereas the prevalence of hypercholesterolaemia was double.

Another comparative study was reported by Loock and co-workers (2006) who conducted a case-control study with the aim of determining the relationship between CHD and its reported risk factors, within urban black South African individuals. The clinical incidence of the (major) risk factors that significantly contributed to the CHD development of the (black African) cases compared to the controls, were similar to the commonly observed CHD risk factors in European (Caucasian) populations (Loock *et al*., 2006). The majority of the cases presented with elevated TC concentrations (≥6.5 $mmol.L^{-1}$), HDL-C concentrations ≤1.2 $mmol.L^{-1}$, LDL-C concentrations ≥3.0 $mmol.L^{-1}$, hypertension [140/90 millimeter of mercury (mm Hg)], and history of tobacco use, which all contributed to a significant Framingham total CVD risk score (Loock *et al*., 2006). Furthermore, it has been suggested that genetics may play an important role in the predisposition of this ethnic group to CHD development, in view of the significant difference ($p<0.0001$) between family history of certain CVD-associated diseases (diabetes, hypertension, stroke and MI) among cases and controls, which make these strong predictors of CHD risk (Loock *et al*., 2006).

The results reported by the recently published South African National Health and Nutrition Examination Survey also reported on the rate of family history of NCD. The rate of self-reported family history of NCD was highest among the Indian population, but the lowest among black South Africans for all four NCD. The prevalence of high blood cholesterol (*i.e*. TC, LDL-C and TG) was significantly higher among the older age groups,

as well as among subjects residing in formal urban and formal rural areas (Shisana *et al*., 2013). In terms of ethnicity and gender, high blood cholesterol mostly affected Indian and coloured males and among females all ethnic groups were affected except for black South Africans (Shisana *et al*., 2013). Furthermore, the black South African group presented with the lowest prevalence of abnormal serum TC and LDL-C concentrations. However, approximately 50% of these subjects presented with abnormal HDL-C concentrations and one out of four black South African respondents presented with elevated serum TG concentrations. Therefore, these results and findings provided evidence against previous notions that black Africans are not in any way predisposed to CHD (Shisana *et al*., 2013; Loock *et al*., 2006; Steyn *et al*., 1991).

Although it is important to take note of all the conventional and additional interactive effects between CVD risk factors, it is even more important for the purpose of the present study to emphasise the role and influence of genetics on these reported risk factors. Given that the anticipated leading cause of mortality worldwide is CVD, it is a global research priority to understand the genetic basis of this disease (Kullo & Ding, 2007). Studies have investigated whether common genetic variants influence common phenotypic traits (which might already be reported CVD risk factors) and this has supported the notion of genetic susceptibility to disease development (Knoblauch *et al.*, 2004; Knoblauch *et al.*, 2002). This stimulated many researchers to begin unravelling the underlying genetic pathogenesis of the disease (Ordovas, 2009; Forouhi & Sattar, 2006), which will be discussed in Chapter three.

CHAPTER THREE

CVD and genetics: the role of the human *Apolipoprotein E* gene

From the previous chapter it is clear that the genetic background of a population also has a pertinent role to play in CVD risk development. Most human genetic material is identical between individuals, with 99.9% agreement. Of the remaining 0.1%, which represents approximately three million genetic variations between individuals, only 7% can possibly be attributed to observable phenotypic differences, *i.e*. in behaviour, health and other phenotypic traits among individuals and groups, since not all genetic factors affect functional abilities (Christensen *et al*., 2000; Cargill *et al*., 1999; Halushka *et al*., 1999; Wang *et al*., 1998; Li & Sadler, 1991). Consequently, a frequently asked question is to what extent genetic variability accounts for the clinically important differences observed in disease outcomes among ethnic groups. An accepted theory of the role of genetics in common disease, is the "common disease/common variant hypothesis", which postulates that most gene variants for complex diseases are common (*i.e.* allele frequency >20%) and, therefore, are present in multiple populations compared to rare, population-specific genetic variants (Bamshad, 2005; Sing *et al*., 2003).

Yet, at the onset of the genetic era, the field of cardiovascular genetics, while promising, was also limited, as most genetic research had no clear application in a clinical setting. This could be ascribed to the fact that all pioneering genetic studies focused on single-gene disorders (Mendelian diseases) and not common, polygenic diseases such as CVD. Prior to the Human Genome Project era, cardiovascular genetics had a rising impact on the definitions, fundamental insights into the disease processes, diagnosis and underlying genetic aetiology for CVD with monogenic origins through the integration of high-technology laboratory investigations and clinical medicine (Milewicz & Seidman, 2000). However, the Human Genome Project brought about the possibility of investigating the genetic origins of common disorders with complex polygenic pathogeneses and ever since, the field of cardiovascular genetics has expanded drastically.

Information on the human genome was published in 2001 along with successive descriptions of gene variation in the human population (Venter *et al*., 2001; Sachidanandam *et al*., 2001). Subsequently, research in this field focused largely on the association between common SNPs in crucial genes and their association with disease, along with other risk factors (Jofre-Monseny *et al*., 2008). These genetically based association studies have resulted in the creation of a better understanding of the aetiology and physiological mechanisms of diseases, as well as the identification of potential genotypic biomarkers, which can be used as predictors of future disease risk (Jofre-Monseny *et al*., 2008).

SNPs are the most prevalent type of genetic variation and can be defined as the single nucleotide variations in genomic DNA that have a minor allele frequency (MAF) of no less than 1% (Cummings, 2011). There are more than 10 million polymorphic sites in the genome most of which have a frequency between 5% and 50% (Mcvean *et al*., 2005). The number of novel SNPs that are being discovered among different populations is continuously increasing (Brookes, 1999). Most SNPs are located outside the protein-coding sequences; however, the SNPs that occur in this region are termed coding SNPs [(c)SNPs] and are more likely to affect the functionality of a gene (Shelling & Ferguson, 2007). Nonetheless, there are numerous other types of genetic variants such as polymorphic inversions, duplications, microsatellites and other forms of heritable variation that can also affect susceptibility to disease (Mcvean *et al*., 2005).

Endemic lifestyle diseases such as CVD are polygenetic diseases and are thus frequently a result of a dysfunctional biological system and not merely a result of one single mutated gene (Mutch *et al*., 2005). Even though this is true, to conduct a study with a systems-biology approach to obtain the necessary information is not always viable. Nonetheless, investigating the SNP variations in a candidate gene is as important in order to be able to understand the bigger picture better and to complete the multifaceted genetic puzzle.

There are a number of candidate genes that play a role in the development of CVD (Figure 3.1). A CVD event usually results from the interplay of multiple pathogenic mechanisms (as discussed in Section 2.1), which each has its own set of determinants, including environmental and underlying genetic determinants (Talmud, 2007). Any genetic variant involved in any of these pathogenic mechanisms and metabolic pathways can possibly result in the over- or under-production of the key proteins, which will disturb the

homeostatic balance (Stephens & Humphries, 2003). Identifying the genes and the genetic variants (within these genes), which are either associated with or causative of the disorder, is important in order to establish disease risk and ultimately to give advice to prevent CVD. Most unfavourable variants of polymorphic genes do not directly cause CHD, but rather affect individual susceptibility (Loktionov *et al.*, 1999). The impact of these genetic variations is influenced by the genetic background of individuals as well as environmental and lifestyle factors to which they are exposed, which are reported to differ among different populations (Loktionov *et al.*, 1999). Other challenges in identifying genetic determinants of CVD include the phenotypic heterogeneity between individuals and populations, gene-gene and gene-environment interactions and the role of single genetic variants and rare genetic variants in CVD development (Kullo & Ding, 2007).

Various candidate genes have been investigated in association studies in order to establish the role these genes play in the pathogenesis of CVD (Hegele, 2009; Jansen *et al*., 2005; Mutch *et al*., 2005; Wang, 2005; Lusis, 2003; Stephens & Humphries, 2003; Knoblauch *et al*., 2002). However, the overall effect a single genetic marker has on disease risk is modest. CVD is more likely to occur owing to a combination of minor genetic alterations in a number of genes that modulate risk than to a single mutation in a single significant gene [as with familial hypercholesterolaemia (FH)], hence CVD is thought of as a “polygenic” disorder (Stephens & Humphries, 2003).

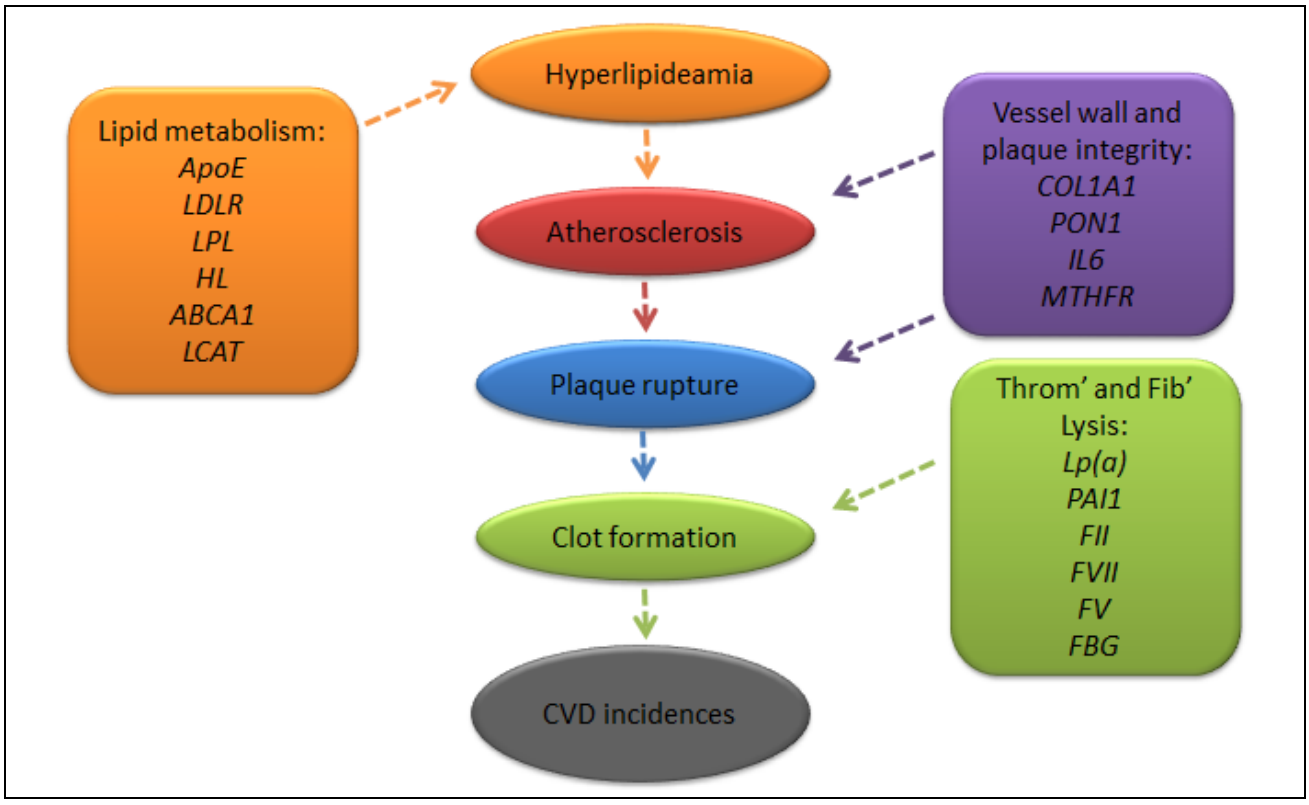
A major breakthrough in this field occurred more recently with the introduction of various genome-wide association studies (GWAS), which have been conducted since 2008. A long with these discoveries, the scientific knowledge regarding important biological implications of the pathways, mechanisms and potential therapeutic strategies involved in lipoprotein metabolism has expanded, even though the clinical application thereof is uncertain (Hegele, 2009). GWAS can screen large numbers of samples from various cohorts and this mechanism has allowed a significant increase in the understanding of complex diseases, *e.g.* CVD (Shelling & Ferguson, 2007).

A number of association studies and genome scans have been conducted to identify genes contributing to CVD development, *i.e*. CVD risk factors (Hegele, 2009; Jansen *et al*., 2005; Wang, 2005; Lusis, 2003; Stephens & Humphries, 2003). Many of these candidate genes contribute to CVD development through pathogenic pathways other than the lipoprotein metabolism, such as inflammation and claudication, but these pathways are not within the scope of this dissertation (Jansen *et al*., 2005). These candidate genes include,

but are not limited to, fibrinogen (*FBG*), factor II Prothrombin (*FII*), *FV* Leiden, *FVII*, *FXIII*, PAI-1, interleukin 6 (*IL6*), methylene-tetrahydrofolate reductase (*MTHFR*), collagen, Type I, Alpha1 (*COL1A1*) and paraoxonase 1 or *PON1* (Lusis *et al*., 2003; Dahlbäck, 2000).

Candidate genes reported in the literature to be involved in the metabolism of lipoproteins, include *CETP*, lipoprotein lipase (*LPL*), hepatic triglyceride lipase (*HL*), LDL-receptor (*LDL-R*), *ApoE*, apolipoprotein A1 (*APOA1*), ATP binding cassette transporter A1 (*ABCA1*) and lecithin-cholesterol acyltransferase or *LCAT* (Jansen *et al*., 2005; Mutch *et al*., 2005; Lusis *et al*., 2003; Stephens & Humphries, 2003; Knoblauch *et al*., 2002). Figure 3.1 recapitulates the role of some of the key candidate genes in the development of CVD, through each of the metabolic pathways mentioned (Stephens & Humphries, 2003).

Figure 3.1 Candidate genes contributing to CVD development



ABCA1 = ATP binding cassette transporter A1; *ApoE* = apolipoprotein E; *FBG* = fibrinogen; *FII* = factor II Prothrombin; *FV* = factor V Leiden; *FVII* = factor VII; *HL* = hepatic triglyceride lipase; *IL6* = interleukin 6; *LCAT* = lecithin-cholesterol acyltransferase; *LDL-R* = low-density lipoprotein receptor; *LPL* = lipoprotein lipase; *Lp(a)* = lipoprotein (a); *MTHFR* = methylene-tetrahydrofolate reductase; *PAI-1* = plasminogen activator inhibitor type 1; *COL1A1* = Collagen, Type I, Alpha1, *PON1* = paraoxonase 1; adapted from Stephens and Humphries (2003); Lusis (2003) and Dahlbäck (2000)

As most plasma lipids and lipoproteins are significant risk factors for CVD development, great interest has arisen in analysing the genetics contributing to the regulation of the lipid profile determinants and these have therefore been extensively investigated over the last three decades (Hegele, 2009). In the mid- 1900s the bulk of the genetic basis of most Mendelian dyslipidaemias had already been discovered and reported. FH is one of these

monogenic diseases that result in atherosclerosis and CHD, with individuals that are heterozygous for *LDL-R* mutations (the molecular root of this disorder) being common in the general population (Caucasians), with an estimated prevalence of 1 in 500 (Jansen *et al*., 2005). A family history of a chronic disease (such as FH) that was the cause of death of a close relative is also regarded as a major risk factor for disease development (Mackay et al., 2004). The genetic causes of less severe, yet more common variations in lipoprotein concentrations were elusive, but have since been identified.

The overall prevalence of FH, which is the most severe lipid disorder, has been estimated to be 1:500 worldwide (Steyn *et al*., 1996). Yet, a much higher prevalence of FH among the Afrikaner population has been estimated at 1:72 (Steyn *et al*., 1996). FH has also been reported to be present among other ethnic groups within South Africa such as Indian and coloured individuals (Kotze *et al*., 1997; Kotze *et al*., 1995). This disease has also been reported to be evident among black South Africans, even though to a lesser extent (Thiart *et al*., 2000). The identification of founder mutations explained the 5 – 10 times higher prevalence of FH in the Caucasian, coloured and Indian populations of South Africa, with a polymorphism that has been noted in the promoter region of the *LDL-R* gene being relatively more frequent in these populations (Hoogendijk et al., 2003). Kotze and co-workers (1993) conducted a study among 276 FH patients and 203 controls and reported that the deleterious effect of the *ApoE* polymorphism was evident in the general population, but the added effect on the lipid profile could not be demonstrated in affected individuals with already very high TC and LDL-C concentrations. This indicates the large effect of high-penetrance mutations underlying this condition among the Afrikaner population of European descent.

Heritability can be defined as the expression of the extent to which the genetic variation among individuals, results in a proportion of phenotypic variation in a population (Cummings, 2011; Elosua et al., 2009; Lusis, 2003). Since the 1980s, studies have been conducted to determine the heritability of certain phenotypes (usually reported CVD risk factors) associated with atherosclerosis (Elosua et al., 2009). Theoretically, the easiest way to examine the heritability rates of traits is to conduct monozygotic (MZ) and dizygotic (DZ) twin studies. MZ twins share 100% of their genes, whereas DZ twins only share half (Lusis, 2003). If one wants to determine whether a genetic component plays a role in the possible development of a specific disease, it is valuable to conduct studies in MZ twins, as they are genotypically alike and must therefore share the same phenotypic components (Lusis, 2003). If these twins do not have identical phenotypes, it can be concluded that

their genetic material is not the only cause of the phenotype of interest and that environmental factors have also contributed to the phenotypic outcome (Lusis, 2003). With that said, various MZ and DZ twin studies have been conducted investigating the role of the environment in CVD development (Lusis, 2003). Such studies merely suggest that in most cases a good environment can overcome the influence of disease risk genes (Lusis, 2003). Draisma *et al*. (2013) reported that MZ twins shared a higher proportion of genetic variation compared to DZ twins, which contributed to the higher similarities in the lipid profile of these MZ twins. These results support the notion that a shared genetic background, along with similarities in age and/or a shared environment, contributes to the similarities in the blood lipid profile of individuals (Draisma et al., 2013). The conclusion from this recent report is in accordance with previous results reported by other twin studies (Iliadou et al., 2001; Christensen et al, 2000; Heller et al, 1994). Even though substantial genetic correlations were reported, no single genetic factor was common to all five lipid profile determinants (Iliadou *et al*., 2001; Christensen *et al.*, 2000; Heller *et al*., 1994).

With the focus of this dissertation being on the lipid profile, heritability estimates for the various determinants will be discussed. Basic estimates have indicated that the conventional risk factors for CVD, *i.e*. the lipid profile, explain approximately 50% of the genetic component that causes CVD (Lusis *et al*., 2003). The variation of TC (as a CVD risk factor) that could be explained by genetic determinants ranged from as low as 28.6% (de Oliveira *et al*., 2008) to as high as 66% (Friedlander *et al*., 1997), with other researchers reporting values within this range (Hegele, 2009; Pilia *et al*., 2006; Lusis, 2003). The heritability estimates for LDL-C ranged between 26.3% (de Oliveira *et al*., 2008) and 65% (Friedlander *et al*., 1997), for HDL-C between 31.2% (de Oliveira *et al*., 2008) and 73% (Friedlander *et al*., 1997) and for TG from 50 to 65% (Friedlander *et al*., 1997), also with other researchers reporting values within these ranges (Hegele, 2009; Pilia *et al*., 2006; Lusis, 2003).

The majority of these studies have reported 40-55% heritability estimates for the lipid profile, supporting the argument that these phenotypic traits have strong genetic underpinnings. Regardless of the validity of this evidence, such information should be interpreted with care, as possible gene-environment interaction(s) could distort the heritability estimates, which is not always taken into account within such studies (Elosua *et al*., 2009).

3.1 JUSTIFICATION FOR INVESTIGATING THE ROLE OF THE *APOE* GENE IN CVD RISK DEVELOPMENT

The large effect of the *ApoE* gene on the lipid profile of Caucasians has been extensively reported (Kolovou & Anagnostopoulou, 2007; Song *et al*., 2004; Lahoz *et al*., 2001; Mahley & Rall, 2000; Miettinen *et al*., 1992; Davignon *et al*., 1988; Mahley, 1988). However, its effect among black South Africans is not well reported and needs further investigation.

The *ApoE* gene is located on human chromosome 19q13.2, is 3.7 kilo-base pairs (Kb) in length and contains three introns and four exons (Lusis *et al.,* 1986; Hagberg *et al.*, 2000). It produces a 299 amino acid polypeptide (Scott *et al.,* 1985; Rall *et al.,* 1982), which plays a fundamental role in the transport and metabolism of plasma cholesterol and TG (Mahley & Huang, 1999), giving the ApoE protein its anti-atherogenic properties (Greenow *et al*., 2005). ApoE also has several other anti-atherogenic roles, which will be discussed and investigated in the sections to follow.

Since the identification of *ApoE* (Siest *et al*., 1995), its association with the risk of developing CVD has been recognised (Kolovou & Anagnostopoulou, 2007; Song *et al*., 2004; Lahoz *et al*., 2001; Mahley & Rall, 2000; Miettinen *et al*., 1992; Davignon *et al*., 1988; Mahley, 1988). The anti-atherogenic roles of *ApoE* seem to be centrally involved in most of the metabolic pathways and/or phenotypic products of the abovementioned candidate genes. *ApoE* has been suggested to play additional roles in the development of CVD related to the reverse cholesterol transport pathway; the inhibition of platelet aggregation (*e.g*. HL, CETP and LCAT); allele-specific antioxidant and immune activities (inhibiting smooth muscle cell and T-lymphocyte activation proliferation and activation, respectively); the hepatic uptake of lipoprotein particles; influencing the hepatic VLDL secretion and processing; activation of enzymes involved in lipoprotein metabolism, inhibition of endothelial cell proliferation and anti-inflammatory activity (Greenow *et al*., 2005; Stannard *et al*., 2001; Riddell *et al*., 1997; Bellosta *et al*., 1995; Shimano *et al*., 1995; Davignon *et al*., 1988).

From the large number of candidate gene association studies that have been conducted, the most consistent evidence has been reported for the *ApoE* gene (Arnett *et al.*, 2007). These studies could describe a large part of the genetic variance of the concentrations of LDL-C and HDL-C on the basis of genetic alterations in a small set of genes, introducing the possibility to use such genetic information to predict CVD risk (Arnett *et al.*, 2007). The

variation in disease risk of certain lipoprotein classes, as defined by specific variations in *ApoE*, resulted in it being identified as one of the many genes contributing to the risk factors of atherosclerosis and to CVD itself (Lusis *et al*., 2003). The genetic impact of the most common and frequently reported *ApoE* genetic alterations on total plasma cholesterol variance is hypothesised to be one of the most powerful (>8%) genetic components in the regulation of cholesterol concentrations at population level (Davignon *et al*., 1988).

The gene locus that had the greatest effect on LDL-C and ApoB variance, in a study conducted by Talmud *et al.* (2009), was the *ApoE* cluster. It has been reported that common polymorphisms within the *ApoE* gene result in modest, but significant contributions to CHD with a greater effect on LDL-C concentrations than any other lipid class (Hegele, 2009). The association between *ApoE* genotypes and LDL-C has also been reported by GWAS (Kathiresan *et al.,* 2008; Willer *et al.*, 2008). However, the association between the *ApoE* genotype and plasma lipid concentrations is complex, as it is also influenced by age, gender, body weight, diet and physical activity (Hagberg *et al.*, 2000).

A meta-analysis conducted by Bennet *et al*. (2007) reassessed the reported association between the *ApoE* genotypes and CHD risk of 203 published studies, including more than 120 000 participants. They concluded that an approximate linear and strongly positive relationship exists between the *ApoE* genotypes and LDL-C concentrations and CVD risk development. Bennet *et al*. (2007) reported a 20% difference in LDL-C concentrations and a 40% variation in CHD risk across the different *ApoE* genotypes. The latter might imply that CHD risk is more strongly associated with genetically determined LDL-C concentrations and years of exposure than to a measurement of LDL-C at a single point in time (Brown & Goldstein, 2006). A meta-analysis by Anand and co-workers (2009) compared the effect of the *ApoE* variants on CHD risk and confirmed that the totality of data presented by the former meta-analysis (Bennet *et al*., 2007) supports their findings that certain genetic variations or *ApoE* haplotypes (ε2 haplotype) lower CHD risk compared to others (*e.g*. ε4 haplotype) that increase risk. However, it is important to consider the possibility that the true genetic effect may be overestimated in the meta-analysis owing to possible publication bias (Song *et al*., 2004).

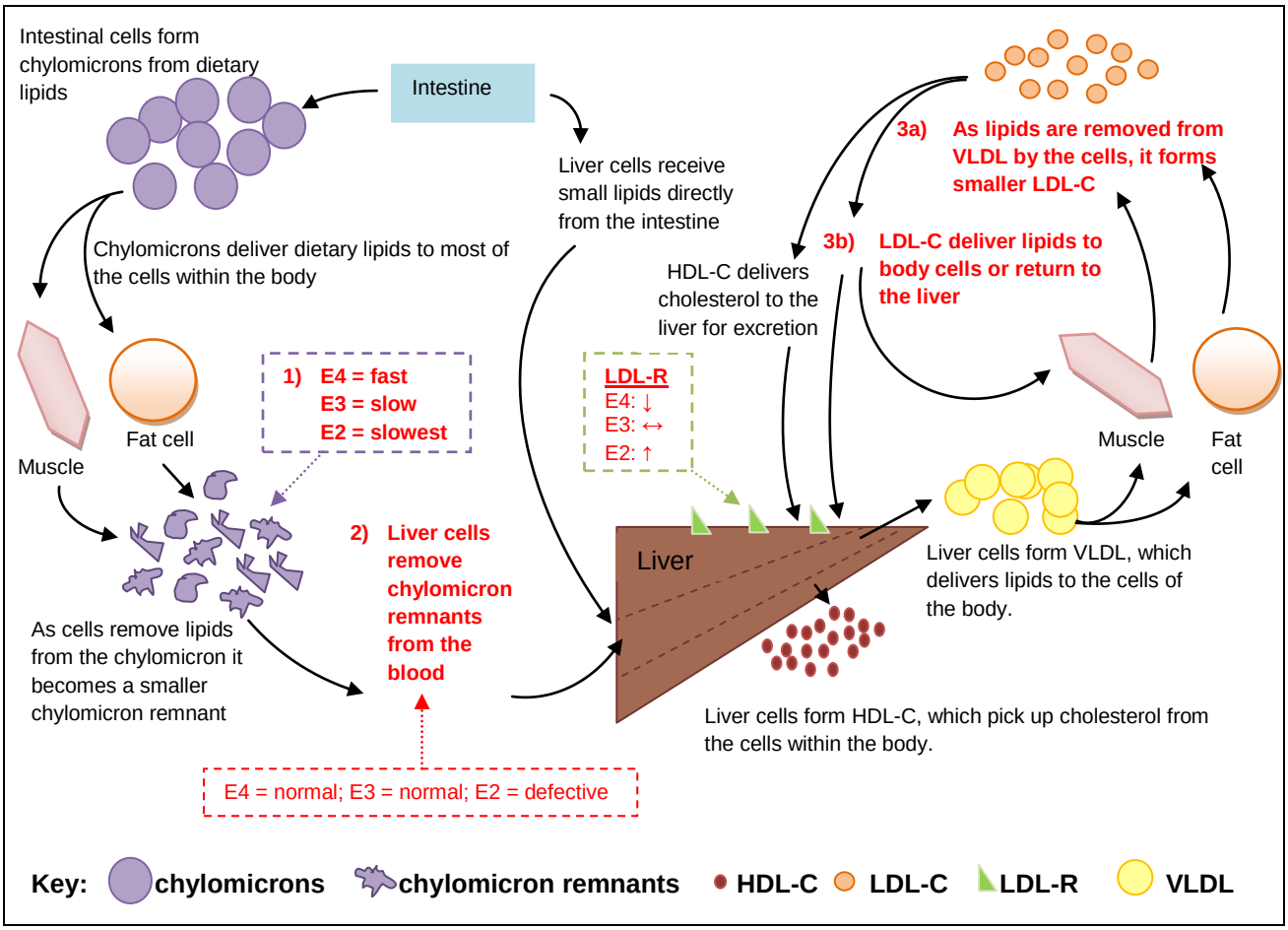
Various mechanistic pathways by which the *ApoE* gene could contribute to CVD development *via* its effects on the blood lipid profile have been identified and these will be

discussed in Section 3.2. In addition, there might also be unidentified mechanisms, which are not yet fully understood, by which *ApoE* gene variations might be linked to CVD risk (Jofre-Monseny *et al*., 2008; Greenow *et al*., 2005).

3.2 THE STRUCTURAL, MECHANISTIC AND FUNCTIONAL ASPECTS OF THE *APOE* GENE

There is a surfeit of evidence that ApoE plays a functional role in the transportation and metabolism of cholesterol (Kolovou & Anagnostopoulou, 2007; Mahley & Rall, 2000; Miettinen *et al*., 1992; Mahley, 1988). Therefore, the vehicles involved in these processes are described in greater detail with reference to Figure 3.2.

Figure 3.2 The lipid transport process via the different lipoproteins as a group



E2 = ApoE 2 isoform; E3 = ApoE 3 isoform; E4 = ApoE 4 isoform; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; LDL-R = low-density lipoprotein receptor; VLDL = very low-density lipoprotein; adapted from Rolfes *et al*. (2012)

Chylomicrons (CM), VLDL, LDL-C, HDL-C and ApoE were initially recognised as the "arginine- (Arg) rich apoproteins" by Shore and Shore in 1973, with ApoE being determined to be a structural component of CM and VLDLs, as well as a subclass of HDL-C (Eichner *et al*., 2002; Curtiss & Boisvert, 2000). The ApoE protein is predominantly produced by the brain and the liver, but is synthesised in a number of other body tissues,

i.e. the kidneys, adrenal glands and monocytes derived from macrophages (Eichner *et al*., 2002). In general, ApoE has a similar role to other apolipoproteins (or apoprotein), which is to maintain the overall structure of lipoproteins. ApoE is involved in the secretion of hepatic lipoproteins and acts as a co-factor in enzymatic reactions in the blood circulation of lipoprotein metabolism and also as a high-affinity ligand for lipoprotein receptors for more efficient cellular lipoprotein uptake (Jofre-Monseny *et al*., 2008; Eichner *et al*., 2002). As the function of ApoE is dictated by its structure, it is important to study the structural changes brought about by the different *ApoE* genetic variants and the functionality of these changes as it pertains to lipid metabolism (Mahley & Rall, 2000) and disease development (Hatters *et al.*, 2006).

As discussed in Section 3.1, the *ApoE* gene is one of many genes encoding for specific proteins that have an influence on blood lipid concentrations (Stengård *et al.*, 2002; Zerba *et al.*, 2000). Several SNPs are reported to exist in the *ApoE* gene and have been reported to change lipid metabolism, thereby altering the serum lipid concentrations, resulting in an altered risk of CVD (Haddy *et al.*, 2002). The focus of the majority of the reports is on the products of the two most common polymorphisms: rs429358 ($T^{3937}\rightarrow C$) and rs7412 ($C^{4075}\rightarrow T$) in exon four of the *ApoE* gene from which the *ApoE* haplotypes (frequently referred to as the *ApoE* alleles) are determined (Nickerson *et al*., 2000). Because of the strong LD between these common non-synonymous SNPs, the three (rather than four) most common and frequently reported protein haplotypes are: ε2, ε3, and ε4 (Campalani *et al.*, 2006; Murrell *et al*., 2006). However, literature generally makes mention of the rs429358 SNP which refers to the ε4 allele of the *ApoE* gene and the rs7412 SNP which refers to the ε2 allele. These polymorphisms have been investigated over the last three decades and have been nominated as “the *ApoE* polymorphisms” (Weisgraber, 1994).

The three common isoforms of the ApoE protein differ in their amino acid sequence at amino acid residues 112 and 158. Arg is encoded at the two CGC codons, which match up with nucleotides 3937 and 4075. When the first base in the CGC codon changes (a C→T transition), it results in a TGC codon, which in turn encodes cysteine (Cys). Therefore, the ε2 haplotype (T^{3937}/T^{4075}) encodes the E2 isoform (Cys^{112}/Cys^{158}), whereas the ε3 haplotype (T^{3937}/C^{4075}) encodes for the E3 isoform (Cys^{112}/Arg^{158}) and the ε4 haplotype (C^{3937}/C^{4075}) encodes for the E4 isoform, referred to as Arg^{112}/Arg^{158} (Seripa *et al*., 2007).

A fourth *ApoE* genetic variation encoded by the C^{3937}/T^{4075} haplotype (GenBank Accession AF261279) also exists, however; this haplotype has seldom been reported in the literature (Seripa *et al*., 2007; Murrell *et al*., 2006; Persico *et al*., 2004). Persico and co-workers (2004) suggested that this allele be referred to as ε3r, while Murrell *et al.* (2006) suggested that this allele be referred to as ε1y, with the addition of the "y" for Yoruba, in order to differentiate between the other E1 proteins previously defined. However, seeing that the ε1 allele encodes for the E1 isoform (Asp^{127}/Cys^{158}) also known as the "Weisgraber allele", the former suggested name, ε3r, is preferred because it encodes for the ApoE3r isoform that is positively charged and has identical physical characteristics to the ApoE3 isoform, only with the Cys and Arg exchanged (ApoE3r = Arg^{112}/Cys^{158}) and this nomenclature will, therefore, be used throughout this dissertation.

The ε3r allele was first reported by Persico and co-workers in 2004, who investigated the enhanced Apoε2 transmission rates in families with autistic probands. An autistic child (male) and his mother presented with unexpected restriction fragment sizes [at base pairs (bp) 33, 48, 72, 81 and 91] and from subsequent DNA sequencing, this (novel) SNP (ε3r) was discovered. The ε3r haplotype has been observed in two ethnic groups: three Caucasian families from Italy (Seripa *et al*., 2007; Persico *et al*., 2004) and one Yoruba family from Nigeria (Murrell *et al*., 2006).

These (common three) haplotypes are co-dominantly inherited and, therefore, six *ApoE* genotypes can possibly arise (ε2/ε2, ε2/ε3, ε3/ε3, ε3/ε4, ε4/ε4, ε2/ε4), which are also the most common and frequently reported variations within the *ApoE* gene (Zannis *et al.*, 1982). In the possible presence of the fourth haplotype (ε3r), additional genotypes are possible *i.e*. ε3r/ε3r, ε2/ε3r, ε3/ε3r, ε4/ε3r.

The association of these haplotypes with disease risk has been extensively investigated (Hegele, 2009; Campalani *et al.*, 2006; Murrell *et al*., 2006; Nickerson *et al*., 2000; Weisgraber, 1994). These analyses revealed that for each haplotype, different functional consequences can arise. The different amino acids lead to structural alterations that have an impact on the functioning of the protein. The substitution of an Arg by a Cys at position 158 in ApoE2 affects a key salt-bridge arrangement in ApoE3 and ApoE2, resulting in a different position of Arg^{150} (Figure 3.3), which alters its ability to interact effectively with LDL-R (Dong *et al.*, 1996). This ultimately results in a 50-100-fold weaker binding affinity

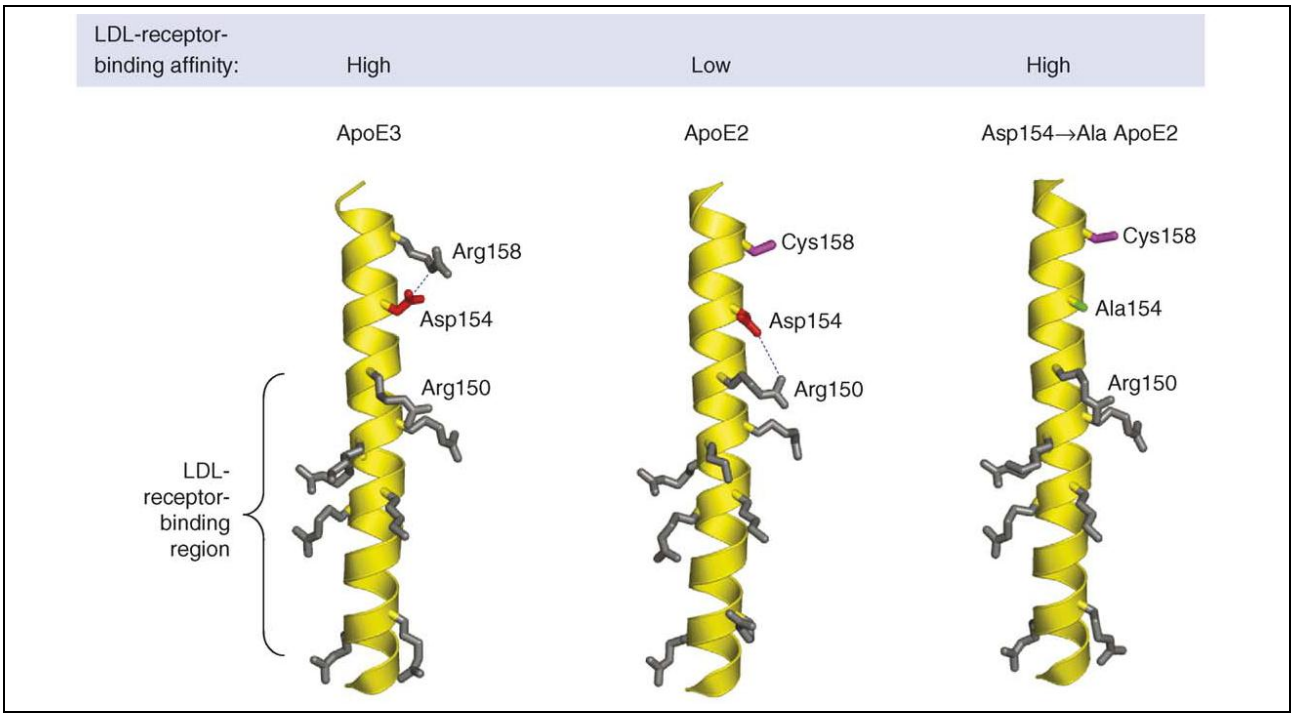
of the protein to LDL-R when it is located on the cell surface (Innerarity *et al.*, 1984; Weisgraber *et al.*, 1982).

As a result of the weak binding affinity, individuals that are homozygous for the Apoε2 haplotype sometimes develop type III hyperlipoproteinaemia (HLP), also reported as dysbetalipoproteinaemia. Since homozygosity for the ε2 haplotype is the primary causal molecular mechanism of developing HLP, HLP can be considered to be a recessive form of FH (Mahley & Rall, 2000). The ε2/ε2 genotype only occurs in 1% of the global population and since other secondary factors are also at play, less than 5% of the individuals with this rare genotype will ultimately develop HLP (Kolovou & Anagnostopoulou, 2007; Mahley & Rall, 2000). For type III HLP to be evident, it requires the homozygosity for the *Apo*ε2 alleles together with secondary factors that cause the sudden development of the disorder (Mahley & Rall, 2000). Such secondary factors include genetic factors *i.e*. monogenic or polygenic gene interactions, hormonal factors (oestrogen being the most significant effect by increasing LDL-R levels and lipolytic activity) and environmental factors such as obesity due to excess dietary intake (Mahley & Rall, 2000). The lipoprotein profile of individuals with HLP is distinctly characterised by the presence of abnormal lipoproteins *i.e*. elevated cholesterol or lipid concentrations attributed to beta very low density lipoprotein (β-VLDL), but slightly lower HDL-C and LDL-C concentrations (Mahley & Rall, 2000). These β-VLDL particles are not only cholesterol-enriched remnants, but are also ApoE enriched (Mahley & Rall, 2000). Individuals carrying the ε2/ε2 haplotype-genotype that do not develop HLP (>90%), are either normolipidaemic or hypolipidaemic owing to the lower HDL-C and LDL-C concentrations.

In the ApoE3 isoform, Arg^{158} forms a salt-bridge with Asp^{154}, which adds little to the structural variation and is therefore reported as the standard isoform. For the ApoE2 isoform, the salt-bridge structure is formed between Asp^{154} and Arg^{150} (as represented by the broken lines illustrated in Figure 3.3). Since the ApoE2 salt-bridge lies within the LDL-R-binding region, this changes its conformation and properties. The ApoE3 and E2 isoforms are inclined to bind to HDL-C, which is a smaller and more phospholipid-enriched lipoprotein particle compared to the E4, which is prone to bind to VLDL, which is a larger and TG-enriched lipoprotein particle (Mahley & Rall, 2000). Therefore, individuals with the ApoE4 isoform tend to have raised TC, LDL-C and apolipoprotein B (ApoB) concentrations and are more prone to develop CVD (Kolovou *et al.*, 2003; Braeckman *et al.*, 1996;

Muros & Rodrćguez-Ferrer, 1996; Luc *et al.*, 1994; Davignon *et al*., 1985). A possible underlying mechanism that can explain the above-mentioned involves the down-regulation of LDL-R, which will be explained in the paragraphs below; however, further research is required to elucidate the full mechanism at play (Mahley & Rall, 2000).

Figure 3.3 Influence of amino acid position 158 on LDL-R-binding activity



Arg = arginine; Asp = aspartic acid; Cys = cysteine; LDL = low-density lipoprotein; reprinted from TRENDS in Biochemical Sciences, Vol 31/no. 8, Hatters *et al*., Apolipoprotein E structure: insights into function, 445-454., Copyright (2006), with permission from Elsevier.

ApoE has been suggested to have additional effects on the development of CVD, other than influencing the lipid metabolism, as discussed in Section 3.1 and 3.2. However, as lipid metabolism forms an integral part of the present study, this will be the primary focus of the discussions ahead.

Four mechanisms are suggested by which the isoforms of ApoE influence lipid metabolism (Kolovou & Anagnostopoulou, 2007). The proposed mechanisms include the differences in the clearance rate of lipoproteins, *i.e.* LDL-C and CM (Gylling *et al.*, 1995), receptor-binding affinities of the different ApoE-containing lipoproteins (Mahley et al., 1999; Mahley, 1988), efficiency of intestinal cholesterol absorption through LDL-C formation and LDL-R regulation (Miettinen *et al.*, 1992) and dietary fat clearance (Weintraub *et al.*, 1987).

One of the mechanisms influencing lipid metabolism is the manner in which the ApoE isoform influences CM clearance (the rate of exogenous fat removal). Weintraub *et al*.

(1987) revealed that ApoE4 carriers clear CM remnants in the liver faster than E3/E3 carriers and twice as fast as E3/E2 carriers, since the E4 isoform does not contain a Cys residue [Figure 3.2: (1)]. This results in the ApoE4 isoform being unable to form disulphide bonds with other proteins, causing the ApoE4 isoform to shift to CMs from HDL-C faster than the other isoforms and thus maintaining low TG concentrations. Moreover, the slow hepatic CM remnant clearance rate by the ApoE2 isoforms can be ascribed to its defective binding ability to the hepatic CM remnant receptors, indicated as the process described at (2) in Figure 3.2, which is, therefore, associated with a delayed catabolism of CM and VLDL remnants leading to elevated TG concentrations (Davignon *et al*., 1988; Weisgraber *et al*., 1982). Ehnholm *et al*. (1984) documented that the presence of the ApoE2 isoform in VLDL particles weakens the conversion between VLDL and intermediate density lipoprotein to LDL-C by a possible indirect influence of the hydrolytic activity of lipoprotein lipase [Figure 3.2 (3a)]. Another hypothesis by Weintraub *et al*. (1987) is based on cholesterol and collective fatty acids in the liver, which is also the foremost site where LDL-C removal takes place. The rate of removal of the endogenous fat-containing particles by the liver can be influenced by the hepatic CM clearance rate. Therefore, a further mechanism lies at the level of LDL-R regulation, which, along with LDL-C conversion, is the third and fourth mechanisms that can be grouped and discussed together as various explanations arise for the (interrelated) pathophysiology of these mechanisms.

The basis for the reduced LDL-C concentrations in the plasma of ApoE2 carriers is less clear. Furthermore, ApoE is not a component of the LDL-C particle, suggesting an indirect influence on LDL-C concentrations. In the fasting state, most plasma ApoE is a component of HDL-C, but it shifts to postprandial particles after fat ingestion (Blum, 1982). ApoE2 isoforms have the greatest stability and the least formation of intermediates, while ApoE4 isoform has an opposite effect, forming a typical melted belt-like configuration, compiled from the helices of ApoE, which wrap around the acyl chains of a bilayer of phospholipids (Hatters *et al.*, 2006; Greenow *et al*., 2005). It is thought that the differences in ApoE protein folding and the different stabilities explain the discrepancy in the lipoprotein binding preferences (Hatters *et al.*, 2006), with ApoE4 binding preferably to larger lipid-rich lipoproteins (VLDL and LDL-C) and ApoE2 and ApoE3 preferring smaller lipoproteins such as HDL-C (Minihane *et al.*, 2007). Therefore, this mechanism has been suggested to be largely responsible for the moderate increases (18%) in LDL-C observed in ApoE4 carriers (Minihane *et al.*, 2007). The slower hepatic clearance of dietary fat in ApoE2/E3 carriers could result in the up-regulation of *LDL-R* and lead to decreased

plasma LDL-C concentrations as an outcome, indicated as (3b) in Figure 3.2 (Weintraub *et al*., 1987). However, ApoE4 carriers are at greater risk of developing CVD, as indicated in previous sections, since these carriers' hepatic clearance of dietary fat is faster, possibly resulting in the down-regulation of *LDL-R*, which in turn results in increased LDL-C concentrations. Another possible explanation for the LDL-C raising effect of the Apoε4 haplotype has been proposed by Bergeron and Havel (1996). This research group hypothesised that the receptor-binding activity of the ApoE4 isoform decreased in response to its over-accumulation on the VLDL and CM remnants, resulting in prolonged dwelling time of these lipoproteins in the circulatory system, which ultimately led to these particles being converted to LDL-C.

In addition, the ApoE protein is one of the many proteins involved in the process of cholesterol absorption. Cholesterol absorption and synthesis are inversely correlated and the cholesterol synthesis therefore increases as the cholesterol absorption efficiency decreases (Moreno *et al.*, 2006; Mattes, 2005). Therefore, ε2 carriers present with higher hepatic cholesterol synthesis and lower intestinal cholesterol absorption than ε4 carriers, with the ε3 carriers reported to be in between those of the ε2 and ε4 carriers (Gylling *et al.*, 1995; Miettinen *et al.*, 1992; Kesäniemi & Miettinen, 1987). According to a study conducted by Miettinen *et al.* (1992), which investigated the effect of different fat-containing diets (37% and 64% reduced fat) on the cholesterol absorption association with the ApoE phenotype, no effect was reported.

The previous section indicated how *ApoE* genetic variations can influence the mechanistic role of the ApoE isoforms in lipid transport and metabolism, resulting in altered circulatory lipoprotein concentrations. These alterations could be beneficial or detrimental in terms of the risk they pose to the development of CVD. The following section will discuss the genetic variants in the *ApoE* gene that have been reported to influence the concentration of *i.e*. plasma ApoE protein, as well as other CVD risk biomarkers such as the lipid profile.

3.3 GENETIC VARIANTS IN THE *APOE* GENE

The *ApoE* gene is polymorphic, as numerous genetic variants have been identified in various ethnic groups (Nickerson *et al*., 2000). However, as mentioned in Section 3.2, most of the genetic variation within a number of populations is attributed to the two common polymorphisms (rs429358 and rs7412) alone resulting in the three well-reported haplotypes (ε2, ε3 and ε4), which are the focus of most studies. Generally, the ε2

haplotype is associated with a lower concentration of LDL-C, but with higher TG and HDL-C concentrations, while the ε4 haplotype is associated with a higher concentration of TC and LDL-C and increased susceptibility to CHD (Ye *et al*., 2003). The following discussion will refer to studies that have investigated the rs429358 and rs7412 loci individually.

Silander *et al*. (2008) reported that the CC genotype group of the rs7412 SNP was associated with increased LDL-C concentrations ($p<0.01$) and the the TT genotype group of the rs429358 variant was associated with elevated CRP concentrations, ($p<0.01$) among a Finnish cohort. Genetic variants previously associated with fasting blood lipids, in an American population (non-Hispanic Caucasian group, non-Hispanic black group and Mexican-American group) have been investigated by Chang *et al.* (2010). For the non-Hispanic black group, rs7412 was strongly associated with both LDL-C and TC concentrations ($p<0.0001$), whereas both *ApoE* rs7412 and rs429358 were significantly associated ($p<0.05$) with LDL-C and TC in the Mexican Americans (Chang *et al*., 2010).

Additional SNPs in the *ApoE* gene have also been reported to influence the concentration of plasma ApoE protein and consequently the lipid profile, albeit to a more limited extent than the two major variants (Geng *et al*., 2011; Abboud *et al*., 2008; Viiri *et al*., 2008; Viiri *et al*., 2005; Moreno *et al*., 2003; Lambert *et al*., 2000; Artiga *et al*., 1998). The following section will discuss the role of these SNPs that have not been investigated as well as the "epsilons" in CVD pathogenesis.

3.4 ADDITIONAL SINGLE NUCLEOTIDE POLYMORPHISMS IN THE *APOE* GENE

Nickerson and co-workers (2000) were one of the first research groups to conduct a systematic investigation of other variations in the *ApoE* gene, as little attention had been paid to characterise polymorphisms in the promoter region, as well as other regions of the gene. To extend the data over the most studied and reported *ApoE* genotypes, the (additional) genetic variants (SNPs), with the following reference sequence (rs) numbers will be investigated in the present study: rs440446, rs405509, rs1081101, rs769449 and rs769452, as summarised in Table 3.1. Therefore, discussions that follow will be focused on these SNPs, specifically. Some of the investigated SNPs, which were determined by sequencing the participants in the current study, were included without any prior clinical hypothesis' and, therefore, with no identified literature (rs1081101) or with scant literature

available (rs769452). The inclusion of these SNPs in the study, regardless of the lack of scientific evidence, is justified in Section 4.3.3, which also addresses one of the objectives of the current study to determine whether an association between these SNPs and the lipid profile exists in a black South African population. The lack of literature for such SNPs might be due to researchers observing no association between the SNP and the lipid profile and, therefore, being unable to publish their results (publication bias), or conducting poorly designed studies with inadequate sample sizes and consequently rejecting the null hypothesis when it is in fact true (type 1 error) or observing real, but weak results (Kathiresan *et al*., 2008; Paré *et al*., 2007).

The -219G/T (rs405509) and +113G/C (rs440446) polymorphisms in the proximal promoter region of the *ApoE* gene were identified for the first time in 1996 by Mui and co-workers and were described and investigated by a number of studies thereafter (Geng *et al*., 2011; Abboud *et al*., 2008; Viiri *et al*., 2008; Viiri *et al*., 2005; Moreno *et al*., 2003; Lambert *et al*., 2000; Artiga *et al*., 1998). The *ApoE* promoter polymorphism -291G/T (rs405509), has been indicated by both *in vitro* and *in vivo* studies to produce variance in the transcriptional activity of the *ApoE* gene (Lambert *et al*., 2000; Artiga *et al*., 1998). More specifically, the G-allele has a higher transcriptional activity than the T-alllele, possibly as a result of the differential binding of nuclear proteins (Artiga *et al*., 1998). This polymorphism has been associated with the incidence of various CVD risk biomarkers and CVD development (Table 3.1).

The study by Viitanen *et al*. (2001) indicated that both the rs405509 promoter SNP and the ε2/ε3/ε4 polymorphisms of the *ApoE* gene are associated with early-onset CHD, which has been reported by other researchers as well (Ye *et al*., 2003; Lambert *et al*., 2000). Since the T-allele of the rs405509 promoter polymorphism and the ε4 haplotype were reported to influence two-hour glucose concentration in an oral glucose tolerance test, this may indicate that insulin resistance is an additional mechanism that is affected, which could explain increased risk of CHD in those harbouring the T-allele of the rs405509 and the ε4 haplotype of the *ApoE* gene polymorphisms. Therefore, the T-allele of the rs405509 has been reported to be associated with MI and premature CHD (Viitanen *et al*., 2001; Lambert *et al*., 2000).

The role of the *ApoE* promoter polymorphisms -219G/T (rs405509) and +113G/C (rs440446) in IS or cerebral atherosclerosis was studied for the first time by Abboud *et al*. (2008) and the results suggest that G-allele carriers at both polymorphic loci are at an

increased risk of IS, most likely independently of the *ApoE* ε2/ε3/ε4 haplotypes. Functional studies on Finns revealed differences in the transcriptional activity of various *ApoE* +113G/C (rs440446) alleles (Viiri *et al*., 2008), demonstrating an association with serum lipid concentrations in men (Viiri *et al*., 2005).

The *ApoE* rs769449 polymorphism, on the other hand, has been associated with fewer CVD risk markers (Table 3.1). Soerensen *et al*. (2012) reported that the frequency of the rs769449 SNP was significantly decreased in the eldest study participants, which indicates a disadvantageous effect on survival. However, rs769449 is in modest LD with rs429358 and, therefore, this association was removed when adjusting for rs429358. This indicates that this association is expected to reflect the well-known effect of rs429358 (Soerensen *et al*., 2012).

The rs769452 variant is referred to as "ε4(-)", because it is slightly more acidic than ε4 (rs429358) and has a substitution at position number 28 in the protein from leucine to proline, CTG-to-CCG (Minihane *et al*., 2007). Limited reports are available on the clinical effect pertaining to rs769452 (Wieland *et al*., 1991). Wieland *et al*. (1991) conducted a study among German-Caucasian subjects and identified a ε4 variant, at rs769452 in the *ApoE* gene, which was not associated with elevated blood lipid concentrations.

Table 3.1 Associations between the (additional) genetic variants and the risk factor and incidence rate of CVD

Study design Study population's health status (#cases;#controls)	Association with CVD	Association with Apo	Other risk biomarkers (p-value)	Citation
rs440446				
Multicentric Belgian case-control study. Middle-aged patients with IS (237; 326)	↑ risk of IS (p=0.01)	---	---	Abboud *et al*. (2008)
Case-cohort study on the two large FINRISK 92 cohort (5,999) and FINRISK 97 cohort (8,141)	---	---	↑ TG (p<0.01) among G-allele carriers	Silander *et al*. (2008)
Randomised controlled trial 799 healthy Chinese subjects	---	---	↓ TC (p=0.05) ↓ LDL-C (p=0.04)	Andreotti *et al*. (2009)
Prospective study 1774 adults in the TARF Study	---	---	↑ Insulin (mIU/L) among CC carriers (p=0.02) ↑ HOMA index among CC carriers (p=0.01)	Komurcu-Bayrak *et al*. (2011)

*ANOVA; $ = direction of difference not reported; Apo = apolipoprotein; AU = arbitrary unit; CHD = coronary heart disease; CHO = carbohydrates, CRP = C-reactive protein; ECTIM = Etude Cas-Témoin de l'Infarctus du Myocarde; FINRISK = surveys carried out every 5 years to assess the prevalence and risk factors of CVD in Finland; GWAS = genome-wide association study; HOMA = homeostatic model assessment; IS = ischaemic stroke; LD = linkage disequilibrium; LDL-C = low-density lipoprotein cholesterol; MI = myocardial Infarction; MUFA= mono unsaturated fatty acids; OGTT = oral glucose tolerance test; SFA = saturated fatty acids; SNPs = TARF = Turkish Adult Risk Factor; TC = total cholesterol; TG = triacylglycerol or triglyceride; TRL= triacylglycerol-rich lipoprotein; WGHS = Women's Genome Health Study; WHR = waist-to-hip-ratio

Table 3.1 (continue) Associations between the (additional) genetic variants and the risk factor and incidence rate of CVD

Study design Study population's health status (#cases;#controls)	Association with CVD	Association with Apo	Other risk biomarkers (p-value)	Citation
rs405509				
Case-control ECTIM Study MI patients (567; 678)	↑ risk of MI (p<0.003)	↓ ApoE (p<0.01)	---	Lambert *et al*. (2000)
Familial case-control Premature CHD within Finish families (118; 110)	---	↓ ApoA1 (p=0.06*)	↑ WHR (p=0.01*) ↑ glucose - 2 hour fasting (p=0.007*) ↑ Insulin 2 hour fasting (p=0.06*)	Viitanen *et al*. (2001)
Postprandial clinical trial (vitamin A fat loading diet) 51 healthy male student volunteers (Spanish)	---	TT genotype group had the lowest postprandial levels of serum ApoE (p=0.05)	TT genotype carriers have higher postprandial response of: large TRL-TG (p=0.03), large TRL-RP (p=0.04) and small TRL-apoB48 (p=0.05) compared to the G-allele carriers	Moreno *et al*. (2003)
Cross-sectional 1170 patients with Angiographically documented CHD	↑ Disease severity (p=0.004)	---	LD with ε4 (p<0.001)	Ye *et al*. (2003)
Clinical trial (3 dietary periods: SFA-rich diet, MUFA-rich diet and a CHO-rich diet) 55 healthy male volunteers (Spanish)	---	Greater ↑ in ApoB among carriers of the *T* allele after the SFA diet compared to the *GG carriers* (p<0.05). Greater ↓ in ApoB among carriers of the *T* allele (compared with *GG* carriers) after changing from the SFA to the CHO diet (p<0.05).	*TT* subjects had a shorter lag time (before propagation of the LDL-C oxidation reaction) after the SFA diet (*p*<0.05) compared with carriers of the *G* allele, A greater ↑ in lag time in the *TT* subjects occurred (compared with G-allelel carriers) after the replacement of the SFA diet by the CHO or MUFA (p<0.05). Greater ↑ in LDL-C among carriers of the *T* allele after the SFA diet compared to the *GG carriers* (p<0.05). Greater ↓ in LDL-C among carriers of the *T* allele (compared with *GG* carriers) after changing from the SFA to the CHO diet (p<0.05).	Moreno *et al*. (2004)
Case control study involving 884 individuals from 142 families as well as 558 case and control subjects from north-eastern Quebec, investigated 1,536 SNPs in 103 candidate genes for CHD	---	ApoB$ concentrations (p=0.02)	---	Paré *et al*. (2007)
Multicentric Belgian case-control study. Middle-aged patients with IS (237; 326)	↑ risk of IS (p=0.01)	---	---	Abboud *et al*. (2008)
Prospective study 1774 adults in the TARF Study	---	↑ ApoE among GG carriers (p=0.07)	↑ Isulin (mIU/L) among TT carriers (p=0.03) ↑ HOMA index among TT carriers (p=0.04)	Komurcu-Bayrak *et al*. (2011)

*ANOVA; $ = direction of difference not reported; Apo = apolipoprotein; AU = arbitrary unit; CHD = coronary heart disease; CHO = carbohydrates, CRP = C-reactive protein; ECTIM = Etude Cas-Témoin de l'Infarctus du Myocarde; FINRISK = surveys carried out every 5 years to assess the prevalence and risk factors of CVD in Finland; GWAS = genome-wide association study; HOMA = homeostatic model assessment; IS = ischaemic stroke; LD = linkage disequilibrium; LDL-C = low-density lipoprotein cholesterol; MI = myocardial Infarction; MUFA= mono unsaturated fatty acids; OGTT = oral glucose tolerance test; SFA = saturated fatty acids; SNPs = TARF = Turkish Adult Risk Factor; TC = total cholesterol; TG = triacylglycerol or triglyceride; TRL= triacylglycerol-rich lipoprotein; WGHS = Women's Genome Health Study; WHR = waist-to-hip-ratio

Table 3.1 (continue) Associations between the (additional) genetic variants and the risk factor and incidence rate of CVD

Study design	Association with CVD	Association with Apo	Other risk biomarkers (p-value)	Citation
rs769449				
GWAS: 6,382 Caucasian women	---	---	LDL-C$^{\$}$ ($p=4.7 \times 10^{-7}$)	Chasman *et al*. (2008)
GWAS: 6,345 apparently healthy women in the WGHS	---	---	↓ CRP$^{\$}$ ($p=8.90 \times 10^{-21}$)	Ridker *et al*. (2008)
GWAS: 17,686 apparently healthy Caucasian women in the WGHS	---	---	↓ CRP ($p=5.98 \times 10^{-57}$) ↓ Fibrinogen ($p=0.25$)	Danik *et al*. (2009)

*ANOVA; $ = direction of difference not reported; Apo = apolipoprotein; AU = arbitrary unit; CHD = coronary heart disease; CHO = carbohydrates, CRP = C-reactive protein; ECTIM = Etude Cas-Témoin de l'Infarctus du Myocarde; FINRISK = surveys carried out every 5 years to assess the prevalence and risk factors of CVD in Finland; GWAS = genome-wide association study; HOMA = homeostatic model assessment; IS = ischaemic stroke; LD = linkage disequilibrium; LDL-C = low-density lipoprotein cholesterol; MI = myocardial Infarction; MUFA= mono unsaturated fatty acids; OGTT = oral glucose tolerance test; SFA = saturated fatty acids; SNPs = TARF = Turkish Adult Risk Factor; TC = total cholesterol; TG = triacylglycerol or triglyceride; TRL= triacylglycerol-rich lipoprotein; WGHS = Women's Genome Health Study; WHR = waist-to-hip-ratio

Being knowledgeable about the reported associations between various genetic variations in the *ApoE* gene with risk biomarkers is fundamental to the present study. However, it is also important to be aware of the evolutionary path that has been hypothesised for the *ApoE* gene, as well as its possible associations or interactions with environmental factors. These aspects will be addressed in the following sections.

3.5 THE ANTHROPOLOGY OF THE *APOE* GENE

Evolutionists have suggested the theory entitled the "Apoε4 ancestral allele theory" because non-human primates harbour the Apoε4 haplotype (Hanlon & Rubinsztein, 1995). According to this theory the Apoε3 and Apoε2 haplotypes were produced by mutation events that later spread among humans after the human and chimpanzee ancestry split. Consequently, the present allele distribution may also have taken place as a result of natural selection, since many functional property differences occur among these three haplotypes. Based on the analysis conducted by Corbo and Scacchi (1999), the Apoε3 haplotype is probably the most prevalent haplotype in all human populations.

According to Gerdes *et al*. (1996), the Apoε3 frequencies are the highest in populations that have historically and habitually performed agricultural activities. It has been hypothesised that the metabolic properties of the E3 isoform were particularly advantageous during the food collection to food production transition (Corbo & Scacchi, 1999). Some populations are characterised by an unusually high frequency of the Apoε4 haplotype, which has been cautiously recognised to be a possible effect of positive selection and its increased prevalence giving an advantage in the

non-westernised environment. In addition, the ApoE gene has been characterised as a ‘‘thrifty gene’’ because of its role in lipid metabolism and transport and, because the Apoε4 haplotype can be considered as a “thrifty allele” (Corbo & Scacchi, 1999).

The “thrifty genotype” hypothesis is defined as an original genetic trait that was advantageous in the past, but became maladaptive (disadvantageous) with the onset of the Westernised abundant diet (Gluckman & Hanson, 2004; Neel, 1962), thus creating a mismatch. This hypothesis could explain why various chronic diseases are becoming increasingly common in developed countries (Williams & Nesse, 1991).

One of the reasons why the ε4 haplotype is considered a “thrifty allele” is that it provides advantages in harsh environments (when diets are low in cholesterol as experienced in some pre-urbanised populations) in that Apoε4 could help in rebalancing cholesterol concentrations by enhancing intestinal cholesterol absorption in situations where it would otherwise have been too low (Scacchi *et al*., 1997). However, when carriers of the Apoε4 haplotype are exposed to an overabundance of food, such as experienced currently in Westernised populations, it increases their susceptibility to developing CHD and thus these “thrifty” haplotypes become disadvantageous (Corbo & Scacchi, 1999). CVD is rare in some developing countries even though the prevalence of the Apoε4 is most frequent (Corbo & Scacchi, 1999), but because of urbanisation one should expect an imminent CVD epidemic in these countries as the lifestyle habits of the indigenous population turn from their traditional to a more Westernised lifestyle. Therefore, owing to the anti-atherogenic role of *ApoE* in lipid metabolism and transport, it is important to investigate the role of genetic variations in this gene on the lipid profile of black (South) Africans (refer to Section 2.1.2) on the edge of a CVD epidemic due to westernisation. Previous studies investigating ApoE and its genetic determinants in Africans (Loktionov *et al*., 1999; Murrell *et al*., 2006) had limited statistical power and thus there is a need for a study with adequate power to be conducted in an African population.

When comparing the ε4 haplotype to the ε2 haplotype, the functional properties of the latter do not seem to offer advantages in any environment (Corbo & Scacchi, 1999). Scarce and scattered data of the ε2 haplotype are available and it seems as though the distribution thereof is not as common as that of the Apoε3 haplotype. Its absence can be traced back to the North-east Asian travellers that peopled the Americas and even now, the ε2 haplotype remains rare among various ethnic groups (Corbo & Scacchi, 1999).

Considering the aforementioned, the necessity of studies conducted in different (urbanising) populations should be highlighted, as they may offer important (additional) insights. Even though the reasons for this variation in (CVD) risk among the various populations remain unknown, there may be genetic or environmental modifiers of the *ApoE* gene (Burchard *et al.*, 2003; Hagberg *et al.*, 2000).

3.6 ASSOCIATIONS AND INTERACTIONS BETWEEN *APOE* GENETIC VARIANTS WITH OTHER ENVIRONMENTAL AND PHENOTYPIC TRAITS IN RELATION TO CVD

The Apoε4 haplotype is associated with a 42% increased risk of CVD (Song *et al.*, 2004). Considering the global CVD burden and the associated role *ApoE* plays in the development of CVD, a reduction in the CVD burden associated with *ApoE* genotypes would be of broad public health benefit (Jofre-Monseny *et al.*, 2008).

By just focusing on molecular genetics, one disregards the dynamic nature of populations with their complex relationships between genes, organisms and environments. Although clinical expression may differ between ethnic groups the identification of genetic risk factors provide information on both current risk (elevated lipid concentrations in the presence of variation in the *ApoE* gene) and future risk (desirable lipid concentrations, but may be hyper-responsive to relevant environmental influences in future). Gene-gene and gene-environment interactions are capable of reducing the effect of a variant, since some genes can control the expression of other genes or influence the structure or function of the protein product encoded, whereas an advantageous environment can change the health effect of an original (disadvantageous) genetic trait (Lanktree et al., 2009; Stephens & Humphries, 2003). Therefore, for individuals harbouring any disadvantageous genetic trait, it is important to establish modifiable lifestyle factors that may reduce their risk of developing CVD, *i.e.* smoking habits and total dietary fat intake, as well as the distribution of the different dietary lipid classes. Since the last mentioned factors are modifiable determinants, it could give the individuals carrying the disadvantageous genetic variant the possible opportunity to decrease CVD risk.

It has been suggested that individuals carrying the ε4 genotype are responsive to altered environmental factors (diet, smoking status, alcohol consumption and physical activity) as an approach to reduce the increased risk of CVD (Haan & Mayeda, 2010; Jofre-Monseny *et al.*, 2008). It is important to establish whether there are interactions between genetic variants and the environment, since the latter is modifiable and can be used to reduce

disease susceptibility. Therefore, one of the objectives of this dissertation was to establish possible gene-environment interaction effects, since the results could provide a scientific reason for individuals carrying disadvantaged genetic variants to decrease their CVD risk. However, it is important to highlight the fact that the ε4 allele is not only a risk factor for CVD alone, but also for Alzheimer's disease (Martin *et al*., 2000).

Humphries *et al*. (2001) investigated whether the effect of the *ApoE* genotype on CVD risk would be worsened by tobacco use and reported increased risk in male carriers of the ε4 haplotype (smoking-genotype interaction) even after adjusting for possible confounders (such as increased BMI, blood pressure, fibrinogen, cholesterol *etc*.). It is important to note that the reported results are only applicable to a specific age group (middle-aged), as it is possible that a different effect may be observed among younger or older men (Humphries *et al*., 2001). Furthermore, a strong genotype-smoking interaction was determined with respect to the fat area in the abdominal aorta for the two transcriptional polymorphisms *i.e*. rs440446 ($p = 0.003$) and rs405509 ($p = 0.01$) by Viiri *et al*. (2008).

Inconsistency exists regarding the alcohol-*ApoE* genotype-CVD association. Corella *et al.* (2001a) articulated that the *ApoE* genotype does not have an influence on LDL-C in male abstainers, but for drinkers, the *ApoE* genotype was associated with differing LDL-C concentrations (Apoε2<Apoε3<Apoε4). On the contrary, Corella *et al.* (2001b) reported that drinking alcohol lowered LDL-C in Apoε4 carriers, but increased LDL-C in Apoε2 carriers. The increase in HDL-C associated with alcohol consumption is stronger among individuals without the Apoε4 haplotype than those with the haplotype (Djoussé *et al.,* 2004).

The link between dyslipidaemia and obesity is well recognised in literature; however, whether ApoE influences and contributes to obesity is not clear (Alvim *et al.*, 2010). According to Alvim *et al.* (2010), obese subjects (from a general population) with a BMI ≥ 30 kg/m^2 were more likely to harbour the ε4 haplotype than the ε2 haplotype. Kolovou *et al.* (2009) reported that the occurrence of obesity in CHD patients was also higher in the Apoε4 haplotype carriers than in the Apoε2 counterparts; however, Liberopoulos *et al.* (2004) did not report such an association. Because of the inconsistency of the studies conducted to date and the lack of knowledge regarding the underlying mechanism, future research is needed to draw a valid conclusion.

Gene-nutrient interactions involved in the development of atherosclerosis are being studied; however, the exact mechanisms are unknown (Kolovou & Anagnostopoulou, 2007). Data indicate that the higher LDL-C concentrations observed in subjects carrying the Apoε4 haplotype are primarily apparent in the presence of an atherogenic diet and that the response to dietary SFA and cholesterol may differ among individuals with different ApoE phenotypes. Therefore, individuals carrying the Apoε4 haplotype have been determined to respond best to reduced SFA, total fat and cholesterol intake regimes (Minihane *et al*., 2007). Many studies have been conducted to prove this hypothesis. Some investigators reported greater plasma lipid responses in subjects carrying the Apoε4 haplotype, whereas others failed to find significant associations between *ApoE* genotype and plasma lipid response (Corella & Ordovas, 2005). These differences could be due to the discrepancies observed in the following variables: gender, age and baseline lipid concentrations, which are all reported to play an important role in the variability of dietary response. Little research has focussed on the association between the *ApoE* gene alterations and environmental factors leading to CVD risk development in a (black) South African population. Therefore, it is important to investigate the gene-nutrient (*ApoE*-dietary fat) interaction in the African population under investigation.

To conclude, the present study will be investigating the effect of gene-environment interactions on the lipid profile in order to establish whether a genetic factor plays a significant role in CVD development. This will be with the intention of possibly adding genetic risk factors to the well-known and existing list of CVD risk factors to support the notion that genetic factors might contribute to or modify the means of disease identification and prevention. Furthermore, the studies conducted on the African population only investigated the role of the well-reported epsilons in the *ApoE* gene and did not investigate the possible role of other and novel SNPs present in the *ApoE* gene, which the present study intends to investigate.

CHAPTER FOUR

Materials and methods

This chapter describes the recruitment and characteristics of the subjects included in this study, as well as the experimental methods used. The student contributed during the data collection phase of the South African arm of the PURE study along with other members of the PURE research team. However, some of the data collection and biochemical analyses of the blood samples obtained were finalised prior to the commencement of the investigation of the role of the *ApoE* gene on the lipid profile of the black South African population that was studied. Detailed discussions on the materials and methods applied to obtain the anthropometric, biochemical, clinical, demographic and nutritional measurements, of relevance to the present study, are provided in Addendum A of this dissertation. For the *ApoE* analysis in particular, the student was involved in the deoxyribonucleic acid (DNA) isolation, SNP selection and genotyping of the DNA samples obtained, statistical analysis of the data and the interpretation thereof.

4.1 PROSPECTIVE URBAN AND RURAL EPIDEMIOLOGICAL STUDY

The current study is a cross-sectional epidemiological study of the baseline cohort of black South African individuals, randomly selected from specific rural and urban areas within the 12-year multi-centre international PURE parent study. The data presented in this dissertation include data collected during the baseline of the parent study in 2005 in the North West Province, South Africa (Teo *et al.*, 2009).

4.1.1 Aim of the study

The main aim of the present dissertation was to investigate the effect of identified SNPs in the *ApoE* gene on the lipid profile of black South Africans. This population is currently undergoing a transition from rural to urban areas. To investigate whether urbanisation and other environmental factors together with the *ApoE* genetic variants increase the markers of the lipid profile as CVD risk factors, possible gene-environment interactions were explored. For the extended list of the specific objectives of the current study the reader is referred to Chapter one.

4.1.2 Participant selection

The parent study focused on the epidemiological transition and its effect on health; therefore, representative volunteers were stratified and recruited from rural and urban areas. The sampling approach was based on power calculations using the data obtained from the Transition, Health and Urbanisation in South Africa study that was conducted in the same area 10 years prior to the planning of the PURE study (Vorster, 2002). Fieldworkers conducted a survey of 1 500 black South African households in each of the four communities (two rural and two urban communities, respectively), situated in the North West Province, South Africa. These households were randomly selected and if the head of the household was willing to participate, a written informed consent form (Addenda B-D) and a detailed demographic questionnaire were completed by the members of the household found suitable to participate in the study. However, in the event of the head of the household not being present or refusing to take part in the study, a non-responder questionnaire was completed, after which the fieldworkers proceeded to the next randomly selected household.

During the household survey, 3 750 questionnaires were collected; however, only 2 010 “apparently healthy” black South African volunteers, aged between 35 and 60 years, qualified to be included in the study and voluntarily participated during the data-collection phase in 2005. These subjects reported no usage of medication for CLD, tuberculosis or reported infection with the human immunodeficiency virus (HIV) and the female participants were not pregnant and/or lactating.

4.2 ETHICAL ASPECTS

This study was approved by the Ethics Committee of the North-West University, Potchefstroom Campus (Ethics Approval number: NWU-00016-10-A1). The recruited subjects were requested to give written informed consent, after which they were informed of all the aspects of the study (including genetic analysis to be conducted in future) in their native language (Tswana) before inclusion (Addenda B-D). A description of how the sensitive data (*i.e*. genetic tests) was handled is provided in the informed consent frorm, which in essence assured participants that all genetic tests and experiments would only focus on genotypes alleged to contribute to an increased risk of NCDs of lifestyle (Addedum D). Furthermore, the subjects had the option to withdraw from the study at any

given stage. Confidentiality and anonymity of the results were assured and results were treated accordingly.

4.3 GENETIC ANALYSIS

Genetic analyses were conducted in order to determine whether specific reported and possibly novel SNPs occured in the *ApoE* genes of the participants included in this study and, furthermore, to determine if they had a significant association with any of the lipid profile determinants under investigation. This section describes the analytical pathway used in order to accomplish SNP identification and selection for genotypic analyses.

4.3.1 Deoxyribonucleic acid isolation

DNA was isolated from buffy coat samples from the PURE individuals collected in 2005. The extraction of DNA was achieved by using a modified protocol of the FlexiGene DNA Isolation Kit (QIAGEN). The cells were lysed by mixing one millilitre (mL) of buffy coat with 2.5 mL of buffer FG1 Cell Lysis Solution, in a 15 mL centrifuge tube. The tube was inverted until thoroughly mixed, followed by centrifugation at 2 000 x *g* for five minutes, at 4°C. Most of the supernatant was removed, without unsettling the visible pellet, after which the tube was left inverted on a clean piece of absorbent paper for at least two minutes. A solution of 0.5 mL of buffer FG2 and 5 microlitres (µl) of QIAGEN protease, was added to the tube containing the pellet in order to remove the cellular and histone proteins, which were bound to the DNA. The tube was vortexed for approximately 10-15 seconds until the white blood cells had been resuspended. On determination of complete homogenisation, the tube was inverted three times, after which it was incubated at 65°C for 10 minutes.

Following the incubation, 0.5 mL of ice-cold isopropanol was added to the tube and mixed by inversion until visualisation of the DNA, which precipitated as a clump. The sample was centrifuged at 2 000 x *g* for three minutes. The DNA was mostly visible as a small (white) pellet. The supernatant was decanted and 0.5 mL of ice-cold 70% ethanol (EtOH) solution was added to the DNA.

The tube was inverted several times to wash the DNA pellet as well as the sides of the centrifuge tube. Centrifugation was repeated as indicated in the previous step. The ethanol was carefully aspirated to prevent the DNA from being lost. The tube was left

inverted on a clean piece of absorbent paper again for five minutes, followed by air-drying in a dessicator for several minutes. Five seconds of low-speed vortexing was performed subsequent to the addition of 200 µL of the DNA rehydration solution. The DNA was incubated at 65°C for one hour or overnight at room temperature (25°C). DNA yield and purity were determined for the DNA isolated. The most common technique to determine DNA yield and purity is the measurement of absorbance (http://community.hirebiomed.com/posts/detail/methods-for-determining-dna-yield-and-concentration-4fbbcb5151fed2d37400001e).

DNA purity (A_{260}/A_{280}), defined as the ratio of the absorbance at 260 nm (indicating DNA concentration) and the absorbance at 280 nm (indicating protein concentration) and the yield were determined and estimated using a NanoDrop® (ND-1000) spectrophotometer. DNA absorbs light most strongly at 260 nm (A_{260}), therefore absorbance measured at this wavelength gives a good indication of the DNA concentration in a sample. The A_{260} reading should be within the linear range of the instrument (generally 0.1–1.0) in order for the numbers obtained to be useful (http://community.hirebiomed.com/posts/detail/methods-for-determining-dna-yield-and-concentration-4fbbcb5151fed2d37400001e). The purity ratio (A_{260}/A_{280}) should be >1.8 and can be calculated, after correcting for turbidity (measured by absorbance at 320 nm) with the use of Equation 4.1.

Equation 4.1 Measurement of DNA purity

$$\text{DNA Purity } (A_{260}/A_{280}) = (A_{260} \text{ reading} - A_{320} \text{ reading}) \div (A_{280} \text{ reading} - A_{320} \text{ reading})$$

DNA = deoxyribonucleic acid

The DNA concentration is estimated by measuring the absorbance at 260 nm (while adjusting the A_{260} measurement for turbidity), multiplying by the dilution factor and using the relationship that an A_{260} of 1.0 = 50 $\mu g.mL^{-1}$ pure DNA (http://community.hirebiomed.com/posts/detail/methods-for-determining-dna-yield-and-concentration-4fbbcb5151fed2d37400001e), as indicated in Equation 4.2. A 10 nanogram per microlitre ($ng.\mu L^{-1}$) final DNA working dilution was prepared with rehydration solution, stored at -20°C and defrosted at room temperature prior to use in the genotyping assays.

Equation 4.2 Measurement of DNA concentration

$$\text{Concentration } (\mu g.mL^{-1}) = (A_{260} \text{ reading} - A_{320} \text{ reading}) \times \text{dilution factor} \times 50\ \mu g.mL^{-1}$$

$\mu g.mL^{-1}$ = micro grams per milli liters

Yield was estimated using Equation 4.3. For this equation the DNA concentration is multiplied by the final total purified sample volume (elution volume).

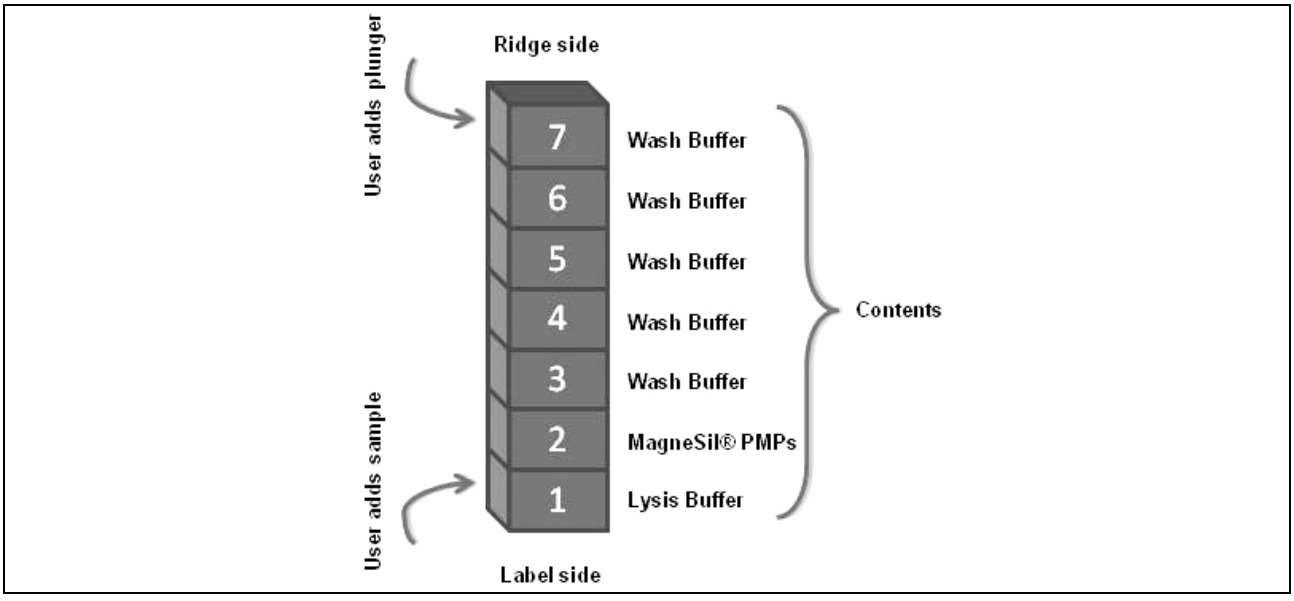
Equation 4.3 Measurement of DNA yield

DNA yield (µg) = DNA concentration × total sample volume (ml)

DNA = deoxyribonucleic acid; ml = milliliters; µg = micrograms

Because of the insufficient DNA yield (<45 ng.µL^{-1}) after the original isolation, certain PURE DNA samples (n=350) were re-isolated using the Maxwell® 16 Instrument, as it is a more effective isolation method in terms of DNA yield than the Flexigene kit. Isolation reactions conducted on the Maxwell® Instrument were set to the standard elution volume (SEV), with the DNA protocol for buffy coat being used. The processing capacity and yield of the buffy coat samples were high, since 500 µL of elution buffer (EB) was added to each 500 µL buffy coat sample for these specific runs. A brief protocol overview for this automated process is as follows: firstly the buffy coat volume was combined with reagents, which were within the Maxwell® 16 SEV Blood DNA Kit [30 µL of Proteinase K (PK) and 300 µL Lysis Buffer]. PK is useful for removing the interaction between DNA and histones, whereas the Lysis buffer was used, as the name implies, to lyse the cell membrane. Sixteen (buffy coat) samples can be processed in each run using the Maxwell® 16 Instrument. The unique plunger design allowed most samples to be purified without the need for pre-processing procedures, *i.e.* PK and other enzyme treatment or centrifugation. The system used preloaded reagent cartridges and plungers (see Figure 4.1).

Figure 4.1 Maxwell® 16 DNA Purification Cartridges



This figure represents the contents of a cartridge for the Maxwell® 16 Blood DNA Purification Kit and a sample is added to well #1 and the plunger is added to well #7; adapted from Promega Corporation (2012).

The basic process that was followed by the Maxwell® 16 Instrument was based on sequential capture and release of MagneSil® Paramagnetic Particles (PMPs), which reversibly bind and purify covalently attached DNA away from cell debris and proteins released upon directed cell lysis, alkaline denaturation and neutralisation, into the wells of the reagent cartridges, which as mentioned, are supplied and loaded with pre-dispensed reagents. The instrument uses a powerful magnet and unique plunger design to lyse and capture the target genomic DNA (gDNA) material while washing away impurities (Kephart *et al*., 2006).

This technique eliminates the need for complex liquid handling and reagent bottles (Shenoi *et al.*, 2007). The reagent cartridges and plunger come into contact with a single sample and are removed and discarded after each run. Single-use reagent cartridges also ensure that new reagents are delivered each time in order to maximise performance and ensure reproducibility. Without cross-contamination of DNA, one blue elution tube for each cartridge was placed into the elution tube slots at the front of the platform, in which 500 μL of EB was added. Once the run of 50 minutes was completed, the elution tubes were removed from the heated elution tube slots and placed into the magnetic elution tube rack from which the DNA was pipetted into storage tubes. An additional 200 μL of EB was added to the same blue elution tube and a second run was completed, to ensure that most of the DNA was obtained.

The downstream analysis of the DNA samples required that the DNA be normalised to a specific concentration. In order to achieve this, the extracted DNA concentrations were measured twice with a NanoDrop® ND-1000 spectrophotometer. The amount of double distilled water (ddH_2O) and DNA from the stock solution that was needed for setting up a working dilution of 50 ng.μL^{-1} was determined by using Equation 4.4 by inserting the mean of the determined measurements.

Equation 4.4 Measurement of DNA absorbance

$$(C \times V)/A = B \qquad V - B = D$$

A = the mean of the two DNA concentration measurements from the stock solution in ng.μL^{-1}; B = quantity of DNA from the stock DNA solution; C = DNA concentration in the working dilution (50 ng.μL-1); D = quantity of ddH_2O to be added to the working dilution; V = final DNA volume (20 μL).

Another absorbance measurement was performed on these normalised DNA samples to ensure that the final working dilution of the samples was within range of 50 ng. μL^{-1}. After

completing the normalisation process, the DNA samples were aliquoted into 96-well plates.

4.3.2 Sequencing of the *ApoE* gene

In general, in order to identify polymorphic sites within a genomic sequence, amplification of the genomic region of interest is required. The *ApoE* gene consists of four exons and three introns (Hagberg *et al.*, 2000; Lusis *et al.,* 1986) and in order to sequence the whole gene, all of these fragments had to be amplified. The amplification of the *ApoE* gene was undertaken in six fragments by conducting polymerase chain reaction (PCR), using six sets of flanking primers for each fragment, which were designed by Dr KR Conradie. The Primer-BLAST program, was used to design the specific primers. Certain conditions had to be adhered to while designing specifc primers to ensure that the abovementioned fragments were successfully amplified. However, one should bear in mind that these conditions were not always attainable and were therefore, used as guidelines for optimal primer design. These conditions included that the melting temperature (T_m) had to be between 52-65 ºC, in order to produce the best results, since a primer with T_m above 65 ºC tends to undergo secondary annealing. Secondly, primer length was between 18-22 bp, which is an efficient length to ensure adequate specificity and easy binding of primers. Furthermore, in order to avoid mispriming, the forward and reverse primers had to have similar T_m, a balanced guanine (G)/ cytosine (C) content (40-60 %) was required and more than three C's or G's in the last five bp at the 3' end of the primer were also avoided to avoid mispriming. For the last condition, the Gibbs free energy (ΔG), which is a measure for the non-specific binding affinity between a candidate primer and its complement primer, was calculated (Han & Kim, 2006). Therefore, to prevent hairpin formation the primer had to adhere to a $\Delta G > -2$ kcal.mol^{-1} at the 3' end. Furthermore, the 3' end self-dimer should optimally adhere to a $\Delta G > -5$ kcal.mol^{-1} or an internal self-dimer with a $\Delta G > -6$ kcal.mol^{-1}. Cross-dimer formation was avoided with a $\Delta G > -5$ kcal.mol^{-1} the 3' end and an internal cross-dimer with a ΔG of > -6 kcal.mol^{-1} was accepted (Ye *et al*., 2012).

Primer sets one to five were optimised by Dr KR Conradie (Table 4.1) before the commencement of the present study. The optimisation of primer set six, along with the amplification of these fragments, was conducted by Ms B. Swanepoel and the author of the present study and were sent for DNA sequencing. The BioRad CFX96 Touch™ Real-Time PCR Detection System was used to amplify the aforementioned fragments in the *ApoE* gene. The optimal annealing temperatures of each primer (for each fragment)

were determined through an optimisation process and the results are summarised in Table 4.1, but will be discussed in greater detail in Section 5.2.

Table 4.1 Primers and PCR protocols used for the sequencing of the *ApoE* gene, per PCR fragment (1 to 6)

PCR Fragment number (size in bp)	Primer	Primer Sequence	Annealing temperature
1 (829 bp)	F R	5'-cct ctc cag att aca ttc-3' 5'-ctc cct tca cat tct aag-3'	56°C
2 (843 bp)	F R	5'-ggt gag gca agc agc agg gg-3' 5'-aac cga ggc cca gag agc gt-3'	67.2°C
3 (918 bp)	F R	5'-tga gat agg agt tag aag ttg-3' 5'-tgt gga aca agt tca agg-3'	53°C
4 (963 bp)	F R	5'-tac ata atg ctt tcc aag tg-3' 5'-caa ctc ctt cat ggt ctc-3'	54.3°C
5 (869 bp)	F R	5'-gcc tcc cac tgt gcg aca cc-3' 5'-ggt cca ccc cag gag gac gg-3'	61.4°C
6 (957 bp)	F R	5'-tct cat cct cac ctc aac ctc ct-3' 5'-aca cag aag gtg ctc cac aaa tgc-3'	73.2°C

ºC = degrees Centigrade; bp = base pairs; µl = microlitres; ddH_2O = double distilled water; DNA = deoxyribonucleic acid; dNTPs = deoxynucleotide triphosphates; $MgCl_2$ = magnesium chloride; mM = millimolar; PCR = polymerase chain reaction; u = units

The PCR reaction consisted of the following reagents: 5 µl buffer, 37.5 nmol $MgCl_2$, 0.5 nmol of each dNTPs, 1.25 U Promega Taq and 50 ng DNA to a final working volume of 15 µL. For fragment six more $MgCl_2$ was used (125 nmol) for reasons discussed in Section 5.2.2. Furthermore, to ensure an optimal PCR reaction the primer concentrations differed per fragment number, with 0.934 nmol forward and reverse primers being added to PCR fragment numbers one, four and five, respectively. For PCR fragment number five, dimethyl sulfoxide was added (62.5 µmol) to inhibit secondary structures in the DNA primers, which often occurred during the optimisation process. For PCR fragment number two, 0.141 nmol forward and reverse primers were added and for PCR fragment three, 0.104 nmol forward and reverse primers. In order to have a final working volume of 15 µL, the mastermix solutions were made up with ddH_2O and, therefore, each PCR fragment's volume of ddH_2O differed.

The thermal cycling conditions of the PCR protocol that was used to achieve amplification were as follows: the initial denaturation step was at 95ºC for two minutes; the cycling conditions were for 40 cycles and included three steps. The first step was the denaturation step at 95ºC for 60 seconds, the second step was the annealing step at the specific annealing temperatures (T_a) for 60 seconds and the third step was elongation at 72ºC for 60 seconds. A final holding step was also included at 15ºC.

Concluding the PCR process, the success of the amplification (of the PCR products) was tested with the use of 2% agarose gel electrophoresis. The samples were loaded with 2 µl of Bioline loading buffer into the gels, which consisted of 2% molecular grade agarose (BioRad), of which the final working Tris/Borate/EDA buffer concentration (1X) consisted of 89mM Tris base, 89mM Boric acid, 2 mM EDTA and 0.5 $\mu g.mL^{-1}$ ethidium bromide, which was then electrophoresed at 100 volts for an hour. Graphical representations of the fragments were obtained using an ultraviolet transilluminator, after which electronic copies of the image were captured, which will be presented and discussed in Section 5.2.3.

The successfully amplified PCR products and primers were sent off to the Central Analytical Facility (Stellenbosch University) for DNA sequencing in both the forward and reverse direction of the *ApoE* gene in 30 randomly selected samples from the 2 010 PURE samples. The NucleoFast® 96 PCR Clean-Up Kits from Machery-Nagel were used to clean the PCR product before performing the sequencing process. The basic principles of these kits are that they are based on ultra-filtration and designed for rapid clean-up of PCR fragments, which entails three steps to complete this procedure. Firstly, the PCR samples are applied to the ultrafiltration membrane, secondly the PCR products collect on the filters while the primers, deoxynucleotide triphosphates (dNTPs) and salts move through the filter and are collected into a waste tube. The amplified products are retained on the filter membrane and are released from the filter membrane after the addition of either nuclease-free water or a low-salt buffer, as the third step (PCR Clean-up User Manual, 2010). These PCR products are ready to use for downstream applications such as sequencing (PCR Clean-up User Manual, 2010).

The purpose of sequencing is to determine the order of the nucleotides (A, T, C and G) in a specific section of a DNA strand. The Central Analytical Facility at Stellenbosch used the ABI Prism®, BIGDye® Terminator version 3 Ready Reaction Cycle Sequencing Kit from Applied Biosystems as a platform to conduct the sequencing reactions. The reaction mixture included the template DNA, free dNTPs and four (A, C, G and T) types of fluorescently labelled 2', 3'-ddideoxyribonucleotide triphosphates (ddNTPs), an enzyme (usually a variant of *Taq* polymerase) and a primer. The sequencing reaction consists of three major steps repeated for 30 to 40 cycles, which are similar to those of a PCR reaction. The reaction is initiated by heating until the two strands of the template DNA separate, after which a specifically designed primer binds to its assigned location and the DNA polymerase begins elongating the primer. This results in a new strand of (replicated) DNA being formed, which is allowed to elongate until it incorporates the fluorescently

labelled ddNTPs. The role of the ddNTPs is to terminate the elongation of the DNA strand, at which point the polymerase then falls away. The basic principle of elongation is that when a specific nucleotide is required to make the new strand, the polymerase enzyme will get the required nucleotide from the mixture. After adding the specific required nucleotide, the enzyme continues to add more nucleotides. However, 5% of the time, the enzyme will get a ddNTP, which will terminate the elongation step. All of the copied DNA will eventually be terminated, but each time the enzyme produces a new copy, the termination point will randomly differ, resulting in various fragment lengths and sizes, which need to be separated and measured.

A variation of the Princeton Separations post-sequencing kit is used to purify the sequenced product, by means of removing unincorporated fluorescently labelled ddNTPs *via* perforated beads, which retain these ddNTPs and dNTPs to allow the larger fragments to pass through. An ultraviolet laser (built into either an ABI3730xl DNA analyser or an ABI3120xl genetic analyser), excites the fragments (from the purified product) as they emerge from the end of the capillaries, and the resultant pulse of fluorescent colours (red, green, blue and yellow, each representing one of the four nucleotides) is measured and used to form a chromatogram.

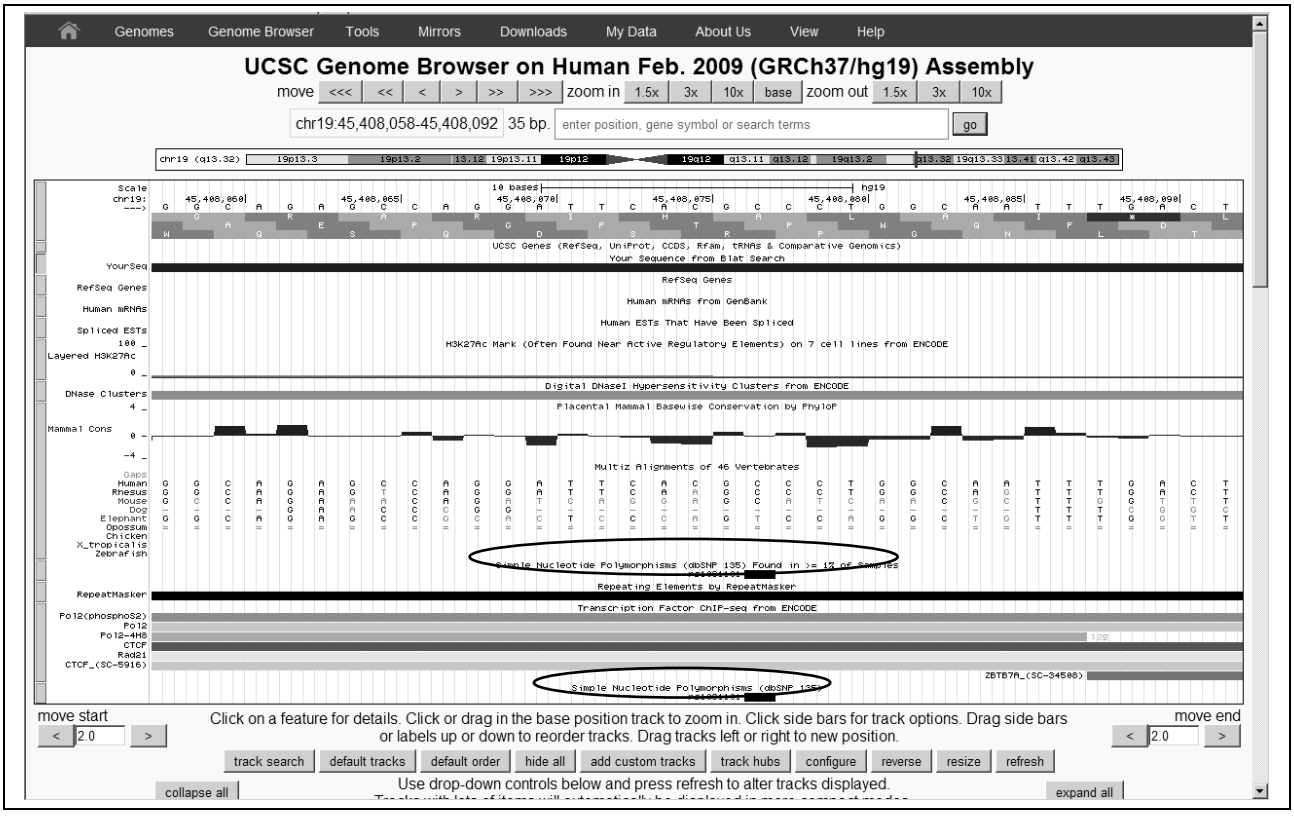
4.3.3 The identification and selection of SNPs within the *ApoE* gene

The process involved in identifying the SNPs in the *ApoE* gene was as follows: the raw forward and reverse sequence electropherograms of 30 randomised samples from the PURE study were received from the Central Analytical Facility at Stellenbosch, which the author of the present study imported into the CLC Workbench sequence viewer program and aligned to the reference *ApoE* gene (Genbank Accession number AF261279) using the alignment functions of the CLC Workbench sequence viewer program. Thereafter, the consensus sequence data for each of the 30 random samples were aligned to the reference gene that was used as a template to identify alterations (SNPs) in the *ApoE* gene through manual inspection. The initial SNPs that were identified, *via* the 30 random samples from the PURE study which were sequenced, were either previously reported SNPs, which had a reference sequence number (rs-number), or novel SNPs.

To determine if an SNP had previously been reported and, therefore, had an rs-number available, the status of the alteration was established using the University of California Santa Cruz (UCSC) Human Genome Assembly Browser program. With the aim of finding

the location of each alteration in the genome, the Human BLAT search tool of the UCSC Human Genome Assembly Browser was used, whereby approximately 20 bp of sequence on either side of the alteration was submitted in the query box. From the BLAT search results the position of the alteration was determined, which was in turn used to determine if the alteration had previously been reported. This was achieved by entering the chromosome location into the genome browser functionality. The dbSNP tract in the resultant output was used to determine if the SNP had previously been reported and was indicated with a black mark along with the rs-number of interest, as illustrated in Figure 4.2. This procedure was followed for each alteration that was identified in the sequenced data of the 30 random samples from the PURE study.

Figure 4.2 Print screen of the UCSC Genome browser used to obtain the rs numbers for identified SNPs



The rs-number obtained for a given alteration within the sequenced data is encircled in red

In addition to the SNPs determined in the 30 random samples from the PURE study, a computer-based literature review of *ApoE* SNPs reported in the literature was also carried out. Since most of the literature is based on the frequently reported "*ApoE* polymorphisms", *i.e*. ε2, ε3, ε4, the need arose to obtain scientific articles that reported on any other SNPs in the *ApoE* gene. The search terms that were used were: "Apolipoprotein E" and "genetics", "Apolipoprotein E" and "single nucleotide polymorphisms", "Apolipoprotein E" and "CVD", "Apolipoprotein E" and "lipid profile". A

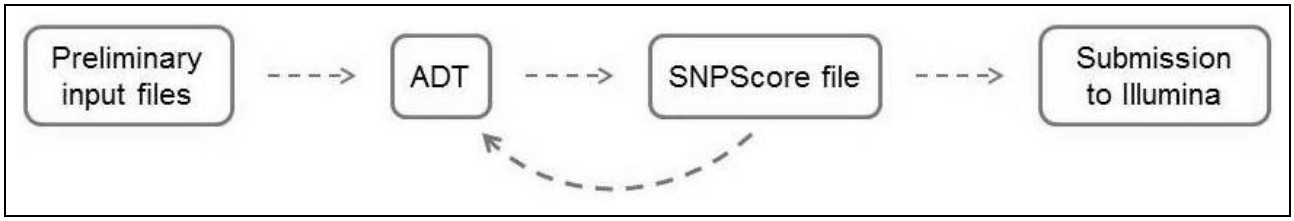
limited number of articles that reported on other SNPs within the *ApoE* gene, besides the frequently reported polymorphisms, stemmed from the search results (Minihane *et al*., 2007; Fullertone *et al*., 2000; Nickerson *et al*., 2000). However, these articles were examined in further detail and all the SNPs that were reported in these articles, which were not identified from the aligned electropherograms of 30 randomised samples from the PURE study, were added to the SNP selection list. After applying these methods, 20 SNPs were initially included in the SNP list.

4.3.4 Process of SNP identification and validation using the BeadXpress® platform

To ensure that the correct SNPs were analysed by the BeadXpress® platform, the custom GoldenGate genotyping assay was used. The GoldenGate genotyping assay is a robust assay for multiplex genotyping and can be analysed on the BeadXpress® platform using VeraCode® technology. A positive attribute of this assay is that it allows researchers to create their own genotyping panels with the assistance of the assay design tool (ADT). The ADT provides a straightforward method for the evaluation of individual loci to design a customised genotyping array.

An SNPScore file provides information on the SNPs to be analysed, such as their validation status, MAF from published studies and designability rankings, along with SNP scores, which in turn indicate predicted success information of the specific SNPs of interest. This array of information was also used as a guideline to determine whether a certain SNP should be rejected or included in the final SNPScore file for analysis. Attaining a final SNPScore file involves a number of steps and procedures, as illustrated in Figure 4.3.

Figure 4.3 Steps involved in the custom GoldenGate genotyping assay design

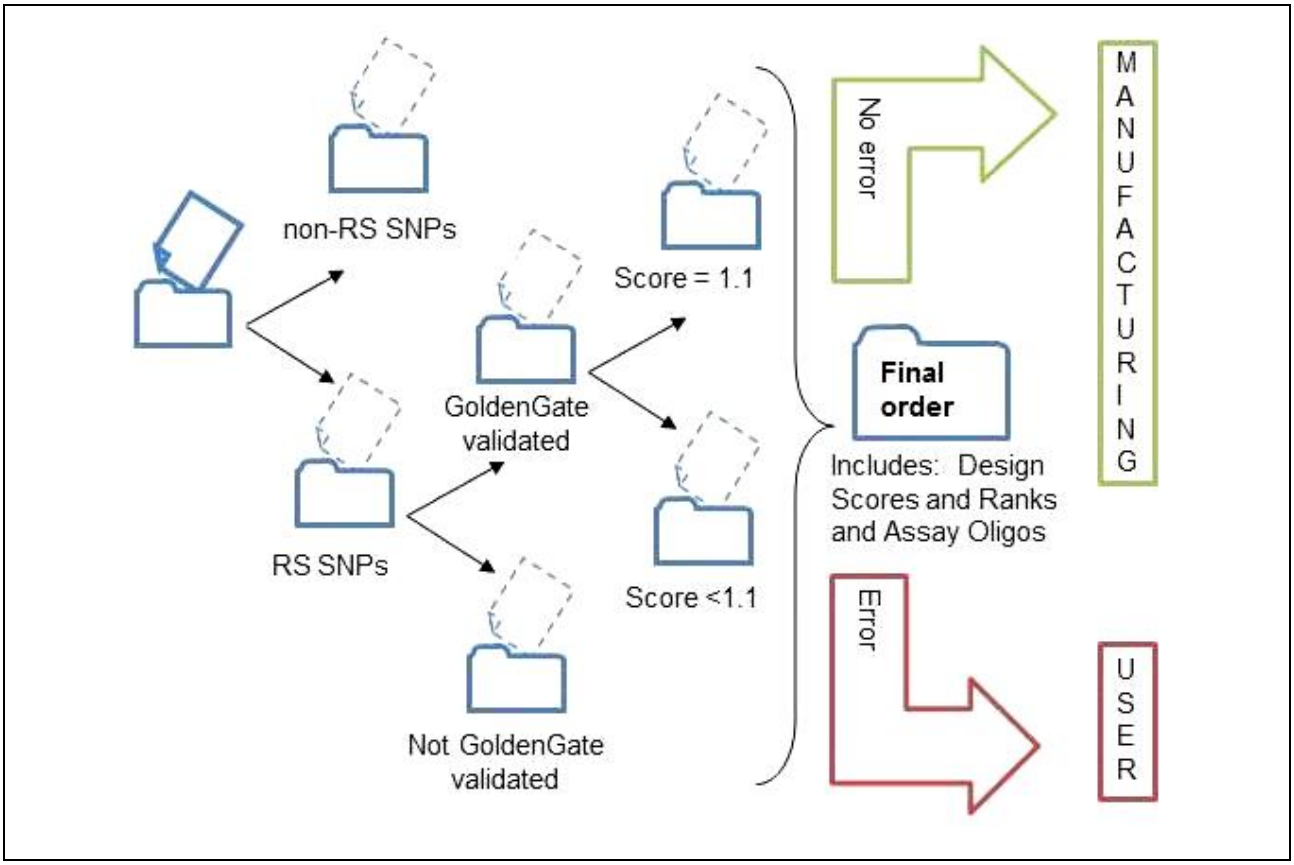


ADT = assay design tool; SNP = single nucleotide polymorphism

The two file formats used for the submission of the SNPs were the RSList (when investigating a reported SNP with a specific rs number) and SequenceList (when investigating a novel SNP). The SequenceList file format was used for the three novel

SNPs identified in the sequenced data. Once these lists had been finalised they were submitted to ADT (as a *.csv file format), where dbSNP was used to obtain information about the locus, followed by the determination of the MAF and validation status of the SNP for each SNP (except for the SNPs listed on the SequenceList) after which the design score and design rank for every SNP were calculated. Designing and ordering a custom oligonucleotide pool assay (OPA) entail further steps (Figure 4.4).

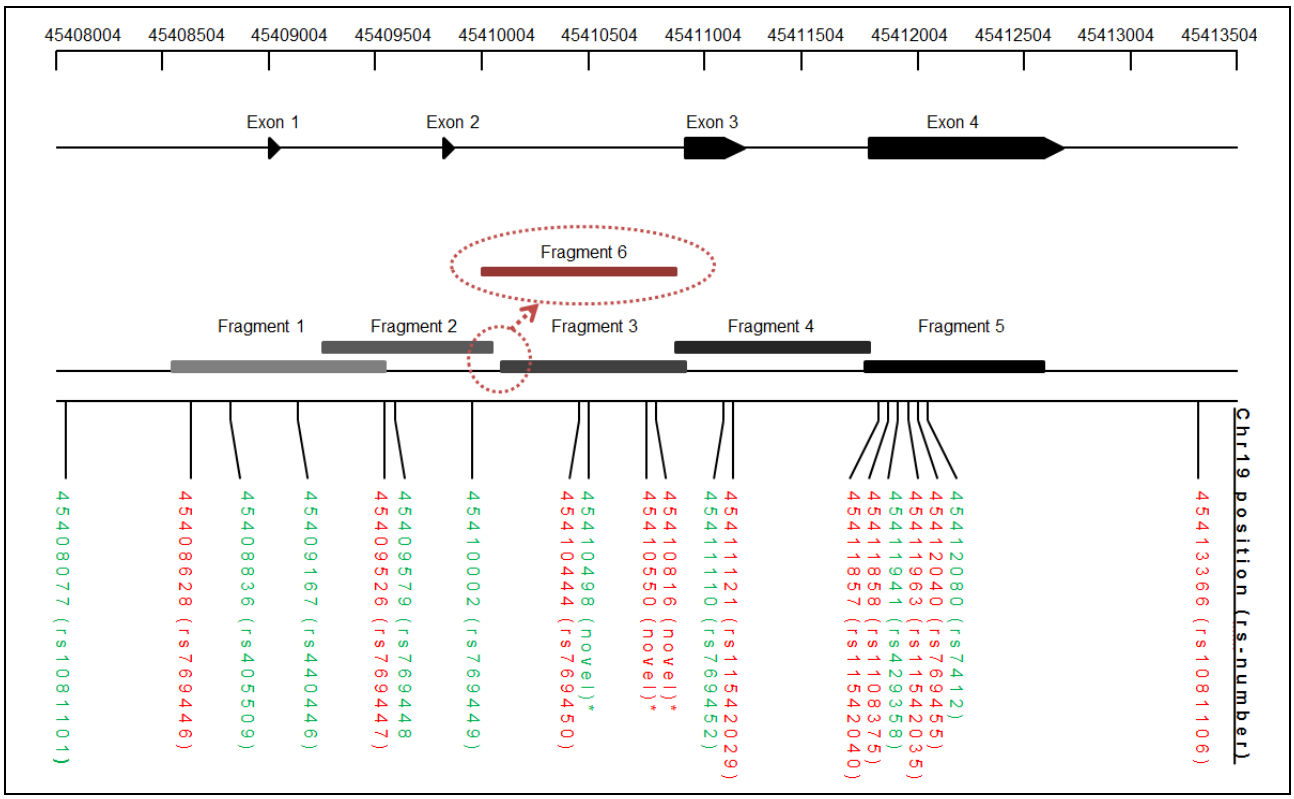
Figure 4.4 Flowchart of the steps involved in ordering a custom OPA



RS = reference sequence; SNP = single nucleotide polymorphism

Figure 4.4 is discussed in detail in Section 5.3.1 referring to the validation scores obtained for the identified SNPs that were to be investigated *via* the GoldenGate® Genotyping assays. The genome locations of the 20 initially identified SNPs (indicated in red) and the nine SNPs selected to be part of the final SNP list (indicated in green) are illustrated in Figure 4.5.

Figure 4.5 Location of the identified and selected SNPs in the ApoE gene



* indicates the location of the novel SNPs identified in the sequenced data; rs = reference number; adapted from Fullertone *et al*. (2000)

From an extended list of identified SNPs, only those listed in Table 4.2 and indicated in green in Figure 4.5 were sent for analysis by the BeadXpress® T Reader. The criteria that were used to determine which SNPs should be analysed by the BeadXpress® T Reader are discussed in Section 5.3.1.

Table 4.2 Summary of the characteristics of the final SNPs in the ApoE gene sent for genotyping by the BeadXpress® T Reader

Chr 19 Position	Strand	Localisation	rs-number	SNP	Type of mutations
45408077	Pos	Flanking	rs1081101	C/T	Transition
45408836	Pos	Promotor	rs405509	A/C	Transversion
45409167	Pos	Intron	rs440446	C/G	Transversion
45409579	Pos	Intron	rs769448	C/T	Transition
45410002	Pos	Intron	rs769449	G/A	Transition
45410498	Pos	Intron	*ApoE*LS01 novel	G/C	Transversion
45411110	Pos	Exon	rs769452	C/T	Transition (Leu/Pro)
45411942	Pos	Exon	rs429358	T/C	Transition (Cys/Arg)
45412080	Pos	Exon	rs7412	C/T	Transition (Arg/Cys)

A = adenine; Arg = arginine; bp = base pair; C = cytosine; Chr = chromosome; Cys = cysteine; G = guanine; Leu = leucine; Pos = positive; Pro = proline; rs-number = reference sequence number; SNP = Single Nucleotide Polymorphism; T = thymine

4.3.5 Genotyping using the BeadXpress® T Reader

In order to determine *ApoE* genotypic status, a DNA-based technique was used. A relatively simple, time-efficient as well as cost-effective high-throughput genotyping method was used to analyse the larger sample set. Genotyping was conducted on the BeadXpress® T Reader from Illumina, at the National Health Laboratory Service (NHLS) at the University of Witwatersrand in Johannesburg, over two days.

The BeadXpress® can analyse 48 to 384 SNPs simultaneously in a 96-well plate owing to the solution-based kinetics of the VeraCode technology and the simplicity of the assay's workflow. The Illumina GoldenGate SNP genotyping assay is analysed on the BeadXpress® platform to accommodate PCR extension and amplification methods that allow a quantitative comparison of bi-allelic SNPs. The assay's workflow for day one includes four steps. The first step in generating genotypic data is the activation of DNA, which is a robust process for which only 250 ng of genomic DNA is needed. This step is necessary in order for genomic DNA samples to bind to paramagnetic particles. During the hybridisation step (step 2), assay oligonucleotides, hybridisation buffer and paramagnetic particles combine with the activated DNA. There are three oligonucleotides [two allele-specific oligonucleotides (ASO) and one locus-specific oligonucleotide (LSO)] that are designed for each SNP locus. The LSO hybridises to numerous bases downstream from the SNP site. The LSO contains a unique address sequence that targets a particular VeraCode microbead type, along with universal PCR primer sites, which all three oligonucleotide sequences contain. The ASOs and LSOs hybridise to the genomic DNA sample bound to the paramagnetic particles during the hybridisation process. A number of wash steps (step 3) are conducted after the hybridisation step in order to remove all excess and mis-hybridised oligonucleotides to guarantee the inclusion of the hybridised oligonucleotides alone. Thereafter, the extension of the applicable ASO and ligation of the extended product to the LSO link the information for the genotype present at the SNP site to the LSO's address sequence. These ligation products serve as PCR templates for the universal PCR primers, namely P1, P2 and P3.

On the second day of determining the genotypic data, five steps are taken, of which step one is to label the P1 and P2 primers with Cy3 and Cy5 dyes, respectively. These Cy3 and Cy5 fluorescently labelled PCR extension products contain different Illumicode address sequences, which are the Illumina GoldenGate Assay products (Table 5.3 provided in Section 5.3.1) and are the starting experimental material for the BeadXpress®

platform. These products then hybridise to small glass bars (VeraCode® "microbeads"), which have imprinted bar codes complementary to the Illumicode (sequence) address, which corresponds to a specific genomic sequence provided in the SNP design file (step 2). The reader pipettes a column (eight samples) of a 96-well plate at a time and aliquots them into a grooved plate that forms the bottom of the system's fluidic cell. The VeraCode microbeads fill the groove plate efficiently and align closely within the grooves through a unique combination of fluid flow, gravity and capillary force. The entire fluidic cell is activated across the optical system and scanned for fluorescent intensity and code classification, after the amplification process has been completed. The BeadXpress® Reader makes use of a dual-detection system (each plate is scanned at two wavelengths simultaneously) in order to identify the unique holographic code on each bead and the signal intensity associated with each bead. Hybridisation of the GoldenGate Assay products onto the VeraCode microbeads separates the assay products for individual SNP genotype readout. The hybridisation process takes place in suspension for several hours, after which the beads are washed, resuspended and loaded onto the reader (step 3). Step 4 entails the use of the BeadXpress® Reader for the identification of microbead codes, as well the detection of the relative Cy3 and Cy5 fluorescent signals. As a result, a signal should be detected in either one of the Cy3 or Cy5 channels for the amplification product for homozygote genotypes and both signals should be detected for a heterozygote genotype, for each SNP that has been genotyped. Lastly, the genotypic data generated by this method are analysed *via* Illumina's GenomeStudio® Genotyping Module v1.0 program, which performs automated genotype clustering and calling. The BeadXpress® has quality control measurements in place to ensure that each assay runs correctly. This is achieved through checking each output from each plate run, using standardised technical parameters (controls) and determining PCR contamination, which is discussed in greater detail in Section 5.3.2.

However, the SNP at rs7412 could not be successfully genotyped by the BeadXpress® platform for reasons discussed in the following chapter, Section 5.3.1. As this SNP formed a central part of the genetic component of the present study, the genotyping was carried out manually by Mr T. Chikowore through a real-time PCR assay using the CFX96 Touch™ Real-Time PCR Detection System, which will be discussed in greater detail in the next section.

4.3.5.1 Genotyping of the *ApoE* polymorphism at rs7412

The real-time PCR Taqman assay was developed for the detection and quantification of genetic material. This method combines a fluorogenic element with that of the traditional PCR methods (Klein, 2002). The probe, a small oligonucleotide segment with a bound fluorescent reporter and quencher, is designed to hybridise to a DNA sequence on the target DNA between two flanking primers. During the annealing phase, the probe hybridises to its specific complementary segment of the targeted DNA sequence. The role of a quencher at the end of each probe sequence is to block the fluorescein so that the real-time PCR cannot detect the probe, until the Taq polymerase enzyme meets and cleaves to the probe after it annealed to the target DNA, causing the probe to break away and become excited. The excitation that occurs at this point is due to the fact that with the cleavage of the probe, the quencher is further away and cannot suppress the excitation of the probe anymore. This excitation is detected in real-time by the thermocycler.

This polymorphism at rs7412 was genotyped in 2 010 subjects, which represents the whole PURE group. The polymorphism is located in exon four. For the SNP at rs7412 located at the bp 4075, two alleles are described. The C-allele represents the “wild-type” allele, whereas the T-allele represents the “mutant”. Therefore, it is possible to distinguish between three genotypes: homozygote 1 (CC for the purpose of the present study), heterozygote 1, 2 (CT) and homozygote 2 (TT).

A probe [6-caboxyfluorescien (FAM™) and 2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein (VIC®)] and primer mix from Applied Biosystem (TaqMan® SNP Genotyping Assays, Human, LG Part Number 4351374) were used in the analyses. The FAM™ probe recognised the C allele (wild type) and the VIC® probe recognised the T allele (mutant). The PCR reaction mixture used in this analysis consisted of 37.5 nmol mastermix, 6.3 nmol primer and probe mix and 50 ng of the patient DNA, along with the ddH_2O that was added to make up a final working volume of 15 µl.

The real-time PCR reactions were carried out in a CFX96 Touch™ Real-Time PCR Detection System, according to the following programme: one cycle for 10 minutes at 95°C (initial denaturation), 40 cycles at 92°C for 15 seconds (denaturation); the specific optimised annealing temperature for the primer set was 62.1°C for 60 seconds (Table 4.3).

Table 4.3 Real-time PCR temperature conditions

Stage	PCR Step	Temperature	Time
Cycle 1 (1x) : Step 1	Initial denaturation	95°C	10 minutes
Cycle 2 (40x): Step 1	Denaturation	92°C	15 seconds
Step 2	Annealing/elongation	62.1°C	1 minute

°C = degrees Centigrade

With the completion of each 96 well reaction, the data were stored in the Bio-Rad CFX Manager© (version 2.0), before being exported to an Excel work sheet. From this work sheet, which contained the genotypic data of the whole PURE cohort of that blood was obtained (2 028), statistical analyses were conducted. The author of the present dissertation and Dr KR Conradie scored the genotypes at the *ApoE* rs7412 loci independently and merged the electronic spreadsheets with the outcomes. Discrepancies in interpretation were resolved by re-determination of the sample DNA genotypes.

4.4 STATISTICAL ANALYSIS

The computer software package Statistica version 10 (Statsoft Inc., Tulsa, Oklahoma, USA) was used for the statistical analysis (Statistica, 2009). A p-value of ≤0.05 was regarded as statistically significant. Furthermore, for the present study any significant result for which a p-value of 0.000000 was obtained will be reported as p = 0.00 in Chapter five, but represents a p-value of <0.001. Only the samples for which all information was available were included in the statistical analysis for the present study. The sample size consisted of 2 010 subjects, but subjects that had either missing phenotypic or genotypic data were excluded from the analyses and, therefore, only 1 458 subjects were ultimately included in the analyses. The outcome variables are the lipid profile determinants: TC, LDL-C, HDL-C and TG concentrations.

4.4.1 Stratification criteria

The PURE data set was divided into two stratums namely, urban and rural. Gender was another stratum categorised into male and female. The HIV status was categorised as follows: 0 was assigned to those infected by the virus, 1 to individuals who did not know their status and 2 to those not infected by HIV. History of tobacco use was categorised into four divisions: 0 = Not answered, 1 = Former smoker, 2 = Current smoker and 3 was assigned to subjects who had never used tobacco before. For the purpose of the present study, stratification will also be conducted for the different genotypes identified in the *ApoE* gene.

4.4.2 Normality testing/distribution

Initially, the continuous data were tested for normality using the Shapiro-Wilk (SW)-test. The data were not normally distributed in the population, even after logarithmic transformation has been conducted. The option of using nonparametric tests is often inconvenient, as these tests are in general less powerful and flexible in terms of the types of conclusions that they can supply. Even though the present study consists of a large sample set, which is appropriate for data transformation, a scientifically safe decision was made not to perform log transformations, as doing it presented no difference in the normality of the data. Although there is a possibility of failure to detect a difference that really exists between the relevant variables, the possibility of being misled into claiming differences determined to be spurious by transforming the data is minimised or eluded. The reason for this is that if data are not transformed, it is more likely that the variation in the data will be overestimated compared to the appropriate transformation (van Emden, 2009). It has been reported that with a larger sample size, the mean tends to be closer to the median (central limit theorem), which implies that normality can be assumed and that mean-based tests can be used (Rice, 2007). Therefore, the statistical analyses presented in the following chapter were conducted according to parametric statistical analysis.

4.4.3 Statistical testing

Descriptive statistics were determined for all continuous variables relevant to the present study (Table 5.1 and Table 5.2) to get a broad overview of the present data set used. The means and standard deviation (±SD) of the identified continuous variables by gender (men against women) and locality (rural against urban) were reported. The mean lipid profile concentrations in each of the categories of the categorical data (BMI, history of tabacco use and HIV status), were calculated by means of cross-tabulation and analysis of variance (ANOVA). Confounding variables for the lipid profile were identified by means of the Spearman Rank test. This nonparametric test was used to determine the correlations, since normality is already a requirement in order do to correlation calculations and this test is not based on means.

The genotype frequencies of the different SNPs under investigation were determined for the whole group and according to locality and gender. A chi-square (χ^2) test was used to compare the observed numbers of each *ApoE* genotype with those expected for a population in Hardy-Weinberg equilibrium (HWE). The equation for the chi-square test is

$\chi^2 = (O-E)^2/E$. A high chi-square value and small p-value (<0.05) will lead to the rejection of the hypothesis that the genotypes were in HWE. The Hardy-Weinberg principle states that both allele and genotype frequencies in a population remain constant — that is, they are in equilibrium — from generation to generation unless specific disturbing influences are introduced. Those disturbing influences include non-random mating, mutations, natural selection, limited population size, "overlapping generations", random genetic drift, gene flow, meiotic drive and migration (Hardy, 1908). It is important to understand that one or more of these "disturbing influences" are always in effect. That is, perfect adherence to the HWE is close to impossible in nature. However, by comparing the observed frequencies with expected genotypic frequencies one is better able to measure the possible causes of the differences present in genotypic distributions between populations. Genetic equilibrium is an ideal state that provides a baseline against which to measure change. Static allele frequencies in a population across generations assumes: no mutation (the alleles do not change), no migration or emigration (no exchange of alleles between populations), infinite population size and no selective pressure for or against any genotypes. Genotype frequencies will also be static when mating is random. Furthermore, the 95% confidence intervals for the frequencies of these genotypes were determined by means of the SAS System Program, to determine the range within which the actual genotype frequency should be in the total population.

To determine whether a significant difference between the genotype groups of a specifc genetic alteration (SNP) and the lipid profile concentrations existed, ANOVA tests were applied. Where an overall significant difference between the above-mentioned groups occurred, a Tukey honestly significant difference *post hoc* test for unequal *n* was used to determine between which groups the significant difference occurred. Lastly, the interactive effects between various variables (locality, gender, diet and each of the confounding variables) against each of the lipid profile determinants, categorised by the three genotypes of each specific genetic variant, were also analysed. For the categorical predictors (gender and locality) factorial ANOVAs from the general linear analyses were used, whereas the general linear model was used where more than two continuous predictors (diet and each of the confounding variables) were present. Where an overall significant difference between all of the above-mentioned groups/ instances was indicated, respectively, a Tukey honestly significant difference *post hoc* test for unequal *n* was used to determine between which groups the significant difference/interaction occurred.

CHAPTER FIVE

RESULTS AND DISCUSSION

The population of South Africa is multi-ethnic and a wide range of cultures are represented within this population. For many years it was believed that black (South) Africans were immune to the development of CVD even when previously reported CVD risk factors were present (Walker & Sareli, 1997). Today, CVD is a major cause of morbidity and mortality in all the cultural groups of South Africa (Klug, 2012; Steyn *et al*., 2005; Akinboboye *et al*., 2003). Steyn *et al*. (2005) reported that the black African population is at an early stage of the CVD epidemic as well as other epidemiological disease transitions, and that this increased disease risk is more prevalent among groups of higher socio-economic and educational status. The change in CVD prevalence can be ascribed to urbanisation and the accompanying unhealthy lifestyle changes (Akinboboye *et al*., 2003).

An unfavourable lipid profile as represented by elevated TC and LDL-C concentrations, in addition to low HDL-C concentrations, are major determinants of CVD (Atkinson *et al*., 2004; Cleeman *et al*., 2001). In 1990 Holme estimated that for every 1% reduction in blood cholesterol there is a corresponding 2.5% reduction in the incidence of CVD (Holme, 1990). Therefore, research endeavours have been undertaken to understand the mechanistic role of certain (risk) factors in the regulation of the blood lipid profile in order to provide evidence that can be used to develop strategies to improve the lipid profile and ultimately reduce CVD morbidity and mortality rates.

However, not all individuals respond to such lipid-lowering measures or strategies in a similar fashion, which can be attributed to the presence of alternative genetic backgrounds in different human populations (Cummings, 1998). Biological insights into novel and reported lipid metabolic pathways can be obtained with the identification of genetic variants that affect lipoprotein concentrations, *i.e*. TG, HDL-C and LDL-C concentrations (Lanktree *et al*., 2009). In the last two decades common variations in many candidate genes have been tested for association with lipoproteins (Chang *et al*., 2010; Lanktree *et al*., 2009; Pallaud *et al*., 2001). The *ApoE* polymorphisms, as described in Section 3.2, have been reported to be one of the potentially most important genetic predictors of plasma lipoprotein concentrations and, therefore, risk of CHD (Song *et al*., 2004; Frikke-Schmidt *et al*., 2000; Mahley & Rall, 2000; Davignon *et al*., 1988).

While the association between specific *ApoE* polymorphisms, environmental factors and well-known CVD risk factors and events in Caucasian individuals has been studied comprehensively, less has been reported about these associations in black (South) African populations. Furthermore, since CVD is a multi-factorial disease, it is crucial to determine whether gene-environment interactions occur at the *ApoE* locus and whether these interactions might possibly explain the variability in the blood lipid profile in black South Africans. Therefore, this study aims to identify key determinants of plasma lipoprotein concentrations and possibly provide insight into the biological mechanisms underlying serum lipid and cholesterol concentrations in a black South African population.

The following sections present the results obtained after the implementation of the techniques and processes as described in Chapter four. The baseline demographic and biochemical characteristics (obtained in 2005) of the PURE study population are examined in Section 5.1. This is followed by Section 5.2, which contains discussions on the DNA isolation process, the PCR and sequencing results of the *ApoE* gene, as well as discussions on the amplification of the various gene fragments in which the polymorphisms were either identified from sequencing this unique population or were identified from literature. The results obtained from the BeadXpress® analysis are discussed in Section 5.3, followed by a similar discussion pertaining to the real-time PCR analyses conducted (Section 5.4). The associations and correlations between the concentrations of the various lipid profile determinants with the various CVD risk markers, *e.g*. BMI and environmental factors (dietary habits), are discussed in Section 5.5, along with the interactive effects these variables have with the *ApoE* genetic variants (SNPs and haplotypes) on the blood lipid profile determinants. This chapter will be concluded with a summary of the most important results (Section 5.6).

5.1 DEMOGRAPHIC AND BIOCHEMICAL CHARACTERISTICS OF THE PURE POPULATION

The baseline characteristics of the PURE study participants are presented in Table 5.1. This study population was representative of both rural and urban areas, with 49.5% of the subjects residing in rural areas and 50.5% residing in urban areas. The study population comprised 37.2% men and 62.8% women.

Table 5.1 Continuous baseline characteristics per gender and locality, within the PURE study population

Variables	Whole group Mean ± SD	Rural Mean ± SD	Urban Mean ± SD	p-value	Men Mean ± SD	Women Mean ± SD	p-value
Number of individuals	1,950	966	984	---	726	1,224	---
Age (years)	49.30 ± 10.40	48.47 ± 9.94	50.20 ± 10.69	0.00	49.76 ± 10.30	49.10 ± 10.40	0.16
BMI ($kg.m^{-2}$)	24.60 ± 6.96	24.16 ± 6.61	25.16 ± 7.29	0.00	20.80 ± 4.04	26.90 ± 7.32	0.00
TC ($mmol.L^{-1}$)	5.01 ± 1.38	4.96 ± 1.37	5.05 ± 1.39	0.17	4.81 ± 1.33	5.13 ± 1.39	0.00
TG ($mmol.L^{-1}$)	1.30 ± 0.79	1.21 ± 0.64	1.38 ± 0.92	0.00	1.22 ± 0.85	1.34 ± 0.75	0.00
LDL-C ($mmol.L^{-1}$)	2.92 ± 1.17	2.92 ± 1.17	2.93 ± 1.18	0.87	2.70 ± 1.15	3.06 ± 1.17	0.00
HDL-C ($mmol.L^{-1}$)	1.52 ± 0.64	1.52 ± 0.62	1.52 ± 0.65	0.92	1.58 ± 0.60	1.48 ± 0.62	0.00
CRP ($mg.L^{-1}$)	8.52 ± 12.45	8.34 ± 12.57	8.70 ± 12.3	0.53	8.33 ± 13.40	8.64 ± 11.90	0.61
PAI-1act ($IU.L^{-1}$)	5.98 ± 7.20	4.96 ± 5.89	7.04 ± 8.22	0.00	4.45 ± 6.88	6.92 ± 7.23	0.00
Hcy ($\mu mol.L^{-1}$)	10.30 ± 4.57	10.82 ± 4.80	9.84 ± 4.26	0.00	11.30 ± 4.47	9.78 ± 4.54	0.00
Fibrinogen ($g.L^{-1}$)	3.70 ± 2.18	3.86 ± 2.21	3.51 ± 2.13	0.00	3.33 ± 2.07	3.92 ± 2.21	0.00
SBP (mmHg)	133.00 ± 24.5	129.72 ± 23.29	137.00 ± 25.10	0.00	136.00 ± 24.10	132.00 ± 24.70	0.00
DBP (mmHg)	87.70 ± 14.6	86.16 ± 14.48	89.30 ± 14.50	0.00	86.90 ± 14.90	88.20 ± 14.30	0.04
Total energy intake (kJ)	7,932 ± 3,783	6,410 ± 2,778	9,426 ± 4,036	0.00	8,639 ± 4,050	7,512 ± 3,551	0.00
Total fat intake (% TE)	22.80 ± 7.63	19.15 ± 6.97	26.40 ± 6.46	0.00	21.50 ± 7.31	23.50 ± 7.72	0.00
Total MUFA intake (% TE)	6.00 ± 2.92	4.43 ± 2.39	7.54 ± 2.55	0.00	5.68 ± 2.79	6.19 ± 2.98	0.00
Total PUFA intake (% TE)	6.93 ± 2.85	6.24 ± 3.02	7.60 ± 2.49	0.00	6.43 ± 2.64	7.22 ± 2.93	0.00
Total SFA intake (% TE)	5.50 ± 2.71	4.26 ± 2.60	6.72 ± 2.22	0.00	5.14 ± 2.44	5.72 ± 2.84	0.00
Total TFA intake (g)	0.41 ± 0.61	0.14 ± 0.20	0.68 ± 0.74	0.00	0.37 ± 0.51	0.44 ± 0.66	0.02
Total cholesterol intake (mg)	187.00 ± 154.00	121.00 ± 124.00	252.00 ± 154.00	0.00	202.00 ± 173.00	179.00 ± 143.00	0.00
Total fibre intake (g)	20.90 ± 10.40	17.50 ± 7.53	24.30 ± 11.70	0.00	22.95 ± 11.70	19.80 ± 9.40	0.00
GGT ($U.L^{-1}$)	97.50 ± 198.00	85.20 ± 146.00	110.00 ± 239.00	0.01	127.00 ± 242.00	79.60 ± 163.00	0.00

% = percentage; BMI = body mass index; DBP = diastolic blood pressure; g = grams; $g.L^{-1}$ = grams per litre; GGT = gamma-glutamyl transferase; CRP = High sensitivity C-reactive protein; Hcy = homocysteine; HDL-C = high density lipoprotein cholesterol; $IU.L^{-1}$ = international units per litre; $kg.m^{-2}$ = kilograms per metres squared; kJ = kilojoules; LDL-C = low density lipoprotein cholesterol; $mmol.L^{-1}$ = millimole per litre; mmHG = millimetre of mercury; MUFA = monounsaturated fatty acids; $PAI\text{-}1_{act}$ = plasminogen activator inhibitor type 1 activity; PUFA = polyunsaturated fatty acids; SBP = systolic blood pressure; SD = standard deviation; SFA = saturated fatty acids; TC = total cholesterol; TE = total energy; TFA = trans fatty acid; TG = triglycerides; $U.L^{-1}$ = units per litre; a p-value of <0.05 is considered significant.

The reason for the preponderance of women can possibly be that more men work during the day and were therefore not at home when the study participants were enrolled. Another reason can be that there are more women than men in South Africa, according to the 2001 and 2011 census data and the sample in this study is, therefore, demographically representative of the South African population.

The means for TC and TG for the entire group in this cohort (as displayed in Table 5.1), were within desirable ranges, as they were below the cut-off values set at 5.17 $mmol.L^{-1}$ and < 1.69 $mmol.L^{-1}$, respectively (Klug, 2012; Butler, 2010). LDL-C and HDL-C concentrations for the whole group also fell within the desirable lipid profile ranges, as they were below the cut-off values considered to indicate increased CVD risk, *i.e*. ≤ 3.00 $mmol.L^{-1}$ and above the cut-off value for HDL-C, *i.e*. ≥ 1.20 $mmol.L^{-1}$ (Butler, 2010). The mean lipid profile of this population is within desirable, healthy ranges; more than half of the population do not harbour an undesirable blood lipid profile. Apart from the TG concentrations that were significantly higher in the urban group ($p = 0.00$) compared to the rural group, no other components of the lipid profile differed between these two groups. The higher TG in the urban group could be due to its higher socio-economic status, which could lead to more adverse lifestyle behaviour, *i.e*. high fat diets (Table 5.1) and increased alcohol intake (Schutte *et al*., 2006; Vorster *et al*., 2005).

The lipid profile determinants were significantly different between the genders, with women presenting with significantly higher TC, TG and LDL-C ($p = 0.00$), as well as significantly lower HDL-C concentrations ($p = 0.00$) compared to men. Even though the lipid profile of the women seems to be unfavourable when compared to the lipid profile of the men, it is still within the desirable ranges. However, this observed lipid profile among women is in agreement with previous reports (de Backer, 2008; Chu *et al*., 2003; Böthig, 1989). The physiological reason for the elevated TC, LDL-C and TG concentrations and lowered HDL-C concentrations among women could be hypothesised to be a result of declining ovarian function, prior to the loss of oestrogen, which is reported reported to have a cardio-protective effect. Yet, even though the ovarian function could be decreasing, it is not completely gone and women would therefore still experience some form of cardio-protection. Another reason for the observed difference between the lipid profile of men and women could be the difference in the body composition between men and women. In the present population the women's BMI was significantly higher than that of the men ($p = 0.00$) and epidemiological studies have reported a direct correlation between increasing BMI and elevated TC, LDL-C and TG and an inverse relationship with HDL-C,

which is also observed in the present population (Shamai *et al*., 2011 ; Krauss *et* al., 1998).

The means for the other CVD risk biomarkers for the total population, such as CRP, PAI-1_{act} and fibrinogen concentrations, are indicated in Table 5.1. Elevated PAI-1 concentrations have been reported to have an effect on atherothrombotic vascular disease and are therefore regarded as a CVD risk factor (Solano *et al*., 2003). Elevated PAI-1 concentrations increase cardiovascular disease risk development through various mechanisms, *i.e*. stimulating clot stability and preventing fibrinolysis in the vessel wall which results in unstable plaque development and mediating inflammatory responses (Belalcazar *et al*., 2011). The mean PAI-1_{act} concentration in the current population was within the reference ranges developed for healthy Caucasian individuals (5.5-10 $IU.L^{-1}$). A Previous studies reported similar PAI-1_{act} concentrations for black South Africans (Greyling *et al*., 2007) and African Americans (Festa *et al*., 2003). However, mean PAI-1_{act} concentrations for rural subjects and men were below the healthy reference ranges, whereas, the mean PAI-1_{act} concentrations for the total population, urban group and the women were within the healthy reference ranges. In addition to this, mean PAI-1_{act} concentrations were significantly lower ($p = 0.00$) when the rural subjects were compared to the urban subjects and when men and women were compared. This observation on the level of urbanisation can possibly be due to the increased risk associated with urbanisation stemming from changes in lifestyle behaviour, which not only result in increasing BMI, but in adverse dietary changes as well, which affect the lipid profile negatively. These consequential factors (*i.e*. unfavourable BMIs and lipid profiles) of urbanisation have been reported to play a major role in determining plasma PAI-1_{act} concentrations and can be a possible explanation for the increased PAI-1_{act} concentrations observed among the urban group of the present cohort (Festa *et al*., 1999; Asplund-Carlson *et al*., 1993; Mehta *et al*., 1987). These factors taken together seem to result in a modest increase in PAI-1_{act} concentrations, but not to the same degree as reported for Caucasians (Greyling *et al*., 2007; Festa *et al*., 2003). The decreased PAI-1_{act} concentrations observed in the men are not in accordance with previous reports for Caucasian populations (Henry *et al*., 1998; Meilahn *et al*., 1996). However, it can be speculated that the lower plasma concentrations of this haemostatic factor can be ascribed to the fact that the men in the present population have a lower BMI than the women, since PAI-1 has been reported to be positively related to upper body obesity (Meilahn *et al*., 1996). Therefore, gender and urbanisation seem to play a major role in determining the PAI-1_{act} concentrations in this black South African cohort and further studies are needed to understand the mechanisms

of regulation of the PAI-1_{act} concentrations better, in particular those related to the influence of gender.

The most extensively studied inflammatory risk marker is CRP, because of numerous studies that have reported CRP to be not only a powerful predictor of metabolic syndrome, but of CVD risk at all levels of the Framingham risk score (Bisoendial *et al*., 2010; Khera *et al*., 2005; Ridker *et al*., 2003; Ridker *et al*., 1997). Therefore, physicians started to measure CRP as a risk biomarker of heart disease in an attempt to prevent CVD events (Ridker, 2003). However, earlier laboratory assay methods, measuring CRP blood concentrations at <10 $mg.L^{-1}$ as the normal ranges, were not sensitive enough and the need for high sensitivity assays to measure high sensitivity CRP (hs-CRP) concentrations were developed (Helal *et al*., 2012). When CRP concentrations are measured with these high sensitivy assays, moderate changes (elevations) can be detected and, therefore, more detailed reference ranges have been developed with CRP concentrations where <1 $mg.L^{-1}$, 1 – 3 $mg.L^{-1}$ and >3 $mg.L^{-1}$, indicates low, moderate and high risk of future CVD events, respectively (Ridker, 2003). The mean CRP concentrations for all four groups were much higher than the reference range for CRP concentrations in healthy Caucasians (<1 $mg.L^{-1}$, 1 – 3 $mg.L^{-1}$ and >3 $mg.L^{-1}$), but this is similar to previous reports indicating that the highest CRP concentrations were reported among African ethnic groups (Khera *et al*., 2005; Albert *et al*., 2004; Ridker *et al*., 1997). However, there was no significant difference between the groups investigated in this study, as presented in Table 5.1. Nonetheless, despite the non-significant differences between groups, the CRP concentrations of the sub-groups that were overweight (urban and female, respectively) were higher than those of the groups that were within normal BMI ranges, which might signify the inflammatory pathways of weight gain (obesity), which contributes to CVD risk development (Ridker *et al*., 2000). Therefore, these results indicate a high risk of CVD events occurring among the present black South African population (Ridker *et al*., 1997).

Elevated homocysteine is considered to be a fundamental CVD risk factor owing to its role in endothelial dysfunction *via* various mechanisms such as cell and tissue damage in arteries, which leads to the development of atherosclerotic plaques and foam cells, as well as the oxidative stress caused by excess LDL-C (McCully, 2007). The plasma Hcy concentrations of the rural group as well as for the male group were significantly higher ($p = 0.00$) when compared to the urban group and to the women. The plasma Hcy concentrations of women were within the reference ranges of 3-14 $\mu mol.L^{-1}$ and below 16 $umol.L^{-1}$, which is considered to be indicative of moderate hyperhomocysteinaemia

(Bhat *et al*., 2008; Ueland *et al*., 1992). This is consistent with the literature in that being male is a biological factor associated with higher Hcy concentrations (Frick *et al.*, 2004). These observed gender differences can possibly be further explained in the light of the elevated GGT concentrations which were observed to be not only significantly higher among men compared to women, but the GGT concentrations in general, which were also higher than the reference range of 8-78 $U.L^{-1}$ in all the groups. Even though the mean alcohol intake was determined by the QFFQ, expressed as reported intake of pure alcohol (ethanol) in grams (g) per day based on the most recent available South African Food Composition Tables for alcohol (Langenhoven *et al*., 1991), this method was not the only one used to evaluate alcohol consumption. Rather, GGT concentrations were also measured as they are a biochemical indicator of alcohol intake and a more objective indication as opposed to self-reported intake according to the QFFQ. Furthermore, it is an indicator of liver damage, which in turn is an important cause of elevated Hcy (Carmel & James, 2002). The significant difference between rural dwellers having higher Hcy than urban dwellers, is in accordance with published literature (Nienaber-Rousseau *et al*., 2013), which ascribes this difference to the lower vitamin B_{12} intake among rural dwellers. Nonetheless, these values are still lower than the cut-off values, which is in agreement with the low Hcy concentrations previously reported among black Africans (Ubbink *et al*., 1999). This suggests that the present cohort is at low risk of the development of the early stages of atherogenesis (endothelial dysfunction).

Hyperfibrinogenaemia is considered an important risk factor for CVD (Grassi *et al*., 2005; Kessler *et al*., 1997), as fibrinogen has an influence on blood viscosity and platelet aggregation (Kessler *et al*., 1997). The fibrinogen concentrations for the total population were slightly above the reference ranges (1.5-3.5 $g.L^{-1}$), with significantly higher concentrations ($p = 0.00$) among rural dwellers and women compared to their respective counterparts. This finding seems to be similar to that reported by other researchers (James *et al*., 2000). In an article that investigated the same data reported in this dissertation, Pieters *et al*. (2011b) reported higher fibrinogen concentrations for the rural group when compared to the urban group and reported that the differences in dietary intakes between these two groups and low socioeconomic factors (*i.e*. employment and level of education) were possible reasons for this discrepancy. However, lower fibrinogen concentrations have been associated with a prudent diet including low intakes of animal-based protein, TFAs and higher levels of plant-based protein, dietary fibre and PUFA:SFA, which is generally representative of dietary intake in rural areas (James *et al*., 2000). The reason for this seeming inconsistency may be ascribed to other factors, such as increased

psychological stress among rural groups (Pieters *et al*., 2011b; Kaptoge *et al*., 2007; Frimerman *et al*., 1997). Psychosocial stressors (*i.e*. low socio-economic status) have been reported to result in increased concentrations of procoagulant molecules (*i.e*. fibrinogen or coagulation factor VII) and reduced fibrinolytic capacity, which causes a hypercoagulable state (Von Känel et al., 2001). In general, fibrinogen concentrations are higher in black South Africans than in Caucasian individuals and the mean fibrinogen concentration of 3.70 ± 2.18 g.L-1 for the whole group is greater than 2.5 $g.L^{-1}$, which is the cut-off value usually associated with low CVD risk (Pieters & Vorster, 2008). The higher fibrinogen concentrations observed in the black South African population compared to those reported for Caucasian populations support the notion that 30–50% of the variance of plasma fibrinogen concentrations are genetically determined (de Maat, 2001).

A positive relationship exists between blood pressure and CVD (de Backer, 2008) and hypertension is one of the primary risk factors for atherosclerosis and CVD, since elevated blood pressure forces the myocardium to work harder to pump blood into the arteries. Furthermore, a single increase of 20/10 mm Hg above the desirable blood pressure values doubles the risk of death from CVD (Rosendorff, 2007). The mean SBP and DBP values for all of the groups were within the pre-hypertensive range *i.e*. 120-139 mmHg/80-89 mmHg (Chobanian *et al*., 2003). The mean SBP and DBP values were significantly higher among the urban group ($p = 0.00$) when compared to the rural group, which emphasises the presence and consequences of unhealthy lifestyle behaviour (Tomson & Lip, 2005). All of the aforementioned reference ranges for CVD risk factors were based on Caucasian individuals, as the respective ranges for black Africans are yet to be identified and interpreting these data in terms of CVD risk in the African population should be done with caution.

Data on the dietary habits of TE intake [in absolute kilojoules (kJ)], total fibre and cholesterol intake (in absolute grams), total fat, SFA, MUFA and PUFA in percentage of TE, were also analysed, as these dietary aspects have been reported to play a role in CVD development (James *et al*., 2000). The mean energy intake for women as well as for men was higher than the recommended daily kJ intake for women (6 720 kJ) and men (8 400 kJ), respectively (Vorster & Nel, 2001; Wolmarans & Oosthuizen, 2001). The significantly higher kJ intake for urban dwellers compared to rural dwellers ($p = 0.00$) was expected and can be attributed to the effect of urbanisation with the accompanying nutrition transition phenomenon, which includes following an energy-dense diet (Misra *et al*., 2010; Popkin, 2006; Alberts *et al*., 2005; Vorster, 2002; Yusuf *et al*., 2001a).

However, another important aspect to take note of in the nutrition transition is that the energy density of food choices increases as the available income to buy food decreases, which in turn can result in higher energy intakes among lower income groups (Temple & Steyn 2009). Furthermore, the significantly lower kJ intake for women when compared to men ($p = 0.00$) was also expected, since the recommended daily energy intakes are also reported to be lower among women.

When comparing the energy derived from fat intake of this population against the recommended AMDRs for fat (20-35%), the urban group as well as both of the gender groups were within the AMDR ranges. The energy distribution for dietary fat intake among the rural group was lower than the AMDR, which can be explained by the consumption of a more traditional diet, known to be lower in fat (Alberts *et al*., 2005; Vorster, 2002; Yusuf *et al*., 2001a) and simply as a result of a lower TE intake per day compared to the urban group (Table 5.1). Furthermore, the energy distribution for total fat was significantly higher for the urban group than the rural group (Table 5.1), again possibly because of the aforementioned reasons.

However, when investigating the distribution of the different dietary fats, the differential vulnerability towards certain disease risks between populations, gender and geographical groups may become apparent. The energy distribution of the different dietary fats was in line with the recommendations on the quantity (total amount) and quality (type) of fat in the diet for optimal health across the life span (Lichtenstein *et al*., 2006). The SFA should provide no more than 10% of the daily energy intake and in the present cohort it provided approximately 6%. PUFA contributed almost 7% of the daily energy intake of the total population and this too is in line with the recommendation of 6-10 % (Lichtenstein *et al*., 2006). Lichtenstein and co-workers (2006) suggested that the remainder of the energy from fat should be provided by MUFAs, whereas TFA should be less than 1% of the daily energy intake (Lichtenstein *et al*., 2006), as it was in the present cohort (Table 5.1). Furthermore, the total cholesterol intake (188 ± 156 mg/day) is also well below the recommended < 300 mg/day (Lichtenstein *et al*., 2006). Therefore, the different dietary fat distributions for this population are within desirable, healthy ranges despite gender or geographical location differences.

The means of all of the different dietary fat types (SFA, TFA, MUFA, and PUFA) were significantly higher for the urban group than the rural group, as well as for women compared to men ($p = 0.00$), which is in accordance with previously reported national data

(Steyn & Nel, 2006). This observation can possibly be explained by the higher fat content and energy-density of the diet that is expected according to the nutrition transition in the urban group. The higher dietary fat intake among women might be explained in the light of their dietary fibre intake. Dietary fibre may promote satiety by slowing gastric emptying, leading to an overall decrease in energy intake (Schneeman, 2002), resulting in lower fat intake. However, the mean total dietary fibre intake of the women was below the dietary reference intake recommended for adult women of the particular age group, which should be 21 g. This might explain why their fat intake was higher. Furthermore, the significantly lower total dietary fibre intake among the rural group corresponds with the published literature that states that dietary fibre intake increases with urbanisation (Vorster *et al*., 2005); however, it does not correspond with the observed higher fat intake among the present urban group. Therefore, the suggestion that other contributing risk factors (*i.e.* lifestyle or genetic) should be taken into consideration as a possible explanation of the results is correct.

Using the cutoff limits as discussed in Supplementary A: Section 1.4, it was determined that 18.2% of the population were considered to be underweight, 42.9% were of normal weight, 17.7% were overweight and 21.2% were obese. According to Table 5.2, the results observed regarding the association of the different lipid profile variables of this population with BMI, were similar to previous reports (Kushner & Blatner, 2005). Firstly, there was a significant difference between the lipid profile determinants for each BMI range/category for the total population as well as by gender and locality grouping ($p = 0.00$). The TC concentrations for the total population, women and rural dwellers ($p = 0.00$) were highest among individuals who were classified as being overweight (according to their BMI, which ranged between 25 and 29.9 $kg.m^{-2}$). The TC concentrations were highest among obese (BMI >30 $kg.m^{-2}$) individuals who resided in an urban area and plateaued among overweight and obese men; $p = 0.00$ (Table 5.2).

LDL-C concentrations increased significantly with increasing BMI categories for the total population, urban dwellers and women ($p = 0.00$). Among men and individuals from the rural setting, the LDL-C concentrations peaked in the overweight BMI category ($p = 0.00$), but plateaued among the obese subjects. The HDL-C concentrations decreased significantly with the increase in BMI for the total population and for the separate groups *i.e*. locality and gender (Table 5.2). Lastly, the TG concentrations increased significantly with each BMI category, with men and urban dwellers having significantly higher TG concentrations ($p = 0.00$). These observations are all concurrent with the literature that

states that increasing body weight is linked to higher TG and LDL-C and lower HDL-C concentrations (Shamai *et al*., 2011).

Table 5.2 Categorical baseline characteristics per gender and locality, in the PURE study population

	BMI (kg.m^{-2})				History of tobacco use			HIV status	
	<18.5	18.5-24.9	25-29.9	≥30	Never	Former	Current	Pos	Neg
TC (mmol.L^{-1})									
Whole	4.61	4.94	5.31	5.27	5.15	4.84	4.90	4.39	5.13
Rural	4.46	4.94	5.35	5.20	5.17	4.75	4.80	4.44	5.07
Urban	4.81	4.95	5.27	5.32	5.13	4.93	5.00	4.34	5.19
Male	4.49	4.85	5.27	5.27	5.00	4.72	4.72	4.24	4.92
Female	4.80	5.03	5.33	5.27	5.21	5.06	5.04	4.47	5.41
LDL-C (mmol.L^{-1})									
Whole	2.42	2.83	3.28	3.30	3.11	2.86	2.77	2.60	2.99
Rural	2.37	2.87	3.29	3.27	3.14	2.90	2.73	2.63	2.98
Urban	2.49	2.76	3.27	3.33	3.08	2.82	2.82	2.55	3.00
Male	2.30	2.74	3.40	3.29	3.00	2.79	2.54	2.46	2.75
Female	2.60	2.91	3.25	3.30	3.16	2.98	2.95	2.67	3.42
HDL-C (mmol.L^{-1})									
Whole	1.73	1.58	1.40	1.30	1.39	1.45	1.59	1.23	1.57
Rural	1.66	1.56	1.46	1.29	1.47	1.29	1.57	1.28	1.56
Urban	1.82	1.61	1.33	1.31	1.42	1.50	1.60	1.18	1.59
Male	1.75	1.58	1.22	1.18	1.44	1.38	1.68	1.23	1.64
Female	1.69	1.58	1.45	1.31	1.45	1.41	1.51	1.23	1.49
TG (mmol.L^{-1})									
Whole	1.02	1.26	1.45	1.49	1.36†	1.31†	1.25†	1.29£	1.30£
Rural	0.97	1.18	1.35	1.42	1.28	1.25	1.16	1.21	1.22
Urban	1.08	1.34	1.54	1.55	1.44	1.36	1.33	1.37	1.39
Male	0.94	1.26	1.53	1.77	1.30#	1.21#	1.17#	1.28@	1.21@
Female	1.13	1.25	1.42	1.47	1.38#	1.47#	1.30#	1.29@	1.12@

All of the variables which differed significantly for the whole group according to the respective categorical variable and by locality and gender has a p-value of = 0.00, except where otherwise indicated: £ = 0.77; @ = 0.02; † = 0.03; # = 0.01; BMI = body mass index; HDL-C = high-density lipoprotein cholesterol; HIV = human immunodeficiency virus; LDL-C = low-density lipoprotein cholesterol; neg = negative; pos = positive; TC = total cholesterol; TG = triglycerides

Unexpected observations were made regarding the history of tobacco use in relation to the lipid profile in the present cohort. Individuals who had never smoked had significantly higher TC, LDL-C and TG concentrations (p = 0.00) compared to individuals who had smoked before or were current smokers. The relationship for TC can be visually described as almost U-shaped, *i.e*. those that reported never having smoked before had the highest TC, but this decreased in individuals who had smoked before and then increased again among current smokers. However, a negative linear relationship was observed for LDL-C and TG in the different tobacco use categories, *i.e.* the highest LDL-C was among individuals who had never smoked before and the lowest LDL-C concentrations were among current smokers (Table 5.2). Furthermore, the HDL-C concentrations were significantly higher among current smokers (p = 0.00) compared to the lower HDL-C

concentrations observed in individuals who had never smoked or had formerly smoked, which is in accordance with the limited number of previous studies conducted among black South Africans (Zatu *et al*., 2011; Alberts *et al*., 2005; Vorster *et al*., 2002). All the aforementioned results are, however, different from those reported in literature for Caucasian individuals (Campbell *et al*., 2008; Ambrose & Barua., 2004; Craig *et al*., 1989). According to the current study the reason for this observation is unclear, but a possible reason might be the role genetic predisposition plays in the development of atherogenesis in individuals exposed to cigarette smoke (Ambrose & Barua., 2004). Previous studies have reported that either the cytochrome P-450 1A1 (*CYP1A1*) MSP polymorphism or certain endothelial nitric oxide synthase intron 4 polymorphisms increased the susceptibility to atherosclerotic diseases related exposure to cigarette smoke (Clouse *et al*., 2000; Mayhan & Sharpe, 1999). However, the prevalence of these particular genetic variants has not been determined on a global level in cigarette smokers and, their biological relevance and importance are therefore unknown (Ambrose & Barua, 2004). Since the health effects of cigarette smoking were more adverse among the non-smokers than the smokers in the present population, it could be hypothesised that other genetic variants might be partially responsible for these observations. However, further research focusing on tobacco use and its effect on the lipid profile in a black South African population is necessary, since studies on this topic are limited in South African populations.

What is also clear from Table 5.2 is that individuals not infected with HIV have significantly higher TC, LDL-C and HDL-C concentrations, which is applicable to all the stratification groups. No difference was observed for the TG concentrations between HIV positive and negative individuals for the total study population. However, the TG concentrations were significantly higher among HIV positive individuals compared to HIV negative individuals, grouped by gender ($p = 0.02$). Furthermore, the HIV negative participants presented with significantly higher TG concentrations compared to the HIV positive participants, when grouped by their area of residence ($p = 0.00$). Even though 16.2% of the population were HIV positive, the subjects recruited to take part in this study were apparently healthy, and those who were HIV positive were not aware of their HIV status prior to the instigation of the study and therefore did not use antiretroviral therapy.

This section described the influence of certain environmental and phenotypic factors in the present cohort. The following sections will discuss the determination of the genetic variations in the ApoE gene.

5.2 GENETIC DATA COLLECTION

The experimental procedures used in the present study included DNA isolation from buffy coat samples, PCR, agarose gel electrophoresis, purification of PCR products, automated sequencing and genetic screening *via* real-time PCR and BeadXpress® analyses. For all of these experimental procedures, commercially available kits were used, which required minimal optimisation. The optimisation of the BeadXpress® and real-time PCR procedures were performed with the use of various controls, which will be discussed in greater detail in Sections 5.3 and 5.4. The results obtained for each of these protocols and processes are discussed in the following sections.

5.2.1 DNA isolation

The samples used for DNA isolation were obtained with written informed consent from each study participant, as outlined in Section 4.2. A modified protocol of the FlexiGene DNA Isolation Kit (QIAGEN®) was used for the manual extraction from buffy coat samples as discussed in Section 4.3.1. After extracting DNA from certain samples using the FlexiGene kit, very low DNA yields were obtained and this might have been due to these individuals simply having low levels of white blood cells. The samples for which an insufficient yield was obtained through the manual DNA isolation method were re-isolated using a different technique *i.e*. the Maxwell® 16 System, as it is a more efficient automated purification system (as discussed in Section 4.3.1). It was important to normalise the concentration of all 2010 DNA samples due to the fact that the genotyping was conducted *via* an automated system where it was crucial to standardise the volumes of the reagents added to the reaction as opposed to manual genotyping methods (*i.e*. PCR) where each samples' volume can be amended individually if need be and, therefore, the researchers of this study decided upon a cut-off value between 45 and 55 $ng.\mu L^{-1}$ to ensure consistent amplification. The yield and purity of the isolated DNA from both techniques were determined, as discussed in Section 4.3.1. The total yield of the manually isolated DNA ranged from 15.2-142.4 μg, while the yield of the DNA obtained from the samples isolated *via* the Maxwell® 16 System ranged between ca. 25.5-128.7 μg. Good quality DNA (without any contamination of ethanol and other proteins) has an A_{260}/A_{280} ratio of 1.7–2.0. The A_{260}/A_{280} ratio of the manually isolated DNA and DNA, obtained through the Maxwell® 16 System as measured by the NanoDrop® ND-1000 spectrophotometer, ranged from 1.74-1.88 and 1.70-1.87, respectively and was therefore of good quality.

5.2.2 Polymerase chain reaction amplification and automated sequencing

The *ApoE* gene in this study was initially amplified in five fragments. After the sequences for these fragments had been aligned, a minor 31 bp gap was identified between fragments two and three. As a result, primer set six (Section 4.3.2) was designed and the amplification process for fragment six was conducted by the author of this study and Ms B Swanepoel, in order to obtain sequence data for this gap. However, even though the primers were successful, a "tg" homopolymeric region repetition in the sequences for fragment six occurred, which resulted in an unreadable sequence for this 31 bp gap (Figure 4.5).

All of these PCR fragments were less than 1 000 bp in size, as this is the maximum length that the Sanger sequence can yield (http://www.biotech.wisc.edu/facilities/dnaseq/sequencing/sangersequencing). In order to ensure that specific primers were designed to amplify these fragments successfully, certain conditions had to be adhered to, as discussed in Section 4.3.2.

The optimised annealing temperatures (T_a) that were used to amplify the five (successful) DNA fragments of the *ApoE* gene are provided and discussed in Section 4.3.2. The polymerisation enzyme Promega *Taq*™ DNA polymerase was used and 37.5 pmol $MgCl_2$ was sufficient for the successful amplification of fragments one to five. As part of the optimisation process of fragment six, the decision was made to increase the actual quantity of $MgCl_2$ (to 125 nmol), which was added to the master mix. The reason for this was to ensure that the activity of the *taq* polymerase was not reduced due to a lack of Mg^{2+} ions for the proper functioning of its active centre (Markoulatos et al., 2002). All the successfully sequenced fragments were amplified as summarised in Chapter 4, Section 4.3.2. The key insight gained with regard to the PCR optimisation procedure, was the need to increase the annealing/extension step to one minute in order to achieve improved amplification results. An easy rule to apply was to consider an extension time (in minutes) equal to the number of kilobase pairs (kb) of the product to be amplified *i.e*. one minute for a 1 kb product (http://medicine.yale.edu/labs/henegariu/www/tavi/p08.html). Another possible reason why the annealing/ extension step of one minute might have resulted in successful amplification can be ascribed to the rich GC content of this region prolonging the binding activity and effectiveness of the Taq polymerase in this region (http://medicine.yale.edu/labs/henegariu/www/tavi/p08.html).

DNA was replaced with nuclease-free ddH_2O (Promega) in a single sample as a negative control for each batch of PCR reactions. Analysis of the success of amplification was achieved by agarose gel electrophoresis. The absence of a visible fragment in the negative control on the electrophoresis gel, served as a positive result for lack of contamination of the samples. For the positive control (containing control DNA), a clearly visible fragment (of the expected length) confirmed that the experimental protocols remained constant and progressed normally. In all the regions, the 50 bp and 100 bp molecular weight markers from Fermentas GeneRuler™ were used interchangeably, based on availability, to estimate the fragment size and are indicated in the figures illustrating photographic representations of the successful amplification of the various fragments in the section below.

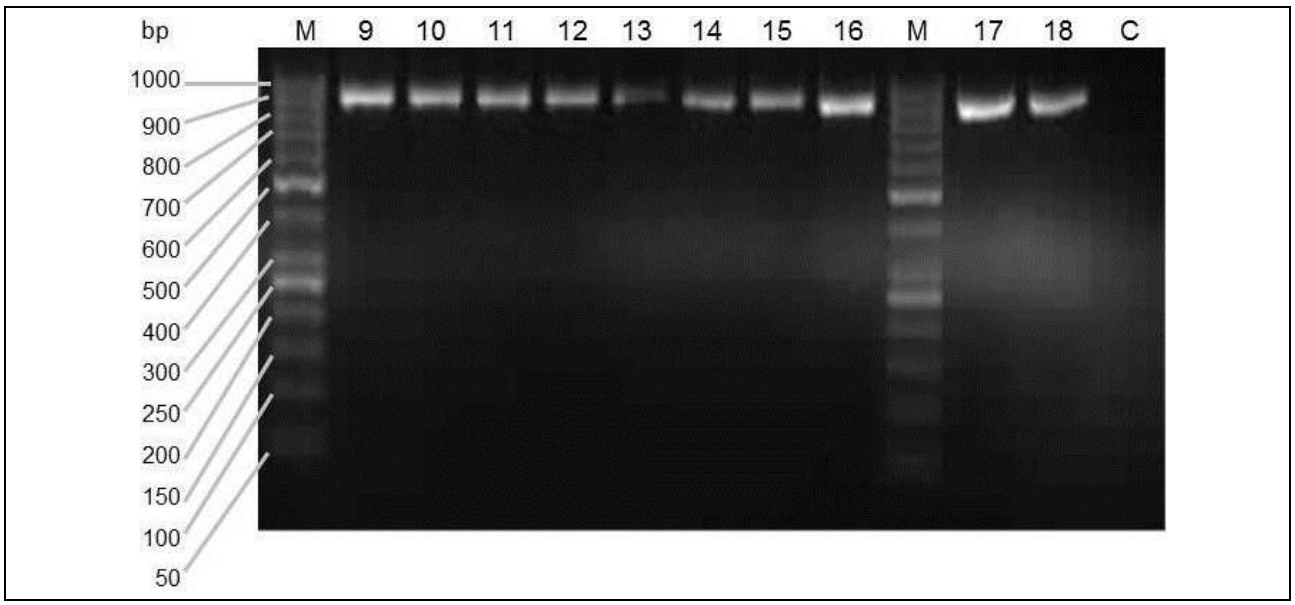
5.2.3 Identification of polymorphisms in the *ApoE* gene

The locations of all the identified SNPs in the *ApoE* gene are illustrated in Figure 5.1. The process used to identify and select the list of SNPs in the *ApoE* gene is described in Section 4.3.3. As mentioned, six fragments were amplified, but successful sequencing data (in both the forward and the reverse direction) were only obtained for five of the fragments (Figure 5.1) and will be discussed subsequently.

5.2.3.1 Amplification of fragment one in the *ApoE* gene

The protocols for the sequencing of the *ApoE* gene as stated in Section 4.3.2 (Table 4.1) were followed at the optimised T_a of 56°C to amplify fragment one of the *ApoE* gene. Fragment one of the *ApoE* gene, was determined to be 829 bp in size and consisted of a combination of regions, *i.e.* promoter, intron one and exon one. Figure 5.1 is a photographic representation of the successful amplification of a subset of the 30 random samples from the PURE study. Furthermore, Figure 5.2 also illustrates that the negative control had no amplification, which indicates that no contamination occurred (Figure 5.1, Lane C).

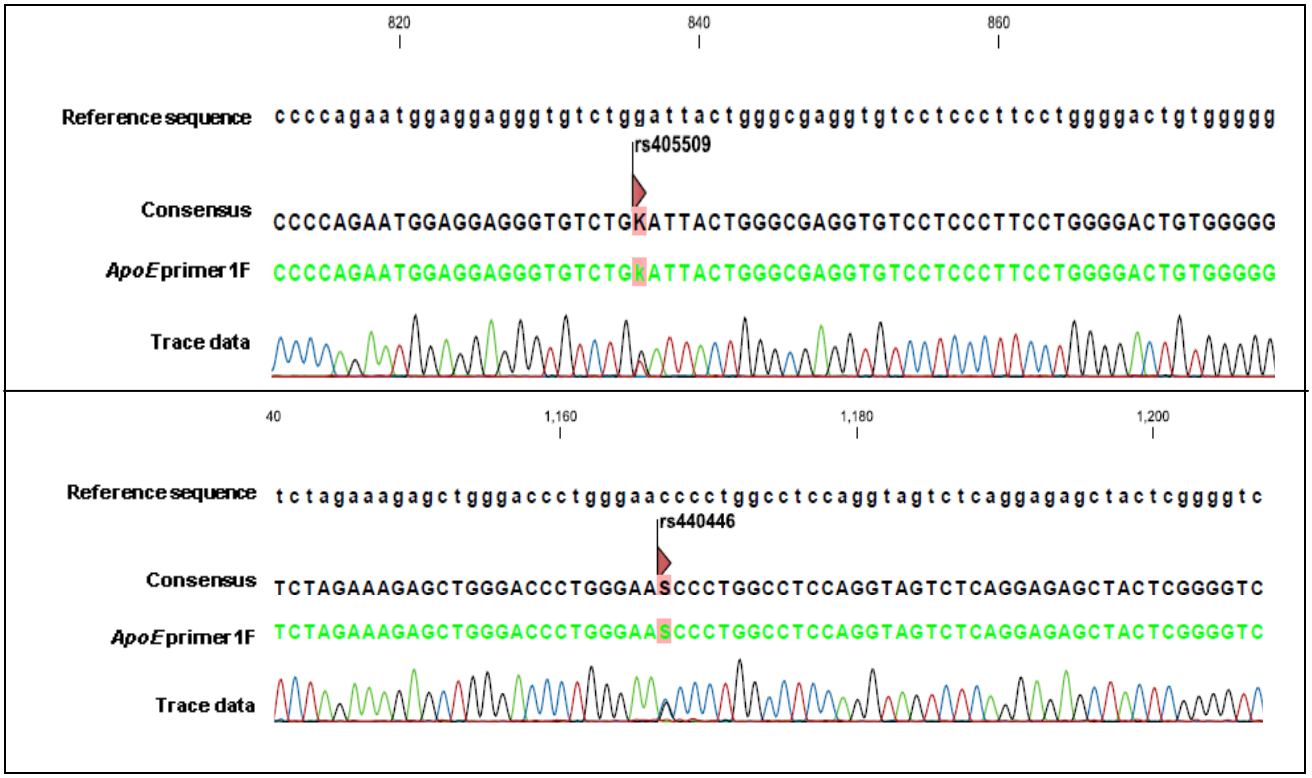
Figure 5.1 Photographic representation of the successful amplification of fragment one of the *ApoE* gene at 56°C



DNA of 30 random samples was electrophoresed on a 2% agarose gel at 100 volts for 60 minutes in 1% Tris/Borate/EDTA (TBE buffer); C: distilled water or dH_2O was used as the negative control; lane M: 50 bp Fermentas GeneRuler™ marker

These amplification products were sequenced as described in Section 4.3.2. The raw sequence electropherograms obtained for the 30 randomised samples were manually aligned to the human genomic reference *ApoE* gene (Genbank Accession number AF261279) using the CLC sequence viewer program (Section 4.3.3). With the reference gene as a template, four SNPs, (rs405509, rs440446, rs769446 and rs769447) were identified in fragment one of the *ApoE* gene (Figure 4.5). However, two of the SNPs (rs769446 and rs769447) were not included in the final list of SNPs to be analysed further, because of low cluster separation scores, as will be discussed in Section 5.3. The two SNPs that did form part of the final SNP list (rs405509 and rs440446), are presented in Figure 5.2, which is a graphic representation of the location where these SNPs were identified in some of the 30 random samples, as well as their respective base pair alterations.

Figure 5.2 Representative electropherogram for the SNPs rs405509 and rs440446 identified in fragment one of the *ApoE* gene

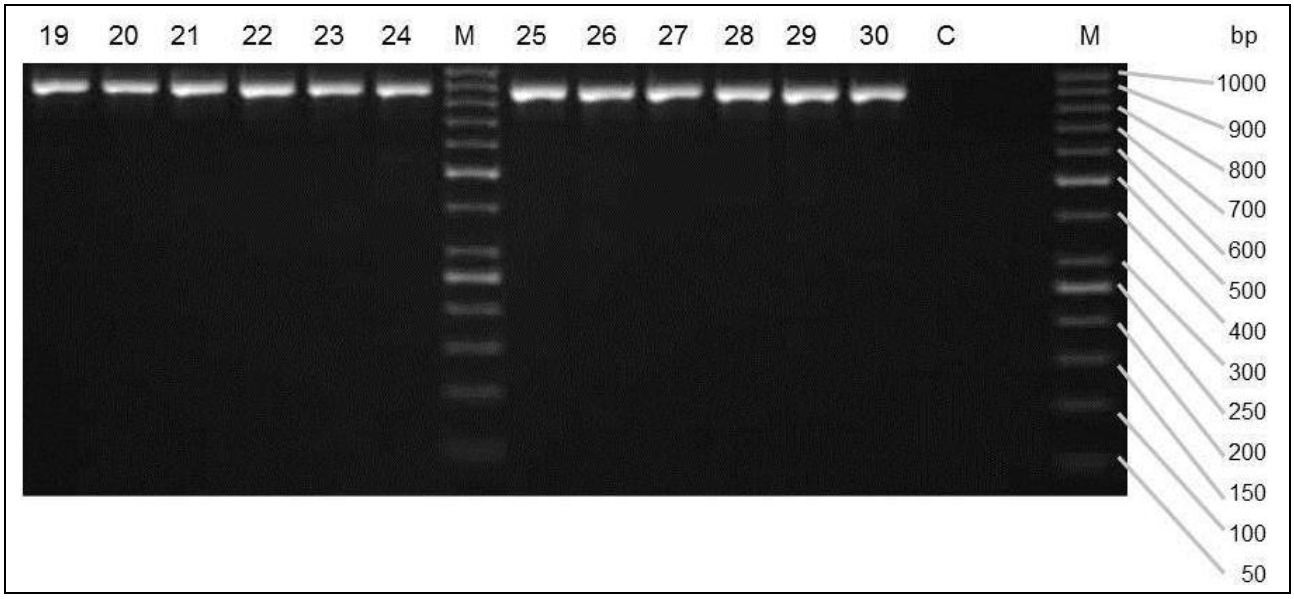


A = adenine; C = cytosine; G = guanine; K = G or T (NC-IUB, 1985); T = thymine. The reported alteration has been indicated with a red highlight.

5.2.3.2 Amplification of fragment two in the *ApoE* gene

The T_a for fragment two of the *ApoE* gene was determined to be 67.2ºC, as per temperature gradient analysis. This fragment was 843 bp in length and was located in an intronic region. Furthermore, an overlap of 295 bp between fragments one and two was observed. From Figure 5.4 it can be determined that fragment two was successfully amplified in a subset of the 30 random samples from the PURE study, after implementing the protocols discussed in Section 4.3.2: Table 4.1. Moreover, this figure indicates that no contamination occurred owing to the lack of amplification in the negative control (Figure 5.3).

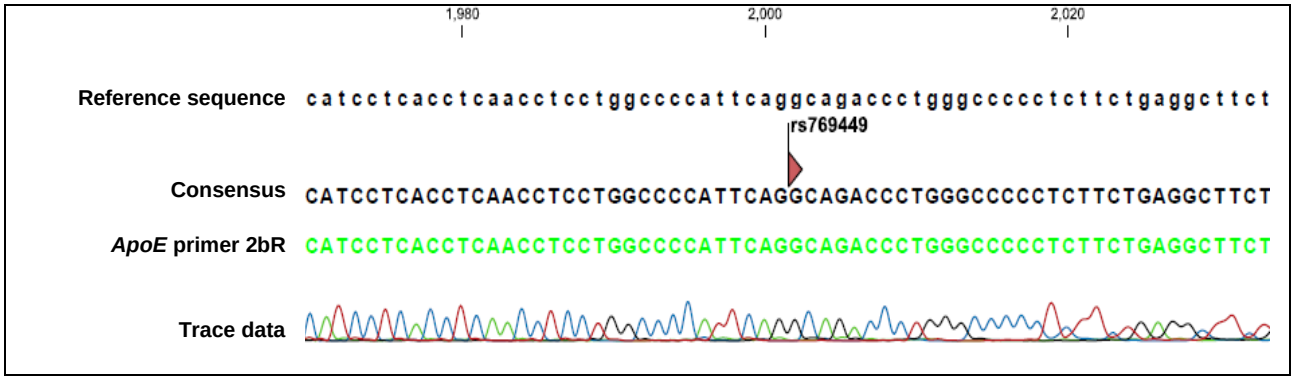
Figure 5.3 Photographic representation of the successful amplification of fragment two of the *ApoE* gene at 67.2°C



Amplified products of a subset of 30 random samples were separated by a 2% agarose gel electrophoresis at 100 volts for 60 minutes in 1% Tris/Borate/EDA (TBE buffer); C: pure water was used as the negative control; lane M: 50 bp Fermentas GeneRuler™ marker

These amplification products were sequenced and aligned to the reference gene as described in Section 4.3.3. Upon sequencing the 30 random samples in the present dataset, it was revealed that all the participants were homozygous for the wild type allele (GG) at the rs769449 locus (Figure 4.5). This SNP is at the end of the readable sequence of this fragment and this might explain the poor sequence quality observed for this particular section (Figure 5.4). However, this SNP, along with another SNP [(rs769448), which was identified through literature] were not included in the final list of SNPs to be analysed further because of low cluster separation scores as discussed in Section 5.3.

Figure 5.4 Representative electropherogram for the SNP rs769449 identified in fragment two within the ApoE gene



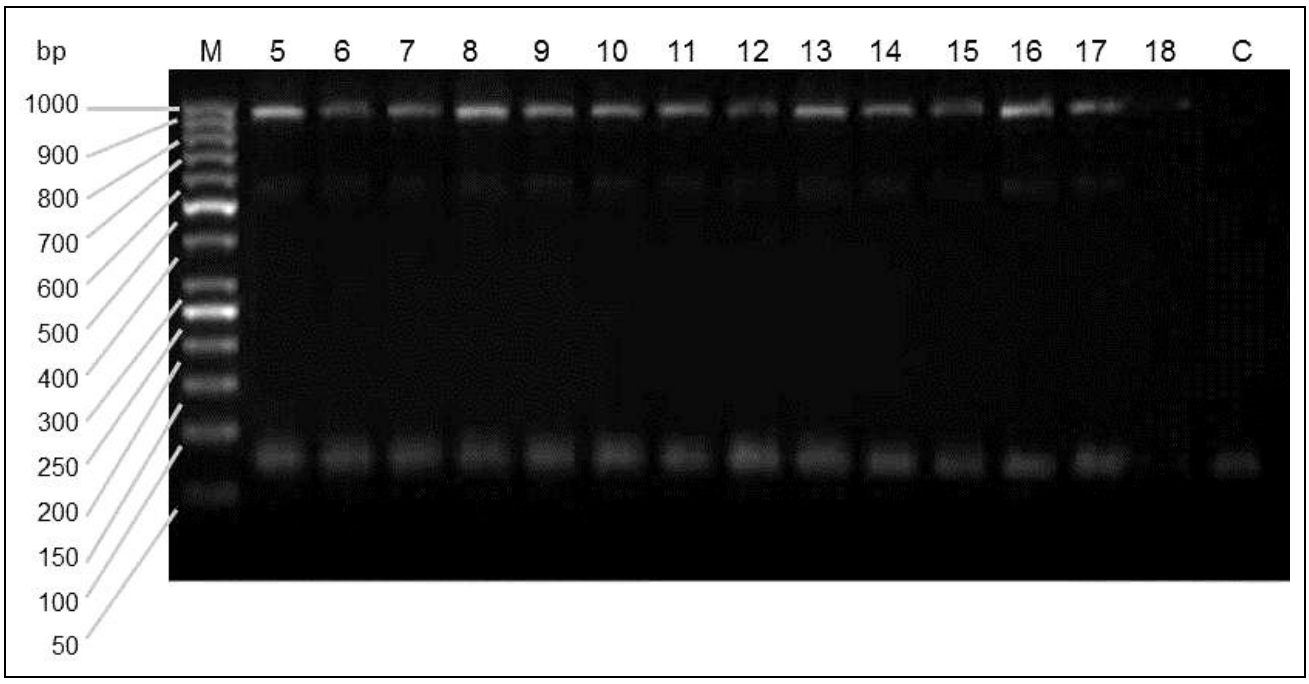
A = adenine; C = cytosine; G = guanine; T = thymine (NC-IUB, 1985); refseq = reference sequence. The reported rs-number has been indicated with a red banner at the specific loci of interest.

5.2.3.3 Amplification of fragment three of the *ApoE* gene

Fragment three in the *ApoE* gene is also located in an intronic region and stretches over 918 bp. As discussed in Section 5.2.2, this fragment did not overlap with the previous fragment, since a gap of 31 bp in length was observed (Figure 4.5). A sixth primer set was designed and the amplification process was optimised to provide the sequence data necessary for these fragments' sequences to overlap. However, only poor sequencing data were obtained for fragment six because of the homopolymeric repetition as discussed in Section 5.2.2, this small gap in the sequence of the *ApoE* gene remained.

The PCR protocols for fragment three, as discussed in Section 4.3.2, were followed in conjunction with a temperature gradient analysis to determine a T_a of 53ºC. Figure 5.5 illustrates the successful amplification of this fragment for a subset of the 30 random samples from the PURE study, using the aforementioned protocol. Moreover, no contamination occurred in this run, as no amplification occurred for the negative control (Figure 5.5). However, the presence of the fragments at the bottom of each lane of the agarose gel indicates the possible presence of primer dimers, which represent the attachment of the primers to one another instead of the amplicon itself. Secondary fragments also occurred, possibly as a result of mispriming. Although an extensive optimisation process was followed in an attempt to prevent the occurrence of these primer dimers and secondary fragments, this remained the best amplification product obtained (Figure 5.5). The actual fragments were therefore removed from the agarose gel and aliquoted into 0.2 µl tubes, which were submitted for gel cleanup and thereafter for sequencing (as discussed in Section 4.3.2).

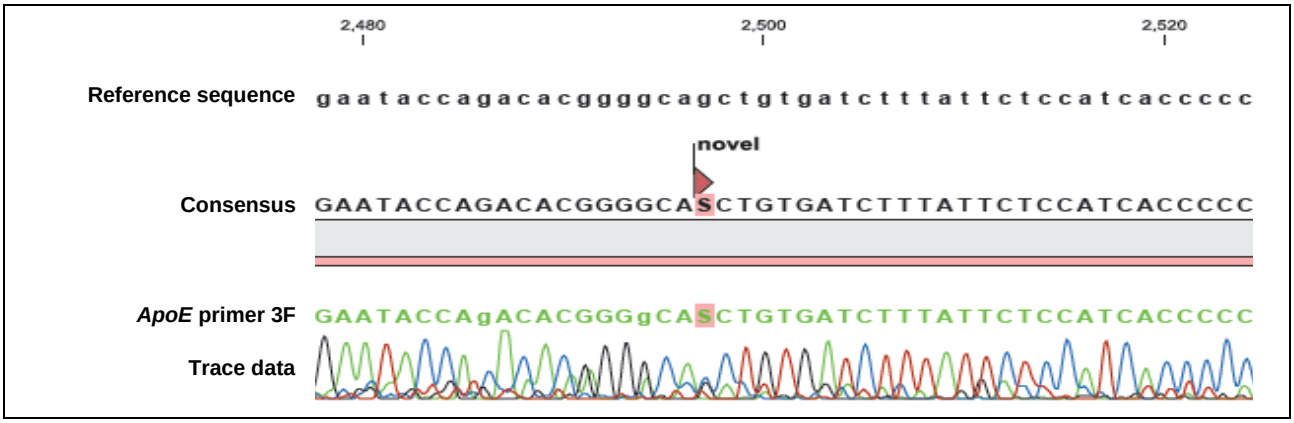
Figure 5.5 Photographic representation of the successful amplification of fragment three of the *ApoE* gene at 53°C



DNA of 30 random samples was separated by a 2% agarose gel electrophoresis at 100 Volts for 60 minutes in 1% Tris/Borate/EDA (TBE buffer); C: pure water as the negative control; lane M: 50 bp Fermentas GeneRuler™ marker

After aligning the sequence products obtained from amplified fragment three of the 30 random samples of the PURE study, three novel alterations were identified. Two of the three novel SNPs were excluded after applying the custom GoldenGate genotyping assay to validate the initial SNP list to be sent for genotypic analyses *via* the BeadXpress® Reader, since these SNPs were less than 60 bp apart (Section 5.3.1). Furthermore, one previously reported alteration, *i.e*. rs769450 (identified *via* the literature research conducted) occurred in this fragment (Figure 4.5). However, after investigating the validation scores for the novel SNP (Figure 5.6) as well as those obtained for the SNP rs769450 (Section 5.3.1: Table 5.3), these alterations within this fragment, were excluded from the final list of SNPs to be analysed further, primarily because of low cluster separation, as discussed in Section 5.3.1.

Figure 5.6 Representative electropherogram for the novel SNP identified in fragment three in the *ApoE* gene

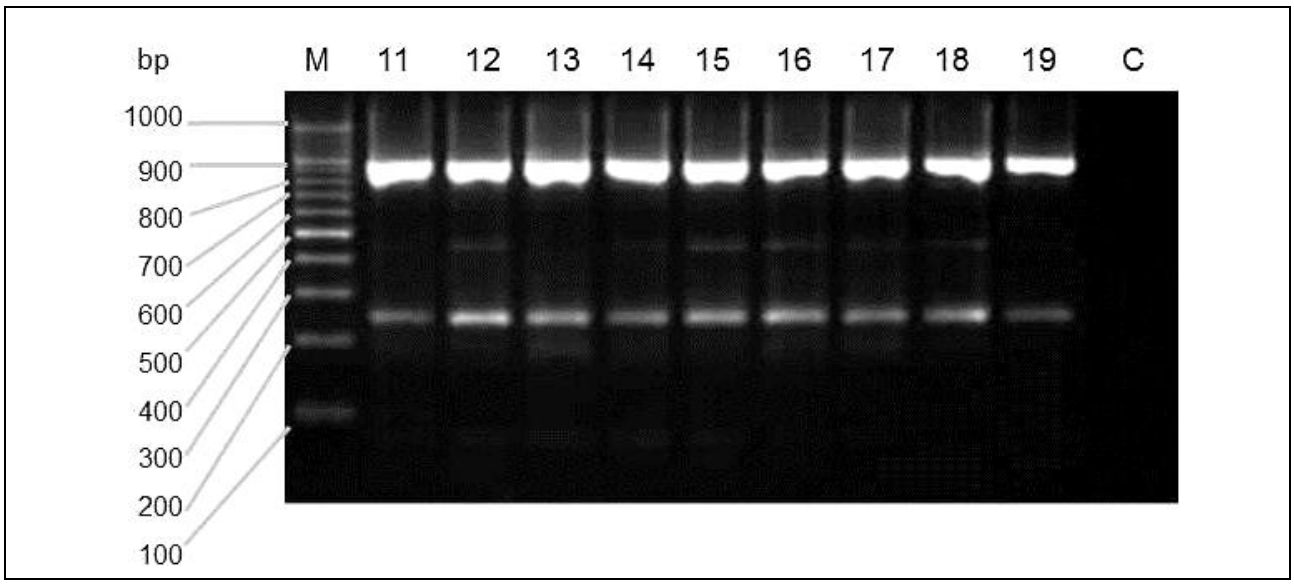


A = adenine; C = cytosine; G = guanine; T = thymine; S = C or G (NC-IUB, 1985); refseq = reference sequence. The novel SNP has been indicated with a red banner at the specific loci of interest.

5.2.3.4 Amplification of fragment four in the *ApoE* gene

Fragment four was amplified *via* the protocol provided in Table 4.1 in Section 4.3.2. This fragment contains exon three of the *ApoE* gene and is 963 bp long, which was confirmed by comparing the fragment length to the 100 bp marker (Fermentas GeneRuler™) that was used during the agarose electrophoresis, as indicated in Figure 5.7. A 138 bp overlap between fragments three and four occurred. From Figure 5.7 it is clear that no amplification occurred in the negative control; however, this figure also indicates that the amplification of fragment four, for the 30 random samples from the PURE study, had issues with secondary fragments which formed. Despite the presence of these secondary fragments, Figure 5.7 illustrates the best amplification product that was obtained after an extensive optimisation process of this fragment. Therefore, the correct fragment was excised from the agarose gel and aliquoted into 0.2 µl tubes. It was then submitted for gel cleanup and subsequent sequencing (Section 4.3.2).

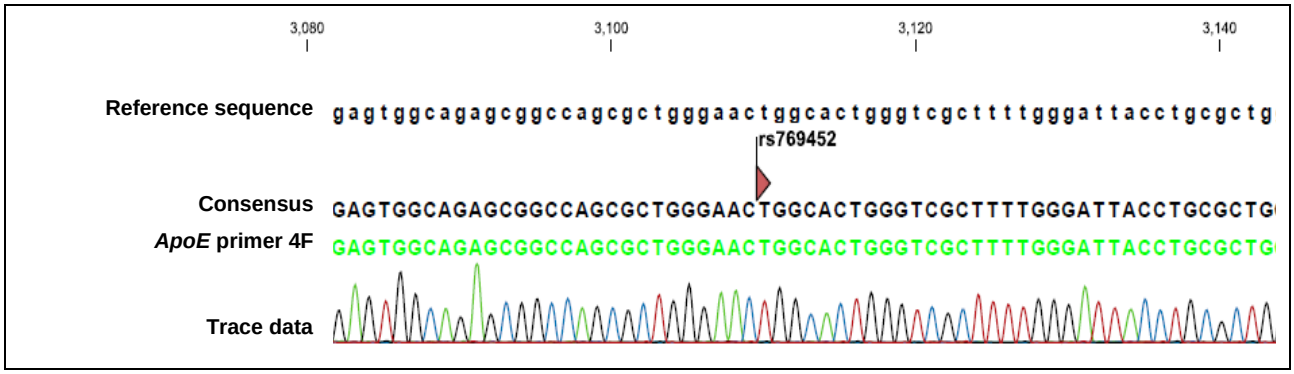
Figure 5.7 Photographic representation of the successful amplification of fragment four of the *ApoE* gene at 54.3°C



The amplified products of fragment four in 30 random samples from the PURE study were separated by 2% agarose gel electrophoresis at 100 volts for 60 minutes in 1% Tris/Borate/EDTA (TBE buffer); C: pure water used as the negative control; lane M: 50 bp Fermentas GeneRuler™ marker

The amplification products of 30 random samples from the PURE study were sent for sequencing, as mentioned in Section 4.3.2, and aligned to the reference sequence in order to identify SNPs in the particular fragment, as discussed in Section 4.3.3. Only one polymorphism was identified, in this fragment, from the 30 random samples from the PURE study (rs769452) and was included in the final SNP list for reasons specified in Section 5.3. Furthermore, one SNP that was identified by literature occurred in this fragment (rs11542029; Figure 4.5). The electropherogram as indicated in the CLC Workbench sequence viewer program for the rs769452 SNP is illustrated in Figure 5.8.

Figure 5.8 Representative electropherogram for the SNP rs769452 identified in fragment four in the *ApoE* gene

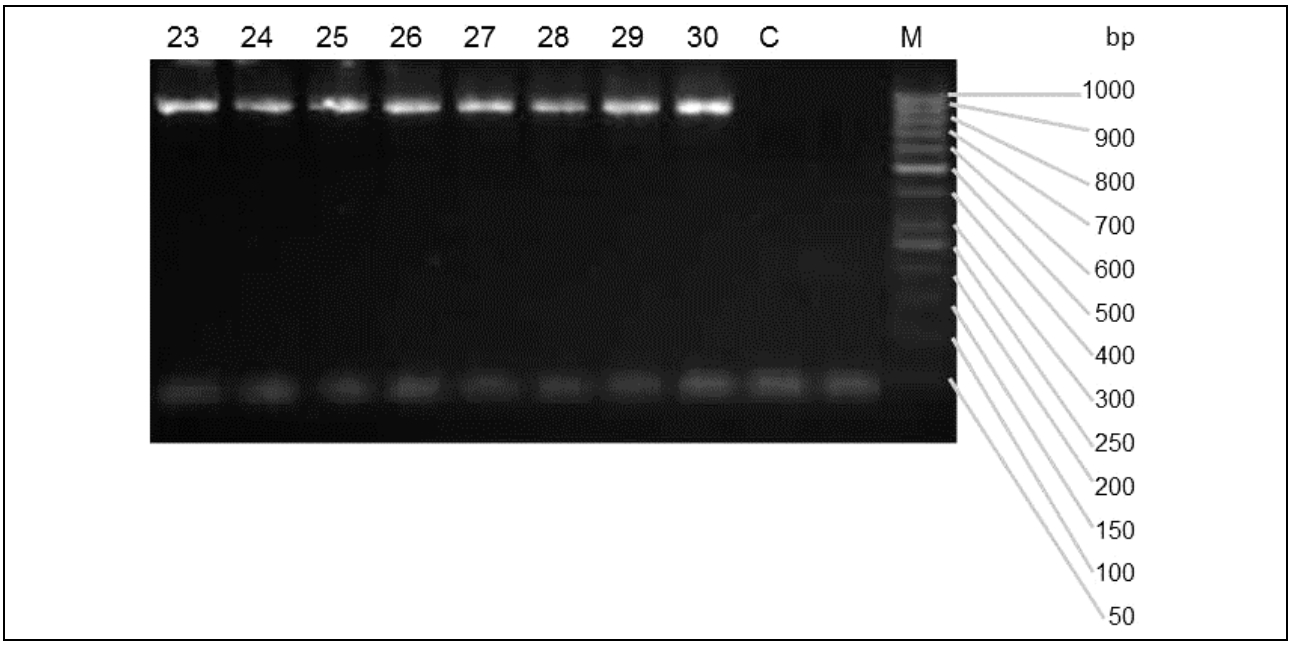


A = adenine; C = cytosine; G = guanine; T = thymine (NC-IUB, 1985). The reported rs-number has been indicated with a red banner at the specific loci of interest.

5.2.3.5 Amplification of fragment five in the *ApoE* gene

Fragment five is 869 bp in length and contains exon four of the *ApoE* gene. A 54 bp overlap occurred between fragment four and five (Figure 4.5). Figure 5.10 illustrates the successful amplification of fragment five in a subset of the 30 random samples from the PURE study using the protocols and T_a provided in Table 4.1 in Section 4.3.2. The absence of any amplified product in the negative control lane indicates that no contamination occurred (Figure 5.9, Lane C). However, the presence of the primer dimers indicated that the primers did not attach as effectively to the amplified region as they did to one another. However, since the primer dimers are only 20-25 bp in length it does not affect the sequencing analyses and the actual fragments do not need to be removed from the agarose gel to obtain effective sequencing data.

Figure 5.9 Photographic representation of the successful amplification of fragment five of the *ApoE* gene at 61.4°C

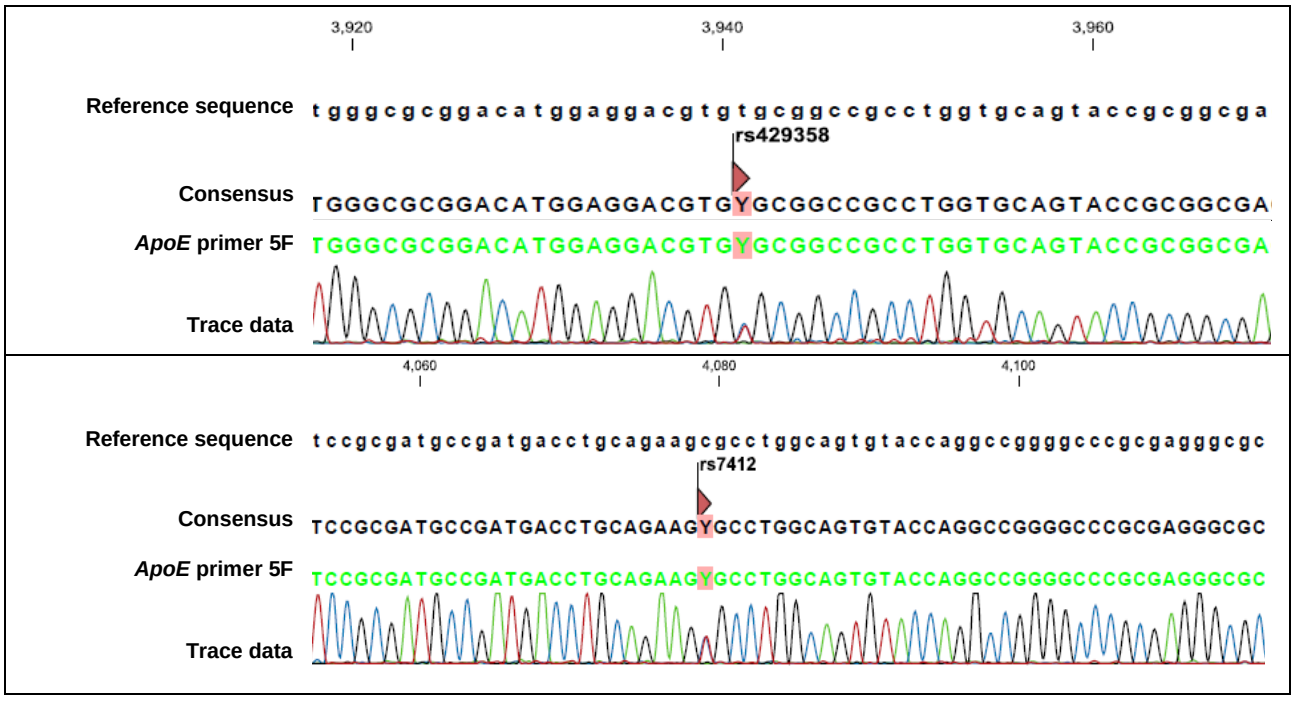


Amplified DNA product of a subset ot the 30 random samples was separated by a 2% agarose gel electrophoresis at 100 Volts for 60 minutes in 1% Tris/Borate/EDTA (TBE buffer); lane 31: pure water as the negative control; lane M: 50 bp Fermentas GeneRuler™ marker

After aligning the sequence products obtained from amplified fragment five, of all 30 randomly selected samples of the PURE study, to the reference sequence, only two genetic alterations were identified: rs429358 and rs7412. Four other genetic alterations (rs11542040, rs1108375, rs11542035, and rs769455) which were identified from a computer-based literature search, also occurred in fragment five (Figure 4.5). These SNPS were, however, already excluded after the custom GoldenGate genotyping assay had been applied to validate the initial SNP list to be included for genotypic analyses *via* the BeadXpress® Reader (Section 4.3.4). Four of the six SNPs that occurred in this

fragment were less than 60 bp apart (rs429358 rs11542035, rs769455 and rs7412). However, it was vital to obtain the genotypic data for SNPs rs429358 and rs7412, since these SNPs are vital in enabling comparisons with the extensive amount of literature available for the haplotype-genotypes that occur as a result of the strong LD between these two SNPs (Section 3.2). Therefore, these SNPs were included in the final SNP list as opposed to the other SNPs. Figure 5.10 is a graphic representation of the location of the two SNPs included (as illustrated by the CLC Workbench sequence viewer program).

Figure 5.10 Representative electropherogram for the SNPs rs429358 and rs7412 identified in fragment five in the *ApoE* gene



A = adenine; C = cytosine; G = guanine; T = thymine; Y = C or T (NC-IUB, 1985). The reported alteration has been indicated with a red highlight

5.2.3.6 Polymorphism in the *ApoE* gene, reported in literature

One of the *ApoE* polymorphisms that was included in the final SNP list, but was not identified by means of sequencing and was, therefore, included as a result of the literature review performed (as discussed in Section 4.3.3) is rs1080110. This SNP occurred in the promoter region of the *ApoE* gene, which was not sequenced in the present study. Furthermore, little attention has been paid to the characterisation of polymorphisms in other regions of the *ApoE* gene (as discussed in Section 3.4). However, Nickerson and co-workers (2000) were one of the first research groups to identify other SNPs in the *ApoE* gene, *i.e*. SNP rs1080110. All the SNPs identified had to apply to the BeadXpress® criteria as well (Section 5.3), which resulted in only the rs1080110 SNP being included in the final SNP set submitted and not the other SNPs that occurred in one of the aforementioned

fragments that were also identified from literature, *i.e*. rs11542029, rs11542040, rs1108375, rs11542035, rs769455 and rs1081106. Section 5.3 will discuss the high-throughput genotyping method *via* the BeadXpress® T Reader from Illumina, which was used to analyse the majority of the SNPs identified in the *ApoE* gene, whereas Section 5.4 will describe the genotyping of the rs7412 SNP with the use of the real-time PCR analyses, conducted on the Bio-Rad CFX96 qPCR apparatus.

5.3 BEADXPRESS® ANALYSES

In order to determine the *ApoE* genotypic status at the loci under investigation, rather than the standard phenotyping method determined by isoelectric focusing followed by SDS-PAGE and immune blotting with anti-*ApoE* (Menzel & Utermann, 1986), a DNA-based technique was used. DNA amplification technologies used for genotyping are much more advanced techniques and less prone to error.

The results obtained from designing the custom GoldenGate® genotyping assays are indicated in the following section. These results were used to determine the final list of SNPs to be investigated further and the results thereof are also discussed in greater detail in the succeeding sections.

5.3.1 Designing custom GoldenGate® genotyping assays

Section 4.3.3 describes the process used to identify and select the list of SNPs in the *ApoE* gene, which were submitted for validation for analysis with the BeadXpress® T-Reader Platform. Initially, a total of 20 SNPs were identified from the literature as well as the sequenced data of the 30 randomly selected samples from the PURE study. This extended list was submitted for genotyping on the BeadXpress® T-Reader Platform. However, from this list, seven SNPs (two novel SNPs at position 45410498 and 45410550 on chromosome 19, as well as rs769447; rs11542029; rs11542040; rs1108375 and rs11542035) were excluded, as these SNPs were less than or equal to 60 bp apart from each other. Further criteria used were that GoldenGate-validated SNPs were favoured above the rest, as these SNPs have been used in previous assays and, therefore, had established primers. GoldenGate®-validated SNPs also have the highest chance of converting into successful assays. Furthermore, the criteria applied when two-hit validated SNPs had to be used was to use the SNPs with the higher SNPScores first.

The criteria and results that were used to determine this final SNP list are discussed in greater detail in the following paragraphs and are presented in Table 5.3. Table 5.3 represents the validation scores for the identified SNPs that were to be investigated *via* the GoldenGate® Genotyping assays. The higher the values of the SNP score (between 0.0-1.1), the greater the ability to design a successful assay. An SNP score of 1.1 represents a GoldenGate®-validated SNP that has previously been designed and has successfully generated polymorphic results on the Illumina® platform. The only GoldenGate®-validated SNP in this analysis was rs440446 (Table 5.3).

Table 5.3 Validation scores for the identified SNPs within the *ApoE* gene investigated in the PURE study

Locus Name*	Final SNP Score	Designability Rank	Failure Codes	Validation Bin	Validation Class
rs1081101	0.522	0.5	--	NonValidated	1
rs769446	0.257	0	360	NonValidated	1
rs405509	0.845	1	--	TwoHit_OneKGenomeValidated	200
rs440446	1.1	1	--	GoldenGate®Validated	900
rs769448	0.701	1	--	NonValidated	1
rs769449	0.583	0.5	--	OneKGenomeValidated	100
rs769450	0.37	0	360	OneKGenomeValidated	100
*ApoE*LS01 novel	0.965	1	340	Unknown	0
rs769452	0.518	0.5	--	NonValidated	1
rs429358	0.418	0.5	--	TwoHit_OneKGenomeValidated	200
rs769455	0.394	0	360; 340	HapMapValidated	120
rs7412	0.572	0.5	340	TwoHitValidated	110
rs1081106	0.399	0	360	NonValidated	1

*dbSNP version 135 was used as the source of SNP information during these analyses; rs = reference sequence; SNP = single nucleotide polymorphism

The SNP score represents the ability to design a successful assay. The designability rank for each SNP is a simplified representation of the SNP score for easily sorting and filtering the results. An SNP score of <0.400 equals a rank of 0 and indicates a low success rate of the OPA, when included in the OPA tube. Four SNPs (rs769450; rs769455; rs1081106; rs769446) obtained a failure score of 360 after the SNP analysis phase, as these SNPs' final SNP scores were less than 0.4. Assays with final SNP scores of <0.4 were avoided, since these assays could decrease the overall performance of all assays. An SNP score of between 0.400 to <0.600 equals a designability rate of 0.5 and points to the particular SNP having a moderate success rate and a moderate risk of failure of the OPA. An SNP score between the values of 0.600 and 1.1 equals a designability rank of 1, which indicates a high success rate of the OPA for that given SNP. Therefore, the SNPs at

rs1081101, rs769452, rs429358, rs7412 and rs769449 had a moderate success rate and moderate risk to the success of the OPA, whereas the SNPs at rs440446, rs405509 and rs769448 had a high success rate and low risk to the success of the OPA (Table 5.3). The SNP at rs7412 and the novel SNP had moderate to high success rates; however, these SNPs obtained a failure score of 340 after the SNP analysis phase, probably because of other SNP(s) in the list being closer than 61 nucleotides away. However, as mentioned in Section 5.2.3.5, SNP rs7412 was still included in the final SNP list to be analysed statistically, as it was genotyped on the Bio-Rad CFX Manager© (version 2.0). Because of limited funds, only one of these SNPs (which could not be genotyped by the BeadXpress®) could be genotyped on the Bio-Rad CFX Manager© (version 2.0). Consequently, the literature is primarily focused on the *ApoE* haplotypes (ε2, ε3, ε4), which are formed in response to the strong LD between the rs7412 SNP and the rs429358 SNP. Therefore, the decision was made rather to obtain the genotypic data for the rs7412 SNP in order to be able to determine the haplotypic data, to compare with published results.

The “unknown” validation status given for the ApoE LS01 SNP is not unexpected, seeing that this was a novel SNP identified in four of the 30 random samples from the PURE study and, therefore, supporting the fact that this SNP is not reported in the dbSNP database (Table 5.3). The other validation bin statuses mean the following: “Non Validated” indicates that the SNPs rs1081101, rs769446; rs769448; rs769452 and rs1061106 have only been analysed by a single method or in a single population before and that even if these SNPs have a high design score, it is probable that they may be monomorphic. The “Two Hit-” validation status ascribed to the SNPs rs7412, rs429358 and rs405509 indicates that both the alleles of these SNPs have been reported by two independent methods or in two or more different populations or have been validated by the HapMap project. The “OneK Genome Validated” status assigned to SNP rs769449 and rs769450 indicates that these SNPs have been validated in the 1000 Genomes Project, which aimed to discover the majority of the genetic variations that occur at population level with a frequency greater than 1%. This was achieved through deep sequencing of approximately 1000 individuals from various populations around the world, using next generation platforms and technologies (Via *et al*., 2010). After careful consideration of all the above-mentioned information, a final SNPscore file was assembled and the SNPs mentioned in Table 5.4 were sent off for final assay design and analysis. Therefore, only those SNPs listed in Table 5.4 were genotyped by either the BeadXpress® T Reader or by using real-time PCR and all discussions and/ or statistical analyses were based on these SNPs alone.

Table 5.4 Summary of the characteristics of the SNPs in the *ApoE* gene that were genotyped in black South African subjects

Chr 19 Position	Strand	Localisation	rs-number	SNP	Type of mutations
45408077	Pos	Flanking	rs1081101	C/T	Transition
45408836	Pos	Promotor	rs405509	A/C	Transversion
45409167	Pos	Intron	rs440446	C/G	Transversion
45409579	Pos	Intron	rs769448	C/T	Transition
45410002	Pos	Intron	rs769449	G/A	Transition
45410498	Pos	Intron	*ApoE*LS01 novel	G/C	Transversion
45411110	Pos	Exon	rs769452	C/T	Transition (Leu/Pro)
45411942	Pos	Exon	rs429358	T/C	Transition (Cys/Arg)
45412080	Pos	Exon	rs7412	C/T	Transition (Arg/Cys)

A = adenine; Arg = arginine; bp = base pair; C = cytosine; Chr = chromosome; Cys = cysteine; G = guanine; Leu = leucine; non-syn = non synonymous mutation; Pos = positive; Pro = proline; rs-number = reference sequence number; SNP = single nucleotide polymorphism; T = thymine

5.3.2 Generation and analysis of the GoldenGate® genotyping data

The Illumina® GoldenGate® Genotyping assay is pre-optimised and flexible as it provides researchers with the ability to design custom panels for high-quality, low to mid-multiplex genotyping. Implementing quality control methods to optimise call rates, evaluate assay performance, ensure sample quality and determine locus performance, ensures optimal final data quality from the GoldenGate® genotyping assay. First, from an initial general assessment of the assay performance that was conducted, it was determined that a certain number of failed or suboptimal samples should be removed, to permit more detailed sample quality evaluations. Subsequently, each locus was evaluated for editing and exclusion from the assay to optimise call rates. The information metrics that were used to ensure that the highest quality data were included were as follows: controls, Gencall Score, Cluster Separation, SNP Call Frequency, and AB Mean for Intensity (R): AB Mean for Theta (T) ratio, all of which are discussed in further detail in the succeeding paragraphs.

The initial step in evaluating assay performance was to analyse the performance of the numerous control analyses performed. Sample-dependent (positive), sample-independent as well as contamination (negative) controls are all built into the GoldenGate® assay to provide a manner in which the quality of the overall performance of samples, reagents, equipment and the BeadChip can be assessed.

The Universal BeadChip platform consists of several individual arrays manufactured on a microscope slide-shaped substrate. Each individual array on the BeadChip contains 1 536 IllumiCodes (each representing a unique 23 bp single-stranded DNA oligonucleotide to identify each DNA sample correctly, as well as each locus being interrogated) attached to a 3-micron bead. Several of these beads, which are represented by the IllumiCode sequence IDs, are included in the analysis and are used as controls. Table 5.5 is a list of the IllumiCode sequence IDs of the different controls included in the assay, with a description and the expected outcome for each sequence ID listed. In the table the terms "U3 match" and "U5 match" are used repeatedly and they represent a signal in the Cy3 and Cy5 channels, which is discussed in greater detail in Section 4.3.5.

Table 5.5 IllumiCode Sequence IDs of the control analyses undertaken in the BeadXpress® system, along with the description and expected outcome of each

IllumiCode sequence ID	Description	Expected outcome
329	AA mismatch	U3 match
1611	CC mismatch	U5 match
1142	GG mismatch	U3 match
279	GT mismatch	U5 match
1742	High AT (31% GC)	U3 match
4824	High GC (62% GC)	U5 match
658	15-base-pair gap	U3 and U5 match
962	First hybridisation controls, 42/57 T_m	U5 match
1209	First hybridisation controls, 57/72 T_m	U5 match
44	Second hybridisation controls	U3 match
278	Second hybridisation controls	U3 match
1112	Second hybridisation controls	U5 match
1632	Second hybridisation controls	U5 match
501	Second hybridisation controls	U3 and U5 match
1003	Second hybridisation controls	U3 and U5 match
3802	Contamination controls	U3 and U5 match
3845	Contamination controls	U3 and U5 match
2832	Contamination controls	U3 and U5 match
3864	Contamination controls	U3 and U5 match

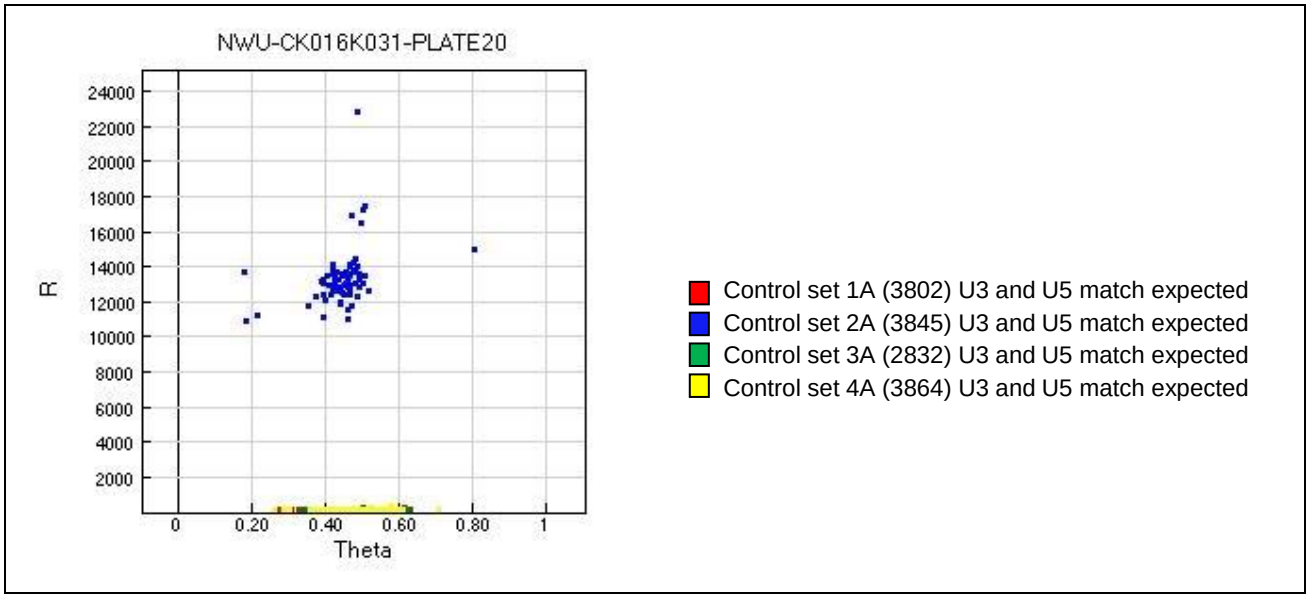
A = adenine; C = cytosine; G = guanine; T = thymine; T_a = annealing temperature; U3 = refers to the Cy3 channel; U5 = refers to the Cy5 channel

The genotype accuracy is determined by a number of factors, such as the quality of the: DNA used, genotype assay and the system itself. The procedures used to ensure good quality DNA have been discussed in Section 4.3.1. To ensure optimal genotyping accuracy a quality control procedure is performed on the DNA, along with the built-in

internal controls that are performed with each reaction (http://support.Illumina®.com/faqs.ilmn#what_has_Illumina®_done_to_ensure_a_high_level_of_genotyping_accuracy).

Contamination controls are included to ensure and assess that no contamination occurred in the plate under investigation. There are four types of PCR contamination detection controls (*i.e.* ILMN-3802, ILMN-3845, ILMN-2832and ILMN-3864), of which only a single type of control should be detected per OPA tube. Therefore, lack of contamination is indicated if only one contamination control type has a high signal, which is visually identified by the number of colours that are displayed. If two or more contamination control types have a high signal, this indicates that significant contamination occurred. Therefore, Figure 5.11 represents a contamination-free environment of a single plate, as only one contamination control type presented a high signal (blue/ ID Sequence 3845). However, this figure is also representative of the graphs for all the other plates. Therefore, carry-over contamination did not occur in any of the plates.

Figure 5.11 GoldenGate® Assay contamination control as displayed in the GenomeStudio® Genotyping Module Controls Dashboard for the current study

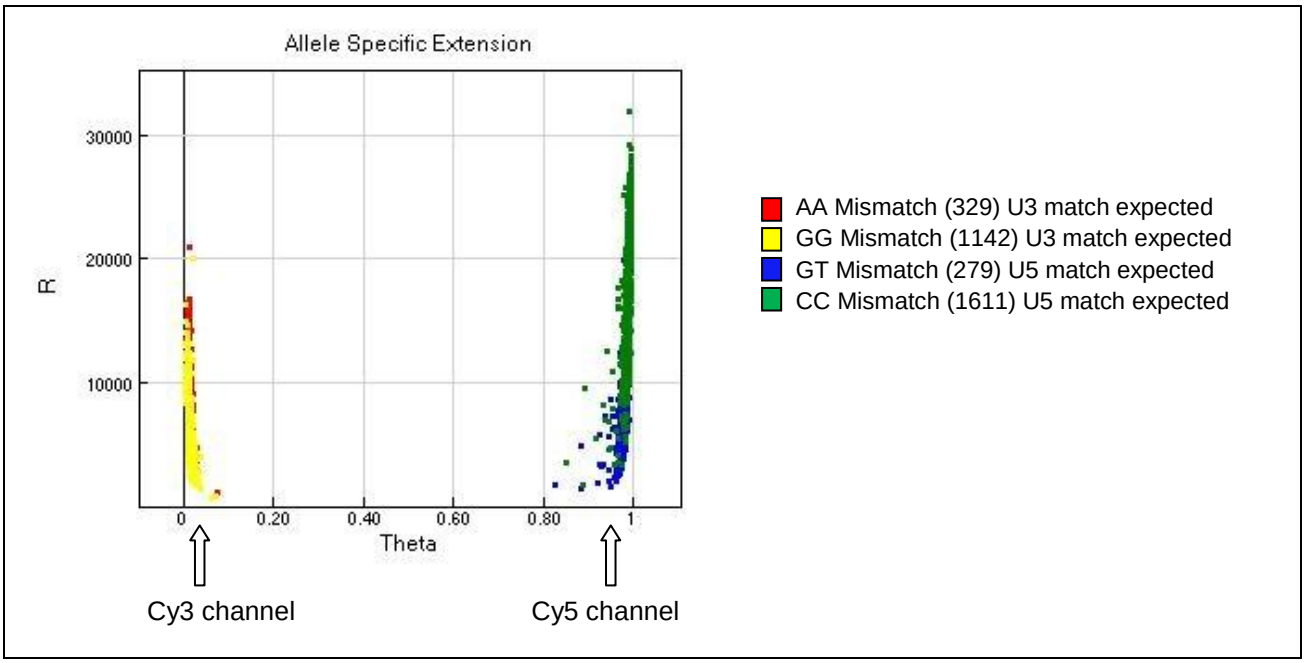


A = adenine; C = cytosine; G = guanine; T = thymine; U3 = refers to the Cy3 channel; U5 = refers to the Cy5 channel

The Allele-Specific extension control is sample-dependent and measures the extension efficiency of the properly matched and mismatched ASO. These controls test for A-A, C-C, G-G, and G-T mismatches corresponding with IllumiCode Sequence IDs 329, 1611, 1142, and 279, respectively. In the event that this control is unsuccessful, it might suggest either a processing failure or poor DNA sample quality. However, this control was

successful in the analyses undertaken in this study, as the sequence IDs 329 and 1142 were predominately Cy3, and Sequence IDs 279 and 1611 were predominately Cy5, as expected (Figure 5.12).

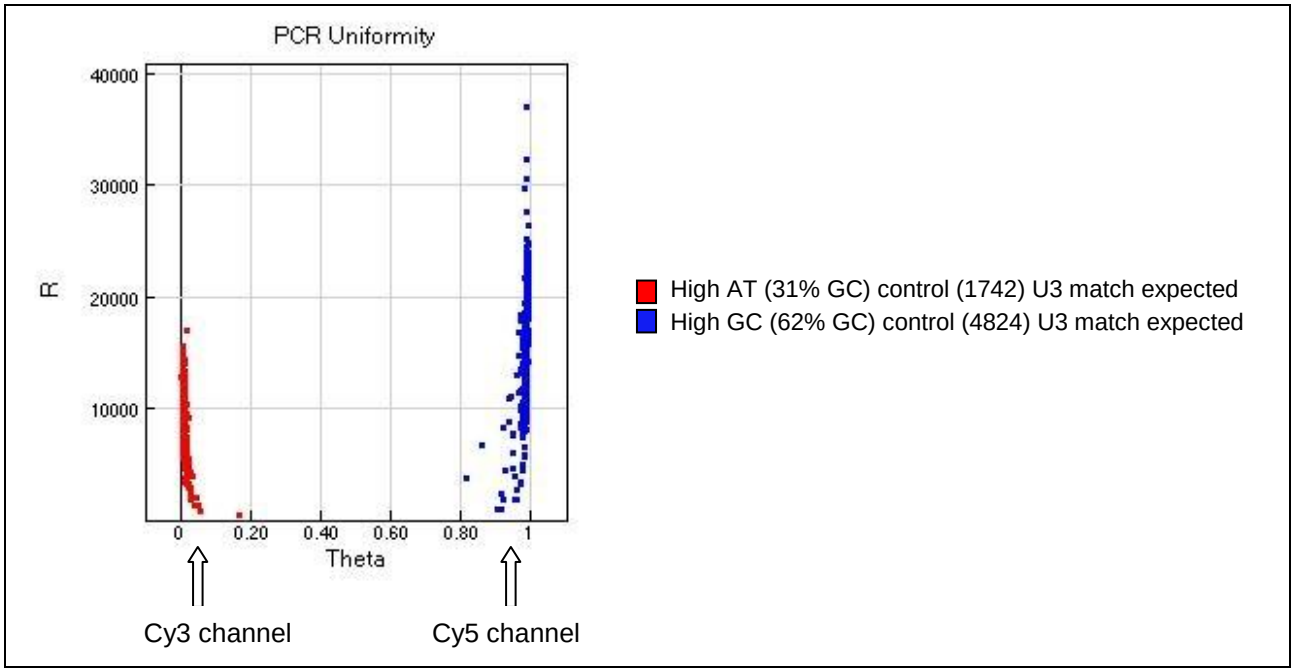
Figure 5.12 GoldenGate® Assay allele specific extension control displayed in the GenomeStudio® Genotyping Module Controls Dashboard for the current study



A = adenine; C = cytosine; G = guanine; T = thymine; U3 = refers to the Cy3 channel; U5 = refers to the Cy5 channel

The amplification efficiency for the high AT and GC-rich regions of the DNA is tested with the use of the PCR uniformity control. Figure 5.13 indicates that the PCR uniformity control was successful, as the expected outcome was met in that the IllumiCode Sequence ID 1742, which ensures amplification efficiency for high AT-rich (31% GC) regions, resulted in a Cy3 signal. Furthermore, the IllumiCode Sequence ID 4824, which amplifies over a high GC-rich (62% GC) region, resulted in a Cy5 signal, as expected (Figure 5.13).

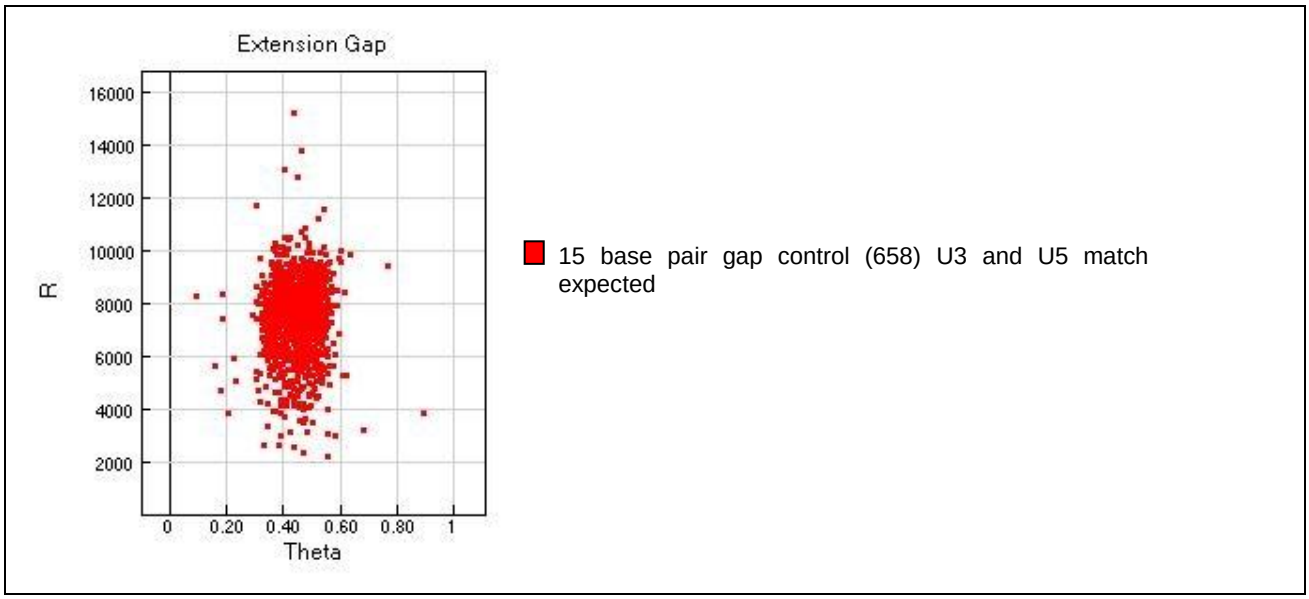
Figure 5.13 GoldenGate® Assay PCR uniformity control displayed in the GenomeStudio® Genotyping Module Controls Dashboard for the current study



A = adenine; C = cytosine; G = guanine; T = thymine; U3 = refers to the Cy3 channel

The extension gap control (IllumiCode Sequence ID 658) tests the efficiency of extending the 15 bases from the 3' end of the ASO to the 5' end of the LSO. Figure 5.14 represents a successful run of the extension gap control as both Cy3 and Cy5 signals were detected in the heterozygote theta range for this control and throughout the investigation, as was expected.

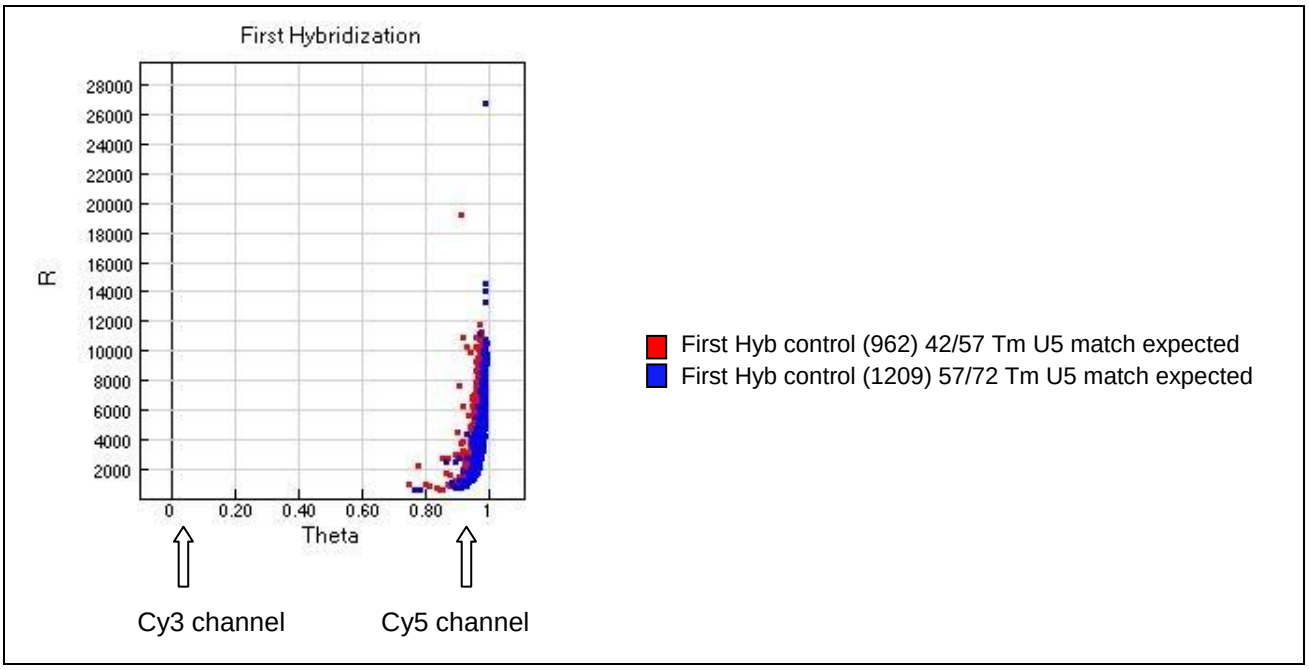
Figure 5.14 GoldenGate® Assay extension gap control displayed in the GenomeStudio® Genotyping Module Controls Dashboard



U3 = refers to the Cy3 channel; U5 = refers to the Cy5 channel

The annealing efficiency of the ASOs at a specific locus is measured by performing the annealing step at two different T_m's, through the use of the first hybridisation control. As graphically presented in Figure 5.15, the IllumiCode Sequence ID 962 and 1209 resulted in the expected Cy5 match. This indicates that the specificity of annealing ASOs at the different T_m's was sufficient for the purposes of this analysis.

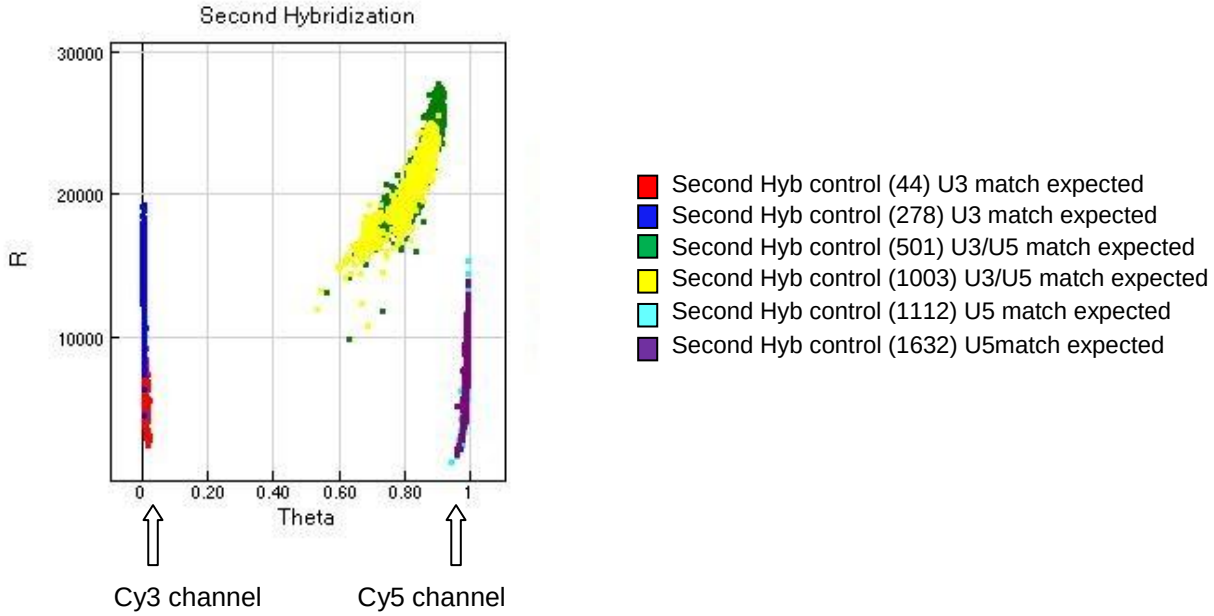
Figure 5.15 GoldenGate® Assay First Hybridisation Control displayed in the GenomeStudio® Genotyping Module Controls Dashboard in this study



Hyb = hybridisation; T_m = melting temperature; U5 = refers to the Cy5 channel

The second hybridisation control tests the hybridisation of the single-stranded assay products to the IllumiCode sequences on the array beads. Figure 5.16 illustrates that the IllumiCode Sequence IDs 44 and 278 resulted in the expected Cy3 signal only, whereas the Sequence IDs 1112 and 1632 resulted in only a Cy5 signal as expected, and Sequence IDs 501 and 1003 contributed to both the Cy3 and Cy5 signal.

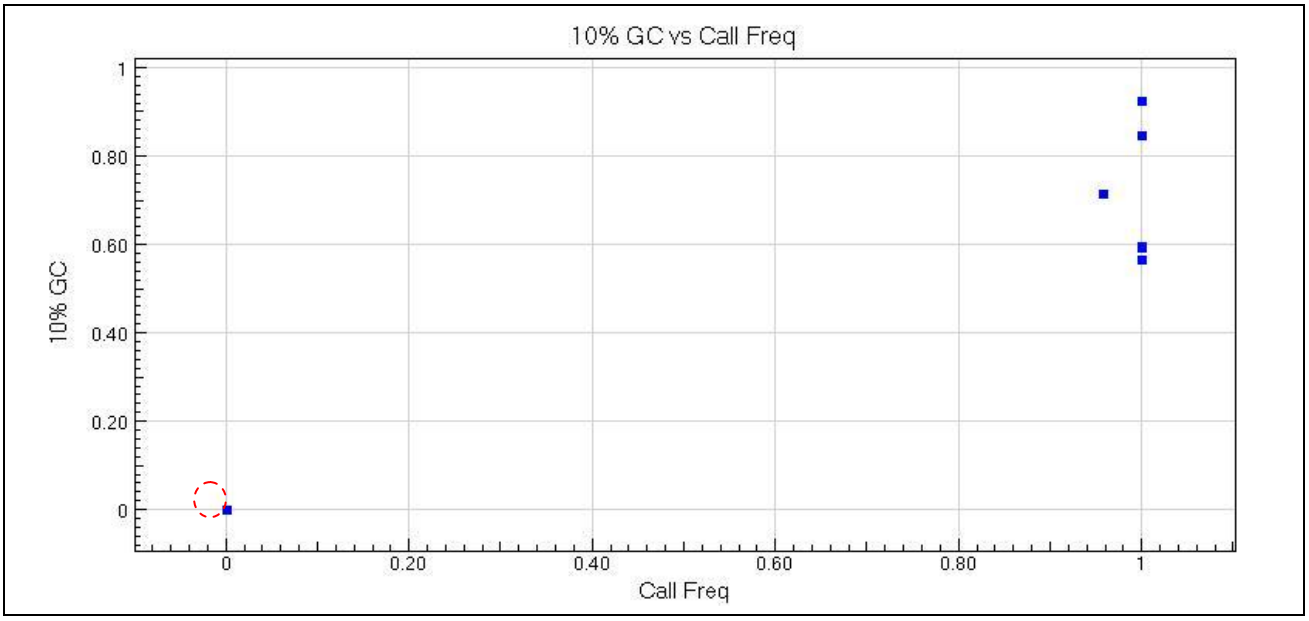
Figure5.16 GoldenGate® Assay Second Hybridisation Control displayed in the GenomeStudio® Genotyping Module Controls Dashboard in this study



Hyb = hybridisation; U3 = refers to the Cy3 channel

In order to establish which samples had performed poorly in the genotyping assay, the quality metric referred to as the GenCall score (ranging from 0-1) was calculated and analysed for each data point. Generally, the GenCall score decreases in value the further the data point is from the centre of a genotype cluster. Furthermore, any of the samples that had a call rate < 0.90 was excluded, since the call rate was a metric applied to ensure the highest quality of data. The call rate provides information on the proportion of all the samples with call scores above the no call threshold at every locus. Therefore, in order to identify problematic samples, a scatter plot is used to compare the call rates as a function of the sample number against a 10% GenCall Score (10% GC) as a function of the sample call rate (Figure 5.17). Samples with a GenCall score of < 0.5 were, therefore, excluded from the project.

Figure 5.17 Scatter plot of 10% GC Score compared to call rates of the investigated SNPs



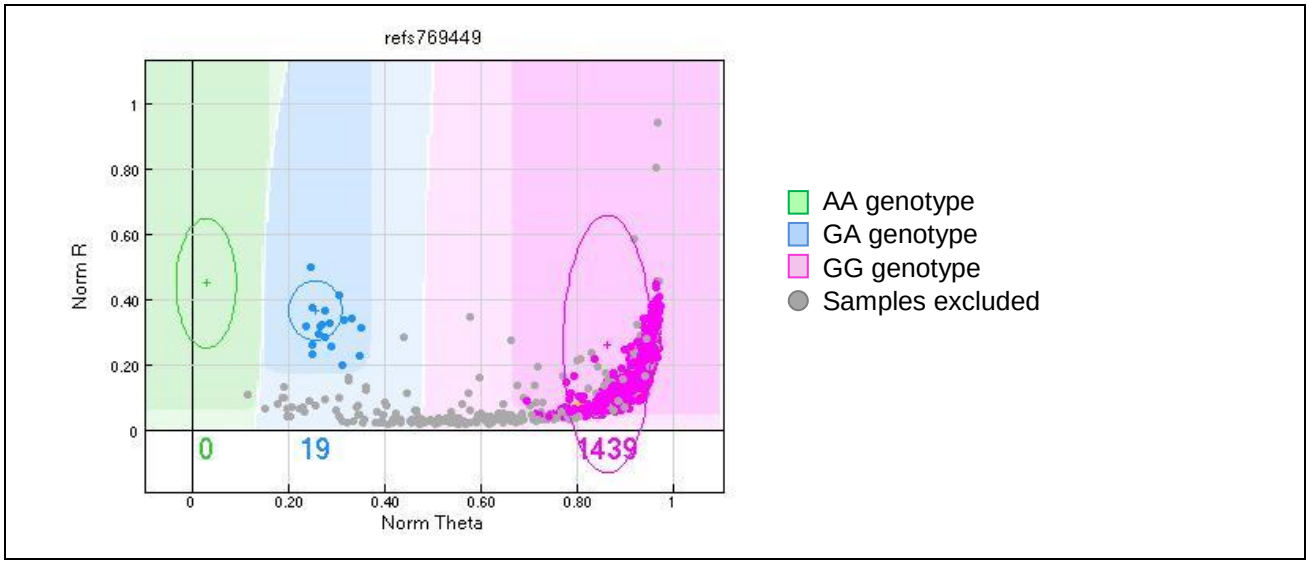
The poorly performing sample (red circle) is an obvious outlier from the majority of samples when 10% GC Score is plotted against sample call rate.

Although the majority of the SNPs had favourable validation scores (Section 5.3.1), some of these SNPs (*i.e*. rs769448, rs769449 and the novel SNP) were excluded (zeroed) from the final GenomeStudio report primarily because of overlapping clusters and ultimately poor genotype distributions, resulting in low cluster separation (< 0.02), hindering the possibility of analysing these SNPs statistically.

Another metric applied in the analyses of the genotyping data was the AB mean for intensity (R) and theta (T). The SNP table was sorted by the AB R mean value [the mean normalised intensity (R) of the heterozygote cluster], which identifies SNPs with low intensity data, with increasing values from 0. SNPs with too low intensities (R<1) were considered to be unreliable and were excluded. The AB T mean (the mean of normalised theta values of the heterozygote cluster) ranges from 0-1. Therefore, the major difference between the AB R mean and the AB T mean is that the former presents the intensity of an SNP (how strong the signal for a particular SNP was), whereas the latter represents the position of datapoints in relation to a specific cluster, *i.e*. the heterozygous cluster). Therefore, when the heterozygote cluster for a specific SNP overlapped with the homozygote cluster, that SNP would have a low AB T mean, *i.e*. between 0-0.2. The genotype signals for the two clusters could, therefore, not be distinguished from each other and thus the SNP had to be excluded.

The specific call region shading of the rs769449 SNP determined by the GenomeStudio™ program is illustrated in Figure 5.19 and will be used to elucidate the previous discussion. Even though the overall GenCall Score for SNP rs769449 was 0.6893, indefinite clusters were formed because of the low AB T mean of 0.255 along with the low cluster seperation value of 0.3841. Furthermore, the intensity of the data was doubtful due to the low AB R mean of 0.367. All this resulted in the exclusion of a large number of samples (indicated in grey), as they were too far from the centre of the cluster (Figure 5.18). Therefore, this SNP was not included in the remaining analyses.

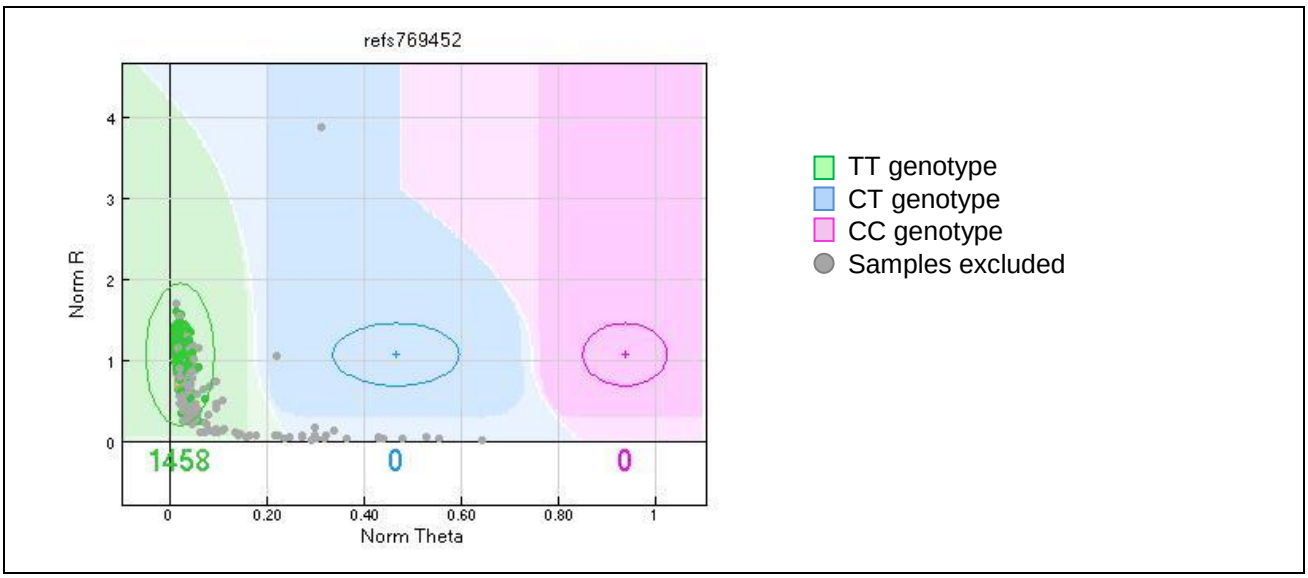
Figure 5.18 Genomestudio® shade call regions for SNP rs769449 as an example for low cluster separation scores



A = adenine; G = guanine

Figure 5.19 serves as another example of the specific reasons why one of the SNPs listed above (*i.e*. rs769452) was not included in the final SNP list. This figure graphically represents the specific call region shading for the rs769452 SNP *via* the GenomeStudio® program. The GenCall Score of 0.5661 for this SNP was low and indicated a poor clustering call for these samples. A large number of samples (indicated in grey in Figure 5.19) were excluded for the same reasons as explained in the former example. Even though, the AB R mean (1.077) value indicated that the intensity of the data was reliable, the AB T mean (0.465) and the cluster separation (0.4359) values indicated that the clusters could only be partially separated and therefore this SNP was not included either in the final SNP list analysed.

Figure 5.19 Genomestudio® shade call regions for SNP rs769452



C = cytosine; T = Thymine

Therefore, based on the aforementioned criterion discussed in the previous and current section, of the nine SNPs analysed, only four in the *ApoE* gene were ultimately included in the statistical analyses. It can be speculated that the high GC content and/or high homopolymeric regions in the *ApoE* sequence resulted in the inefficiency of the genotypic analysis conducted for most of the SNPs identified in the *ApoE* gene. However, SNP rs7412 was genotyped through real-time PCR analyses and the established control measurements for this analysis are discussed in the following section. The graphical illustrations and implications of the aforementioned findings are presented and discussed in Section 5.5.

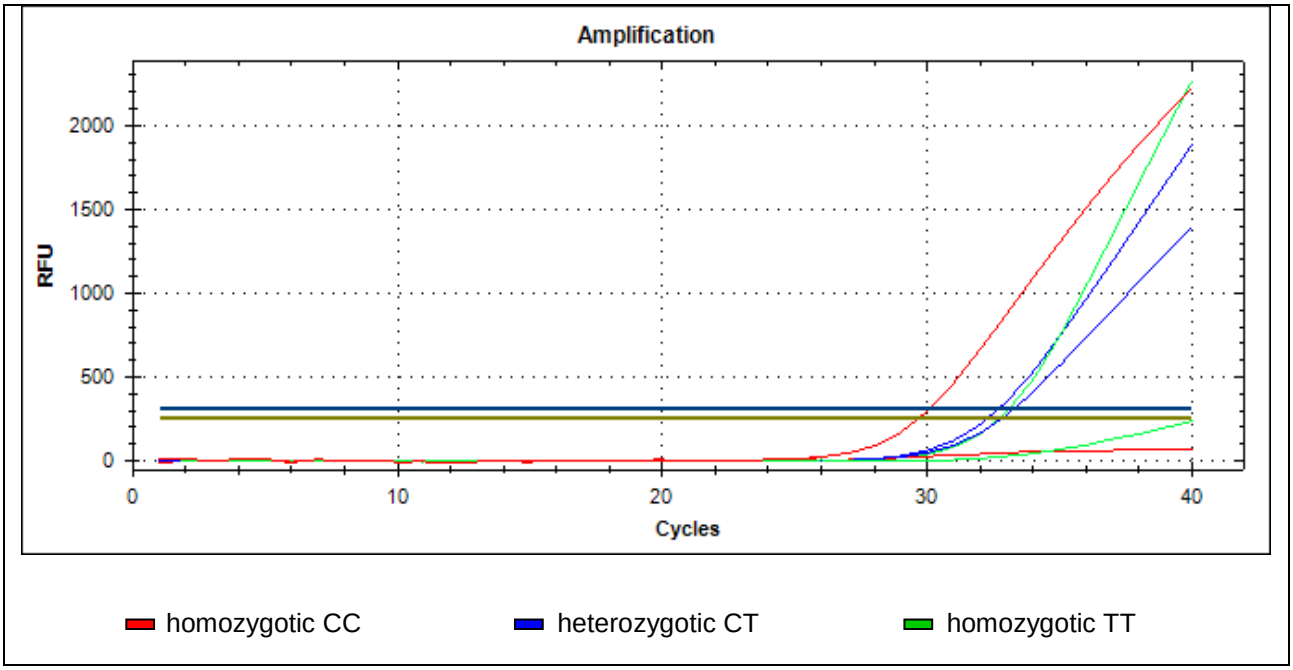
5.4 REAL-TIME PCR ANALYSES

The real-time PCR analyses were conducted on the Bio-Rad CFX96 qPCR apparatus (as discussed in Section 4.3.5.1) in a 96-well format. The control measures implemented for these analyses included the insertion of three no-template controls (NTC) in each batch (96-well plate) undertaken, positioned at A1, E6 and H12, respectively, in each plate, where the DNA was replaced with nuclease-free ddH_20. Control individuals of the reported genotype (homozygote AA, heterozygote AT and homozygote TT, respectively) were included in each plate and genotyping was carried out blinded to plasma lipid profile concentrations (Figure 5.20).

Figure 5.20 is a graphical representation of how the FAM and VIC probes are indicated by the Bio-Rad CFX Manager© (version 2.0) that was used. The sole FAM excitation

spectrum represents a homozygous C genotype (CC), whereas the sole VIC excitation spectrum represents the homozygous T genotype (TT). Therefore, if excitation of both probes occurred, the heterozygous genotype (CT) was present. The data obtained were stored and interpreted by the author of the study and Dr KR Conradie, as discussed in Section 4.3.5.1.

Figure 5.20 Controls used in the genotype analyses of the rs7412 SNP via the Bio Rad CFX Manager© (version 2.0)



C = cytosine; RFU = Relative Fluorescence Units; T = thymine; Blue graph represents the heterozygote genotype (AT), red graph represents the wild type homozygote (AA) and the green graph represents the mutant heterozygote (TT).

5.5 ALLELE FREQUENCY DETERMINATION

The obtained MAF of each SNP, was compared to the MAF reported by dbSNP for other etnic groups. Genetic association analyses were conducted for each of the genetic variants (SNPs and haplotypes) investigated with the four determinants of the blood lipid profile investigated in this study. Adherence to the assumptions of HWE was determined for each SNP by means of the chi-square test for goodness of fit in order to determine whether the genotype frequencies remained constant within a population and as a quality control measure. Differences in the concentrations of the lipid profile determinants were determined between each genotype group per SNP for the total cohort, as well as for the rural *versus* urban groups and male *versus* female groups individually using the factorial ANOVA test. Prior to these analyses, possible confounders that may have influenced the associations were identified and were adjusted for in the statistical analysis. According to the literature age, gender, BMI, HIV-status, tobacco use, alcohol consumption, physical activity, dietary fat intake and fibre intake have all been reported to be confounders of the

lipid profile of individuals (Bazzano, 2005). In the present cohort, a number of these confounding variables were determined to correlate differently with each of the different lipid profile determinants. Table 5.6 is a summary of the results obtained for the Spearman Rank Order correlations assessment between the lipid profile determinants and the possible confounding variables. The r-values indicated in red had a p-value of less than 0.05 (Table 5.6).

Table 5.6 Spearman Rank Order correlations between identified variables and the lipid profile in the current PURE cohort

Variables	TC	LDL-C	HDL-C	TG
Age	0.16	0.11	0.06	0.20
Total energy intake	0.02	-0.05	0.11	0.02
Total fibre intake	0.02	-0.03	0.10	0.01
Total fat intake	0.06	0.04	0.04	0.07
Total SFA intake	0.06	0.05	0.01	0.08
Total MUFA intake	0.06	0.04	0.02	0.08
Total PUFA intake	0.05	0.02	0.05	0.05
Total TFA intake	0.05	0.03	0.01	0.07
Total cholesterol intake	0.08	0.05	0.05	0.07
GGT	0.16	-0.03	0.29	0.16
PAI-1_{act}	0.11	0.14	-0.18	0.29
CRP	0.03	0.07	-0.14	0.13
SBP	0.17	0.08	0.15	0.14
DBP	0.15	0.07	0.13	0.13
Hcy	0.06	-0.05	0.19	0.001
Fibrinogen	0.08	0.13	-0.09	0.07
BMI	0.19	0.28	-0.24	0.29

BMI = body mass index; DBP = diastolic blood pressure; GGT = gamma-glutamyl transferase; CRP = C-reactive protein; HDL-C = high density lipoprotein cholesterol; Hcy = homocysteiene; LDL-C = low density lipoprotein cholesterol; MUFA = monounsaturated fatty acids; PAI-1_{act} = plasminogen activator inhibitor type 1; PUFA = polyunsaturated fatty acids; SBP = systolic blood pressure; SFA = saturated fatty acids; TC = total cholesterol; TFA = trans fatty acids; TG = triglycerides.

Various variables had weak ($r \leq 0.1$), but significant correlations with TC, LDL-C, HDL-C and TG (Table 5.6). Large correlation values ($r > 0.2$) for continuous variables with the various lipid profile determinants were considered to be confounders. Furthermore, a number of smaller, yet significant correlations (≈ 0.15) were also observed for which some of these variables were included in the list of possible confounders, if it was argued that the correlation could be of practical significance (Table 5.7). However, these small, but significant correlations could be ascribed to the large sample size of this study and are most likely not of biological significance. When each of the variables of importance per lipid profile determinant (as indicated in Table 5.7) was corrected for, the initial significant associations remained and, therefore, only the unadjusted values are reported, unless otherwise stated.

Table 5.7 Variables confounded for per lipid profile determinant in the present cohort

Lipid profile determinant	Co-variants
TC	age, GGT, SBP, BMI
LDL-C	BMI
HDL-C	GGT, PAI-1_{act}, SBP, Hcy, BMI
TG	age, GGT, PAI-1_{act}, BMI

BMI = body mass index; GGT = gamma-glutamyl transferase; HDL-C = high density lipoprotein cholesterol; Hcy = homocysteiene; LDL-C = low density lipoprotein cholesterol; PAI-1_{act} = plasminogen activator inhibitor type 1; SBP = systolic blood pressure; TC = total cholesterol; TG = triglycerides.

Lastly, the interactive effects between various variables (locality, gender, diet and each of the confounding variables) against each of the lipid profile determinants, categorised by the three genotypes of each specific SNP/haplotype, were also analysed and only the significant interactions were examined in greater detail in the succeeding discussions. Furthermore, a two-grouping approach (running these tests in an amalgamated heterozygote/homozygote mutant group *vs.* a homozygote wild type group) was followed when the homozygote mutant group included fewer than 30 samples.

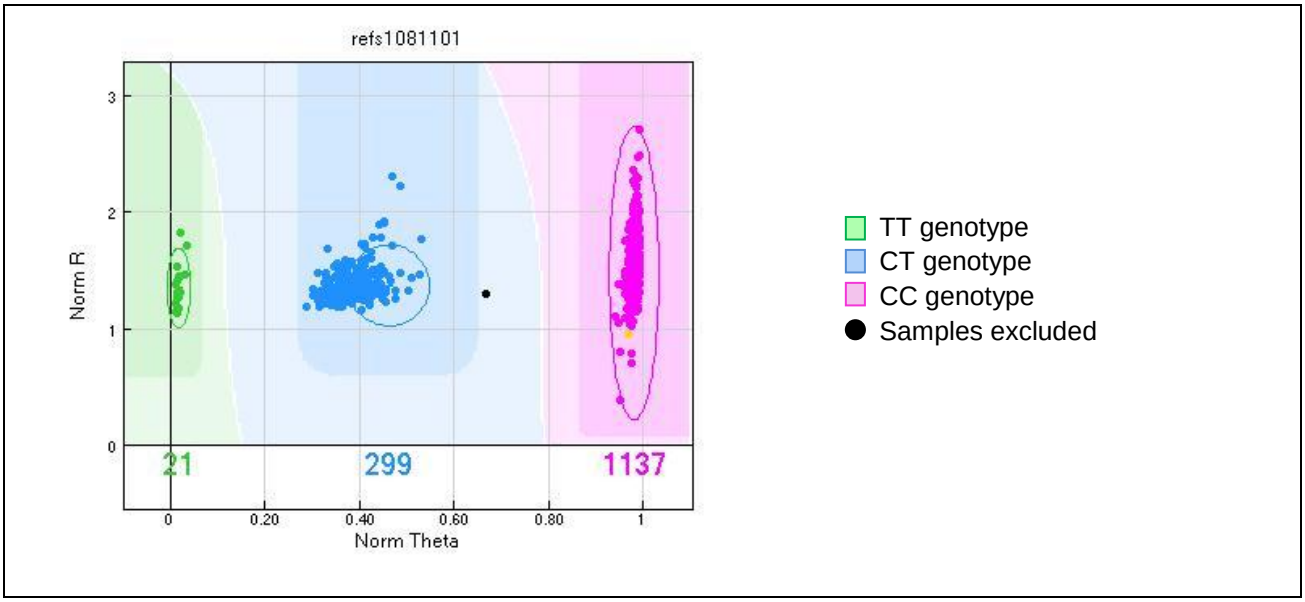
Therefore, all gene interactions pertaining to gender were referred to as gene-gender interactions, whereas the gene interactions related to residential locality were referred to as gene-environment interactions. The term “genotype-phenotype” interaction was used when describing gene interactions with any of the confounding variables and “gene-diet” when any of the continuous dietary data were analysed. All of the aforementioned genetic characteristics of the study population will be discussed per identified SNP that is under investigation, in order to assist in the logical flow and presentation of the data. It is important to highlight that these analyses were exploratory in nature and, therefore, the results obtained, represented in Table 5.8 to Table 5.19 (which present a summary of the analytical findings for each genetic variant), will be discussed with great care in the following sections.

5.5.1 SNP rs1081101

Figure 5.21 represents the call region shading of the rs1081101 SNP as determined through the GenomeStudio® program. The GenCall Score for this SNP was 0.9165, indicating a reliable clustering call and that the majority of the samples are in the centre of the cluster. The values of the AB R mean (1.372) and AB T mean (0.462) of this particular SNP was optimal, indicating that the intensity of the data was reliable and also that separate and definite clusters could be outlined. The cluster separation for rs1081101 was

0.71, which also supports the assertion that the data obtained was reliable. The process followed to ensure that the correct alleles (for each of the investigated SNPs) were labelled accurately from the BeadXpress® data demanded close attention to the strand from which the assay was designed, which was then correlated with that of the actual alleles of each of the SNPs according to the dbSNP website.

Figure 5.21 GenomeStudio® shade call regions for SNP rs1081101 in the PURE dataset



C = cytosine; T = thymine

Cytosine (C) is the ancestral allele for rs1081101 and thymine (T) is the variant allele. The MAF of this SNP in the present black South African population is 0.23, which is higher than that reported for sub-Saharan Africans (0.07) as well as Europeans with a MAF as low as 0.004 (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1081101). Possible reasons for the MAF being more prevalent among the black South African population are provided in the next paragraphs in this section.

The high p-value ($p > 0.05$) of the chi-square test indicates that the rs1081101 SNP adheres to the assumptions of HWE (Table 5.8) and that no migration, in-breeding, positive selection *etc*. occurred within the present cohort group (Hardy, 1908). The difference between the MAF of the two African populations can be ascribed to the internal diversity of the African population. As a result, it has been suggested that all other populations arose from the African population, which is commonly referred to as the "out-of-Africa" theory (Campbell & Tishkoff, 2008; Jorde & Wooding, 2004). The higher MAF among the black South African population could, therefore, be ascribed to the varied

allele frequencies often observed throughout the African population, which could possibly be a result of genetic drift and natural selection (Cavalli-Sforza *et al*., 1994).

However, the larger MAF difference between the present black South African population and the European population cannot be ascribed to the aforementioned factors, but rather due to the out-of-Africa theory. This theory postulates that all modern humans initially arose from Africa and, therefore, those that remained in Africa have the highest degree of genetic diversity whereas the European population often present with lower MAFs because of factors such as positive selection (Cummings, 2011; Jorde & Wooding, 2004).

Furthermore, the 95% CI for these genotypes were determined by means of the SAS System Program. The observed genotype frequencies for SNP rs1081101 were within the 95% CI ranges specified in Table 5.8. Therefore, this sample population was sufficiently representative of the larger population in view of the narrow range of the 95% CI, which provided confidence in the results from the sample set investigated and confirmation that no sampling bias occurred.

Table 5.8 Adherence to the assumptions of Hardy-Weinberg Equilibrium of the rs1081101 locus in the present PURE population using chi-square test analyses

rs1081101 (C/T)					
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions					
Chi-square test	Genotypes	CC	CT	TT	p-value (χ^2)
	Observed numbers (O)	1,137	299	21	--
	Frequency (%)	77.97	20.59	1.44	
	Expected numbers (E)	1,135.07	301.86	20.07	
	$\chi^2 = (O-E)^2/E$	0.00	0.01	0.04	0.81 (0.06)
Genotype distribution	95% CI range (%)	75.84-80.10	18.51-22.67	0.83-2.05	--

% = percentage; χ^2= Chi-square value; C = cytosine; CI = confidence intervals; E = expected numbers; O = observed numbers; rs = reference sequence; T = thymine

None of the lipid profile determinants was significantly different between the genotypes in the total population or the other gender and locality groups (Table 5.9), indicating that this SNP is unlikely to have an effect on the blood lipid profile. The exploration for possible interaction effects revealed no significant gene-gender, gene-environment (locality) or gene-diet interactions for the rs1081101 SNP on any of the various lipid profile determinants (Table 5.9). However, one significant genotype-phenotype interaction was observed (Figure 5.22).

Table 5.9 Associations of the lipid profile determinant concentrations with rs1081101 locus in the present PURE population

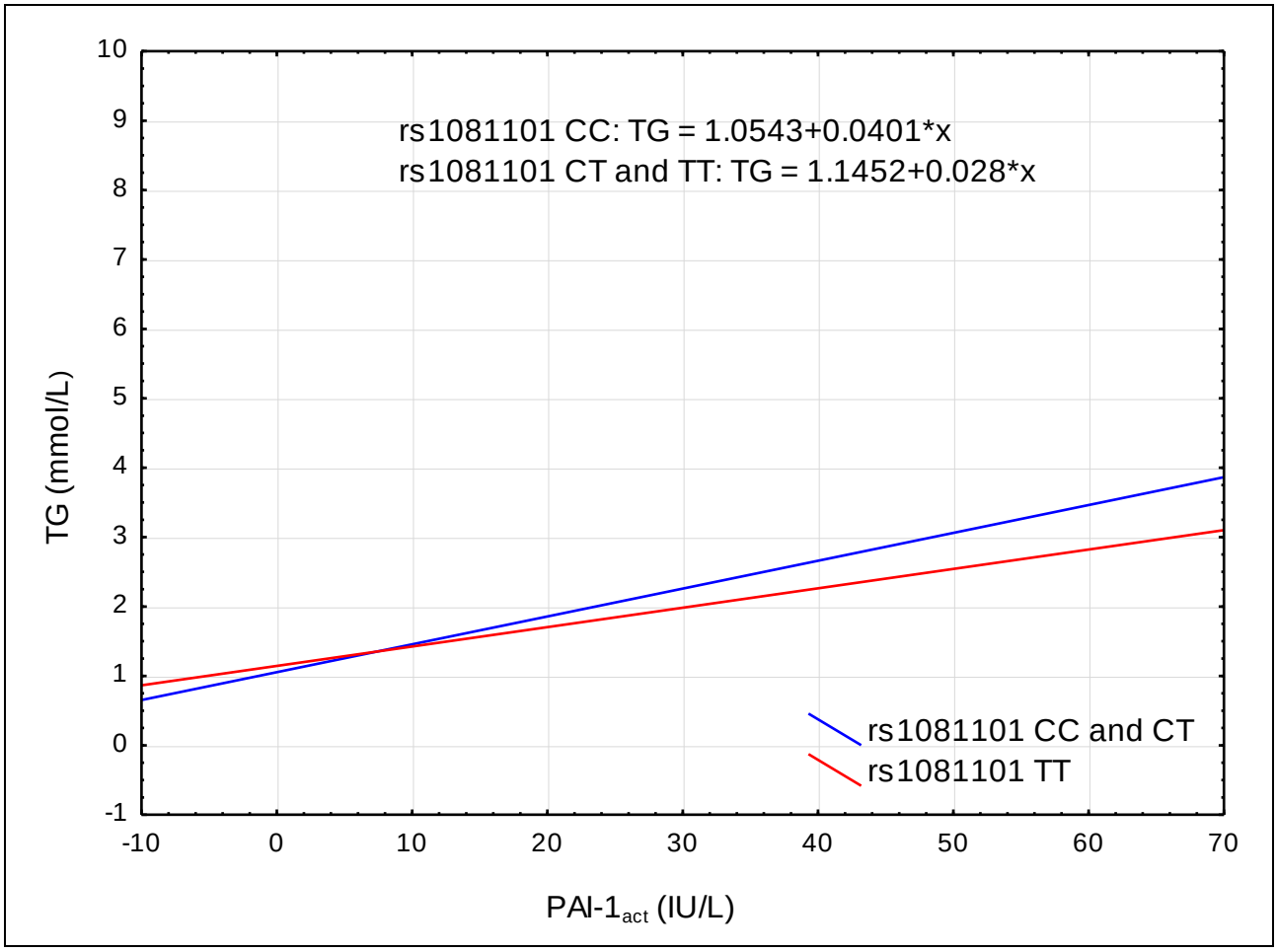
rs1081101 (C/T)						
Lipid profile	**Sub-groups**	**CC**	**CT**	**TT**	**p-value[$]**	**p-value***
[TC]	Whole group	5.02	5.03	4.86	0.88	--
	Male	4.79	4.74	4.95	0.87	0.69
	Female	5.16	5.16	4.80	0.68	
	Rural	4.97	5.02	4.97	0.88	0.65
	Urban	5.07	5.07	4.78	0.72	
[LDL-C]	Whole group	2.93	2.94	2.59	0.41	--
	Male	2.71	2.61	2.49	0.66	0.69
	Female	3.06	3.08	2.66	0.47	
	Rural	2.92	2.96	2.81	0.89	0.69
	Urban	2.94	2.91	2.41	0.30	
[HDL-C]	Whole group	1.53	1.52	1.65	0.69	--
	Male	1.57	1.62	1.95	0.16	0.20
	Female	1.51	1.48	1.42	0.75	
	Rural	1.52	1.52	1.71	0.65	0.65
	Urban	1.54	1.51	1.59	0.87	
[TG]	Whole group	1.25	1.28	1.40	0.67	--
	Male	1.13	1.15	1.14	0.84	0.39
	Female	1.31	1.34	1.62	0.35	
	Rural	1.18	1.20	0.99	0.62	0.26
	Urban	1.31	1.37	1.73	0.32	

[$] = indicates the p-values of the association of the genotypes at the rs1081101 locus and each of the lipid profile determinants for the total population as well as by gender and locality groups; * = indicates the p-values of the interaction effect between rs1081101 and gender/rural and urban group; C = cytosine; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; rs = reference sequence; T = thymine; TC = total cholesterol; TG = triglycerides.

Figure 5.22 is a graphical representation of the significant interaction between PAI-1_{act} and the rs1081101 genetic variant ($p = 0.01$) on TG concentrations. Since the frequency of the homozygous mutant genotype was low, the heterozygote and homozygote mutant groups were combined.

In univariate analysis, no significant association was observed between the SNP single-handedly ($p = 0.22$) and TG concentrations, but a significant positive correlation was observed between PAI-1_{act} and TG concentrations ($p = 0.00$) in the present cohort (Figure 5.22). The latter is similar to that reported for Caucasians (Krebs *et al*., 2003), but conflicting to that reported for South Africans (Greyling *et al*., 2007) and African-Americans (Solano *et al*., 2003). When harbouring the C-allele, the association of TG with PAI-1_{act} was stronger than when harbouring the variant genotypes (CT and TT). However, after adjusting for multiple testing and unequal groups with the *post hoc* test this effect was no longer significant.

Figure 5.22 Scatterplot of PAI-1_{act} against triglyceride concentrations categorised by rs1081101



C = cytosine; IU/L = international units per litre; mmol/L = millimole per litre; rs = reference sequence; PAI-1_{act} = plasminogen activator inhibitor type 1 activity; T = thymine; TG = triglyceride concentrations

PAI-1_{act} had previously been associated with the blood lipid profile, which has been reported as causative of PAI-1 concentrations (Festa *et al*., 1999; Tremoli *et al*., 1993; Stiko-Rahm *et al*., 1990). Krebs *et al*. (2003) with other researchers (Asplund-Carlson *et al*., 1993; Stiko-Rahm *et al*., 1990; Mehta *et al*., 1987) have reported that one of the major determinants of plasma PAI-1_{act} concentrations are the TG concentrations of normo- and hypertriglyceridaemic patients.

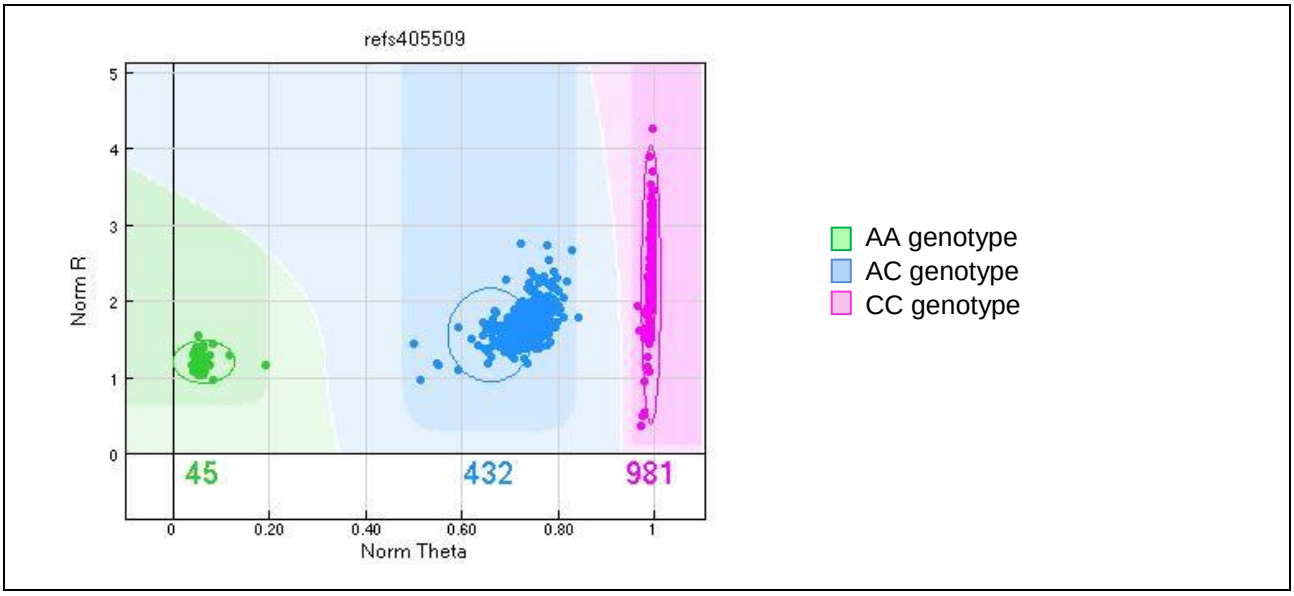
Stiko-Rahm *et al*. (1990) reported a possible explanation for the clinical association between serum TG and plasma PAI-1 concentrations. They determined that PAI-1 secretions were increased in cultured hepatocytes and endothelial cells, which may have been due to an intracellular signal that is generated when VLDL binds to its surface receptor in these cells. Krebs *et al*. (2003) added to this hypothesis and suggested that free fatty acids (FFA) may in fact be the mediator responsible for elevated PAI-1 concentrations, since FFA stimulate the production of PAI-1 in hepatocytes (Banfi *et al*., 1997).

Black South Africans have been reported to have a favourable lipid profile (Vorster, 2002), which has been suggested as a possible reason for the lower PAI-1 concentrations determined in this population (Pieters & Vorster, 2008). However, PAI-1 concentrations have been reported to increase with urbanisation in black South Africans (but not to the elevated concentrations observed in Caucasians), possibly owing to the accompanying adverse dietary changes that result in a deteriorating lipid profile (Pieters & Vorster, 2008; Greyling *et al*., 2007). Pieters & Vorster (2008) and Greyling *et al*, (2007) speculated that PAI-1 concentrations in black South Africans are strongly regulated by genetic factors, although it remains unproven at this time.

5.5.2 SNP rs405509

The call region shading for the rs405509 SNP, as determined by the GenomeStudio® program, is presented in Figure 5.23. The GenCall Score for this SNP was 0.59, indicating that the majority of the samples had reliable clustering calls and were not too far from the centre of the cluster. Furthermore, the values of the AB R mean (1.56) and AB T mean (0.66) of this particular SNP were optimal, which indicated that the intensity of the data was reliable and also that separate and definite clusters could be formed. This is confirmed by the high cluster separation value of 0.58.

Figure 5.23 Genomestudio® shade call regions for SNP rs405509



A = adenosine; C = cytosine

The ancestral allele at the rs405509 locus is cytosine (C). The MAF in this population was lower (0.18) than that reported for sub-Saharan Africans (0.21) and Europeans with a MAF of 0.51 (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=405509). The slight

MAF differences observed among the different African populations can be ascribed to the low occurrence of genetic drift, selection or migration among and between these African cohorts (Cummings, 2011). Furthermore, the MAF differences between the present cohort and Europeans can be ascribed to the out-of-Africa theory discussed in the previous section and also emphasises the differences in evolutionary factors that have affected the distribution of the alleles (Tishkoff *et al.,* 2009; Campbell & Tishkoff, 2008).

However, the sample size of the present black South African population was smaller than previous reported analyses for sub-Saharan Africans, which might give rise to the better estimation of the MAF for the sub-Saharan African population. It is, therefore, important to bear this is mind when interpreting MAF results and comparing the genotype frequencies between different populations. Based on the chi-square test conducted (Table 5.10), this SNP ($p = 0.76$; $\chi^2 = 0.09$ and d.f. = 2) adhered to the assumptions of HWE (Hardy, 1908). According to the 95% CI, all the genotype frequencies for SNP rs405509 were within the expected ranges (Table 5.10), indicating that the study population is representative of the larger population and, therefore, sampling error was unlikely.

Table 5.10 Adherence to the assumptions of Hardy-Weinberg Equilibrium of the rs405509 locus in the present PURE population using chi-square test analyses

rs405509 (A/C)					
Chi-square test for goodness of fit to the Hardy-Weinberg proportions					
Chi-square test	Genotypes	CC	AC	AA	p-value$^{\$}$ (χ^2)
	Observed numbers (O)	981	432	45	
	Frequency (%)	67.26	29.65	3.09	
	Expected numbers (E)	982.72	428.56	46.72	--
	$\chi^2 = (O-E)^2/E$	0.00	0.03	0.06	0.76 (0.09)
Genotype distribution	95% CI range (%)	65.85-69.67	27.30-32.00	2.20-3.98	--

% = percentage; χ^2= Chi-square value; A = adenosine; C = cytosine; CI = confidence intervals; E = expected numbers; O = observed numbers; rs = reference sequence

No significant associations were determined between the concentrations of the lipid profile determinants and the different genotype groups at the rs405509 locus for either the whole group or the gender and locality group (Table 5.11). These results are in accordance with previous studies, which reported that this polymorphism either did not modify plasma lipid or lipoprotein concentrations (Lambert *et al.,* 2000) or indicated a non-significant influence on TC or LDL-C concentrations (Viitanen *et al*., 2001). However, the C-allele has been reported to be associated with MI and premature CHD (Viitanen *et al*., 2001; Lambert *et al*., 2000) and furthermore, it has been associated with the incidence of various CVD risk biomarkers such as various apolipoproteins (Komurcu-Bayrak *et al*., 2011; Moreno

et al., 2003; Viitanen *et al*., 2001; Lambert *et al*., 2000), insulin (Komurcu-Bayrak *et al*., 2011; Viitanen *et al*., 2001) and waist:hip ratio (Viitanen *et al*., 2001). As no significant association has been reported between this polymorphism and the lipid profile, it suggests that the impact of the C-allele may be *via* an alternative mechanism, such as that of atherosclerotic plaque formation, as opposed to having a major effect on circulating lipids (Lambert *et al*., 2000).

Furthermore, these results could imply that the association of the *ApoE* gene with atherothrombosis is possibly due to the contributing role of more genetic and environmental factors than just this locus *per se* (Moreno *et al*., 2004). Furthermore, one of the objectives of this study was to explore whether possible interaction effects of environmental or physiological factors with specific *ApoE* genetic variations could influence the lipid profile in the present black South African cohort.

No significant gene-gender or gene-diet interaction effects were observed for SNP rs405509 with any of the lipid profile determinant concentrations investigated. However, a significant rs405509-locality interaction effect on TG concentrations was observed (Table 5.11), even though the rs405509 SNP itself did not have a significant effect on the TG concentrations ($p = 0.73$) although it was determined that locality (rural/urban grouping) did have an effect ($p = 0.02$; see Table 5.11). This interaction still occured after adjusting for the respective confounding variables (Section 5.5; Table 5.7).

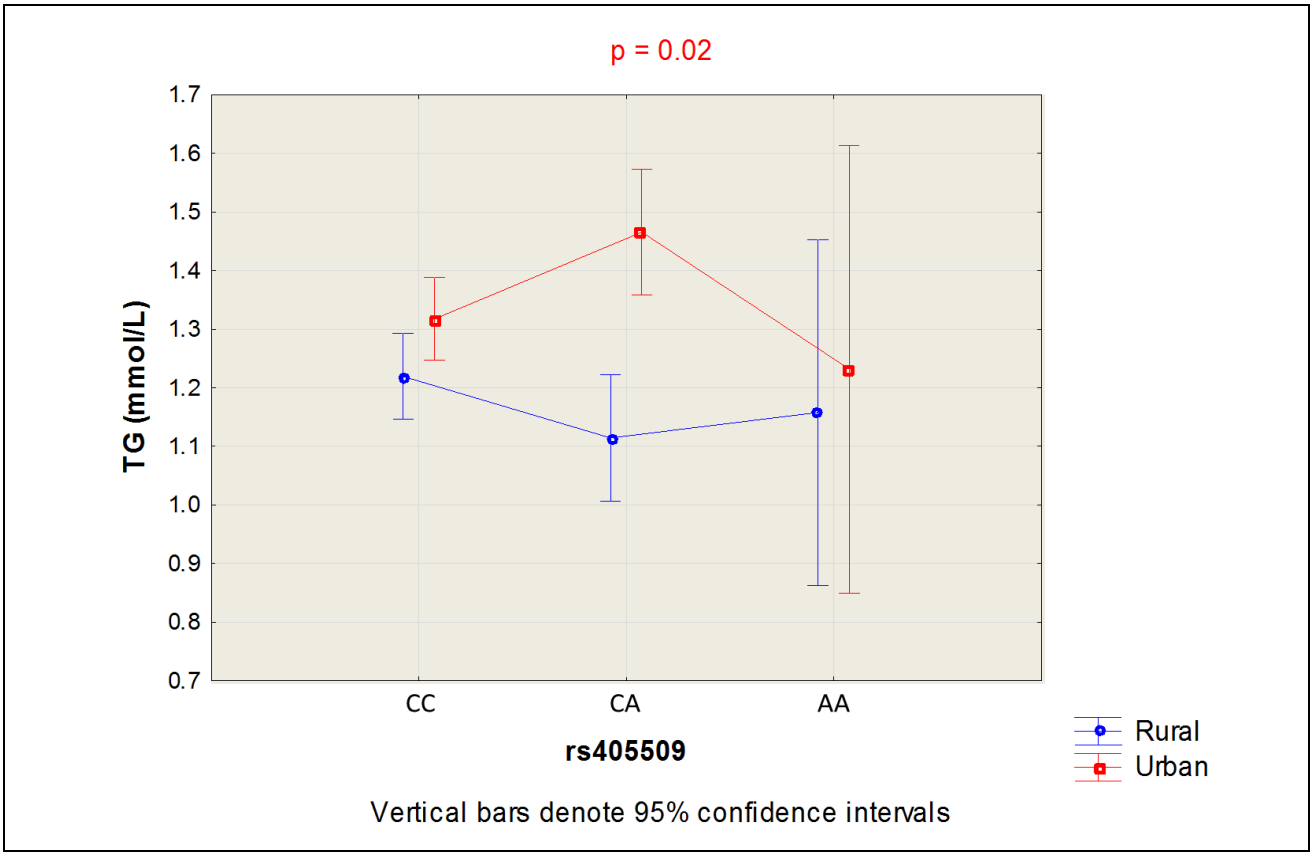
Table 5.11 Associations of the lipid profile determinant concentrations with the rs405509 locus in the present PURE population

rs405509 (G/T)						
Lipid profile	**Stratification**	**CC**	**CA**	**AA**	**p-value$^{\$}$**	**p-value***
[TC]	Whole group	4.99	5.10	4.93	0.37	--
	Male	4.75	4.88	4.53	0.42	0.84
	Female	5.12	5.23	5.19	0.63	
	Rural	4.97	4.97	5.13	0.82	0.26
	Urban	5.01	5.22	4.63	0.09	
[LDL-C]	Whole group	2.89	3.01	2.94	0.21	--
	Male	2.63	2.80	2.66	0.31	0.87
	Female	3.03	3.13	3.11	0.50	
	Rural	2.91	2.95	3.06	0.75	0.46
	Urban	2.87	3.07	2.74	0.10	
[HDL-C]	Whole group	1.54	1.53	1.45	0.66	--
	Male	1.60	1.57	1.38	0.36	0.70
	Female	1.50	1.50	1.49	1.00	
	Rural	1.53	1.51	1.54	0.95	0.61
	Urban	1.55	1.54	1.30	0.32	
[TG]	Whole group	1.26	1.25	1.22	0.54	--
	Male	1.15	1.12	1.09	0.85	0.70
	Female	1.32	1.33	1.30	0.39	
	Rural	1.20	1.14	1.16	0.32	0.02
	Urban	1.32	1.36	1.31	0.11	

$^{\$}$ = indicates the p-values of the association rs405509 and each of the lipid profile determinants for the total population as well as by gender and locality groups; * = indicates the p-values of the interaction effect between rs405509 and gender/rural and urban; A = adenosine; C = cytosine; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; rs = reference sequence; TC = total cholesterol; TG = triglycerides

Furthermore, unequal N *post hoc* tests were conducted to determine where the significant difference between the genotype-environment interaction effects on TG occurred. These analyses revealed that individuals harbouring the heterozygous genotype (CA) in an urbanised setting had significantly higher TG concentrations than the individuals harbouring the same heterozygous genotype in a rural setting ($p = 0.00$), as well as the individuals harbouring the wild type genotype (CC) in a rural setting ($p = 0.00$) (Figure 5.24). However, these elevated concentrations are still within the desirable ranges for TG (≤ 1.5 mmol.L^{-1}). These observations suggest that the reported adverse effect of urbanisation on the lipid profile (Section 5.1), is increased in individuals harbouring the heterozygote genotype (Table 5.11).

Figure 5.24 Significant genotype-environment interaction effects on TG concentrations for the three genotype groups of rs405509



A = adenosine; C = cytosine; mmol/L = millimole per litre; rs = reference sequence; TG = triglyceride concentrations

One possible explanation for these observations might be that urbanised subjects harbouring the wild type genotype (CC) have the ability to metabolise TG better than those harbouring the heterozygote genotype (CA), which might result in the significantly increased TG concentrations among the latter group. Another possible reason can possibly be that the heterozygote genotype (CA) has an effect on certain structural aspects of the ApoE protein/isoform. This could result in the isoform being unable to form disulphide bonds with other proteins, which would in turn result in this isoform mobilising CMs faster than the other isoforms (which are the proposed products of the homozygote wild type and mutant genotypes, respectively) and, therefore, maintaining low TG concentrations (Davignon *et al*., 1988; Weisgraber *et al*., 1982). Furthermore, it can be hypothesised that dietary lipids enhance the affinity of ApoE isoforms for receptors (Fisher *et al*., 2004). Since the TG concentrations are higher among the urban group for all three genotypes, it can be hypothesised that these elevated concentrations trigger conformational changes (similar to those mentioned above) in the ApoE protein products of the CC and AA genotype carriers, which results in higher binding affinities of these isoforms to specific hepatic receptors, resulting in reduced TG concentrations among the CC and AA genotype carriers. On an environmental level, the increased TG

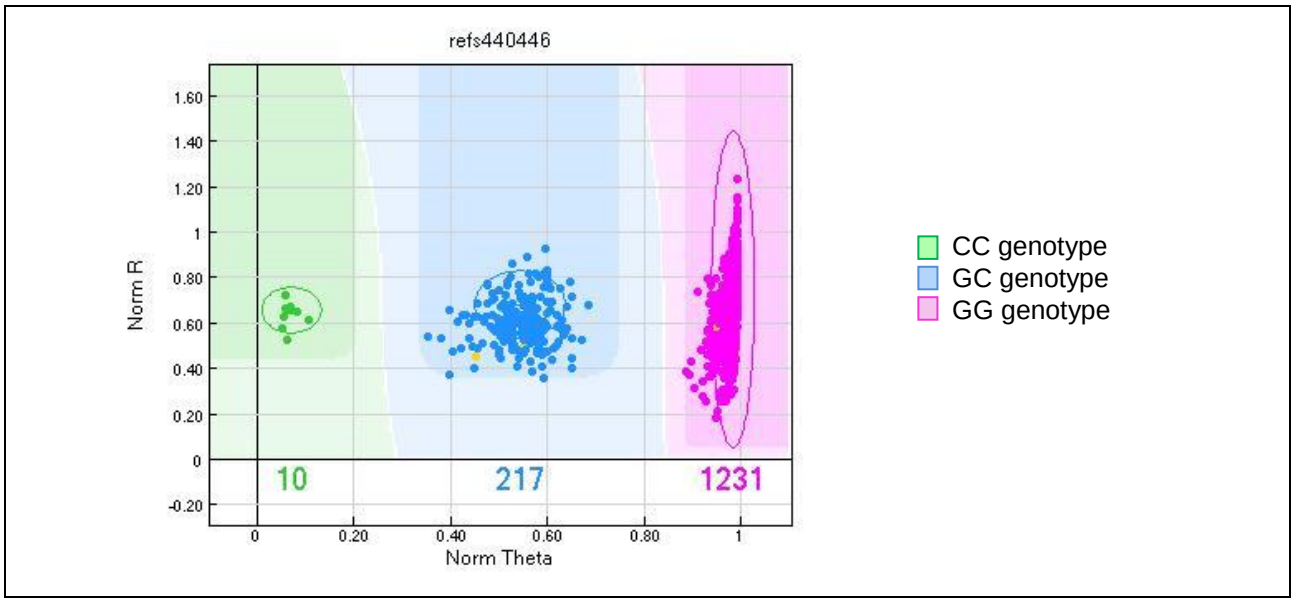
concentrations can be ascribed to the atherogenic diet associated with urbanisation (Vorster, 2002).

However, as mentioned in the previous section, these results should be interpreted with caution, as they might simply be a result of multiple testing bias or reporting a type 1 error due to the prevalence of outliers within an already small sample size. Therefore, further investigations are necessary to clarify not only the mechanistic role this interaction might have, but also whether this interaction will persist in a sample size of greater power.

5.5.3 SNP rs440446

The GenomeStudio® program was used to generate the specific call region shading for rs440446 that is illustrated in Figure 5.25. The GenCall Score for this SNP was within the desirable range (0.85), but all samples that were too far from the centre of the cluster consequently had insufficient GenCall scores (< 0.5) and were excluded. Furthermore, the value of the AB R mean was 0.68, which indicated that the intensity of the data was reliable. The AB T mean value was 0.54 and the cluster separation value was 0.61, which indicated that separate and definite clusters could be formed (Figure 5.25).

Figure 5.25 Genomestudio® shade call regions for SNP rs440446



C = cytosine; G = guanine

The ancestral allele for rs440446 was determined to be the G-allele and the C-allele was the variant allele. The MAF in the population (0.08) investigated was lower than the MAF observed in other ethnic groups *i.e.* Caucasians (0.32) and African-Americans (0.17). However, the MAF among sub-Saharan African and Bantu populations was 0.00

(http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=440446). These MAF differences among the present black South African population and the populations outside Africa emphasise the effect of the well-known out-of-Africa theory, as discussed previously. Furthermore, a conclusion that can be drawn from the Hardy-Weinberg law is that the larger the MAF difference between two populations is, the more factors such as: genetic drift, migration and selection are operating within these populations. However, the slight MAF differences observed among the different African populations can be ascribed to the low occurrence of genetic drift, selection or migration among and between these African cohorts (Cummings, 2011).

A chi-square test was used to determine whether the population adhered to the assumptions of HWE (Table 5.12). The insignificant p-value (0.90) of the chi-square test indicated that at the rs440446 locus this population did adhere to these assumptions and the genotypic distribution was therefore not affected by any of the HWE-altering factors, as discussed in Section 4.4.3 (Hardy, 1908). Furthermore, the prevalence of the genotypes at the rs440446 SNP were all within the 95% CI ranges reported in Table 5.12. Therefore, this sample population was sufficiently representative of the larger population, because of the narrow ranges determined, which provides confidence in the results from the samples and highlights that no sampling bias occurred.

Table 5.12 Adherence to the assumptions of Hardy-Weinberg Equilibrium of the rs440446 locus in the present PURE population using chi-square test analyses

rs440446 (G/C)					
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions					
Chi-square test	Genotype	GG	GC	CC	p-value[$] ($\chi^2$)
	Observed numbers (O)	1231	217	10	--
	Frequency (%)	84.42	14.89	0.69	
	Expected numbers (E)	1230.63	217.74	9.63	
	$\chi^2 = (O-E)^2/E$	0.00	0.00	0.01	0.90 (0.02)
Genotype distribution	95% CI range (%)	82.55-86.28	13.06-16.72	0.26-1.11	--

% = percentage; χ^2= Chi-square value; C = cytosine; CI = confidence intervals; E = expected numbers; O = observed numbers; rs = reference sequence; T = thymine

No significant associations were observed between any of the lipid profile determinants, (TC, LDL-C, HDL-C and TG) and the different genotype groups of rs440446 for the whole group or the different gender or locality groups (Table 5.13). This indicates that, despite stratification, this SNP is unlikely to have an effect on the blood lipid profile.

Table 5.13 Associations of the lipid profile determinant concentrations with the rs440446 locus in the present PURE population

rs440446 (G/C)						
Lipid profile	**Stratification**	**GG**	**GC**	**CC**	**p-value[$]**	**p-value***
[TC]	Whole group	5.02	5.01	5.06	0.97	--
	Male	4.78	4.79	4.72	0.99	0.82
	Female	5.16	5.12	5.75	0.68	
	Rural	4.97	5.02	5.01	0.90	0.64
	Urban	5.08	4.99	5.24	0.71	
[LDL-C]	Whole group	2.92	2.93	2.88	0.99	--
	Male	2.69	2.69	2.55	0.96	0.89
	Female	3.06	3.05	3.55	0.76	
	Rural	2.92	2.94	2.87	0.99	0.95
	Urban	2.93	2.92	2.92	1.00	
[HDL-C]	Whole group	1.53	1.51	1.62	0.77	--
	Male	1.58	1.60	1.69	0.91	0.91
	Female	1.51	1.47	1.47	0.77	
	Rural	1.52	1.57	1.69	0.55	0.19
	Urban	1.55	1.46	1.39	0.28	
[TG]	Whole group	1.26	1.26	1.26	0.49	--
	Male	1.14	1.11	1.07	0.91	0.55
	Female	1.32	1.33	1.63	0.26	
	Rural	1.19	1.15	1.02	0.53	0.07
	Urban	1.33	1.35	2.09	0.10	

[$] = indicates the p-values of the association rs440446 and each of the lipid profile determinants for the total population as well as by gender and locality groups; * = indicates the p-values of the interaction effect between rs440446 and gender/rural and urban; C = cytosine; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; rs = reference sequence; T = thymine; TC = total cholesterol; TG = triglycerides

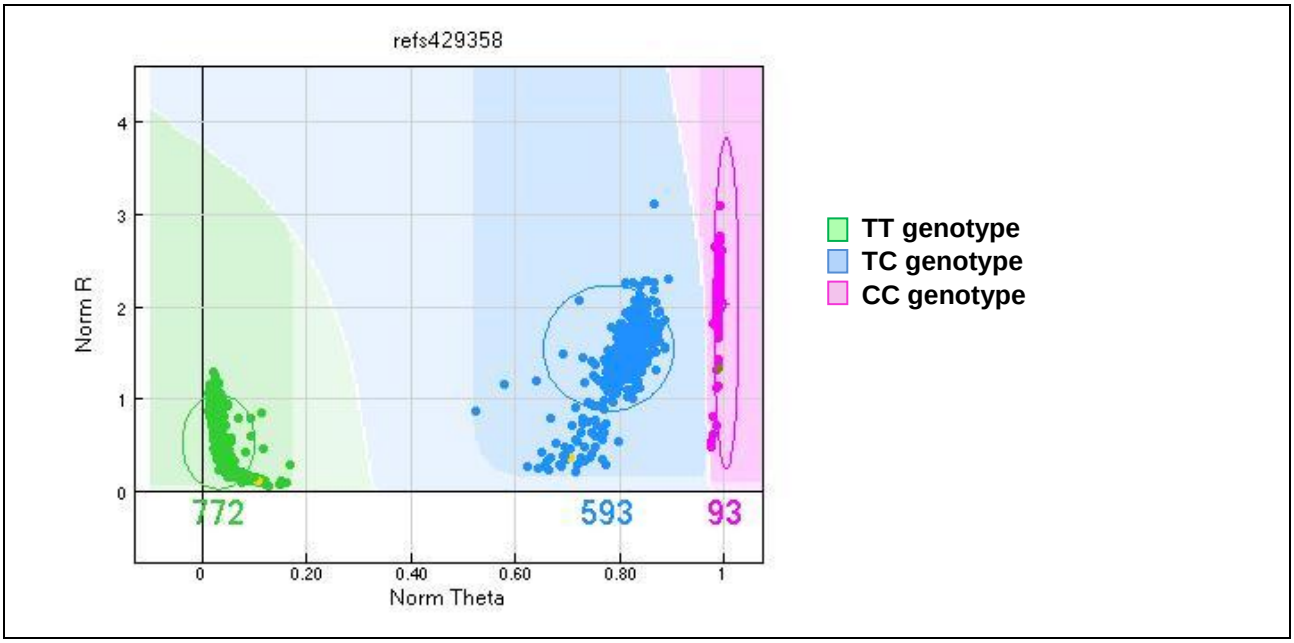
Previous studies have reported that the promoter polymorphisms (which include rs440446) have an effect on the transcriptional activity of the *ApoE* gene and that the C-allele possesses transcription-enhancing capabilities (Viiri *et al.,* 2008). This could still be true for the present cohort; however, since the ApoE concentrations were not measured in the present study, the possible effect the rs440446 SNP could have on the transcriptional activity of *ApoE* gene in the present cohort group could not be determined.

Silander and co-workers (2008) reported a strong association ($p \leq 0.001$) between the rs440446 variant and TG in Finnish men and women analysed together; however, such significant results were not observed in the present black South African population. Furthermore, no significant gene-gender, gene-environment (rural/urban) or gene-diet interaction effects with any of the lipid profile determinants were observed within the present cohort (Table 5.13).

5.5.4 SNP rs429358

Figure 5.26 represents the call region shading of the rs429358 SNP as determined by the GenomeStudio® program. The overall GenCall Score for this SNP was 0.54. A large number of samples were too far from the centre of the cluster, indicating an unreliable clustering call. This was confirmed by their GenCall score, which was too low (< 0.5). The causes of the low GenCall score could possibly be ascribed to the quality of the specific location of the SNP within the gene. However, no general interpretation guidelines for GenCall scores exist, since it depends on the clustering and intensity of the samples. The intensity of the data was within standard ranges which indicates that it is reliable (AB R mean of 1.61), along with the fact that separate and definite clusters could be drawn (AB T mean of 0.71; Cluster Sep of 0.50). Furthermore, as mentioned in Section 5.2.3.5, the genotypic data for this SNP (along with those for the rs7412 SNP) are needed in order to be able to determine the well-reported *ApoE* haplotypic data (ε2, ε3, ε4). Therefore, based on the aforementioned reasons, the genotypic data for this SNP were still used to conduct statistical analyses.

Figure 5.26 Genomestudio® shade call regions for SNP rs429358



C = cytosine; T= thymine

The C allele is the ancestral (minor) allele and the T allele is the variant allele for rs429358 (Figure 5.26; Table 5.14). The MAF in the present cohort was calculated as 0.27. This is higher than that reported for sub-Saharan Africans (0.02) as well as Caucasians with a MAF of 0.15 (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=429358). The higher MAF among the black South African population, compared to the sub-Saharan

African population, could be explained by the different population sample sizes *i.e*. the current MAF is based on 120 sub-Saharan Africans *vs* 1 458 black South African participants from the present study. Factors that influence genotype distribution (*e.g*. genetic drift, natural selection and migration) differ between population groups and this could explain the different MAF observed between the African and Caucasian groups.

The results of the chi-square test for SNP rs429358 was: p = 0.14 for a chi-square value of 2.19 and 2 d.f. (Table 5.14). Therefore, this SNP was in HWE and adhered to the assumptions of HWE (Hardy, 1908). Furthermore, no sampling errors occurred, as the genotype frequencies for the present study population were all within the 95% CI ranges based on the frequencies expected in the total population (Table 5.14).

Table 5.14 Adherence to the assumptions of Hardy-Weinberg Equilibrium of the rs429358 locus within the present PURE population using chi-square test analyses

rs429358 (T/C)					
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions					
Chi-square test	Genotype	TT	TC	CC	p-value (χ^2)
	Observed numbers (O)	772	593	93	--
	Frequency (%)	52.99	40.63	6.38	
	Expected numbers (E)	783.05	570.89	104.05	
	$\chi^2 = (O-E)^2/E$	0.16	0.86	1.17	0.14 (2.19)
Genotype distribution	95% CI range (%)	50.42-55.55	38.11-43.16	5.13-7.64	--

% = percentage; χ^2= Chi-square value; C = cytosine; CI = confidence intervals; E = expected; O = observed numbers; rs = reference sequence; T = thymine

A number of significant associations (p < 0.05) have been reported between rs429358 and other CVD risk biomarkers before, such as the ApoB/A1 ratio (Anand *et al.,* 2009); CRP (Silander *et al.,* 2008) and ApoB concentrations (Paré *et al.,* 2007). However, as the focus of the present study was on the lipid profile, these previous associations were not investigated. Chang *et al*. (2010) investigated the genetic variants previously associated with fasting blood lipids in an American population. Different ethnic groups were included in the study *i.e*. non-Hispanic Caucasians, non-Hispanic blacks and Mexican-Americans. The results reported that rs429358 associated significantly (p < 0.05) with LDL-C and TC in the Mexican Americans (Chang *et al.,* 2010). The present study observed the same results for LDL-C in a black South African population.

An increase in TC and LDL-C concentrations, for the whole group, occurred with each addition of a C-allele, resulting in the highest concentrations in those harbouring the CC

genotype; however, these observations were only significant for LDL-C. Furthermore, a significant gender difference was observed with elevated LDL-C concentrations among females which were not present in males (Table 5.15). However, after adjusting for the confounding variables (Section 5.5, Table 5.7), this effect was lost. Furthermore, the *post hoc* unequal N test did not generate any significant results and it was therefore not possible to determine the genotype groups between which the LDL-C differences occurred. No other significant associations were determined between the concentrations of the other lipid profile determinants and the different genotype groups of rs429458 for either the whole group or the gender/locality groups (Table 5.15).

Table 5.15 Associations of the lipid profile determinant concentrations with rs429358 locus in the present PURE population

rs429358 (C/T)						
Lipid profile	**Stratification**	**TT**	**TC**	**CC**	**p-value$^{\$}$**	**p-value***
[TC]	Whole group	4.97	5.07	5.15	0.21	--
	Male	4.76	4.80	4.90	0.78	0.72
	Female	5.08	5.24	5.31	0.17	
	Rural	4.97	4.97	5.12	0.70	0.26
	Urban	4.96	5.18	5.18	0.08	
[LDL-C]	Whole group	2.84	2.99	3.17	0.01	--
	Male	2.62	2.72	2.98	0.19	0.74
	Female	2.96	3.16	3.29	0.02	
	Rural	2.85	2.96	3.20	0.11	0.77
	Urban	2.83	3.02	3.14	0.05	
[HDL-C]	Whole group	1.56	1.50	1.46	0.15	--
	Male	1.64	1.55	1.42	0.14	0.39
	Female	1.52	1.47	1.49	0.50	
	Rural	1.58	1.48	1.42	0.06	0.37
	Urban	1.55	1.53	1.52	0.92	
[TG]	Whole group	1.25	1.28	1.14	0.16	--
	Male	1.13	1.15	1.11	0.70	0.82
	Female	1.32	1.35	1.16	0.18	
	Rural	1.20	1.16	1.13	0.31	0.05
	Urban	1.31	1.40	1.15	0.07	

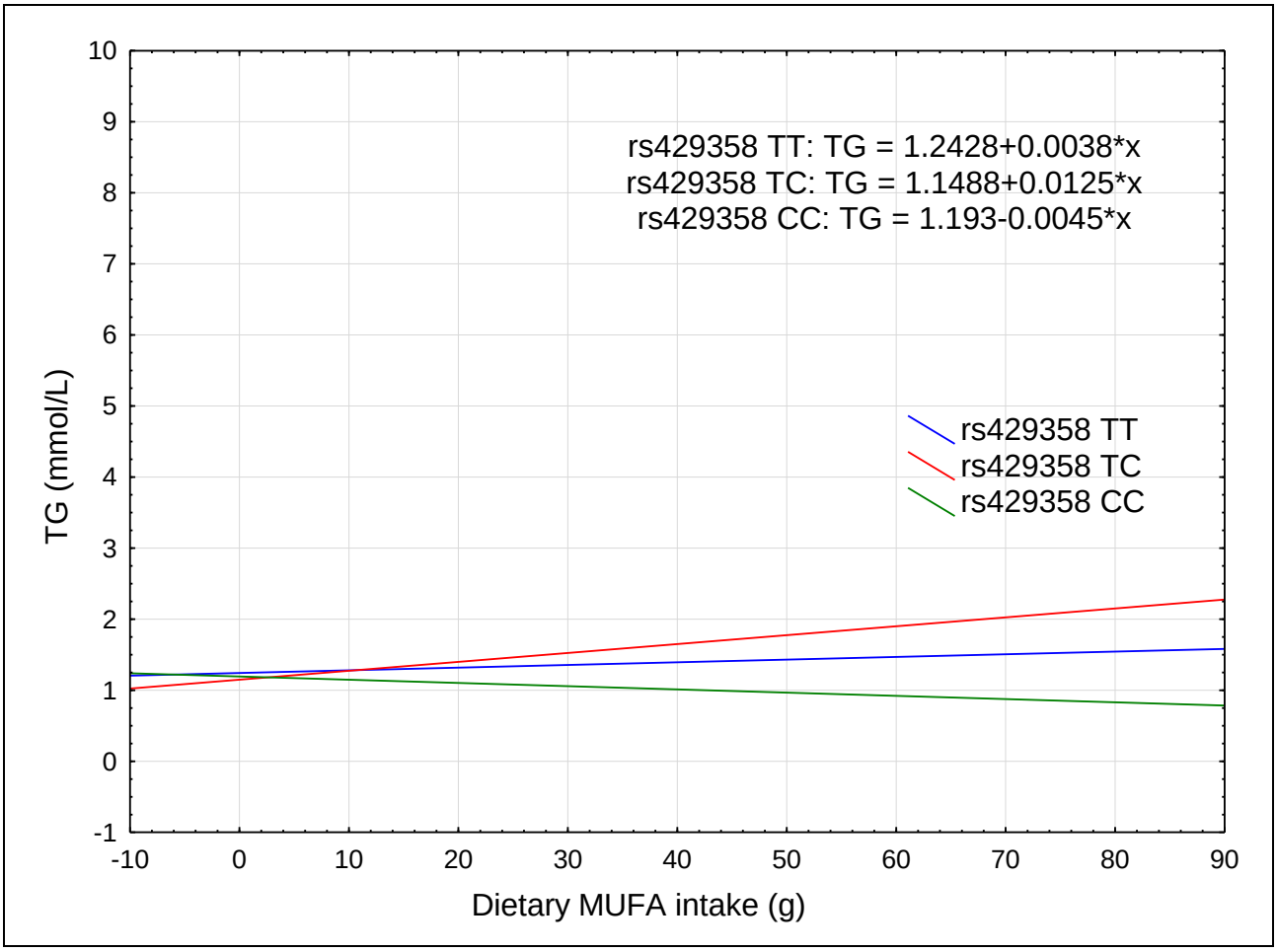
$^{\$}$ = indicates the p-values of the association between rs429358 and each of the lipid profile determinants for the total population as well as by gender and locality groups; * = indicates the p-values of the interaction effect between rs429358 and gender/rural and urban; C = cytosine; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; rs = reference sequence; T = thymine; TC = total cholesterol; TG = triglycerides

No significant interaction effect was observed between rs429358 and gender or level of urbanisation, for TC, LDL-C or HDL-C (Table 5.15). However, a weak rs429358-locality interaction (p = 0.05) was observed for TG, with the TG concentrations seemingly higher

among urbanised individuals carrying the TC genotype. However, the unequal N *post hoc* tests were not significant, indicating that the interaction was likely to be a type 1 error. Furthermore, the slight significance of this interaction was also lost when the respective confounding variables were adjusted for (Section 5.5; Table 5.7).

According to the literature, an inverse relationship exists between dietary MUFA intake and TG concentrations (Kris-Etherton, 1999), whereas positive correlations exist between dietary TFA and SFA intake and TG concentrations (Mozaffarian & Willett, 2009). Even though these relationships were observed in the present study, none was significant (Section 5.5; Table 5.6). Significant gene-diet interactions were observed for the rs429358 SNP. According to Table 5.15, lower TG concentrations were observed among those harbouring the CC genotype, but this was non-significant ($p = 0.16$) compared to those harbouring other genotypes. However, a significant gene-diet interaction with reference to the dietary MUFA intake was observed ($p = 0.04$), so that individuals harbouring the CC genotype had decreasing TG concentrations with increasing MUFA intake (Figure 5.27). A positive relationship between MUFA intake and TG concentrations was established for those individuals carrying the T- allele.

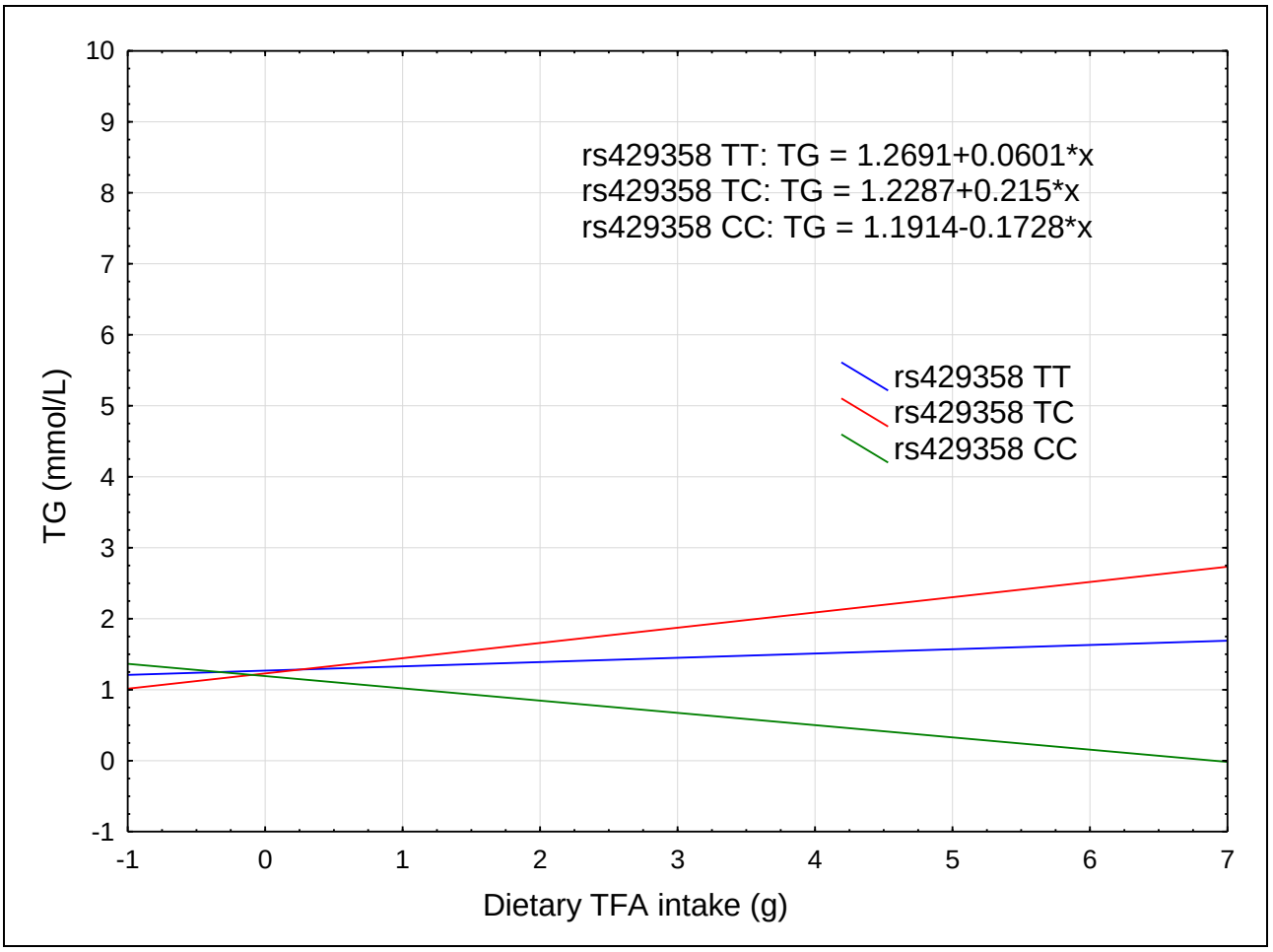
Figure 5.27 The interactive effect of dietary MUFA intake with rs429358 on triglyceride concentrations for each of the genotype groups



C = cytosine; g = grams; mmol/L = millimole per litre; MUFA = mono-unsaturated fatty acids; rs = reference sequence; T = thymine; TG = triglycerides

Furthermore, the significant interactions between dietary TFA ($p = 0.04$) and SFA ($p = 0.03$) with the SNP rs429358 resulted in inverse TFA:TG and SFA:TG relationships, specifically among individuals harbouring the CC genotype (Figure 5.28). However, separately neither the SNP ($p = 0.63$) nor the dietary TFA intake ($p = 0.61$) was significantly correlated with TG concentrations, according to the univariate analysis (Figure 5.28).

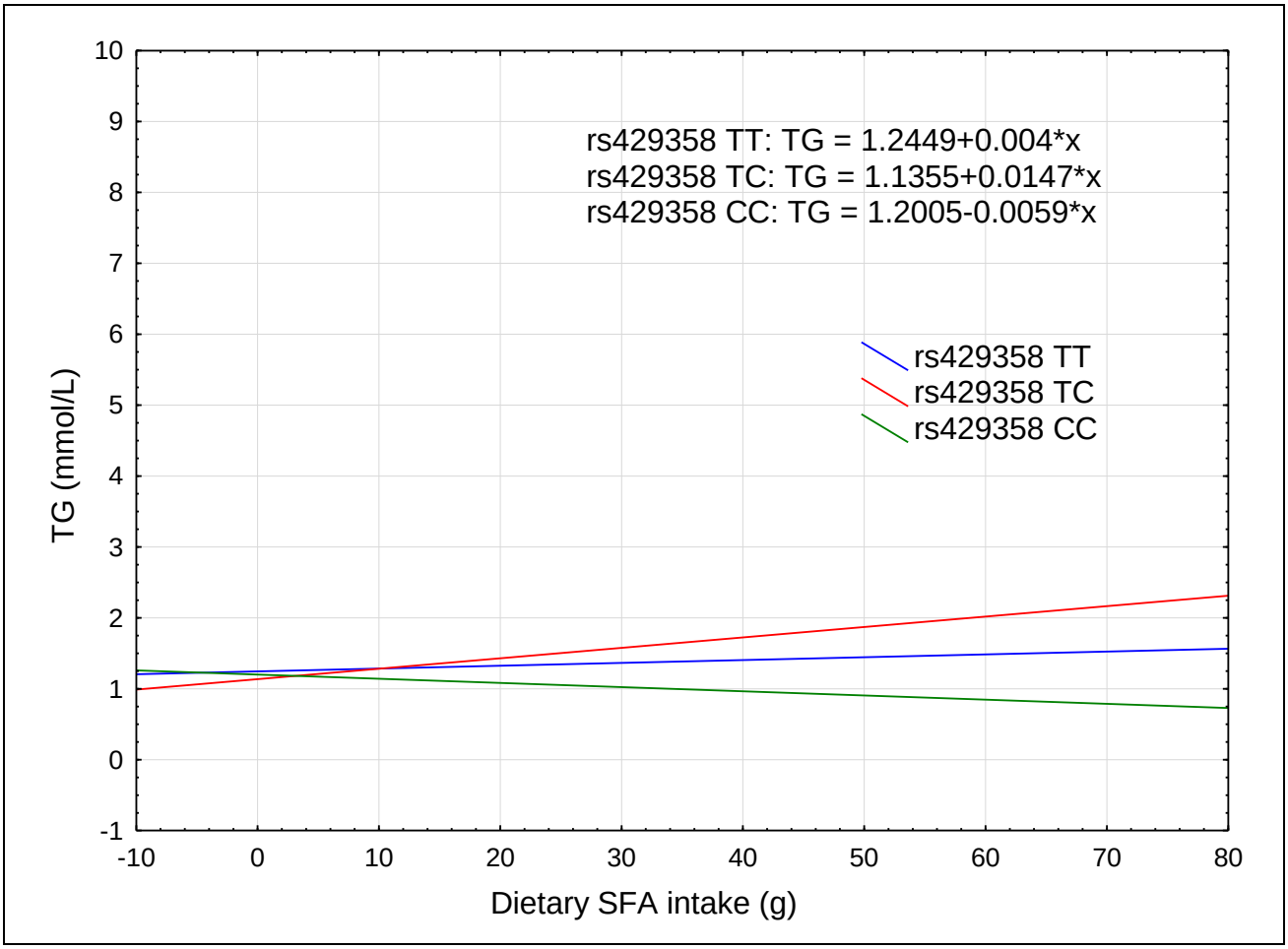
Figure 5.28 The interactive effect of dietary TFA intake with rs429358 on triglyceride concentrations for each of the genotype groups



C = cytosine; g = grams; mmol/L = millimole per litre; rs = reference sequence; T = thymine; TFA = trans fatty acids; TG = triglycerides

Similar results were reported for the univariate analysis conducted for the rs429358-SFA interaction, where the SNP itself (p = 0.34) and the dietary SFA intake (p = 0.20) also did not significantly correlate with TG concentrations (Figure 5.29). These results suggest that these interactive gene-diet effects possibly have a mechanistic role in the metabolism of TG concentrations and further studies should be conducted to investigate this. However, the unequal N *post hoc* tests were not significant for any of these three gene-diet interactions and therefore the differences observed in Figures 5.28-5.30 cannot be ascribed to a specific genotype (CC) group.

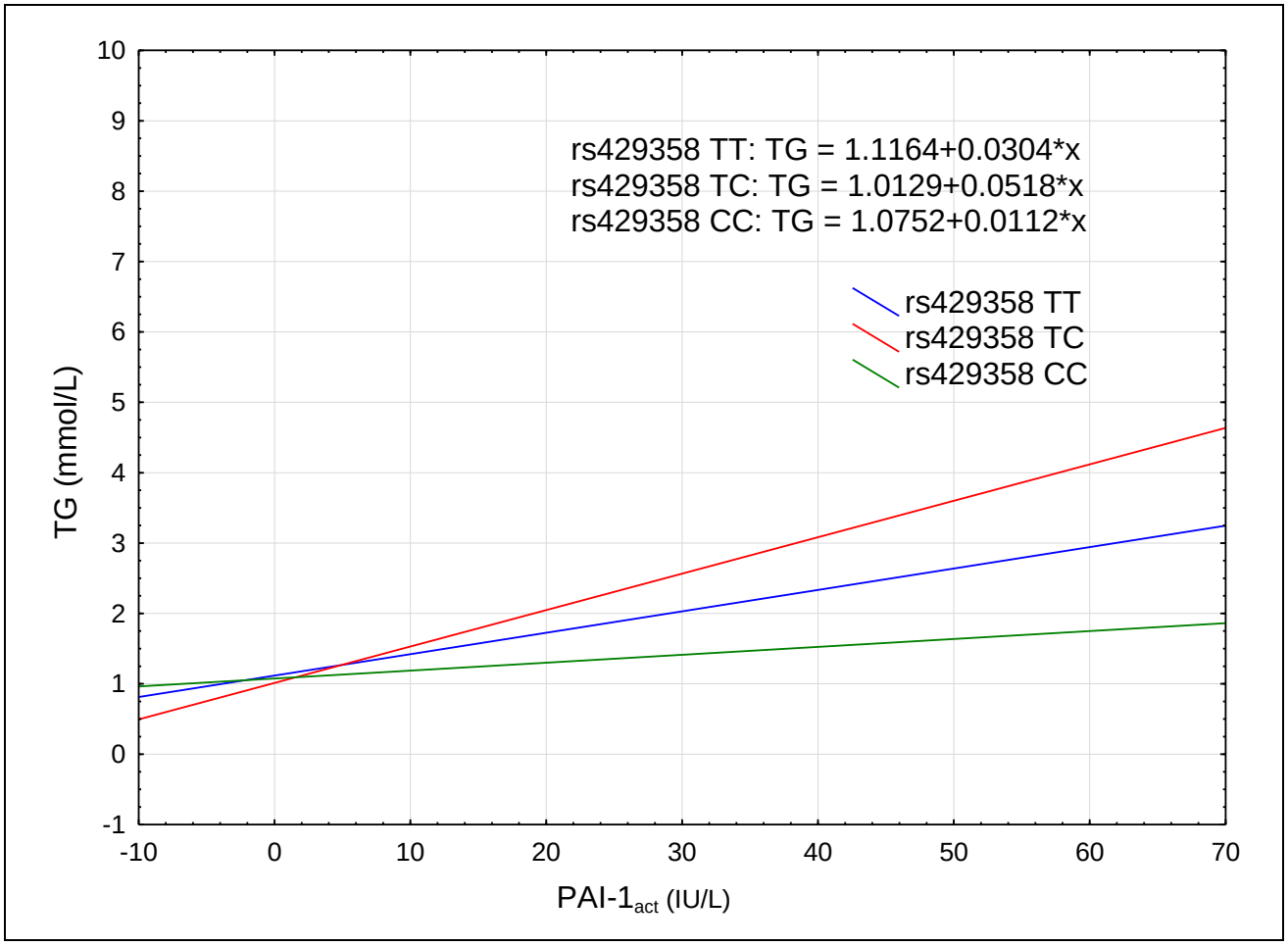
Figure 5.29 The interactive effect of dietary SFA intake with rs429358 on triglyceride concentrations for each of the genotype groups



C = cytosine; g = grams; mmol.L^{-1} = millimole per litre; rs = reference sequence; SFA = saturated fatty acids; T = thymine; TG = triglycerides

Lastly, a significant genotype-phenotype interaction effect between PAI-1$_{act}$ and the rs429358 SNP on TG was observed ($p = 0.00$), but no significant difference could be ascribed between the different genotypes, as the unequal N *post hoc* test was not statistically significant (Figure 5.30). The univariate test for significance indicated that separately the SNP ($p = 0.13$) was not significantly correlated with TG concentrations, but PAI-1$_{act}$ ($p = 0.00$) was. This phenomenon observed for SNP rs429358 is similar to that observed in Section 5.5.1 for rs1081101 and this interactive effect can therefore be added to the list of genetics-related factors that may be causative of the increased PAI-1 concentrations among black South Africans.

Figure 5.30 The interactive effect of PAI-1_{act} concentrations with rs429358 on triglyceride concentrations for each of the genotype groups



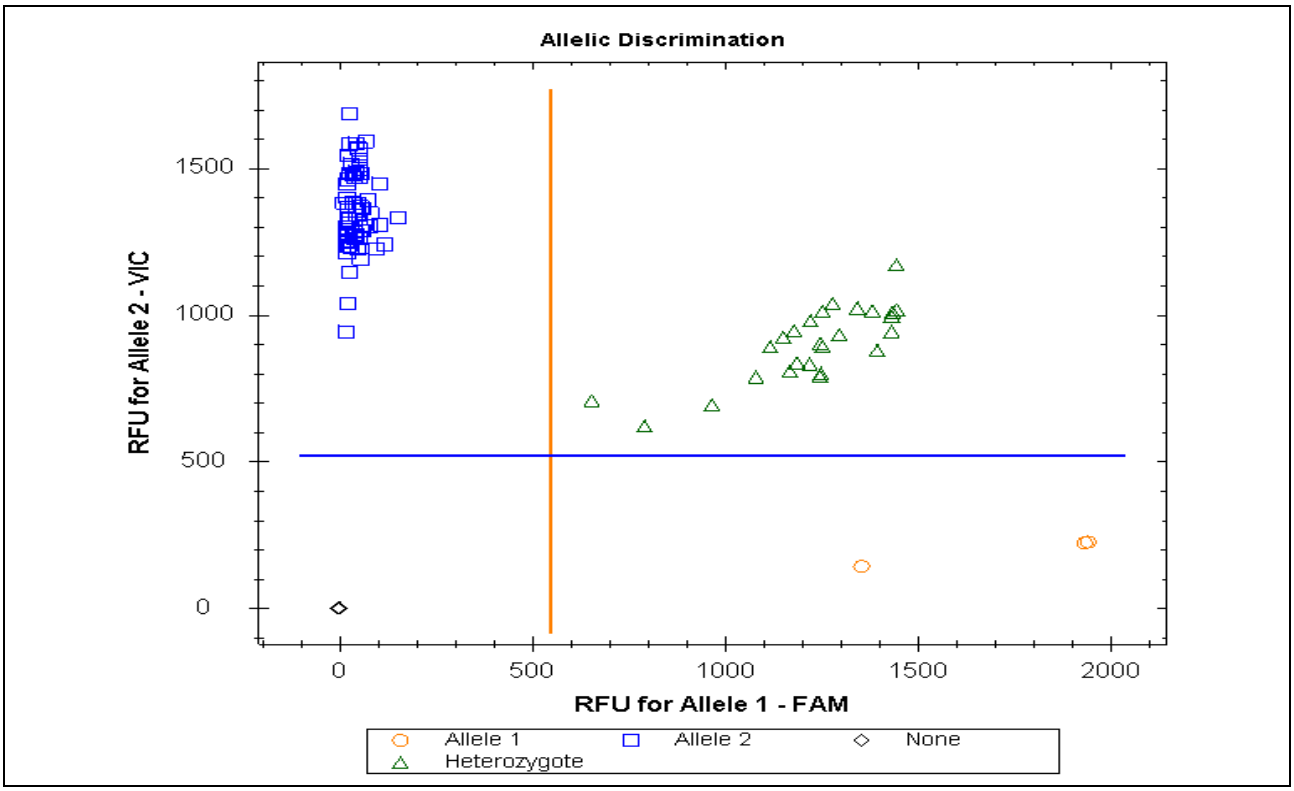
C = cytosine; g = grams; $IU.L^{-1}$ = international units per litre; mmol.L-1 = millimole per litre; PAI-1_{act} = plasminogen activator inhibitor type 1 activity; rs = reference sequence; T = thymine; TG = triglycerides

5.5.5 SNP rs7412

Data output for this SNP differed from the other SNPs, as a different method was applied to conduct the genotypic analysis of SNP rs7412, as discussed in Section 5.4. The allelic discrimination (AD) window generated for the current cohorts were all similar to Figure 5.31 (conducted on a Bio-Rad CFX96 qPCR apparatus). The homozygous alleles located in the top left quadrant represent the CC genotypes, as the VIC dye is attached to the C-TaqMan® allele probe. The homozygous alleles located in the bottom right quadrants represent the TT genotypes, since the FAM dye is attached to the T-TaqMan® allele probe. The heterozygotes (CT) are located in the upper right quadrant. The NTCs or any samples with no DNA amplification are located in the lower left quadrant. Normalised relative fluorescence units (NRFU) for each allele (dye) were plotted against the opposite allele, as presented in Figure 5.31. Furthermore, normalisation is accomplished relative to the NTC wells. Figure 5.31 is a graphical representation of the annotated computer screen view of the Bio-Rad CFX Manager© (version 2.0).

Furthermore, this figure is specifically an illustrative example of what the AD graph looked like after the analysis had been conducted on the Bio-Rad CFX96 qPCR machine and after the data had been normalised based on the NTC results.

Figure 5.31 Bio-Rad CFX Manager© allelic discrimination for SNP rs7412



Allele 1 = T; Allele 2 = C; RFU = Relative fluorescence units; VIC = 2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein FAM = 6-caboxyfluorescien

The ancestral allele for rs7412 is the C allele with the T allele being the variant allele. The MAF was 0.14 in the present cohort which is slightly higher than the 0.10 among sub-Saharan Africans and 0.09 among (American) Caucasians (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=7412) previously reported. Possible reasons for these differences are provided in the following paragraphs.

This SNP was also in HWE according to the chi-square test results, with d.f of 2 (Table 5.16), indicative of the fact that this SNP adhered to the assumptions of HWE (Hardy, 1908). As previously discussed, the study population was representative of the larger black South African population, as the genotype frequencies fell in the 95% CI ranges (presented in Table 5.16) and therefore no sampling error occurred. The differences between the MAF of the two African populations can be ascribed to the internal diversity in these populations (Jorde & Wooding, 2004), whereas the MAF difference between the present black South African cohort and the Caucasian population can be ascribed to the aforementioned out-of-Africa theory and the fact that Caucasians present

with lower MAFs because of factors such as positive selection (Cummings, 2011; Jorde & Wooding, 2004).

Table 5.16 Adherence to the assumptions of Hardy-Weinberg Equilibrium of the rs7412 locus in the present PURE population using chi-square test analyses

rs7412 (C/T)					
Chi-square test for goodness of fit to the Hardy-Weinberg proportions					
Chi-square test	Genotype	CC	CT	TT	p-value$^{\$}$ (χ^2)
	Observed numbers (O)	1494	469	43	--
	Frequency (%)	74.47	23.38	2.14	
	Expected numbers (E)	1489.39	478.22	38.39	
	$\chi^2 = (O-E)^2/E$	0.01	0.18	0.55	0.39 (0.75)
Genotype distribution	95% CI ranges (%)	72.57-76.39	21.53-25.23	1.51-2.78	--

% = percentage; χ^2= Chi-square value; C = cytosine; CI = confidence intervals; E = expected; O = observed numbers; rs = reference sequence; T = thymine

A number of significant associations were observed between the rs7412 genotypes and the different lipid profile determinants in the different groupings (Table 5.17). The TC and LDL-C concentrations were significantly different between the three genotypes for the total population, gender and rural/urban groups (Table 5.17), whose effects remained after adjusting for the respective confounding variables (Section 5.5; Table 5.7). Unequal N *post hoc* tests indicated that the TC concentrations of the CC genotype carriers were significantly higher than the CT ($p < 0.01$) and TT genotype ($p = 0.03$) carriers, for females as well as the total population, respectively. Furthermore, the CC genotype carriers had significantly higher ($p < 0.05$) TC concentrations, than the CT genotype carriers for rural as well as urban dwellers. Similar results were observed for the LDL-C concentrations which were higher among the CC genotype carriers than the CT ($p = 0.00$) and TT genotype ($p < 0.004$) carriers, for the total population and women as well as rural and urban dwellers. Furthermore, the LDL-C concentrations were significantly higher among the CC genotype carriers than the CT genotype carriers ($p < 0.01$) among men and urban dwellers. Therefore, from these results it seems that overall the CC genotype carriers have elevated TC and LDL-C concentrations compared to the other genotype groups.

Even though the HDL-C concentrations were significantly associated with the different genotype classes in the total population, women and the urban group, the increased HDL-C concentrations observed among the TT genotype carriers, were not significantly

different from the other genotypes after unequal N *post hoc* tests had been conducted, possibly because of the relatively small sample size observed in the TT genotype group. However, the significance of this association was also lost when the confounding variables were adjusted for (Section 5.5; Table 5.7). A significant association was observed between the TG concentrations and the rs7412 genotypes among the rural group (Table 5.17); however, no significant associations between the genotype classes were determined, since the *post hoc* tests were not significant either.

Table 5.17 Associations of the lipid profile determinant concentrations with the rs7412 locus in the present PURE population

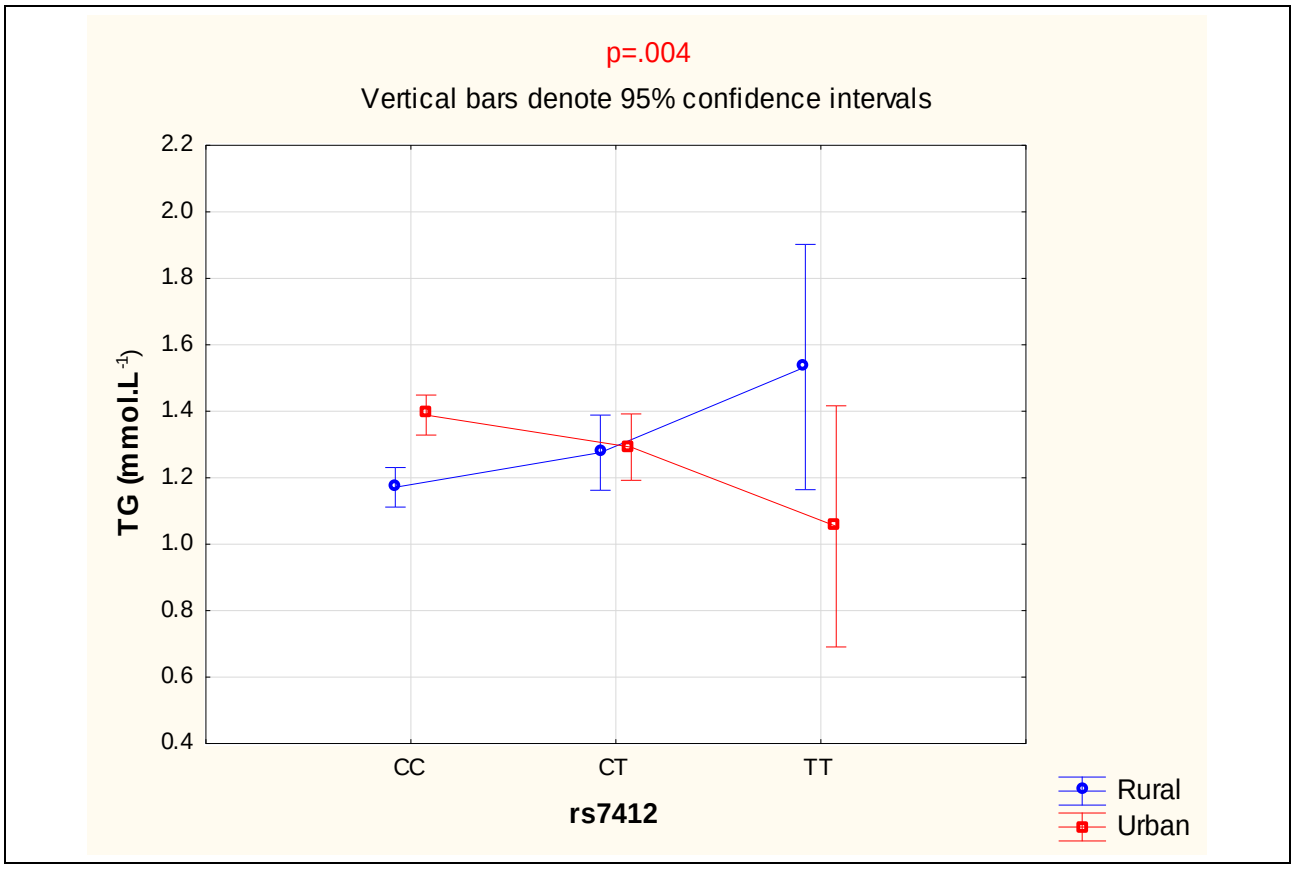
rs7412 (C/T)						
Lipid profile	**Stratification**	**CC**	**CT**	**TT**	**p-value$^{\$}$**	**p-value***
[TC]	Whole group	5.10	4.79	4.30	0.00	--
	Male	4.89	4.61	4.35	0.04	0.63
	Female	5.23	4.91	4.28	0.00	
	Rural	5.07	4.76	4.25	0.00	0.99
	Urban	5.14	4.82	4.36	0.002	
[LDL-C]	Whole group	3.04	2.62	2.05	0.00	--
	Male	2.80	2.45	2.03	0.00	0.50
	Female	3.18	2.73	2.05	0.00	
	Rural	3.03	2.62	1.91	0.00	0.74
	Urban	3.05	2.63	2.20	0.00	
[HDL-C]	Whole group	1.50	1.59	1.66	0.01	--
	Male	1.58	1.63	1.79	0.33	0.63
	Female	1.45	1.57	1.59	0.02	
	Rural	1.50	1.58	1.65	0.28	0.78
	Urban	1.49	1.61	1.67	0.03	
[TG]	Whole group	1.25	1.28	1.33	0.95	--
	Male	1.14	1.18	1.17	0.97	0.98
	Female	1.32	1.35	1.41	0.83	
	Rural	1.18	1.25	1.54	0.02	0.00
	Urban	1.33	1.31	1.10	0.20	

$^{\$}$ = indicates the p-values of the association rs7412 and each of the lipid profile determinants for the total population as well as by gender and locality groups; * = indicates the p-values of the interaction effect between rs7412 and gender/rural and urban; C = cytosine; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; rs = reference sequence; T = thymine; TC = total cholesterol; TG = triglycerides

A significant gene-locality interaction was observed and graphically, it seems as if the heterozygote (CT) and homozygote mutant genotypes (TT) are beneficial in an urbanised environment, but that the TT genotype has an adverse effect on TG concentrations in a rural environment (Figure 5.32). Nonetheless, these results are speculative, since no statistical significance has been reported for them in the present cohort and these

observations will therefore not be discussed in greater detail. However, the unequal N *post hoc* tests did indicate that the CC genotype carriers from urbanised areas presented with significantly higher TG concentrations than the CC genotype carriers from rural areas (p = 0.00). These results could imply the contributing role environmental factors play in the lipid profile of urban groups that have been reported to have adopted a more westernised diet that is high in dietary fats (Alberts *et al*., 2005; Vorster, 2002; Yusuf *et al*., 2001a). Another possibility to consider is that the urban carriers of the CC genotype are not equipped to metabolise the (increased) TG concentrations as effectively as those harbouring the other genotypes.

Figure 5.32 The interactive effect of locality with rs7412 on triglyceride concentrations for each of the genotype groups



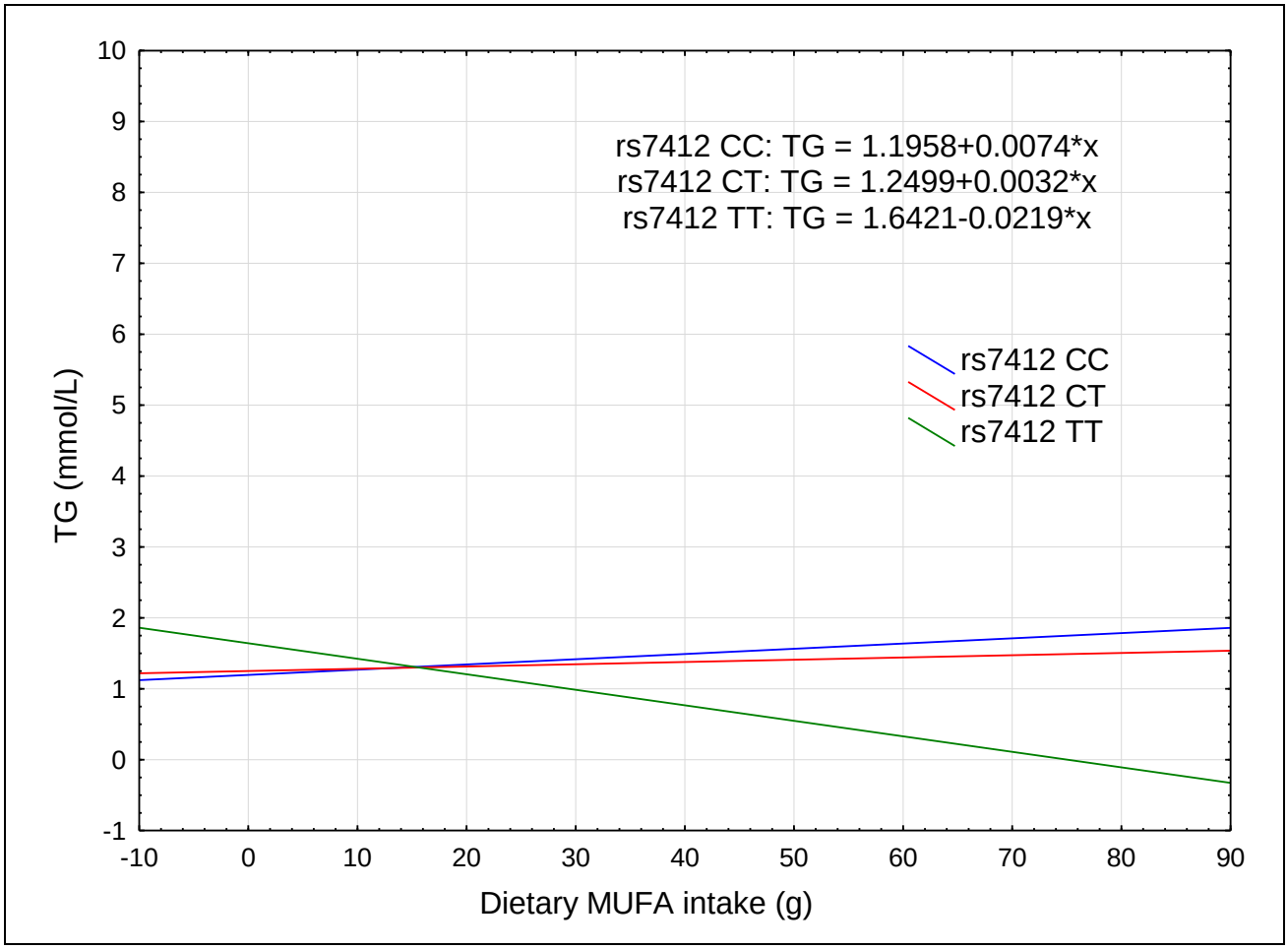
C = cytosine; mmol.L-1 = millimole per litre; rs = reference sequence; T = thymine; TG = triglycerides

Furthermore, significant gene-diet interactions were observed for rs7412, but after *post hoc* tests had been conducted these significant effects were lost, possibly as a result of multiple testing or the small sample size observed in the TT genotype group. However, as the nature of the study was exploratory, the possible implication of these interaction effects can only be speculated on and will be discussed in the following paragraphs.

According to the literature, TG concentrations decrease with the consumption of MUFAs (Kris-Etherton, 1999), but increase with increasing dietary TFA and cholesterol consumption (Mozaffarian & Clarke, 2009; Mozaffarian & Willett, 2009). The observations made for the interactive effects of the respective dietary variables (cholesterol and TFA) with the rs7412 SNP on TG concentrations were in accordance with those of the publications quoted above, in that an inverse correlation between dietary MUFAs and TG concentrations was observed among individuals harbouring the TT genotype (Figure 5.33). On the other hand, the interactive effects of dietary cholesterol and TFA with the rs7412 SNP on TG concentrations were not in accordance with these publications, as there was an inverse linear correlation between dietary cholesterol and TFA intake and TG concentration among individuals harbouring the TT genotype instead of the expected positive linear correlation (Figure 5.34).

In Section 3.2 four mechanisms that influence lipid metabolism were discussed, one of which was the manner in which the ApoE isoforms influence the rate of exogenous fat removal. The ApoE2 isoforms (which are a product of the strong LD between the rs429358 and rs7412 SNPs) have been reported to have defective binding ability to the hepatic CM remnant receptors. This results in a slow hepatic CM remnant clearance rate which is in turn associated with a delayed catabolism of CM and VLDL remnants leading to elevated TG concentrations (Davignon *et al*., 1988; Weisgraber *et al*., 1982). Therefore, one can hypothesise that the TT genotype of the rs7412 SNP counteracts this ApoE protein characteristic, by possibly increasing the CM remnant clearance rate *via* increasing receptor-binding activities, resulting in reduced TG concentrations in this specific cohort. The ApoE protein is also involved in the process of cholesterol absorption and ε2 carriers have been reported to present with higher hepatic cholesterol synthesis and lower intestinal cholesterol absorption than ε4 carriers (Gylling *et al.*, 1995; Miettinen *et al.*, 1992; Kesäniemi & Miettinen, 1987). It can therefore be hypothesised from these facts that the effect of the rs7412 SNP alone, may possibly have opposite effects on the metabolism of lipids when not in LD with the rs429358 SNP and that the TT genotype of the rs7412 SNP causes these effects. However, the results from the univariate tests indicated that neither the SNP ($p = 0.07$) nor the dietary MUFA intake ($p = 0.36$) solely had a significant effect on the TG concentrations, but the interactive effect of these two variables on the TG concentrations was significant ($p = 0.03$). Furthermore, owing to possible multiple testing bias and inadequate power, the true effect will require further study.

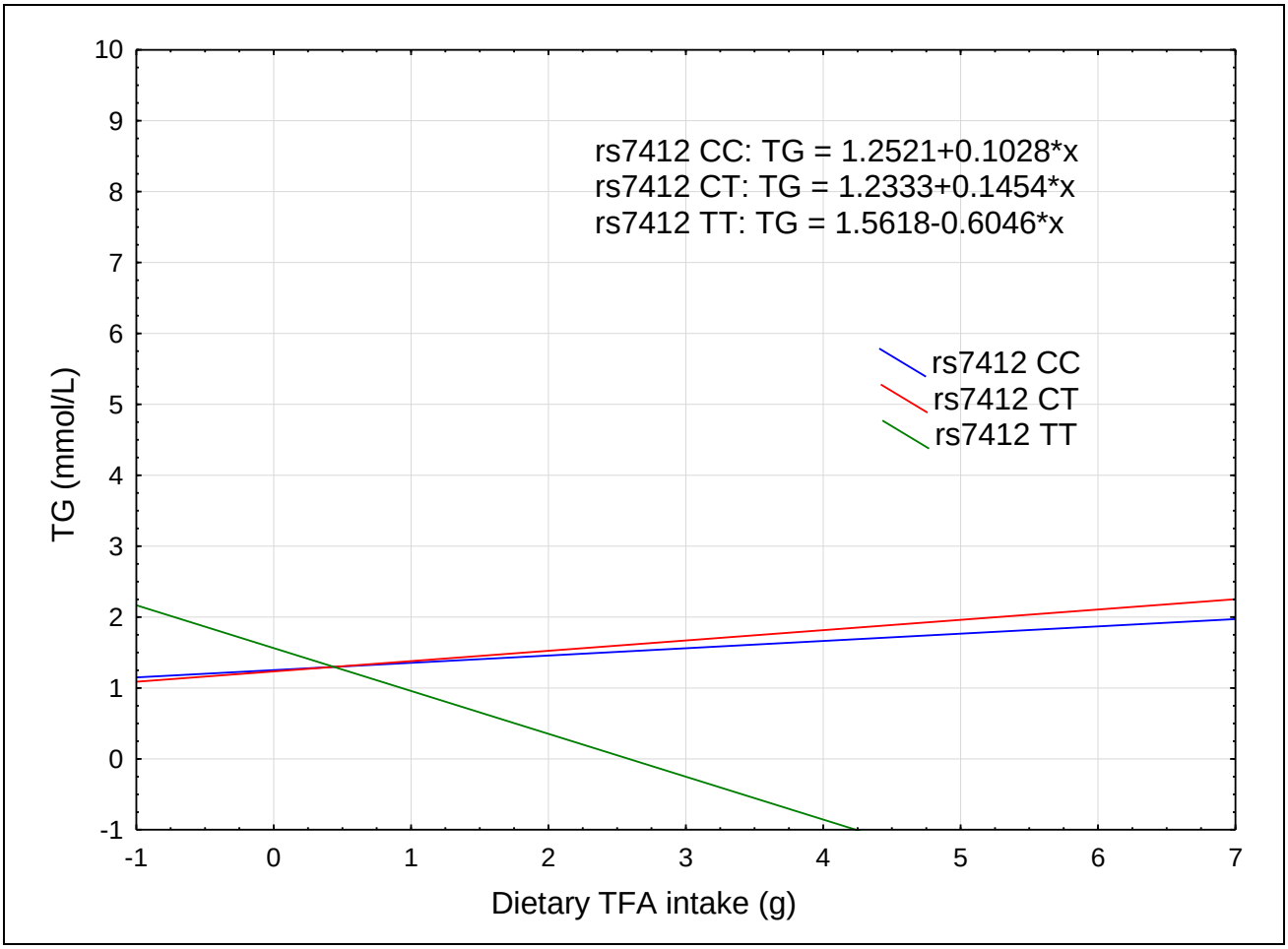
Figure 5.33 The interactive effect of dietary MUFA intake with rs7412 on triglyceride concentrations for each of the genotype groups



C = cytosine; g = grams; mmol.L^{-1} = millimole per litre; MUFA = mono-unsaturated fatty acids; rs = reference sequence; T = thymine; TG = triglycerides

The univariate test indicated a significant interactive effect between the dietary with the rs7412 SNP on the TG concentrations ($p = 0.03$). However, these variables were not significantly associated with TG concentrations on their own, $p > 0.05$ (Figure 5.34). Furthermore, the *post hoc* tests for this interaction effect were also not significant. This implies that this effect is not true in the present cohort and should be interpreted with care, for the aforementioned reasons.

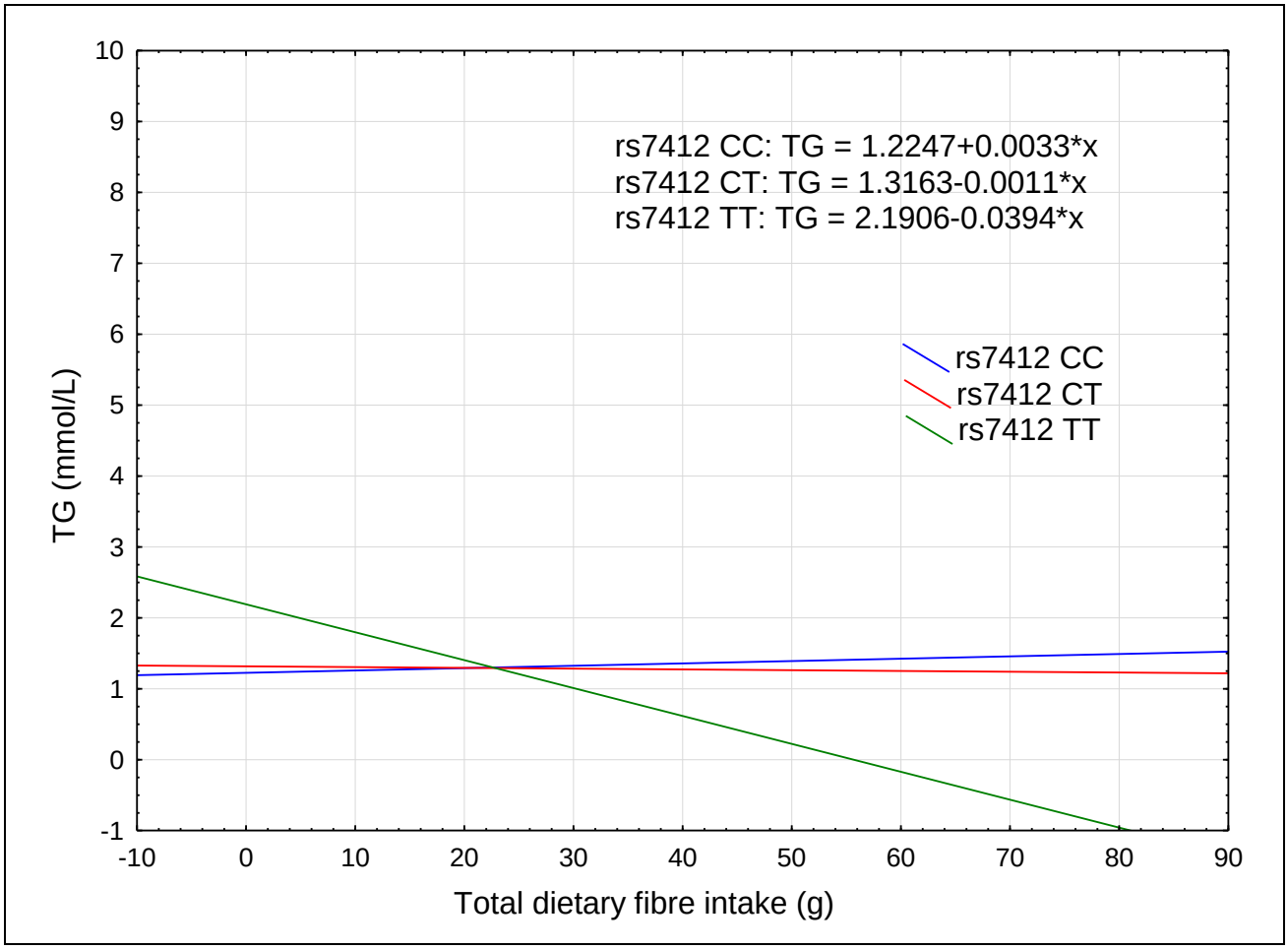
Figure 5.34 The interactive effect of dietary TFA intake with rs7412 on triglyceride concentrations for each of the genotype groups



C = cytosine; CHOL = cholesterol; g = grams; mmol.L^{-1} = millimole per litre; rs = reference sequence; T = thymine; TFA = trans fatty acids; TG = triglycerides

Figure 5.35 represents the significant interactive effect of dietary fibre intake with rs7412 on TG concentrations for each of the genotype groups ($p = 0.01$). Results from the univariate tests indicated that the dietary fibre ($p = 0.02$) and the SNP ($p = 0.01$), also had a single-handed significant effect on the TG concentrations. A meta-analysis on the cholesterol-lowering effects of dietary fibre concluded that TG and HDL-C were not significantly influenced by (soluble) fibre intake (Brown *et al*., 1999). However, an inverse association between dietary fibre and serum TG and a positive association between dietary fibre and HDL-C concentrations have been reported by other researchers (Berg *et al*., 2008; Sonneneberg *et al*., 1996). According to Figure 5.38, such an inverse correlation between TG concentrations and total dietary fibre intake appears to be present among the homozygote mutant (TT) carriers. However, as the *post hoc* tests were not significant, this result should be interpreted with care.

Figure 5.35 The interactive effect of dietary fibre intake with rs7412 on triglyceride concentrations for each of the genotype groups



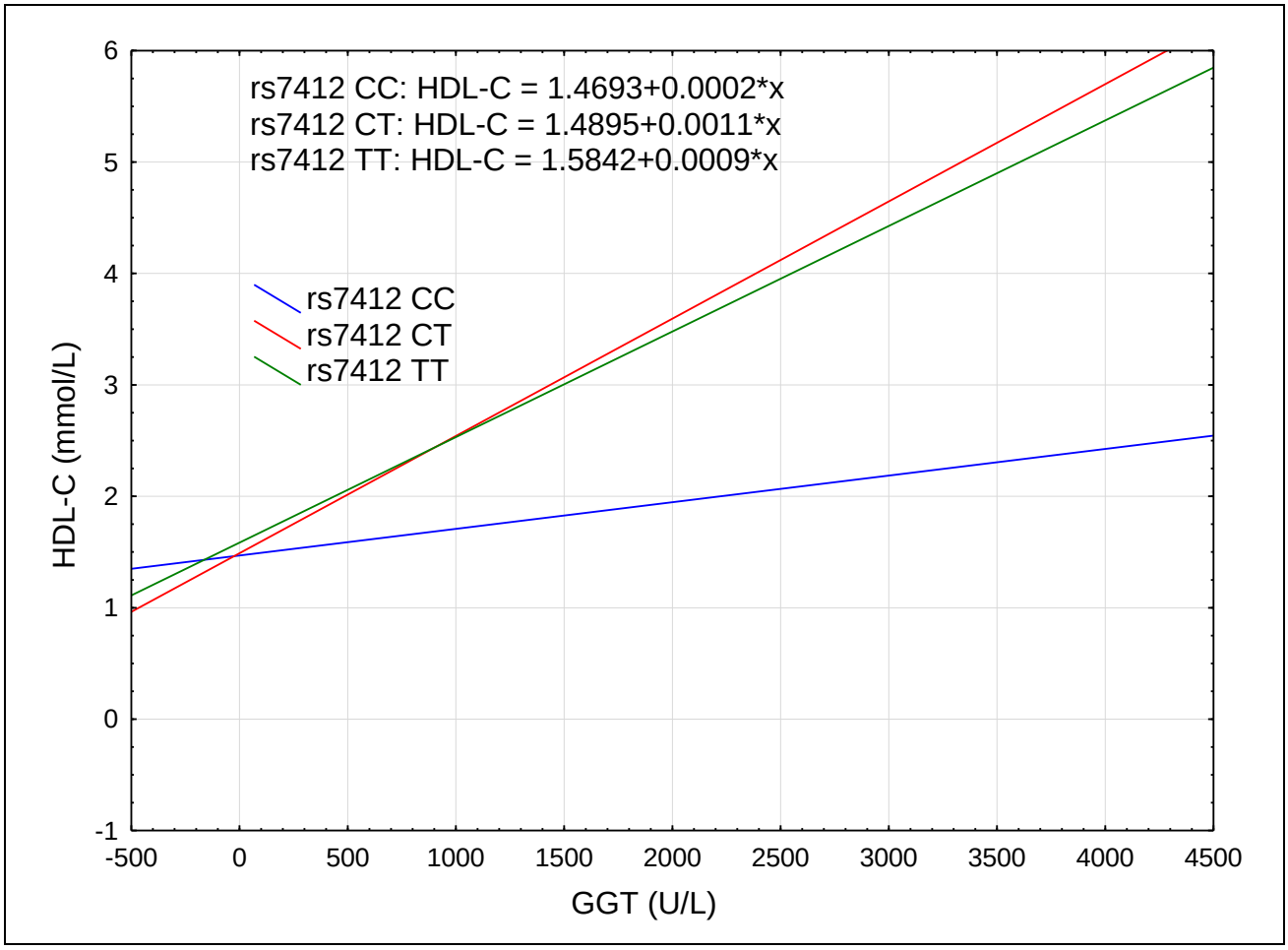
C = cytosine; g = grams; mmol.L^{-1} = millimole per litre; rs = reference sequence; T = thymine; TG = triglycerides

This gene-diet interaction might imply that an additional TG lowering effect can possibly be achieved among TT genotype carriers through increased total dietary fibre intake. Similar correlations and interaction effects have been reported for other *ApoE* genetic variants before, but not for the rs7412 SNP itself (Ordovas, 1999). Some researchers have argued that the cholesterol-lowering effect of fibre is independent of ApoE, as fibre results in increased bile acid excretion and consequently the liver makes use of the stored cholesterol to produce new bile acids. The bacterial by-products of fibre fermentation in the colon have also been determined to inhibit cholesterol synthesis in the liver (Eastwood, 1992). Therefore, based on this evidence it is clear that this interaction effect is not a result of a reported biological mechanistic link. It is therefore unexpected and needs further investigation (Wu *et al*., 2007). The small sample size harbouring the TT genotype and multiple testing bias indicate that these results should be interpreted with care and their true effect should be confirmed in powered studies.

Although the obvious dietary factors implicated in gene-diet interactions affecting plasma lipid concentrations are dietary fats, cholesterol and fibre intake, other dietary components such as alcohol intake have revealed significant interactions as well (Ordovas, 2008). As mentioned in Section 5.1, GGT was measured as a marker of alcohol consumption and has been reported to correlate positively with HDL-C in both men and women from the present cohort group (Pisa *et al*., 2010). Furthermore, other researchers have reported that no *ApoE*-alcohol interaction effect on lipid concentrations (*i.e*. HDL-C) occur among Caucasian populations (Chen *et al*., 2009; Djoussé *et al*., 2004).

In this cross-sectional, exploratory study significant interaction ($p = 0.00$) between rs7412 and GGT concentrations on HDL-C concentrations was observed (Figure 5.36), which remained significant after *post hoc* testing (Bonferonni, $p = 0.02$). The *post hoc* test furthermore indicated that the HDL-C concentrations were significantly higher in the heterozygote genotype group (CT) compared to the wild type homozygote genotype group, TT ($p = 0.02$), indicating that the mutant allele might in this case be advantageous while interacting with GGT concentrations. To the knowledge of the researcher a limited number of analyses have been performed on the interaction between the *ApoE* genotype and alcohol consumption and its modulating effect on various lipids. Therefore, the results presented in this study contribute to existing literature and the mechanistic role of this interaction needs to be investigated further.

Figure 5.36 The interactive effect of GGT concentrations with rs7412 on HDL-C concentrations for each of the genotype groups



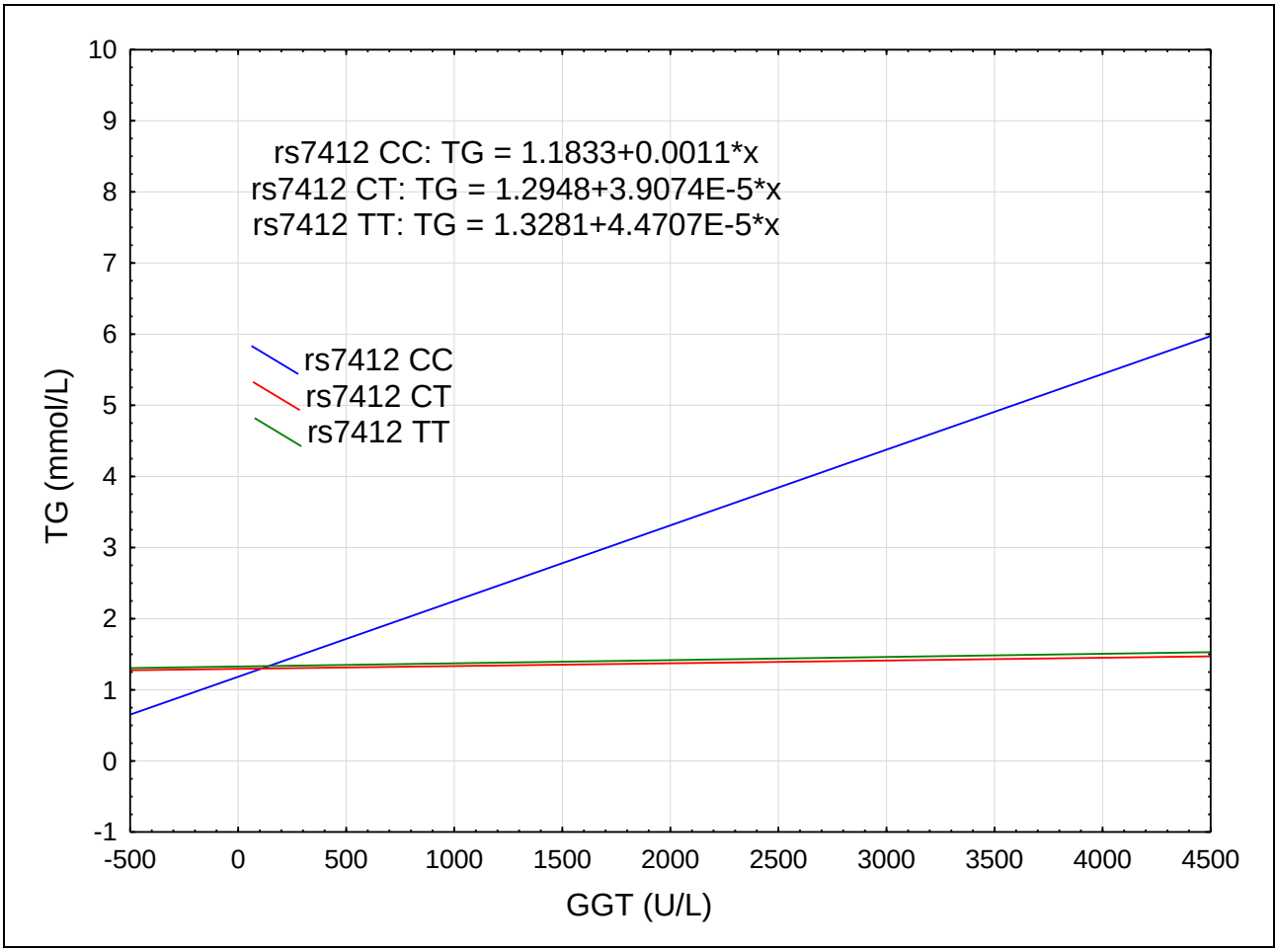
C = cytosine; GGT = gamma-glutamyl transferase; HDL-C = high-density lipoprotein cholesterol; mmol.L^{-1} = millimole per litre; rs = reference sequence; T = thymine; U/L = units per litre

Although the positive correlation between alcohol consumption and HDL-C concentrations is well known, its effect on the other lipid profile determinants, *i.e*. LDL-C and TG, is still unclear (Ordovas, 2008). A possible hypothesis to explain why these correlations with other lipid profile determinants could not have been confirmed might be derived from existing interactive effects between genetic variants and alcohol consumption. The significant interaction ($p = 0.00$) between the GGT concentrations and the TG concentrations in the different rs7412 genotypes observed in the present study is supportive of this hypothesis (Figure 5.37).

Pisa *et al*. (2010) reported a significant positive correlation between the TG and TC concentrations and GGT concentrations for both men and women in the black South African cohort under investigation. Furthermore, as indicated in Section 5.1, the TG as well as GGT concentrations were highest among urban dwellers. According to Table 5.17, the TG concentrations are elevated among CC genotype carriers and in addition to this, the present study observed significant interaction ($p = 0.00$) between the GGT

concentrations and the TG concentrations in the different rs7412 genotypes (Figure 5.37). Furthermore, increased TG concentrations were observed among the CC genotype in Figure 5.37 and elevated TG concentrations were observed among urban dwellers (Figure 5.32). These results might imply that even though urban dwellers consume more alcohol than rural dwellers and as a result present with significantly higher GGT concentrations ($p = 0.01$), individuals harbouring the CC genotype have significantly higher TG. The data suggest that in urbanised individuals, variability at this locus modulates the effects of consuming alcoholic beverages on TG concentrations.

Figure 5.37 The interactive effect of GGT concentrations with rs7412 on TG concentrations for each of the genotype groups



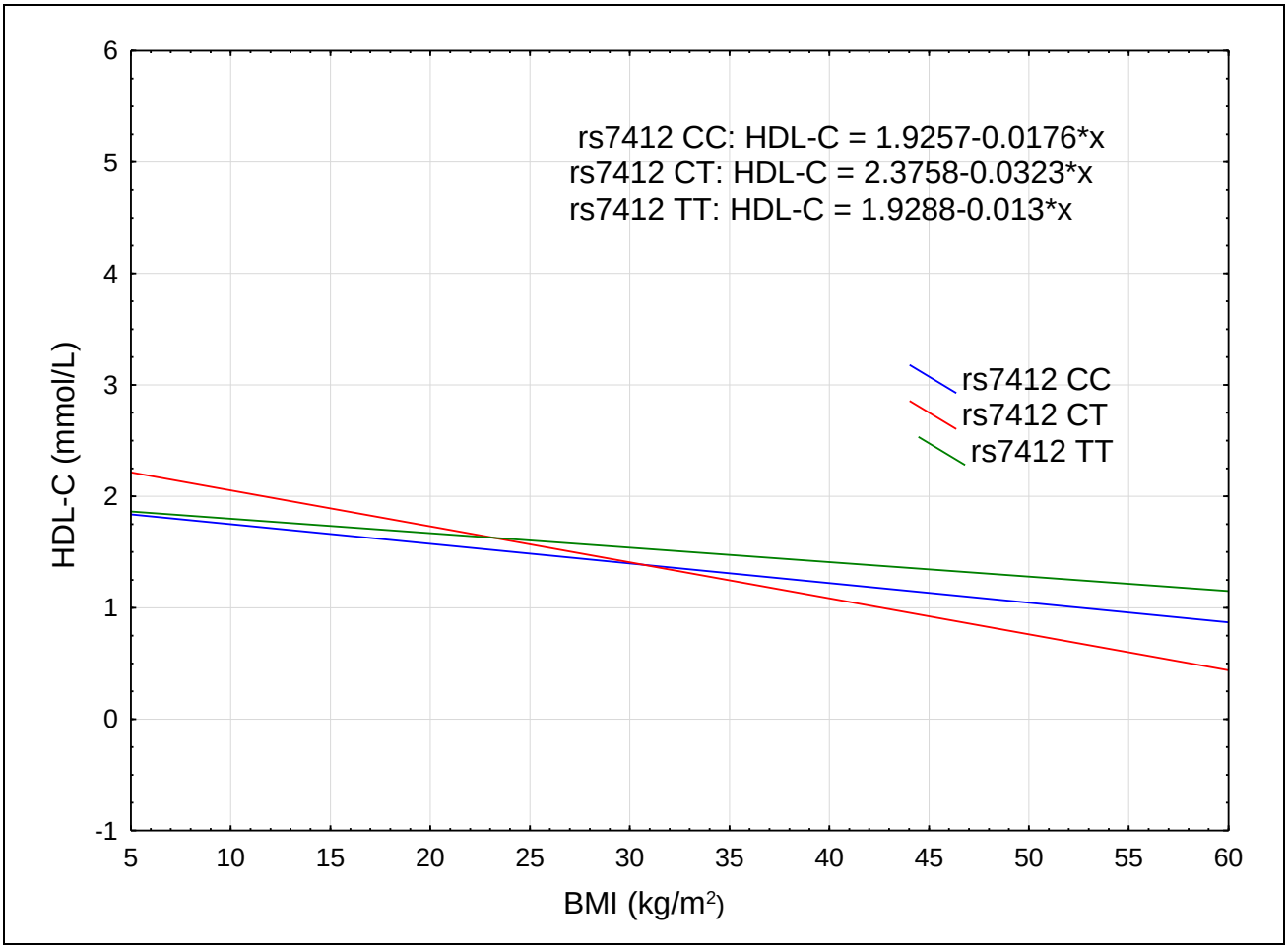
C = cytosine; mmol/L = millimole per litre; rs = reference sequence; T = thymine; TG = triglyceride; U.L-1 = units per litre

Significant gene-phenotype interaction was also observed for the rs7412 SNP. The inverse correlation between BMI and HDL-C is supported by literature (Shamai *et al.,* 2011; Krauss *et al.,* 1998). A number of obesity metrices exist, such as waist circumference, waist-to-hip ratio, waist-to-height ratio and BMI, of which the latter has been reported to be the simplest and most clinically practical and accepted measure of obesity (Litwin, 2008). According to Figure 5.38, it seems that the significant interaction

($p = 0.02$) between the rs7412 SNP and BMI might be a modifying factor for the lower HDL-C concentrations observed in the C-allele carriers of the rs7412 SNP, as the *post hoc* tests were determined to be significant (Bonferonni: $p = 0.01$).

This observation is in accordance with the suggestion that with the inclusion of gene-obesity interaction effects on HDL-C, the total variance of HDL-C concentrations is explained (Lamina *et al*., 2012). This research group reported results based on SNP scores, which indicated that a substantial part of the genetic regulation of HDL-C is modified by increasing obesity, as the combined effect of HDL-C-increasing alleles is attenuated by increasing obesity levels (Lamina *et al*., 2012). Furthermore, the reported results of the present study are in conflict with Mi *et al*. (2011), who reported no significant BMI-*ApoE* interaction on HDL-C concentrations in a small number of non-Caucasians. However, Mi *et al*. (2011) might not have been able to detect such a significant effect owing to lack of power. A proposed mechanism for this observed interaction is provided in Section 5.5.6 pertaining to the modulating effect of the *ApoE* (ε2/ ε2)-BMI interaction on HDL-C concentrations (Figure 5.40).

Figure 5.38 The interactive effect of BMI with rs7412 on HDL-C concentrations for each of the genotype groups



BMI = body mass index; C = cytosine; HDL-C = high-density lipoprotein cholesterol; $kg.m^{-2}$ = kilograms per square meters; $mmol.L^{-1}$ = millimolle per litre; rs = reference sequence; T = thymine

5.5.6 *ApoE* haplotype-genotypes (the ε2/ ε3/ ε4 epsilons)

As discussed in the literature review chapter (Section 3.2) most reported data pertaining to genetic variation in the *ApoE* gene refer to the ε2/ ε3/ ε4 haplotype-genotypes. These haplotype-genotypes result from the strong LD between the genotypes at the rs429358 (Section 5.5.4) and rs7412 loci (Section 5.5.5), as discussed extensively in Section 3.2. Ten combinations of these *ApoE* haplotype-genotypes occurred in the present cohort (Table 5.18), as opposed to the six well-known haplotype-genotype combinations, determined in Zannis *et al.* (1982). The reason for the additional four combinations is the presence of the fourth (rare) allele (ε3r) in the South African cohort investigated, which was unexpected, since the occurrence of this allele has only been reported four times before (Section 3.2). However, the observed LD does not support the formation of this rare ε3r allele due to intragenic recombination in an ε2/ε4 heterozygote, but rather favours the formation of this allele through a T-C mutation at nucleotide 3937 in an ε2 allele, or through a C-T mutation at nucleotide 4075 in an ε4 allele (Seripa *et al*., 2007), as discussed in Section 3.2.

The haplotypic distribution was tested for HWE (Table 5.18). The chi-square test indicated that these haplotypes did not adhere to the assumptions of HWE with a significant p-value (0.00) and a d.f. of 9 (Hardy, 1908). This might indicate that as this population is undergoing an epidemiological transition; this form of migration may explain the change in the haplotype frequencies observed. Another explanation may be positive selection, which might be at play in the present cohort, as one halpotype-genotype (ε3 carriers) seems to be more prevalent than the other halpotype-genotypes, which might indicate that the former halpotype-genotype has an "increased" ability to survive and reproduce (Cummings, 2011). However, according to the analysis that was conducted to measure the range within which the actual haplotype-genotype frequency is, no sampling error occurred, as the genotype frequencies for each of the haplotype-genotypes were within the 95% CI ranges (Table 5.18).

Table 5.18 Adherence to the assumptions of Hardy-Weinberg Equilibrium of the various haplotype genotypes in the present PURE population using chi-square test analyses

Haplotype-genotype						
Chi-square test for goodness of fit to the Hardy-Weinberg proportions						
Haplotype genotype	Observed numbers (O)	Frequency (%)	Expected numbers (E)	χ^2= (O-E)χ^2/E	p-value (χ^2)	95% CI ranges
ε2/ε2	28	1.94	13.28	16.33368	0.00	1.23-2.65
ε2/ε3	221	15.29	176.84	11.02802		13.44-17.15
ε2/ε4	0	--	53.10	53.09965		--
ε2/ε3r	0	8.51	11.79	11.78927		--
ε3/ε3	519	35.92	588.93	8.303913		33.44-38.39
ε3/ε4	463	32.04	353.68	33.79133		29.63-34.45
ε3/ε3r	123	--	78.52	25.19089		7.07-9.95
ε4/ε4	91	6.30	53.10	27.05171		5.04-7.55
ε4/ε3r	0	--	23.58	23.57855		--
ε3r/ε3r	0	--	2.62	2.617474		--
Total	1445	100	1355.44	212.79	---	---

ε = epsilon; % = percentage; χ^2= Chi-square value; CI = confidence intervals; E = expected; O = observed numbers

The distribution of the *ApoE* haplotype-genotype frequencies in the present cohort is presented in Table 5.18. Overall, 38.34% of subjects were carriers of the *ApoE* ε4 allele, whereas 17.23% of subjects were carriers of the *ApoE* ε2 allele. In the present cohort, the ε4 and ε2 allele distributions are higher than the common 20% and 10% reported among Caucasians. However, these results are similar to those reported by other researchers who have also investigated the *ApoE* allele distribution in a black South African population. A study conducted by Loktionov *et al*. (1999) reported a difference in the *ApoE* allele distribution between black South Africans and UK Caucasians. The *ApoE* distribution

present in the UK Caucasians was similar to that previously reported in other Caucasian studies, with an approximate 70% frequency of the ε3/ε3 genotype, 20% of the ε4-bearing subjects and 10% harbouring the ε2 allele. The incidence of the Apoε4 allele among the black South African population (Loktionov *et al*., 1999) was higher (28.5%) compared to the 11.2% prevalence in the Caucasian group. In 1999, Loktionov *et al*. reported the highest frequency (14.5%) ever observed for the Apoε2 allele, although this was in a smaller sample of black South Africans than in the current study (this allele is associated with lower CHD risk). Furthermore, Loktionov *et al*. (1999) reported the lowest frequency of ε3/ε3 carriers (30%), lower frequencies for the ε4/ε4 haplotype-genotype (9%) and a frequency of 0% for the ε2/ε2 haplotype-genotype, whereas the expected haplotype frequency for ε2/ε2 was reported to be 2%, among the black South African population. The ε4/ε4 haplotype frequency in the present cohort (6.30%) was less than that reported previously for this ethnic group. Furthermore, the ε2/ε2 haplotype was observed in the present cohort, but at a low frequency (± 2%), even though the expected frequency was much higher, 13.28% (Table 5.18). The ε3/ε3 distribution was slightly higher in the present cohort (approximately 6%, Table 5.18) than that reported by Loktionov *et al*. (1999).

Therefore, according to these results, it is suggested that the potential genetic risk, as defined by the high-risk ε4 allele, is high in the black South African population, which is reported to have a generally low CVD incidence rate. Even though this may seem contradictory, it in actual fact emphasises the role that environmental factors play in CVD development. In essence the unfavourable *ApoE* allele frequencies reported in the black South African population may result in a rapid increase in CHD incidence among this population in the event that their exposure to adverse environmental factors (*i.e*. high-fat Western diets and a sedentary lifestyle, which results from urbanisation) should also increase. Loktoniov *et al.* (1999) also investigated other quantitative CVD risk biomarkers (*i.e.* blood lipid concentrations) and determined that the TC and LDL-C concentrations are dependent on the *ApoE* genotype in black South Africans; however, these concentrations were always 20-30% lower compared to the Caucasian subjects, with similar *ApoE* genetic variants. The adverse effect of *ApoE* polymorphisms on lipid concentrations among rural, black South African subjects is similar to that in Caucasians, even though the absolute values of the lipid concentrations (particularly TC and LDL-C) were 20-30% lower among the rural South African population than in the Caucasian population that presented with similar *ApoE* allele frequencies (Loktionov *et al*., 1999). These results emphasise the role of other contributing social and behavioural factors (*i.e.* urbanisation), which may play a

vital part in the development of a multi-factorial disease such as CVD (Loktionov *et al.*, 1999).

Further statistical analyses were conducted on the haplotype-genotype groups for which a sufficient amount of data were available. These included the following groups: ε2/ε2; ε3/ε2; ε3/ε3; ε3/ε3r; ε4/ε4 and ε3/ε4. In general, the results presented in Table 5.19 are similar to the published data related to the *ApoE* genotype-lipid profile relationship in other populations *i.e.* non-Hispanic blacks; Mexican-Americans and Tunisians (Chang *et al.,* 2010; Lanktree *et al.,* 2009; Jemaa *et al.,* 2006; Atkinson *et al.,* 2004).

Table 5.19 Associations of lipid profile determinant concentrations with haplotype genotypes in the present PURE population

		Haplotype-genotypes							
		ε3/ε4	**ε3/ε3**	**ε3/ε3r**	**ε3/ε2**	**ε4/ε4**	**ε2/ε2**	**p-value$^{\$}$**	**p-value***
[TC]	Whole group	5.11	5.10	4.85	4.75	5.14	4.19	0.00	--
	Male	4.86	4.91	4.50	4.50	4.90	4.09	0.05	0.95
	Female	5.27	5.20	5.04	4.89	5.30	4.25	0.01	
	Rural	5.02	5.17	4.73	4.63	5.13	3.98	0.00	0.28
	Urban	5.21	5.04	4.96	4.86	5.15	4.40	0.14	
[LDL-C]	Whole group	3.05	3.02	2.74	2.56	3.20	1.88	0.00	--
	Male	2.79	2.78	2.41	2.37	2.98	1.93	0.00	0.63
	Female	3.21	3.14	2.92	2.67	3.34	1.85	0.00	
	Rural	3.03	3.06	2.69	2.51	3.21	1.73	0.00	0.85
	Urban	3.07	2.98	2.78	2.60	3.18	2.04	0.00	
[HDL-C]	Whole group	1.50	1.52	1.51	1.64	1.43	1.74	0.00	--
	Male	1.55	1.62	1.54	1.64	1.42	1.65	0.57	0.35
	Female	1.46	1.46	1.49	1.63	1.44	1.79	0.00	
	Rural	1.48	1.57	1.45	1.59	1.41	1.63	0.44	0.07
	Urban	1.52	1.47	1.56	1.68	1.46	1.85	0.00	
[TG]	Whole group	1.25	1.26	1.34	1.23	1.14	1.26	0.42	--
	Male	1.19	1.19	1.22	1.17	1.11	1.12	0.99	0.96
	Female	1.33	1.33	1.41	1.30	1.15	1.34	0.42	
	Rural	1.13	1.20	1.30	1.19	1.13	1.40	0.36	0.07
	Urban	1.39	1.33	1.38	1.28	1.14	1.13	0.28	

$^{\$}$ = indicates the p-values of the haplotype-genotype association and each of the lipid profile determinants for the total population as well as by gender and locality groups; * = indicates the p-values of the interaction effect between haplotype-genotypes and gender / rural, urban; ε = epsilon; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; TC = total cholesterol; TG = triglycerides

In the present black South African population, one-way ANOVAs indicated that the TC, LDL-C and HDL-C concentrations differed significantly between the different haplotype-genotype groups in the total population (Table 5.19). After adjusting for the respective confounding variables (Section 5.5., Table 5.7), the significant associations for TC and LDL-C remained unchanged. The ε4 haplotype carriers presented with a more unfavourable lipid profile compared to the other genotype carriers (Table 5.19). The addition of the ε4 haplotype, resulted in an increase in TC and LDL-C, but a decrease in HDL-C and TG, whereas the opposite was observed for the genotypes harbouring the ε2 allele (Table 5.19). According to the *post hoc* tests, only the LDL-C concentrations varied significantly between the haplotype-genotype groups for the total population. The ε3 carriers (ε3/ε4 and ε3/ε3) had significantly lower LDL-C concentrations compared to the ε4/ε4 carriers ($p < 0.001$), but higher than the ε2/ε2 carriers ($p = 0.003$). Furthermore, the ε3/ε2 carriers had significantly lower LDL-C concentrations than the ε4/ε4 carriers ($p = 0.004$), with the latter presenting with significantly higher LDL-C concentrations than the ε2/ε2 carriers, $p = 0.00$ (Table 5.19). The same results for the *post hoc* tests were obtained for women and rural dwellers. These results are in accordance with previous reports indicating that the *ApoE* genotype (ε2<ε3<ε4) was associated with increasing LDL-C concentrations among the different sub-groups (Corella *et al*., 2001; Weintraub *et al.*, 1987). Furthermore, the *post hoc* tests indicated that the significant association between LDL-C concentrations and the different haplotype-genotype groups was also present in the urban group, with the ε3/ε4 carriers presenting with significantly higher LDL-C concentrations than the ε3/ε2 carriers; $p = 0.04$ (Table 5.19). In addition, the *post hoc* test indicated that the ε3/ε3 carriers had significantly higher TC concentrations than the ε3/ε2 carriers residing in a rural area ($p = 0.03$), indicating that the possibility of the ε2 haplotype's protective function is also applicable to the present cohort.

The TG concentrations were lowest among the ε4/ε4 genotype carriers compared to the other genotype groups and even though this observation did not attain statistical significance it is contradictory to the literature, which reports significantly higher TG among ε4 carriers compared to non-carriers (Rahilly-Tierney *et al.,* 2011; Jemaa *et al.,* 2006; Frikke-Schmidt *et al*., 2000). An interesting observation was made regarding the haplotype-genotype harbouring the *rare allele* (haplotype) ε3/ε3r, which presented with the highest TG concentrations compared to the other groups. Even though this did not attain significance between the haplotype-genotype groups according to the *post hoc* tests either, the association itself is different from the normal lipid profile generally associated with this haplotype (Murrell *et al.,* 2006). Nonetheless, the presence of the ε3r allele has

only been reported a few times before and this association (though not significant) therefore seems to be a novel finding and contributes to the scarce literature pertaining to this rare allele.

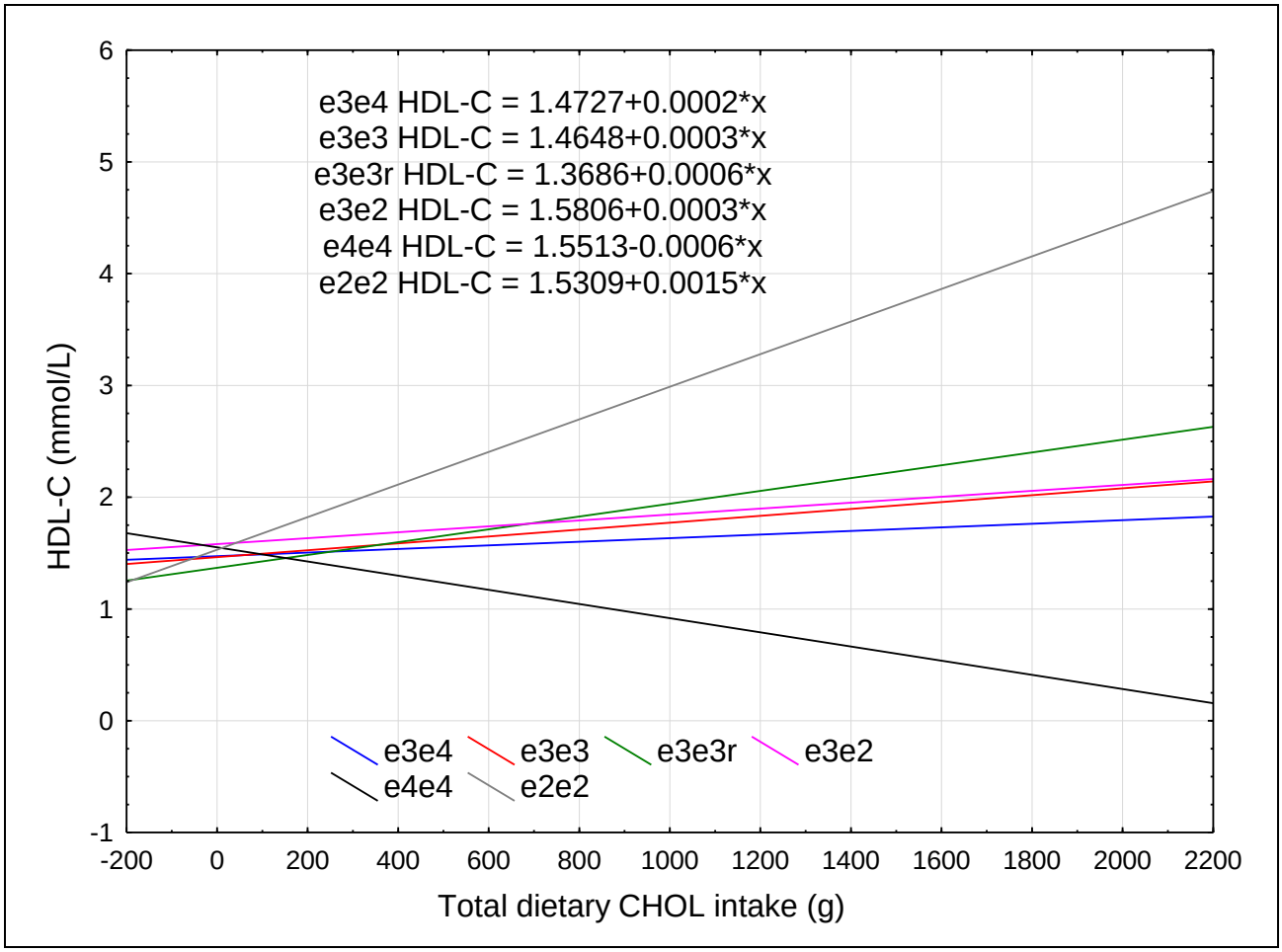
In general the majority of ε2/ε2 carriers (> 90%) are hypolipidaemic and present with lower HDL-C and LDL-C concentrations (Mahley & Rall, 2000). However, elevated HDL-C concentrations were observed for ε2/ε2 carriers in the total population (Table 5.19), which might on a mechanistic level explain the healthy lipid profile (desirable HDL-C) observed in the present cohort (Section 5.1). However, the ε2/ε2 carriers were not especially prevalent in the present cohort, with a frequency of approximately 2% and it is therefore unlikely that they had a population effect. Furthermore, the HDL-C concentrations among the ε2/ε2 carriers were not significantly different from the HDL-C concentrations observed in the other haplotype-genotype groups, as the *post hoc* test were not significant.

Contradictory results have been reported regarding the effect of the *ApoE* genotype on HDL-C (Jemaa *et al.,* 2006; Frikke-Schmidt *et al.,* 2000; Ikewaki *et al.,* 1994). However, it seems from the significant positive association observed between the *ApoE* haplotype-genotype groups and HDL-C for the whole group in the present cohort, that these data support the notion of such an association (Frikke-Schmidt *et al.,* 2000; Ikewaki *et al.,* 1994). Although the mechanism of this relationship remains to be clarified, it has been hypothesised that gene-environment interactions contribute to the inconsistencies between studies, as it is clear that environmental and phenotypic factors interact with genetic factors in determining HDL-C in each of the *ApoE* genotypes (Jemaa *et al.,* 2006). Such interactions are observed in the present cohort and are, therefore, supportive of this notion (Figure 5.39-5.41).

The gene-diet interaction represented in Figure 5.39 that was observed in the present cohort was of borderline significance ($p < 0.05$) and the true impact may therefore be questioned. From Figure 5.39 it seems that harbouring the ε2/ε2 genotype may have a beneficial effect on the lipid profile, as HDL-C concentrations increase with an increase in dietary fat intake. However, the univariate test (factorial ANOVA) indicated that a significant positive correlation exists between dietary cholesterol intake and HDL-C concentrations ($p = 0.001$), as expected, but no significant association between this genetic variation alone and HDL-C was observed ($p = 0.55$). Furthermore, the *post hoc* tests were not significant. Therefore, it seems that dietary cholesterol intake has a greater

effect on HDL-C concentrations compared to the effect of the genetic variant or the gene-diet interaction effect.

Figure 5.39 The interactive effect of dietary cholesterol intake with the *ApoE* haplotype-genotypes on HDL-C concentrations for each of the genotype groups



% = percentage; ε = epsilon; CHOL = dietary cholesterol intake; HDL-C = high-density lipoprotein cholesterol

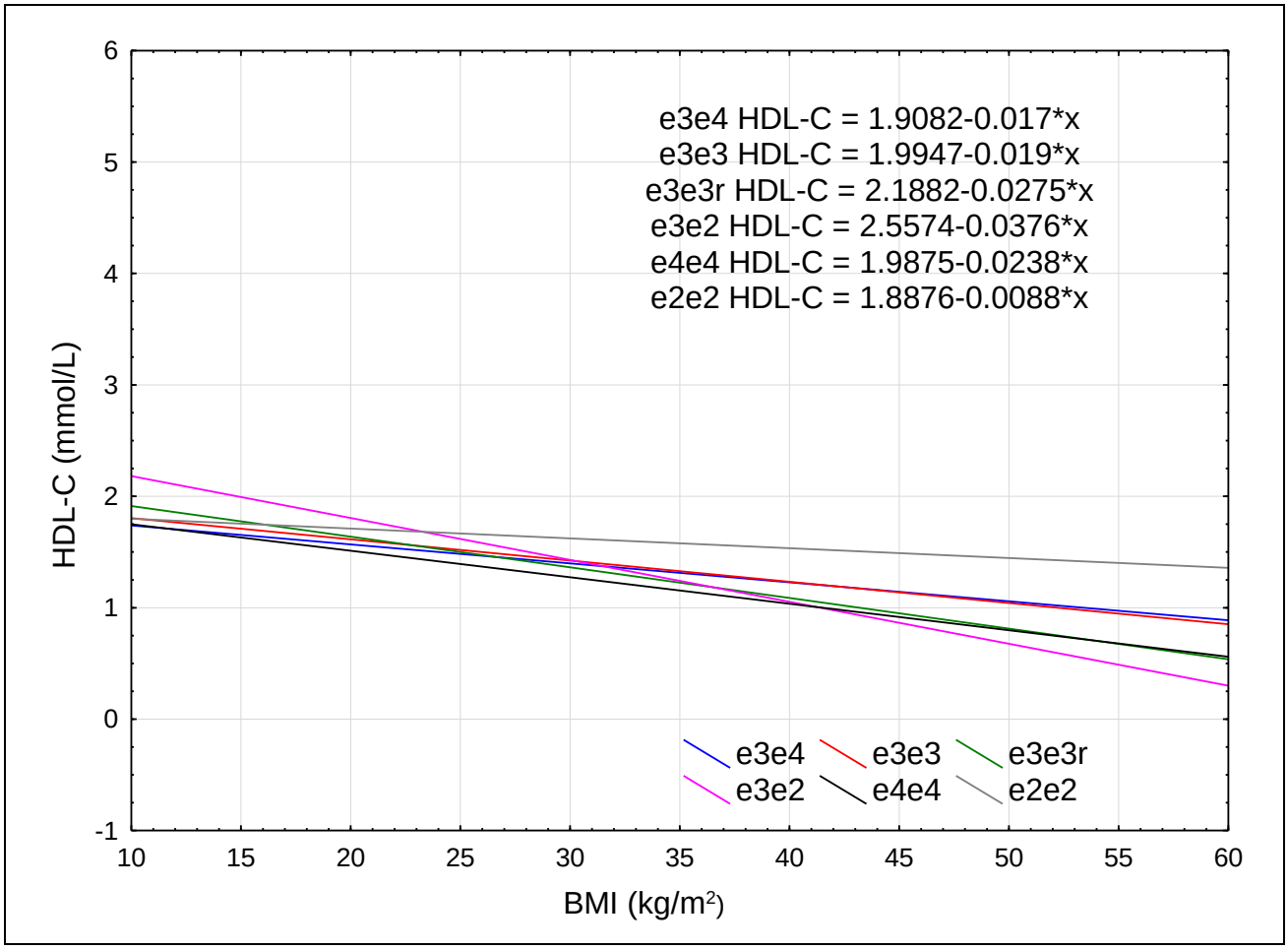
A number of researchers have reported a stronger positive association between HDL-C concentrations and alcohol consumption in *Apoε2* carriers compared to the other haplotype carriers (Corella *et al*., 2011; Lu *et al.,* 2010), but inconsistent results were reported regarding the statistical significance of this interaction over a follow-up period and its effect on CHD risk (Corella *et al*., 2011; Lu *et al.,* 2010). However, this interaction was not observed in the present cohort.

Two genotype-phenotype interactions were observed for the *ApoE* haplotype-genotypes with BMI ($p = 0.02$) and PAI-1_{act} concentrations ($p < 0.004$) on HDL-C and TG concentrations (Figure 5.40 and Figure 5.41). The univariate tests (factorial ANOVA) indicated that a significant correlation exists between BMI and HDL-C concentrations ($p = 0.00$), as expected, and a significant association between the *ApoE*

haplotype-genotypes and HDL-C was also observed ($p = 0.002$). As mentioned before, the inverse relationship between BMI and HDL-C is well known (Section 5.5.5). According to Figure 5.40, this relationship seems to be prevalent among all the haplotype-genotypes as well, except for the ε2/ε2 group, for which the regression line is more level, compared to the other haplotype-genotypes' regression lines. Therefore, the favourable ε2/ε2 genotype possibly plays a protective role by preventing decreasing HDL-C concentrations with increasing BMI. Furthermore, the mean HDL-C concentrations among ε4 carriers were significantly lower than those among non-carriers and are in accordance with those reported in a recently published study by Rahilly-Tierney *et al*. (2011). However, the aforementioned study, along with other investigations, has reported no significant interaction between the *ApoE* genotype and the continuous BMI variable on HDL-C concentrations (Srinivasan *et al.,* 2001; Lussier-Cacan *et al.,* 2000). It was hypothesised that the ε4 carrier status modified the well-established association between increasing BMI and decreasing HDL-C concentrations, but this hypothesis has been rejected (Rahilly-Tierney *et al.,* 2011).

The anticipated mechanism for the interactions presented in Figure 5.39 and 5.40 is that these environmental (dietary cholesterol intake) and phenotypic factors (adiposity) possibly enhance the reported mechanistic role of the ApoE protein, synthesised by the liver as part of VLDL particles. During cholesterol transport, the ApoE protein serves as a ligand for several processes that cycle TG from VLDL into HDL subtypes. The ε4 isoform of ApoE is less effective in binding to lipoprotein receptors than the ε2 or ε3 isoforms, so in individuals with the *ApoE* ε4 haplotype, the pathway between TG-rich VLDL particles and small, denser, more easily metabolised HDL particles is less effective (Mahley & Rall, 2000). Further research would be necessary to support any speculation on the metabolic pathway between BMI, *ApoE*, and HDL-C concentrations. However, the findings reported here reject the suggested mechanism that the effectiveness of the ApoE (protein) as a ligand decreases, as the pathway between VLDL and HDL subtypes become overwhelmed by the TG-rich environment owing to adiposity (Rahilly-Tierney *et al*., 2011). In fact, it can rather be suggested that the binding efficacy of the *ApoE* ε4 is stronger in the present black South African cohort, compared to that observed in Caucasians, which could result in more effective lipid metabolism (even though a small increase in LDL-C has occurred) which in turn can be further enhanced by the observed gene-environment and genotype-phenotype interactions (Figure 5.39 and 5.40).

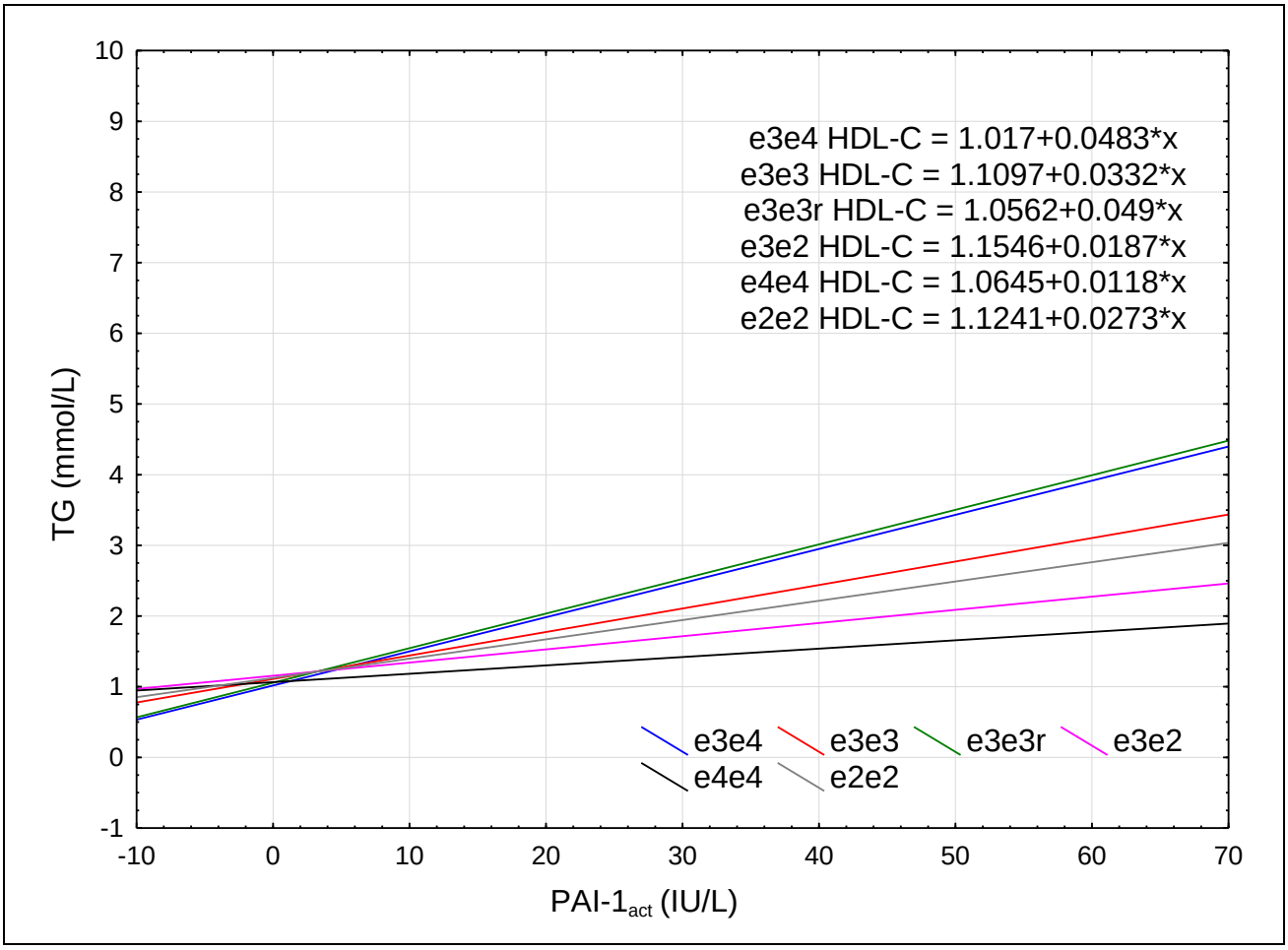
Figure 5.40 The interactive effect of BMI with the *ApoE* haplotype-genotypes on HDL-C concentrations for each of the genotype groups



% = percentage; ε = epsilon; BMI = body mass index; HDL-C = high-density lipoprotein cholesterol

Cummings (1998) were the first to report the interaction between PAI-1_{act} and the *ApoE* genotype. These researchers reported that the ε4 allele was not only associated with the high CVD risk, but carriers of this allele were determined to have higher PAI-1 activity compared to the low-risk ε2/ε3 group (Cummings, 1998). These findings proposed a new mechanism for the association between the presence of the ε4 allele and increased CVD risk. Furthermore, elevated TG concentrations have been suggested to be atherogenic by themselves and are associated with increased PAI-1_{act} concentrations (Cummings, 1998). The literature therefore supports the researcher's observation that significant PAI-1_{act}-*ApoE* genotype interaction occurs in relation to TG (Figure 5.41). However, as the *post hoc* tests were not significant, it indicates that these interactions should be interpreted with care and that further powered studies should confirm the true impact of these interactive effects.

Figure 5.41 The interactive effect of PAI-1$_{act}$ with the *ApoE* haplotype-genotypes on TG concentrations for each of the genotype groups



ε = epsilon; HDL-C = high-density lipoprotein cholesterol; PAI-1 = plasminogen activator inhibitor type 1; TG = triglycerides

5.6 SUMMARY OF THE RESULTS

The results obtained in this study will be summarised and discussed in line with the specific objectives of this research project (Section 1.2), in order to determine to what extent they were met. The results reported were based on five SNPs and one set of haplotype-genotype combinations and their associations with the lipid profile determinants. Initially nine SNPs were included in the final SNP list to be genotyped, of which one SNP (novel) failed to be genotyped and three SNPs were excluded from the study because of validation and quality issues pertaining to the BeadXpress® platform (Section 5.3). Nonetheless, such effects are common in experiments based on medium throughput genotyping analyses and the present study remains the largest study to investigate the effect of polymorphisms in the *ApoE* gene on the lipid profile of black South Africans.

All the SNPs investigated adhered to the assumptions of the HWE, but the haplotype-genotype distribution did not adhere to the assumptions of HWE, possibly as a result of positive selection. However, as the frequency distributions of these

haplotype-genotype combinations were within the 95% CI ranges, it is most likely that it was not a result of sampling (selection) bias. Furthermore, the MAF distributions differed between those reported for the present black South African cohort and other ethnic groups as discussed per investigated SNP in Section 5.5, which suggests that the present cohort under investigation presents with a unique genetic profile. Even though the novel alterations identified in the present cohort could not be successfully genotyped other novel alterations were unexpectedly observed in the present cohort, *i.e*. the fourth *rare allele* (ε3r). Even though negative results were reported for the rs1081101 SNP (Section 5.5.1; Table 5.9), these results are novel, as no prior clinical significance has been reported for this SNP before. The investigation of all of these aforementioned aspects addressed the first specific objective of the study, which was to determine the genotype distribution of various reported as well as novel alterations in the *ApoE* gene in a black South African population.

The second specific objective was to assess whether the identified polymorphisms in the *ApoE* gene are associated with the lipid profile of a black South African population. The well-reported rs429358 and rs7412 SNPs (as the constituent SNPs of the haplotype-genotypes) presented with the strongest associations with various components of the blood lipid profile in the black South African cohort under investigation. The variety of associations between these particular SNPs and the blood lipid profile determined in the present cohort strongly indicates that it is integral to any public health investigation into CVD development in the present cohort group.

The last three specific objectives of this research study can be addressed together, since the discussion and interpretation of one interactive effect (*i.e*. gene-environment, gene-phenotype and gene-diet) on the lipid profile are often not meaningful without the others. From Section 5.5 it is clear that a number of significant interactive effects between various SNPs and environmental, phenotypic and dietary factors were present in relation to the blood lipid profile in this cohort. However, after *post hoc* tests had been conducted, the significance of the majority of these interactions was lost. This can possibly be ascribed to inadequate power of the present sample size per genotype group. Therefore, future studies with greater power are needed to confirm whether these results have a true effect on the lipid profile of black South Africans. Nonetheless, the interaction effect between various environmental, phenotypic and dietary traits on the lipid profile in the presence of certain *ApoE* genetic variants that were determined and for which significant

results were obtained are summarised in Table 5.20 and will be discussed in the following paragraphs.

Two gene-environment interaction effects on TG remained significant after *post hoc* tests had been conducted. From the results it was evident that individuals residing in an urbanised environment and harbouring the heterozygote genotype for the rs405509 SNP had an increased risk of elevated TG concentrations. The reason for this is that these individuals are reported to consume a more westernised diet that is high in dietary fat compared to a more traditional (rural) diet and it could be suggested that these subjects are unable to metabolise dietary TG as well, which results in elevated serum TG concentrations. Therefore, increased dietary fat intake among these individuals might aggravate the existing adverse dietary effects of an urbanised diet on the lipid profile. The same implications apply to the rs7412-locality interaction effect on TG concentrations as urban dwellers harbouring the wild-type homozygous genotype have significantly higher TG concentrations than individuals harbouring the same genotype, but residing in a rural area.

Whether the aforementioned results are of significant public health concern might be highlighted if the percentages of the total population that harbour the heterozygote genotype for the rs405509 SNP and reside in an urban area (15%) or harbour the homozygote wild-type genotype for the rs7412 SNP and reside in an urban area (36%) are considered. From these percentages it might not seem particularly remarkable; however, these percentages double if calculated according to the total number of individuals who are urbanised and harbour the aforementioned SNPs.

Table 5.20 Summary of all the significant results observed in thepresent study

	TC	LDL-C	HDL-C	TG
Associations				
Whole group	rs7412 haplotype-genotype	rs429358 rs7412 haplotype-genotype	rs7412 haplotype-genotype	---
Men	rs7412 haplotype-genotype	rs7412 haplotype-genotype	rs7412	---
Women	rs7412 haplotype-genotype	rs429358 rs7412	rs7412 haplotype-genotype	---
Rural	rs7412 haplotype-genotype	rs7412 haplotype-genotype	---	rs7412
Urban	rs7412 haplotype-genotype	rs429358 rs7412 haplotype-genotype	haplotype-genotype	---
Interactions				
Gene - environment	---	---	---	rs405509 – locality* rs429358 - locality rs7412 – locality*
Gene – diet	---	---	rs7412-GGT* Haplotype.genotype – cholesterol	rs429358-MUFA rs429358-SFA rs429358-TFA rs7412-MUFA rs7412-TFA rs7412-fibre rs7412-GGT
Gene - phenotype	---	---	rs7412-BMI* haplotype-genotype – BMI	rs1081101- PAI-1_{act} rs429358 - PAI-1_{act} Haplotype.genotype - PAI-1_{act}

* = significant interactions that remained after *post hoc* tests were conducted; BMI = body mass index; GGT = gamma-glutamyl transferase; HDL-C = high-density lipoproteins; LDL-C = low-density lipoprotein cholesterol; MUFA = mono-unsaturated fatty acids; PAI-1_{act} = plasminogen activator inhibitor type 1; PUFA = poly-unsaturated fatty acids; rs = reference sequence; TC = total cholesterol; TFA = trans fatty acids; TG = triglycerides; tHcy = total homocysteine

Significant interactions were determined between various *ApoE* genetic variants and a number of phenotypic traits that were expected (BMI and PAI-1_{act}). The fact that the interaction between PAI-1_{act} and the respective genetic alterations on TG concentrations was determined for more than one SNP calls for further investigation, despite the insignificant *post hoc* tests determined for these interactions. PAI-1 concentrations have been reported to increase with urbanisation in black South Africans, possibly owing to the accompanying adverse dietary changes, which result in a deteriorating lipid profile (Pieters & Vorster, 2008; Greyling *et al*., 2007). Since three genetic variants have been determined to have an interactive effect with PAI-1_{act} concentrations on TG concentrations (Table 5.20), it is suggested that further functional analysis should be conducted to determine whether these interactive effects with these genetic variants are causative of the increased PAI-1 concentrations among black South Africans.

The rs7412 SNP had a significant interactive effect with BMI on HDL-C concentrations (Table 5.20). Even though the interactive responses of individuals carrying the CC genotype, as well as those carrying the CT genotype, and BMI on HDL-C concentrations were similar, the individuals harbouring the homozygote wild type genotype (CC) are at increased risk of lowered HDL-C concentrations as they become more overweight. Furthermore, the wild-type genotype presents with a high-frequency distribution in the present cohort. The implication of this result is that 93% of individuals harbouring this genotype in the present cohort have increased susceptibility to reduced HDL-C concentrations as they become overweight, as discussed before. Similar results were observed for the rs7412-GGT interaction effect on HDL-C concentrations in that the CC carriers had significantly lower HDL-C concentrations than the CT carriers. A possible suggestion might be that the favourable effects of moderate alcohol consumption on HDL-C for CC carriers are not as great as for the carriers of the CT genotype or the general public. However, this specific result should be interpreted with care, as the GGT biomarker is rather used to verify alcohol abuse or even liver damage and not necessarily alcohol consumption itself.

With the focus of the present study mainly on nutrigenetics, an important result to highlight was the rs429358-MUFA interaction effect on TG concentrations, which might indicate that carriers of the CC genotype at the rs429358 locus should consume dietary PUFAs rather than dietary MUFAs to reduce TG concentrations. Even though the QFFQ used was verified for this population, the reported dietary fat intake should be interpreted with care and these diet-gene interactions should be analysed with the use of the blood values of the dietary data, as this might give more accurate results. However, this was beyond the scope of this dissertation.

To conclude, the significant gene-locality interaction effect on the lipid profile supports the scientific evidence indicating that urbanisation, along with other lifestyle changes, could possibly explain the hastening and increasing incidence rates of metabolic and vascular dysfunction observed in certain ethnic groups (Forouhi & Sattar, 2006; Steyn *et al*., 2005; Yusuf *et al*., 2001a; Walker & Sareli, 1997; Steyn *et al*., 1991). Furthermore, the significant genotype-phenotype interaction effects on the lipid profile discussed might be an approach to reduce the increased risk of CVD among the individuals harbouring these specific genotypes, as they are receptive to modifiable factors (*i.e*. alcohol intake/GGT, BMI) being altered (Haan & Mayeda, 2010; Jofre-Monseny *et al*., 2008). Therefore, from these results it is evident that the CVD biomarker under investigation, *i.e*. the lipid profile of

black South Africans, is most definitely influenced by not only genetic variations in the *ApoE* gene and certain environmental factors, but by the interaction between these factors as well. However, it is not certain whether the significant results observed will withstand other factors, such as gene-gene interactions, since it is reported that the effect of one genetic alteration can be altered by another along the gene. The contributions and limitations of the present study will be discussed in the concluding chapter to follow.

CHAPTER SIX

CONCLUSION

One of the most recognised multi-factorial diseases is CVD, given that no one single (risk) factor is causative of the development of this disease. The interactive effect between various lifestyle and phenotypic factors with genetic factors could influence susceptibility to developing CVD among individuals and ultimately populations. The lipid profile is a central biomarker in the development of CVD and the *ApoE* gene is the best-studied candidate gene. Therefore, the main aim of this study was to investigate the associations between specific *ApoE* SNPs and the lipid profile of a black South African population, taking into account certain environmental and phenotypic factors in order to explore the interaction effect between these variables.

The fact that this study was conducted using a multi-factorial approach and a large sample size (n = 1 458 with genotyping data) of black South African individuals, adds to the novelty of the research presented. Furthermore, accurate estimates of the frequency of each of the genetic variants that was investigated were made possible by to the large sample size of this study. When compared to the literature, the frequency distributions of the investigated genetic variants in the present cohort differed not only from non-African populations, but from different African populations as well. This emphasises how genetically diverse human beings truly are and that any genetic associations that have been reported in one ethnic group cannot be assumed to be present or significant on a population scale. Furthermore, most genetic association studies are conducted in non-African populations, but the internal diversity between African populations is much greater and the need for genetic association studies to be conducted in African populations is therefore more pronounced. Since it has frequently been reported that CVD risk differs between ethnic groups, it is of relevance to investigate not only the genetic factors, but also the environmental factors along with the interaction effect of these factors that contribute to the development of CVD in different populations.

In the present study, the results indicated that the rs405509 and rs7412 SNPs had a true significant interactive effect with the environment (area of residence) on TG concentrations. From these results it could be suggested to advocate that individuals

residing in an urbanised environment and harbouring the heterozygote genotype for the rs405509 SNP should be particularly conscious of their dietary fat intake and adhere to the dietary recommendations made to the general public, *i.e.* to lower dietary (saturated) fat intake for an improved lipid profile. The same recommendations can be advocated for urban dwellers harbouring the wild-type genotype for the rs7412 SNP. Since CVD is a multi-factorial disease, the prevalence of these gene-environment interactions in the present cohort was not unexpected.

The rs7412 SNP had a significant interactive effect with BMI as well as GGT on HDL-C and it could be hypothesised that these interactions might possibly explain the lipid variability observed in black African populations to a certain extent. Even though it seems as if this SNP might be important in a public health context pertaining to weight management, the health recommendation for the individuals harbouring the wild type genotype for the rs7412 SNP will be the same as it would be to the general population, *i.e*. to be aware of the health implications of being overweight. For this reason, it would not be wise to limit this recommendation to those harbouring the wildtype genotype for the rs7412 SNP, but to merely emphasise this recommendation among the broader public.

Definitive gene-diet interactions (nutrigenetics) were observed, suggestive of the fact that lifestyle approaches such as increasing fibre and MUFA intake and reducing the intake of alcohol, dietary cholesterol, dietary SFA and TFA could to some extent negate the negative effect that certain *ApoE* genotypes have on the lipid profile. Therefore, it is important to highlight that even if an individual harbours a favourable *ApoE* genetic variant, the necessary lifestyle modifications associated with reduced CHD risk should still be implemented and advocated, since environmental (risk) factors remain the greatest contributor to disease development. Furthermore, these gene-diet interaction effects could possibly provide valuable information on the aetiology of CVD.

One should be cautious to imply that these observed associations and interactions regarding the lipid profile are of biological relevance, as it is possible that the matter might be more complex than initially hypothesised (Marques-Vidal *et al*., 2003). It is difficult to interpret an uncomplicated benefit-risk model in the field of nutrigenetics because of the conflicting interactions between different genes (Ordovas, 2006). Knowledge and understanding of these gene-gene interactions have expanded over the last few years in the field of nutritional science; however, a lot is still to be discovered and is still not fully understood. Furthermore, various other interaction types *i.e*. direct interactions, epigenetic

interactions and genetic variations, could partially describe nutrient-gene links observed and should be considered when interpreting interactive effects.

The fact that results that were determined in the present study conflict with previously reported ones is a common occurrence in genetic association studies (Chang *et al*., 2010). These inconsistencies might be due to a number of factors, *i.e*. the variation in study design, the ethnicity of the population and the geographic region of the population; different statistical methods used and the power thereof (especially when certain genotypes are rare) and lastly the modifiable effect of gene-environment interactions. The observed differences in associations observed between different ethnicities are more likely to be due to differences in allele frequencies or haplotype structure than differences in the real effect (Lanktree *et al.,* 2009).

A fundamental issue when conducting multiple statistical analyses is that the results that are obtained or determined may be due to an increased risk of type 1 errors. The reason why a large number of interaction effects were tested for between each of the genetic variants of the *ApoE* gene under investigation and many other variants was to remove the possibility of omitting interactions that could have explained certain content-dependent effects. In addition to the statistical p-values obtained per interaction, *post hoc* tests were conducted and 2D scatterplots were graphed to inspect the interactions visually. This information for each of the significant interactions determined was evaluated for biological relevance and consistency with literature to determine whether the observed interactions were plausible in the present black South African cohort.

Another factor that should also be taken into account and considered for future (powered) exploratory studies is the option of other genetically defined confounders such as gene-therapy interactions, gene-gene interactions, LD related aspects, epigenetic mechanisms and pleiotropic genes. These might contribute to the understanding of the CVD burden associated with *ApoE* and might be as insightful as the aforementioned gene-environment interactions observed in the present black South African cohort. Furthermore, the results reported in the present cross-sectional study should not only be tested again with the follow-up data of the larger PURE prospective dataset, which has recently been made available, but also be replicated in other ethnic groups in an interventional study design as well. Since the PURE study is a prospective longitudinal study, it is recommended that future analyses determine whether the significant reported

associations between the identified *ApoE* genetic variants with the blood lipid profile contribute to the CVD incidence rates in a black South African population.

The results reported here contribute to the existing literature, which is mainly focused on the well-known *ApoE* epsilon data rather than the SNPs in the *ApoE* gene and reported for non-African groups. The *rare ApoE allele* (ε3r) was detected in the present cohort and although a possible explanation for the formation of ε3r has been proposed, an explanation for the rarity and the physiological role of this allele remains unavailable to date. The present study addressed some of the gaps in the literature, as it is to the knowledge of the researcher the first study to report the presence of this *rare allele* (although its frequency is still low) in a population rather than in single individuals as previously reported. The present study corroborates previous statements that the *ApoE* polymorphism is potentially one of the most important genetic predictors of plasma lipoprotein concentrations and, therefore, risk of CVD. This study further contributes to the growing body of literature by identifying key determinants of plasma lipoprotein concentrations, *i.e*. urbanisation, alcohol consumption, certain dietary factors and BMI, and provides some insight into the biological mechanisms underlying serum lipid and cholesterol concentrations in a black South African population.

In conclusion, the specific objectives of the present study have been addressed, but many questions on the role and effect of the genetic variants in the *ApoE* gene investigated on CVD development in the cohort under investigation, remains to be answered. It is very important that the nutritional genetic results that have been reported advance to nutrigenomic-based studies in order to translate these observational results into molecular mechanisms. Therefore, it is necessary to apply more robust methods in order to obtain more robust data. The outcomes of applying this genomic approach in future nutritional research could be far-reaching, with the possibility of developing dietary and lifestyle recommendations on a population level, which could in turn have an effect on major public health issues, *i.e.* obesity and NCD. However, to achieve this, a multidisciplinary approach needs to be followed and the nutritional research field would need to break away from the more traditional methods used.

ADDENDUM A

1 DATA COLLECTION

During the first data collection phase, various forms of data were gathered through different data collection methods. Standardised, interviewer-based questionnaires were used to capture various characteristics, *i.e.* demographic information and dietary composition. Specific experimental assays (discussed in this Addendum A: Section 1.2) were also conducted, which included blood sample collection along with various analyses measuring CVD risk biomarkers and anthropometrical variables. Methods pertaining to genetic analysis are discussed in Chapter four, Section 4.3.

1.1 Questionnaires

Trained fieldworkers interviewed the eligible participants in their native language in order to collect information *via* the PURE questionnaires. The data that were relevant to the present study, were recorded for each individual and included age, gender, location of inhabitance (rural or urban), ethnicity, medical history (*inter alia*: a history of CLDs diagnosed by a doctor and HIV status) and lifestyle behaviour (*inter alia*: history of tobacco use, alcohol consumption and dietary intake).

1.1.1 Quantitative food frequency questionnaire

Validated quantitative food frequency questionnaires were completed during interviews using food portion books designed and standardised for use in the black South African population, to determine the dietary intake of participants (MacIntyre *et al*., 2001). The dietary data were computerised by using the *Foodfinder3*® program [Medical Research Council (MRC), Tygerberg, 2007], after which the data were sent to the MRC of South Africa for nutrient analyses. The dietary reference intake (DRI) for fat, according to the acceptable macronutrient distribution range (AMDR), is 20-35% of the TE intake (from mostly PUFA and MUFA sources). Therefore, a low-fat diet was defined as one in which less than 20% of the TE intake was derived from fat, whereas a high-fat diet was defined as a diet in which more than 35% of the TE intake was derived from fat (Vorster & Nel, 2001; Wolmarans & Oosthuizen, 2001).

1.2 Measurements of phenotypic characteristics

This section describes the experimental methods used to determine the phenotypic characteristics of the PURE subjects. Laboratory processes and anthropometric techniques will be described in further detail. Subjects were requested to fast for at least eight hours before their arrival at the site of testing.

1.2.1 Blood collection

Registered nurses were responsible for the collection of the blood samples. Whole blood was drawn from the antecubital vein, using a disposable needle, while the patients were seated, and with free blood flow to ensure minimal trauma. Each collection tube (vacutainer) was filled to its capacity. Mixing of the contents of the tubes was achieved through gentle inversion. The procedure for the collection of serum was that the blood was allowed to clot in a serum tube at room temperature for 30 minutes, after which it was centrifuged at 2 000 x gravitational force (*g*) for 15 minutes at 10 degrees centigrade (°C), within two hours after collection. For the plasma samples the blood was collected in ethylene diamine tetra-acetic acid (EDTA) tubes and centrifuged at 2 000 x *g* for 15 minutes at 4°C, within two hours of collection. Both the resultant serum and plasma samples were aliquotted and stored at -70°C until analyses were conducted. The buffy coat collection (for DNA isolation purposes) entailed centrifuging citrate-treated whole blood at 2 000 x *g* for 15 minutes at room temperature, after which the leukocyte layer was aliquotted and stored in a storage tube at -70°C until DNA was extracted.

1.2.2 Measurement of total, low-density lipoprotein and high-density lipoprotein cholesterol

The Konelab20i™ auto analyser was used for quantitative determination of serum HDL-C concentrations (Thermo Fisher Scientific Oy, Vantaa, Finland). It is a clinical chemistry analyser based on the principle of homogeneous enzymatic colourimetric, immunoturbidimetric and ion-selective electrodes (Gordon *et al*., 1977). Water-soluble complexes containing LDL-C, VLDL and CM are formed in the presence of magnesium ions and dextran sulphate, but are resistant to polyethylene glycol (PEG) modified enzymes. At that point, the HDL-C concentration is measured enzymatically by cholesterol esterase and cholesterol oxidase coupled with PEG to the amino groups. Cholesterol esterase breaks the cholesterol esters down quantitatively into free cholesterol and fatty acids. Furthermore, in the presence of oxygen, cholesterol is oxidised by

cholesterol oxidase to form delta (Δ) 4 cholestenone and hydrogen peroxide. A blue pigment is formed when hydrogen peroxide, in conjunction with aminoantipyrine, sodium N (2-hydro-3-sulfopropyl)-3,5-dimethoxyaniline, hydrogen and water, reacts with peroxidase. The HDL-C concentration is related to the colour intensity of the blue quinoneimine dye, which is measured at an absorbance of 583 nm.

The Konelab20i™ auto analyser was also used to analyse TC *via* oxidase-peroxidase treatment with phenol aminoantipyrine reagents (Thermo Fisher Scientific, Vantaa, Finland). Cholesterol esterase hydrolyses the cholesterol esters to cholesterol and free fatty acids, after which cholesterol oxidase oxidises the free cholesterol to cholest-4-en-3-one and hydrogen peroxide. The hydrogen peroxide combines with hydroxybenzoic acid and 4-aminoantipyrine to form a chromophore (quinoneimine dye) which is measured at 500-550 nm (Allain *et al*., 1974).

The LDL-C particles were not measured directly, but rather estimates were determined using the Friedewald equation. This was done by subtracting the amount of cholesterol associated with other particles, such as HDL-C, assuming a prolonged fasting state in $mmol.L^{-1}$ (LDL-C = TC – HDL-C – TG/2.2), provided the TG did not exceed 4.5 $mmol.L^{-1}$ (Klug, 2012).

1.2.3 Measurement of triglyceride concentrations

The Sequential Multiple Analyser Computer (SMAC) method was applied to measure the TG concentrations, using the Konelab20i™ auto analyser. The TGs were hydrolysed to fatty acids and glycerol, whereupon the glycerols were phosphorylated to glycerol-3-phosphate, which was oxidised to dihydroxyacetone phosphate and hydrogen peroxide. Once the formed hydrogen peroxide reacts with the 4-aminoantipyrine and 4-chlorophenol, a quinoneimine dye is formed. The absorbance is measured at 510 nm, after which the TG concentration is calculated by using a calibration curve (Thermo Electron Corporation, Vantaa, Finland).

1.2.4 Measurement of plasminogen activator inhibitor type-1

A chromogenic assay kit was used for the determination of plasminogen activator inhibitor type-1 (PAI-1). A fixed amount of tissue plasminogen activator (t-PA) was added to a plasma sample and then acidified to destroy potential plasmin inhibitors, and subsequently

diluted. A mixture of Glu-plasminogen, poly-D-lysine and chromogenic substrate was added. The quantity of active t-PA remaining in the sample catalysed the conversion of plasminogen to plasmin, which hydrolysed the chromogenic substrate. The amount of colour that developed was relative to the amount of active t-PA in the sample. The PAI-1 activity ($PAI\text{-}1_{act}$) content of the sample was the difference between the amount of t-PA added to the sample and the amount of t-PA recovered (Spectrolyse pL PAI-1, Biopool, Trinity Biotech, Ireland).

1.2.5 Measurement of total homocysteine concentrations

Venous EDTA treated plasma samples were used to determine the total homocysteine (Hcy) concentrations (which included the sum of free and protein-bound Hcy, homocystine and Hcy-cysteine mixed disulphide) on the Abbott automated immunoassay analyser (AxSYM) by an external pathology company. The determination of tHcy concentration was based on fluorescence polarisation immunoassay- (FPIA) –based technology with the CV for all assays equal to 4.52%. The procedure required that Hcy, mixed disulphide and protein-bound forms of Hcy in the sample were reduced to form free thiol by the use of dithiothreitol. Thereafter, the free Hcy was enzymatically converted to S-adenosyl-L-Hcy (SAH) by the use of SAH hydrolase and excess adenosine. SAH hydrolase converts SAH to Hcy, under physiological conditions. The excess adenosine (in the pretreatment solution) initiates the conversion of Hcy to SAH by the use of the recombinant SAH hydrolase (Sundrehagen E/Axis Biochemicals ASA). Thereafter, specific monoclonal antibodies and the fluoresceinated SAH analog tracer start the FPIA detection system from which the tHcy concentrations are calculated by the created calibration curve.

1.2.6 Measurement of gamma-glutamyl transferase as the biochemical marker of alcohol intake

The Konelab™ auto analyser (Thermo Fisher Scientific Oy, Vantaa, Finland), was used and the SMAC method was applied to measure the serum gamma-glutamyl transferase (GGT) concentrations. The principle of this kit is that the transfer of glutamic acid to acceptors such as glycylglycine is catalysed by GGT. During this process 5-amino-2-nitrobenzoate, which absorbs light at a wavelength of 405 nm, is released and the increase in absorbance is directly related to GGT activity (Thermo Electron Corporation, Vantaa, Finland). The cut-off values used for men and women were 80 $U.L^{-1}$ and 50 $U.L^{-1}$, respectively.

1.2.7 Blood pressure

The OMRON HEM-757 automatic digital blood pressure monitor from Omron Healthcare (Kyoto, Japan) was used to measure the blood pressure (BP) of the participants using the brachial artery. While the measurement was taken, the participants were seated upright with their right arm supported at heart level and the exact readings for systolic BP and diastolic BP were reported in millimetres of mercury (mm Hg). The participants had to be rested, relaxed and calm for five minutes and should not have climbed stairs in the 15 to 30 minutes before the measurement was taken. The participants should also not have smoked, exercised or eaten in the last 30 minutes before the measurement was taken.

1.3 Determination of human immunodeficiency virus status

Subjects who gave informed consent for HIV testing were also given an option to be informed of the results obtained. Each of the tests was done according to the protocol stipulated by the National Department of Health, South Africa. Each subject received pre-test counselling in groups of ten, as well as individual post-test counselling, for those who wanted to be informed of the results. The presence of HIV antibodies was determined using the rapid First Response® HIV card test 1-2.0 (Transnational Technologies Inc.) and if the results were positive, the test was repeated with the Pareeshak® (BHAT Biotech India) card test.

1.4 Anthropometrical assessment

The subjects, wearing minimal clothing, were measured by qualified level 2 International Society for the Advancement of Kinanthropometry accredited anthropometrists. The height of the participants was measured in metres with a stadiometer (IP 1465, Invicta, London, UK) and body mass was determined using a portable electronic scale (Precision Health Scale, A&D Company, Tokyo, Japan). All anthropometrical measurements were taken twice, unless a difference of larger than 5% occurred between the first two measurements, when a third measurement was taken to assure accuracy. BMI was calculated by dividing their body mass (kilograms) by their height squared (metres^2). A BMI below 18.5 kg.m^{-2} was interpreted as being underweight, between 18.5-24.9 kg/m^{-2} to be within the normal range, between 25.0-29.9 kg/m^{-2} to be overweight and ≥30 kg/m^{-2} as obese.

ADDENDUM B

PURE-SA Project

INFORMED CONSENT FORM (PHASE 1)

I, the undersigned ……(full names) understand that the only information that will be asked from me is the family census and household questionnaires. I understand that a field worker from the PURE-study will ask me the questions and that all the information gained from me will be kept confidential.

I indemnify the University, also any employee or student of the University, of any liability against myself, which may arise during the course of the project.

I will not submit any claims against the University regarding personal detrimental effects due to the project, due to negligence by the University, its employees or students, or any other subjects.

……..

(Signature of the subject)

Signed at ..... on ……..

Witnesses

1. .....

2. ….....

Signed at ..... on ……

ADDENDUM C

PURE-SA project

INFORMATION TO COMMUNITIES

Dear Participant

Thank you for being willing to help us in this very important project. We are sure that the project will contribute to improve health of all the people of the North West Province.

The aim of the project is to get enough information regarding the development of chronic diseases like Diabetes, Stroke, Lung disease and heart disease with urbanisation to plan appropriate health and nutrition intervention strategies.

For this study we need 2 000 subjects whom we can follow for 12 years. The baseline survey will be done from April 2005 to November 2005. The subjects must be from rural as well as urban communities. Therefore, 500 subjects from 4 different levels of urbanisation will be needed. Ganyesa and Tlakgameng were chosen for the rural and semi-rural areas because they are still under tribal law with a good infra structure and stability. We also spoke to Chief M. Letlhogile and the mayor Mr. E. Tladinyane and both gentlemen gave us permission to do the research in these two communities. Ikageng and the informal Ikageng were chosen as they are convenient and near the University. Cllr GG Megalanyane and Cllr Mahesh Roopa are informed about the study.

All the questionnaires will be filled out at your houses by trained research field workers who are from your communities. After a household survey and a family census on most of the households in your community to give us an overview of the total community, 250 men and 250 women from all four sites (Ganyesa, Tlakgameng, Ikageng, and the Informal Ikageng) will be asked to proceed with the study. These subjects should be:

- Older than 35 years
- Healthy – which means that they must not be aware of any disease and do not take any chronic medication.

These 2 000 subjects will be asked to fill out the adult questionnaire, the food frequency questionnaire, the health questionnaire and the physical activity questionnaire. We will also make an appointment with each subject to take some measurements such as weight, height, skinfold thicknesses, ECG (test for heart abnormalities), lung functions, blood pressure, blood glucose, blood samples and a urine sample.

It is very important that we gather quality data and knowledge. Because HIV/AIDS is such a devastating illness and affects almost all aspects of health, it is necessary to know if HIV is absent before we analyse the data. Therefore, we will ask questions about your HIV status which you are allowed not to answer.

It is also very important to us that you feel free to participate in this study and that you understand what the study is all about. The fieldworker will ask you to sign this form after you have read and understood it.

Kind regards

Dr ANNAMARIE KRUGER

Contact details: 082 7715778 / 018 2994037(W) / 018 2907024(H)

ADDENDUM D

PURE-SA Project

INFORMED CONSENT FORM (Phase 2)

I, the undersigned ……(full names) read / listened to the information on the project in PART 1 and PART 2 of this document and I declare that I understand the information. I had the opportunity to discuss aspects of the project with the project leader and I declare that I participate in the project as a volunteer. I hereby give my consent to be a subject in this project.

I agree to be tested for HIV **.....**	Yes	No
I want to know my HIV-status **.....**	Yes	No
I agree to give a blood sample **.....**	Yes	No

I hereby also declare that I am aware that:

1. this blood sample will be used for the purpose of
    a. Isolating DNA to look at genetic factors that are currently associated with Type 2 Diabetes (i.e. the Calpain10, Adiponectin, Leptin and Leptin Receptor genes), or genetic factors that may be associated with non-communicable diseases in the future. We give the assurance that all genetic tests and experiments will only focus on genotypes suspected to contribute to an increased risk of non-communicable diseases of lifestyle.
    b. Testing for liver function by determining liver enzymes such as AST, GGT,
    c. Analyses of other than genetic parameters for Diabetes Mellitus such as HbA_1C, Blood glucose and Insulin
    d. Analyses of clotting factors and hypertension markers
    e. Analyses of bone health, iron and nutrition status
    f. And may be stored until such time as the above measurements/analyses will be done.
2. A two hour glucose tolerance test will be done
3. Body measurements such as height, weight, skinfold thicknesses, arm and leg circumferences will be taken

4. Electrocardiograph be taken
5. Blood pressure to be taken
6. Pulse wave velocity measurements will be made
7. A urine sample to be collected to analyse for the presence of heavy metals such as lead and mercury,
8. A Spirometer test to be performed to determine lung function
9. A handgrip test to be performed to test muscle strength
10. A hair sample to be taken to test for fumonisin mycotoxins.

…….....……..

(Signature of the subject)

Signed at **... Potchefstroom / Ganyesa ...** (delete not applicable option) on **……/……/ 2005**

Witnesses

1. ..... 2. .....……

Signed at **... Potchefstroom / Ganyesa ...** (delete not applicable option) on **……/……/ 2005**

PART 1

1. School/Institute:
   Faculty of Health Sciences, North-West University

2. Title of project/trial:
   PURE: Prospective Urban and Rural Epidemiological study

3. Full names, surname and qualifications of project leader:
   Dr. Annamarie Kruger, Ph.D. (Nutrition)

4. Rank/position of project leader:
   Research Manager

5. Aim of this project

   PURE's aim is that understanding the different lifestyle and health transitions of individuals in response to societal changes will elucidate societal and individual adaptive strategies that could diminish the adverse health effects of industrialisation and urbanisation on health, while retaining its benefits.

6. Explanation of the nature of all procedures, including identification of new procedures:

   Each participant will have to fill in a number of questionnaires (Adult questionnaire, Physical activity questionnaire, Food frequency questionnaire, Health questionnaire) with the help of field workers. A blood and urine sample will be taken. Physical measures will be performed, including anthropometric measures (such as weight, height, and waist circumference), blood pressure, lung capacity and lung volume and an ECG will be performed.

7. Description of the nature of discomfort or hazards of probable permanent consequences for the subjects which may be associated with the project: (Including possible side-effects of and interactions between drugs or radio-active isotopes which may be used.)

   It will take each participant quite a while (about two hours) to complete all the tests and discomfort may be experienced with the taking of blood samples. No measures will have permanent damage or consequences for the participants.

8. Precautions taken to protect the subjects:

The research nurse will be present at all times, and will be responsible for the blood sampling. She is very experienced and has performed these procedures numerous times in previous studies.

9. Description of the benefits which may be expected from this project:

When measures with immediate results are taken, such as blood glucose levels or blood pressure, the information will be communicated to the individual to seek professional help. Since this study is a longitudinal study, subjects that are high at risk will be identified from the dataset and personal feedback will be given.

10. Alternative procedures which may be beneficial to the subjects:

There will be tested for HIV/AIDS, therefore pre-test counselling will be given. If the subject wants to know his/her status and he/she tests positive, post counselling will also be given.

PART 2

To the subject signing the consent:

You are invited to participate in a research project. It is important that you read/listen to and understand the following general principles, which apply to all participants in our research project:

1. Participation in this project is voluntary.
2. It is possible that you personally will not derive any benefit from participation in this project, although the knowledge obtained from the results may be beneficial to other people.
3. You will be free to withdraw from the project at any stage without having to explain the reasons for your withdrawal. However, we would like to request that you would rather not withdraw without a thorough consideration of your decision, since it may have an effect on the statistical reliability of the results of the project.
4. The nature of the project, possible risk factors, factors which may cause discomfort, the expected benefits to the subjects and the known and the most probable permanent consequences which may follow from your participation in this project, are discussed in Part 1 of this document.
5. We encourage you to ask questions at any stage about the project and procedures to the project leader or the personnel, who will readily give more information. They will discuss all procedures with you.
6. The University staff will use standardised procedures and take all possible precaution to protect the subject from risks.
7. All information will be kept CONFIDENTIAL and no personal information will be published without my consent.

Dr ANNAMARIE KRUGER

Contact details: 082 7715778 / 018 2994037(W) / 018 2907024(H)

REFERENCE LIST

Abboud, S., Viiri, L.E., Lütjohann, D., Goebeler, S., Luoto, T., Friedrichs, S., Desfontaines, P., Gazagnes, M., Laloux, P. & Peeters, A. 2008. Associations of apolipoprotein E gene with ischemic stroke and intracranial atherosclerosis. *European journal of human genetics,* 16(8):955-960.

Akinboboye, O., Idris, O. & Akinkugbe, O. 2003. Trends in coronary artery disease and associated risk factors in sub-Saharan Africans. *Journal of human hypertension,* 17(6):381-387.

Albert, M.A., Glynn, R.J., Buring, J. & Ridker, P.M. 2004. C-reactive protein levels among women of various ethnic groups living in the united states (from the women's health study). *The American journal of cardiology,* 93(10):1238-1242.

Alberts, M., Urdal, P., Steyn, K., Stensvold, I., Tverdal, A., Nel, J.H. & Steyn, N.P. 2005. Prevalence of cardiovascular diseases and associated risk factors in a rural black population of South Africa. *European journal of cardiovascular prevention & rehabilitation,* 12(4):347-354.

Allain, C.C., Poon, L.S., Chan, C.S.G., Richmond, W. & Fu, P.C. 1974. Enzymatic determination of total serum cholesterol. *Clinical Chemistry*, 20(4): 470-475.

Alvim, R.O., Freitas, S., Ferreira, N.E., Santos, P., Cunha, R.S., Mill, J.G., Krieger, J.E. & Pereira, A.C. 2010. APOE polymorphism is associated with lipid profile, but not with arterial stiffness in the general population. *Lipids health disease,* 9(128):1-7.

Ambrose, J.A. & Barua, R.S. 2004. The pathophysiology of cigarette smoking and cardiovascular disease: an update. *Journal of the American college of cardiology,* 43(10):1731-1737.

Anand, S.S., Xie, C., Paré, G., Montpetit, A., Rangarajan, S., Mcqueen, M.J., Cordell, H.J., Keavney, B., Yusuf, S. & Hudson, T.J. 2009. Genetic variants associated with myocardial infarction risk factors in over 8000 individuals from five ethnic groups the INTERHEART genetics study. *Circulation: Cardiovascular genetics,* 2(1):16-25.

Anderson, K.M., Odell, P.M., Wilson, P.W. & Kannel, W.B. 1991. Cardiovascular disease risk profiles. *American heart journal,* 121(1):293-298.

Andreotti, G., Menashe, I., Chen, J., Chang, S. C., Rashid, A., Gao, Y. T., & Hsing, A. W. 2009. Genetic determinants of serum lipid levels in Chinese subjects: a population-based study in Shanghai, China. *European journal of epidemiology*, 24(12), 763-774.

Armstrong, P., Lekezwa, B. & Siebrits, K. 2008. Poverty in South Africa: A profile based on recent household surveys. *The university of Stellenbosch: Stellenbosch economic working papers,* 4(08)

Arnett, D.K., Baird, A.E., Barkley, R.A., Basson, C.T., Boerwinkle, E., Ganesh, S.K., Herrington, D.M., Hong, Y., Jaquish, C. & Mcdermott, D.A. 2007. Relevance of genetics and genomics for prevention and treatment of cardiovascular disease. A scientific statement from the American heart association council on epidemiology and prevention, the stroke council, and the functional genomics and translational biology interdisciplinary working group. *Circulation,* 115(22):2878-2901.

Artiga, M.J., Bullido, M.J., Sastre, I., Recuero, M., García, M.A., Aldudo, J., Vázquez, J. & Valdivieso, F. 1998. Allelic polymorphisms in the transcriptional regulatory region of apolipoprotein E gene. *FEBS letters,* 421(2):105-108.

Asplund-Carlson, A., Hamsten, A., Wiman, B. & Carlson, L. 1993. Relationship between plasma plasminogen activator inhibitor-1 activity and VLDL triglyceride concentration, insulin levels and insulin sensitivity: Studies in randomly selected normo-and hypertriglyceridaemic men. *Diabetologia,* 36(9):817-825.

Atkinson, C., Oosthuizen, W., Scollen, S., Loktionov, A., Day, N.E. & Bingham, S.A. 2004.Modest protective effects of isoflavones from a red clover-derived dietary supplement on cardiovascular disease risk factors in perimenopausal women, and evidence of an interaction with ApoE genotype in 49–65 year-old women. *The journal of nutrition,* 134(7):1759-1764.

Austin, M.A., Hokanson, M., & Edwards, K.L. 1998. Hypertriglyceridemia as a cardiovascular risk factor. *The American journal of cardiology,* 81(4):7B-12B.

Bamshad, M. 2005. Genetic influences on health. *The journal of the American medical association,* 294(8):937-946.

Banfi, C., Rise, P., Mussoni, L., Galli, C., & Tremoli, E. 1997. Linoleic acid enhances the secretion of plasminogen activator inhibitor type 1 by HepG2 cells. *Journal of lipid research*, 38(5): 860-869.

Beaglehole, R. 2001. Global cardiovascular disease prevention: Time to get serious. *The Lancet,* 358(9282):661-663.

Belalcazar, L.M., Ballantyne, C.M., Lang, W., Haffner, S.M., Rushing, J., Schwenke, D.C., Pi-Sunyer, F.X. & Tracy, R.P. 2011. Metabolic factors, adipose tissue, and plasminogen activator inhibitor-1 levels in type 2 diabetes findings from the look AHEAD study. *Arteriosclerosis, thrombosis, and vascular biology,* 31(7):1689-1695.

Bellosta, S., Mahley, R., Sanan, D., Murata, J., Newland, D., Taylor, J. & Pitas, R. 1995. Macrophage-specific expression of human apolipoprotein E reduces atherosclerosis in hypercholesterolemic apolipoprotein E-null mice. *Journal of clinical investigation,* 96(5):2170-2179.

Bennet, A.M., Di Angelantonio, E., Ye, Z., Wensley, F., Dahlin, A., Ahlbom, A., Keavney, B., Collins, R., Wiman, B. & De Faire, U. 2007.Association of apolipoprotein E genotypes with lipid levels and coronary risk. *The journal of the American medical association,* 298(11):1300-1311.

Berg, C. M., Lappas, G., Strandhagen, E., Wolk, A., Torén, K., Rosengren, A., & Lissner, L. 2008. Food patterns and cardiovascular disease risk factors: the Swedish INTERGENE research program. *The American journal of clinical nutrition*, 88(2), 289-297.

Bergeron, N., & Havel, R. J. 1996. Prolonged postprandial responses of lipids and apolipoproteins in triglyceride-rich lipoproteins of individuals expressing an apolipoprotein epsilon 4 allele. *Journal of Clinical Investigation*, 97(1), 65.

Bersamin, A., Luick, B.R., King, I.B., Stern, J.S. & Zidenberg-Cherr, S. 2008. Westernizing Diets Influence Fat Intake, Red Blood Cell Fatty Acid Composition, and Health In Remote Alaskan Native Communities in the Center for Alaska Native Health Study. *Journal of the American Dietetic Association,* 108:266-273.

Bhat, T. A., Mir, M. R., Qasim, I., Misra, S. S., & Kirmani, M. A. 2008. Genetic Polymorphism of 5, 10-Methylenetetrahydrofolate Reductase C677T in Kashmiri Population. *Biotechnology*, 7(4).

Bhatt, D.L., Steg, P.G., Ohman, E.M., Hirsch, A.T., Ikeda, Y., Mas, J., Goto, S., Liau, C., Richard, A.J. & Röther, J. 2006. International prevalence, recognition, and treatment of cardiovascular risk factors in outpatients with atherothrombosis. *The journal of the American medical association,* 295(2):180-189.

Bisoendial, R.J., Boekholdt, S.M., Vergeer, M., Stroes, E.S.G. & Kastelein, J.J.P. 2010. C-reactive protein is a mediator of cardiovascular disease. *European heart journal,* 31(17):2087-2091.

Blum, C. B., Deckelbaum, R. J., Witte, L. D., Tall, A. R., & Cornicelli, J. 1982. Role of apolipoprotein E-containing lipoproteins in abetalipoproteinemia. *Journal of Clinical Investigation*, 70(6), 1157.

Böthig, S. 1989. WHO MONICA Project: objectives and design. *International journal of epidemiology*, 18(3), S29.

Bourne, L.T., Lm1bert, E.V. & Steyn, K. 2002. Where does the black population of South Africa stand on the nutrition transition? *Public health nutrition,* 5(1a):157-162.

Braeckman, L., De Bacquer, D., Rosseneu, M. & De Backer, G. 1996. Apolipoprotein E polymorphism in middle-aged Belgian men: Phenotype distribution and relation to serum lipids and lipoproteins. *Atherosclerosis,* 120(1):67-73.

Bazzano, L. A. 2005. *Dietary intake of fruit and vegetables and risk of diabetes mellitus and cardiovascular diseases*. Geneva: WHO.

Brown, L., Rosner, B., Willett, W. W., & Sacks, F. M. 1999. Cholesterol-lowering effects of dietary fiber: a meta-analysis. *The American journal of clinical nutrition*, 69(1), 30-42.

Brown, M.S. & Goldstein, J.L. 2006. Lowering LDL—not only how low, but how long? *Biomedicine,* 311, 1721-1723.

Brouwer, I.A., Wanders, A.J. & Katan, M.B. 2010. Effect of animal and industrial trans fatty acids on HDL and LDL cholesterol levels in humans–a quantitative review. *PLoS one,* 5(3):e9434.

Burchard, E.G., Ziv, E., Coyle, N., Gomez, S.L., Tang, H., Karter, A.J., Mountain, J.L., Pérez-Stable, E.J., Sheppard, D. & Risch, N. 2003. The importance of race and ethnic background in biomedical research and clinical practice. *New England journal of medicine,* 348(12):1170-1175.

Butler, N. 2010. National guidelines at a glance: Hypercholesterolaemia. *Professional nursing today,* 14(5):26-31.

Campalani, E., Allen, M., Fairhurst, D., Young, H., Mendonca, C., Burden, A., Griffiths, C., Crook, M., Barker, J. & Smith, C. 2006.Apolipoprotein E gene polymorphisms are associated with psoriasis but do not determine disease response to acitretin. *British journal of dermatology,* 154(2):345-352.

Campbell, M. C., & Tishkoff, S. A. 2008. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual review of genomics and human genetics*, 9, 403.

Campbell, S.C., Moffatt, R.J. & Stamford, B.A. 2008. Smoking and smoking cessation—the relationship between cardiovascular disease and lipoprotein metabolism: A review. *Atherosclerosis,* 201(2):225-235.

Cappuccio, F.P. 1997. Ethnicity and cardiovascular risk: Variations in people of African ancestry and south Asian origin. *Journal of human hypertension,* 11(9):571-576.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P., Kalyanaraman, N. & Nemesh, J. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature genetics,* 22(3):231-238.

Casas, J.P., Shah, T., Cooper, J., Hawe, E., McMahon, A.D., Gaffney, D., Packard, C.J., O'Reilly, D.S., Juhan-Vague, I. & Yudkin, J.S. 2006. Insight into the nature of the CRP–coronary event association using Mendelian randomization. *International journal of epidemiology,* 35(4):922-931

Carmel, R. & James, S. J. 2002. Alcohol abuse: an important cause of severe hyperhomocysteinemia. *Nutrition Review*, 60(7):215–221

Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. 1994. History and geography of human genes. Princeton, New Jersey: Princeton University Press.

Chang, M., Yesupriya, A., Ned, R.M., Mueller, P.W. & Dowling, N.F. 2010. Genetic variants associated with fasting blood lipids in the US population: Third national health and nutrition examination survey. *BMC medical genetics,* 11(1):62.

Chardigny, J., Destaillats, F., Malpuech-Brugère, C., Moulin, J., Bauman, D.E., Lock, A.L., Barbano, D.M., Mensink, R.P., Bezelgues, J. & Chaumont, P. 2008. Do trans fatty acids from industrially produced sources and from natural sources have the same effect on cardiovascular disease risk factors in healthy subjects? Results of the trans fatty acids collaboration (TRANSFACT) study. *The American journal of clinical nutrition,* 87(3):558-566.

Chen, Y. C., Lee-Chen, G. J., Wu, Y. R., Hu, F. J., Wu, H. C., Kuo, H. C., & Chen, C. M. 2009. Analyses of interaction effect between apolipoprotein E polymorphism and alcohol use as well as cholesterol concentrations on spontaneous deep intracerebral hemorrhage in the Taiwan population. *Clinical Chimica Acta*, 408(1), 128-132.

Christensen, K., Mcgue, M., Yashin, A., Iachine, I., Holm, N.V. &Vaupel, J.M. 2000. Genetic and Environmental Influences on Functional Abilities in Danish Twins Aged 75 Years and Older. *Journal of Gerontology: Medical Science.* 55A(8): 446–452.

Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo, J. L., & Roccella, E. J. 2003. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. *Hypertension*, 42(6), 1206-1252.

Clouse, W. D., Rud, K. S., Hurt, R. D., & Miller, V. M. 2000. Short-term treatment with transdermal nicotine affects the function of canine saphenous veins. *Vascular Medicine*, 5(2), 75-82.

Cleeman, J. I., Grundy, S. M., Becker, D., & Clark, L. T. 2001. Expert panel on Detection, Evaluation and Treatment of High blood Cholesterol in Adults. Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Adult Treatment Panel (ATP III). *Jama*, 285(19), 2486-2497.

Connor, M., Rheeder, P., Bryer, A., Meredith, M., Dubb, A., & Fritz, V. 2008. The South African stroke risk in general practice study. *South African Medical Journal*, 95(5), 334.

Corbo, R. & Scacchi, R. 1999. Apolipoprotein E (APOE) allele distribution in the world.is APOE* 4 a 'thrifty'allele? *Annals of human genetics,* 63(4):301-310.

Corella, D. & Ordovas, J.M. 2005. Single nucleotide polymorphisms that influence lipid metabolism: Interaction with dietary factors. *Annual review nutrition,* 25: 341-390.

Corella, D., Sabater, A., Cortina, S., Folch, J., & Ordovas, J. M. 2001a. Environmental factors modulate the effect of the APOE genetic polymorphism on plasma lipid concentrations: Ecogenetic studies in a Mediterranean Spanish population. *Metabolism*, 50(8), 936-944.

Corella, D., Tucker, K., Lahoz, C., Coltell, O., Cupples, L.A., Wilson, P.W., Schaefer, E.J. & Ordovas, J.M. 2001b. Alcohol drinking determines the effect of the APOE locus on LDL-cholesterol concentrations in men: The Framingham offspring study. *The American journal of clinical nutrition,* 73(4):736-745.

Corella, D., Portolés, O., Arriola, L., Chirlaque, M. D., Barrricarte, A., Francés, F., & Moreno-Iribas, C. 2011. Saturated fat intake and alcohol consumption modulate the association between the APOE polymorphism and risk of future coronary heart disease: a nested case-control study in the Spanish EPIC cohort. *The Journal of Nutritional Biochemistry*, 22(5), 487-494.

Craig, W.Y., Palomaki, G.E. & Haddow, J.E. 1989. Cigarette smoking and serum lipid and lipoprotein concentrations: An analysis of published data. *British medical journal,* 298(6676):784.

Cummings, J.H. 1998. Apolipoprotein E genotype modulates the effect of black tea drinking on blood lipids and blood coagulation factors: A pilot study. *British journal of nutrition,* 79133-139.

Cummings, M. R. 2011. *Human Heredity: Principles & Issues*. Cengage Learning.

Curtiss, L.K. & Boisvert, W.A. 2000. Apolipoprotein E and atherosclerosis. *Current opinion in lipidology,* 11(3):243-251.

Dahlöf, B. 2010. Cardiovascular disease risk factors: Epidemiology and risk assessment. *The American journal of cardiology,* 105(1):3A-9A.

Danik, J. S., Paré, G., Chasman, D. I., Zee, R. Y., Kwiatkowski, D. J., Parker, A., & Ridker, P. M. 2009. Novel Loci, Including Those Related to Crohn Disease, Psoriasis, and Inflammation, Identified in a Genome-Wide Association Study of Fibrinogen in 17 686 Women The Women's Genome Health Study. *Circulation: Cardiovascular Genetics*, 2(2), 134-141.

Davignon, J., Gregg, R.E. & Sing, C.F. 1988. Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis, thrombosis, and vascular biology,* 8(1):1-21.

De Backer, G.G. 2008. Risk factors and prevention of cardiovascular disease: A review. *Dialogues in cardiovascular medicine,* 13(2):83-99.

De Caterina, R., Zampolli, A., Del Turco, S., Madonna, R., & Massaro, M. 2006. Nutritional mechanisms that influence cardiovascular disease. *The American journal of clinical nutrition*, 83(2), 421S-426S.

De Oliveira, C.M., Pereira, A.C., De Andrade, M., Soler, J.M. & Krieger, J.E. 2008. Heritability of cardiovascular risk factors in a Brazilian population: Baependi heart study. *BMC medical genetics,* 9(1):32.

De Maat, M. P. 2001. Effects of diet, drugs, and genes on plasma fibrinogen levels. *Annals of the New York Academy of Sciences*, 936(1), 509-521.

Djoussé, L., Pankow, J.S., Arnett, D.K., Eckfeldt, J.H., Myers, R.H. & Ellison, R.C. 2004.Apolipoprotein E polymorphism modifies the alcohol-HDL association observed in the national heart, lung, and blood institute family heart study. *The American journal of clinical nutrition,* 80(6):1639-1644.

Dong, L., Parkin, S., Trakhanovl, S.D., Rupp, B. & Simmons, T. 1996.Binding of apolipoprotein E2 in type ll hyperlipoproteinemia. *Nature structural biology,* 3(8).

Draisma, H. H., Reijmers, T. H., Meulman, J. J., van der Greef, J., Hankemeier, T., & Boomsma, D. I. 2013. Hierarchical clustering analysis of blood plasma lipidomics profiles from mono-and dizygotic twin families. *European Journal of Human Genetics*, 21(1): 95-101.

Dzau, V.J., Antman, E.M., Black, H.R., Hayes, D.L., Manson, J.E., Plutzky, J., Popma, J.J. & Stevenson, W. 2006. The cardiovascular disease continuum validated: Clinical evidence of improved patient outcomes part i: Pathophysiology and clinical trial evidence (risk factors through stable coronary artery disease). *Circulation,* 114(25):2850-2870.

Eastwood, M. A. 1992. The physiological effect of dietary fiber: an update. *Annual review of nutrition*, 12(1), 19-35.

Ehnholm, C., Mahley, R. W., Chappell, D. A., Weisgraber, K. H., Ludwig, E., & Witztum, J. L. 1984. Role of apolipoprotein E in the lipolytic conversion of beta-very low density lipoproteins to low density lipoproteins in type III hyperlipoproteinemia. *Proceedings of the National Academy of Sciences*, 81(17), 5566-5570.

Eichner, J.E., Dunn, S.T., Perveen, G., Thompson, D.M., Stewart, K.E. & Stroehla, B.C. 2002. Apolipoprotein E polymorphism and cardiovascular disease: A HuGE review. *American journal of epidemiology,* 155(6):487-495.

Elosua, R., Lluis, C. & Lucas, G. 2009. Research into the genetic component of heart disease: From linkage studies to genome-wide genotyping. *Revistaespañola de cardiología,* 9(Supl. B):24-38.

Ezzati, M., Lopez, A.D., Rodgers, A., Vander Hoorn, S. & Murray, C.J. 2002.Selected major risk factors and global and regional burden of disease. *The lancet,* 360(9343):1347-1360.

Ezzati, M., Vander Hoorn, S., Lawes, C.M., Leach, R., James, W.P.T., Lopez, A.D., Rodgers, A. & Murray, C.J. 2005.Rethinking the "diseases of affluence" paradigm: Global patterns of nutritional risks in relation to economic development. *PLoS medicine,* 2(5):e133.

Ezzati, M., Vander Hoorn, S., Rodgers, A., Lopez, A.D., Mathers, C.D. & Murray, C.J. 2003.Estimates of global and regional potential health gains from reducing multiple major risk factors. *The Lancet,* 362(9380):271-280.

Festa, A., D'agostino, R., Mykkänen, L., Tracy, R., Howard, B.V. & Haffner, S.M. 1999. Low-density lipoprotein particle size is inversely related to plasminogen activator inhibitor-1 levels the insulin resistance atherosclerosis study. *Arteriosclerosis, thrombosis, and vascular biology,* 19(3):605-610.

Festa, A., D'Agostino, R., Rich, S. S., Jenny, N. S., Tracy, R. P., & Haffner, S. M. 2003. Promoter (4G/5G) Plasminogen Activator Inhibitor-1 Genotype and Plasminogen Activator Inhibitor-1 Levels in Blacks, Hispanics, and Non-Hispanic Whites The Insulin Resistance Atherosclerosis Study. *Circulation*, 107(19), 2422-2427.

Fisher, C., Abdul-Aziz, D. & Blacklow, S.C. 2004. A two-module region of the low-density lipoprotein receptor sufficient for formation of complexes with apolipoprotein E ligands. *Biochemistry,* 43(4):1037-1044.

Forouhi, N.G. & Sattar, N. 2006. CVD risk factors and ethnicity—A homogeneous relationship? *Atherosclerosis supplements,* 7(1):11-19.

Forrester, T. 2004. Historic and Early Life Origins of Hypertension in Africans. *American Society for Nutritional Sciences,* 134: 211–216.

Frick, B., Schroecksnadel, K., Neurauter, G., Leblhuber, F., & Fuchs, D. 2004. Increasing production of homocysteine and neopterin and degradation of tryptophan with older age. *Clinical biochemistry*, 37(8), 684-687.

Friedlander, Y., Austin, M.A., Newman, B., Edwards, K., Mayer-Davis, E.J. & King, M. 1997. Heritability of longitudinal changes in coronary-heart-disease risk factors in women twins. *The American journal of human genetics,* 60(6):1502-1512.

Frikke-Schmidt, R., Nordestgaard, B.G., Agerholm-Larsen, B., Schnohr, P. & Tybjærg-Hansen, A. 2000. Context-dependent and invariant associations between lipids, lipoproteins, and apolipoproteins and apolipoprotein E genotype. *Journal of lipid research,* 41(11):1812-1822.

Frimerman, M., Miller, M., Hylton, I., Laniado, M. & Keren, M. 1997. Changes in hemostatic function at times of cyclic variation in occupational stress. *The American journal of cardiology,* 79(1):72-75.

Fullerton, S.M., Clark, A.G., Weiss, K.M., Nickerson, D.A., Taylor, S.L., Stengård, J.H., Salomaa, V., Vartiainen, E., Perola, M. & Boerwinkle, E. 2000. Apolipoprotein E variation at the sequence haplotype level: Implications for the origin and maintenance of a major human polymorphism. *The American journal of human genetics,* 67(4):881-900.

Gelehrter., T.D., Collins, F.S. & Ginsburg, D. 1998. Principles of medical genetics. 2nd *ed.* Baltimore, MD: Lippincott Williams & Wilkins. p410.

Geng, H., Law, P.P., Ng, M.C., Li, T., Liang, L., Ge, T., Wong, K., Liang, C., Ma, R.C. & SO, W. 2011. APOE genotype-function relationship: Evidence of− 491 A/T promoter polymorphism modifying transcription control but not type 2 diabetes risk. *PloS one,* 6(10):e24669.

Gerdes, L.U. 2003. The common polymorphism of apolipoprotein E: Geographical aspects and new pathophysiological relations. *Clinical chemistry and laboratory medicine,* 41(5):628-631.

Gerdes, L. U., Gerdes, C., Hansen, P. S., Klausen, I. C., Faergeman, O., & Dyerberg, J. 1996. The apolipoprotein E polymorphism in Greenland Inuit in its global perspective. *Human genetics*, 98(5), 546-550.

Gluckman, P.D. & Hanson, M.A. 2004. Living with the past: Evolution, development, and patterns of disease. *Science,* 305(5691):1733-1736.

Gordon, T., Castelli, W. P., Marthana, C. H., William, B. K. & Thomas, R. D. 1977. High density lipoprotein as a protective factor against coronary heart disease: The Framingham study. *The American journal of medicine,* 65(5): 707-714.

Grassi, M., Assanelli, D., Mozzini, C., Albertini, F., Salvadori, G., Archetti, S., Negrini, R., Galeazzi, G. & Pezzini, A. 2005. Modeling premature occurrence of acute coronary syndrome with atherogenic and thrombogenic risk factors and gene markers in extended families. *Journal of thrombosis and haemostasis,* 3(10):2238-2244.

Greenland, P., Knoll, M.D., Stamler., J., Neaton, J.D., Dyer, Ar., Garside, D.B. & Wilson, P .W. 2003. Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. *The journal of the American Medical Association,* 290(7):891-897.

Greenow, K., Pearce, N.J. & Ramji, D.P. 2005.The key role of apolipoprotein E in atherosclerosis. *Journal of molecular medicine,* 83(5):329-342.

Greyling, A., Pieters, M., Hoekstra, T., Oosthuizen, W. & Schutte, A. 2007. Differences in the association of PAI-1 activity with the metabolic syndrome between African and Caucasian women. *Nutrition, metabolism and cardiovascular diseases,* 17(7):499-507.

Grundy, S.M., Bilheimer, D., Chait, A., Clark, L.T., Denke, M., Havel, R.J., Hazzard, W.R., Hulley, S.B., Hunninghake, D.B. & Kreisberg, R.A. 1993. Summary of the second report of the national cholesterol education program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel II). *The journal of the American medical association,* 269(23):3015-3023.

Grundy, S.M., Pasternak, R., Greenland, P., Smith, S. & Fuster, V. 1999. Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations, a statement for healthcare professionals from the American heart association and the American college of cardiology. *Journal of the American college of cardiology,* 34(4):1348-1359.

Gylling, H., Kontula, K., & Miettinen, T. A. 1995. Cholesterol absorption and metabolism and LDL kinetics in healthy men with different apoprotein E phenotypes and apoprotein B Xba I and LDL receptor Pvu II genotypes. *Arteriosclerosis, thrombosis, and vascular biology*, 15(2), 208-213.

Haan, M. N., & Mayeda, E. R. 2010. Apolipoprotein E genotype and cardiovascular diseases in the elderly. *Current cardiovascular risk reports*, 4(5), 361-368.

Haddy, N., De Bacquer, D., Chemaly, M.M., Maurice, M., Ehnholm, C., Evans, A., Sans, S., Martins, M., De Backer, G. & Siest, G. 2002. The importance of plasma apolipoprotein E concentration in addition to its common polymorphism on inter-individual variation in lipid levels: Results from ApoEurope. *European Journal of Human Genetics*, 10(12).

Hagberg, J.M., Wilund, K.R. & Ferrell, R.E. 2000.APO E gene and gene-environment effects on plasma lipoprotein-lipid levels. *Physiological genomics,* 4(2):101-108.

Halushka, M.K., Fan, J., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R. & Chakravarti, A. 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature genetics,* 22(3):239-247.

Hanlon, C.S. & Rubinsztein, D.C. 1995. Arginine residues at codons 112 and 158 in the apolipoprotein E gene correspond to the ancestral state in humans. *Atherosclerosis,* 112(1):85-90.

Hardy, G. H. 1908. Mendelian proportions in a mixed population. *Science*, 28(706), 49-50.

Harris, P., Nagy, S., & Vardaxis, N. 2006. Mosby's Dictionary of Medicine, Nursing and Health Professions, Australian and New Zealand edition. *Clinical and Experimental Optometry*, 89(6), 402-404.

Hatters, D.M., Peters-Libeu, C.A. & Weisgraber, K.H. 2006. Apolipoprotein E structure: Insights into function. *Trends in biochemical sciences,* 31(8):445-454.

Hawkesworth, S., Dangour, A.D., Johnston, D., Lock, K., Poole, N., Rushton, J., Uauy, R. & Waage, J. 2010.Feeding the world healthily: The challenge of measuring the effects of agriculture on health. *Philosophical transactions of the royal society B: Biological sciences,* 365(1554):3083-3097.

Hegele, R.A. 2009. Plasma lipoproteins: Genetic influences and clinical implications. *Nature reviews genetics,* 10(2):109-121.

Heller, D.A., Pedersen, N.L., De Faire, U., & Mcclearn, G.E. 1994. Genetic and Environmental Correlations among Serum Lipids and Apolipoproteins in Elderly Twins Reared Together and Apart. *American Journal of Human Genetics*, 55: 1255-1267.

Henry, M., Tregouet, D., Alessi, M., Aillaud, M., Visvikis, S., Siest, G., Tiret, L. & Juhan-Vague, I. 1998. Metabolic determinants are much more important than genetic polymorphisms in determining the PAI-1 activity and antigen plasma concentrations a family study with part of the stanislas cohort. *Arteriosclerosis, thrombosis, and vascular biology,* 18(1):84-91.

Hernandez, L.M., & Blazer, D.G. (Eds.). 2006. Genes, behaviour, and the social environment: moving beyond the nature/nurture debate. *National Academies Press*.

Holme, I. 1990. An analysis of randomized trials evaluating the effect of cholesterol reduction on total mortality and coronary heart disease incidence. *Circulation,* 82(6):1916-1924.

Holvoet, P., Jenny, N.S., Schreiner, P.J., Tracy, R.P. & Jacobs, D.R. 2007. The relationship between oxidized LDL and other cardiovascular risk factors and subclinical CVD in different ethnic groups: The multi-ethnic study of atherosclerosis (MESA). *Atherosclerosis,* 194(1):245-252.

Hoogendijk, C. F., Scholtz, C. L., Pimstone, S. M., Ehrenborg, E., Kastelein, J. J., Defesche, J. C. & Kotze, M. J. 2003. Allelic variation in the promoter region of the LDL receptor gene: analysis of an African-specific variant in the FP2 *cis*-acting regulatory element. *Molecular and cellular probes*, 17(4): 175-181.

Hopkins, P.N. & Williams, R.R. 1981. A survey of 246 suggested coronary risk factors. *Atherosclerosis,* 40(1):1-52.

Humphries, S.E., Talmud, P.J., Hawe, E., Bolla, M., Day, I.N. & Miller, G.J. 2001. Apolipoprotein E4 and coronary heart disease in middle-aged men who smoke: A prospective study. *The Lancet,* 358(9276):115-119.

Hunt, S. A., Abraham, W. T., Chin, M. H., Feldman, A. M., Francis, G. S., Ganiats, T. G., & Riegel, B. 2005. ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the adult a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure): developed in collaboration with the American College of Chest Physicians and the International Society for Heart and Lung Transplantation: endorsed by the Heart Rhythm Society. *Circulation*, 112(12), e154-e235.

Iliadou, A., Lichtenstein, P., de Faire, U., & Pedersen, N. L. 2001. Variation in genetic and environmental influences in serum lipid and apolipoprotein levels across the lifespan in Swedish male and female twins. *American journal of medical genetics*, 102(1): 48-58.

Illumina. 2012. An introduction to Illumina next generation sequencing technology for microbiologists. Illumina Inc.

Innerarity, T., Weisgraber, K., Arnold, K., Rall, S. & Mahley, R. 1984. Normalization of receptor binding of apolipoprotein E2.evidence for modulation of the binding site conformation. *Journal of biological chemistry,* 259(11):7261-7267.

Ikewaki, K., Rader, D. J., Zech, L. A., & Brewer, H. B. 1994. In vivo metabolism of apolipoproteins AI and E in patients with abetalipoproteinemia: implications for the roles of apolipoproteins B and E in HDL metabolism. *Journal of lipid research*, 35(10), 1809-1819.

Ioannidis, J.P., Ntzani, E.E. & Trikalinos, T.A. 2004. Racial'differences in genetic effects for complex diseases. *Nature genetics,* 36(12):1312-1318.

James, S., Vorster, H.H., Venter, C.S., Kruger, H.S., Nell, T.A., Veldman, F.J. & Ubbink, J.B. 2000. Nutritional status influences plasma fibrinogen concentration: Evidence from the THUSA survey. *Thrombosis research,* 98(5):383-394.

Jansen, A.C., Van Aalst-Cohen, E.S., Tanck, M.W., Cheng, S., Fontecha, M.R., Li, J., Defesche, J.C. & Kastelein, J.J. 2005. Genetic determinants of cardiovascular disease risk in familial hypercholesterolemia. *Arteriosclerosis, thrombosis, and vascular biology,* 25(7):1475-1481.

Jemaa, R., Elasmi, M., Naouali, C., Feki, M., Kallel, A., Souissi, M., & Kaabachi, N. 2006. Apolipoprotein E polymorphism in the Tunisian population: frequency and effect on lipid parameters. *Clinical biochemistry*, 39(8), 816-820.

Jofre-Monseny, L., Minihane, A. & Rimbach, G. 2008. Impact of apoE genotype on oxidative stress, inflammation and disease risk. *Molecular nutrition & food research,* 52(1):131-145.

Jorde, L.B. & Wooding, S.P. 2004. Genetic variation, classification and 'race'. *Nature genetics,* (36): S28-S33.

Kahn, K., Tollman, S.M., Collinson, M.A., Clark, S.J., Twine, R., Clark, B.D., Shabangu, M., Gómez-Olivé, F.X., Mokoena, O. & Garenne, M.L. 2007. Research into health, population and social transitions in rural South Africa: Data and methods of the Agincourt health and demographic surveillance System1. *Scandinavian journal of public health,* 35(69):8-20.

Kaptoge, S., White, I.R., Thompson, S.G., Wood, A.M., Lewington, S., Lowe, G.D. & Danesh, J. 2007. Associations of plasma fibrinogen levels with established cardiovascular disease risk factors, inflammatory markers, and other characteristics: Individual participant meta-analysis

of 154,211 adults in 31 prospective studies: The fibrinogen studies collaboration. *American journal of epidemiology,* 166(8):867-879.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y. & Tanaka, T. 2008. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature genetics,* 41(1):56-65.

Keil, J.E., Sutherland, S.E., Knapp, R.G., Lackland, D.T., Gazes, P.C. & Tyroler, H.A. 1993. Mortality rates and risk factors for coronary disease in black as compared with white men and women. *New England journal of medicine,* 329(2):73-78.

Kephart, D., Krueger, S., Grunst, T., & Shenoi, H. 2006. Introducing the Maxwell® 16 Instrument: A Simple, Robust and Flexible Tool for DNA Purification. *Promega Notes*, 92, 20-23.

Kesäniemi, Y. A., & Miettinen, T. A. (1987). Cholesterol absorption efficiency regulates plasma cholesterol level in the Finnish population. *European journal of clinical investigation*, 17(5), 391-395.

Kessler, C., Spitzer, C., Stauske, D., Mende, S., Stadlmüller, J., Walther, R. & Rettig, R. 1997. The Apolipoprotein E and ß-Fibrinogen G/A-455 Gene Polymorphisms Are Associated With Ischemic Stroke Involving Large-Vessel Disease. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 17:2880-2884.

Khera, A., Mcguire, D.K., Murphy, S.A., Stanek, H.G., Das, S.R., Vongpatanasin, W., Wians, F.H., Grundy, S.M. & De Lemos, J.A. 2005. Race and gender differences in C-reactive protein levels.*Journal of the American college of cardiology,* 46(3):464-469.

Klein, D. 2002. Quantification using real-time PCR technology: applications and limitations. *Trends in molecular medicine*, 8(6), 257-260.

Klug, E.Q. 2012. South African dyslipidaemia guideline consensus statement. *South African medical journal,* 102(3):178-187.

Knoblauch, H., Bauerfeind, A., Krähenbühl, C., Daury, A., Rohde, K., Bejanin, S., Essioux, L., Schuster, H., Luft, F.C. & Reich, J.G. 2002. Common haplotypes in five genes influence genetic variance of LDL and HDL cholesterol in the general population. *Human molecular genetics,* 11(12):1477-1485.

Knoblauch, H., Bauerfeind, A., Toliat, M. R., Becker, C., Luganskaja, T., Günther, U. P., & Reich, J. G. 2004. Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. *Human molecular genetics*, 13(10), 993-1004.

Kolovou, G.D. & Anagnostopoulou, K.K. 2007. Apolipoprotein E polymorphism, age and coronary heart disease.*Ageing research reviews,* 6(2):94-108.

Kolovou, G.D., Daskalova, D.C., Hatzivassiliou, M., Yiannakouris, N., Pilatis, N.D., Elisaf, M., Mikhailidis, D.P., Cariolou, M.A. & Cokkinos, D.V. 2003. The epsilon 2 and 4 alleles of apolipoprotein E and ischemic vascular events in the Greek population—implications for the interpretation of similar studies. *Angiology,* 54(1):51-58.

Kolovou, G., Damaskos, D., Anagnostopoulou, K. & Cokkinos, D.V. 2009. Apolipoprotein E gene polymorphism and gender. *Annals of clinical & laboratory science,* 39(2):120-133.

Komurcu-Bayrak, E., Onat, A., Yuzbasiogullari, B., Mononen, N., Laaksonen, R., Kähönen, M., Hergenc, G., Lehtimäki, T. & Erginel-Unaltuna, N. 2011. The APOE − 219G/T and 113G/C polymorphisms affect insulin resistance among Turks. *Metabolism,* 60(5):655-663.

Kotze, M. J., Loubser, O., Thiart, R., Villiers, J. N. P., Langenhoven, E., Theart, L., Steyn, K., Marais, A.D., & Raal, F. J. 1997. CpG hotspot mutations at the LDL receptor locus are a frequent cause of familial hypercholesterolemia among South African Indians. *Clinical genetics*, 51(6): 394-398.

Kotze, M. J., Langenhoven, E., Theart, L., Loubser, O., Micklem, A., & Oosthuizen, C. J. J. 1995. Recurrent LDL-receptor mutation causes familial hypercholesterolaemia in South African coloureds and Afrikaners.

Kotze, M. J., De Villiers, W. J., Steyn, K., Kriek, J. A., Marais, A. D., Langenhoven, E., Herbert, J.S., Graadt van Roggen, J.F., van der Westhuyzen, D.R. & Coetzee, G. A. 1993. Phenotypic variation among familial hypercholesterolemics heterozygous for either one of two Afrikaner founder LDL receptor mutations. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 13(10): 1460-1468.

Krauss, R. M., Winston, M., Fletcher, B. J., & Grundy, S. M. 1998. Obesity: impact on cardiovascular disease. *Circulation*, 98(14), 1472-1476.

Krebs, M., Geiger, M., Polak, K., Vales, A., Schmetterer, L., Wagner, O. F., & Roden, M. 2003. Increased plasma levels of plasminogen activator inhibitor-1 and soluble vascular cell adhesion molecule after triacylglycerol infusion in man. *Thrombosis and haemostasis*, 90(3), 422-428.

Kris-Etherton, P. M. 1999. AHA science advisory: monounsaturated fatty acids and risk of cardiovascular disease. *The Journal of nutrition*, 129(12), 2280-2284.

Kris-Etherton, P. M., Innis, S., & Ammerican, D. A. 2007. Position of the American Dietetic Association and Dieticians of Canada: dietary fatty acids. *Journal of the American Dietetic Association*, 107(9), 1599.

Kruger, H.S., Venter, C.S. & Vorster, H.H. 2001. Obesity in African women in the North West Province, South Africa is associated ,vith an increased risk of non-communicable diseases: the THUSA study. Transition and Health during Urbanisation of South Africans. *The British journal of nutrition,* 86(6):733-740.

Kullo, I.J. & Ding, K. 2007. Mechanisms of disease: The genetic basis of coronary heart disease. *Nature clinical practice cardiovascular medicine,* 4(10):558-569.

Kushner, R. F., & Blatner, D. J. 2005. Risk assessment of the overweight and obese patient. *Journal of the American Dietetic Association*, 105(5), 53-62.

Lahoz, C., Schaefer, E.J., Cupples, L.A., Wilson, P.W., Levy, D., Osgood, D., Parpos, S., Pedro-Botet, J., Daly, J.A. & Ordovas, J.M. 2001. Apolipoprotein E genotype and cardiovascular disease in the Framingham heart study. *Atherosclerosis,* 154(3):529-537.

Lambert, J. C., Brousseau, T., Defosse, V., Evans, A., Arveiler, D., Ruidavets, J. B., & Amouyel, P. 2000. Independent association of an APOE gene promoter polymorphism with increased risk of myocardial infarction and decreased APOE plasma concentrations—the ECTIM study. *Human molecular genetics*, 9(1), 57-61.

Lamina, C., Forer, L., Schönherr, S., Kollerits, B., Ried, J. S., Gieger, C., & Kronenberg, F. 2012. Evaluation of gene-obesity interaction effects on cholesterol levels: A genetic predisposition score on HDL-cholesterol is modified by obesity. *Atherosclerosis*, 225: 363-369

Langenhoven, M., Conradie, P.J., Wolmarans, P., & Faber, M. 1991. MRC (Medical Research Council) Food Quantities Manual, 2nd Ed.

Lanktree, M. B., Anand, S. S., Yusuf, S., & Hegele, R. A. 2009. Replication of genetic associations with plasma lipoprotein traits in a multiethnic sample. *Journal of lipid research*, 50(7), 1487-1496.

Leeder, S.R. 2004. A race against time: The challenge of cardiovascular disease in developing economies. Columbia University.

Li, W. & Sadler, L.A. 1991. Low nucleotide diversity in man. *Genetics,* 129(2):513-523.

Libby, P. 2003. Vascular Biology of Atherosclerosis: Overview and State of the Art. *The American Journal of Cardiology*, 1(suppl):3A–6A

Libby, P., Ridker, P.M. & Hansson, G.K. 2011. Progress and challenges in translating the biology of atherosclerosis. *Nature,* 473(7347):317-325.

Liberopoulos, E., Miltiadous, G., Hatzivassiliou, M., Ayrton, N., Bairaktari, E., Cariolou, M. & Elisaf, M. 2004. Apolipoprotein E polymorphism in northwesterngreece: Frequency and effect on lipid parameters. *Annals of clinical & laboratory science,* 34(3):347-354.

Lichtenstein, A.H., Appel, L.J., Brands, M., Carnethon, M., Daniels, S., Franch, H.A., Franklin, B., Kris-Etherton, P., Harris, W.S. & Howard, B. 2006. Diet and lifestyle recommendations revision 2006. A scientific statement from the American heart association nutrition committee. *Circulation,* 114(1):82-96.

Litwin, S.E. 2008. Which measures of obesity best predict cardiovascular risk? *Journal of the American college of cardiology,* 52(8):616-619.

Loktionov, A., Vorster, H., O'Neill, I.K., Nell, T., Bingham, S.A., Runswick, S.A. & Cummings, J.H. 1999. Apolipoprotein E and methylenetetrahydrofolatereductase genetic polymorphisms in relation to other risk factors for cardiovascular disease in UK Caucasians and black South Africans. *Atherosclerosis,* 145(1):125-135.

Loock, M., Steyn, K., Becker, P. & Fourie, J. 2006. Coronary heart disease and risk factors in black South Africans: A case-control study. *Ethnicity and disease,* 16(4):872.

Lu, Y., Feskens, E. J., Boer, J., Imholz, S., Verschuren, W. M., Wijmenga, C., & Dollé, M. E. 2010. Exploring genetic determinants of plasma total cholesterol levels and their predictive value in a longitudinal study. *Atherosclerosis*, 213(1), 200-205.

Luc, G., Bard, J., Arveiler, D., Evans, A., Cambou, J., Bingham, A., Amouyel, P., Schaffer, P., Ruidavets, J. & Cambien, F. 1994. Impact of apolipoprotein E polymorphism on lipoproteins and risk of myocardial infarction. The ECTIM study. *Arteriosclerosis, thrombosis, and vascular biology,* 14(9):1412-1419.

Lusis, A., Heinzmann, C., Sparkes, R., Scott, J., Knott, T., Geller, R., Sparkes, M. & Mohandas, T. 1986. Regional mapping of human chromosome 19: Organization of genes for plasma lipid transport (APOC1,-C2, and-E and LDLR) and the genes C3, PEPD, and GPI. *Proceedings of the national academy of sciences,* 83(11):3929-3933.

Lusis, A.J. 2003. Genetic factors in cardiovascular disease: 10 questions. *Trends in cardiovascular medicine,* 13(8):309-316.

Lussier-Cacan, S., Bolduc, A., Xhignesse, M., Niyonsenga, T., Connelly, P. W., & Sing, C. F. 2000. Impact of age and body size on inter-individual variation in measures of lipid metabolism: influence of gender and apolipoprotein E genotype. *Clinical genetics*, 57(1) 35-47.

Macintyre, U. E., Venter, C. S., Vorster, H. H., & Steyn, H. S. 2001. A combination of statistical methods for the analysis of the relative validation data of the quantitative food frequency questionnaire used in the THUSA study. *Public Health Nutrition*, 4(1), 45-52.

Mackay, J., Mensah, G.A. & Greenlund, K. 2004. The atlas of heart disease and stroke. World Health Organization.

Mahley, R.W. 1988. Apolipoprotein E: Cholesterol transport protein with expanding role in cell biology. *Science,* 240(4852):622-630.

Mahley, R.W. & Huang, Y. 1999. Apolipoprotein E: From atherosclerosis to alzheimer's disease and beyond. *Current opinion in lipidology,* 10(3):207-218.

Mahley, R.W., Huang, Y. & Rall, S.C. 1999. Pathogenesis of type III hyperlipoproteinemia (dysbetalipoproteinemia): Questions, quandaries, and paradoxes. *Journal of lipid research,* 40(11):1933-1949.

Mahley, R.W. & Rall Jr, S.C. 2000. Apolipoprotein E: Far more than a lipid transport protein. *Annual review of genomics and human genetics,* 1(1):507-537.

Marques-Vidal, P., Bongard, V., Ruidavets, J. B., Fauvel, J., Hanaire-Broutin, H., Perret, B., & Ferrières, J. 2003. Obesity and alcohol modulate the effect of apolipoprotein E polymorphism on lipids and insulin. *Obesity research*, 11(10), 1200-1206.

Martin, E. R., Lai, E. H., Gilbert, J. R., Rogala, A. R., Afshari, A. J., Riley, J., Finch, K. L., Stevens, J. F., Livak, K. J.,Slotterbeck, B.D., Slifer, S.H., Warren, L.L., Conneally, P.M., Schmechel, D.E., Purvis, I., Pericak-Vance, M.A., Roses, A.D. & Vance, J. M. 2000. SNPing Away at Complex Diseases: Analysis of Single-Nucleotide Polymorphisms around APOE in Alzheimer Disease. *The American Journal of Human Genetics*, 67(2): 383-394.

Masemola, M.L., Alberts, M. & Urdal, P. 2007. Apolipoprotein E genotypes and their relation to lipid levels in a rural South African population 1. *Scandinavian journal of public health,* 35(69 suppl):60-65.

Mattes, R. D. 2005. Fat taste and lipid metabolism in humans. *Physiology & behavior*, 86(5), 691-697.

Mayhan, W. G., & Sharpe, G. M. 1999. Chronic exposure to nicotine alters endothelium-dependent arteriolar dilatation: effect of superoxide dismutase. *Journal of Applied Physiology*, 86(4), 1126-1134.

Mayosi, B.M., Flisher, A.J., Lalloo, U.G., Sitas, F., Tollman, S.M. & Bradshaw, D. 2009. The burden of non-communicable diseases in South Africa. *The Lancet,* 374(9693):934-947.

Mccully, K.S. 2007. Homocysteine, vitamins, and vascular disease prevention. *The American journal of clinical nutrition,* 86(5):1563S-1568S.

Mckeigue, P., Ferrie, J., Pierpoint, T. & Marmot, M. 1993. Association of early-onset coronary heart disease in South Asian men with glucose intolerance and hyperinsulinemia. *Circulation,* 87(1):152-161.

Mcvean, G., Spencer, C.C. & Chaix, R. 2005. Perspectives on human genetic variation from the HapMap project. *PLoS genetics,* 1(4):e54.

Medical Research Council, Inc. 2007. FoodFinder 3 (food analysis software system, version 3. www.foodfinder3.co.za.

Mehta, J., Mehta, P., Lawson, D. & Saldeen, T. 1987. Plasma tissue plasminogen activator inhibitor levels in coronary artery disease: Correlation with age and serum triglyceride concentrations. *Journal of the American college of cardiology,* 9(2):263-268.

Meilahn, E.N., Cauley, J.A., Tracy, R.P., Macy, E.O., Gutai, J.P. & Kuller, L.H. 1996. Association of sex hormones and adiposity with plasma levels of fibrinogen and PAI-1 in postmenopausal women. *American journal of epidemiology,* 143(2):159-166.

Menzel, H. J., & Utermann, G. 1986. Apolipoprotein E phenotyping from serum by Western blotting. *Electrophoresis*, 7(11), 492-495.

Mi, X., Eskridge, K. M., George, V., & Wang, D. 2011. Structural equation modeling of gene–environment interactions in coronary heart disease. *Annals of human genetics*, 75(2), 255-265.

Micha, R. & Mozaffarian, D. 2008.Trans fatty acids: Effects on cardiometabolic health and implications for policy. *Prostaglandins, leukotrienes and essential fatty acids,* 79(3):147-152.

Miettinen, T.A., Gylling, H., Vanhanen, H. & Ollus, A. 1992. Cholesterol absorption, elimination, and synthesis related to LDL kinetics during varying fat intake in men with different apoprotein E phenotypes. *Arteriosclerosis, thrombosis, and vascular biology,* 12(9):1044-1052.

Milewicz, D.M. & Seidman, C.E. 2000.Genetics of cardiovascular disease. *Circulation,* 102 (suppl 4):Iv-103-Iv-111.

Minihane, A., Jofre-Monseny, L., Olano-Martin, E. & Rimbach, G. 2007. ApoE genotype, cardiovascular risk and responsiveness to dietary fat manipulation.*Proceedings of the nutrition society,* 66(2):183-197.

Misra, A., Singhal, N. & Khurana, L. 2010. Obesity, the metabolic syndrome, and type 2 diabetes in developing countries: Role of dietary fats and oils. *Journal of the American college of nutrition,* 29(sup3):289S-301S.

Mollentze, W., Moore, A., Steyn, A., Joubert, G., Steyn, K., Oosthuizen, G. & Weich, D. 1995. Coronary heart disease risk factors in a rural and urban orange free state black population. *South African medical Journal,* 85(2):90-96.

Moreno, J. A., López-Miranda, J., Marín, C., Gómez, P., Pérez-Martínez, P., Fuentes, F., & Pérez-Jiménez, F. 2003. The influence of the apolipoprotein E gene promoter (− 219G/T) polymorphism on postprandial lipoprotein metabolism in young normolipemic males. *Journal of lipid research*, 44(11), 2059-2064.

Moreno, J.A., Pérez-Jiménez, F., Marín, C., Gómez, P., Pérez-Martínez, P., Moreno, R., Bellido, C., Fuentes, F. & López-Miranda, J. 2004. The effect of dietary fat on LDL size is influenced by apolipoprotein E genotype in healthy subjects. *The journal of nutrition,* 134(10):2517-2522.

Moreno, R., Perez-Jimenez, F., Marin, C., Moreno, J. A., Gomez, P., Bellido, C., & Lopez-Miranda, J. 2006. A single nucleotide polymorphism of the apolipoprotein A–V gene− 1131T> C modulates postprandial lipoprotein metabolism. *Atherosclerosis*, 189(1), 163-168.

Mozaffarian, D., & Clarke, R. 2009. Quantitative effects on cardiovascular risk factors and coronary heart disease risk of replacing partially hydrogenated vegetable oils with other fats and oils. *European Journal of Clinical Nutrition*, 63, S22-S33.

Mozaffarian, D., Aro, A. & Willett, W. 2009. Health effects of trans-fatty acids: Experimental and observational evidence. *European journal of clinical nutrition,* 63, S5-S21.

Muros, M. & Rodrćguez-Ferrer, C. 1996.Apolipoprotein E polymorphism influence on lipids, apolipoproteins and lp (a) in a Spanish population under expressing APO E4. *Atherosclerosis,* 121(1):13-21.

Murrell, J.R., Price, B.M., Baiyewu, O., Gureje, O., Deeg, M., Hendrie, H., Ogunniyi, A. & Hall, K. 2006.The fourth apolipoprotein E haplotype found in the Yoruba of Ibadan. *American journal of medical genetics part B: Neuropsychiatric genetics,* 141(4):426-427.

Mutch, D.M., Wahli, W. & Williamson, G. 2005. Nutrigenomics and nutrigenetics: The emerging faces of nutrition. *The FASEB journal,* 19(12):1602-1616.

Navas-Nacher, E. L., Colangelo, L., Beam, C., & Greenland, P. 2001. Risk factors for coronary heart disease in men 18 to 39 years of age. *Annals of internal medicine*, 134(6): 433-439.

Neel, J.V. 1962. Diabetes mellitus: A “thrifty” genotype rendered detrimental by “progress”? *American journal of human genetics,* 14(4):353.

Nickerson, D.A., Taylor, S.L., Fullerton, S.M., Weiss, K.M., Clark, A.G., Stengård, J.H., Salomaa, V., Boerwinkle, E. & Sing, C.F. 2000. Sequence diversity and large-scale typing of SNPs in the human apolipoprotein E gene. Genome research, 10(10):1532-1545.

Nienaber-Rousseau, C., Ellis, S. M., Moss, S. J., Melse-Boonstra, A., & Towers, G. W. 2013. Gene-environment and gene-gene interactions of specific MTHFR, MTR and CBS gene variants in relation to homocysteine in black South Africans. *Gene,* 530 (1): 113-118.

Norman, R., Bradshaw, D., Steyn, K., Gaziano, T., & the South African Comparative Risk Assessment Collaborating Group. 2007. Estimating the burden of disease attributable to high cholesterol in South Africa in 2000. *South African Medical Journal.* 97: 708-715.

Ohman, E.M., Bhatt, D.L., Steg, P.G., Goto, S., Hirsch, A.T., Liau, C., Mas, J., Richard, A., Röther, J. & Wilson, P.W. 2006. The Reduction of atherothrombosis for continued health (REACH) registry: An international, prospective, observational investigation in subjects at risk for atherothrombotic events-study design. *American heart journal,* 151(4):786. e1- e10.

Okrainec, K., Banerjee, D.K. & Eisenberg, M.J. 2004. Coronary artery disease in the developing world. *American heart journal,* 148(1):7-15.

Onat, A., Sari, İ., Can, G., Hergenç, G. & Avci, G.Ş. 2006. Plasma triglycerides, an independent predictor of cardiovascular disease in men: A prospective study based on a population with prevalent metabolic syndrome. *International journal of cardiology,* 108(1):89-95.

Oosthuizen, W., Vorster, H., Kruger, A., Venter, C., Kruger, H. & De Ridder, J. 2002. Impact of urbanisation on serum lipid profiles-the THUSA survey. *South African medical journal,* 92(9):723-728.

Opie, L.H. & Mayosi, B.M. 2005. Cardiovascular disease in sub-Saharan Africa. *Circulation,* 112(23):3536-3540.

Ordovas, J. M. (1999). The genetics of serum lipid responsiveness to dietary interventions. *Proceedings of the Nutrition Society*, 58(1), 171-188.

Ordovas, J.M. 2009. Genetic influences on blood lipids and cardiovascular disease risk: tools for primary prevention. *American Journal of clinical nutrition,* 89(suppl):1509S–17S.

Ordovas, J. M. 2008. Genotype–phenotype associations: modulation by diet and obesity. *Obesity*, 16(S3), S40-S46.

Ordovas, J.M. 2006. Nutrigenetics, plasma lipids, and cardiovascular risk. *Journal of the American dietetic association,* 106(7):1074-1081.

Ordovas, J.M. & Mooser, V. 2004. Nutrigenomics and nutrigenetics. *Current opinion in lipidology,* 15(2):101-108.

Pallaud, C., Gueguen, R., Sass, C., Grow, M., Cheng, S., Siest, G. & Visvikis, S. 2001. Genetic influences on lipid metabolism trait variability within the stanislas cohort. *Journal of lipid research,* 42(11):1879-1890.

Paré, G., Serre, D., Brisson, D., Anand, S.S., Montpetit, A., Tremblay, G., Engert, J.C., Hudson, T.J. & Gaudet, D. 2007. Genetic analysis of 103 candidate genes for coronary artery disease and associated phenotypes in a founder population reveals a new association between endothelin-1 and high-density lipoprotein cholesterol. *The American journal of human genetics,* 80(4):673-682.

Persico, A., D'agruma, L., Zelante, L., Militerni, R., Bravaccio, C., Schneider, C., Melmed, R., Trillo, S., Montecchi, F. & Elia, M. 2004. Enhanced APOE2 transmission rates in families with autistic probands. *Psychiatric genetics,* 14(2):73-82.

Pieters, M., Dolman, R., Ntyintyane, L., Jerling, J., & Raal, F. 2011a. Risk factor profile of coronary artery disease in black South Africans: coronary artery disease in black South Africans. *SA Heart*, 8(1): 4-11.

Pieters, M., De Maat, M.P., Jerling, J.C., Hoekstra, T. & Kruger, A. 2011b. Fibrinogen concentration and its role in CVD risk in black South Africans-effect of urbanisation. *Thrombosis and haemostasis,* 106(3):448.

Pieters, M. & Vorster, H.H. 2008. Nutrition and hemostasis: A focus on urbanization in South Africa. *Molecular nutrition & food research,* 52(1):164-172.

Pilia, G., Chen, W., Scuteri, A., Orrú, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M. & Loi, P. 2006. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS genetics,* 2(8):e132.

Pisa, P. T., Kruger, A., Vorster, H. H., Margetts, B. M., & Loots Du, T. 2010. Alcohol consumption and cardiovascular disease risk in an African population in transition: the Prospective Urban and Rural Epidemiology (PURE) study. *South African Journal of Clinical Nutrition*, 23(2): 29-37

Popkin, B.M. 2006. Global nutrition dynamics: The world is shifting rapidly toward a diet linked with non-communicable diseases. *The American journal of clinical nutrition,* 84(2):289-298.

Rahilly-Tierney, C. R., Arnett, D. K., North, K. E., Pankow, J. S., Hunt, S. C., Ellison, R. C., & Djoussé, L. 2011. Apolipoprotein ε4 polymorphism does not modify the association between body mass index and high-density lipoprotein cholesterol: a cross-sectional cohort study. *Lipids in health and disease*, 10(1), 167.

Rall, S., Weisgraber, K.H. & Mahley, R.W. 1982. Human apolipoprotein E. the complete amino acid sequence. *Journal of biological chemistry,* 257(8):4171-4178.

Reddy, K.S. & Yusuf, S. 1998. Emerging epidemic of cardiovascular disease in developing countries. *Circulation,* 97(6):596-601.

Reilly, J. T., & Cawley, J. C. 1996. Haematological disease. *Medicine. Oxford, UK: Blackwell Science*.

Rice, J. A. 2007. *Mathematical statistics and data analysis*. Cengage Learning.

Riddell, D.R., Graham, A. & Owen, J.S. 1997. Apolipoprotein E inhibits platelet aggregation through the l-arginine: Nitric oxide pathway implications for vascular disease. *Journal of biological chemistry,* 272(1):89-95.

Ridker, P.M., Cushman, M., Stampfer, M.J., Tracy, R.P. & Hennekens, C.H. 1997. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *New England journal of medicine,* 336(14):973-979.

Ridker, P.M., Hennekens, C.H., Buring, J.E. & Rifai, N. 2000. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *New England journal of medicine,* 342(12):836-843.

Ridker, P.M., Buring, J.E., Cook, N.R. & Rifai, N. 2003. C-reactive protein, the metabolic syndrome, and risk of incident cardiovascular events: An 8-year follow-up of 14 719 initially healthy American women. *Circulation,* 107(3):391-397.

Ridker, P.M., Brown, N.J., Vaughan, D.E., Harrison, D.G. & J.L. 2004. Established and emerging plasma biomarkers in the prediction of first atherothrombotic events. *Circulation,* 109(25 Supp11):N6-N19.

Ridker, P.M., Pare, G., Parker, A., Zee, R.Y., Danik, J.S., Buring, J.E., Kwiatkowski, D., Cook, N.R., Miletich, J.P. & Chasman, D.I. 2008. Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: The women's genome health study. *The American journal of human genetics,* 82(5):1185-1192.

Rosendorff, C. 2007. Hypertension and coronary artery disease: a summary of the American Heart Association scientific statement. *The Journal of Clinical Hypertension*, 9(10), 790-795.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J. & Willey, D.L. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409(6822):928-933.

Scacchi, R., Corbo, R. M., Rickards, O., Mantuano, E., Guevara, A., & De Stefano, G. F. 1997. Apolipoprotein B and E genetic polymorphisms in the Cayapa Indians of Ecuador. *Human biology*, 375-382.

Schachter, M. 1997. The pathogenesis of atherosclerosis. *International journal of cardiology*, *62*, S3-S7.

Schneeman, B.O. 2002.Gastrointestinal physiology and functions. *British journal of nutrition,* 88(S2):S159-S163.

Schuster, S.C., Miller, W., Ratan, A., Tomsho, L.P., Giardine, B., Kasson, L.R., Harris, R.S., Petersen, D.C., Zhao, F., Qi, J., Alkan, C., Kidd, J.M., Sun, Y., Drautz, D.I., Bouffard, P., Muzny, D.M., Reid, J.G., Nazareth, L.V., Wang, Q., Burhans, R., Riemer, C., Wittekindt, N.E., Moorjani, P., Tindall, E.A., Danko, C.G., Teo, W.S., Buboltz, A.M., Zhang, Z., Ma, Q., Oosthuysen, A., Steenkamp, A.W., Oostuisen, H., Venter, P., Gajewski, J., Zhang, Y., Pugh, B.F., Makova, K.D., Nekrutenko, A., Mardis, E.R., Patterson, N., Pringle, T.H., Chiaromonte, F., Mullikin, J.C., Eichler, E.E., Hardison, R.C., Gibbs, R.A., Harkins, T.T. & Hayes, V.M. 2010. Complete Khoisan and Bantu genomes from southern Africa. *Nature*, 463(7283):943-947.

Schutte, A., Van Vuuren, D., Van Rooyen, J., Huisman, H., Schutte, R., Malan, L. & Malan, N. 2006. Inflammation, obesity and cardiovascular function in African and Caucasian women from South Africa: The POWIRS study. *Journal of human hypertension,* 20(11):850-859.

Scott, J., Knott, T., Shaw, D. & Brook, J. 1985. Localization of genes encoding apolipoproteins CI, CII, and E to the p13→ cen region of human chromosome 19.*Human genetics,* 71(2):144-146.

Seripa, D., Matera, M. G., Daniele, A., Bizzarro, A., Rinaldi, M. Gravina, C., Bisceglia, L., Corbo, R. M., Panza, F., Solfrizzi, V., Fazio, V.M., Dal Forno, G., Masullo, C., Dallapiccola, B., & Pilotto, A. 2007. The Missing ApoE Allele. *Annals of Human Genetics*, 71:496–500.

Shamai, L., Lurix, E., Shen, M., Novaro, G.M., Szomstein, S., Rosenthal, R., Hernandez, A.V. & Asher, C.R. 2011. Association of body mass index and lipid profiles: Evaluation of a broad spectrum of body mass index patients including the morbidly obese. *Obesity surgery,* 21(1):42-47.

Shelling, A.N. & Ferguson, L.R. 2007. Genetic variation in human disease and a new role for copy number variants. *Mutation Research/Fundamental and molecular mechanisms of mutagenesis,* 622(1):33-41.

Shenoi, H., Kephart, D., Bjerke, M., Shultz, S. & Cowan, C. 2007. Maxwell® 16 buffy coat genomic DNA purification application.

Shetty, P.S. 2002. Nutrition transition in India. *Public health nutrition,* 5(1A):175-182.

Shimano, H., Ohsuga, J., Shimada, M., Namba, Y., Gotoda, T., Harada, K., Katsuki, M., Yazaki, Y. & Yamada, N. 1995. Inhibition of diet-induced atheroma formation in transgenic mice expressing apolipoprotein E in the arterial wall. *Journal of clinical investigation,* 95(2):469.

Shisana, O., Labadarios, D., Rehle, T., Simbayi, L., Zuma, K., Dhansay, A., & Faber, M. the SANHANES-1 Team (2013) South African National Health and Nutrition Examination Survey (SANHANES-1).

Shore, V.G., & Shore, B. 1973. Heterogeneity of Human Plasma Very Low Density Lipoproteins. Separation of Species Differing in Protein Components. *Biochemistry*, 12(3): 502-507.

Siest, G., Pillot, T., Régis-Bailly, A., Leininger-Muller, B., Steinmetz, J., Galteau, M. & Visvikis, S. 1995. Apolipoprotein E: An important gene and protein to follow in laboratory medicine. *Clinical chemistry,* 41(8):1068-1086.

Silander, K., Alanne, M., Kristiansson, K., Saarela, O., Ripatti, S., Auro, K., Karvanen, J., Kulathinal, S., Niemelä, M. & Ellonen, P. 2008. Gender differences in genetic risk profiles for cardiovascular disease. *PloS one,* 3(10):e3615.

Sing, C.F., Stengård, J.H. & Kardia, S.L.R. 2003. Genes, environment, and cardiovascular disease. *Arteriosclerosis, thrombosis, and vascular biology,* 23(7):1190-1196.

Sliwa, K., Wilkinson, Hansen, C., Ntyintyane, L., Tibazarw A, K., Becker, A & Stewart, S. 2008. Spectrum of heart disease and risk factors in a black urban population in South Africa (the Heart of Soweto Study): a cohort study. *The Lancet,* 371(9616):915-922.

Soerensen, M., Dato, S., Tan, Q., Thinggaard, M., Kleindorp, R., Beekman, M., Suchiman, H.E.D., Jacobsen, R., Mcgue, M. & Stevnsner, T. 2012. Evidence from case–control and longitudinal studies supports associations of genetic variation in APOE, CETP, and IL6 with human longevity. *Age,* 1-14.

Solano, M.P., Perry, A.C., Wang, X., Ross, R. & Goldberg, R.B. 2003. Insulin resistance but not visceral adipose tissue is associated with plasminogen activator inhibitor type 1 levels in overweight and obese premenopausal African-American women. *International journal of obesity,* 27(1):82-87.

Song, Y., Stampfer, M.J. & Liu, S. 2004. Meta-analysis: Apolipoprotein E genotypes and risk for coronary heart disease. *Annals of internal medicine,* 141(2):137-147.

Sonnenberg, L. M., Quatromoni, P. A., Gagnon, D. R., Cupples, L. A., Franz, M. M., Ordovas, J. M., Wilson, P.W.F., Schaefer, E.J. & Millen, B. E. 1996. Diet and plasma lipids in women. II. Macronutrients and plasma triglycerides, high-density lipoprotein, and the ratio of total to high-density lipoprotein cholesterol in women: the Framingham nutrition studies. *Journal of clinical epidemiology*, 49(6): 665-672.

Srinivasan, S. R., Ehnholm, C., Elkasabany, A., & Berenson, G. S. 2001. Apolipoprotein E polymorphism modulates the association between obesity and dyslipidemias during young adulthood: the Bogalusa Heart Study. *Metabolism*, 50(6): 696-702.

Stannard, A.K., Riddell, D.R., Sacre, S.M., Tagalakis, A.D., Langer, C., Von Eckardstein, A., Cullen, P., Athanasopoulos, T., Dickson, G. & Owen, J.S. 2001. Cell-derived apolipoprotein E (ApoE) particles inhibit vascular cell adhesion molecule-1 (VCAM-1) expression in human endothelial cells. *Journal of biological chemistry,* 276(49):46011-46016.

Statistica. 2009. StatSoft, Inc. STATISTICA (data analysis software system), version 10. www.statsoft.com.

Statistics South Africa. 2010. Mortality and causes of death in South Africa, 2010: Findings from death notification. Statistics South Africa.

Stengård, J. H., Clark, A. G., Weiss, K. M., Kardia, S., Nickerson, D. A., Salomaa, V., Ehnholm, E.B., & Sing, C. F. 2002. Contributions of 18 additional DNA sequence variations in the gene encoding apolipoprotein E to explaining variation in quantitative measures of lipid metabolism. *The American Journal of Human Genetics*, 71(3): 501-517.

Stengård, J.H., Weiss, K.M. & Sing, C.F. 1998. An ecological study of association between coronary heart disease mortality rates in men and the relative frequencies of common allelic variations in the gene coding for apolipoprotein E. *Human genetics,* 103(2):234-241.

Stephens, J.W., Bain, S.C. & Humphries, S.E. 2008. Gene–environment interaction and oxidative stress in cardiovascular disease. *Atherosclerosis*, 200: 229–238.

Stephens, J. & Humphries, S. 2003. The molecular genetics of cardiovascular disease: Clinical implications. *Journal of internal medicine,* 253(2):120-127.

Steyn, K. 2007. Heart disease in South Africa. Editor, Heart and Stroke Foundation South Africa.

Steyn, N., Burger, S., Monyeki, K., Alberts, M. & Nthangeni, G. 2001. Seasonal variation in dietary intake of the adult population of Dikgale. *South African journal of clinical nutrition,* 14(4):140-145.

Steyn, K., Jooste, P., Bourne, L., Fourie, J., Badenhorst, C., Bourne, D., Langenhoven, M., Lombard, C., Truter, H. & Katznellenbogen, J. 1991.Risk factors for coronary heart disease in the black population of the cape peninsula. *South African medical journal,* 79(8):480-485.

Steyn, K., Sliwa, K., Hawken, S., Commerford, P., Onen, C., Damasceno, A., Ounpuu, S. & Yusuf, S. 2005. Risk factors associated with myocardial infarction in Africa the INTERHEART Africa study. *Circulation,* 112(23):3554-3561.

Steyn, K., Goldberg, Y. P., Kotze, M. J., Steyn, M., Swanepoel, A. S. P., Fourie, J. M., ... & Van der Westhuyzen, D. R. 1996. Estimation of the prevalence of familial hypercholesterolaemia in a rural Afrikaner community by direct screening for three Afrikaner founder low density lipoprotein receptor gene mutations. *Human genetics*, 98(4): 479-484.

Steyn, N.P. & Nel, J.H. 2006. Dietary intake of adult women in South Africa, Kenya and Nigeria with a focus on the use of spreads. Accessed at http://www.mrc.ac.za/chronic/kenyareport.pdf

Stiko-Rahm, A., Wiman, B., Hamsten, A. & Nilsson, J. 1990. Secretion of plasminogen activator inhibitor-1 from cultured human umbilical vein endothelial cells is induced by very low density lipoprotein. *Arteriosclerosis, thrombosis, and vascular biology,* 10(6):1067-1073.

Sundrehagen E/Axis Biochemicals ASA. Enzymatic Assay for Homocysteine and a Kit Therefor. EP 623174/US 5631127.

Talayero, B.G. & Sacks, F.M. 2011. The Role of Triglycerides in Atherosclerosis. *Current Cardiology Reports*, 13(6):544–552

Tall, A.R., Yvan-Charvet, L., Terasaka, N., Pagler, T. & Wang, N. 2008. HDL, ABC transporters, and cholesterol efflux: Implications for the treatment of atherosclerosis. *Cell metabolism,* 7(5):365-375.

Talmud, P.J. 2007. Gene–environment interaction and its impact on coronary heart disease risk. *Nutrition, metabolism and cardiovascular diseases,* 17(2):148-152.

Talmud, P.J., Drenos, F., Shah, S., Shah, T., Palmen, J., Verzilli, C., Gaunt, T.R., Pallas, J., Lovering, R. & Li, K. 2009. Gene-centric association signals for lipids and apolipoproteins identified via the Human CVD BeadChip. *The American journal of human genetics,* 85(5):628-642.

Temple, N. & Steyn, N.P. 2009. Food prices and energy density as barriers to healthy food patterns in Cape Town, South Africa. *Journal of Hunger Environmental Nutrition* 4(2): 203-13

Teo, K., Chow, C.K., V Az, M., Rangarajan, S., Yusuf, S. & Pure Investigators writing Group. 2009. The Prospective Urban Rural Epidemiology (PURE) study: examining the impact of societal influences on chronic non communicable diseases in low-, middle-, and high-income countries. *American heart journal,* 158(1):1-7.e1.

Thiart, R., Scholtz, C. L., Vergotine, J., Hoogendijk, C. F., de Villiers, J. N. P., Nissen, H., Brusgaard, K., Gaffney, D., Hoffs, M.S., Vermaak, W.J.H., Kotze, M. J. 2000. Predominance of a 6 bp deletion in exon 2 of the LDL receptor gene in Africans with familial hypercholesterolaemia. *Journal of medical genetics*, 37(7): 514-519.

Thorogood, M., Connor, M., Tollman, S., Hundt, G.L., Fowkes, G. & Marsh, J. 2007. A cross-sectional study of vascular risk factors in a rural South African population: Data from the southern African stroke prevention initiative (SASPI). *BMC public health,* 7(1):326.

Tibazarwa, K., Ntyintyane, L., Sliwa, K., Gerntholtz, T., Carrington, M., Wilkinson, D. & Stewart, S. 2009. A time bomb of cardiovascular risk factors in South Africa: Results from the heart of Soweto study “Heart awareness days”. *International journal of cardiology,* 132(2):233-239.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J. & Doumbo, O. 2009. The genetic structure and history of Africans and African Americans. *Science,* 324(5930):1035-1044.

Tomson, J. & Lip, G.Y. 2005. Blood pressure demographics: Nature or nurture, genes or environment? *BMC medicine,* 3(1):3.

Tremoli, E., Camera, M., Maderna, P., Sironi, L., Prati, L., Colli, S., Piovella, F., Bernini, F., Corsini, A. & Mussoni, L. 1993. Increased synthesis of plasminogen activator inhibitor-1 by cultured human endothelial cells exposed to native and modified LDLs. an LDL receptor-independent phenomenon. *Arteriosclerosis, thrombosis, and vascular biology,* 13(3):338-346.

Ubbink, J. B., Christianson, A., Bester, M. J., Van Allen, M. I., Venter, P. A., Delport, R., & Vermaak, W. J. 1999. Folate status, homocysteine metabolism, and methylene tetrahydrofolate reductase genotype in rural South African blacks with a history of pregnancy complicated by neural tube defects. *Metabolism*, 48(2), 269-274.

Ueland, P.M., Refsum, H. & Brattstrom, L. 1992. Plasma homocysteine and cardiovascular disease. *Atherosclerotic cardiovascular disease, hemostasis, and endothelial function. New York: Marcel dekker,* 183.

Van Emden, H. 2009. *Statistics for terrified biologists*. Wiley. com.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A. & Holt, R.A. 2001.The sequence of the human genome. *Science,* 291(5507):1304-1351.

Via, M., Gignoux, C. & Burchard, E.G. 2010. The 1000 genomes project: New opportunities for research and social challenges. *Genome med,* 2(3).

Viiri, L.E., Loimaala, A., Nenonen, A., Islam, S., Vuori, I., Karhunen, P.J. & Lehtimäki, T. 2005. The association of the apolipoprotein E gene promoter polymorphisms and haplotypes with serum lipid and lipoprotein concentrations. *Atherosclerosis,* 179(1):161-167.

Viiri, L.E., Viiri, K.M., Ilveskoski, E., Huhtala, H., Mäki, M., Tienari, P.J., Perola, M., Lehtimäki, T. & Karhunen, P.J. 2008. Interactions of functional apolipoprotein E gene promoter polymorphisms with smoking on aortic atherosclerosis. *Circulation: Cardiovascular genetics,* 1(2):107-116.

Viitanen, L., Pihlajamäki, J., Miettinen, R., Kärkkäinen, P., Vauhkonen, I., Halonen, P., Kareinen, A., Lehto, S. & Laakso, M. 2001. Apolipoprotein E gene promoter (–219G/T) polymorphism is associated with premature coronary heart disease. *Journal of molecular medicine,* 79(12):732-737.

Von Känel, R., Mills, P. J., Fainman, C., & Dimsdale, J. E. 2001. Effects of Psychological Stress and Psychiatric Disorders on Blood Coagulation and Fibrinolysis A Biobehavioral Pathway to Coronary Artery Disease?. *Psychosomatic Medicine*, 63(4), 531-544.

Vorster, H.H. & Nel, T.A. 2001. Make starchy foods the basis of most meals. *South African journal of clinical nutrition*, 14(3):17-23.

Vorster, H.H., Bourne, L.T., Venter, C.S. & Oosthuizen, W. 1999. Contribution of nutrition to the health transition in developing countries: A framework for research and intervention. *Nutrition reviews,* 57(11):341-349.

Vorster, H. H., Jerling, J. C., Steyn, K., Badenhorst, C. J., Slazus, W., Venter, C. S., & Bourne, L. T. 1998. Plasma fibrinogen of black South Africans: the BRISK study. *Public health nutrition*, 1(03), 169-176.

Vorster, H.H., Venter, C.S., Wissing, M.P. & Margetts, B.M. 2005. The nutrition and health transition in the north west province of South Africa: A review of the THUSA (transition and health during urbanisation of South Africans) study. *Public health nutrition-cab international-,* 8(5):480.

Vorster, H. 2002. The emergence of cardiovascular disease during urbanisation of Africans. *Public health nutrition,* 5(1A):239-244.

Walker, A. & Sareli, P. 1997. Coronary heart disease: Outlook for Africa. *Journal of the royal society of medicine,* 90(1):23.

Wang, Q. 2005. Molecular genetics of coronary artery disease. *Current Opinion in Cardiology*, 20(3): 182-188.

Wang, D.G., Fan, J., Siao, C., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E. & Spencer, J. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome.*Science,* 280(5366):1077-1082.

Ware, J.H. 2006.The limitations of risk factors as prognostic tools.*N engl J med,* 355(25):2615-2617.

Weintraub, M.S., Eisenberg, S. & Breslow, J.L. 1987. Dietary fat clearance in normal subjects is regulated by genetic variation in apolipoprotein E. *Journal of clinical investigation,* 80(6):1571.

Weisgraber, K. H. 1994. Apolipoprotein E: structure-function relationships. *Advances in protein chemistry*, 45, 249-302.

Weisgraber, K.H., Innerarity, T. & Mahley, R. 1982. Abnormal lipoprotein receptor-binding activity of the human E apoprotein due to cysteine-arginine interchange at a single site. *Journal of biological chemistry,* 257(5):2518-2521.

Wieland, H., Funke, H., Krieg, J., Luley, C. 1991. ApoE3-Freiburg and apoE4-Freiburg are two genetic apoE variants which are caused by exchanges of uncharged amino acids and do not appear to be associated with lipid disorders or heart disease. Abstract Book of the Ninth International Symposium on Atherosclerosis. Rosemont, Illinois, p164.

Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath, S.C., Timpson, N.J., Najjar, S.S. & Stringham, H.M. 2008. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature genetics,* 40(2):161-169.

Williams, G.C. & Nesse, R.M. 1991. The dawn of Darwinian medicine. *Quarterly review of biology,* 1-22.

Wilson, P.W. 2000. Lipids, lipases, and obesity does race matter? *Arteriosclerosis, thrombosis, and vascular biology,* 20(8):1854-1856.

Wilson, P.W.F., D'agostino, P.B., Levy, D., Belanger, A.M., Silbershatz, H. & Kannel, W.B. 1998. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Journal of the American Heart Association,* 97:1837-1847.

Winkleby, M. A., Robinson, T. N., Sundquist, J., & Kraemer, H. C. 1999. Ethnic variation in cardiovascular disease risk factors among children and young adults. *The journal of the American Medical Association*, 281(11), 1006-1013.

Wolmarans, P. & Oosthuizen, W. 2001. Eat fats sparingly - implications for health and disease. *South African journal of clinical nutrition,* 14(3):48-54.

World Health Organization. 2010. *World health statistics 2010*. World Health Organization.

World Heatlh Organization. 2011. *Noncommunicable Diseases Country Profiles 2011*. World Health Organization.

Wu, K., Bowman, R., Welch, A. A., Luben, R. N., Wareham, N., Khaw, K. T., & Bingham, S. A. 2007. Apolipoprotein E polymorphisms, dietary fat and fibre, and serum lipids: the EPIC Norfolk study. *European heart journal*, 28(23), 2930-2936.

Wyndham, C. 1979. Mortality from cardiovascular diseases in the various population groups in the republic of South Africa.

Yach, D., Hawkes, C., Gould, C.L. & Hofman, K.J. 2004. The global burden of chronic diseases. *The journal of the American medical association,* 291(21):2616-2622.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, 13(1), 134.

Ye, S., Dunleavey, L., Bannister, W., Day, L.B., Tapper, W., Collins, A.R.; Day, I.N.M., & Simpson, I. 2003. Independent effects of the -219G4T and e2/e3/e4 polymorphisms in the apolipoprotein E gene on coronary artery disease: The Southampton Atherosclerosis Study. *European Journal of Human Genetics*, 11: 437–443.

Yusuf, S., Hawken, S., Ôunpuu, S., Dans, T., Avezum, A., Lanas, F., Mcqueen, M., Budaj, A., Pais, P. & Varigos, J. 2004. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): Case-control study. *The lancet,* 364(9438):937-952.

Yusuf, S., Reddy, S., Ôunpuu, S. & Anand, S. 2001a. Global burden of cardiovascular diseases part I: General considerations, the epidemiologic transition, risk factors, and impact of urbanization. *Circulation,* 104(22):2746-2753.

Yusuf, S., Reddy, S., Ôunpuu, S. & Anand, S. 2001b. Global burden of cardiovascular diseases part II: Variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies. *Circulation,* 104(23):2855-2864.

Zannis, V.I., Breslow, J.L., Utermann, G., Mahley, R.W., Weisgraber, K.H., Havel, R.J., Goldstein, J.L., Brown, M.S., Schonfeld, G. & Hazzard, W. 1982.Proposed nomenclature of apoEisoproteins, apoE genotypes, and phenotypes. *Journal of lipid research,* 23(6):911-914.

Zatu, M., Van Rooyen, J. & Schutte, A. 2011. Smoking and vascular dysfunction in Africans and Caucasians from South Africa. *Cardiovascular journal of Africa,* 22(1):18.

Zerba, K., Ferrell, R., & Sing, C. 2000. Complex adaptive systems and human health: the influence of common genotypes of the apolipoprotein E (ApoE) gene polymorphism and age on the relational order within a field of lipid metabolism traits. *Human genetics*, 107(5), 466-475.

Zoratti, R. 1998. A review on ethnic differences in plasma triglycerides and high-density-lipoprotein cholesterol: Is the lipid pattern the key factor for the low coronary heart disease rate in people of African origin? *European journal of epidemiology,* 14(1):9-21.

INTERNET SOURCES:

(http://www.afesis.org.za/Sustainable-Settlements-Articles/beyond-the-rural-and-urban-development-frameworks-theorising-the-relationship, 10 October 2013).

IHS Global Insight Southern Africa:
http://www.ihsglobalinsight.co.za/News/news.asp?id=1026

Statistical release (P0302). Mid-year population estimates 2011**:**
(http://www.statssa.gov.za/publications/P0302/P03022011.pdf

Statistical release (P03093). Mortality and causes of death in South Africa, 1997–2003:
http://beta2.statssa.gov.za/publications/P03093/P030932005.pdf
http://www.statssa.gov.za/Publications/P03093/P030932010.pdf

South African Demographic and Health Survey, 2003:
http://www.doh.gov.za/docs/reports/2003/sadhs2003/part1.pdf

NucleoFast 96 PCR Clean**-**up User Manual - Clontech Laboratories:
www.clontech.com/xxclt_ibcGetAttachment.jsp?cItemId=17420

Statistical release (P0301.4). Census 2011 results:
http://www.statssa.gov.za/Publications/statsdownload.asp?PPN=P0301.4

Census 2001 Key results: http://www.statssa.gov.za/Publications/CensusKey/CensusKey2001.pdf

http://community.hirebiomed.com/posts/detail/methods-for-determining-dna-yield-and-concentratio

n-4fbbcb5151fed2d37400001e

http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1081101

http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=405509

http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=440446

http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=429358

http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=7412)