

The association of LDLR and PCSK9 variants with LDL-c levels in a black South African population in epidemiological transition

T van Zyl
10795626

Thesis submitted for the degree Doctor of Philosophy in
Dietetics at the Potchefstroom Campus of the North-West
University

Promoter: Prof JC Jerling
Co-Promoter: Dr KR Conradie
Assistant promoter: Prof EJM Feskens

November 2013

ABSTRACT

Background

Elevated concentrations of low-density lipoprotein cholesterol (LDL-c) are a major risk factor for the development of coronary artery disease (CAD) because of their role in the progression of atherosclerosis. The black South African population is known to have had historically low LDL-c and in the past there was almost no CAD in the population. However, as this population moves through the nutrition transition, LDL-c levels are increasing. LDL-c levels are regulated by the LDL receptors, which is the major protein involved with transporting cholesterol across cell membranes in humans. Proprotein convertase subtilisinlike/kexin type 9 (PCSK9) is another protein involved with the regulation of LDL-c through its role in assisting with the degradation of the LDL receptor. Variants in both genes can cause elevated or lowered LDL-c levels. Very little information is available on the frequency or presence of variants in the low-density lipoprotein receptor (LDLR) and PCSK9 gene in the black South African population and on how these variants associate with LDL-c. The main aim of the study was thus to determine novel and existing genetic variants in these two genes and to describe the manner in which they associate with plasma LDL-c levels in a black South African population undergoing an epidemiological transition.

Methods

The 2005 baseline data from the Prospective Urban and Rural (PURE) study population were used in this study. The study population consisted of apparently healthy black volunteers from the North West province of South Africa, aged 35 to 60 years. Thirty individuals were randomly chosen from the 1860 volunteers to determine the presence of known and novel variants in these genes by automated bidirectional sequencing. The promoter region, exons and flanking regions were sequenced and variants were identified utilising CLC DNA Workbench. Deoxyribonucleic acid (DNA) samples for 1500 individuals of the PURE study population were genotyped by means of a Golden Gate Genotyping Assay. Analyses of covariance (ANCOVA) were used to test for associations between the different genotypes in both the LDLR and PCSK9 genes and LDL-c levels. Haplotypes were generated by using the confidence intervals on the software programme, HaploView. A genetic risk score (GRS) was determined by including variants which associated significantly with LDL-c. The GRS, the haplotypes and the variants that associated significantly with LDL-c were used in separate linear regression models with variants which correlated with LDL-c to determine how all these variables contribute to the differences in LDL-c levels.

Results and discussion

Novel and known variants were identified in both the genes and in total 52 variants were genotyped. Rare variants such as rs17249141 and rs28362286 were detected in the study population and are associated with low levels of LDL-c. The variants identified in the LDLR gene were situated largely in regulatory regions such as the promoter, intron and 3'untranslated regions. Haplotypes in the *LDLR* gene with the highest frequency associated with lower LDL-c levels, which could contribute to the study population's low mean LDL-c level. Haplotypes identified in the PCSK9 gene had a weaker association with LDL-c levels. The minor allele frequencies of many of the variants differed from those of the European population and therefore the importance of population-specific research cannot be sufficiently emphasised. The GRS, haplotypes and variants used in the regression models to determine whether they contributed to predicting the variance in LDL-c in the study population made a small contribution to explaining this. BMI best explained the variance in LDL-c levels. Older women with a body mass index (BMI)>25kg/m² were identified as being at greater risk of developing elevated LDL-c levels than the rest of the study population. Heterozygote carriers of variant, rs28362286, had 0.787 mmol/L lower LDL-c than carriers of the wild type and this is associated with a reduced risk of developing CAD.

Conclusion and recommendation

When considering the results mentioned above, adding genetic analysis to explaining the variance in LDL-c levels seems to have its limitations, but the study included only two of many genes that play a role in the metabolism and regulation of LDL-c levels. Incorporating more genes and more variants into analyses and prediction models will add greater value to defining LDL-c levels. Rarer variants with a large impact on protein function, such as rs28362286, have a greater effect on LDL-c levels and could predict the variance better than the common variants. Risk factors such as BMI can also still be trusted to indicate which individuals or groups are at risk of developing elevated LDL-c levels. Health advice should be given to appropriate target groups such as older women with a BMI >25kg/m² in order to prevent CAD from becoming a burden in this population.

KEY WORDS: Low-density lipoprotein cholesterol, LDLR gene, PCSK9 gene

OPSOMMING

Agtergrond

Verhoogde lae digtheid lipoproteïen cholesterol (LDL-c) konsentrasies is een van die hoof risiko faktore vir die ontwikkeling van koronêre arteriële siekte (CAD) deur die rol wat dit speel in die verloop van aterosklerose. Die swart Suid-Afrikaner populasie het histories lae LDL-c gehad en CAD was byna afwesig. Verstedeliking is egter besig om plaas te vind in hierdie populasie en daarmee saam verhoog LDL-c vlakke. LDL-c vlakke word gereguleer deur LDL reseptore, hierdie proteïen is hoofsaaklik verantwoordelik vir die transport van LDL-c oor selmembrane. Proprotein convertase subtilisinlike/kexin type 9 (PCSK9) is nog 'n proteïen wat betrokke is by die regulering van LDL-c deur die rol wat die proteïen speel in die afbraak van die LDL reseptor. Variante in beide die gene veroorsaak verhoogde en verlaagde LDL-c vlakke. Inligting rakende die frekwensie en teenwoordigheid en assosiasie met LDL-c vlakke van variante in die lae digtheid lipoproteïen cholesterol reseptor (*LDLR*) en *PCSK9* gene in die swart Suid-Afrikaner populasie is skaars. Die hoofdoel van hierdie projek was om te bepaal of daar nuwe en bekende variante in hierdie twee gene teenwoordig is en hoe hierdie variante met LDL-c assosieer in 'n swart Suid-Afrikaner populasie wat besig is om te verstedelik.

Metode

Die 2005 basislyn data van die Prospective Urban and Rural (PURE) studiepopulasie is in hierdie studie gebruik. Die PURE populasie bestaan uit klaarblyklike gesonde vrywilligers vanuit die Noordwes Provinsie van Suid-Afrika tussen die ouderdomme van 35 en 60 jaar. Dertig individue is wederkerig gekies vanuit die PURE studiepopulasie om te bepaal of nuwe en bekende variante teenwoordig is deur gebruik te maak van geoutomatiseerde DNS volgorde bepaling. Die deoksiribonukleïnesuur (DNS) volgorde van die promoter area, eksons en die ekson-intron grense is bepaal en die variante is geïdentifiseer deur gebruik te maak van die sagteware, CLC DNA Workbench. DNS monsters vir 1500 individue van die PURE populasie is gegenotipeer met behulp van 'n Golden Gate Genotyping Assay. Analise van ko-variantsie (ANCOVA) is gebruik om te toets vir assosiasies tussen die verskillende genotipes in beide die *LDLR* en *PCSK9* gene en LDL-c vlakke. Haplotipes is gegenereer met behulp van vertrouensintervalle op die sagteware, Haploview. 'n Genetiese risikotelling is bepaal deur die variante in te sluit wat noemenswaardig met LDL-c vlakke assosieer. Die genetiese risikotelling, haplotipes en die variante wat noemenswaardig met LDL-c geassosieer het, is in afsonderlike liniêre regressie modelle gebruik

saam met veranderlikes wat noemenswaardig met LDL-c gekorreleer het om te bepaal tot watter mate al hierdie veranderlikes bydra tot variasie in LDL-c vlakke.

Resultate en bespreking

Nuwe en bekende variante is in beide die gene geïdentifiseer en die studiepopulasie is in totaal vir 52 variante gegenotipeer. Seldsame variante soos rs17249141 en rs28362286 is in die studiepopulasie geïdentifiseer en is met lae LDL-c vlakke geassosieer. Die variante wat in die LDLR geen geïdentifiseer is, is meestal geleë in regulerende areas soos die promoter, intron en 3' ongetransleerde streek. Haplotipes in die LDLR geen met die hoogste frekwensie het met laer LDL-c vlakke geassosieer wat moontlik kan bydra tot die studiepopulasie se gemiddelde lae LDL-c vlakke. Haplotipes wat in die PCSK9 geen gevind is, het swakker assosiasies met LDL-c vlakke getoon. Die frekwensies van die mindere allele het verskil van die Europese populasie en daarom kan populasies spesifieke genetiese navorsing nie genoeg beklemtoon word nie. Die genetiese risikotelling, haplotipes en variante wat in die regressie modelle gebruik is, het min bygedra tot die verduideliking in verskille in LDL-c vlakke in die studiepopulasie. LMI het die verskil in LDL-c vlakke die beste verduidelik. Ouer vroue (>50 jaar) met 'n liggaamsmassa indeks (LMI) >25 kg/m² is ook geïdentifiseer as die groep met die grootste risiko om verhoogde LDL-c vlakke te ontwikkel. Heterosigoot draers van die variant, rs28362286, het LDL-c vlakke van 0.787 mmol/L laer as die wilde-tipe gehad en hierdie verlaging word geassosieer met 'n verminderde risiko om CAD te ontwikkel.

Gevolgtrekking en aanbeveling

Wanneer die bogenoemde resultate oorweeg word wil dit voorkom asof die byvoeging van genetiese analise min bydra tot die verduideliking van die variasie in LDL-c vlakke, maar hierdie studie sluit slegs twee van vele gene in, wat 'n rol speel in die metabolisme en regulasie van LDL-c vlakke. Deur meer gene en variante in analyses en voorspellingsmodelle in te sluit sal groter waarde toevoeg tot die definiëring van LDL-c vlakke. Seldsame variante met 'n groter impak op proteïenfunksie, soos rs28362286, het 'n groter effek op LDL-c vlakke en hierdie variante verduidelik die verskille beter as die variante wat meer algemeen voorkom. Risiko faktore soos LMI kan ook nog met vertroue gebruik word om aan te dui watter individue of groepe 'n risiko loop om verhoogde LDL-c vlakke te ontwikkel. Gesondheidsriglyne moet ook aan die gepaste teikengroepe deurgegee word soos ouer vroue met 'n LMI >25 kg/m² om sodoende te verhoed dat CAD 'n probleem in hierdie populasie word.

ACKNOWLEDGEMENTS

I want to thank God, my Heavenly Father, for this wonderful opportunity of growth and self-actualisation. Thank You for blessing me with the abilities to complete this challenge.

I would also like to thank the following people, without whom completion of this thesis would not have been possible:

- My promoter, Prof. Johann Jerling, for the opportunity to develop independently through his guidance.
- My co-promoter, Dr Karin Conradie, for her guidance and the indispensable support with the laboratory work.
- Prof. Edith Feskens, my assistant promoter, for her assistance with the statistical analysis. Meeting you has changed my view on statistics in nutrition research.
- Prof Annamarie Kruger and her team from the South African PURE study for the opportunity to work with the data from the PURE study as well as every single person that volunteered and participated in the study.
- Marius Smuts and Marlien Pieters for their support, understanding and general guidance in the PhD process.
- Wayne Towers for obtaining funding for the project and assisting when the genetics got the better of me.
- My colleagues in the Nutrition Department for their continuous support and words of encouragement.
- Mary Hoffman for the language editing and her patience as the chapters were sent through gradually.
- My parents for always believing that I can overcome every challenge I put my mind to and for having laid the steadfast foundation of my faith in God. Without faith and hope I would not have been able to achieve this milestone.
- My brothers and their families for their support and encouragement.
- My husband for all his support and for suffering through all the moods swings and times of frustration. Thank you for reminding me that there is life beyond this thesis.

I will lift up my eyes to the hills.

From whence comes my help?

My help comes from the Lord,

Who made heaven and earth.

Psalms 121 (NKJV)

CONTENT

ABSTRACT	i
OPSOMMING	iii
ACKNOWLEDGEMENT	v
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF ANNEXURES	xii
LIST OF ABBREVIATIONS	xiii
 CHAPTER 1 – INTRODUCTION	 1
1.1 BACKGROUND AND MOTIVATION	1
1.2 AIMS AND OBJECTIVE	3
1.3 STRUCTURE OF THESIS	4
1.4 CONTRIBUTION OF THE AUTHORS	5
 CHAPTER 2 – LITERATURE REVIEW	 6
2.1 NUTRITION TRANSITION	7
2.2 DIETARY/LIFESTYLE CHANGES DURING THE NUTRITION TRANSITION	8
2.3 EMERGENCE OF CVD IN THE BLACK SOUTH AFRICAN POPULATION	10
2.4 THE ROLE OF LDL IN CAD	12
2.4.1 Atherosclerosis is a complex and multifactorial disease	13
2.4.1.1 LDL-c and the vascular endothelium	13
2.4.1.2. LDL-c and extracellular matrix	15

2.4.1.3. Plaque development.....	16
2.4.1.4 Plaque stability	18
2.4.1.5 Platelet aggregation and atherothrombosis.....	19
2.4.1.6 Activation of the coagulation cascade.....	20
2.5 DETERMINANTS OF LDL-C LEVELS.....	21
2.5.1 The low-density lipoprotein (LDL) receptor	25
2.5.2 The <i>LDLR</i> gene	27
2.5.3 Classification of <i>LDLR</i> mutations	29
2.5.4 Frequency of <i>LDLR</i> variants in the black South African population.....	30
2.5.5 Regulation of the LDLR receptor by PCSK9	35
2.5.5.1 The <i>PCSK9</i> gene	35
2.5.5.2 PCSK9 is a serine protease.....	35
2.5.5.3 PCSK9 function	36
2.5.5.4 Classification of PCSK9 mutations	38
2.5.5.5 Circulating PCSK9.....	38
2.5.5.6 PCSK9 crystal structure	40
2.6 SUMMARY	40
 CHAPTER 3 – COMMON AND RARE SINGLE NUCLEOTIDE POLYMORPHISMS IN THE <i>LDLR</i> GENE ARE PRESENT IN A BLACK SOUTH AFRICAN POPULATION AND ASSOCIATE WITH LDL-c LEVELS	 41
ABSTRACT	43
INTRODUCTION.....	43

METHODS AND MATERIALS.....	44
RESULTS.....	47
DISCUSSION.....	53
ACKNOWLEDGEMENTS.....	59
REFERENCES.....	59

CHAPTER 4 – THE ASSOCIATION OF COMMON AND RARE VARIANTS IN THE *PCSK9* GENE WITH LDL-CHOLESTEROL IN A BLACK SOUTH AFRICAN POPULATION..... 65

ABSTRACT	67
INTRODUCTION.....	67
METHODS AND MATERIALS.....	68
RESULTS.....	71
DISCUSSION.....	78
ACKNOWLEDGEMENTS.....	83
REFERENCES.....	84

CHAPTER 5 – PREDICTING LDL-CHOLESTEROL LEVELS IN A BLACK SOUTH AFRICAN POPULATION IN TRANSITION..... 89

ABSTRACT	91
INTRODUCTION.....	91
METHODS AND MATERIALS.....	92
RESULTS.....	96
DISCUSSION.....	107

ACKNOWLEDGEMENTS.....	111
REFERENCES.....	111
 CHAPTER 6 – GENERAL DISCUSSION, CONCLUSION AND RECOMMENDATIONS.....	117
6.1 INTRODUCTION.....	117
6.2 DETERMINING IMPORTANT GENETIC VARIANTS IN THE <i>PCSK9</i> AND <i>LDLR</i> GENE....	118
6.3 GENOTYPING THE STUDY POPULATION FOR SNPS IDENTIFIED AND DETERMINING HARDY-WEINBERG EQUILIBRIUM	120
6.4 THE ASSOCIATION BETWEEN LDL-c LEVELS AND THE DIFFERENT GENOTYPES OF EACH SNP IN THE <i>PCSK9</i> AND <i>LDLR</i> GENES AND MAF COMPARISONS.....	121
6.5. THE PRESENCE OF HAPLOTYPES AND THE ASSOCIATION BETWEEN LDL-c LEVELS AND THE DIFFERENT HAPLOTYPES	123
6.6 FACTORS PREDICTING LDL-c LEVELS.....	123
6.7 STRENGTHS AND LIMITATIONS.....	126
6.8 RECOMMENDATIONS	126
6.9 GENERAL CONCLUSION.....	127
 REFERENCES.....	128
ANNEXURES	149

LIST OF TABLES

Chapter 1

Table 1.1 List of members within the research team and their contributions	5
--	---

Chapter 2

Table 2.1 Summary of major genes affecting plasma LDL-c concentrations	23
Table 2.2 <i>LDLR</i> variants identified in black South Africans	34

Chapter 3

Table1 Population characteristics	48
Table 2 Minor allele frequency of the various SNPs in <i>LDLR</i> in different populations	49
Table 3 <i>LDLR</i> genotype distribution and association with LDL-c	50
Table 4 Haplotypes associated with LDL-c within the PURE population	54

Chapter 4

Table1 Population characteristics	72
Table 2 Information on <i>PCSK9</i> variants found in the PURE population	73
Table 3 Haplotypes associated with LDL-c within the PURE population	78

Chapter 5

Table 1 Population characteristics	96
Table 2 LDL-c levels and BMI according to age and gender	100
Table 3 Linear regression model for BMI, gender, age with LDL-c as the dependent variable	102
Table 4 Variants in the <i>LDLR</i> and <i>PCSK9</i> genes associated with LDL-c	103
Table 5 Linear regression model for GRS with LDL-c as the dependent variable	104
Table 6 Summary of risk alleles and the mean LDL-c levels per group	105
Table 7 Linear regression model for SNPs with LDL-c as the dependent variable	106
Table 8 Linear regression model for haplotypes with LDL-c as the dependent variable	107

LIST OF FIGURES

Chapter 2

Figure 2.1 The sequence of events generating the fatty streak lesion	14
--	----

Chapter 5

Figure 1 Mean LDL-c according to BMI and age categories in women	101
--	-----

LIST OF ANNEXURES

ANNEXURE A: ETHICAL APPROVAL	149
ANNEXURE B: INFORMED CONSENT FORM PHASE 1.....	149
ANNEXURE C: INFORMED CONSENT FORM PHASE 2	150
ANNEXURE D: VARIANTS IDENTIFIED THROUGH SEQUENCING IN THE <i>LDLR</i> GENE	152
ANNEXURE E: PRIMERS DESIGNED FOR <i>LDLR</i> PCR AND SEQUENCING	154
ANNEXURE F: LD PLOT FROM HAPLOVIEW FOR THE <i>LDLR</i> GENE.....	155
ANNEXURE G: HAPLOTYPES IDENTIFIED IN THE <i>LDLR</i> GENE BY MEANS OF HAPLOVIEW	156
ANNEXURE H: VARIANTS IDENTIFIED THROUGH SEQUENCING IN THE <i>PCSK9</i> GENE ..	157
ANNEXURE I: PRIMERS DESIGNED FOR <i>PCSK9</i> PCR AND SEQUENCING	159
ANNEXURE J: LD PLOT FROM HAPLOVIEW FOR THE <i>PCSK9</i> GENE	160
ANNEXURE K: HAPLOTYPES IDENTIFIED IN THE <i>PCSK9</i> GENE BY MEANS OF HAPLOVIEW	261
ANNEXURE L: CORRELATIONS AND LINEAR REGRESSIONS FOR CHAPTER 5	162
ANNEXURE M: GUIDE FOR AUTHORS FOR THE JOURNAL OF HUMAN GENETICS...	174
ANNEXURE N: GUIDE FOR AUTHORS FOR THE EUROPEAN JOURNAL OF HUMAN GENETICS.....	189
ANNEXURE O: GUIDE FOR AUTHORS FOR GENES AND NUTRITION	202

LIST OF ABBREVIATIONS

3'UTR	3'untranslated region
A	Adenine
ACAT	Acyl-CoA cholesterol acyltransferase
ADP	Adenosine diphosphate receptor
AMI	Acute myocardial infarction
APOB	Apo lipoprotein B
APOE	Apo lipoprotein E
AREs	AU-rich elements
ARH	Adaptor protein
ANOVA	Analysis of variance
ANCOVA	Analysis of covariance
BMI	Body mass index
BRISK	Coronary Heart Disease Risk Factor Study
C	Cytosine
Ca ²⁺	Calcium
CAD	Coronary artery disease
CCR2	CC chemokine receptor 2
CHD	Coronary heart disease

CHO	Carbohydrates
CRIBSA	Coronary risk in black South Africans Study
CVD	Cardiovascular disease
CXCR2	CX chemokine receptor 2
DNA	Deoxyribonucleic acid
DNS	Deoksiribonukleïnesuur
DQS	Diet quality scores
ECM	Extracellular matrix
ER	Endoplasmic reticulum
EGF	Epidermal growth factor
EGF-A	Epidermal growth factor-like repeat homology domain
FH	Familial hypercholesterolaemia
G	Guanine
GGGT	Golden Gate Genotyping
GOF	Gain-of-function
GRS	Genetic risk score
HDL-C	High density lipoprotein cholesterol

HMG CoA	3-hydroxy-3-methylglutaryl coenzyme A
HL	Hepatic lipase
HWE	Hardy-Weinberg equilibrium
HIV/AIDS	Human immunodeficiency virus/Acquired immune deficiency syndrome
ICAM-1	Intracellular adhesion molecule 1
IDL	Intermediate-density lipoproteins
IHD	Ischaemic heart disease
kDa	Kilo dalton
kg	Kilogram
LD	Linkage disequilibrium
LDL	Low density lipoprotein
LDL-c	Low density lipoprotein cholesterol
<i>LDLR</i>	Low density lipoprotein receptor gene
LOF	Loss-of-function
LMI	Liggaamsmassa indeks
LPL	Lipoprotein lipase
LRP-1	Low-density lipoprotein receptor-related protein-1
MAF	Minor allele frequency

MCP-1	Monocyte chemo-attractant protein 1
MCSF	Monocyte colony-stimulating factor
MMP	Matrix-metalloproteinase (Homer <i>et al.</i> , 2008)
mRNA	Messenger ribonucleic acid
MTP	Microsomal triglyceride transfer protein
NA	Not available
NF	Necrosis factor
NO	Nitric oxide
NPC1L1	Niemann-Pick C 1-like 1 protein
NR-NCDs	Nutrition-related non-communicable diseases
PAK	p21-activated kinase
PAR	Protease-activated receptors
PC	Proprotein convertases
PCR	Polymerase chain reaction
<i>PCSK9</i>	Proprotein convertase subtilisinlike/kexin type 9
pKa	Acid dissociation constant
PPAR α	Peroxisome proliferator-activated receptor alpha
PUFA	Polyunsaturated fatty acids
PURE	Prospective Urban and Rural Epidemiological study

SATFAT	Saturated fatty acids
Ser	Serine
SD	Standard deviation
SIFT	Sorting tolerant from intolerant
SMAC	Sequential Multiple Analyzer Computer
SMCs	Smooth muscle cells
SNP	Single nucleotide polymorphism
SRE-1	Sterol regulating element-1
SREBP	Sterol regulating element-binding proteins
T	Tyrosine
TC	Total cholesterol
TE	Total energy
TFG- β	Transforming growth factor- β
TG	Triglycerides
Thr	Threonine
THUSA	Transition in Health during Urbanization in South Africa Study
VCAM-1	Vascular cell-adhesion molecule 1
VLDL-c	Very low density lipoprotein cholesterol
vWf	Von Willebrand factor

CHAPTER 1 – INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Cardiovascular diseases (CVD) have only recently emerged in the black South African population (Sliwa *et al.*, 2008) and coronary artery disease (CAD) and hypertensive heart failure have been identified as ‘newer’ forms of heart disease in this population (Stewart *et al.*, 2011). These newer forms of heart disease are emerging in this population accompanied by the historically prevalent diseases such as rheumatic heart disease, the dilated cardiomyopathies, pulmonary heart disease, arrhythmias and infectious forms of heart disease as the population moves from rural to urban areas (Akinboboye *et al.*, 2003; Stewart *et al.*, 2011; Vorster, 2002). Changes in lifestyle and dietary intakes occur with urbanisation and these changes are accompanied by an increase in CAD risk factors such as hypercholesterolaemia (Oosthuizen *et al.*, 2002). Both obesity and the intake of foods high in total fat and saturated fats are associated with low-density lipoprotein cholesterol (LDL-c) levels, as noted in the INTERHEART study (Steyn *et al.*, 2005) and the THUSA study (conducted in the North West province of South Africa) (Steyn *et al.*, 2005; Vorster *et al.*, 2005). Hypercholesterolaemia is a major risk factor for the development of CAD through its role in the aetiology of atherosclerosis (Steinberg, 2002). Patients with familial hypercholesterolemia develop accelerated atherosclerosis as a result of their elevated LDL-c levels and this leads to mature development of CAD (Nordlie *et al.*, 2005). Various studies which lowered cholesterol levels with statin therapy have shown that a 10% reduction in cholesterol levels will reduce deaths from CAD by 15% (Gould *et al.*, 2007; MRC/BHF Heart protection study, 2002; Gould *et al.*, 1998). Although the prevalence of CAD in the black African population has been historically low (Walker & Sareli, 1997) current research suggest that elevated LDL-c levels are also present in black South African CAD patients (Dolman *et al.*, 2011) and therefore these individuals are not exempted from this risk factor. Current guidelines on dyslipidaemia include rigorous management of LDL-c levels to prevent the progression of CAD (Reiner *et al.*, 2011) as a recent study showed that life-long reduction of LDL-c lowers the risk of developing CAD (Ference *et al.*, 2012).

LDL-c levels are affected by weight (Haslam & James, 2005) and diet, as previously mentioned, as well as by physical activity (Grundy, 2005), age (McNamara *et al.*, 1987) and gender (Heiss *et al.*, 1980). Apart from these factors, LDL-c levels are also regulated on a genetic level and as much as 50% of the differences between individuals are due to genetics (Burnett & Hooper, 2008). Various genes have been identified as having an effect on plasma LDL-c levels and the *LDLR* gene

produces the most important importation protein of LDL-c, namely the LDL receptor (Brown & Goldstein, 1986), which facilitates the binding, removal and degradation of lipoproteins such as LDL and intermediate density lipoprotein (IDL). Variants within this gene were first identified as the cause of familial hypercholesterolaemia (Brown *et al.*, 1981) and currently more than 1200 variants have been identified (Fokkema *et al.*, 2005). The frequency of variants in the *LDLR* gene in a black South African population is not known as the variants that have been identified were either in single individuals or in the probands and their families (Leitersdorf *et al.*, 1988, Peeters *et al.*, 1998, Scholtz *et al.*, 1999, Thiart *et al.*, 2000). Some known variants are the cause of hypercholesterolaemia (Leitersdorf *et al.*, 1988) and some found in regulatory regions such as the promoter have a LDL-c-lowering effect (Scholtz *et al.*, 1999). Determining the frequency and association of variants in the *LDLR* gene would give more insight into the role that this gene plays in regulating LDL-c levels in this population.

Another gene identified as having a direct effect on LDL-c was the *PCSK9* gene. This gene is responsible for producing the protein, PCSK9, which is involved in the post-translational degradation of the LDL receptor (Zaid *et al.*, 2008). PCSK9 binds to the LDL receptor on the cell surface and the LDLR-PCSK9 complex is internalised and moved to the lysosome where the whole complex is degraded, prohibiting the recycling of the LDL receptor back to the cell surface. Gain-of-function (GOF) variants identified in the *PCSK9* gene are associated with elevated LDL-c levels as fewer LDL receptors are available on the cell surface to clear LDL-c from the circulation. Loss-of-function (LOF) variants cause defective proteins that are either not released into the circulation or have a less effective function; these variants are associated with lower levels of LDL-c as a greater number of receptors are available on the cell surface to remove LDL-c from the circulation (Abifadel *et al.*, 2003, Abifadel *et al.*, 2009). Rare variants that cause FH have been identified in various populations (Abifadel *et al.*, 2003, Homer *et al.*, 2008, Kotowski *et al.*, 2006, Miyake *et al.*, 2008) and rare variants that cause very low LDL-c levels have been identified in individuals of African descent (Cohen *et al.*, 2005). Limited information is available on the frequency of variants in the *PCSK9* gene and how these associate with LDL-c.

The main aim of the study is to determine novel and existing genetic variants in the *PCSK9* and *LDLR* genes and to describe the manner in which they associate with plasma LDL-c levels in a black South African population undergoing an epidemiological transition. This aim will be addressed by investigating the entire study population from the Prospective Urban and Rural Epidemiological (PURE) study. The PURE study is a large-scale epidemiological study that recruited volunteers from rural and urban settings to track changes in lifestyles in 17 countries in

transition. Two communities (urban and rural) in the North-West Province of South Africa were chosen and over 2000 volunteers were randomly selected in 2005. Baseline data and blood samples for genetic tests were collected in 2005. The presence of known and novel variants, as well as the frequency and association with LDL-c levels, will be determined in both genes. Haplotypes, a genetic risk score (GRS) and the variants that associate significantly will be used to determine their contribution to the prediction of the variance in LDL-c, together with factors such as diet, physical activity, BMI, age and gender.

The results of this study will give us a better perspective on the contribution of these two genes in determining LDL-c levels in a population which is rapidly becoming urbanised. The results will also elucidate any differences between the frequencies of variants in the study population and populations of European descent. This will indicate whether population-specific genetic tools should be developed for the study population. The results will also show the value of using variants from two genes directly involved with LDL-c metabolism to predict variance in LDL-c levels together with the other factors (diet, BMI, age, gender and physical activity) that affect LDL-c levels.

1.2 AIMS AND OBJECTIVES

The main aim of the study is to determine novel and existing genetic variants in the *PCSK9* and *LDLR* genes and to describe the manner in which they associate with plasma LDL-c levels in a black South African population undergoing an epidemiological transition.

The objectives of the study are to:

- Determine the genetic variants of importance in the *PCSK9* and *LDLR* genes in 30 randomly chosen black South African individuals from the South African PURE study by means of bidirectional automated cycle sequencing.
- Screen the population of 2000 individuals from the South African PURE study for the selected single nucleotide polymorphisms (SNPs) in the *PCSK9* and *LDLR* genes identified through bidirectional automated cycle sequencing by means of BeadXpress array technology and to determine whether the population is in Hardy-Weinberg equilibrium for each gene.
- Investigate the association between LDL-c levels and the different genotypes in each examined SNP in the *PCSK9* and *LDLR* genes.

- Compare the minor allele frequencies (MAF) of these SNPs with those of the European population from the 1000 Genomes project.
- Investigate the presence of haplotypes in the *PCSK9* and *LDLR* genes and the association between LDL-c levels and the different haplotypes.
- Investigate which factors (diet, BMI, age, gender, genetic risk score, haplotypes or genotypes) best predict LDL-c levels in the study populations.

1.3 STRUCTURE OF THESIS

This thesis has been compiled according to the article format. Chapter 2 follows this introductory chapter and includes the literature review where relevant literature is presented with the focus on the following: the nutrition transition and dietary and lifestyle changes during the transition, CVD in the black South African population, LDL-c and its role in CAD, and regulation of LDL-c by *LDLR* and *PCSK9*. The additional information in this chapter will assist in interpreting the data reported in the articles included in the thesis.

Chapter 3 is an article with the title: Common and rare single nucleotide polymorphisms in the *LDLR* gene are present in a black South African population and associate with low-density lipoprotein cholesterol levels. This article has been published in the online version of the Journal of Human Genetics. The article reports on the variants that were genotyped in the *LDLR* gene and their association with LDL-c. The minor allele frequency (MAF) comparisons and haplotype data from the *LDLR* gene are also covered in this article

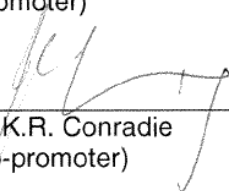
The article in Chapter 4 has the following title: The association of common and rare variants in the *PCSK9* gene with LDL-cholesterol in a black South African population. This article has been submitted to the European Journal of Human Genetics. Variants in the *PCSK9* gene and their associations with LDL-c levels which were genotyped are addressed in this article and the MAF comparisons and haplotype data are also reported in this article.

Chapter 5 is an article with the title: Predicting LDL-cholesterol levels in a black South African population in transition. This article has been submitted to Genes and Nutrition. In this article, factors that correlate with LDL-c levels in the study population are identified and used together with the haplotypes, SNPs and GRS identified in Chapters 3 and 4 to create a model to predict the difference in LDL-c levels.

1.4 CONTRIBUTION OF THE AUTHORS

The contributions of the researchers involved in the articles presented in this thesis are detailed in the following table.

Table 1.1: List of members within the research team and their contributions

Name and signature	Affiliation	Role in the study
Mrs Tertia van Zyl (PhD candidate) 	Centre of Excellence for Nutrition (CEN) within the School for Physiology, Nutrition and Consumer Science of the North-West University	Responsible for designing primers for both the LDLR and PCSK9 genes. Generating PCR products for sequencing. Identifying novel and know variants from sequences. Responsible for choosing variants for GoldenGate genotyping assay and preparing final genotyped data for statistical analysis. Responsible for literature reviews, statistical analysis, interpretation of data and writing up of publications and thesis
Prof. J.C. Jerling (Promoter) 	Centre of Excellence for Nutrition (CEN) within the School for Physiology, Nutrition and Consumer Science of the North-West University	Supervisor of T. van Zyl, assisted with data collection of the PURE study, interpretation of results and co-authored Chapters 3, 4 and 5.
Dr. K.R. Conradie (Co-promoter) 	Centre of Excellence for Nutrition (CEN) within the School for Physiology, Nutrition and Consumer Science of the North-West University	Co-Supervisor of T. van Zyl, assisted with data collection of the PURE study, laboratory analysis, genotyped data, interpretation of results and co-authored Chapters 3, 4 and 5.
 Prof. E.J.M. Feskens	Division of Human Nutrition; Wageningen University; Wageningen; The Netherlands	Assisted with statistical analysis, co-authored Chapters 3, 4 and 5.
*With my signature I declare that I approved the above-mentioned articles, that my role in the study, as indicated above, is representative of my actual contribution and that I hereby give consent that it may be published as part of the PhD thesis of Mrs T van Zyl.		

CHAPTER 2 - LITERATURE REVIEW

Cardiovascular disease (CVD) is a leading cause of death worldwide and account for the majority of deaths amongst all diseases and causes. Most of these deaths are caused by coronary heart disease (CHD) and stroke (WHO, 2013). CHD is caused by atherosclerosis as a result of elevated serum cholesterol levels (Steinberg, 2002). Hypercholesterolemia is a major risk factor for developing CVD especially CAD through the role it plays in the development of fatty streaks and fibrous plaques (Gould *et al.*, 2007).

Unhealthy dietary habits which include the intake of an excess fat especially saturated fat and a low intake of fibre are associated with elevated LDL-c levels and the development of atherosclerosis. The nutrition transition brings about a change in dietary intake especially in fat intake, fibre and fruits are also consumed in smaller amounts. Fats are consumed in greater amounts and the intake of saturated fat increases accordingly (MacIntyre *et al.*, 2002). Therefore the role of the nutrition transition (urbanisation) and the changes that take place in dietary intake during this phenomenon will be discussed first to set the scene for the emergence of CAD in the black South African population. The cholesterol levels of the black South African population have historically been reported as low and not deemed as a risk factor for developing CAD (Walker & Sareli, 1997). However as this population is rapidly becoming urbanised evidence have proven that risk factors for the development of CAD are emerging in this population (Akinboboye *et al.*, 2003). Patients with CAD have higher levels of LDL-c than their counterparts without CAD (Dolman *et al.*, 2011).

As LDL-c is a major risk factor for the early development of CAD this literature study will focus on LDL-c and the role it plays in the development of atherosclerosis. LDL-c levels are determined by various factors such as diet (Grundy, 2005), body mass index (BMI) (Haslam & James, 2005) and genetics (Burnett & Hooper, 2008). Up to 50% of the variance in LDL-c between individuals is of genetic origin (Burnett & Hooper, 2008). The *LDLR* (Brown & Goldstein, 1986) and the *PCSK9* (Zaid *et al.*, 2008) genes are directly involved with regulating serum LDL-c levels and therefore the remainder of the literature study will focus on the role these two genes play in determining LDL-c levels.

2.1 NUTRITION TRANSITION

Omran (1971) described the epidemiological transition as a shift in disease patterns of a population as mortality falls during the demographic transition (Omran, 1971). The nutrition transition, caused by a change in the types of foods consumed as well as a change in living environment, has an impact on the health of populations, communities and individuals. Changes in physical activity and lifestyle also have an effect on the health of individuals and are therefore not limited to dietary changes alone (Popkin, 2002). As reviewed by Amuna & Zotor (2008), this shift in disease patterns is associated with changes in behaviours, lifestyle, diets, physical activity, smoking and alcohol consumption. During this transition acute infectious disease lessens and chronic diseases of lifestyle, such as type II diabetes mellitus, cardiovascular diseases (CVD), stroke, high blood pressure, gout and certain cancers, increase, causing a steady shift in the age pattern of mortality, moving it from younger to older ages (Amuna & Zotor, 2008). The historical stages of and changes during the nutrition transition have been described by Popkin & Caballero (2002) as follows: Pattern 1 is the age of collecting food (diet high in carbohydrate and fibre and low in fat), Pattern 2 is the age of famine (acute scarcity of food and low dietary variety), Pattern 3 is the age of receding famine (fewer carbohydrates and more fruits and vegetables in the diet), Pattern 4 is the age of degenerative disease (nutrition-related chronic diseases of lifestyle, diet high in animal fat, cholesterol, sugar and refined carbohydrates) and Pattern 5 is the age of behavioural change to revise the diet and to reduce degenerative diseases and prolong health (healthier diet with less fat, more unrefined carbohydrates, fruits and vegetables) (Popkin & Caballero, 2002).

Developed countries such as Japan, the United States and the countries of Western Europe are in the later stages of the nutrition transition whereas developing countries are at different stages of the transition within their population, communities and households (Popkin, 2002). The developing countries like South Africa are moving much more quickly through these stages, with little time for health systems, policies and economies to adjust for the increase in nutrition-related non-communicable diseases (NR-NCDs) (Popkin, 2002). Simultaneously, other health problems occur in South Africa, so that the country has a double burden of disease, where infectious diseases co-exist with NR-NCDs. In countries such as South Africa, the primary health care system is not properly equipped to deal even with current health issues such as HIV/AIDS and tuberculosis and thus an increase in NR-NCDs will weigh heavily on such an inadequately established system (Popkin, 2002).

The changes that occur during the nutrition transition and which contribute to the rise in NR-NCDs will be discussed in the following paragraph.

2.2 DIETARY/LIFESTYLE CHANGES DURING THE NUTRITION TRANSITION

The nutrition transition results in an increased intake of more palatable foods containing high amounts of fat and sugar (Kruger *et al.*, 2001; Vorster *et al.*, 2000). The intake of foods high in fibre such as unrefined cereals, fruits and vegetables is inadequate and this low intake is associated with NR-NCDs (Kruger *et al.*, 2001; Vorster *et al.*, 2000). Fruit and vegetables may be more available in the urban areas but these food items are considered luxuries to poorer households that cannot afford them (Hawkes, 2006). In the South African setting, high-fat, high-energy foods are also available from local street vendors and not just from commercial outlets and these street vendors are more affordable than the commercial outlets, making these food items available to all socioeconomic classes (Feeley *et al.*, 2012). The cost of a healthy diet in South Africa is much higher than the unhealthy alternatives, as demonstrated by Temple & Steyn (2011).

The Transition, Health and Urbanisation in South Africa (THUSA) study conducted in the North West province of South Africa demonstrated how the dietary and nutrient intakes differed between the various groups at different stages of urbanisation (MacIntyre *et al.*, 2002). Across the five groups representing the stages of urbanisation they found that changes in dietary patterns occurred. The contribution of carbohydrates towards energy intake decreased from 67.2% in the rural and farm groups to 56.5% in the upper class urban group. Maize meal is the staple food of this population and the consumption patterns of maize as well as other cereal-based foods are different amongst the five groups, where, for the upper class group, the intake of maize meal contributes less to total energy than in the other groups (MacIntyre *et al.*, 2002). The lower contribution to energy intake from carbohydrates in the upper class urban group is caused by a higher intake of protein and fats. The intake of plant protein decreases as the intake of animal protein increases, and with that, the intake of fat also increases. Sugar intake was mostly the same across all groups, with the biggest difference being between those living on farms and the upper class urban group. In the upper class urban group the intake of sweets, cakes and sugary beverages was higher than in the rural and farm groups, which is indicative of a lack of variety in the diets of the rural and farm groups as they can afford mostly bulk purchases of basic foods such as sugar and maize on a monthly basis (MacIntyre *et al.*, 2002).

Total fat, saturated fat and monounsaturated fat intake as well as the proportion of energy provided by fat increased with urbanisation within the THUSA study. Fat intake in the rural group was 23.3% of total energy while in the upper class urban group it was 31.2% of total energy. The increases in the intakes of these nutrients were due to the high consumption of red meat in the upper class urban group (MacIntyre *et al.*, 2002).

The energy intake was similar in the rural groups and lower than in the urban groups of the THUSA study (MacIntyre *et al.*, 2002) although other researchers did report higher energy intakes (Vorster *et al.*, 1997). Increased energy intakes are associated with an increase in BMI (Stubbs & Lee, 2004). The prevalence of obesity in the participants in the THUSA study was determined mainly by household income, total energy intake, fat intake and low physical activity, as found by Kruger *et al.*, (Kruger *et al.*, 2002, Kruger *et al.*, 2003, Kruger *et al.*, 2001). Obesity was also associated with a number of cardiovascular risk factors such as increased total cholesterol (Kruger *et al.*, 2003; Oosthuizen *et al.*, 2002). Individuals most at risk were women who were physically inactive and of higher socioeconomic standing (Kruger *et al.*, 2003).

The small increase in fibre intake across the various groups was not significant. Even though more fruit and vegetables were part of the upper class urban group's diet, their fibre intake did not increase to the recommended amounts (MacIntyre *et al.*, 2002).

The study on the prevalence of coronary risk factors (BRISK) in the 1990s in the population of the Cape Peninsula, South Africa, found trends similar in the diet patterns and nutrient intakes of this urban population (Bourne *et al.*, 1993) to those in the urban group of the THUSA study. Carbohydrate intakes were reduced, fat intakes increased and fibre intakes also decreased in the BRISK study population, and their protein intakes also changed to a higher intake of animal protein (Bourne *et al.*, 1993). Evidence from the Birth to Twenty cohort also indicates that dietary habits and eating practices of adolescents living in urban areas include high-fat, high energy-dense snack foods (Feeley *et al.*, 2012). In the comparison of macronutrient intakes between urban and rural South Africans by Vorster *et al.* (2011) from 1975 to 1996, 1998 and 2005, there is also a noticeable increase in the consumption of dietary fat and a decrease in carbohydrate consumption (Vorster *et al.*, 2011).

Form the previously mentioned data it is clear that the nutrition transition is leaving its mark on the South African population as the data are in accordance with the changes in diet and lifestyle as described by Popkin (Popkin, 1998). These changes (dietary and lifestyle) are associated with disturbed lipoprotein profiles as well as other risk factors (obesity) that have the capacity to lead to the development of NR-NCDs such as CVD. CVD has become a global burden and the South African population has not been spared this non-communicable disease.

2.3 EMERGENCE OF CVD IN THE BLACK SOUTH AFRICAN POPULATION

Cardiovascular disease (CVD) is one of the leading causes of death in the world. Each year more people die of CVD than from any other cause. In 2008, an estimated 17.3 million people died of CVD, representing 30% of all global deaths. Of these deaths an estimated 7.8 million were due to coronary artery disease (CAD) and 8.8 million to stroke (WHO, 2013). As a group of disorders of the heart and blood vessels, cardiovascular diseases include: coronary artery disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease, congenital heart disease, deep vein thrombosis and pulmonary embolism (WHO, 2013). The Heart of Soweto Study gives an indication of the types of heart diseases that are prevalent amongst the black population in Soweto. In a group of newly diagnosed patients at the Chris Hani Baragwanath Hospital, the primary diagnoses were 19% for hypertension, 44% for heart failure, 17% for valvular heart disease, 10% for coronary artery disease and 9% for other diagnoses (Sliwa *et al.*, 2008). Further analysis of the Heart of Soweto data by Stewart *et al.* (2011) indicated that hypertensive heart failure and CAD are the two “newer” forms of heart disease that are starting to replace the prevalence of the historical forms of heart disease such as rheumatic heart disease, dilated cardiomyopathies, pulmonary heart disease, arrhythmias and infectious forms of heart disease (Stewart *et al.*, 2011).

The historically low prevalence and near absence of CAD in the black African population came to the attention of researchers as early as the 1960s. Since these first observations, this phenomenon has been followed during the past decades and different African countries have reported the low incidence of CAD in patients (Walker & Sareli, 1997).

However, over the last few years the emergence of CAD amongst the black African population has increased and several studies in Africa have documented changes in CAD patterns and their risk factors, especially in urban areas (Akinboboye *et al.*, 2003; Hakim *et al.*, 1995; Mamo & Oli, 2001; Steyn *et al.*, 1991). Data from the Heart of Soweto Study indicates an increased number of cases of CVD of which a small but significant number (10%) are coronary artery disease (CAD) cases (Sliwa *et al.*, 2008). Forty-two percent of women and 29% of men participating in this study had LDL-c levels of ≥ 3.0 mmol/L, suggesting an increase in the risk factors for developing CAD (Tibazarwa *et al.*, 2009). The THUSA study showed a difference between risk factors (such as total cholesterol levels) for developing CAD when comparing Africans from rural and urban environments. The urban professionals had higher intakes of energy and fat, as represented by their increased total cholesterol levels and BMI (Vorster, 2002).

The theory that the black African population is resistant to developing acute myocardial infarction in the presence of rising risk factors (Walker & Sareli, 1997) has been refuted by the data from the INTERHEART Africa study (Steyn *et al.*, 2005). The increase in risk factors associated with acute myocardial infarction (AMI) in this study can be explained by the fact that subjects are exposed to an unhealthy lifestyle during their entire lifetime, during which CVD risk factors have sufficient time to impact on the development of the disease, whereas, in previous reports, risk factors in the black African population which had not been exposed to these factors for an entire lifespan were just beginning to emerge and increase. This is supported by the fact that African-Americans had lower occurrences of CAD from 1960 to 1980, whereas their mortality rates have now reached those of the American Caucasian population, which is explained by the extended exposure to risk factors (Keil *et al.*, 1993).

The black South African group from the INTERHEART study demonstrate a spectrum of CVD risk factors unique to a population in the early stages of the epidemiological transition. This trend is further illustrated by higher risk for AMI in the more affluent part of society (Steyn *et al.*, 2005). This trend was also found at the beginning of the CVD epidemic in European societies in the early 20th century (Marmot *et al.*, 1978). The THUSA study also found that individuals in the “upper class”, representing professional people, had an increase in risk factors associated with the development of NR-NCD (Vorster, 2002). At the beginning of an epidemiological transition the more affluent individuals of a population are at greater risk of developing CVD. In most countries, CVD is also in transition and may vary within a country by region or by socioeconomic status, potentially moving in both directions (Yusuf *et al.*, 2001).

As previously mentioned, the THUSA study reflects data from 1996 and 1998 collected in the North West Province, from which it was also clear that the difference in fat intake was significant; the rural participants' intake was 23.3% and that of the urban participants was 31.2% (MacIntyre *et al.*, 2002). Total cholesterol levels of the participants in the THUSA project differed significantly between the urban and rural strata (3.91mmol/L and 4.79mmol respectively). The urban group also had higher LDL-c levels (3.10mmol/L) compared with the rural group (2.33mmol/L) (Oosthuizen *et al.*, 2002). The higher LDL-c levels as a result of following a more atherogenic diet could lead to an increased prevalence of CVD in this population.

Data from Dolman *et al.* (2011) showed that CAD patients have higher mean LDL-c levels than the controls, which would indicate that hyperlipidaemia is a risk factor for developing CAD in the black population. A comparative study by Peer *et al.* (2013) demonstrates that LDL-c levels have

increased from 2.3 to 2.9mmol/L in the black population of Cape Town, South Africa, over the last 18 to 19 years. Even though the mean levels of LDL-c were lower than 3mmol/L, raised LDL-c levels were present in the population, of 37.8% in the men and 47.0% in the women. As the protective mechanism of a favourable lipid profile is diminishing in this population, the raised LDL-c levels put them at greater risk of developing atherosclerosis, an unfavourable outcome of which is ischaemic heart disease (IHD). An association between elevated LDL-c and increased fat intake ($\geq 30\%$ of total energy intake) was also detected (Peer *et al.*, 2013).

The role of increased fat intake and its effects on blood lipid values is clear from the above evidence. It is also clear that these changes can lead to increases in the prevalence of NR-NCDs and risk factors. The effect of genetics on LDL-c regulation in the nutrition transition has not been investigated extensively, especially in the black South African population. Increased fat intake associated with urbanisation may also have an effect on genes when a variant occurs within the lipid metabolism pathway which can lead to a greater risk of developing CVD. This diet-gene interaction is complex when it comes to chronic diseases (such as CVD) and it is not clearly understood. However, the relationship between hypercholesterolaemia and the development of CVD, especially in relation to atherosclerosis, has been unravelled and research is clear about the causal effect of hypercholesterolaemia on the development of atherosclerosis.

As previously mentioned, hypercholesterolaemia is one of the major risk factors for developing CAD in all populations. The role of LDL-c in the pathogenesis of these diseases is of great importance and its role in CAD will be discussed in the following paragraphs.

2.4 THE ROLE OF LDL IN CAD

CVD, but especially CAD, results from atherosclerosis of the coronary arteries leading to decreased coronary circulation, and in the late stages of the disease an artery or numerous arteries may be completely blocked. Atherosclerosis results from the development of lipid-filled lesions in the arterial wall causing stiffening, thickening and narrowing of the arteries, which reduces blood flow through the affected vessels and increases the risk of thrombosis (Nordlie *et al.*, 2005). Widespread, focal lesions are present and restricted mostly to large elastic and muscular arteries such as the aorta, the epicardial coronary, femoral and carotid arteries whereas small arteries such as the intracerebral and intramyocardial branches of the coronary arteries are not affected (Woolf, 1999). LDL particles, in particular, contribute to the development and progression of

atherosclerosis, as this is the primary type of lipid taken up in atherosclerotic lesions as the disease progresses (Nordlie *et al.*, 2005). Genetic, pathological and epidemiological studies have shown clearly that plasma levels of LDL-c are directly related to the incidence of coronary events and cardiovascular death. Elevated concentrations of LDL-c, particularly, are associated with an increased risk of developing atherosclerotic coronary heart disease (CHD) (Babiak & Rudel, 1987).

2.4.1 Atherosclerosis is a complex and multifactorial disease

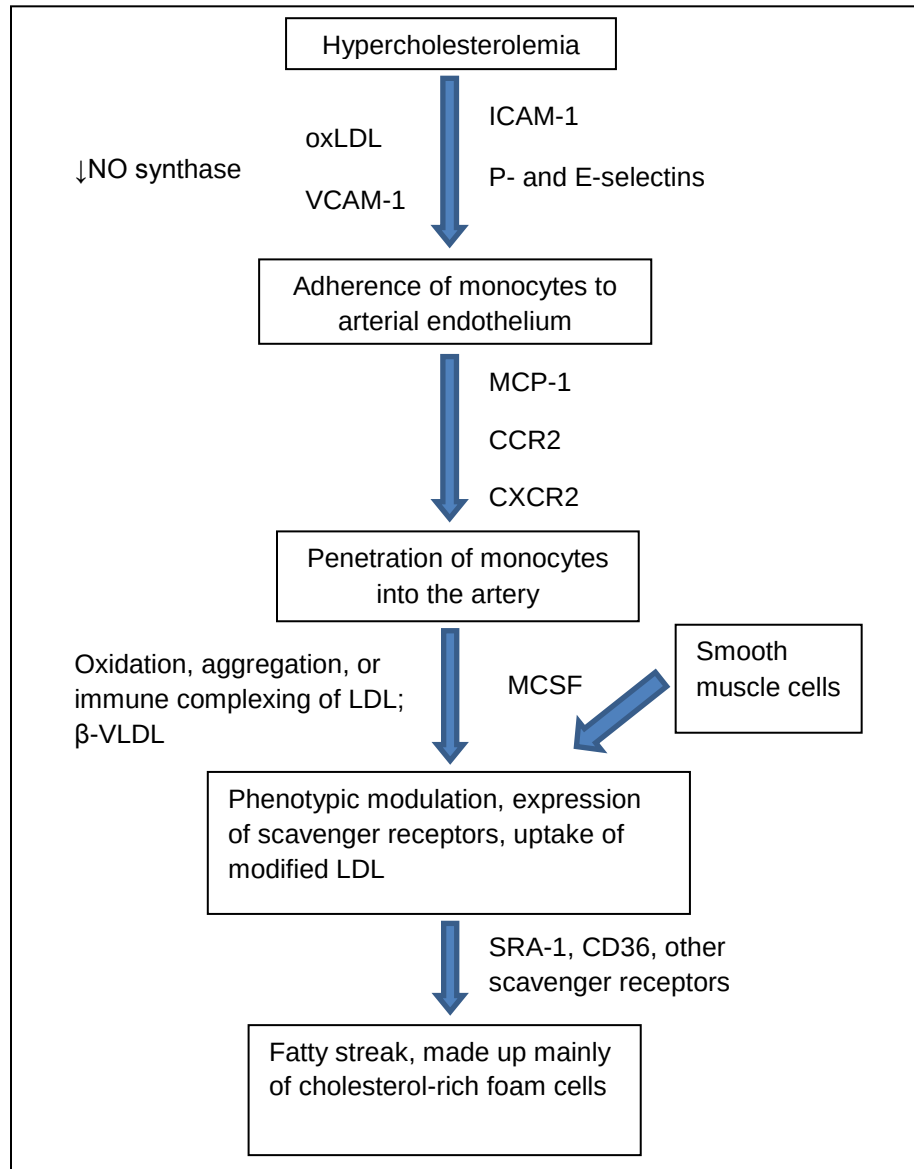
Atherosclerosis is a multifactorial disease that consists of many facets and in combination these facets add to the development of the disease. The focus of the next section is mainly on the roles which hyperlipidaemia and LDL-c play in the development of the disease.

2.4.1.1 LDL and the vascular endothelium

Hypercholesterolaemia triggers the events that lead to the development of the fatty streak, which is the generation of the first visible lesion of atherosclerosis (Steinberg, 2002). These events are summarised in Figure 2.1.

Hypercholesterolaemia results in an increased movement of LDL through the vascular endothelium, especially in regions that are predisposed towards atherosclerotic lesions (such as the aorta). The LDL particles accumulate in the subendothelial space and interact with proteins and proteoglycans that stimulate modification such as aggregation, glycosylation, enzymatic proteolysis and oxidation. This increases their atherogenicity and retention in the vascular intima (Gleissner *et al.*, 2007); (Llorente-Cortés & Badimon, 2005). LDL particle size also has the ability to affect atherosclerosis, the smaller, denser LDL having a greater tendency to be taken up by the vascular endothelium than the larger LDL subfractions (Bjornheden *et al.*, 1996). The small, dense LDL is also more susceptible to oxidative stress, as *in vitro* studies show (de Graaf *et al.*, 1993).

Hypercholesterolaemia induces an increase in the expression of vascular cell-adhesion molecule-1 (VCAM-1) on the endothelial surface lining the major arteries (Steinberg, 2002). Intracellular adhesion molecule 1 (ICAM-1) and the P- and E-selectins are also expressed at higher levels in inflammatory conditions. These adhesion molecules are key adhesion molecules for monocytes and T-cells (Hwang *et al.*, 1997). Hypercholesterolaemia also causes an increase in the expression of monocyte chemoattractant protein-1 (MCP-1), a key chemotactic factor in the artery wall, and increases the expression of its receptor on monocytes (Steinberg, 2002).

Figure 2.1 The sequence of events generating the fatty streak lesion.

oxLDL, oxidized low-density lipoprotein, ICAM-1, intracellular adhesion molecule 1, VCAM-1, vascular cell-adhesion molecule 1, MCP-1, monocyte chemo-attractant protein 1, CCR2, CC chemokine receptor 2, CXCR2, CX chemokine receptor 2, VLDL, very low-density lipoproteins, LDL, low-density lipoprotein, MCSF, monocyte colony-stimulating factor, SRA-1, scavenger receptor class A type 1, CD36, thrombospondin receptor (Steinberg, 2002).

As a multifactorial organ the vascular endothelium can change its functional status to contribute to the homeostasis of the vascular wall (Michiels, 2003). Under normal physiological conditions the endothelium has antithrombotic properties and various substances are exchanged between the blood and tissues (Khazaei *et al.*, 2008). The vascular endothelium produces and releases nitric oxide (NO), which contributes to the antiatherogenic and antithrombotic properties of the endothelium (Khazaei *et al.*, 2008). NO inhibits the aggregation of platelets, has strong vasodilatory activity, has an important anti-inflammatory function and blocks the expression of proinflammatory molecules such as necrosis factor (NF) κ B (Napoli *et al.*, 2006) and adhesion molecules (ICAM-1, VCAM-1) (Carreau *et al.*, 2011), as well as leukocyte adhesion and infiltration (Kubes *et al.*, 1991). Tight junctions and gap junctions which are essential cell–cell connections are important to ensure the regulation of endothelial permeability. The presence and functionality of connections regulate the formation of gap junctions. The expression of connections is altered during the formation of atherosclerotic lesions (Chadjichristos & Kwak, 2007). Gap junctions, in addition to favouring intercellular signalling processes, regulate NO-dependent vasodilation. The bioavailability of endothelial NO decreases in the presence of atherogenic concentrations of LDL-c. The presence of native LDL (Vidal *et al.*, 1998) or modified LDL particles (Liao *et al.*, 1995) as well as the formation of superoxide anions (O_2^-) reduces the concentration and/or activation of NO synthase, the enzyme that produces NO (Pritchard *et al.*, 1995). Elevated plasma LDL-c concentrations increase the permeability of the endothelium and have been associated with p21-activated kinase (PAK) through a mechanism mediated by protein kinase G and Ser/Thr kinase Akt (Orr *et al.*, 2007).

2.4.1.2 LDL and the extracellular matrix

The composition and structure of the extracellular matrix (ECM) influence the vascular remodelling that takes place during the development and complication of atherosclerotic plaques. The main components of the ECM are produced mostly by smooth muscle cells (SMCs), such as proteoglycans, collagen, and elastin, as well as a large number of proteins responsible for the equilibrium between synthesis (lysyl oxidase) and degradation (metalloproteinases, plasminogen activators) of the ECM during the atherogenic process. LDL particles are modified when they interact with ECM components such as oxidants (e.g. 15-lipoxygenase and myeloperoxidase), and/or proteolytic enzymes (e.g. chymase and tryptase), lipolytic enzymes (sphingomyelinase) and hydrolytic enzymes (phospholipase A_2). Therefore, different types of modified LDL particles are generated by the different components (Öörni *et al.*, 2000). LDL particles interact with

proteoglycans and are trapped in the arterial intima. Versican-like proteoglycans are abundant in the ECM and have an affinity with LDL, and these particles are most likely the main structure of the intima interacting with particles that enter this region. These proteoglycans are considered to be important atherogenic elements because they strongly interact with, retain and aggregate cholesterol-rich lipoproteins. Small dense LDL particles have the greatest capacity to interact with proteoglycans, especially with the chondroitin sulphate proteoglycans. Insoluble complexes form between these molecules and LDL particles are trapped in the arterial intima. Collagen is essential for maintaining the integrity and elasticity of the vascular wall but its glycosylated forms are an important factor in atherogenesis as they favour LDL retention (Barnes and Farndale, 1999). Modified forms of LDL particles are internalised in cells by different types of receptors such as low-density lipoprotein receptor-related protein-1 (LRP-1). The LDLR, LRP-1 and scavenger receptors have been detected in different cell types such as monocytes, SMCs and platelets with key functions in the development of atherosclerotic lesions. LDL aggregates form because of the structural changes in these lipoproteins after the interaction of versican with LDL (Camejo *et al.*, 1998).

Native LDL particles are internalised by the LDL-receptor whereas the modified LDL particles are internalised in cells by different types of receptors such as low-density lipoprotein receptor related protein-1 (LRP-1), scavenger receptors such as SRAI, SR-AII, CD36, LOX-1 or CXCL16 for oxidised LDL. Various cell types such as monocytes, SMCs and platelets which also play a pivotal role in the development of atherosclerosis also contain LDLR, LRP-1 and scavenger receptors (Badimón *et al.*, 2009).

2.4.1.3 Plaque development

Infiltration of blood monocytes into the vascular intima plays an important part in the development of atherosclerotic lesions. The modified forms of LDL particles increase the expression of soluble chemotactic compounds (MCP-1, interleukin [IL]) and enhance the expression of adhesion molecules such as VCAM-1, integrins and selectins, which are exposed on the surface of activated endothelial cells and favour leukocyte (monocyte and T-cell) recruitment, adhesion and transmigration. The simultaneous expression of these molecules indicates an intensive activation of different genes through a common transcription factor such as NF- κ B. In addition to the effects of the modified LDL particles on the vascular endothelium, it favours the entry of monocytes into the vascular wall. Migration of monocytes takes place through the junctions between endothelial cells,

especially in places with high concentrations of modified LDL particles. The infiltrated monocytes differentiate into macrophages and express scavenger receptors (such as CD36 and LOX-1) induced by macrophage colony-stimulating factor (M-CSF), which internalises many of the cholesterol molecules and cholesterol esters contained in modified LDL particles. Cholesterol internalisation leads to the formation of foam cells. Foam cells secrete pro-inflammatory cytokines, growth factors, tissue factor, interferon δ , MMP and reactive oxygen species that maintain the chemotactic stimulus for leukocytes to adhere to the vascular endothelium, increase the expression of scavenger receptors, enhance macrophage replication and regulate SMC accumulation in the intima (Libby, 2002; Libby *et al.*, 2010).

A key process in the development of atherosclerotic lesions is the migration of SMC from the vascular media to the vascular intima and this is also modulated by certain components of the LDL-related receptors by means of signalling processes mediated by cytokines and activation of proteinases. SMCs overexpress receptors such as LRP-1 in the presence of LDL aggregates. These receptors internalise LDL particles which cause the transformation to foam cells; they also act as receptors for different ligands and participate in signalling processes. Foam cells are thus derived from SMCs and/or macrophages (Badimón *et al.*, 2009).

Platelet function is also affected by hypercholesterolaemia. Native LDL particles do not bind to the platelet surface but affect platelet function through activation of transduction signals or lipid exchange. The phospholipid composition of the platelet membrane is transformed by the native LDL either by inducing the production or translocation of the membrane phospholipids or favouring the insertion of phospholipids from the circulation (Engelmann *et al.*, 1996). Binding of platelets to oxidised LDL particles induces activation, morphological changes and platelet aggregation and this contributes to the formation of thrombi, especially after plaque rupture (Maschberger *et al.*, 2000). Fibrin is also deposited on the surface of plaques after rupture and as the fibrin layer is covered by endothelium and SMC are taken up, the plaque size expands (Smith *et al.*, 1992). Endothelial dysfunction is able to induce adhesion of platelets to the vascular wall. The endothelium is able to secrete an ample amount of von Willebrand factor (vWf) in reaction to different inflammatory stimuli and so promote platelet recruitment. After vascular damage, the dynamics of platelet deposition and the consequent thrombus formation is regulated locally by (1) the severity of stenosis, (2) the type of lesion, and (3) the composition of the atherosclerotic plaque (Badimón *et al.*, 2009). Fissure or rupture of the atherosclerotic plaque in the coronary arteries and thrombus formation is essential for the development of acute ischaemic syndromes. Plaque rupture maybe the result of the excessive degradation of the extracellular matrix structure caused by inhibited collagen production,

collagen degradation and the superficial erosion of the plaque caused by apoptosis of endothelial cells (Libby, 2013). In asymptomatic patients and those with stable angina, the organisation of the thrombus is important in the progression of atherosclerosis. Rupture of the atherosclerotic plaque exposes components of the vascular matrix to the bloodstream, thereby favouring the interaction between platelets and the vascular wall (Virmani *et al.*, 2005). Platelets, just like red blood cells, can penetrate the atherosclerotic plaque through rupture of angiogenic capillaries. The close interaction between platelets and monocytes favours the differentiation of these cells into macrophages and their transformation into mast cells. Research has shown that activated platelets can release cholesterol, which is internalised and stored as lipid droplets by SMCs and macrophages and can induce the formation of foam cells (Daub *et al.*, 2006).

2.4.1.4 Plaque stability

Deterioration of the plaque from a stable to a vulnerable form and the formation of a thrombus thereafter are involved in causing this life-threatening condition for the patients. The type rather than the size of the plaque increases the risk of plaque rupture (Badimón *et al.*, 2009). The main factors that affect the stability of the plaques are the ratio of ECM to lipid content as well as the types of cells that make up the plaque. Vulnerable plaques which are prone to rupture have a large lipid centre, a thin fibrous cap with little collagen, many inflammatory cells (Libby, 2002) and a few SMCs (Badimón *et al.*, 2009). The normal human coronary artery consists of three layers. The endothelial cells are in contact with the blood in the arterial lumen and rest on a basement membrane. The intimal layer in adult humans normally contains a small amount of smooth muscle cells distributed within the intimal extracellular matrix. The internal elastic lamina forms the barrier between the tunica intima and the underlying tunica media. The media consists of multiple layers of smooth muscle cells which are packed more tightly than in the intima and are embedded within a matrix enriched with elastin and collagen. The lipid-rich core forms in early atherogenesis, together with the recruitment of inflammatory cells. The artery enlarges in an outward, abluminal direction to accommodate the expansion of the intima. As dyslipidaemia and inflammatory conditions persist, the lipid core can grow, the extracellular matrix is degraded by proteinases secreted by the activated leukocytes and the synthesis of new collagen is limited by interferon- γ (Libby, 2002). SMC are strongly inhibited by interferon- γ , which is a product of activated T-cells to make new collagen required to repair and maintain the integrity of the fibrous cap (Amento *et al.*, 1991). Collagen may also be degraded by proteinases that have the capability of catalysing the first steps to break down the thin fibres of collagen.

These enzymes belong to the matrix-metalloproteinase (MMP) family and include MMP-1; MMP-8 and MMP-13, which are produced by macrophages (Shah *et al.*, 1995). These changes can thin the cap and render it vulnerable and susceptible to rupture. When the plaque ruptures, blood coming into contact with the tissue factor in the plaque coagulates. Thrombin is generated by the coagulation cascade and activates platelets, which instigate thrombus formation as they make contact with the intimal compartment. An acute myocardial infarction can result, if the thrombus blocks the vessel permanently. The thrombus may eventually resorb as a result of endogenous or therapeutic thrombolysis. However, thrombin triggers a wound healing response that can stimulate smooth muscle proliferation. Platelet-derived growth factor (PDGF) released from activated platelets stimulates smooth muscle migration. Transforming growth factor- β (TFG- β), also released from activated platelets, stimulates interstitial collagen production. This increased migration, proliferation and extracellular matrix synthesis by SMC thickens the fibrous cap and causes the intima to further expand. The expansion is now inwards, which constricts the lumen. Stenotic lesions produced by the luminal intrusion of the fibrosed plaque may restrict flow, particularly under situations of cardiac demand, leading to ischaemia, commonly provoking symptoms such as angina pectoris. Advanced stenotic plaques, being more fibrous, may prove less susceptible to rupture and renewed thrombosis. Lipid lowering can reduce lipid content and calm the initial inflammation response, yielding a more stable plaque with a thick fibrous cap and a preserved lumen (Libby, 2002).

Plaques with a thin fibrous layer covering the lipid core and high tissue factor (TF) content are those with the greatest potential to rupture and cause thrombosis. LDL particles are implicated in TF cell expression and research has shown that the interaction between LRP-1 and LDL aggregates is one of the mechanisms that induce TF expression (Badimón *et al.*, 2009). Macrophages are another source of TF where the expression is regulated by PPAR α (Camino-López *et al.*, 2007).

2.4.1.5 Platelet aggregation and atherothrombosis

After being activated, platelets undergo a conformational change and release intraplatelet Ca^{2+} . These events are responsible for the subsequent platelet activation process, which includes platelet degranulation, eicosanoid synthesis, expression of adhesion proteins, and exposure of a procoagulant platelet surface. A change in the morphology of the platelet membrane occurs as it changes from a smooth disc to a prickly sphere. This alteration exposes phospholipoprotein

components rich in phosphatidylserine on the external side of the membrane (Sims & Wiedmer, 2001). The exposure of the phospholipoprotein components facilitates the binding and activation of the coagulation factors and subsequent thrombin formation. Thrombin amplifies the initial activator effect of collagen and further stimulates the activation, release and recruitment of platelets. These events lead to the formation of eicosanoids and the appearance of fibrin filaments, primarily in the external portion of the platelet thrombus, but also at the interstitial sites between adhered platelets. The binding of thrombin to fibrin or the artery wall masks thrombin receptors for antithrombin III, heparin and cofactor II, so that the mural thrombus acts as a potent thrombogenic stimulus – dependent on thrombin and relatively resistant to heparin – that favours thrombus growth. In the final stage of thrombus formation, fibrinogen is converted to fibrin by thrombin, and this leads to stabilisation of the platelet aggregates. In addition, activation of the signalling pathway occurs from outside in, thereby causing amplification of the initial signal and greater platelet activation and recruitment. This forms a mass that expands and continues to recruit more platelets as these reach the prothrombotic microenvironment (Badimón *et al.*, 2009).

2.4.1.6 Activation of the coagulation cascade

The coagulation cascade is activated after vascular rupture. TF expressed in foam cells is present in the lipid centre of atherosclerotic lesions and is a potent inducer of the coagulation cascade (Moons *et al.*, 2002). Activated platelets, endothelial dysfunction and TF have important thrombogenic potential, promoting the coagulation cascade and the formation of fibrin. The dysfunctional endothelium changes to a procoagulant state simultaneously as surface cofactors are exposed on platelets and catalyse the formation of thrombin from prothrombin.

Thrombin is the final product of the activation of the coagulation cascade and is a strong platelet agonist and an important component in the pathogenesis of the atherothrombotic process. Four protease-activated receptors (PAR) are present on the various atherosclerotic cell types through which thrombin signalling takes place (Peterson *et al.*, 2003). Receptors present on SMC are PAR1 and PAR2, which favour their growth, whereas PAR1, PAR2 and PAR4 are present in macrophages and induce inflammatory pathways. The receptors present on platelets are PAR1 and PAR4 and their signalling ends in the activation of the fibrinogen receptor (Badimón *et al.*, 2009). The progression of atherosclerotic lesions takes place through thrombin and its binding to PAR (Vu *et al.*, 1991). The fibrinogen receptor is activated by endothelial receptors, collagen GPIIb/IX/V and GPVI, receptors coupled to G proteins (PAR1, PAR4) or ADP receptors, which

enhance GPIIb/IIIa-dependent platelet aggregation as well as ensuing thrombus formation (Badimón *et al.*, 2009).

In the above section the role of elevated LDL-c levels in the development of atherosclerosis has been described. As LDL-c levels are important in the progression of atherosclerosis, the determinants of LDL-c levels will be discussed in the following paragraphs.

2.5 DETERMINANTS OF LDL-C LEVELS

Lipids are hydrophobic molecules which include triglycerides, cholesterol and its esters, and these molecules do not circulate freely in the blood. Instead, they are transported in lipoproteins, spherical complexes of lipids and apoproteins that stabilise the lipid emulsion (Burnett & Hooper, 2008). There is a lipoprotein transport system in place to deliver the lipid particles to the appropriate cells by means of the blood circulation in a regulated and direct mode. The lipids from the diet and those endogenously produced are delivered to the correct tissue sites in the precise amounts by means of control mechanisms. Since most cells have a limited capacity to store cholesterol and triglycerides, steps need to be taken to limit the uptake when the stores have reached their capacity. Intake of an excess of requirements because of the body's inability to inhibit absorption leads to the accumulation of lipids in the circulation (Packard & Shepherd, 1999). Accumulation of LDL-c, especially, as discussed under the section on atherosclerosis, is detrimental to an individual's health. Lipoprotein transport regulation is most effective at low plasma lipid levels. The LDL receptor, which mediates the uptake of LDL-c, is able to meet cellular cholesterol requirements when LDL-c is present at 1mmol/L. In most developed countries the mean LDL-c is approximately 4mmol/L, and at these levels the receptor is down-regulated to prevent intracellular sterol accumulation (Packard & Shepherd, 1999).

LDL is a cholesterol-rich lipoprotein and the metabolic product of VLDL, a triglyceride-rich lipoprotein secreted by the liver. ApoB-100 is the structural backbone of LDL-c and it is essential for the assembly and secretion of triglyceride-rich lipoproteins (Burnett & Hooper, 2008). VLDL is produced in a two-step model. Step one is represented by the independent and simultaneous formation of two VLDL precursors: an apoB-containing precursor and a VLDL sized lipid droplet that lacks apoB. In the second step, these two precursors fuse to form a mature VLDL. The apoB-containing precursor is formed during the translation and associated translocation of the protein to the lumen of the endoplasmic reticulum (ER) (Olofsson *et al.*, 1999). At the luminal side of the ER

in the inner membrane, phospholipids and free cholesterol are present, providing a site for apoB incorporation. Small amounts of cholesterol esters and triglycerides are added at this point and a primary, dense lipoprotein is formed (Boren *et al.*, 1993). The process is dependent on the microsomal triglyceride transfer protein (MTP). MTP stabilises the newly produced apoB within the cell ER and aids the transfer of lipids from the ER to apoB (Hussain *et al.*, 2003).

As VLDL circulates, the core triglycerides of the VLDL are hydrolysed by lipoprotein lipase (LPL) and intermediate-density lipoproteins (IDL) are formed. Hepatic lipase (HL) further metabolises IDL to LDL, which is mostly cleared by the LDLR into the liver or peripheral tissue (Brown & Goldstein, 1986). HDL transport lipids from the peripheral tissues to the liver, a process known also as reverse cholesterol transport (Lewis & Rader, 2005).

It has been established that both the environment and genetics affect LDL-c levels. A diet high in saturated fats, for example, is associated with elevated total cholesterol and LDL-c levels (Frayn & Stanner, 2005). Population variation in LDL-c levels differs over a three-fold range. Genetic variance is responsible for up to 50% of this inter-individual variation in plasma (Burnett & Hooper, 2008). The variation is also polygenic, with not one locus only but numerous loci responsible for the differences (Burnett & Hooper, 2008). Table 2.1 contains a summary of the major genes that affect LDL-c plasma concentrations.

The prevalence, associations and effects of SNPs on LDL-c levels have been studied in different populations (Kathiresan *et al.*, 2007; Knoblauch *et al.*, 2002) but studies in the black African population have been few and far between. Where studies identified genes with variants associated with disease there were no follow-up studies to verify the findings in larger populations. Therefore, genetic variants associated with LDL-c levels must still be determined to enable us to compare these variants with what has been discovered in the European population. The detection of the genetic variations present in African individuals associated with LDL-c levels can give us insight into the role they might play in affecting LDL-c levels during urbanisation of the black African population.

Table 2.1 Summary of major genes affecting plasma LDL-c concentrations (adapted from Burnett & Hooper, 2008)

Gene, gene locus	Protein	Lipid disorder	Effect of variant on protein	Effect on LDL-c levels	Reference
APOB, 2p24	Apolipoprotein B	Familial hypo-betalipoproteinaemia	Prematurely truncated apoB protein	Levels are 25 to 33% of normal levels	(Whitfield <i>et al.</i> , 2004)
		Familial ligand-defective apoB-100	Affects the binding affinity for the LDL receptor	Increase levels, less severe than FH levels	(Whitfield <i>et al.</i> , 2004)
MTP, 4q22-q24	Microsomal triglyceride transfer protein	Abetalipoproteinaemia	A truncated protein is formed and the assembly of LDL-c is affected	Levels are 25 to 33% of normal levels	(Whitefield <i>et al.</i> , 2004; Berriot-Varoqueaux <i>et al.</i> , 2000; Narcisi <i>et al.</i> , 1995)
APOE, 19q13.2	Apolipoprotein E	ApoE4 modulates the risk for other disorders	ApoE4 – influence the stability of the protein as well as the folding. Protein has a greater affinity for VLDL and LDL	LDL-C elevated by 8% increments	(Hatters <i>et al.</i> , 2006; Minihane <i>et al.</i> , 2007)
		ApoE2 homozygotes – Type III hyperlipoproteinaemia	Weakens binding affinity of the protein for LDL-receptors on the cell surface	Elevated levels of triglycerides	(Weisgraber <i>et al.</i> , 1982)
ARH, 1p36-p35	Adaptor protein	Autosomal recessive hypercholestrolaemia	Adaptor protein can be absent, impaired interaction with LDL receptor, clusters LDL receptors in clatherin-coated pits	Increase LDL-c levels markedly	(Soutar <i>et al.</i> , 2003)

Table 2.1 continues...

Gene, gene locus	Protein	Lipid disorder	Effect of variant on protein	Effect on LDL-c levels	Reference
NPC1L1, 7p13	Niemann-Pick C 1-like 1 protein	Not defined	Variants could possibly decrease the protein's function	Decrease levels with 10%	(Cohen <i>et al.</i> , 2006)
PCSK9, 1p34.1-p32	Proprotein convertase subtilisin/kexin-type 9	Familial hypercholesterolaemia	Gain-of-function variants affect the affinity of the protein towards the LDL receptor	Increase LDL-c levels markedly	(Abifadel <i>et al.</i> , 2003)
		Hypocholesterolaemia	Loss-of-function variants produce truncated proteins or proteins with less activity	Up to 40% lower levels	(Cohen <i>et al.</i> , 2005)
LDLR, 19p13.2	LDL receptor	Familial hypercholesterolaemia	Defective LDLR production, function or recycling leads to reduce clearance of LDL	200-300% increase (heterozygote) 500-1000% increase (homozygote)	(Hobbs <i>et al.</i> , 1990)

The LDL receptor is the main regulator of LDL-c levels in plasma as this protein mediates the binding, uptake and degradation of LDL-c. This receptor is encoded by the *LDLR* gene and multiple mutations in this gene have been identified as having an effect on the function of the LDL receptor and the development of atherosclerosis. Therefore, the *LDLR* gene was selected to be investigated in the study population to determine the prevalence and association of SNPs with LDL-c. The LDL receptor and *LDLR* gene will be discussed in the following paragraphs.

2.5.1 The low-density lipoprotein (LDL) receptor

The LDL receptor is a cell surface glycoprotein which is the major protein to transport cholesterol across the human cell membrane. The receptor mediates the specific binding, uptake and degradation of plasma lipoproteins containing either apolipoprotein (apo) B or apoE. Remnants of triglyceride-rich lipoprotein metabolism (also known as IDL), which are the precursors of plasma LDL and LDL itself, are recognised by the LDL receptor. It regulates plasma cholesterol by mediating the endocytosis of LDL-c by means of clathrin-coated pits (Brown *et al.*, 1981). The LDL receptor was identified in 1973 as the cause of familial hypercholesterolaemia (FH) in a study comparing the fibroblasts of normal and FH patients. Normal fibroblasts can synthesise the cholesterol they need for growth intracellularly from acetyl CoA or it can be taken up from plasma LDL. Normal cells grown in the presence of serum containing LDL produce low amounts of cholesterol. There is low activity of the rate-limiting enzyme in cholesterol synthesis, 3-hydroxy-3-methylglutaryl CoA reductase (HMG CoA reductase), in these cells. Growing cells in serum deficient of lipoproteins increases the activity of the HMG CoA reductase and more cholesterol is produced. In FH-homozygote fibroblasts, however, the same high levels of HMG CoA reductase are maintained in the presence or absence of lipoproteins because of their inability to take up LDL-c (Hobbs *et al.*, 1990). In 1982 the LDL receptor was purified and its complementary DNA was cloned by Russell *et al.*, (1983) and Yamamoto *et al.*, (1984). The *LDLR* gene was isolated and characterised in 1985 by Sühdof *et al.*

The following paragraphs explain the different steps in the pathway of LDL receptor-mediated uptake and degradation of lipoproteins in cultured cells based on the work of Goldstein, Brown and colleagues. The LDL receptor is produced as a precursor that is processed to its mature form as it is transported to the cell membrane. As the newly synthesised protein moves through the Golgi,

the core N- and O-linked sugar residues which are added during translation on the endoplasmic reticulum are trimmed and modified; this process is very sensitive to changes in the structure of the protein (Tolleshaug *et al.*, 1982). Substantial stability is added to the mature LDL receptor by numerous disulphide bonds (Goldstein *et al.*, 1985). The mature receptor protein is then randomly inserted into the cell membrane, where it is available to bind lipoprotein ligands containing either apoB or apoE. As a single molecule of apoB is present and multiple copies of apoE are present the LDL receptor has a greater affinity for lipoproteins containing apoE. The receptor recognition sites on VLDL and lipoprotein (a) are concealed on their cell surfaces as a result of the conformation of the apoB in these lipoproteins. Therefore, even though both contain apoB they are poor ligands for the LDL receptor (Hui *et al.*, 1984).

After the LDL receptor binds to its ligand, the receptor-ligand complexes cluster in clathrin-coated pits that turn inward to form intracellular endosomes (Goldstein *et al.*, 1979). These endosomes consequently combine with lysosomes to form secondary lysosomes where the low pH causes the ligand to detach from the receptor. The ligand is then degraded but the receptor recycles to the surface where it can perform its function for further rounds of endocytosis (Brown *et al.*, 1983). The LDL receptor must contain sufficient structural information to be distinguished from other receptors where the ligand is not degraded or where the receptor is degraded with the ligand. The receptor must also contain a ligand-binding site and more structural information for the protein to be glycosylated in such a way that it is transported efficiently to the cell surface. Once inserted in the cell membrane it must be directed to cluster in clathrin-coated pits for internalisation (Soutar, 1999).

The protein content of the ligand is degraded within the lysosomal compartment to its principal amino acids and the cholesterol ester (main constituent of LDL) is hydrolysed to form free cholesterol and fatty acids. The free cholesterol formed through this process is used either for the synthesis of a new cell membrane or in the appropriate cells, as a substrate for synthesis of bile acids or steroid hormones. Regulatory measures exist to prevent the accumulation of excess cholesterol in the cells. LDL receptor activity is firstly reduced by decreasing the rate of transcription of the *LDLR* gene (Russell *et al.*, 1983) and by decreasing cholesterol synthesis by down-regulation of the HMG CoA reductase (Chin *et al.*, 1982). Secondly, any free and surplus cholesterol is sequestered as cholesterol esters and stored safely within the cell, by the enzyme acylCoA cholesterol acyltransferase (ACAT) (Goldstein *et al.*, 1974).

2.5.2 The *LDLR* gene

The human *LDLR* gene is 45kb long and consists of 18 exons and 17 introns and it encodes a protein of 860 amino acids (Südhof *et al.*, 1985a). The gene is located in the distal short arm of chromosome 19p13.2 (Lindgren *et al.*, 1985).

Figure 2.6 shows how the exons of the gene correlate with the functional domains of the mature protein. The *LDLR* promoter contains most if not all of the cis-acting DNA sequences responsible for the ubiquitous and regulated expression of the gene in animal cells (Goldstein & Brown, 1990, Südhof *et al.*, 1987). Within 200bp of the initiator methionine codon are three imperfect direct repeats of 16bp each, two A/T-rich sequences and a cluster of mRNA initiation sites, all of which function in transcription. Two of the direct repeat sequences (nr 1 and 3) interact with the general Sp1 (a ubiquitous transcription factor) to promote transcription. These sequences alone are not adequate for high level expression. The contribution of the third direct repeat is required, which is a conditional-positive sterol regulatory element, SRE-1, that is 10 nucleotides in length. In the absence of sterols this element synergises with the two Sp1 sites (specificity protein 1 is a transcription factor) to promote transcription. In the presence of sterols the SRE-1 loses its activity. Cells suppress receptor mRNA through this regulatory mechanism and develop a corresponding decrease in cell surface LDL receptors when cultured in the presence of sterols (Hobbs *et al.*, 1990).

Exon 1 encodes the signal sequence which consists of a short 5' untranslated region and 21 hydrophobic amino acids. This sequence is cleaved from the protein during translocation into the endoplasmic reticulum (ER) and this leaves a mature protein of 839 amino acids (Hobbs *et al.*, 1990).

Exons 2 to 6 encode the ligand-binding domain which is made up of seven repeats of 40 amino acids each. Each of the repeats is encoded by a single exon except for the third, fourth and fifth repeats which are joined in one exon (Südhof *et al.*, 1985a). Each of the seven repeats contains six cysteine residues that form three intra-repeat disulfide bonds. Also important for ligand binding is the negatively charged triplet, Ser-Asp-Glu (acidic residues) that forms part of the COOH-terminal end of each repeat (Hobbs *et al.*, 1990).

The substitutions of a single base pair in this domain can have a noticeable effect on the ability of the protein to bind ligands and could affect the intracellular transport of the newly synthesised protein (Esser & Russell, 1988). Still, not all the repeats are equally important for high-affinity

binding of different lipoproteins that contain the two ligands that are recognised by the LDL receptor, as shown by analysing the mutations at specific sites in detail. Repeats four and five are of key importance for binding apoE-containing lipoproteins, but repeats two, three, six and part of the seventh domain, another type of cysteine-rich protein motif known as the a growth-factor repeat, must also be present for normal binding of LDL. Mutations in exon 4 of the gene that affect single amino acid residues in repeats four and five are particularly common in FH patients (Hobbs *et al.*, 1986).

Exons 7 to 14 of the gene encode a 400-amino acid sequence that is 33% identical with a portion of the human epidermal growth factor (EGF) precursor gene, namely the EGF precursor-like domain (Südhof *et al.*, 1985a, Südhof *et al.*, 1985b). This region includes three growth factor repeats, which are 40 amino acid cysteine-rich sequences that differ from the cysteine-rich sequences in the ligand-binding domain. The first two growth factor repeats, designated A (encoded by exon 7) and B (encoded by exon 8), are nearby and separated from the third (C, encoded by exon 14) by a 200-amino acid sequence. This sequence is encoded by exon 9 to 13 and contains five copies of a conserved motif (YWTD) repeated once each 40-60 amino acids. The EGF precursor homology domain is needed for the acid-dependent dissociation of lipoproteins from the receptor in the endosome during receptor recycling. This domain also functions to position the ligand-binding domain in order to bind LDL-c on the cell surface (Davis *et al.*, 1986b). Single amino acid changes in the EGF precursor-like domain are mostly class 2 mutations and this indicates that the modification of the sugar residues as the protein moves through the Golgi is particularly sensitive to minor changes in the secondary structure of this region (Soutar, 1999).

Exon 15 encodes 58 amino acids that are rich in serine and threonine residues, of which many serve as binding sites for O-linked carbohydrate chains (Davis *et al.*, 1986a). This domain is one of the least well-conserved parts of the protein sequence, both between species and members of the *LDLR* gene family. The function of this domain has not been elucidated and in cultured cells the protein produced by cDNA lacking this sequence appears to have a normal function. Even so, FH patients (Kajinami *et al.*, 1988; Mozas *et al.*, 2004) with a mutation in exon 15 have been identified, suggesting that there must be a damaging effect on the receptor *in vivo*. The effect is not, however, described by any of the authors but Kajinami *et al.* (1988), which suggests that it could be that the hepatic function of the mutant protein may be different from that of the function in fibroblasts in which the original research (Brown & Goldstein, 1986) was done.

Exon 16 and the 5' end of exon 17 encode 22 hydrophobic amino acids that comprise the membrane-spanning domain (Hobbs *et al.*, 1990). This domain anchors the receptor at the cell surface. Receptor mediated endocytosis cannot be catalysed by a truncated protein that lacks the whole or part of the transmembrane domain (Soutar, 1999). Single base-pair mutations found in this domain have not been characterised (Hobbs *et al.*, 1992; Bunn *et al.*, 2002).

The rest of exon 17 and the 5' end of exon 18 encode the 50 amino acids that make up the cytoplasmic domain, which is the only intracellular part of the protein when it is present on the cell surface. This domain is important because it ensures that the receptor is localised in the coated pits on the cell surface (Chen *et al.*, 1990). Mutations in this domain usually are of the internalisation-defective phenotype (class 4) and result mostly from truncations. Four amino acid residues have been identified, NPVY, which form an internalisation signal that contains the information needed to facilitate interaction between the receptor and clathrin in the coated pits. Experiments in transgenic mice have revealed that sequences between residues 812 and 828 are necessary to direct the receptor protein to the basolateral surface of hepatocytes (Soutar, 1999).

The 2.6-kb 3' untranslated region of the mRNA is specified by the remainder of exon 18. The LDL receptor protein is highly conserved, especially the cytoplasmic domain, followed by the EGF precursor domain, followed by the ligand-binding and the trans membrane domain (Hobbs *et al.*, 1990). The introns of the human *LDLR* gene and the untranslated part of the final exon contain numerous *alu* repeats. These have been implicated in homologous recombination events that have resulted in some of the major gene deletions and rearrangements that have been observed at this locus in FH patients (Soutar, 1999).

2.5.3 Classification of Mutations

A classification system based on the nature of the biochemical defect observed in cultured cells has been devised that includes five different categories of mutation. These five classes are thus based on the phenotypic effects on the protein (Soutar, 1999). Because class 1 mutations fail to produce immune-detectable protein, this class of mutation disrupts the synthesis of the protein in the ER (null allele). Class 2 mutations are caused by transport-defective alleles which produce proteins that are either partially or completely blocked in the transport between the ER and the Golgi (Hobbs *et al.*, 1990). Class 2A mutations completely block the transport of the protein, while class 2B mutations produce proteins that are transported at a measurable, but markedly reduced, amount

(Goldstein *et al.*, 1985). The bind-defective alleles are grouped into the class 3 mutations and produce proteins that are synthesised and transported to the cell surface but do not bind LDL-c normally. Class 4 mutations encode receptors that move to the cell surface and bind LDL-c normally, but are unable to cluster in clathrin-coated pits and thus do not internalise LDL-c (internalisation-defective alleles). These mutations affect the cytoplasmic domain alone (4A) or also the membrane-spanning region (4B) (Goldstein *et al.*, 1985). Class 5 mutations are caused by recycling-defective alleles and produce proteins that bind and internalise LDL-c in coated pits but do not discharge the LDL-c in the endosome and the receptor is not recycled to the cell surface. Other mechanisms that are involved with the mutations are the truncation of the receptor before the membrane-spanning domain (Lehrman *et al.*, 1985) which is shed into the cell medium and incorrect localisation (Anderson *et al.*, 1977)

2.5.4 Frequency of *LDLR* variants in the black South African population

In Table 2.2 there is a summary of the variants that have been found in the black South African population. The frequency of these variants in the broad population is not known as most of the studies were done in patients with familial hypercholesterolaemia (FH) or patients who presented the phenotype for FH. More than 1125 variants have been reported on the website of the British Heart Foundation (www.ucl.uk/ldlr/LOVDv.1.1.0/) but only the variants summarised in the table have been identified in the black South African population.

Eleven variants within coding and regulatory regions have been identified in the black South African population and these influence the function and transcription of the receptor. These variants have a direct association with LDL-c levels. The transcription of the LDL receptor is regulated by sterols in the circulation. Transcription is up-regulated when plasma is depleted of sterol through the interaction of the sterol response element (SRE-1) with other regulatory elements, which promotes the synthesis of the LDL receptor mRNA. The regulation is intercellular and it modifies the expression of genes. The majority of circulating LDL moves into the liver where the intracellular regulation takes place. Transcription is inhibited when the sterol regulatory element-binding protein (SREBP) interacts with SRE-1, rendering it inactive in the presence of sterols (Soutar, 1999). In 2003 another protein was identified that reduces the number of LDL receptors on the hepatic cell surface, namely proprotein convertase subtilisin kexin 9 (PCSK9). Variants within the *PCSK9* gene affect the number of LDL receptors on the cell surface which also affects circulating LDL-c levels

(Abifadel *et al.*, 2003). The role of *PCSK9* and its variants in the degradation of LDL receptors will be described in the following section.

Table 2.2 *LDLR* variants identified in black South Africans (Summarised from www.ucl.ac.uk/ldlr/LOVDv.1.1.0/)

	<i>LDLR</i> variant	rs number	Region	Changes in DNA sequence	LDL receptor function	Effect on LDL-c	References
1	c.-268G>T	rs17249134	Promoter	G→T at position -268	7% decrease in LDL receptor transcription	Increase	(Scholtz <i>et al.</i> , 1999) (Whittall <i>et al.</i> , 2002) (Hoogendijk <i>et al.</i> , 2003)
2	c.-217C>T	rs17249141	Promoter, FP1 cis-acting regulatory element	C→T at position -217	50–160% increase in transcription activity, increase LDL receptors	Decrease	(Scholtz <i>et al.</i> , 1999) (mixed ancestry)
3	c.-185_-183delCTT (FH-Pedi 2)	None	Promoter, SRE repeat 1	3 bp deleted (CTT)	5–15% LDL receptor activity when heterozygous compound with FH Pedi	Increase	(Peeters <i>et al.</i> , 1998)
	c. 137_142 deletion FH-Cape Town-1	None	Exon 2 (Repeat 1)	6 bp deletion GCGATG	15–30% activity in homozygotes	Increase	(Leitersdorf <i>et al.</i> , 1988)
4	c.172 del G FH-Pedi 1	None	Exon 2	1 bp (G) deletion at 172	5–15% LDL receptor activity when heterozygous compound with FH Pedi	Increase	(Hobbs <i>et al.</i> , 1992)
5	c.313 + 1G→A	None	Intron 3	313 + 1G→A	Disrupts normal splicing	Increase	(Leren <i>et al.</i> , 1994) (Lombardi <i>et al.</i> , 1995) (Jensen <i>et al.</i> , 1996) (Deiana <i>et al.</i> , 2000)

	LDLR variant	rs number	Region	Changes in DNA sequence	LDL receptor function	Effect on LDL-c	References
6	c.514G>C D172H FH Swazi	None	Exon 4 (repeat 4)	G→C at position 514	Effect on function has not been defined	Increase	(Thiart <i>et al.</i> , 2000)
7	c.757C>T R253W	None	Exon 5 (repeat 6)	C→T at position 757	Effect on function has not been defined	Affect might be altered by other genetic/environmental factors	(Khoo <i>et al.</i> , 2000) (Fouchier <i>et al.</i> , 2001) (Thiart <i>et al.</i> , 2000) (Nauck <i>et al.</i> , 2001)
8	c.1217G>A R406Q	None	Exon 9 (EGF spacer)	G→A at position 1217	Effect on function has not been defined	Increase	(Thiart <i>et al.</i> , 2000) (Tosi <i>et al.</i> , 2007)
9	c.1222G>A E408K FH-Algeria 1, FH-Osaka	None	Exon 9 (EGF spacer)	G→A at position 1222	<2% LDLR activity in homozygote 5–10% LDLR activity in homozygote	Increase	(Hobbs <i>et al.</i> , 1992) (Webb <i>et al.</i> , 1996) (Lind <i>et al.</i> , 1998) (Amsellem <i>et al.</i> , 2002)
10	c.2096C>TP699L	None	Exon 14 (repeat C)	C→T at position 2096	Effect on function has not been defined	Increase	(Schuster <i>et al.</i> , 1995) (Thiart <i>et al.</i> , 2000) (Fouchier <i>et al.</i> , 2001) (Van Gaal <i>et al.</i> , 2001)

	LDLR variant	rs number	Region	Changes in DNA sequence	LDLR function	Effect on LDL-c	References
11	c. 2441G>A R814Q FH Xhosa	None	Exon 17 (cytoplasmic region)	G→A at position 2441	Reduction in LDL catabolism but effect on function has not been defined	Increase	(Arca and Jokinen 1998) (Thiart <i>et al.</i> , 2000) (Fouchier <i>et al.</i> , 2001) (Humphries <i>et al.</i> , 2006)

2.5.5 Regulation of the LDL receptor by PCSK9

2.5.5.1 The *PCSK9* gene

The gene which encodes the PCSK9 protein was identified by Abifadel *et al.*, in 2003 as the third locus for FH in patients that presented with an autosomal dominant hypercholesterolaemia phenotype (Abifadel *et al.*, 2003). This French research team identified missense gain-of-function (GOF) mutations that were associated with increased LDL-c levels. The gene is situated on chromosome 1p32 and consists of 12 exons and 11 introns. Loss-of-function (LOF) mutations were later identified by Cohen *et al.*, (2005). These two LOF mutations were caused by nonsense mutations and were found predominantly in African Americans and were associated with hypocholesterolaemia (Cohen *et al.*, 2005). Since the first discovery of PCSK9, a number of GOF and LOF mutations have been identified in various populations but little is known about these variants in the black South African population. One hundred and one variants have been reported on the website of the British Heart Foundation (www.ucl.uk/ldlr/LOVDv.1.1.0/) and one of these variants (S127R) was found in South African patients of mixed ancestry with FH (Homer *et al.*, 2008).

2.5.5.2 PCSK9 is a serine protease

Proprotein convertase subtilisin kexin 9 (PCSK9) is a subtilase (enzymes that cleave peptide bonds in proteins (Hedstrom, 2002)) that is expressed mainly in the liver, the intestine and, to a lower extent, in the kidney, skin and brain (Seidah *et al.*, 2003) but only the liver secretes it into the plasma (Zaid *et al.*, 2008). The PCSK9 protein is synthesised as a 72kDa pro-enzyme in the endoplasmic reticulum and consists of a signal peptide (illustrated in Figure 7) (residues 1 to 30; first part of exon 1), a pro-domain (residues 31 to 152; end part of exon 1 to first part of exon 3) and a catalytic domain (residues 153 to 451; end part of exon 3 to middle part of exon 8), followed by a C-terminal domain (residues 452 to 692; exon 9 to 12) that are rich in cysteines and histidines (Horton *et al.*, 2007). The C-terminal domain is also termed the V domain (Piper *et al.*, 2007). The protein undergoes autocatalytic intra-molecular processing at the FAQ152↓SIP site to form a 63kDa mature enzyme found in the ER and Golgi (Seidah *et al.*, 2003).

The pro-domain serves as a chaperone for folding as well as an inhibitor of catalytic activity. The autocatalytic cleavage between the pro-domain and the catalytic domain separates them from each other but the pro-domain stays bound, blocking the catalytic site (Kwon *et al.*, 2008). The last four

amino acids of the prodomain are the part which covers the catalytic triad, causing the restricted access. The next domain is the C-terminal domain where variant C679X prevents the folding of the C-terminal domain but does not prevent autocatalytic cleavage, suggesting that cleavage is cotranslational (Horton *et al.*, 2007). Other proprotein convertases (PC) are catalytically cleaved for a second time to release the pro-domain and expose the catalytic activity of the protein, thereby activating it. No second activation cleavage site has been identified for PCSK9 (Kwon *et al.*, 2008). The mature protein and the N-terminal pro-domain which stays tightly attached to it throughout the whole secretory pathway undergo glycosylation (Seidah *et al.*, 2003), phosphorylation (Dewpura *et al.*, 2008) and tyrosine sulfation before secretion (Benjannet *et al.*, 2006). These modifications are not required for PCSK9 to be excreted and their role in PCSK9 function is still unclear (Horton *et al.*, 2009).

PCSK9 binds to LDL receptors extracellularly to be internalised into cells. The LDL receptor-PCSK9 complex, together with the adapter protein ARH, is internalised. PCSK9 moves with the LDL receptor to the endosome and the lysosome but the protein does not destroy the LDL receptor (Seidah *et al.*, 2003).

2.5.5.3 PCSK9 protein function

PCSK9 functions by reducing the number of LDL receptors in hepatocytes (Zaid *et al.*, 2008). As previously mentioned, GOF and LOF mutations have been detected in various studies. GOF mutations increase the concentration of the PCSK9 protein in the circulation or they increase the affinity of the protein for LDL receptors; therefore, more LDL receptors are degraded, leaving fewer receptors on the cell surface to internalise LDL-c, which, in turn, causes an increase in circulating LDL-c. The opposite is true for LOF mutations and the diminished function of the PCSK9 protein results in an increased number of LDL receptors on the cell surface. The increased number of LDL receptors is very effective in internalising LDL-c, and lower levels of circulating LDL-c are detected in affected individuals (Abifadel *et al.*, 2003, Abifadel *et al.*, 2009). Inactivation of PCSK9 does not lower plasma cholesterol level in LDLR^{-/-} mice, which suggests that the effect of PCSK9 on LDL-c is facilitated through LDL receptors only (Zaid *et al.*, 2008). PCSK9 is secreted from cells, and after the extracellular PCSK9 binds to the LDL receptor, the complex is internalised, thereby promoting its degradation. For the internalisation process to take place, the ARH adaptor protein needs to be present (Horton *et al.*, 2007). A study by Lagace *et al.* (2006) shows that the overexpression of PCSK9 also promotes the same LDL receptor degradation in ARH knockout and control mice. This

could indicate that PCSK9 also acts on LDL receptors after its biosynthesis and before it reaches the surface of the hepatocyte (Lagace *et al.*, 2006). This intracellular mode of action of PCSK9 is supported by the observation that two naturally occurring PCSK9 mutants (S127R and D129G) that are not secreted from the cells were still associated with ADH (Homer *et al.*, 2008). Whether this occurs only when PCSK9 activity is extremely up-regulated or also under normal conditions remains to be tested (Niesen *et al.*, 2008).

It would appear that the intracellular trafficking of PCSK9 from the ER to downstream sites is controlled by the LDL receptor but the manner of interaction between the receptor and PCSK9 seems to differ in the ER and at the cell surface (Nassoury *et al.*, 2007). The region where interaction takes place between the secreted PCSK9 and the extracellular domain of the receptor was found to be the first epidermal growth factor-like repeat homology domain (EGF-A) in human LDL receptors (Li *et al.*, 2007) to bind in a calcium-dependent manner (Zhang *et al.*, 2007). LDL receptor regions which are needed for PCSK9 degradation but do not make contact with PCSK9 are the β -propeller domain and at least three ligand binding repeats (Zhang *et al.*, 2008). The C-terminal domain of PCSK9 is another region required for LDL receptor degradation even though the region does not bind to the receptor (Zhang *et al.*, 2008).

It was found that mutations that interfere with PCSK9 binding in the EGF-A domain of the LDL receptor inhibit PCSK9-mediated degradation of the receptor. The recent crystal structure of the PCSK9/EGF-A complex shows that the EGF-A domain of the LDL receptor binds to the apex of the catalytic domain of PCSK9, and that this binding is highly specific (Kwon *et al.*, 2008). PCSK9 also binds and targets two other transmembrane proteins closely related to LDL receptor for degradation *in vitro*, namely the VLDLR and the apoE2 (Poirier *et al.*, 2008). Another *in vitro* study showed that LDL particles compete with PCSK9 for LDL receptor binding, thereby inhibiting PCSK9 function (Fisher *et al.*, 2007). PCSK9 dimers are also more effective in degrading LDL receptors than monomers, as discovered in an *in vitro* study by Fan (2008). The gain-of-function variant, D374Y, displays greater self-association than the wild-type protein; it also has greater affinity for the receptor and greater LDL receptor degradation activity. This study also showed that HDL particles inhibited PCSK9 dimerisation, whereas VLDL augmented it, suggesting that lipoproteins may modulate PCSK9 functions (Fan *et al.*, 2008).

The ligand-binding repeats 4 and 5 closely associate with the β -propeller domain at an acidic pH, as indicated by the crystal structure of the LDL receptor. The receptor changes from an open to a closed conformation in the acid pH, releasing the LDL particles (Rudenko *et al.*, 2002). The affinity

of PCSK9 towards LDL receptor increases radically at acidic endosomal pH versus neutral pH at the cell surface. This is consistent with the findings that LDL receptors bound to PCSK9 are targeted for lysosomal degradation after endocytosis and they are not recycled to the cell surface (Fisher *et al.*, 2007; Zhang *et al.*, 2007).

2.5.5.4 Classification of *PCSK9* mutations

The mutations in the *PCSK9* gene can be divided into five groups according to their effects on the synthesis and secretion of the PCSK9 protein (Horton *et al.*, 2007). Class 1 mutations are the class in which no detectable protein is produced, and so far, only one null allele variant has been identified, namely Y142X. Class 2 is the processing defective group where the mutations interfere with the autocatalytic cleavage of the prodomain; therefore the protein is not secreted into the circulation but functions between the endoplasmic reticulum and the cell membrane. The transport of PCSK9 is delayed from the ER to the cell surface in class 3 mutations and all class 2 mutations also delay this transport. The truncated protein formed by the nonsense mutation C679X is cleaved normally but is misfolded and retained in the ER (Benjannet *et al.*, 2006; Zhao *et al.*, 2006b).

Class 4 mutations include mutations that affect the stability of PCSK9 and are found amongst LOF mutations. In cultured cells, a small number of the PCSK9 that are synthesised undergoes a second, membrane-bound furin-mediated cleavage event. LOF mutations such as A443T could possibly be more susceptible to furin cleavage (Benjannet *et al.*, 2006).

The affinity of PCSK9 for the LDL receptor or other proteins that promote receptor degradation could also be affected by various mutations and this concludes the class 5 mutations. Variant D374Y is a GOF mutation which binds with greater affinity to the LDL receptor and reduces LDL receptor activity to approximately ten times that of the wild-type protein (Lagace *et al.*, 2006). Another GOF mutation that could also fall into class 5 is F216L, which is predicted to reside very close to D374Y (Zhao *et al.*, 2006a).

2.5.5.5 Circulating PCSK9

No measurable PCSK9 can be detected in mice livers in which PCSK9 has been inactivated and this suggests that the liver is the major source of circulating PCSK9 (Zaid *et al.*, 2008). ELISAs to

measure plasma human PCSK9 have been developed by several laboratories (Careskey *et al.*, 2008, Horton *et al.*, 2007; Lambert *et al.*, 2008; Mayne *et al.*, 2007). Due to the difference in antibody specificity and the standards used in these assays the mean PCSK9 concentration varies extensively, ranging from 500ng/ml to 4000ng/ml.

PCSK9 levels in the multi-ethnic population of the Dallas Heart Study varied from 33 to 2988 ng/ml (Lakoski *et al.*, 2009). In a study on children and adolescents of the Quebec Child and Adolescent Health and Social Survey (QCAHSS), PCSK9 levels ranged from 17.6 to 212µg/L (Baass *et al.*, 2009). Serum PCSK9 levels in a large Han Chinese population varied from 12.9 to 222ng/ml (Cui *et al.*, 2010). For each of these studies a different ELISA was used, which may explain the difference in the ranges; it could also be that PCSK9 levels differ between populations. In a review by Lambert *et al.*, (2008), the different methods used in various studies were also noted, as well as the differences in the levels detected in these studies (Alborn *et al.*, 2007; Careskey *et al.*, 2008; Lagace *et al.*, 2006; Lambert *et al.*, 2008; Mayne *et al.*, 2007)

PCSK9 levels correlate with LDL-c but not with HDL-C (Careskey *et al.*, 2008; Lambert *et al.*, 2008). PCSK9 is transcriptionally regulated by sterol regulating element-binding proteins (SREBP) (Horton *et al.*, 2007), which regulate genes required for fatty acid synthesis (SREBP-1c) and cholesterol synthesis (SREBP-2) (Brown & Goldstein, 2009). The activity of SREBP-2 is modulated by cellular sterol levels (Horton *et al.*, 2003). The SREBP binding site is in the promoter region of the *PCSK9* gene (Jeong *et al.*, 2008). SREBP-2 plays a more pivotal role in PCSK9 regulation than SREBP-1c (Costet *et al.*, 2006), especially under physiological conditions (Horton *et al.*, 2007). Fasting or cholesterol feeding that suppresses SREBP-2 is associated with a decrease in PCSK9 mRNA levels in mice livers (Jeong *et al.*, 2008). Mice without liver SREBP-2 have lower PCSK9 mRNA levels whereas mice that lack SREBP-1c do not (Horton *et al.*, 2009). Statins increase PCSK9 protein levels in the plasma of hypercholesterolaemia patients, which shows that PCSK9 blocks the efficiency of this drug (Careskey *et al.*, 2008). Fibrates decrease the expression of PCSK9 in the liver and its stimulation by statins. The expression of inactivators of PCSK9 such as furin and PC5/6A is also increased by fibrates (Kourimate *et al.*, 2008). Chenodeoxycholic acid, a bile acid (Langhi *et al.*, 2008), and berberine, a natural hypocholesterolaemic agent, also repress PCSK9 expression but the pathways involved are still unclear (Cameron *et al.*, 2008).

The expression of PCSK9 promotes the degradation of liver LDL and therefore plasma LDL-c and PCSK9 levels may be directly associated. The levels of these two proteins are not always linked because treatment with high-dose statins reduces plasma LDL-c levels but it increases the levels of

circulating PCSK9 (Careskey *et al.*, 2008). The half-life of recombinant PCSK9 in the blood of wild-type mice is five minutes and disabling the LDL receptor increases the half-life of PCSK9 to 15 minutes, therefore implicating the LDL receptor as a major pathway for PCSK9 removal (Grefhorst *et al.*, 2008).

2.5.5.6 PCSK9 crystal structure

Three groups have reported similar structures of PCSK9 (Cunningham *et al.*, 2007; Fisher *et al.*, 2007; Hampton *et al.*, 2007). The catalytic domain is similar to that of other subtilisin-like serine proteases except for the high-affinity interactions with the prodomain that shields the active site from exogenous substrates. The C-terminal domain is a unique structure that contains three six-stranded β -sheet subdomains arranged with quasi-threefold symmetry (Horton *et al.*, 2009).

2.6 SUMMARY

In summary it can be seen that the nutrition transition has detrimental effects on the health of a population. Unhealthy lifestyles such as an increased intake of saturated fats cause a greater risk of the development of CVD. Research has clearly shown the importance of hypercholesterolaemia in the development of atherosclerosis and the risk of CAD. Because a large part of the black South African population is in the early stages of the nutrition transition, it is of utmost importance to determine what factors have an effect on CAD risk factors such as LDL-c levels. Developing countries move at a faster pace through the nutrition transition and it is important to elucidate what role LDL-c plays in developing non-communicable disease. The genetic background and the effect on the LDL-c levels of this population have not been studied in depth and therefore more information on these topics is needed to make better decisions on public health strategies in South Africa. *LDLR* and *PCSK9* are two genes that have a direct association with LDL-c. As hypercholesterolaemia has been identified as one of the major risk factors for developing CAD, it is necessary to study the effects that these two genes have on this very important risk factor.

CHAPTER 3

COMMON AND RARE SINGLE NUCLEOTIDE POLYMORPHISMS IN THE *LDLR* GENE ARE PRESENT IN A BLACK SOUTH AFRICAN POPULATION AND ASSOCIATE WITH LOW DENSITY LIPOPROTEIN CHOLESTEROL LEVELS

Tertia van Zyl,^{1,*} Johann C. Jerling,¹ Karin R. Conradie,¹ Edith J.M. Feskens²

¹ Centre of Excellence for Nutrition, North-West University, 2531, Potchefstroom, South Africa; ² Division of Human Nutrition, Wageningen University, 6700, Wageningen, The Netherlands

Running title: *Common and rare LDLR SNPs in an African population*

Published in The Journal of Human Genetics (online 28 November 2013)

Tertia Van Zyl - Journal of Human Genetics, DOI: JHG.2013.123, Online Publication Notification

From: <a.yoshida@nature.com>
To: <tertia.vanzyl@nwu.ac.za>
Date: 2013/11/28 10:08 AM
Subject: Journal of Human Genetics, DOI: JHG.2013.123, Online Publication Notification

Dear Ms Tertia Van Zyl,

I am pleased to inform you that your paper 'Common and rare single nucleotide polymorphisms in the LDLR gene are present in a black South African population and associate with low density lipoprotein cholesterol levels' is now published online in *JHG* <http://www.nature.com/doifinder/10.1038/jhg.2013.123> in our Advance Online Publication (AOP) service.

Please be aware that your published article is the definitive version and we are unable to correct errors after publication. If errors are found which affect the sense of the article, or the indexing and citation, a corrigendum may be published.

Your article will benefit from services including:

- **Wide exposure** - A large worldwide readership is ensured through the journal's institutional site licenses. Your article will also be freely available to scientists in the developing world through the WHO-brokered HINARI initiative.
- **High visibility of content** - All content is fully searchable and accessible through <http://www.nature.com>, with articles frequently highlighted on NPG's Research and Reviews section subject areas. Thousands of scientists also see current journal content through Table of Contents e-alerts, delivered each month, highlighting the articles in that month's issue.
- **Marketing support** - All NPG journals receive specific and tailored marketing to further ensure content is seen by the correct audience. Activities include conference attendance, direct mail, advertising, online promotion and NPG's press release service.
- **Abstracting and indexing** - Journals published by NPG are included in the leading abstracting and indexing services including ISI, MEDLINE, EMBASE and BIOSIS. Extensive reference linking to MEDLINE, ISI, Chemport and others via services such as CrossRef and DOI numbers provide seamless online linking between articles and databases.

More information about Nature Publishing Group can be found at <http://npg.nature.com/authors>.

Nature Publishing Group

DISCLAIMER: This e-mail is confidential and should not be used by anyone who is not the original intended recipient. If you have received this e-mail in error please inform the sender and delete it from your mailbox or any other storage mechanism. Neither Macmillan Publishers Limited nor any of its agents accept liability for any statements made which are clearly the sender's own and not expressly made on behalf of Macmillan Publishers Limited or one of its agents. Please note that neither Macmillan Publishers Limited nor any of its agents accept any responsibility for viruses that may be contained in this e-mail or

ABSTRACT

The LDL receptor plays an essential role in regulating plasma LDL-c levels. Genetic variation in the *LDLR* gene can be associated with either lower or moderately raised plasma levels of LDL-c, or may cause familial hypercholesterolemia. The prevalence of single nucleotide polymorphisms in the *LDLR* in the black South African population is not known and therefore we aimed to determine the genotypic variation of the *LDLR* in the study population as well as to define the association of the different genotypes with plasma LDL-c levels. A random selection of 1 860 apparently healthy black South African volunteers aged 35 to 60 years was made in a cross-sectional study. Novel SNPs were identified in a subset of 30 individuals by means of automated sequencing before screening the entire cohort by means of the Illumina VeraCode™ GoldenGate Genotyping Assay on a BeadXpress™ Reader system. Twenty-five SNPs were genotyped, two of which were novel. A very rare SNP, rs17249141, in the promoter region was significantly associated with lower levels of LDL-c. Four other SNPs (rs2738447, rs14158, rs2738465 and rs3180023) were significantly associated with increased levels of LDL-c. We can conclude that some of the various SNPs identified do indeed associate with LDL-c levels.

Keywords: LDLR/LDL-c/African population/haplotypes

INTRODUCTION

The LDL receptor is a cell surface glycoprotein which plays a focal role in regulating low-density lipoprotein cholesterol (LDL-c) levels in humans as it is the major protein involved in cholesterol transport. The receptor mediates the specific binding, uptake and degradation of plasma lipoproteins containing either apolipoprotein (apo) B or apoE. Remnants of triglyceride-rich lipoprotein metabolism (also known as intermediate-density lipoprotein (IDL), which are the precursors of plasma LDL and LDL itself, are recognised by the LDL receptor. It regulates plasma cholesterol by mediating the endocytosis of LDL-c by means of clathrin-coated pits.¹ An elevated LDL-c level is a major risk factor for the development of coronary artery disease (CAD) through the role it plays in the aetiology of atherosclerosis.² The LDL-c levels of the black South African population are thought to be historically low³ and the prevalence of CAD is minimal.⁴ Current research indicates that the prevalence of CAD risk factors such as hypercholesterolaemia is increasing in this population.^{40,41} Genetic variation influences LDL-c levels in a population and 50% of the inter-individual variability is due to this genetic component.⁵ A large number of genetic variants have been identified that cause a range of LDL-c levels in various populations.⁶ Some variants, such as the single nucleotide polymorphism (SNP) found in the promoter region of the *LDLR* gene, are able to increase the

promoter activity by approximately 50% and cause lower LDL-c levels.⁷ The question arises whether there are genetic variants present in the South African Prospective Urban and Rural Epidemiological (PURE) study population that contribute to the different levels of LDL-c, and to what degree these variants associate with LDL-c levels. The aim of this study was thus to determine whether single nucleotide polymorphisms in the *LDLR* gene are present in the study population and to what extent these variants associate with LDL-c.

METHODS AND MATERIAL

Study population

The study population is part of the South African Prospective Urban and Rural Epidemiological (PURE) study and consists mainly of self-identified Tswana-speaking individuals from the North West Province of South Africa. The PURE study is a prospective cohort study that follows changes in lifestyle, risk factors and chronic disease by means of periodic standardised data collection in two urban (Ikageng and Sonderwater extension 7 and 11) and two rural (Ganyesa and Tklagameng) areas over a period of 12 years. The central hypothesis of the study states that lifestyle (physical activity and nutrition) and psychosocial transitions secondary to urbanisation and industrialisation lead to behaviours that cause the development of chronic non-communicable disease, such as CVD, diabetes, high blood pressure and obesity. Baseline data used in this study was collected in 2005. Two thousand eligible volunteers were randomly selected from the four areas by door to door recruitment. Individuals were suitable to take part in the study if they were between the ages of 35 and 60 years and if they were not using medication for chronic diseases. One participant using lipid lowering medication was excluded from this study. To avoid bias from population stratification, the urban and rural cohorts were matched for ethnicity, gender and age. The design of the PURE study has been described previously in greater detail.⁸ The guidelines laid down in the Declaration of Helsinki were followed throughout the study. Coded data was used to perform all analysis and all data were treated confidentially. Informed consent was obtained from each participant in the PURE study for the collection of blood samples and ethical approval for the study was given by the Ethics Committee of the North-West University (Ethics number: NWU-00016-10-A1). DNA samples for 1 860 of the volunteers were available for analysis and genotyping of the *LDLR* gene was successful for 1530 individuals.

Biochemical analysis

Blood was sampled after eight hours of fasting. Serum lipid (total cholesterol (TC), HDL and TG) measurements were carried out by a Sequential Multiple Analyzer Computer (SMAC) using a Konelab™ auto analyser (Thermo Fisher Scientific, Vantaa, Finland). LDL-c levels were determined by using the Friedewald equation $[LDL-c = TC - HDL-C - (TG)/2.17]^9$ and missing values were assigned to individuals with a TG measurement of $>4.5\text{mmol/L}$.

DNA analysis

Deoxyribonucleic acid (DNA) isolation was performed by using Qiagen® Flexigene™ DNA extraction kits (QIAGEN, USA: catalogue number 51206). A sub-sample of 30 randomly chosen individuals was used to identify known and novel mutations in the exons and flanking regions of the *LDLR* by means of Sanger sequencing. Thirty individuals would give a total sample size of 60 chromosomes and, as communicated by Collins and Schwartz, using 40 to 65 chromosomes will detect polymorphisms at a frequency of 0.05 with a power of 0.80 to 0.95.¹⁰ Polymerase chain reaction (PCR) amplification of the proximal promoter region, the exons and the exon-intron boundaries of the *LDLR* gene was performed on an iCycler (BioRad). Primers were designed to amplify and sequence the various regions of the *LDLR* and these are described in supplementary documentation (Annexure E). NucleoFast® 96 PCR Clean-Up Kits from Machery-Nagel were used for the clean-up of the PCR product prior to performing the sequencing process. Automated cycle sequencing was performed according to the basic principles of Sanger sequencing.¹¹ The standard ABI Prism®, BIGDye® Terminator version 3 Ready Reaction Cycle Sequencing Kit supplied by Applied Biosystems was used as a platform for the sequencing reactions. The purified product underwent electrophoresis on a 50cm capillary either on an ABI3730xl DNA analyser or an ABI3120xl genetic analyser according to standard protocols.

Genotyping

The sequences that were generated were examined for the presence of variants using the CLC DNA Workbench version 6.0.2. Variants were confirmed by repeat sequencing in the opposite direction. Seventy variants were identified within both the coding and non-coding areas of the *LDLR* gene, including one insertion (a 1bp insertion, c.*1218_1219Gins, in the 3'UTR not previously reported), one deletion (rs17249078) and 68 single nucleotide polymorphisms (these are available in Annexure D). Five of the SNPs were novel and not previously reported. Genotyping of the population was done by means of the Illumina VeraCode™ GoldenGate

Genotyping Assay on a BeadXpress™ platform using the VeraCode™ technology. All 68 SNPs (rs numbers for known SNPs and sequences for novel SNPs) were sent through to Illumina Technical Support for evaluation using the Assay Design Tool. Each SNP was scored (varying from 0 to 1) by the Assay Design Tool based on compatibility with successful GoldenGate genotyping. SNPs with a score above 0.4 were chosen for genotyping. Fourteen SNPs fell out for selection because they were located less than 60bp from another SNP (this is a prerequisite for the assay design) and 10 SNPs had a design score of less than 0.4. Twenty-nine SNPs were then selected on the basis of their frequency (>0.02) in the sub-sample, novelty and whether they were previously reported to associate with LDL-c. A further four SNPs (rs3745677, novel 5, rs2304182, rs5742911) failed during the execution of the assay on the BeadXpress platform. The study population was finally genotyped for 25 SNPs (of which two were novel and not previously reported). SNPs with a Gene Call Score of >0.5 and samples with a call rate of ≥ 0.9 were included in the analysis. The A of the initiating ATG codon is designated as nucleotide c.1 and the reference sequence *LDLR*: NM_001195798.1 was used.

Haplotype frequencies

Haplotype frequencies were determined by means of the software package Haploview version 4.2.25 to determine the haploblocks. Markers with a Hardy-Weinberg p-value of ≥ 0.05 and a minor allele frequency of $\geq 1\%$ were included in the analysis. Association analyses were conducted using LDL-c of $\leq 3\text{mmol/L}$ for controls and $>3\text{mmol/L}$ for cases.

Measurements

The anthropometric measurements such as weight (kilograms), height (metres) and body mass index ($\text{BMI} = \text{weight in kilograms}/\text{height}^2$ in metres) were supervised by a level 3 anthropometrist according to guidelines adopted at the Airlie Conference sponsored by the National Institutes of Health. Standardised and calibrated equipment was used for measurements on volunteers wearing minimal clothing.¹²

Diet

Dietary intakes of the volunteers were determined by means of previously developed¹³ and validated¹⁴ culturally sensitive quantified food frequency questionnaires. The nutrient intakes

were analysed by using the Foodfinder3® program (Medical Research Council, Tygerberg, South Africa), which is based on the South African Food Composition tables¹⁵.

Statistical analysis

Descriptive statistics on the study population were computed on the statistical program SPSS version 20 and are expressed in means and standard deviations. The observed genotype frequencies were tested for deviation from Hardy-Weinberg equilibrium by using the Chi-square test of independence (χ^2). Analyses of covariance were used to test association between different genotypes and LDL-c levels. Adjustments for the confounders age, gender and body mass index (BMI) were made in all analyses. The Bonferroni correction was used to deal with the issue of multiple testing. P-values of 0.05 or lower were considered statistically significant.

RESULTS

The characteristics of the study population are shown in Table 1. The mean LDL-c levels for the men were 2.70mmol/L (SD 1.12) and for the women, 3.06mmol/L (SD 1.17). The mean age for both men and women was similar at 49.9 yrs (SD 10.4) and 49.1 yrs (SD 10.4), respectively. Twenty-five of the 70 variants identified by bidirectional sequencing of PCR products were finally genotyped for this study population. Two of these variants were novel and have not been previously reported.

The European and African populations from the 1000 Genomes Project were used to compare the minor allele frequency of the study population (Table 2). The following rare variants were absent in the European population: rs17249141, rs148698650, rs116405216 and rs72658862.

All four previously mentioned variants were present in the study population but, according to the 1000 Genomes database, rs148698650 is not present in the African populations. Rare and common variants such as rs17242759, rs10423288, rs11669576, rs5929, rs4508523, rs6413503, rs3826810 and rs2738464 have a minor allele frequency that is greater in the study population and the African population than in the European population. Variants with a minor allele frequency greater than that of the study population and the other African populations that are present in the European population are as follows: rs2228671, rs5930, rs2738447, rs688, rs5925, rs14158, rs1433099 and rs3180023. Three common variants, rs1569372, rs5927 and rs2738465, have similar minor allele frequencies across the three different populations.

Table1 Population characteristics

Variables	Men		Women	
	n	Mean (SD)	n	Mean (SD)
Age (years)	696	49.9 (10.4)	1168	49.1 (10.4)
Total Cholesterol (mmol/L)	696	4.79 (1.31)	1168	5.13 (1.39)
LDL-c (mmol/L)	696	2.70 (1.12)	1168	3.06 (1.17)
HDL-C (mmol/L)	696	1.58 (0.65)	1168	1.48 (0.62)
TG (mmol/L)	696	1.15 (0.58)	1168	1.33 (0.70)
Energy intake (kJ)	675	8039 (3762)	1133	7231 (3373)
%Fat of TE	675	23.17 (7.25)	1133	24.55 (7.72)
%SFA of TE	675	5.53 (2.50)	1133	5.93 (2.78)
BMI categories:				
1 (<19 kg/m ²)	196	17 (1.15)	130	16.8 (1.25)
2 (19 - 24.99 kg/m ²)	373	20.9 (1.70)	390	22 (1.85)
3 (25 - 29.99 kg/m ²)	68	26.8 (1.18)	246	27.4 (1.48)
4 (>30 kg/m ²)	22	34.1 (3.63)	338	35.5 (4.81)
Urban/Rural	399/348		603/657	
LDL-c (<3 mmol/L / >3mmol/L)	453/243		618/550	

Abbreviations: LDL-c, Low density lipoprotein cholesterol; HDL-C, high density lipoprotein cholesterol; TG triglycerides; TE, total energy; SFA, saturated fatty acids; BMI, body mass index

Table 2 Minor allele frequency of the various SNPs in *LDLR* in different populations

SNP	Base pair	Minor (Major) Allele	PURE population frequency	MAF 1000 Genomes AFR ^a	MAF 1000 Genomes EUR ^a
rs369850745 ^b	11199911	T (C)	0.01	NA	NA
rs17249141	11200008	T (C)	0.01	0.02	0.00
rs17242759	11200309	A (C)	0.09	0.06	0.01
rs2228671	11210912	T (C)	0.03	0.02	0.09
rs10423288	11215846	C (T)	0.14	0.16	0.02
rs148698650	11218079	A (G)	0.01	0.00	0.00
rs116405216	11221324	A (G)	0.06	0.02	0.00
rs11669576	11222300	A (G)	0.16	0.24	0.05
rs72658862	11223929	T (C)	0.06	0.02	0.00
rs5930	11224265	A (G)	0.14	0.21	0.41
rs1569372	11224491	G (A)	0.42	0.29	0.47
rs5929	11226800	T (C)	0.10	0.12	0.04
rs4508523	11226944	T (C)	0.44	0.46	0.12
rs2738447	11227480	A (C)	0.12	0.16	0.43
rs688	11227602	T (C)	0.04	0.06	0.47
rs5925	11230881	C (T)	0.19	0.17	0.45
rs5927	11233941	A (G)	0.27	0.35	0.27
rs369402076 ^b	11234029	G (T)	0.02	NA	NA
rs6413503	11241904	A (G)	0.11	0.16	0.01
rs14158	11242044	A (G)	0.11	0.12	0.22
rs3826810	11242133	A (G)	0.13	0.13	0.03
rs2738464	11242307	G (C)	0.32	0.38	0.13
rs2738465	11242469	A (G)	0.26	0.25	0.26
rs1433099	11242658	C (T)	0.48	0.40	0.71
rs3180023	11243209	A (C)	0.001	0.01	0.08

Abbreviations: NA, Not available; MAF, minor allele frequency; ^aMAF in individuals from African (AFR) and European (EUR) descent taken from <http://browser.1000genomes.org>. ^bNovel single nucleotide polymorphisms

Variant c.-314 C→T (rs369850745) was found in the promoter region of the gene but the association with LDL-c levels was not statistically significant ($p=0.85$) (Table 3). The other novel SNP found in intron 15, c.2311+9 T→G (rs369402076), was also not associated with LDL-c levels. For both variants only heterozygote carriers were detected and both variants were in HWE. Both these novel SNPs had a very low minor allele frequency of 0.01 and 0.02 for novel rs369850745 and rs369402076 respectively. A very rare variant in the promoter region of the *LDLR* gene, rs17249141 (c.-217 C→T), had a significant association with decreased LDL-c levels ($p=0.05$). For variant rs5930, detected in exon 10, the mean LDL-c level (3.29mmol/L) for the AA genotype was higher than for the other two genotypes although this was not statistically significant ($p=0.09$); the mean difference in LDL-c level for the AA genotype is 13.4% compared to the other two genotypes. In exon 11, variant rs5929 had a 3.5–5% difference in LDL-c levels for each mutant allele even though there was no significant association with LDL-c levels. Intron 11 harboured the variant rs2738447, where the AA genotype was significantly associated with elevated LDL-c levels with a 23.7% difference in mean LDL-c levels when the individual had both the mutant alleles ($p=0.004$). SNP rs5925 was detected in exon 13, where the CC genotype had a mean LDL-c level greater than the TT and the TC genotypes although the association did not reach statistical significance ($p=0.06$). In the 3' untranslated region (3'UTR), rs14158 was significantly associated ($p=0.05$) with mean LDL-c levels with a difference of 27.9% when both of the mutant alleles (AA) were present. Variants rs2738465 and rs3180023, also in the 3'UTR, were both significantly associated with elevated LDL-c levels ($p=0.05$ and $p=0.006$, respectively). All the SNPs were in HWE except for rs6413503.

Table 3 *LDLR* genotype distribution and association with LDL-c

SNP number	Nucleotide substitution / Position	p (HWE)	Frequency	LDL-c (mmol/L) 95% CI	ANCOVA p-value ^k
rs369850745	c.-314 C>T	0.91			
CC	(novel)		1501	2.94 (2.88-2.99)	0.852
CT	Promoter		34	2.90 (2.53-3.28)	
TT			0	-	
rs17249141	c.-217 C>T	0.98			
CC	Promoter		1521	2.95 ^a (2.89-3.00)	0.050
CT			16	2.4 ^a (1.85-2.95)	
TT			0	-	
rs17242759	c.67+18 C>A	0.42			
CC	Intron 1		1270	2.94 (2.88-3.00)	0.120
CA			258	2.91 (2.77-3.04)	
AA			8	3.73 (2.95-4.50)	

Table 3 continues

SNP number	Nucleotide substitution / Position	p (HWE)	Frequency	LDL-c (mmol/L) 95% CI	ANCOVA p-value ^k
rs2228671	c.81 C>T	0.19			
CC	(C27C)		1457	2.95 (2.89-3.01)	
CT	Exon 2		77	2.81 (2.56-3.06)	0.474
TT			3	2.59 (1.33-3.85)	
rs10423288	c.314-50	0.99			
TT	T>C		1132	2.93 (2.87-2.99)	
TC	(IVS3-50)		373	2.97 (2.86-3.09)	0.444
CC	Intron 3		30	2.71 (2.31-3.11)	
rs148698650	c.829 G>A	0.93			
GG	(E277K)		1507	2.94 (2.88-2.99)	
GA	Exon 6		29	3.05 (2.64-3.45)	0.590
AA			0	-	
rs116405216	c.941-4 G>A	0.90			
GG	Intron 6		1364	2.94 (2.88-3.00)	
GA			164	2.92 (2.77-3.11)	0.945
AA			6	3.04 (2.15-3.93)	
rs11669576	c.1171 G>A	0.77			
GG	(A391T)		1082	2.91	
GA	Exon 8		412	3.01	0.320
AA			34	3.02	
rs72658862	c.1187-25	0.69			
CC	C>T		1346	2.94 (2.88-3.00)	
CT	Intron 8		163	2.94 (2.77-3.11)	0.653
TT			7	3.33 (2.50-4.15)	
rs5930	c.1413 A>G	0.98			
AA	(R471R)		29	3.29 (2.89-3.70)	
AG	Exon 10		357	3.04 ^b (2.93-3.16)	0.023
GG			1146	2.90 ^b (2.83-2.96)	
rs1569372	c.1586+53	0.99			
AA	A>G		518	2.99 (2.89-3.01)	
AG	Intron 10		747	2.90 (2.82-2.98)	0.376
GG			266	2.95 (2.82-3.04)	
rs5929	c.1617 C>T	0.97			
CC	(P539P)		1239	2.97 (2.91-3.03)	
CT	Exon 11		281	2.82(2.69-2.95)	0.077
TT			17	2.72 (2.19-3.25)	
rs4508523	c.1705+56	0.81			
CC	C>T		493	2.99 (2.89-3.09)	
CT	Intron 11		739	2.92 (2.84-3.00)	0.446
TT			296	2.90 (2.77-3.03)	
rs2738447	c.1706-55A>C	0.64			
CC	Intron 11		1188	2.9 ^c (2.84-2.96)	
CA			322	3.03 ^d (2.92-3.16)	0.001
AA			27	3.59 ^{cd} (3.18-4.02)	

Table 3 continues

SNP number	Nucleotide substitution / Position	p (HWE)	Frequency	LDL-c (mmol/L) 95% CI	ANCOVA p-value ^k
rs688	c.1773 C>T	0.82			
CC	(N591N)		1381	2.95 (2.89-3.00)	0.051
CT			132	2.83 ^e (2.64-3.02)	
TT			2	4.62 ^e (3.08-6.16)	
rs5925	c.1959 T>C	0.92			
TT	(V653V)		1001	2.97 (2.90-3.04)	0.020
TC	Exon 13		475	2.84 ^f (2.74-2.94)	
CC			60	3.19 ^j (2.91-3.47)	
rs5927	c.2232 A>G	0.18			
AA	(R744R)		96	2.90 (2.82-2.97)	0.212
AG	Exon 15		629	3.00 (2.92-3.09)	
GG			806	2.94 (2.72-3.17)	
rs369402076	c.2311+9 T>G	0.67			
TT	(novel)		1466	2.94 (2.88-2.99)	0.854
TG	Intron 15		68	2.96 (2.69-3.23)	
GG			0	-	
rs6413503	c.2548-42	<0.01			
GG	G>A		1267	2.91 (2.85-2.97)	0.089
GA	Intron 17		166	3.10 (2.93-3.27)	
AA			88	3.01 (2.78-3.24)	
rs14158	c.*52 G>A	0.83			
GG	3'UTR		1220	2.9 ^h (2.84-2.96)	0.001
GA			282	3.05 ^g (2.92-3.18)	
AA			19	3.71 ^{gh} (3.21-4.21)	
rs3826810	c.*141 G>A	0.13			
GG	3'UTR		1120	2.94 (2.88-3.01)	0.899
GA			360	2.94 (2.82-3.05)	
AA			17	2.82 (2.28-3.34)	
rs2738464	c.*315 G>C	0.59			
GG	3'UTR		165	2.87 (2.70-3.04)	0.620
GC			650	2.93 (2.85-3.02)	
CC			722	2.96 (2.88-3.04)	
rs2738465	c.*504 G>A	0.71			
GG	3'UTR		855	2.90 ⁱ (2.83-2.98)	0.051
GA			573	2.95 (2.86-3.04)	
AA			107	3.18 ⁱ (2.96-3.39)	
rs1433099	c.*666 T>C	0.20			
TT	3'UTR		403	2.92 (2.80-3.04)	0.917
TC			800	2.95 (2.87-3.02)	
CC			330	2.95 (2.84-3.06)	

Table 3 continues

SNP number	Nucleotide substitution / Position	p (HWE)	Frequency	LDL-c (mmol/L) 95% CI	ANCOVA p-value ^k
rs3180023	c.*1217 C>A	1.00			
CC	3'UTR		1535	2.94 ⁱ (2.88-2.99)	
CA			2	5.09 ^j (3.56-6.64)	0.006
AA	Exon 12		0	-	

^{a,b,c,d,e,f,g,h,i,j} Means with the same vowels differ significantly; ^kp-value after adjustment for age, gender and BMI; SNP, single nucleotide polymorphism, HWE, Hardy-Weinberg equilibrium, LDL-c, low density lipoprotein cholesterol, ANCOVA, analysis of covariance, C, cytosine, A, adenine, T, thymine, G, guanine

Haplotype analysis

Four blocks were identified from the haplotype analysis performed with the software package Haploview version 4.2.25 (results in Table 4). From the association analysis, all four blocks contained haplotypes that were associated with either elevated or lower LDL-c levels. Block 1 revealed two haplotypes, namely GCGGC and GCAAC, where the first haplotype was associated with lower LDL-c ($p=0.03$) and the second haplotype was associated with increased LDL-c ($p<0.01$). Block 2 also contained two haplotypes where the CT haplotype was associated with lower LDL-c as more controls presented with the haplotype. The other haplotype in Block 2 was the AT haplotype, which was associated with elevated LDL-c. Three haplotypes were calculated in Block 3 and two of these were associated with elevated LDL-c levels, namely AGG ($p=0.03$) and GAG ($p=0.01$). The third haplotype (GGG) in Block 3 was more prevalent amongst the controls and is therefore associated with lower levels of LDL-c. Only one of the haplotypes (GC) identified in Block 4 was associated with lower levels of LDL-c. Annexure F and G contain the graphical representations of the LD plot and the haplotypes of the *LDLR*.

DISCUSSION

For this study we sequenced the exons and flanking intron regions of the *LDLR* gene to determine the genotypic variation of the population within the gene. Of the 70 SNPs identified by the bidirectional sequencing, 25 were genotyped within the entire study population, only two of which were novel (rs369850745 and rs369402076). We also performed haplotype analysis to determine whether there were any associations with mean LDL-c levels.

Table 4 Haplotypes associated with LDL-c within the PURE population

Haplotype	Frequency	Case, control frequencies	P-value
Block 1 - rs11669576; rs72658862; rs5930; rs1569372; rs5929			
GCGGC	0.36	0.34, 0.37	0.03
GCGAC	0.19	0.18, 0.20	0.16
ACGAC	0.16	0.17, 0.15	0.22
GCAAC	0.14	0.16, 0.12	<0.01
GCGAT	0.11	0.10, 0.11	0.21
GCGAT	0.06	0.06, 0.05	0.29
Block 2 - rs2738447; rs5925			
CT	0.69	0.66, 0.70	0.01
CC	0.19	0.19, 0.20	0.42
AT	0.12	0.15, 0.10	<0.01
Block 3 - rs5927; rs14158; rs3826810			
GGG	0.50	0.46, 0.52	0.03
AGG	0.27	0.29, 0.25	0.03
GGA	0.13	0.13, 0.14	0.34
GAG	0.10	0.12, 0.09	0.01
Block 4 - rs2738465; rs1433099			
GT	0.53	0.54, 0.52	0.34
AC	0.25	0.27, 0.25	0.20
GC	0.22	0.20, 0.23	0.01

Cases were defined as individuals with LDL-c >3mmol/L and controls were defined as individuals with LDL-c <3mmol/L

There were no variants found in exons three, four and five in the sub-sample of 30 individuals. Together with exons two and six, the above-mentioned exons code for the ligand-binding domain which is of great importance in binding apoB-containing lipoproteins such as LDL. Mutations within repeats four and five of the ligand-binding domain (present in exon 4) are common in FH patients.¹⁶

We found that four rare variants (rs17249141, rs148698650, rs116405216 and rs72658862) were present within the study population as well as in the African populations from the 1000 Genomes database but they were absent from the European population. SNP rs148698650

was not detected in the 1000 Genomes African samples but it was identified in a woman (heterozygote) in the 1000 Genomes population of Mexican ancestry from Los Angeles. The SNP was originally detected in individuals of Spanish, Cuban and Swedish descent.¹⁷⁻¹⁹ This could possibly be the first report of rs148698650 in an African population. The minor allele frequencies of the SNPs in the study population are more comparable to the African populations than the European populations. Variants such as rs688 and rs5925, which were used in a genetic risk score to define the predictive value of genetics in determining cholesterol levels²⁰, would not be an obvious choice in the study population as their frequencies are much lower than in the European population. The frequencies and associations of variants in different populations should be investigated before they are used in further analyses to ensure their relevance.

The two novel SNPs (rs369850745 and rs369402076) identified had no meaningful effect on the function of the LDL receptor and there was no association with LDL-c levels. Variant rs369850745 is located more than 100bp from the SREBP1 binding site which could indicate that the SNP is too far away to have any effect on the transcription of the protein. In intron 15 the variant rs369402076 is most likely located just after the region that would influence the splicing of the mRNA or the base change does not have any effect on the splice function. Splice donor mutations 2311+2T>G²¹, 2311+1G>C²² and 2311+1G>A²³ are located close to rs369402076 (2311+9T>G) and these variants are associated with hypercholesterolemia.

Sixteen heterozygote carriers of variant rs17249141 (c.-217C>T) were identified in the study population. This variant was first reported by Scholtz et al. in 1999 in a subject of mixed ancestry. The promoter activity of this variant increased to 160% of normal in cells depleted of sterol and 50% in cells supplemented with a sterol-containing medium in the transfection study⁷. In this study the effect of one allele was an 18.6% difference in LDL-c levels, which is in accordance with an increased transcriptional rate of the gene.

In exon 11 the silent variant rs5929 (c.1617C>T; P539P) was detected and the base change from a cytosine to a thymine causes a synonymous mutation. As previously reported, the SNP is not associated with LDL-c levels²⁴ and the small lowering effect of the mutant allele in the study population might be caused by a variant with which the SNP is in linkage or it might be a specific phenomenon related to the study population.

The intronic variant rs2738447 (c.1706-55A>C) had a statistically significant association with elevated LDL-c where the homozygote mutants (genotype AA) had a mean LDL-c level of 3.59mmol/L. No genotype-phenotype information could be found for this variant. Analyses of this variant in the Human Splicing Finder indicate that the SNP could form part of various

enhancer motifs and therefore influence the splicing efficiency of the pre-mRNA, which leads to alternative splicing.²⁵ We can hypothesise that the function of the different proteins (receptors) formed by the alternative splicing could be less effective in internalising LDL-c than receptors synthesised by normal splicing, which would then lead to increased plasma levels of LDL-c.²⁵

Variants rs688 (c.1773 C>T) and rs5925 (c.1959 T>C) are both synonymous mutations and have been identified as exon-splicing enhancers associated with increased LDL-c levels by Rafiq et al. (2011). These variants alter the efficiency of LDLR splicing and therefore have an elevated effect on LDL-c.²⁶ As previously mentioned, these two variants have been used in a genetic risk score to determine the predictive value of genetics on plasma total cholesterol levels where the results demonstrated that it could be beneficial in identifying individuals with elevated plasma cholesterol levels.²⁰ In the study done by Lu et al.²⁰, the minor allele frequencies of both these SNPs were 0.43 in their European population. The minor allele frequencies of the SNPs in the PURE population were much lower (rs688 = 0.04; rs5925 = 0.19). Rs5925 and rs688 was in strong linkage disequilibrium (LD) with $r^2=0.99$ ²⁰; however, the LD of these two SNPs in the PURE population were much weaker, with $r^2=0.20$. The lower r^2 in the study population is also an indication that the two SNPs have been separated through recombination.²⁷ LD between most of the analysed SNPs (data not shown) is weak when r^2 is used as a measure of LD, which is a characteristic of the African genome where more recombination events have taken place.²⁸ A multiallelic D' of 0.35 computed between haplotype block 2 and block 3 also indicates that historical recombination has taken place between the different blocks.²⁷ Rs688 and rs5925 are situated within haploblock 2 (Table 4) as identified by Haploview analyses using Gabriel et al. (2002) confidence intervals. Differences in LD among populations are possibly attributable to differences in demographic history²⁹, as the biological determinants of LD (rates of recombination, mutation, gene conversion) are estimated to be constant across groups.²⁸ The above mentioned variants together with rs5930 and rs5927 are synonymous mutations which have an effect on the translational rates of the receptor which may be due to abundance of tRNA or because of the stability of the mRNA.⁴²

The 27.9% difference in LDL-c levels in individuals homozygous for rs14158 is indicative of the role the SNP plays in the expression of the gene, as variants found in the 3'UTR of a gene have been proven to affect the regulation of mRNA.³⁰ LDLR gene expression is controlled mainly by regulatory sequences in the 3'UTR as it regulates changes in mRNA stability at the transcriptional and the post-transcriptional level.^{31,32} Three mRNA destabilising elements, known as AU-rich elements (AREs), responsible for the rapid turnover of LDLR mRNA, have been identified.^{32,33} Variant, rs14158 (c.*52G>A) is present in the 2.5-kb-long 3'UTR of the LDLR gene and, together with rs3826810 (c.*141G>A), rs2738464 (c.*315G>C), rs2738465

(c.*504G>A), rs1433099 (c.*666T>C) and rs2738466 (c.*773A>G), it forms a common haplotype reported by Muallem et al. (2007).³⁴ Rs2738466 was not genotyped in the PURE study population (SNP is too close to rs1433099 to be analysed on the BeadXpress platform) and therefore no comparison could be made between the haplotypes formed within the current study population and those used by Muallem et al. (2007). However, analysis of the five SNPs that were genotyped revealed seven haplotypes, of which only two had significant associations with LDL-c (Table 4). Haplotype GGCGC was associated with lower LDL-c levels with a mean level of 2.69mmol/L and haplotype AGCAC was associated with elevated LDL-c levels with a mean value of 3.71mmol/L ($p<0.01$).

The AA genotype (homozygote mutant) of variant rs2738465 was associated with increased levels of LDL-c ($p=0.05$). The SNP is also found in the 3'UTR that forms part of the above-mentioned haplotype and therefore plays a role in the regulation of mRNA. When the mutant allele is included in a haplotype (AGCAC) it is associated with increased LDL-c levels. Two heterozygotes were detected for the very rare SNP, rs3180023, and, despite the small number, the association with LDL-c did reach statistical significance ($p=0.006$). The variant is also more prevalent in the European population than in the study population (MAF=0.001 versus MAF=0.08). Another rare variant, rs2228671, found in exon 2, is a synonymous mutation where the base change of a cytosine to a thymine does not cause the cysteine at position 27 to change to a different amino acid. Even though the association of the SNP in this population is not statistically significant ($p=0.47$), it has been reported on in various other populations such as a German population where there was a significant association with lower LDL-c and the T allele of the variant. LDL-c was 0.19mmol/L lower for every T allele carried by the German population³⁵. In the PURE population such an association was not present but for the TT genotype, LDL-c was 22% lower than for the CC genotype. Rs11669576 (A391T original amino acid nr A370T) was also detected in this study population but with no significant association with LDL-c as determined in various other studies.³⁶⁻³⁸

The haplotype analysis according to the confidence intervals by Gabriel et al. (2002) revealed four haploblocks.²⁸ Interestingly, the haplotypes from each block with the highest frequency (GCGGC = 0.36; CT = 0.69; GGG = 0.50; GC = 0.22) were associated with lower levels of LDL-c except for block 4 (Table 4). These haplotypes include the alleles associated with lower LDL-c, as can be seen from the first haplotype in Block 1, where the G alleles from rs5930 and rs1569372 have lower mean LDL-c levels. Of the haplotypes in block 4, the GC haplotype is the only haplotype for which the association with lower LDL-c reached statistical significance, where the G allele in rs2738465 has a lower mean LDL-c level. The T and C alleles in rs1433099 have similar mean LDL-c levels.

The only variant that was not in HWE was rs6413503, where there was an excess of homozygote mutants as a result of the genotyping assay and a large number of heterozygotes were not included in the analysis. This is also the only variant that is not in HWE and therefore we do not suspect that it is due to genetic drift, non-random mating, selection or population structure.

There are few studies on the association of rare and common SNPs with LDL-c in an African population and these results give us a greater insight into how genetics play a role in LDL-c at population level in this particular ethnic group. This was, however, a cross-sectional study and functional information for the novel and lesser-known SNPs is not available. Further research into the specific function of each of the SNPs would be of great value in determining to what extent they regulate LDL-c levels. All the SNPs identified by the automated sequencing could not be genotyped within the entire study population because of restriction within the genotyping assay and it would be of great value to genotype these SNPs as well in the study population as 22 of these variants were detected within the 3'UTR. The predictive value of the significantly associated SNPs can be determined at a later stage, however, as the study is prospective with a follow-up of 12 years. It would also be of great value to determine the prevalence and frequency of SNPs in the intron regions of the *LDLR* that were not analysed as current research indicates the importance of variants found within these regions.²⁵

In summary, we have shown that there are SNPs present in the study population that associate with LDL-c levels. For many of these SNPs the prevalence was low and only the individuals harbouring these SNPs had significantly higher or lower LDL-c levels. The study population appears to be healthy with normal mean LDL-c levels and the effect of some variants does not raise LDL-c to pathological levels. Unfortunately the clinical significance of the results is absent because of the observational nature of the study. The SNPs that significantly associated with LDL-c levels were situated mostly within regulatory regions such as the promoter area and the 3'UTR, which affects either transcriptional rate or gene expression (regulatory SNPs). Several of the SNPs which were situated in the coding areas (exons) of the gene were also prevalent in a few individuals and either had an association with elevated LDL-c or had no effect on the levels. In this population it appears, therefore, that the regulatory SNPs make a greater contribution towards LDL-c levels than do SNPs in the coding regions. Interestingly, there were no variants present in exons three, four, five, nine and 17 of the gene. The absence of variants in these exons could be a reason for the low mean LDL-c as well as the lower prevalence of FH in this population. Exons four and nine have been identified as hot spots for mutations causing FH whereas exons 13 and 15 contain a smaller number of mutations and are classified as FH cold spots.³⁹ In agreement with previous evidence³⁹ we found only one SNP in

each of exon 13 and 15. However, the eligibility criteria for the PURE study were that the volunteers should be apparently healthy and this could have had an influence on detecting FH-related variants because individuals harbouring these mutations could have been excluded.

Haplotypes were also identified with significant associations mostly with lower LDL-c, and the low mean LDL-c levels of the study population could possibly be attributed to the presence of these haplotypes in a large percentage of the population (Table 4). This needs to be further investigated, however, to determine whether other confounding factors exist that could result in the low mean LDL-c levels of the study population. Minimal linkage disequilibrium between the various SNPs exists, which is an indication of historical recombination, a characteristic of the African genome.

We also showed that the minor allele frequencies of the identified SNPs in the black South African study population are different from those of the European population, with only a handful of SNPs sharing frequencies across the populations; conversely, the frequencies are very similar to those of other African populations. This indicates that genetic information should be population-specific when used in downstream analysis.

Further research on the other 45 variants identified in the sub-sample of 30 individuals is warranted in order to give us even greater insight into the role that the LDL receptor plays in determining LDL-c levels in this study population.

ACKNOWLEDGEMENTS

We thank the organisers of the South African PURE study, especially Prof. A Kruger and her team, for providing the opportunity to work with the data and samples from the study population. Without their contributions this study would not have been possible. Wayne Towers is thanked for his support with regard to the genetic aspects of the study as well as for obtaining funding for the project. Funding was received from the North-West University and we would like to thank the institution for all of its contributions.

REFERENCES

- [1] Brown, M. S., Kovanen, P. T. & Goldstein, J. L., Regulation of plasma cholesterol by lipoprotein receptors, *Science* **212**, 628-635 (1981).
- [2] Burnett, J. R. & Hooper, A. J., Common and Rare Gene Variants Affecting Plasma LDL Cholesterol, *Clin. Biochem. Rev* **29**, 11-26 (2008).
- [3] Steyn, K., Jooste, P. L., Bourne, L., Fourie, J., Badenhorst, C. J., Bourne, D. E., Langenhoven, M. L., Lombard, C. J. *et al.*, Risk factors for coronary heart disease in the black population of the Cape Peninsula. The BRISK study, *S. Afr. Med. J.* **79**, 480-485 (1991).
- [4] Walker, A. R. P. & Sareli, P., Coronary heart disease: outlook for Africa, *J. R. Sc. Med.* **90**, 23-27 (1997).
- [5] Heller, D. A., de Faire, U., Pedersen, N. L., Dahlen, G. & McClearn, G. E., Genetic and environmental influences on serum lipid levels in twins, *N. Engl. J. Med.* **328**, 1150-1156 (1993).
- [6] Soutar, A. K. & Naoumova, R. P., Mechanisms of disease: genetic causes of familial hypercholesterolemia, *Nat. Clin. Pract. Card.* **4**, 214-225 (2007).
- [7] Scholtz, C., Peeters, A., Hoogendijk, C., Thiart, R., Villiers, J., Hillermann, R., Liu, J., Marais, A. *et al.*, Mutation -59c-->t in repeat 2 of the LDL receptor promoter: reduction in transcriptional activity and possible allelic interaction in a South African family with familial hypercholesterolaemia, *Hum. Mol. Genet.* **8**, 2025-2030 (1999).
- [8] Teo, K., Chow, C. K., Vaz, M., Rangarajan, S., Yusuf, S. & PURE Investigators-Writing Group, 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, *Am. Heart. J.* **158**, 1-7.e1 (2009).
- [9] Friedewald, W. T., Levy, R. I. & Fredrickson, D. S., Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge, *Clin. Chem.* **18**, 499-502 (1972).
- [10] Collins, J. S. & Schwartz, C. E., Detecting polymorphisms and mutations in candidate genes, *Am. J. Hum. Genet.* **71**, 1251 (2002).

- [11] Sanger, F., Nicklen, S. & Coulson, A. R., DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467 (1977).
- [12] Lohman, T. G., Roche, A. F. & Martorell, R. - *Anthropometric standardization reference manual* 1st edn (Human kinetic books, Champaign, Illinois, 1988).
- [13] MacIntyre, U., Venter, C. & Vorster, H., A culture-sensitive quantitative food frequency questionnaire used in an African population: 1. Development and reproducibility, *Public Health Nutr.* **4**, 53 (2001).
- [14] MacIntyre, U., Venter, C. & Vorster, H., A culture-sensitive quantitative food frequency questionnaire used in an African population: 2. Relative validation by 7-day weighed records and biomarkers, *Public Health Nutr.* **4**, 63 (2001).
- [15] Langenhoven, M., Kruger, M., Gouws, E. & Faber, M., *Medical Research Council Food Composition Tables* (1998).
- [16] Soutar, A. K. in *Low-density lipoprotein receptors* (eds Betteridge, D. J., Illingworth, D. R. & Shepherd, J.) 303-322 (Arnold Publishers, London, 1999).
- [17] Cenarro, A., Jensen, H. K., Casao, E., Civeira, F., González-Bonillo, J., Rodríguez-Rey, J. C., Gregersen, N. & Pocovi, M., Identification of recurrent and novel mutations in the LDL receptor gene in Spanish patients with familial hypercholesterolemia, *Hum. Mutat.* **11**, 413-413 (1998).
- [18] Pereira, E., Ferreira, R., Hermelin, B., Thomas, G., Bernard, C., Bertrand, V., Nassiff, H., Del Castillo, D. M. *et al.*, Recurrent and novel LDL receptor gene mutations causing heterozygous familial hypercholesterolemia in La Habana, *Hum. Genet.* **96**, 319-322 (1995).
- [19] Ekström, U., Abrahamson, M., Sveger, T., Lombardi, P. & Nilsson-Ehle, P., An efficient screening procedure detecting six novel mutations in the LDL receptor gene in Swedish children with hypercholesterolemia, *Hum. Genet.* **96**, 147-150 (1995).
- [20] Lu, Y., Feskens, E. J., Boer, J., Imholz, S., Verschuren, W., Wijmenga, C., Vaarhorst, A., Slagboom, E. *et al.*, Exploring genetic determinants of plasma total cholesterol levels and their predictive value in a longitudinal study, *Atherosclerosis* **213**, 200-205 (2010).

- [21] Amsellem, S., Briffaut, D., Carrié, A., Rabès, J., Girardet, J., Fredenrich, A., Moulin, P., Krempf, M. *et al.*, Intronic mutations outside of Alu-repeat-rich domains of the LDL receptor gene are a cause of familial hypercholesterolemia, *Hum. Genet.* **111**, 501-510 (2002).
- [22] Humphries, S. E., Whittall, R. A., Hubbart, C. S., Maplebeck, S., Cooper, J. A., Soutar, A., Naoumova, R., Thompson, G. R. *et al.*, Genetic causes of familial hypercholesterolaemia in patients in the UK: relation to plasma lipid levels and coronary heart disease risk, *J. Med. Genet.* **43**, 943-949 (2006).
- [23] Lelli, N., Garuti, R., Ghisellini, M., Tiozzo, R., Roller, M., Aimale, V., Ginocchio, E., Naselli, A. *et al.*, Occurrence of multiple aberrantly spliced mRNAs of the LDL-receptor gene upon a donor splice site mutation that causes familial hypercholesterolemia (FHBenevento)., *J. Lipid Res.* **36**, 1315-1324 (1995).
- [24] Zakharova, F., Damgaard, D., Mandelshtam, M., Golubkov, V., Nissen, P., Nilsen, G., Stenderup, A., Lipovetsky, B. *et al.*, Familial hypercholesterolemia in St.-Petersburg: the known and novel mutations found in the low density lipoprotein receptor gene in Russia, *BMC Med. Genet.* **6**, 6 (2005).
- [25] Desmet, F., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M. & Bérout, C., Human Splicing Finder: an online bioinformatics tool to predict splicing signals, *Nucleic Acids Res.* **37**, e67-e67 (2009).
- [26] Rafiq, S., Ahmed, N., Soutar, A. & Qamar, R., The genetic characterization of familial hypercholesterolemia in Pakistan, *J. Basic and Appl. Sci.* **7**, 21-25 (2011)
- [27] Bansal, A., Vignal, C. & McGinnis, R. in *Tools for Statistical Genetics* 2nd edn(ed Barnes, M. R.) 217-246 (Wiley, West Sussex, England, 2008).
- [28] Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M. *et al.*, The structure of haplotype blocks in the human genome, *Science* **296**, 2225-2229 (2002).
- [29] Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R. *et al.*, Linkage disequilibrium in the human genome, *Nature* **411**, 199-204 (2001).

- [30] Chen, J., Férec, C. & Cooper, D. N., A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview, *Hum. Genet.* **120**, 1-21 (2006).
- [31] Nakahara, M., Fujii, H., Maloney, P. R., Shimizu, M. & Sato, R., Bile acids enhance low density lipoprotein receptor gene expression via a MAPK cascade-mediated stabilization of mRNA, *J. Biol. Chem.* **277**, 37229-37234 (2002).
- [32] Wilson, G., Vasa, M. & Deeley, R., Stabilization and cytoskeletal-association of LDL receptor mRNA are mediated by distinct domains in its 3' untranslated region, *J. Lipid Res.* **39**, 1025-1032 (1998).
- [33] Knouff, C., Malloy, S., Wilder, J., Altenburg, M. K. & Maeda, N., Doubling expression of the low density lipoprotein receptor by truncation of the 3'-untranslated region sequence ameliorates type iii hyperlipoproteinemia in mice expressing the human apoe2 isoform, *J. Biol. Chem.* **276**, 3856-3862 (2001).
- [34] Muallem, H., North, K. E., Kakoki, M., Wojczynski, M. K., Li, X., Grove, M., Boerwinkle, E., Wilhelmsen, K. C. *et al.*, Quantitative effects of common genetic variations in the 3' UTR of the human LDL-receptor gene and their associations with plasma lipid levels in the Atherosclerosis Risk in Communities study, *Hum. Genet.* **121**, 421-431 (2007).
- [35] Linsel-Nitschke, P., Götz, A., Erdmann, J., Braenne, I., Braund, P., Hengstenberg, C., Stark, K., Fischer, M. *et al.*, Lifelong Reduction of LDL-cholesterol Related to a Common Variant in the LDL-Receptor Gene Decreases the Risk of Coronary Artery Disease--A Mendelian Randomisation Study, *PLoS ONE* **3**, 1-9 (2008).
- [36] Gudnason, V., Patel, D., Sun, X., Humphries, S., Soutar, A. & Knight, B., Effect of the Stul polymorphism in the LDL receptor gene (Ala 370 to Thr) on lipid levels in healthy individuals, *Clin. Genet.* **47**, 68-74 (1995).
- [37] Vieira, J. R. S., Whittall, R. A., Cooper, J. A., Miller, G. J. & Humphries, S. E., The A370T variant (Stul polymorphism) in the LDL receptor gene is not associated with plasma lipid levels or cardiovascular risk in UK men, *Ann. Hum. Genet.* **70**, 697-704 (2006).
- [38] El Messal, M., Chihab, K. A., Chater, R., Vallvé, J. C., Bennis, F., Hafidi, A., Ribalta, J., Varret, M. *et al.*, Familial hypercholesterolemia in Morocco: first report of mutations in the LDL receptor gene, *J. Hum. Genet.* **48**, 199-203 (2003).

- [39] Varret, M., Rabés, J., Thiart, R., Kotze, M. J., Baron, H., Cenarro, A., Descamps, O., Ebhardt, M. *et al.*, LDLR Database: new additions to the database and the software, and results of the first molecular analysis, *Nucleic Acids Res.* **26**, 248-252 (1998).
- [40] Akinboboye, O., Idris, O. & Akinkugbe, O., Trends in coronary artery disease and associated risk factors in sub-Saharan Africans, *J. Hum. Hypertens.* 17, 381 (2003).
- [41] Dolman, R. C., Jerling, J. C., Pieters, M., Ntyintyane, L. & Raal, F., Risk factor profile of coronary artery disease in black South Africans, *SA Heart Journal* 8, 4-11 (2011).
- [42] Sauna, Z. E. & Kimchi-Sarfaty, C., Understanding the contribution of synonymous mutations to human disease, *Nature Reviews Genetics* 12, 683-691 (2011).

CHAPTER 4

THE ASSOCIATION OF COMMON AND RARE VARIANTS IN THE PCSK9 GENE WITH LDL-CHOLESTEROL IN A BLACK SOUTH AFRICAN POPULATION

Tertia van Zyl,^{1,*} Johann C. Jerling,¹ Karin R. Conradie,¹ Edith J.M. Feskens²

¹ Centre of Excellence for Nutrition, North-West University, 2531, Potchefstroom, South Africa; ² Division of Human Nutrition, Wageningen University, 6700, Wageningen, The Netherlands

Submitted to the European Journal of Human Genetics

Tertia Van Zyl - 632-13-EJHG Receipt of New Paper by European Journal of Human Genetics

From: <ejhg@lumc.nl>
To: Tertia.VanZyl@nwu.ac.za
Date: 2013/08/01 03:29 PM
Subject: 632-13-EJHG Receipt of New Paper by European Journal of Human Genetics

Leiden, 1st Aug 2013

Dear Ms Van Zyl,

Thank you for submitting your manuscript to the European Journal of Human Genetics. Please reply to ejhg@lumc.nl in any correspondence regarding this manuscript, quoting the number 632-13-EJHG.

To view the progress of your manuscript please click on the URL below:

[http://mts-ejhg.nature.com/cgi-bin/main.plex?
el=A5Bd3Cbn3A5Wku3J6A9uWKAipV8ajApTdBoSMBBoAZ](http://mts-ejhg.nature.com/cgi-bin/main.plex?el=A5Bd3Cbn3A5Wku3J6A9uWKAipV8ajApTdBoSMBBoAZ)

You can now use a single sign-on for all your Nature Publishing Group accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity. Please check your account regularly and ensure that we have your current contact information.

In addition, NPG encourages all authors and reviewers to associate an Open Researcher and Contributor Identifier (ORCID) to their account. ORCID is a community-based initiative that provides an open, non-proprietary and transparent registry of unique identifiers to help disambiguate research contributions.

Again, thank you for your submission.

Yours sincerely,

Prof. Dr. G-J B van Ommen
Department of Human Genetics
Building 2, Zone S-04
PO Box 9600
2300 RC Leiden
The Netherlands
Tel: (31) (0)71 526 9403
Fax: (31) (0)71 526 8285
Email: EJHG@lumc.nl

Visit us at: www.nature.com/ejhg
Subscribe at: www.nature.com/ejhg/prices.html
Submit at: www.mts-ejhg.nature.com

Abstract

Variants in the proprotein convertase subtilisinlike/kexin type 9 (*PCSK9*) gene are associated with either higher or lower low density lipoprotein cholesterol (LDL-c) levels. Neither the genotypic distribution of these SNPs nor their association with LDL-c levels is known in the black South African population. The aim of this study was to investigate the genotypic distribution of novel and known variants in the *PCSK9* gene and determine their association with LDL-c in a black South African cohort. A cross-sectional study was conducted on 1 840 randomly selected, apparently healthy black South African volunteers aged 35 to 60 years. A subset of 30 individuals was used to identify novel SNPs in the *PCSK9* gene after automated sequencing of the exons and flanking intronic regions. Screening of these variants in the entire cohort was achieved by the use of the Illumina VeraCode™ Golden Gate Genotyping Assay on a BeadXpress™ Reader system. The variants were tested for adherence to Hardy-Weinberg equilibrium after which ANCOVAs were used to determine the association with LDL-c levels. Twenty-seven SNPs were genotyped and three of these were novel. Variants, rs7552350:C>G, rs28362269:C>T and rs499718:T>C were significantly associated with elevated LDL-c whereas rs28362263:G>A and rs28362286:C>A were significantly associated with lower LDL-c levels. Other common variants detected were not significantly associated with LDL-c in the population. These results confirm that rare and common variants in *PCSK9* gene are present in black South African population and are associated with LDL-cholesterol.

Keywords: PCSK9, LDL-c, single nucleotide polymorphisms, African population

Introduction

Elevated low density lipoprotein cholesterol (LDL-c) levels are a major risk factor for the development of cardiovascular disease (CVD) whereas life-long lower levels of LDL-c result in a reduced risk of developing CVD.^{1,2,3} Research has provided sufficient evidence that elevated LDL-c levels is the initiator of atherosclerosis.⁴ LDL-c is a metabolic product of very low density lipoprotein (VLDL) formed in the circulation and is cleared from the circulation by means of LDL-receptor mediated endocytosis in the liver.⁵ Plasma levels of LDL-c are determined by the relative tempo at which LDL-c is produced and cleared from the circulation and this is predominantly mediated by the LDL-receptor.⁶ PCSK9 was identified in 2003 as the third locus for familial hypercholesterolemia (FH)⁷ and because of its involvement in post-translational degradation of the hepatic LDL-receptor⁸ it has been identified as a potential drug target for lowering LDL-c. More than 100 variants have been identified in different populations across the globe^{7, 9-11} and new variants are continuously identified.¹² Gain-of-function variants are

associated with elevated LDL-c and some known variants include S127R^{7, 13}, F216L¹² and N374Y.^{14, 15} Loss-of-function mutations which cause low levels of LDL-c such as the C679X and Y142X variant have been found in individuals of African descent but only in small percentages (2 - 6%). These variants were detected in less than 0.1% of Caucasian populations.¹ Other LOF variants have been identified in Caucasian populations and include R93C¹¹, R46L^{7, 16} and A443T.^{10, 17} Variants Y142X and C679X are associated with a 40% reduction in mean LDL-c levels in the Dallas Heart Study.¹ In the ARIC study the variants lowered LDL-c levels by 1mmol/L (28%) and were associated with an 88% reduction in CHD incidence.¹⁸ The Y142X variant was not found in a study done in Zimbabwean women.¹⁹ Although mostly loss-of-function mutations were found in the African-American populations studied, gain-of-function variants such as H553R have also been discovered.^{10, 14, 15, 20} The genotypic distribution of variants in the *PCSK9* gene is unknown in the black South African population as previous research in a South African population was done on a small number of FH patients.⁹ The aim of the study was to determine the distribution of novel as well as known common and rare variants in the *PCSK9* gene in an apparently healthy black South African cohort and to determine the association of these variants with LDL-c levels.

Methods and Materials

Study population

The baseline data for the South African Prospective Urban and Rural Epidemiological (PURE) study was collected in 2005 over a twelve-week period and was used in this study. The study population consists mainly of self-identified Tswana speaking individuals. Two thousand eligible volunteers older than 35 years were randomly selected and recruitment was done on a door to door basis. Two rural and urban areas in the North West Province of South Africa were used as recruitment sites. From these sites, 1004 volunteers were recruited from Ikageng and Sonderwater extension 7 and 11 (urban areas) and 1004 from Ganyesa and Tklagameng (urban areas). Apparently healthy volunteers were recruited and individuals on chronic medication for non-communicable diseases and/or any self-reported illness were excluded. To avoid bias due to population stratification the urban and rural cohorts were matched for gender and age. More detail on the study design of the PURE study is available in greater detail elsewhere.²¹ The guidelines given by the Declaration of Helsinki were followed throughout the duration of this study. All procedures regarding human volunteers were approved by the Ethics Committee of the North-West University, South Africa (Ethics number: NWU-00016-10-A1). Volunteers were kept anonymous and all data was treated confidentially. All analysis was

performed with coded data. DNA samples for 1860 of the volunteers were available for analysis and successful genotyping of 1530 individuals were achieved for most variants.

Biochemical analysis

Blood was sampled after eight hours of fasting. Serum lipid (total cholesterol, HDL and TG) measurements were taken by means of a Sequential Multiple Analyzer Computer (SMAC) using a Konelab™ auto analyser (Thermo Fischer Scientific, Vantaa, Finland). LDL-c levels were determined by using the Friedewald equation [$\text{LDL-c} = \text{TC} - \text{HDL-C} - (\text{TG})/2.17$] and missing values were assigned to individuals with a TG measurement of $>4.5\text{mmol/L}$.

DNA analysis

Informed consent for the collection of blood samples was obtained from each participant in the PURE study. Deoxyribonucleic acid (DNA) isolation was performed by using Qiagen® Flexigene™ DNA extraction kits (QAIKEN, Valencia, CA, USA, catalogue number 51206).

A sub-sample of 30 randomly chosen individuals was used to identify known and novel mutations in the promoter region, exons and exon-intron boundaries of the *PCSK9* gene. Thirty individuals would give a total sample size of 60 chromosomes and as communicated by Collins and Schwartz using 40 to 65 chromosomes will detect polymorphisms at a frequency of 0.05 with a power of 0.80 to 0.95.²² Polymerase chain reaction (PCR) amplification of the proximal promoter region, the exons and exon-intron boundaries of the *PCSK9* gene was performed on a BioRad iCycler (Hercules, CA, USA). The primers designed for amplification and sequencing of the various regions are described in Annexure I. NucleoFast® 96 PCR Clean-Up Kits from Machery-Nagel (Düren, Germany) were used for the clean-up of the PCR product prior to performing the sequencing process. Automated cycle sequencing was performed according to the basic principles of Sanger sequencing.²³ The standard ABI Prism®, BIGDye® Terminator version 3 Ready Reaction Cycle Sequencing Kit supplied by Applied Biosystems (Carlsbad, CA, USA) was used as a platform for the sequencing reactions. The purified product was subjected to electrophoresis on a 50cm capillary on either an ABI3730xl DNA analyser or an ABI3120xl genetic analyser according to standard protocols.

Genotyping

The sequences that were generated were examined for the presence of variants using the CLC DNA Workbench version 6.0.2 (Aarhus, Denmark). Variants were confirmed by repeat sequencing in the opposite direction. Seventy one variants were identified within the coding as well as non-coding areas of the *PCSK9* gene which includes two insertions (rs45454392:GCT ins, a 1bp insertion not previously reported in the 3'UTR: c.*448_*449insG), two deletions (rs45597632:-C and rs57780528:-CGGAGGG) and 67 single nucleotide polymorphisms (these are available in Annexure H). Seven of the SNPs were novel and not previously reported. Genotyping of the population was done by means of Illumina's VeraCode™ Golden Gate Genotyping (GGGT) Assay on a BeadXpress™ platform using the VeraCode™ technology. All 67 SNPs (rs numbers for known SNPs and sequences for novel SNPs) were sent through to Illumina Technical Support for evaluation using the Assay Design Tool. Each SNP was scored (varying from 0 to 1) by the Assay Design Tool based on compatibility with successful GGGT. SNPs with a score above 0.4 were chosen for genotyping. Four SNPs had a score below 0.4. Twenty-one SNPs fell out for selection because they were located less than 60bp from another SNP (this is a prerequisite for the GGGT assay design). Thirty-one SNPs were then selected on the basis of their frequency (>0.02) in the sub sample, novelty and whether it was a loss-of-function or gain-of-function mutation. A further five SNPs (rs151193009: C>T rs113138552:G>A, rs72646516:G>T, rs479910:A>G and rs141502002:C>T) failed during the execution of the GGGT assay. The study population was finally genotyped for 25 SNPs by means of GGGT (of which three were novel and not previously reported). Genotype calls for one of the novel SNPs (c.997-22A>C) were very poor (75% call frequency) and this SNP was also omitted from the statistical analysis and associations. SNPs with a gene call score of < 0.5 and samples with a call rate of ≤0.9 were included. The reliability of these filters was tested by comparing sequenced data with the GGGT assay data. The A of the initiating ATG codon is designated as nucleotide c.1 and the reference sequence *PCSK9*: NM_174936.3 was used. Two variants (rs28362263:G>A and rs28362286:C>A) were analysed by using Taqman SNP Genotyping Assays by Applied Biosystems (Carlsbad, CA, USA) (rs28362286:C>A - Assay ID: C_63745519_10) (rs28362263:G>A - Assay ID: C_25934978_20). Prediction software SIFT²⁴ (Sorting tolerant from intolerant) and Polyphen-2²⁵ were used to determine whether the novel SNPs identified in the coding regions of the *PCSK9* gene would have an effect on the protein function. These variants have been submitted to dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>).

Haplotype frequency

Haplotype frequencies were determined by means of the software package Haploview version 4.2 (Broad Institute, Cambridge, Massachusetts, USA) and the confidence intervals as described by Gabriel et al.²⁶ were used to form haploblocks.²⁷ To determine the haploblocks markers with a Hardy-Weinberg p-value of ≥ 0.05 and a minor allele frequency (MAF) of $\geq 1\%$ were included in the analysis (the LD plot is available in Annexure J and the identified haplotypes are available in Annexure K). Association analysis was done by using LDL-c of $<3\text{mmol/L}$ for controls and $>3\text{mmol/L}$ for cases.

Measurements

The anthropometric measurements such as weight (kilograms), height (metres) and body mass index ($\text{BMI} = \text{weight in kilograms}/\text{height}^2$ in metres) were supervised by a level 3 anthropometrist according to guidelines adopted at the Arlie Conference sponsored by the National Institutes of Health. Standardised and calibrated equipment was used for measurements on volunteers wearing minimal clothing.²⁸

Statistical analysis

Descriptive statistics on the study population were calculated on the statistical program SPSS version 20 and are expressed in means and standard deviations. The observed genotype frequencies were tested for deviation from Hardy-Weinberg equilibrium by using the Chi-square test of independence (χ^2). ANOVA was used to test association between different genotypes and LDL-c levels. Where significant associations were found ANCOVA was used to adjust for the confounders: age, gender and body mass index (BMI) as LDL-c levels correlate with these variables. The Bonferroni correction was used to deal with issue of multiple testing. P-values of 0.05 or lower were considered statistically significant.

Results

The characteristics of the study population are shown in Table 1. The mean LDL-c levels for the men were 2.70mmol/L (SD 1.12) and for the women 3.06mmol/L (SD 1.17). The mean age for both men and women was similar at 49.9 yrs. (SD 10.4) and 49.1 yrs. (SD 10.4), respectively.

Table1 Population characteristics

Variables	Men		Women		T-test
	n	Mean (SD)	n	Mean (SD)	P-value
Age (years)	696	49.9 (10.4)	1168	49.1 (10.4)	0.161
Total Cholesterol (mmol/L)	696	4.79 (1.31)	1168	5.13 (1.39)	<0.001
LDL-c ^a (mmol/L)	696	2.70 (1.12)	1168	3.06 (1.17)	<0.001
HDL-C ^b (mmol/L)	696	1.58 (0.65)	1168	1.48 (0.62)	0.002
TG ^c (mmol/L)	696	1.15 (0.58)	1168	1.33 (0.70)	<0.001
BMI ^d (kg/m ²)	696	20.8 (4.04)	1168	26.9 (7.32)	<0.001
Rural/Urban	348/399		657/603		
LDL-c (<3mmol/L / >3mmol/L)	453/243		618/550		

^a Low density lipoprotein cholesterol, ^b high density lipoprotein cholesterol, ^c triglycerides, ^dbody mass index

Of the 71 variants (available in Annexure H) identified in the *PCSK9* gene by automated, bidirectional sequencing 27 were successfully screened in the study population. Of these 27 SNPs, three are novel variants which have not previously been reported and there were no significant association between these novel variants and LDL-c. Five of the 27 variants were significantly associated with LDL-c levels after adjusting for confounders and multiple testing. Two variants (rs28362263:G>A and rs28362286:C>A) were associated with lower LDL-c and three (rs7552350:C>G, rs28362269:C>T and rs499718:T>C) with elevated levels of LDL-c. All the variants were in HWE except for rs7552350:C>G. The MAF of the variants are reported in Table 2. The African and European populations from the 1000 Genomes Project were used for comparing MAF with the study population. Eight rare variants were identified in the study population and are also present in the African population but not in the European population. The MAF of all other variants differed between the European population and the study population except for variant, rs2483205, which have a MAF that is similar across the three populations. Table 2 also contains the results on the frequency of the variants identified in the study population as well as the association with LDL-c.

Table 2 Information on PCSK9 variants found in the PURE population

SNP number	Variant information	Frequency	LDL-c (mmol/L)	ANOVA /ANCOVA ^a	Bonferroni p-value ^b	MAF	Novel Yes/No	Previous studies	Similar results
rs28362202	c.-26 G>A	n=1530					No	Reported in Yoroba population from 1000 Genomes Project	No association results with LDL-c available
GG	5'UTR	1305	2.93 (2.87-2.99)	0.139	0.597 (GG vs GA)	^c P: 0.08			
GA	p (HWE) = 0.80	218	3.03 (2.88-3.18)	0.260	0.670 (GA vs AA)	^d A: 0.02			
AA		7	2.51 (1.68-3.33)		0.973 (AA vs GG)	^e E: 0.00			
rs2495480	c.208-161 C>T	n=1530					No	Abifadel et al., 2003	Yes
TT	Intron 1	1053	2.91 (2.84-2.97)	0.119	0.479 (TT vs TC)	P: 0.17		Shioji et al., 2004	
TC	p (HWE) = 0.71	438	2.99 (2.89-3.10)	0.064	0.411 (CC vs TC)	A: 0.20			
CC		39	3.27 (2.92-3.67)		0.132 (TT vs CC)	E: 0.06			
rs2495482	c.207+15 A>G	n=1523					No	Leren et al., 2004	No association results with LDL-c available
GG	Intron 1	1154	2.92 (2.87-2.94)	0.157	0.851 (GG vs AG)	P: 0.13		Humphries et al., 2006	
AG	p (HWE) = 0.77	347	2.99 (2.89-3.11)	0.244	0.906 (AG vs AA)	A: 0.18		Huijgen et al., 2012	
AA		22	3.25 (2.78-3.71)		0.523 (AA vs GG)	E: 0.06			
rs200109442	c.400-24 C>T	n=1534					No	Reported in African Americans in 1000 Genomes project	No association results with LDL-c available
CC	Intron 2	1483	2.95 (2.89-3.01)	0.239		P: 0.02			
CT	p (HWE) = 0.80	51	2.70 (2.39-3.00)	0.116	0.116	A: NA			
TT		0	-			E: 0.001			
rs4927193	c.399+165 T>C	n=1535					No	Shioji et al., 2004	No association results with LDL-c available
TT	Intron 2	709	2.94 (2.86-3.03)	0.383	1.00 (TT vs TC)	P: 0.33		Huijgen et al., 2012	
TC	p (HWE) = 0.24	647	2.97 (2.86-3.06)	0.287	0.344 (TC vs CC)	A: 0.20		Postmus et al., 2013	
CC		179	2.82 (2.66-2.99)		0.586 (TT vs CC)	E: 0.13			
rs624612	c.524-68 C>G	n=1537					No	Abifadel et al., 2003	No association results with LDL-c available
CC	Intron 3	1351	2.95 (2.89-3.01)	0.412		P: 0.06		Blesa et al., 2008	
GC	p (HWE) = 0.99	180	2.87 (2.71-3.03)	0.65	1.00	A: 0.07			
GG		6	3.12 (2.23-4.01)			E: 0.41			

SNP number	Variant information	Frequency	LDL-c (mmol/L)	ANOVA /ANCOVA ^a	Bonferroni p-value ^b	MAF	Novel Yes/No	Previous studies	Similar results
rs499718	c.523+230 T>C	n=1533					No	Ding and Kullo et al., 2008	No association results with LDL-c available
TT	Intron 3	369	2.88 (2.75-2.96)	0.030	0.129 (TT VS TC)	P: 0.51		Postmus et al., 2013	
TC	p (HWE) = 0.95	760	3.02 (2.94-3.10)	0.024	0.047 (TC vs CC)	A: 0.46			
CC		404	2.88 (2.76-2.99)		1.00 (TT vs CC)	E: 0.17			
rs2483205	c.658-7 C>T	n=1534					No	Abifadel et al., 2003	No association results with LDL-c available
CC	Intron 4	421	2.85 (2.74-2.95)	0.159	0.359 (CC vs CT)	P: 0.47		Leren et al., 2004	
CT	p (HWE) = 0.60	784	2.95 (2.88-3.03)	0.072	0.814 (CT vs TT)	A: 0.49		Humphries et al., 2007	
TT		329	3.03 (2.91-3.15)		0.072 (CC vs TT)	E: 0.48		Miyake et al., 2008 Postmus et al., 2013	
rs7552350	c.657+114 C>A	n=1535					No	Huijgen et al., 2012	No association results with LDL-c available
CC	Intron 4	867	2.91 (2.83-2.98)	0.018	1.00 (CC vs CA)	P: 0.26			
CA	p (HWE) = 0.02	545	2.9 (2.81-3.00)	0.001	0.001 (CA vs AA)	A: NA			
AA		123	3.31 (3.12-3.51)		0.001 (CC vs AA)	E: NA			
rs2495477	c.799+3 A>G	n=1617					No	Abifadel et al., 2003	No association results with LDL-c available
AA	Intron 5	121	2.89 (2.69-3.08)	0.597		P: 0.27		Leren et al., 2004	
AG	p (HWE) = 0.36	580	2.94 (2.84-3.03)	0.828	1.00	A: 0.28		Miyake et al., 2008	
GG		835	2.95 (2.88-3.03)			E: 0.59		Humphries et al., 2006 Huijgen et al., 2012	
rs494198	c.799+64 C>A	n=1505					No	Abifadel et al., 2003	No association results with LDL-c available
CC	Intron 5	275	2.87 (2.74-3.00)	0.301	1.00 (CC vs CA)	P: 0.58		Leren et al., 2004	
CA	p (HWE) = 0.22	698	2.94 (2.85-3.02)	0.443	1.00 (CA vs AA)	A: 0.51			
AA		532	2.98 (2.88-3.07)		0.610 (CC vs AA)	E: 0.37			
rs28362256	c.996+379 C>T	n=1537					No	Reported in Bushman, Yoroba and African American populations available on dbSNP	No association results with LDL-c available
CC	Intron 6	1336	2.93 (2.87-2.99)	0.304		P: 0.07			
CT	p (HWE) = 0.92	195	3.01 (2.85-3.16)	0.65	1.00	A: 0.04			
TT		6	3.00 (2.11-3.89)			E: 0.00			

SNP number	Variant information	Frequency	LDL-c (mmol/L)	ANOVA /ANCOVA ^a	Bonferroni p-value ^b	MAF	Novel Yes/No	Previous studies	Similar results
rs374941781	c.829 G>C (V277L)	n=1534					Yes		
GG	Exon 6	6	2.65 (1.75-3.54)	0.426	1.00 (GG vs GC)	P: 0.07			
GC	p (HWE) = 0.96	191	2.86 (2.70-3.02)	0.463	0.859 (GC vs CC)	A: NA		-	-
CC		1337	2.95 (2.90-3.01)		1.00 (GG vs CC)	E: NA			
rs533375	c.1180+174 A>G	n=1537					No	Ding and Kullo et al., 2008	No association results with LDL-c available
AA	Intron 7	803	2.93 (2.66-3.03)	0.231	1.00 (AA vs AG)	P: 0.29			
AG	p (HWE) = 0.31	599	2.98 (2.89-3.07)	0.420	0.654 (AG vs GG)	A: 0.25			
GG		135	2.85 (2.85-3.01)		1.00 (GG vs AA)	E: 0.18			
rs584626	c.1354+102 C>T	n=1536					No	Abifadel et al., 2003 Shioji et al., 2004	No association results with LDL-c available
CC	Intron 8	689	2.95 (2.87-3.04)	0.988		P: 0.34			
CT	p (HWE) = 0.84	672	2.93 (2.83-3.00)	0.695	1.00	A: 0.23			
TT		175	2.97 (2.81-3.14)			E: 0.18			
rs585131	c.1355-56 C>T	n=1536					No	Abifadel et al., 2003 Huijgen et al., 2012	No association results with LDL-c available
CC	Intron 8	845	2.93 (2.86-3.01)	0.333		P: 0.26			
CT	p (HWE) = 0.59	578	2.97 (2.88-3.06)	0.652	1.00	A: 0.21			
TT		113	2.86 (2.66-3.07)			E: 0.18			
rs28362261	c.1274 A>G (N425S)	n=1537					No	Pisciotta et al., 2006 Kotowski et al., 2006 Wu et al., 2013	Yes
AA	Exon 8	1504	2.94 (2.88-3.00)	0.770	0.986	P: 0.01			
AG	p (HWE) = 0.91	33	2.94 (2.55-3.32)	0.986		A: 0.03			
GG		0	-			E: 0.00			
rs28362263	c.1327G>A (A443T)						No	Haung et al., 2009 Wu et al., 2013 Coram et al., 2013	Yes
GG	Exon 8	1514	2.94 (2.89-2.99)	0.078	0.043 (GG vs GA)	P: 0.04			
GA	p (HWE) = 0.50	114	2.68 (2.48-2.89)	0.047	1.00 (GA vs AA)	A: 0.09			
AA		4	2.77 (1.68-3.87)		1.00 (GG vs AA)	E: 0.00			

SNP number	Variant information	Frequency	LDL-c (mmol/L)	ANOVA /ANCOVA ^a	Bonferroni p-value ^b	MAF	Novel Yes/No	Previous studies	Similar results
rs28362269	c.1504-16 C>T	n=1533					No	Lye et al., 2013	Yes, but loose significance after Bonferroni correction in PURE population
CC	Intron 9	1374	2.92 (2.86-2.98)	0.018	0.601 (CC vs CT)	P: 0.05			
CT	p (HWE) = 0.95	154	3.01 (2.87-3.22)	0.029	0.110 (CT vs TT)	A: 0.06			
TT		5	4.11 (3.14-5.08)		0.055 (CC vs TT)	E: 0.00			
rs28362270	c.1658 A>G (H553R)	n=1533					No	Kotowski et al., 2006	In DHS association with increased LDL-c, No association in Cook County (Kotowski et al., 2006)
AA	Exon 10	1425	2.92 (2.86-2.98)	0.043	0.094 (AA vs AG)	P: 0.04			
AG	p (HWE) = 0.99	106	3.16 (2.95-3.38)	0.087	1.00 (AG vs GG)	A: 0.01			
GG		2	2.55 (1.03-4.11)		1.00 (GG vs AA)	E: 0.00			
rs483462	c.1681+64 G>A	n=1533					No	Abifadel et al., 2003 Leren et al., 2004	No association results with LDL-c available
GG	Intron 10	892	2.98 (2.52-3.03)	0.113	1.00 (GG vs GA)	P: 0.23			
GA	p (HWE) = 0.35	568	2.90 (2.81-2.99)	0.170	0.541 (GA vs AA)	A: 0.39			
AA		73	2.77 (2.91-3.05)		0.375 (AA vs GG)	E: 0.69			
rs111563724	c.1696 T>C (W566R)	n=1537					No	Found in a Bushman from 1000 Genomes Project	No association results with LDL-c available
TT	Exon 11	1438	2.94 (2.88-3.00)	0.639		P: 0.03			
TC	p (HWE) = 0.58	96	2.98 (2.75-3.20)	0.707	1.00	A: NA			
CC		3	2.45 (1.18-3.71)			E: NA			
rs372506466	c.1781 C>A (A594D)	n=1522					Yes		
CC	Exon 11	1382	2.97 (2.91-3.02)	0.072	0.097 (CC vs CA)	P: 0.05			
CA	p (HWE) = 0.37	134	2.75 (2.56-2.94)	0.081	1.00 (CA vs AA)	A: NA		-	-
AA		6	2.64 (1.74-3.52)		1.00 (AA vs CC)	E: NA			
rs28362286	c.2037C>A(C679X)	n=1428					No	Cohen et al., 2005 Hallman et al., 2007	Yes
CC	Exon 12	1350	2.99 (2.93-3.05)	<0.001		P: 0.03			
CA	p (HWE) = 0.57	78	2.17 (1.92-2.41)	<0.001	<0.001	A: 0.01			
AA		0	-			E: 0.00			

SNP number	Variant information	Frequency	LDL-c (mmol/L)	ANOVA /ANCOVA ^a	Bonferroni p-value ^b	MAF	Novel Yes/No	Previous studies	Similar results
rs505151	c.2009 G>A	n=1505					No	Naureckiene et al., 2003	Yes
AA	Exon 12	737	2.93 (2.85-3.01)	0.717		P: 0.30		Abifadel et al., 2003	
AG	p (HWE) = 1.00	633	2.96 (2.87-3.04)	0.756	1.00	A: 0.29		Chen et al., 2005	
GG		135	3.00 (2.81-3.19)			E: 0.03		Humphries et al., 2006 Kotowski et al., 2006	
rs662145	c.*571C>T	n=1534					No	Postmus et al., 2013	No association results with LDL-c available
CC	3'UTR	795	2.99 (2.75-3.18)	0.184	0.182 (CC vs CT)	P: 0.27			
CT	p (HWE) = 0.47	632	2.88 (2.79-2.97)	0.168	1.00 (CT vs TT)	A: 0.42			
TT		107	2.97 (2.91-3.07)		1.00 (TT vs CC)	E: 0.75			
rs369066144	c.*500 C>T	n=1536					Yes		
CC	3'UTR	1431	2.95 (2.89-3.01)	0.250	1.00 (CC vs CT)	P: 0.04			
CT	p (HWE) = 0.30	101	2.85 (2.63-3.07)	0.278	0.718 (CT vs TT)	A: NA		-	-
TT		4	2.18 (1.09-3.28)		0.511 (TT vs CC)	E: NA			

^a ANOVA p-value reported first then ANCOVA p-value, ^b Bonferroni correction for ANCOVA is reported, ^c MAF of the study population, ^d MAF of the African population from the 1000 Genomes Project (<http://browser.1000genomes.org>), ^e MAF of the European population from the 1000 Genomes Project, LDL-c, low density lipoprotein cholesterol; MAF, minor allele frequency; p (HWE), Hardy-Weinberg p-value; 3'UTR, 3 untranslated region

Haplotype analysis**Table 3** Haplotypes associated with LDL-c within the PURE population

Haplotype	Frequency	Case, control frequencies	P-value
Block 1 - rs2495480; rs4927193			
TT	0.51	0.49, 0.51	0.18
TC	0.32	0.33, 0.33	0.78
CT	0.17	0.18, 0.16	0.15
Block 2 - rs249548; rs4927193			
AC	0.52	0.54, 0.50	0.03
CC	0.42	0.40, 0.43	0.05
AG	0.06	0.06, 0.07	0.61
Block 3 – rs584626; rs585131			
TT	0.67	0.67, 0.67	0.89
CC	0.26	0.26, 0.26	0.97
CT	0.07	0.07, 0.07	0.75
Block 4 - rs28362269; rs483462; rs505151; rs662145			
CGAC	0.43	0.42, 0.43	0.78
CGGC	0.25	0.25, 0.24	0.76
CAAT	0.23	0.21, 0.24	0.06
TGGC	0.05	0.07, 0.50	0.02
CGAT	0.04	0.05, 0.04	0.18

From the haplotype analysis four blocks were identified. The association analysis revealed significant association in block 2 and 4 (Table 3). In block 2 haplotype AC ($p=0.03$, frequency of the haplotype = 0.52) was associated with increased levels of LDL-c as more cases than controls exhibited this haplotype. Haplotype CC ($p=0.05$, frequency of the haplotype = 0.42) however was associated with lower LDL-c levels, more controls than cases with this haplotype were observed. The haplotype TGGC in block 4 was associated ($p=0.02$, frequency of the haplotype = 0.05) with increased LDL-c levels as more cases than controls presented with this haplotype.

Discussion

In this study we sequenced the promoter region, exons and exon-intron boundaries of the *PCSK9* gene to determine the presence of known and novel variants. Of the 71 variants detected 27 SNPs were genotyped and their association with LDL-c was determined. Three of these SNPs have not previously been reported in the literature and are considered to be novel.

Comparison with European and African populations of minor allele frequencies of the different variants found in the study population revealed that the frequencies are more comparable with the African population than with the European. Variant, rs2483205:C>T, is the only common variant with similar MAF across the three populations. Some of the very rare variants were also only present in the study and African populations from the 1000 Genomes project (Table 2). These differences illustrate how important it is not to infer genetic results from one population to another and to have a clear description of the study population you work with. Rare variants are also more prevalent in the African population.

Novel SNP rs374941781 (c.829 G>C; V277L) was found in exon 6, which forms part of the catalytic domain of the protein. The G>C substitution at base 829 causes a missense mutation and an amino acid change from a valine (GTC) to a leucine (CTC). A study by Cameron et. al.¹³ identified R272Q as a novel variant and determined that the mutant PCSK9 from this variant was secreted normally and in amounts comparable with that of the wild type PCSK9. Internalisation of LDL-receptor by this variant was also affected in the same manner as the wild type PCSK9. Therefore, the variant may be a benign polymorphism with no effect on PCSK9 function. In variant V277L, valine and leucine are both hydrophobic amino acids and it is predicted that the substitution of one hydrophobic amino acid with another is normally tolerated.²⁹ A SIFT score of 0.36 also indicates that this variant is tolerated²⁴ and this prediction was confirmed as being benign by a Polyphen-2²⁵ score of 0.005.

Novel SNP rs372506466 (A594D) was found in exon 11 and this forms part of the C-terminal domain. A594D is a missense mutation caused by a C>A substitution at base 1781 and this results in an amino acid change from an alanine (GCC) to an aspartic acid (GAC). The SIFT prediction for this variant was that it would affect protein function (score=0.02)¹⁹, which was also confirmed by a Polyphen-2 prediction of being possibly damaging (score=0.66).²⁵ This is also a rare variant found in only 10% of the population, with a 5% difference in LDL-c per allele. The variant is found in the C-terminal domain, as with various variants found by Miyake et al.¹¹ which were associated with lower LDL-c. Another variant identified by this group was A239D, which was found in a hypocholesterolemic individual. This variant also has an alanine replaced with an aspartic acid. This specific amino acid change could also have an effect on the protein function. Loss-of-function variant C679X is also found in the C-terminal domain and is associated with lower LDL-c levels. The C-terminal domain binds to the ligand-binding domain of the LDL-receptor³⁰ and the possible effect that variant A594D has on the function of this domain could cause the association with lower mean LDL-c levels.

The study population was not genotyped for the loss-of-function, nonsense mutation Y142X (rs67608943:C>G) as a study done on 653 women from Zimbabwe did not reveal any individuals harbouring this variant.¹⁹ The results of the study on the Zimbabwean women were also a motivation for not including the SNP in the analysis after it was not detected in our sub-sample of 30 individuals.

An association between rs2495480:C>T and LDL-c levels has previously been reported by Shioji et al.³¹ in a Japanese population. The association was not statistically significant but the CT/TT genotypes had lower mean LDL-c than the wild type (CC genotype). In the PURE population the association was also not statistically significant and the CC genotype presented with higher mean LDL-c levels (3.27 mmol/L). Although the association was not statistically significant increasing LDL-c levels above 3mmol/L will have a significant effect on the development of atherosclerosis. There was also no significant association between rs2495482:A>G and LDL-c levels even though the AA genotype had increased LDL-c levels (3.25mmol/L). This SNP has previously been reported in a study³² but there were no results on the association with LDL-c levels. Heterozygote carriers of another intronic variant, rs200109442:C>T, presented with LDL-c levels which were 0.25mmol/L lower than those of non-carriers of the variant, whereas with variant rs2483205:C>T the carriers of the ancestral allele C presented with a LDL-c level which were 0.15mmol/L lower than those of non-carriers. These intronic variants are possibly located close to splice site junctions that have an effect on the transcription of the mRNA and therefore have an effect on the PCSK9 protein produced or may be in linkage disequilibrium with other variants that cause the association. Functional studies on these variants are recommended for a better understanding of their effects.

Variant rs7552350 (c.657+114C>A) was detected in the intronic region between exon 4 and 5. Only the homozygous variant (AA genotype) was significantly associated with higher LDL-c levels. The effect of this variant on the protein function is currently unknown. Two other variants (c.657+76C>A and c.657+82A>G) identified by Abifadel et al.³³ are located 38 and 32 base pairs respectively from this variant and were described as common polymorphisms with no effect on protein function. The splice site prediction algorithm from the Human Splicing Finder³⁴ predicted that there would be no effect on mRNA splicing for these two variants as reported by Huijgen et al.³⁵ The departure from HWE in rs7552350 could possibly be caused by heterozygote deficiency and the excess of homozygote individuals due to the performance of the genotyping assay. Caution should thus be exercised in interpreting these results.

The rare variant A443T (rs28362263:G>A) found in the catalytic domain was also detected in the study population. Mean LDL-c levels for carriers of the variant were 8.8% lower than that of non-carriers. This variant is a missense mutation where a nucleotide change (guanine to adenine) at position 1327 on the gene causes an amino acid change at position 443 on the protein from an alanine (GCC) to a threonine (ACC). Various studies have also reported on the association of A443T with lower LDL-c but none reported on the how the variant effects protein function.^{10, 17, 36} Zhao et al. did however determine that the variant has no effect on how the protein is processed or secreted.³⁷ The variant might be more sensitive to furin cleavage which affects the stability of the protein.³⁸

Another rare variant present in only the African populations, namely rs28362269, also found in the intronic region, was significantly associated with increased LDL-c levels; homozygote individuals harbouring this variant had mean LDL-c levels 1.19mmol/L higher than those of non-carriers of the variant. According to information from dbSNP this variant is non-pathogenic but no studies on the function of the variant could be found.

The gain-of-function variant H553R (rs28362270) was not significantly associated with LDL-c levels and the two individuals with the homozygous genotype (GG) had very low LDL-c levels. The low levels in these individuals suggest that another variant may play a causative role as the physical attributes of the two individuals differ dramatically (very low versus a high BMI). The mean LDL-c level of the AG genotype, however was 8.2% higher than the wild type (AA). Kotowski et al. also reported on this variant, it had a MAF of 0.013 in the black participants of the Dallas Heart Study and there was a significant association with elevated LDL-c levels but no association in the Cook County sample.¹⁰

The rare loss-of-function variant, rs28362286 (C679X), was present at a minor allele frequency of 0.03 and carriers of the variant had mean LDL-c levels 27% lower than those of non-carriers. This variant has previously been reported by various researchers with results that are similar to what we have found.^{1, 19, 39} The cystein at residue 679 is replaced with a stop codon and results in a protein which fails to fold properly and this protein does not exit the endoplasmic reticulum (ER) and it is not secreted.³⁷

The amino acid change at position 566 on the protein from tryptophan (TGG) to arginine (CGG) of rs111563724 is caused by a T>C substitution at base 1696. This is another missense mutation but there is no association with LDL-c. A SIFT prediction of this substitution states that it is tolerated with a score of 0.46²⁴ and the Polyphen-2 score of 0.001 also confirmed a benign affect.²⁵

Polymorphism rs505151 (E670G), a non-synonymous missense mutation, has also been previously reported in various studies.^{10, 11, 32, 36, 40-42} In most of these studies no association was reported or none was found. The SNP is located in the cysteine-rich C-terminal domain, which appears to be involved in the auto-regulation process as the deletion of this domain leads to the accumulation of processed PCSK9 in the ER. The protein is thus not secreted from the ER and is unable to proceed with its normal function. *In silico* studies determined that an energy difference caused by the polymorphism does not alter the proper folding of the domains. The protein structure and flexibility are affected, however and this could influence the pKa balance, which in turn could have an effect on the affinity of the protein for its substrates, such as the LDLR.⁴³ This variant has been reported to form part of a haplotype that associates with LDL-c.⁴⁰ Haplotype analysis of this study population furthermore identified that rs505151 was part of a haplotype block (TGGC) that significantly associated ($p=0.02$) with elevated LDL-c levels, with a frequency of 0.05 (Table 3).

Haplotype analysis

Four haplotype blocks were identified with various lengths as seen in Table 3. From the association analysis three haplotypes were identified as having an association with LDL-c. The first haplotype in block 2 (AC) is associated with lower LDL-c as both these alleles in rs2495480 and rs4927193 have lower mean LDL-c levels than the T alleles. The second haplotype in block 2 (CC) is associated with higher LDL-c levels as these alleles in rs494198 and c.829 G>C (V277L) have higher mean LDL-c levels than their counterparts. In block 4 a haplotype TGGC was identified which is associated with higher mean LDL-c levels as most of the alleles have a mean LDL-c level above 2.88mmol/L.

General conclusion

This study is one of few studies on genetics and LDL-c in an African population. Even though the candidate gene approach was taken to identify variants that associate with LDL-c levels it gives more insight into the genetic diversity of an African population and how known and novel variants detected in the study population associate with LDL-c levels. Unfortunately, there is no functional information available for the novel variants and future research on these is recommended. The study population could also not be genotyped for all the variants identified by sequencing as the genotyping assay had restrictions when it came to the analysis. It would be of great value to genotype the study population for the other variants to gain even greater insight into the function of PCSK9 in determining LDL-c levels. Being a cross-sectional study has its limitations although it is

also prospective, and the results of follow-up values will enable us to determine the longitudinal effects of the variants.

From these results it is clear that LOF and GOF variants in the PCSK9 gene are present in the study populations. These variants are rare, however and do not seem to have a significant impact on the LDL-c levels of the entire population; only the few affected individuals with the variants rs28362286 and rs28362263 benefit from the lower LDL-c levels. Only two of the common polymorphism detected in this study had a significant association with LDL-c and the one, rs7552350, was not in HWE therefore it can be concluded from this study that it seems that variants in the PCSK9 gene do not affect LDL-c at population level. The genetic diversity between the European and African populations are also confirmed with the difference in MAF between these populations. As the PURE study only included apparently healthy individuals the detection of GOF variants might be underrepresented as this criterion could have excluded individuals with very high LDL-c levels.

Conflict of interest

The authors declare no conflict of interest.

Supplemental information

Supplemental information is available at the website of The European Journal of Human Genetics.

Acknowledgements

We thank the organizers of the South African PURE study, especially Prof. A Kruger and her team for providing the opportunity to work with the data and samples from the study population. Without their contributions this study would not have been possible. Wayne Towers is thanked for his support on the genetic aspects of the study as well as obtaining funding for the project. Funding was received from the North-West University and we would like to thank the institution for all of their contributions.

References

- 1 Cohen J, Pertsemlidis A, Kotowski I, Graham R, Garcia C, Hobbs HH: Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* 2005; **37**: 161-165.
- 2 Rudel LL, Parks JS, Johnson FL, Babiak J: Low density lipoproteins in atherosclerosis. *Journal of Lipid Research* 1986; **27**: 465-474.
- 3 Ference BA, Yoo W, Alesh I *et al*: Effect of Long-Term Exposure to Lower Low-Density Lipoprotein Cholesterol Beginning Early in Life on the Risk of Coronary Heart DiseaseA Mendelian Randomization Analysis. *J Am Coll Cardiol* 2012; **60**: 2631-2639.
- 4 Steinberg D: Atherogenesis in perspective: Hypercholesterolemia and inflammation as partners in crime. *Nature Medicine* 2002; **8**: 1211-1217.
- 5 Goldstein JL, Hobbs H, Brown MS: Familial hypercholesterolemia; in Scriver C, Beaudet A, Sly W, Valle D (eds): *The Metabolic and Molecular Bases of Inherited Disease*. Volume II. New York, 2001, pp 2863-2913.
- 6 Horton J, Cohen J, Hobbs H: Molecular biology of PCSK9: its role in LDL metabolism. *Trends in Biochemical Sciences - Regular Edition* 2007; **32**: 71-77.
- 7 Abifadel M, Varret M, Rabès J *et al*: Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* 2003; **34**: 154-156.
- 8 Horton JD, Cohen JC, Hobbs HH: PCSK9: a convertase that coordinates LDL catabolism. *Journal of Lipid Research* 2009; **50**: S172-S177.
- 9 Homer V, Marais A, Charlton *et al*: Identification and characterization of two non-secreted PCSK9 mutants associated with familial hypercholesterolemia in cohorts from New Zealand and South Africa. *Atherosclerosis (00219150)* 2008; **196**: 659-666.
- 10 Kotowski I, Pertsemlidis , Luke *et al*: A Spectrum of PCSK9 Alleles Contributes to Plasma levels of Low-Density lipoprotein Cholesterol. *Am J Hum Genet* 2006; **78**: 410-422.

- 11 Miyake Y, Kimura R, Kokubo Y *et al*: Genetic variants in PCSK9 in the Japanese population: Rare genetic variants in PCSK9 might collectively contribute to plasma LDL cholesterol levels in the general population. *Atherosclerosis* 2008; **196**: 29-36.
- 12 Abifadel M, Guerin M, Benjannet S *et al*: Identification and characterization of new gain-of-function mutations in the PCSK9 gene responsible for autosomal dominant hypercholesterolemia. *Atherosclerosis* 2012; **223**: 394-400.
- 13 Cameron J, Holla ØL, Laerdahl JK *et al*: Characterization of novel mutations in the catalytic domain of the PCSK9 gene. *J Intern Med* 2008; **263**: 420-431.
- 14 Leren TP: Mutations in the PCSK9 gene in Norwegian subjects with autosomal dominant hypercholesterolemia. *Clin Genet* 2004; **65**: 419-422.
- 15 Timms KM, Wagner S, Samuels ME *et al*: A mutation in PCSK9 causing autosomal-dominant hypercholesterolemia in a Utah pedigree. *Hum Genet* 2004; **114**: 349.
- 16 Berge KE, Ose L, Leren TP: Missense Mutations in the PCSK9 Gene Are Associated With Hypocholesterolemia and Possibly Increased Response to Statin Therapy. *Arterioscler Thromb Vasc Biol* 2006; **26**: 1094-1100.
- 17 Allard D, Amsellem S, Abifadel M *et al*: Novel mutations of the PCSK9 gene cause variable phenotype of autosomal dominant hypercholesterolemia. *Hum Mutat* 2005; **26**: 497.
- 18 Cohen J, Boerwinkle, Mosley T, Hobbs H: Sequence Variations in PCSK9, Low LDL, and Protection against Coronary Heart Disease. *N Engl J Med* 2006; **354**: 1264-1272.
- 19 Hooper A, Marais A, Tanyanyiwa D, Burnett J: The C679X mutation in PCSK9 is present and lowers blood cholesterol in a Southern African population. *Atherosclerosis (00219150)* 2007; **193**: 445-448.
- 20 Abifadel M, Rabes JP, Devillers M *et al*: Mutations and polymorphisms in the proprotein convertase subtilisin kexin 9 (PCSK9) gene in cholesterol metabolism and disease. *Hum Mutat* 2009; **30**: 520-529.

- 21 Teo K, Chow CK, Vaz M, Rangarajan S, Yusuf S, PURE Investigators-Writing Group: The Prospective Urban Rural Epidemiology (PURE) study: examining the impact of societal influences on chronic noncommunicable diseases in low-, middle-, and high-income countries. *Am Heart J* 2009; **158**: 1-7.e1.
- 22 Collins JS, Schwartz CE: Detecting polymorphisms and mutations in candidate genes. *Am J Hum Genet* 2002; **71**: 1251.
- 23 Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977; **74**: 5463-5467.
- 24 Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* 2009; **4**: 1073-1081.
- 25 Adzhubei IA, Schmidt S, Peshkin L: A method and server for predicting damaging missense mutations. *Nature Methods* 2010; **7**: 248-249.
- 26 Gabriel SB, Schaffner SF, Nguyen H *et al*: The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225-2229.
- 27 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263-265.
- 28 Lohman TG, Roche AF, Martorell R: Anthropometric Standardization Reference Manual. Champaign, Illinois, Human kinetic books, 1988.
- 29 Ng PC, Henikoff S: SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**: 3812-3814.
- 30 Yamamoto T, Lu C, Ryan RO: A two-step binding model of PCSK9 interaction with the low density lipoprotein receptor. *J Biol Chem* 2011; **286**: 5464-5470.
- 31 Shioji K, Mannami T, Kokubo Y *et al*: Genetic variants in PCSK9 affect the cholesterol level in Japanese. *Journal of Human Genetics* 2004; 109-114.

32 Humphries SE, Whittall RA, Hubbart CS *et al*: Genetic causes of familial hypercholesterolaemia in patients in the UK: relation to plasma lipid levels and coronary heart disease risk. *J Med Genet* 2006; **43**: 943-949.

33 Abifadel M, Varret M, Rabes J *et al*: Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* 2003; **34**: 154.

34 Desmet F, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C: Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research* 2009; **37**: e67-e67.

35 Huijgen R, Sjouke B, Vis K *et al*: Genetic variation in APOB, PCSK9, and ANGPTL3 in carriers of pathogenic autosomal dominant hypercholesterolemic mutations with unexpected low LDL-C Levels. *Hum Mutat* 2012; **33**: 448-455.

36 Fasano T, Cefalu AB, Di Leo E *et al*: A Novel Loss of Function Mutation of PCSK9 Gene in White Subjects With Low-Plasma Low-Density Lipoprotein Cholesterol. *Arterioscler Thromb Vasc Biol* 2007; **27**: 677-681.

37 Zhao, Tuakli-Wosornu, Lagace *et al*: Molecular characterization of loss-of-function mutations in PCSK9 and identification of a compound heterozygote. *Am J Hum Genet* 2006; **79**: 514-523.

38 Benjannet S, Rhainds D, Hamelin J, Nassoury N, Seidah NG: The Proprotein Convertase (PC) PCSK9 Is Inactivated by Furin and/or PC5/6A. *Journal of Biological Chemistry* 2006; **281**: 30561-30572.

39 Huang C, Fornage M, Lloyd-Jones DM, Wei GS, Boerwinkle E, Liu K: Longitudinal association of PCSK9 sequence variations with low-density lipoprotein cholesterol levels the coronary artery risk development in young adults study. *Circulation: Cardiovascular Genetics* 2009; **2**: 354-361.

40 Chen SN, Ballantyne CM, Gotto AM, Tan Y, Willerson JT, Marian AJ: A common PCSK9 haplotype, encompassing the E670G coding single nucleotide polymorphism, is a novel genetic marker for plasma low-density lipoprotein cholesterol levels and severity of coronary atherosclerosis. *J Am Coll Cardiol* 2005; **45**: 1611-1619.

41 Benjannet S, Rhainds D, Essalmani R *et al*: NARC-1/PCSK9 and its natural mutants. *Journal of Biological Chemistry* 2004; **279**: 48865-48875.

42 Leren TP, Berge KE: Identification of mutations in the apolipoprotein B-100 gene and in the PCSK9 gene as the cause of hypocholesterolemia. *Clinica Chimica Acta* 2008; **397**: 92-95.

43 Norata GD, Garlaschelli K, Grigore L *et al*: Effects of PCSK9 variants on common carotid artery intima media thickness and relation to ApoE alleles. *Atherosclerosis* 2010; **208**: 177-182.

CHAPTER 5

PREDICTING LDL-CHOLESTEROL LEVELS IN A BLACK SOUTH AFRICAN POPULATION IN TRANSITION

Tertia van Zyl,^{1*} Johann C. Jerling,¹ Karin R. Conradie,¹ Edith J.M. Feskens²

¹ Centre of Excellence for Nutrition, North-West University, 2531, Potchefstroom, South Africa; ² Division of Human Nutrition, Wageningen University, 6700, Wageningen, The Netherlands

*Correspondence: tertia.vanzyl@nwu.ac.za; telephone number: +27 018 2852292; fax number: +27 018 2992464

This article has been submitted to Genes and Nutrition

From: "Genes and Nutrition (GENU)" <santhi.natraj@springer.com>
To: "Tertia Van Zyl" <tertia.vanzyl@nwu.ac.za>
Date: 2013/11/28 09:51 AM
Subject: GENU: Submission Confirmation

Dear Tertia,

Thank you for submitting your manuscript, Predicting LDL-cholesterol levels in a black South African population in transition, to Genes & Nutrition.

During the review process, you can keep track of the status of your manuscript by accessing the following web site:

<http://genu.edmgr.com/>

Your username is: Tertia
Your password is: vanzyl667

Should you require any further assistance please feel free to contact the Editorial Office by clicking on the "contact us" in the menu bar to send an email to us.

With kind regards,
Springer Journals Editorial Office
Genes & Nutrition

Now that your article will undergo the editorial and peer review process, it is the right time to think about publishing your article as open access. With open access your article will become freely available to anyone worldwide and you will easily comply with open access mandates. Springer's open access offering for this journal is called Open Choice (find more information on www.springer.com/openchoice). Once your article is accepted, you will be offered the option to publish through open access. So you might want to talk to your institution and funder now to see how payment could be organized; for an overview of available open access funding please go to www.springer.com/oafunding. Although for now you don't have to do anything, we would like to let you know about your upcoming options.

Abstract

Background: Elevated LDL-c levels promote the development of atherosclerosis and are a major risk factor in the progression of coronary heart disease. Dietary factors, weight, physical activity, age, gender and genetics are important factors that affect plasma LDL-c levels. The aim of the study was to investigate which of these factors best predict the variance in LDL-c levels in a black South African population.

Methods: The PCSK9 and LDLR genes in 1 530 volunteers, aged 35 to 60 years, of the South African PURE study population were screened for 52 variants. From these SNPs we determined a genetic risk score and haplotypes. Validated quantified food frequency questionnaires were used to determine the dietary intakes of the volunteers. Spearman's correlations were used to identify which factors correlated best with LDL-c levels. Separate linear regression models were used to determine the predictive value of each of the variables.

Results: The GRS, selected SNPs and haplotypes explained 1.4%, 3.4% and 2.7% of the variance respectively in plasma LDL-c. BMI was the factor with the largest predictive value and could explain 6.6% of the variance in LDL-c levels. There were no significant correlation between LDL-c levels and dietary intakes.

Conclusion: From these results we can conclude that BMI is a factor with a large impact on determining the variance in mean LDL-c levels in this study population. Genetic results showed that rare variants with greater effect on protein function have a better predictive value than the more common variants. The contribution of genetics in predicting LDL-c levels can be strengthened by including more genes as lifestyle factors contributed more towards the prediction of LDL-c levels in this study. The importance of the genetic variants in predicting LDL-c levels was less than the lifestyle factors used in this study.

Keywords Low-density lipoprotein, genetic risk score, haplotypes, LDLR, PCSK9, BMI, fat intake

Introduction

The role of elevated LDL-c levels in the development of coronary artery disease (CAD) has been well established (Pearson et al. 2002). Current information on early exposure to elevated LDL-c levels has also highlighted the importance of managing LDL-c levels and lowering the risk of CAD development (Ference et al. 2012). The South African black population has had historically low LDL-c levels (Steyn et al. 1991) but, as this population moves through the epidemiological transition, these levels are expected to increase, as established by Vorster et al. (Vorster 2002). The INTERHEART Africa study identified dyslipidemia as a single risk factor

that associated strongly with the development of CVD in Africa (Steyn et al. 2005). Recent data from the CRIBSA study showed that the black population in the Western Cape Province of South Africa is increasingly at risk of developing CAD as LDL-c levels have risen over the last 18 years (Peer et al. 2013). LDL-c levels are affected by various factors, including diet, weight, physical activity, age, gender (Grundy 2005) and genetics (Lusis 1988). Various dietary components have an effect on LDL-c levels such as fat, fatty acids and fiber. There is a positive linear relationship between saturated fatty acid intake, trans fatty acid intake and LDL-c concentration (Institute of Medicine 2005) with saturated fatty acids as the distinct dietary factor that has the largest negative effect on LDL-c concentration (Hu et al. 2001). Fiber, especially viscous fiber, can impede the absorption of dietary fat and cholesterol, and it can interfere with the enterohepatic recirculation of cholesterol and bile acids, which possibly will result in reduced blood cholesterol concentrations (Institute of Medicine 2005). Body mass index (BMI) has a positive association with LDL-c and it has been found that dyslipidemia develops gradually with a BMI $\geq 21\text{kg/m}^2$ (Haslam and James 2005). LDL-c levels also increase with age (McNamara et al. 1987) and lipoprotein levels differ between men and women (Heiss et al. 1980), with genetics adding an inter-individual variability of at least 50% to the equation (Burnett and Hooper 2008). The importance of each of these factors may differ in various populations and could depend on how fast the population is moving through the different stages of the epidemiological transition (Popkin 2002). Genes such as the low-density lipoprotein gene (*LDLR*) and the proprotein convertase subtilisin kexin 9 (*PCSK9*) are directly involved in the metabolism of LDL-c (Abifadel et al. 2003, Brown et al. 1981) and thus these genes were chosen in order to determine their association with LDL-c levels in the study population. The aim of this study, therefore, was to investigate which of the factors (diet, age, gender, BMI, physical activity and genetics) could add the most predictive value to explaining the variance in LDL-c levels in a black South African population from the North West Province. The results from this investigation would indicate which factor(s) should be addressed in the study population to prevent LDL-c levels from increasing to such levels that it would add to the progression of CAD.

Materials and methods

Study population

Individuals from the South African Prospective Urban and Rural Epidemiological (PURE) study were included in this research project. The study population from the North West Province of South Africa consists mainly of self-identified Tswana-speaking volunteers. The PURE study is

a prospective cohort study that follows changes in lifestyle, risk factors and chronic disease in the North West Province of South Africa over a period of no fewer than 10 years. In 2005, at baseline, eligible volunteers over the ages of 35 years were randomly selected from two urban and two rural areas. Only 2010 of the 3750 volunteers from which questionnaires were collected were eligible to take part in the study. Ikageng and Sonderwater extension 7 and 11 were included as the urban areas and Ganyesa and Tklagameng were selected as the rural areas and volunteers were recruited on a door to door basis. Urban and rural cohorts were matched for gender and age in order to avoid bias from population stratification. The baseline data collected in 2005 over a 12 week period was used in this study. Inclusion criteria were apparently healthy individuals between the ages of 35 and 60 years. Further information on the design of the PURE study is available from Teo et al. (Teo et al. 2009). DNA samples for 1 860 of the volunteers were available for analysis. Eventually 1 530 individuals were genotyped for the 52 variants and, after the exclusion of all individuals with missing values, a sample size of 1 169 remained to be used in the multiple regression analysis. All participants gave informed consent for the collection of blood samples and the performance of genetic analysis. Ethical approval for the study was obtained from the North-West University's Ethics Committee (Ethics number: NWU-00016-10-A1). Guidelines from the Declaration of Helsinki were followed throughout the study where humans were involved. Blood was sampled after eight hours of fasting. Serum lipid (total cholesterol (TC), HDL and TG) measurements were carried out by a Sequential Multiple Analyzer Computer (SMAC) using a KonelabTM auto analyser (Thermo Fisher Scientific, Vantaa, Finland). LDL-c levels were determined by using the Friedewald equation [$\text{LDL-c} = \text{TC} - \text{HDL-C} - (\text{TG})/2.17$] (Friedewald et al. 1972) and missing values were assigned to individuals with a TG measurement of $>4.5\text{mmol/L}$.

Gene and SNP selection and genotyping

The LDLR and PCSK9 genes were selected as candidate genes as they have a direct effect on LDL-c levels in humans (Zhang et al. 2008). All the exon and flanking regions of both genes were sequenced bidirectionally in a randomly chosen sub-sample of 30 individuals by means of automated cycle sequencing. Gene and SNP selection procedures as well as the method for genotyping analysis have been described in detail in Chapters 3 and 4.

Measurements

A level 3 anthropometrist supervised anthropometric measurements such as weight (kilograms), height (meters) and body mass index ($\text{BMI} = \text{weight in kilograms}/\text{height}^2 \text{ in meters}$) according to

guidelines adopted at the Airlie Conference sponsored by the National Institutes of Health. Standardized and calibrated equipment was used for measurements of volunteers, who wore minimal clothing (Lohman et al. 1988).

Questionnaires

Dietary intakes of the volunteers were determined by means of previously developed (MacIntyre et al. 2001a) and validated (MacIntyre et al. 2001b) culturally sensitive quantified food frequency questionnaires. The nutrient intakes were analyzed by means of the *Foodfinder3®* program (Medical Research Council, Tygerberg, South Africa), which is based on the South African Food Composition Tables (Langenhoven et al. 1998). The physical activity index (PAI) of the PURE populations was tested by means of PAI questionnaires developed and tested in the THUSA study, as previously described by Kruger and colleagues (Kruger et al. 2002).

Statistical analysis

Statistical analyses were performed with SPSS version 21 software (IBM). Descriptive statistics for the study population are reported in means and standard deviations as determined by Student's t-tests. Spearman's correlations determined which of the variables (urbanization, age, gender, BMI, physical activity and dietary intakes) were correlated with LDL-c and two-tailed p-values were used. The following dietary factors were included in the correlations: fat intake, fatty acid intake, cholesterol-, fiber-, carbohydrate- and protein intake. A variable with a correlation coefficient of > 0.1 was included in the regression models and the p-value had to be < 0.05 . The predictive value of the various contributors was determined by step-wise linear regression models. All assumptions for multiple linear regression analyses were met therefor all observations were independent with residuals normally distributed and there were no multicollinearity between variables. The SNPs, GRS and haplotypes were used in separate models to determine which of these genetic factors, together with the covariates, could best predict the variance in mean LDL-c levels. SNPs that were significantly associated ($p < 0.05$) with LDL-c levels after Bonferroni correction and have a minor allele frequency (MAF) of > 0.01 were used in the linear regression model to determine the predictive effect of single SNPs on LDL-c variance. Haplotypes and haplotype frequencies were determined by means of the software package Haploview version 4.2.25 (Broad Institute, Cambridge, Massachusetts, USA) by means of the confidence intervals as described by Gabriel et al. (Gabriel et al. 2002). Markers with a Hardy-Weinberg p-value of ≥ 0.05 and a MAF of $\geq 1\%$ were included in the analysis. Haploblocks with a significant association with mean LDL-c were included in the

regression model ($p < 0.05$). Haplotypes to be included in the regression model were identified by adding haploblocks into a backward regression model. The observed genotype frequencies were tested for deviation from Hardy-Weinberg equilibrium by using the Chi-square test of independence (χ^2). Statistical analyses were conducted on 1 169 individuals after all missing values were excluded.

Genetic risk score computation and analysis

A count genetic risk scores (GRSs) was calculated from the SNPs in the PCSK9 and LDLR gene that were significantly associated with LDL-c levels. We assumed that the SNPs were independently associated with LDL-c as well as an additive genetic model for each SNP, applying a linear weighting of 0, 1 and 2 to genotypes containing 0, 1 or 2 risk alleles (raising LDL-c levels), respectively. We assumed that each SNP in the count score contribute equally to the LDL-c increasing effect and it was calculated by summing the total number of risk alleles, producing a maximal score of 8. Individuals with no risk alleles were used as the reference group. Six SNPs had significant associations with elevated LDL-c levels, namely rs2738447, rs14158, rs2738465 and rs3180023 in LDLR and rs7552350 and rs28362269 in PCSK9. Variant rs3180023 was excluded from the SNP score because of the low frequency of the mutant allele (< 0.01) and rs7552350 was excluded because the SNP is not in HWE ($p = 0.02$).

Results

The characteristics of the study population are reported in Table 1 for men and women separately which were divided into rural and urban groups; means and standard deviations are reported. The ages of both groups are very similar with men at a mean age of 49.7 years and women at 49.3 years. There is a significant difference in the lipid profile of the men and women. The mean LDL-c levels of the women are much higher (3.11mmol/L) than those of the men (2.70mmol/L). There is also a significant difference (<0.001) between the mean LDL-c levels of the rural and urban women. The women also have a mean BMI of 26.8 kg/m², which is considered to be overweight and is also higher than the mean normal BMI (20.5 kg/m²) of the men. There is again a significant difference ($p=0.002$) between the BMI of the rural and urban women. From Table 1 it is also evident that the rural women are more physically active than their urban counterparts. Even though the dietary intakes differ significantly between men and women and rural and urban, for the most part their macronutrient intakes are within guidelines (National Research Council 2005). Fiber intake is also low in between genders and the urban group has a higher fiber intake.

Table 1 Population characteristics

Variable	Study Population ^a	Men ^a	Women ^a	T-test between genders
	Total (n=1169)	Total (n=440)	Total (n=729)	
	Rural (n=608)	Rural (n=211)	Rural (n=397)	
	Urban (n=561)	Urban (n=229)	Urban (n=332)	P-value
Age (years)				
Total	49.5 (10.3)	49.7 (10.07)	49.3 (10.5)	0.527
Rural	48.5 (9.98)	49.3 (10.2)	48.0 (9.83) [¥]	0.128
Urban	50.5 (10.6)	50.1 (9.93)	50.9 (11.03) [¥]	0.367
BMI (kg/m²)				
Total	24.4 (6.76)	20.5 (3.73)	26.8 (7.08)	<0.001
Rural	23.9 (6.27)	20.4 (3.59)	25.8 (6.58) [¥]	0.002
Urban	24.9 (7.21)	20.6 (3.85)	27.9 (7.49) [¥]	<0.001

Table 1 Population characteristics continues

Variable	Study Population ^a	Men ^a	Women ^a	T-test between genders
TC (mmol/l)				
Total	5.03 (1.38)	4.77 (1.31)	5.18 (1.39)	<0.001
Rural	4.98 (1.38)	4.71 (1.31)	5.12 (1.39) [¥]	<0.001
Urban	5.08 (1.37)	4.84 (1.32)	5.24 (1.38) [¥]	<0.001
LDL-c (mmol/l)				
Total	2.96 (1.15)	2.70 (1.10)	3.11 (1.15)	<0.001
Rural	2.96 (1.14)	2.70 (1.07)	3.09 (1.15) [¥]	<0.001
Urban	2.96 (1.15)	2.70 (1.14)	3.14 (1.14) [¥]	<0.001
HDL-C (mmol/l)				
Total	1.50 (0.62)	1.56 (0.64)	1.46 (0.61)	0.005
Rural	1.47 (0.59)	1.50 (0.64)	1.45 (0.57) [¥]	0.377
Urban	1.53 (0.65)	1.62 (0.65)	1.47 (0.65) [¥]	0.005
TAG (mmol/l)				
Total	1.26 (0.65)	1.12 (0.53)	1.34 (0.70)	<0.001
Rural	1.22 (0.59)	1.12 (0.49)	1.27 (0.63) [¥]	0.001
Urban	1.30 (0.71)	1.13 (0.56)	1.42 (0.77) [¥]	<0.001
Total energy (kJ)				
Total	7248 (3355)	7908 (3614)	6850 (3124)	<0.001
Rural	6202 (2543)	6659 (2745)	5958 (2397) [¥]	0.002
Urban	8382 (3739)	9058 (3927)	7916 (3535) [¥]	<0.001
Protein % of TE				
Total	12.2 (2.07)	12.5 (2.09)	12.0 (2.03)	<0.001
Rural	11.4 (1.90)	11.7 (1.90) [¥]	11.3 (1.89) [¥]	0.033
Urban	13.0 (1.93)	13.3 (1.96) [¥]	12.8 (1.89) [¥]	0.004

Table 1 Population characteristics continues

Variable	Study Population ^a	Men ^a	Women ^a	T-test between genders
CHO % of TE				
Total	63.9 (9.15)	64.6 (8.86)	63.5 (9.31)	0.049
Rural	68.5 (8.01)	69.4 (7.87) [¥]	68.1 (8.06) [¥]	0.048
Urban	58.8 (7.50)	60.1 (7.24) [¥]	58.0 (7.56) [¥]	0.001
Fat % of TE				
Total	23.7 (7.60)	23.0 (7.30)	24.2 (7.76)	0.007
Rural	20.3 (6.87)	19.3 (6.58) [¥]	20.8 (6.97) [¥]	0.01
Urban	27.5 (6.48)	26.4 (6.20) [¥]	28.3 (6.55) [¥]	<0.001
MUFA % of TE				
Total	6.18 (2.96)	5.97 (2.82)	6.30 (3.04)	0.065
Rural	4.67 (2.38)	4.45 (2.33) [¥]	4.79 (2.40) [¥]	0.092
Urban	7.81 (2.66)	6.35 (2.84) [¥]	8.11 (2.73) [¥]	0.001
PUFA % of TE				
Total	7.27 (2.95)	7.01 (2.76)	7.43 (3.04)	0.017
Rural	6.72 (3.22)	6.35 (2.84) [¥]	6.92 (3.39) [¥]	0.03
Urban	7.86 (2.48)	7.61 (2.54) [¥]	8.04 (2.44) [¥]	0.043
SATFAT % of TE				
Total	5.71 (2.76)	5.39 (2.50)	5.91 (2.88)	0.002
Rural	4.47 (2.51)	4.16 (2.32) [¥]	4.64 (2.59) [¥]	0.022
Urban	7.05 (2.36)	6.52 (2.10) [¥]	7.42 (2.46) [¥]	<0.001
Cholesterol intake (mg)				
Total	171.1 (137.8)	189.6 (151.3)	159.9 (127.8)	<0.001
Rural	116.9 (101.7)	122.2 (96.2) [¥]	114.0 (104.5) [¥]	0.341
Urban	229.8 (147.5)	251.7 (165.7) [¥]	214.7 (131.6) [¥]	0.005

Table 1 Population characteristics continues

Variable	Study Population ^a	Men ^a	Women ^a	T-test between genders
Total fibre intake (g)				
Total	20.09 (10.0)	22.4 (11.2)	18.7 (8.92)	<0.001
Rural	17.8 (7.76)	19.2 (8.74) [¥]	17.12 (7.20) [¥]	0.003
Urban	22.5 (11.5)	25.4 (12.3) [¥]	20.54 (10.41) [¥]	<0.001
PAI score				
Total	2.96 (0.87)	3.06 (1.35)	2.90 (0.63)	0.034
Rural (n=132/274)	3.01 (0.54)	2.96 (0.55)	3.02 (0.54) [¥]	0.269
Urban (n=121/204)	2.88 (1.28)	3.16 (1.86)	2.73 (0.71) [¥]	0.017

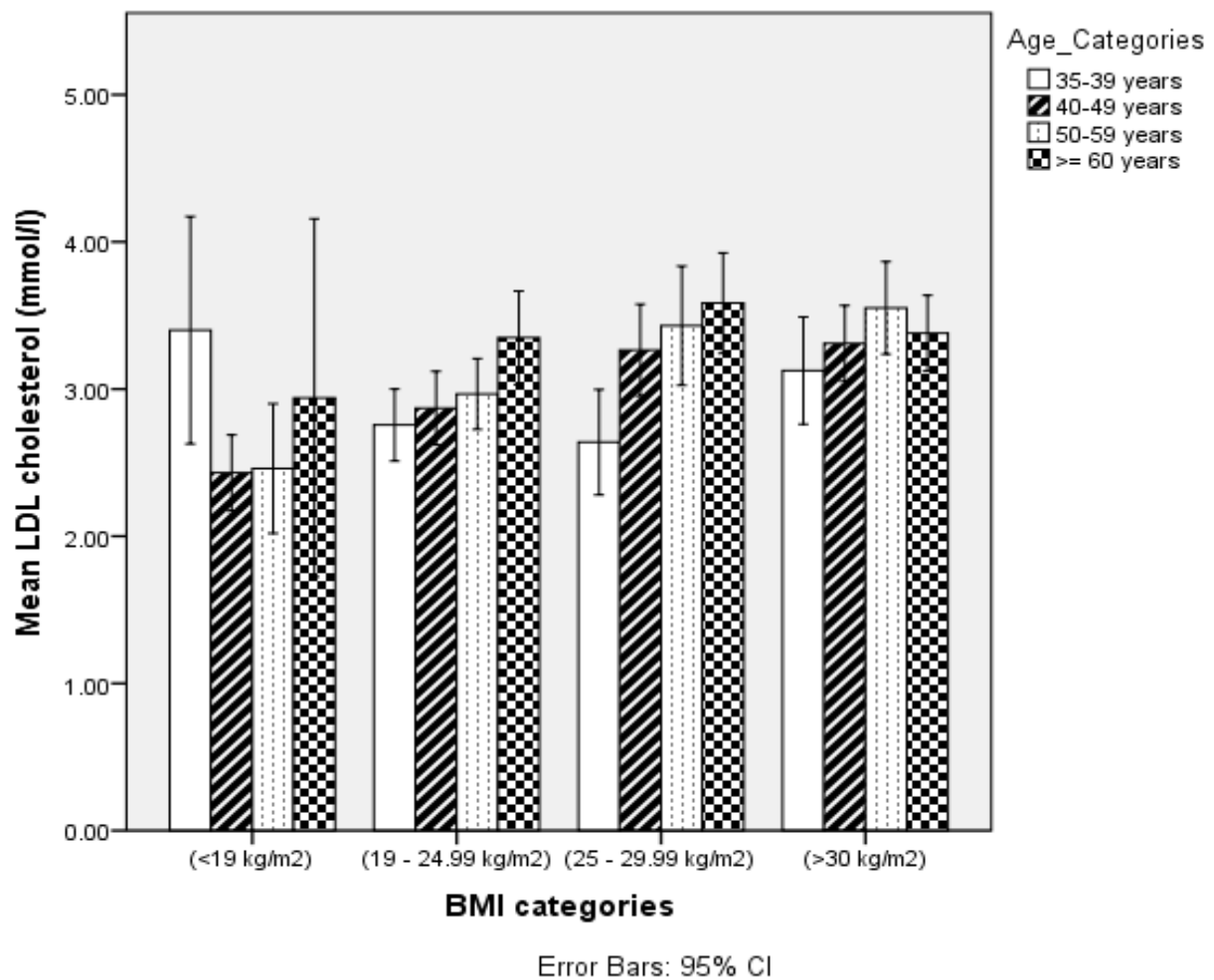
^a Means and standard deviations are reported. [¥]Indicates a significant difference between rural and urban groups for each variable. BMI, body mass index; TC, total cholesterol; LDL-c, low-density lipoprotein cholesterol; HDL-C high-density lipoprotein cholesterol; TAG, triglycerides; CHO, carbohydrates; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids; SATFAT, saturated fatty acids; PAI, physical activity index

In Table 2, LDL-c levels were analyzed by age and gender and it is clear from these results that LDL-c levels increase as the population ages. This increase is more prominent, however, in women over 50 years of age. The percentage of women with an LDL-c level >3mmol/L is also higher than that of the men. Mean BMI remains almost constant across the age groups in both women and men. In Figure 1 the differences between the various age categories and BMI categories in women are clearly noticeable.

Table 2 LDL-c levels and BMI according to age and gender

Age categories (years)					
Men	N	LDL-c (mmol/L) (SD) ^a	LDL-c ≤3mmol/L (%) ^a	LDL-c >3mmol/L (%) ^a	BMI (kg/m ²) (SD) ^a
35-39	77	2.71 (0.13)	68.8	31.2	20.2 (3.49)
40-49	162	2.65 (0.09)	63.6	36.4	20.3 (3.44)
50-59	125	2.69 (0.10)	64.0	36.0	20.5 (4.21)
>60	76	2.84 (0.13)	63.2	36.8	21.3 (3.65)
		p=0.669 ^b		p=0.858 ^b	p=0.250 ^b
Women					
35-39	138	2.92 (0.10)	58.7	41.3	27.0 (7.57)
40-49	283	3.03 (0.07)	55.8	44.2	26.3 (6.91)
50-59	184	3.20 (0.08)	44.6	55.4	26.9 (6.94)
>60	124	3.38 (0.10)	36.3	63.7	27.4 (7.11)
		p=0.004 ^b		p<0.001 ^b	p=0.504 ^b

^aMeans and standard deviations are reported. ^b ANOVA p-values. LDL-c, low-density lipoprotein cholesterol; BMI, body mass index, SD, standard deviation

Figure 1 Mean LDL-c according to BMI and age categories in women

From the correlations (Annexure L) that were done, only age, gender and BMI had a significant relationship with mean LDL-c levels. None of the dietary intake components or any of the other factors correlated significantly with mean LDL-c levels. BMI, age and gender were used in the linear regression models together with the genetic variables to determine which best predict LDL-c levels. Urbanization was not used as a predictor as there were no correlation between LDL-c and urbanization as demonstrated by identical means of LDL-c (2.96mmol/L) in the rural and urban groups. These three variables explained 7.7% of the variance in LDL-c levels, with BMI adding the most value (6.6%) to explaining the variance (Table 3).

Table 3 Linear regression model for BMI, gender, age with LDL-c as the dependent variable

Step 1	B	SE B	β	p-value
Constant	1.885	0.12		0.000
BMI	0.044	0.005	0.256	0.000
r ² = 0.066 (Step 1)				
Step 2	B	SE B	β	p-value
Constant	1.473	0.189		0.000
BMI	0.043	0.005	0.253	0.000
Age	0.009	0.003	0.077	0.005
r ² = 0.072 (Step 2)				
Step 3	B	SE B	β	p-value
Constant	1.486	0.189		0.000
BMI	0.037	0.005	0.219	0.000
Age	0.009	0.003	0.081	0.004
Gender	0.183	0.073	0.077	0.013
r ² = 0.076 (Step 3)				

BMI, Body mass index, Coding for gender: men=0; women=1

The study population was genotyped for 52 SNPs in the LDLR and PCSK9 genes (complete table is available in Annexure L) as described in Chapters 3 and 4. Four SNPs (rs2738447; $p < 0.001$, rs14158; $p = 0.001$, rs2738465; $p = 0.05$ and rs28362269; $p = 0.03$) were included in a GRS that significantly associated with elevated LDL-c levels. None of these SNPs were in linkage disequilibrium. Adding the GRS to the regression model accounted for a further 1.4% of the variance in LDL-c levels (Table 5). Table 6 contains a summary of the number of individuals in each risk allele group with the corresponding mean LDL-c levels.

Table 4 Variants in the LDLR and PCSK9 genes associated with LDL-c

SNP number	Nucleotide substitution / MAF	Gene / Position	p (HWE)	Frequency	LDL-c (mmol/L)	ANCOVA P-value ^a
rs499718	c.523+230 T>C	PCSK9	0.95	n=1533		
TT	MAF:0.51	Intron 3		369	2.88	
TC				760	3.02	0.047
CC				404	2.88	
rs7552350	c.657+114 C>A	PCSK9	0.02	n=1535		
CC	MAF:0.26	Intron 4		867	2.91	
CA				545	2.9	0.001
AA				123	3.31	
rs28362263	c.1327G>A	PCSK9	0.50	n=1632		
GG	(A443T)	Exon 8		1514	2.94 ^e	
GA	MAF:0.04			114	2.68 ^e	0.047
AA				4	2.77	
rs28362269	c.1504-16 C>T	PCSK9	0.95	n=1533		
CC	MAF:0.05	Intron 9		1374	2.92	
CT				154	3.01	0.05
TT				5	4.11	
rs28362286	c.2037C>A	PCSK9	0.57	n=1428		
CC	(C679X)	Exon 12		1350	2.99	
CA	MAF:0.03			78	2.17	<0.001
AA				0	-	
rs17249141	c.-217 C>T	LDLR	0.98	n=1535		
CC	MAF:0.01	Promoter		1521	2.95	
CT				16	2.4	0.050
TT				0	-	
rs5930	c.1413 A>G	LDLR	0.98	n=1532		
AA	MAF:0.14	Exon 10		29	3.29	
AG				357	3.04	0.023
GG				1146	2.90	
rs2738447	c.1706-55A>C	LDLR	0.64	n=1537		
CC	MAF:0.12	Intron 11		1188	2.9	
CA				322	3.03	0.001
AA				27	3.59	
rs14158	c.*52 G>A	LDLR	0.83	n=1521		
GG	MAF:0.11	3'UTR		1220	2.9	
GA				282	3.05	0.001
AA				19	3.71	

Table 4 continues

SNP number	Nucleotide substitution / MAF	Gene / Position	p (HWE)	Frequency	LDL-c (mmol/L)	ANCOVA P-value ^a
rs2738465	c.*504 G>A	LDLR	0.71	n=1535		
GG	MAF:0.26	3'UTR		855	2.90	
GA				573	2.95	0.050
AA				107	3.18	
rs3180023	c.*1217 C>A	LDLR	1.00	n=1537		
CC	MAF:0.001	3'UTR		1535	2.94	0.006
CA				2	5.09	
AA				0		

^a P-values after adjustment for age, gender, body mass index and multiple testing (Bonferroni), SNP, single nucleotide polymorphism, HWE, Hardy Weinberg equilibrium, LDL-c, low density lipoprotein cholesterol, ANCOVA, analysis of co-variance, MAF, minor allele frequency, PCSK9, proprotein convertase subtilisin kexin 9, LDLR, low-density lipoprotein receptor, 3'UTR, 3 untranslated region, A, adenine, G, guanine, C, cytosine, T, thymine

Table 5 Linear regression model for BMI, gender, age and GRS with LDL-c as the dependent variable

Step 1	B	SE B	β	p-value
Constant	1.468	0.194		0.000
BMI	0.037	0.005	0.220	0.000
Gender	0.180	0.075	0.076	0.017
Age	0.009	0.003	0.086	0.002
r ² =0.077 (step 1)				
Step 2	B	SE B	β	p-value
Constant	1.390	0.196		0.000
BMI	0.037	0.005	0.217	0.000
Gender	0.165	0.075	0.070	0.021
Age	0.010	0.003	0.090	0.003
GRS 1 allele	0.108	0.077	0.043	0.473
GRS 2 allele	0.036	0.101	0.011	0.224
GRS 3 allele	0.134	0.114	0.035	0.191
GRS 4 allele	0.486	0.204	0.068	0.099
GRS 5 allele	0.177	0.309	0.016	0.839
GRS 6 allele	1.58	0.552	0.080	0.001
GRS 7 allele	1.24	0.779	0.045	0.187
r ² =0.091 (step 2)				

BMI, body mass index, GRS, genetic risk score, Coding for gender: men=0; women=1, Zero risk alleles were used as reference in the analysis

Table 6 Summary of participants, number of risk alleles and the mean LDL-c levels per group

Number of alleles in GRS	N	Mean (95%CI)
0	514	2.88 (2.78-2.97)
1	338	3.00 (2.88-3.12)
2	153	2.95 (2.77-3.13)
3	114	3.05 (2.84-3.26)
4	31	3.43 (3.03-3.83)
5	13	3.03 (2.40-3.65)
6	4	4.52 (3.40-5.64)
7	2	4.15 (2.56-5.73)

GRS, genetic risk score, CI, confidence intervals

We also added the SNPs which associated significantly with LDL-c levels, separately to the regression model to determine their predictive effects (Table 7). The following SNPs were included in this analysis: rs14158, rs2738447, rs28362269 and rs28362286. In a χ^2 analysis rs5930 had a strong association with LDL-c; however, this SNP is in linkage disequilibrium with rs2738447 ($D'=0.948$; $r^2=0.80$; $LOD=315.5$) and was therefore not included in the regression model. When added together, the four SNPs explained 3.2% of the variance in mean LDL-c levels, with rs28362286 having the greatest contribution of 2.4%. The other SNPs separately made very small contributions to explaining the variance in LDL-c levels.

Haplotype analysis of both genes revealed four haploblocks in each gene. From a backward regression analysis, eight haplotypes were identified for inclusion in the regression model. The first two blocks in the LDLR gene had no haplotypes which could explain the variance in LDL-c levels significantly. One haplotype (GGG) in block 3 and one haplotype in block 4 (GC) of the LDLR gene were retained for analysis in the regression model. The haplotypes in block 3 of the LDLR gene were constructed of rs5927, rs14158 and rs3826810. LDLR haploblock 4 consisted of rs2738465 and rs1433099. No haplotypes in blocks 2 and 3 of the PCSK9 gene could significantly add to explaining the variance in LDL-c levels. In block 1 of the PCSK9 gene there were two haplotypes (CT/CT and TT/CC) that could contribute to explaining the variance in LDL-c levels. This block contained the SNPs rs2495480 and rs4927193. In block 4, which consisted of rs28362269, rs483462, rs505151 and rs662145, three haplotypes (CGAC, TGGC and CGAT/CAAT) which added significantly to explaining variance in LDL-c were identified. These haplotypes were used in the regression model with the covariates (BMI, age and gender) and together they had a predictive value of 10.4%, of which 2.7% could be explained by the haplotypes alone (Table 8).

Table 7 Linear regression model for BMI, gender, age and SNPs with LDL-c as the dependent variable

Step 1	B	SE B	β	p-value
Constant	1.468	0.194		0.000
BMI	0.037	0.005	0.22	0.000
Gender	0.18	0.075	0.076	0.016
Age	0.009	0.003	0.085	0.003
r ² =0.077 (step 1)				
Step 5	B	SE B	β	p-value
Constant	1.477	0.192		0.000
BMI	0.037	0.005	0.216	0.000
Gender	0.188	0.074	0.080	0.011
Age	0.009	0.003	0.083	0.003
rs28362286	-0.787	0.147	-0.149	0.000
rs14158 het	0.079	0.091	0.026	0.384
rs14158 homo	0.558	0.304	0.055	0.067
rs2738447 het	0.071	0.085	0.025	0.407
rs2738447 homo	0.340	0.295	0.034	0.249
rs28362269	0.100	0.103	0.027	0.330
r ² =0.108 (step 5)				

BMI, Body mass index; het, heterozygote; homo, homozygote, SNPs, single nucleotide polymorphism, Coding for gender: men=0; women=1; the wild type was used as reference in the analysis From the literature (Cohen *et al.*, 2005) it is known that rs28362286 has a large effect on LDL-c levels and therefor this variant was added first to the model. From the current research rs14158 carriers had mean LDL-c levels 27.9% higher than non-carriers and were therefore added secondly and rs2738447 was added thirdly because the carriers had mean LDL-c levels 23% higher than non-carriers. Rs28362269 were added last because very little is known about the variant.

Table 8 Linear regression model for BMI, gender, age and haplotypes with LDL-c as the dependent variable

Variables	B	SE B	β	p-value
Constant	1.405	0.198		0.000
BMI	0.038	0.005	0.222	0.000
Gender	0.174	0.074	0.073	0.019
Age	0.009	0.003	0.085	0.002
LDLR Haplotype Block3 GGG	-0.113	0.068	-0.049	0.095
LDLR Haplotype Block4 GC	-0.116	0.069	-0.049	0.095
PCSK9 Haplotype Block1 CT/CT	0.423	0.21	0.055	0.05
PCSK9 Haplotype Block1 TT/CC	0.204	0.109	0.052	0.065
PCSK9 Haplotype Block4 Pairing	0.179	0.080	0.065	0.028
PCSK9 Haplotype Block4 Hm CGAC	0.324	0.088	0.109	0.000
PCSK9 Haplotype Block4 Ht TGGC	0.257	0.129	0.058	0.046
PCSK9 Haplotype Block4 CGAT/CAAT	0.598	0.238	0.071	0.012
r ² = 0.104				

Abbreviations: BMI, body mass index; LDLR, low density lipoprotein receptor; PCSK9, proprotein convertase subtilisinlike/kexin type 9, Coding for gender: men=0; women=1

Discussion

In this study we aimed to investigate the predictive value of dietary factors, BMI, physical activity, age, gender and genetics on the variance in LDL-c levels. We identified BMI, age and gender as variables with significant correlation with LDL-c and used these, along with haplotypes, a GRS and a grouping of SNPs from the LDLR and PCSK9 genes, in various regression models to determine their contribution in predicting variation in LDL-c levels in the study population. This is one of few studies that utilize genetics to predict the variance in LDL-c in a black South African population.

The only dietary component that correlated ($r=0.072$) with mean LDL-c levels was saturated fat intake, but after adjusting for age, gender and BMI the association was weakened. From the dietary intake data it is clear that the 23.7% of energy from fat and 5.71% energy from saturated fat intake of the study population is lower than the recommended maximum level of 35% for total fat and 10% for saturated fat (Elmadfa and Kornsteiner 2009). The saturated fat intake of the high LDL-c group (5.9% energy from saturated fat) is also low and it would appear that contribution from saturated fat is not the cause of the elevated LDL-c levels. Although we did not detect an association between fat intake and LDL-c levels, Peer et al.(2013) found an association between high fat intake and elevated LDL-c levels in their cross-sectional investigation (Peer et al. 2013).

Associations between fat intake and LDL-c were also reported in a Lithuanian rural population where decreasing fat intake lead to lower LDL-c levels over 20 years. The percentage energy from total fat intake in the Lithuanian population was however above the suggested 35% (Ramazauskiene et al. 2011). Residual confounding is present in observational studies and this fact could explain the differences in associations between studies as confounders vary within populations (Smith and Ebrahim 2003). Evidence from randomized controlled trials gives a clearer association between fat and saturated fat intake where saturated fat intake elevate LDL-c levels (Mensink and Katan 1992). The fat and saturated fat intake in the study population might also be too low to have any association with LDL-c levels. The urban population does have a 20% higher mean intake of fat than their rural counterparts which is of concern and indicates that urbanisation is having an effect on the urban population. Two diet quality scores (DQS) for the PURE population were determined by Dolman et al. (2013) but these DQS had no correlation with LDL-c. These two scores indicated that the PURE population has a low intake of both micro- and macronutrients and that the urban group has a better diet quality than the rural group. The urban group also followed a diet that is low in fat even though their intakes of all foods were higher than that of the rural group (Dolman et al. 2013).

The Framingham Offspring study (Schaefer et al. 1994) and the CRIBSA study (Peer et al. 2013) both identified a significant association between older age and LDL-c levels and this is similar to our results. Analysis of the different age categories showed a significant association with LDL-c in the age category over 60 years. A regression analysis also shows a stronger relationship between older age and LDL-c in women (data available in Annexure L; $p=0.038$ for gender x age predictor).

In all linear regression analyses, the age, gender and BMI explained 7.7% of the variance in LDL-c, of which BMI could explain 6.6%. The relationship between BMI and LDL-c has been established within the South African context. The elevated circulating levels of LDL-c in obese individuals are the result of increased endogenous cholesterol synthesis which suppresses the synthesis of the LDL receptor and this result in lower clearance of the LDL particles (Katan and Beynen 1987, Keys et al. 1965). The THUSA study, conducted in the same province as the PURE study, identified the relationship between obesity and LDL-c (Vorster et al. 2005). More recently this association was found in the CRIBSA study conducted in a population from Cape Town. Peer et al. (2013) reported that increased BMI (OR 1.03, 95% CI 1.01-1.05; $p=0.003$) was significantly associated with LDL-c this compares well with the significant association that we found between LDL-c and increased BMI (OR, 1.07, 95% CI 1.050-1.089; $p=0.000$). In the Framingham Offspring study the association

found with LDL-c levels were also significant in both men and women with an OR 1.11 (95% CI; 1.07-1.15; $p=0.0001$) (Lamon-Fava et al. 1996).

We also investigated the effectiveness of genetic tools such as a GRS, haplotypes and SNPs in predicting variance in LDL-c levels. A GRS was constructed and had a very small predictive value of 1.4%. There were a total of four SNPs within the GRS. Lu et al. (2010) reported that, using 12 SNPs in their GRS, they could explain 6.9% of the variance in total cholesterol levels (Lu et al. 2010). An earlier study by Talmud et al. (2009) included 24 SNPs and had a predictive value of 14.6%. These findings suggest that a GRS which includes a large number of SNPs from various genes involved with LDL-c metabolism and regulation would be most beneficial in providing a screening method for individuals with elevated LDL-c levels. The GRS for elevated LDL-c levels consisted mostly of common variants and it would appear from our results that the predictive value of these variants is not very effective in the model we used in this study population. The GRS associated with lower LDL-c levels included the SNP rs28362286, which is a loss-of-function variant found in the *PCSK9* gene. PCSK9 is a protein that binds to the LDL receptor and is involved with the degradation of the receptor in the lysosome by preventing its recycling to the cell surface (Horton et al. 2009). A truncated protein is formed by the variant and, as the protein loses its ability to function, an increased number of receptors is present on the cell surface, allowing for an increased uptake of LDL-c, which causes the lower LDL-c levels (Cohen et al. 2005).

The haplotypes had the smallest predictive effect of all the genetic risk factors and could predict 2.4% of the variance in LDL-c levels. The haplotypes consisted mainly of common variants which have a smaller effect on LDL-c levels than the rare variants. Each of the eight haplotypes used in the regression model made small contributions towards predicting LDL-c levels. This indicates that these haplotypes will have little predictive value when used separately. The variant rs28362268 was not included in any of the haplotypes determined by Haploview according to Gabriel's confidence intervals (Gabriel et al. 2002). From the multiple regression analysis with the haplotypes we can also conclude that for this population there is little to no relationship between LDL-c levels and common variants for the chosen SNPs.

From the separate SNP input we were able to detect that variant rs28363386 contributed most (2.4%) to explaining the variation in LDL-c levels. As mentioned above, it is a rare variant, with a large effect on PCSK9 protein function, which causes lower LDL-c levels (Cohen et al. 2005). From the multiple regression models it can be seen that the heterozygote carriers (CA) of the variant had 0.787 mmol/L lower LDL-c levels than the non-carriers. Lower LDL-c levels are associated with a

lower risk for developing CAD where a 1mmol/L decrease in LDL-c reduces the risk of CHD events with 26.6% (Gould et al. 2007). Therefore heterozygote carriers of rs28362286 in this population have a 20% lower risk for CHD events.

We can therefore conclude that BMI adds the most significant value to predicting LDL-c levels in this population. Older women from urban areas with a BMI $>25\text{kg/m}^2$ are also the group most at risk of developing elevated LDL-c levels. Previous studies in South Africa as well as in other populations have found similar results (Peer et al. 2013, Kruger et al. 2001, Haslam and James 2005). Incorporating genetic variables to help predict LDL-c levels seems to have its limitations in this study but it does have an additive effect on explaining LDL-c variance in this population. It might be of more benefit to include a larger number of variants in a GRS from genes such as the APOE, APOB, ARH, NPC1L1 and MTP as suggested by the results of Lu et al. and Talmud et al. (Lu et al. 2010, Talmud et al. 2009) as multiple genes are involved with regulating LDL-c. Detection of other types of variants can also be included in future analysis as insertions, deletions and even splice variations and copy number variations also affect the function of the proteins. The effect of rare variants with a large impact on the function of the protein involved with LDL-c levels is much greater than that of the common variants that were included in the haplotypes, for example. These results would then rather support the common disease rare variant hypothesis where the rare variants with greater penetrance are stronger associated with a disease (Schork et al. 2009). In this study population factors such as age, gender and BMI that affect LDL-c levels can still be used with confidence to predict LDL-c levels. On the other hand, genetics will most certainly give us a better understanding of the differences that exist between populations.

Compliance with Ethics Guidelines

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all volunteers for being included in the study.

Conflicts of interest

Tertia van Zyl declares she has no conflict of interest.

Johann Jerling declares he has no conflict of interest.

Karin Conradie declares she has no conflict of interest.

Edith Feskens declares she has no conflict of interest.

Acknowledgements

We thank the organizers of the South African PURE study, especially Prof. A Kruger and her team, for providing the opportunity to work with the data and samples from the study population. Without their contributions this study would not have been possible. Wayne Towers is thanked for his support with regard to the genetic aspects of the study as well as for obtaining funding for the project. Funding was received from the North-West University and we would like to thank the institution for all of its contributions.

Appendix A. Supplementary data

LDLR and PCSK9 variants and associations with LDL-c

References

- Abifadel M, Varret M, Rabès J, Allard D, Ouguerram K, Devillers M, Cruaud C, Benjannet S, Wickham L, Erlich D (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* 34:154-156
- Brown MS, Kovanen PT, Goldstein JL (1981) Regulation of plasma cholesterol by lipoprotein receptors. *Science* 212:628-635
- Burnett JR, Hooper AJ (2008) Common and Rare Gene Variants Affecting Plasma LDL Cholesterol. *Clin Biochem Rev* 29:11-26
- Cohen J, Pertsemlidis A, Kotowski I, Graham R, Garcia C, Hobbs HH (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* 37:161-165

Collins JS, Schwartz CE (2002) Detecting polymorphisms and mutations in candidate genes. *Am J Hum Genet* 71:1251

Dolman RC, Wentzel-Viljoen E, Jerling JC, Feskens EJ, Kruger A, Pieters M (2013) The use of predefined diet quality scores in the context of CVD risk during urbanization in the South African Prospective Urban and Rural Epidemiological (PURE) study. *Public Health Nutr*:1-11

Elmadfa I, Kornsteiner M (2009) Fats and fatty acid requirements for adults. *Ann Nutr Metab* 55:56-75

Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, Kahn J, Afonso L, Williams KA, Flack JM (2012) Effect of Long-Term Exposure to Lower Low-Density Lipoprotein Cholesterol Beginning Early in Life on the Risk of Coronary Heart DiseaseA Mendelian Randomization Analysis. *J Am Coll Cardiol* 60:2631-2639

Friedewald WT, Levy RI, Fredrickson DS (1972) Estimation of the Concentration of Low-Density Lipoprotein Cholesterol in Plasma, Without Use of the Preparative Ultracentrifuge. *Clin Chem* 18:499-502

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225-2229

Gould AL, Davies GM, Alemao E, Yin DD, Cook JR (2007) Cholesterol reduction yields clinical benefits: meta-analysis including recent trials. *Clin Ther* 29:778-794

Grundey SM (2005) Prevention of atherosclerotic cardiovascular disease: why are the benefits of lifestyle therapies neglected?. *Dialogues Cardiol Med* 10:73

Grundey SM, Denke MA (1990) Dietary influences on serum lipids and lipoproteins. *J Lipid Res* 31:1149-1172

Haslam DW, James WPT (2005) Obesity. *The Lancet* 366:1197-1209

Heiss G, Tamir I, Davis CE, Tyroler HA, Rifkind BM, Schonfeld G, Jacobs D, Frantz Jr ID (1980) Lipoprotein-cholesterol distributions in selected North American populations: the lipid research clinics program prevalence study. *Circulation* 61:302-315

Horton JD, Cohen JC, Hobbs HH (2009) PCSK9: a convertase that coordinates LDL catabolism. *J Lipid Res* 50:S172-S177

Hu FB, Manson JE, Willett WC (2001) Types of dietary fat and risk of coronary heart disease: a critical review. *J Am Coll Nutr* 20:5-19

Katan MB, Beynen AC (1987) Characteristics of human hypo- and hyperresponders to dietary cholesterol. *American Journal of Epidemiology* 125:387-399

Keys A, Anderson JT, Grande F (1965) Serum cholesterol response to changes in the diet: IV. Particular saturated fatty acids in the diet. *Metab Clin Exp* 14:776-787

Kruger HS, Venter CS, Vorster HH, Margetts BM (2002) Physical inactivity is the major determinant of obesity in black women in the North West Province, South Africa: the THUSA study. *Nutrition* 18:422-427

Kruger H, Venter C, Vorster H (2001) Obesity in African women in the North West Province, South Africa is associated with an increased risk of non-communicable diseases: the THUSA study. *Br J Nutr* 86:733-740

Lamon-Fava S, Wilson PWF, Schaefer EJ (1996) Impact of Body Mass Index on Coronary Heart Disease Risk Factors in Men and Women: The Framingham Offspring Study. *Arteriosclerosis, Thrombosis, and Vascular Biology* 16:1509-1515

Langenhoven M, Kruger M, Gouws E, Faber M (1998) Medical Research Council Food Composition Tables

Lohman TG, Roche AF, Martorell R (1988) Anthropometric standardization reference manual. Human kinetic books, Champaign, Illinois

Lu Y, Feskens EJ, Boer J, Imholz S, Verschuren W, Wijmenga C, Vaarhorst A, Slagboom E, Müller M, Dollé ME (2010) Exploring genetic determinants of plasma total cholesterol levels and their predictive value in a longitudinal study. *Atherosclerosis* 213:200-205

Lusis A (1988) Genetic factors affecting blood lipoproteins: the candidate gene approach. *J Lipid Res* 29:397-429

MacIntyre U, Venter C, Vorster H (2001a) A culture-sensitive quantitative food frequency questionnaire used in an African population: 1. Development and reproducibility. *Public Health Nutr* 4:53

MacIntyre U, Venter C, Vorster H (2001b) A culture-sensitive quantitative food frequency questionnaire used in an African population: 2. Relative validation by 7-day weighed records and biomarkers. *Public Health Nutr* 4:63

McNamara J, Campos H, Ordovas J, Peterson J, Wilson P, Schaefer E (1987) Effect of gender, age, and lipid status on low density lipoprotein subfraction distribution. Results from the Framingham Offspring Study. *Arterioscler Thromb Vasc Biol* 7:483-490

Mensink RP, Katan MB (1992) Effect of dietary fatty acids on serum lipids and lipoproteins. A meta-analysis of 27 trials. *Arterioscler Thromb Vasc Biol* 12:911-919

National Research Council (2005) Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein, and amino acids. *Natl Academy Pr*, Washington, DC

Pearson TA, Blair SN, Daniels SR, Eckel RH, Fair JM, Fortmann SP, Franklin BA, Goldstein LB, Greenland P, Grundy SM (2002) AHA guidelines for primary prevention of cardiovascular disease and stroke: 2002 update consensus panel guide to comprehensive risk reduction for adult patients without coronary or other atherosclerotic vascular diseases. *Circulation* 106:388-391

Peer N, Steyn K, Lombard C, Gaziano T, Levitt N (2013) Alarming rise in prevalence of atherogenic dyslipidaemia in the black population of Cape Town: the Cardiovascular Risk in Black South Africans (CRIBSA) study. *Eur J Prev Cardiol*
<http://cpr.sagepub.com/content/early/2013/07/23/2047487313497865>

Popkin BM (2002) The shift in stages of the nutrition transition in the developing world differs from past experiences!. *Public Health Nutr* 5:205-214

Ramazauskiene V, Petkeviciene J, Klumbiene J, Kriaucioniene V, Sakytė E (2011) Diet and serum lipids: changes over socio-economic transition period in Lithuanian rural population. *BMC Public Health* 11:447

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463-5467

Schaefer EJ, Lamon-Fava S, Cohn SD, Schaefer MM, Ordovas J, Castelli W, Wilson P (1994) Effects of age, gender, and menopausal status on plasma low density lipoprotein cholesterol and apolipoprotein B levels in the Framingham Offspring Study. *J Lipid Res* 35:779-792

Schork NJ, Murray SS, Frazer KA, Topol EJ (2009) Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 19:212-219

Sing CF, Davignon J (1985) Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am J Hum Genet* 37:268-285

Smith GD, Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?. *Int J Epidemiol* 32:1-22

Steyn K, Jooste PL, Bourne L, Fourie J, Badenhorst CJ, Bourne DE, Langenhoven ML, Lombard CJ, Truter H, Katzenellenbogen J (1991) Risk factors for coronary heart disease in the black population of the Cape Peninsula. The BRISK study. *S Afr Med J* 79:480-485

Steyn K, Sliwa K, Hawken S, Commerford P, Onen C, Damasceno A, Ounpuu S, Yusuf S, for the INTERHEART Investigators in Africa, (2005) Risk Factors Associated With Myocardial Infarction in Africa: The INTERHEART Africa Study. *Circulation* 112:3554-3561

Talmud PJ, Drenos F, Shah S, Shah T, Palmen J, Verzilli C, Gaunt TR, Pallas J, Lovering R, Li K (2009) Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip. *Am J Hum Genet* 85:628-642

Teo K, Chow CK, Vaz 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 noncommunicable diseases in low-, middle-, and high-income countries. *Am Heart J* 158:1-7.e1

Vorster HH, Venter CS, Wissing MP, Margetts BM (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 Nutr* 8:480-490

Vorster HH (2002) The emergence of cardiovascular disease during urbanisation of Africans. *Public Health Nutr* 5:239-243

Zhang D, Garuti R, Tang W, Cohen JC, Hobbs HH (2008) Structural requirements for PCSK9-mediated degradation of the low-density lipoprotein receptor. *Proceedings of the National Academy of Sciences* 105:13045-13050

CHAPTER 6 – GENERAL DISCUSSION, CONCLUSION AND RECOMMENDATIONS

6.1 INTRODUCTION

In this concluding chapter the main findings of the three articles which form part of this thesis will be summarised and conclusions will be drawn based on these results. Chapters 3, 4 and 5 contain the discussion and interpretation of the results of this research project as well as comparisons made with relevant literature. This chapter aims to address the objectives of this study and therefore the main aim and objectives are included below. The general discussion, conclusion, strengths and limitations of the research and recommendations for future research will then follow.

Aims and objectives

The main aim of the study is to determine novel and existing genetic variants in the PCSK9 and LDLR genes and to describe the manner in which they associate with plasma LDL-c levels in a black South African population undergoing an epidemiological transition.

The objectives of the study are to:

- Determine the genetic variants of importance in the PCSK9 and LDLR genes in 30 randomly chosen black South African individuals from the South African PURE study by means of bidirectional automated cycle sequencing.
- Screen the population of 2000 individuals from the South African PURE study for the selected single nucleotide polymorphisms (SNPs) in the PCSK9 and LDLR genes identified through bidirectional automated cycle sequencing by means of BeadXpress array technology and to determine whether the population is in Hardy-Weinberg equilibrium for each gene.
- Investigate the association between LDL-c levels and the different genotypes in each examined SNP in the PCSK9 and LDLR genes.
- Compare the minor allele frequencies (MAF) of these SNPs with those of the European population from the 1000 Genomes project.

- Investigate the presence of haplotypes in the PCSK9 and LDLR genes and the association between LDL-c levels and the different haplotypes.
- Investigate which factors (diet, BMI, age, gender, genetic risk scores, haplotypes or genotypes) best predict LDL-c levels in the study populations.

6.2 DETERMINING IMPORTANT GENETIC VARIANTS IN THE PCSK9 AND LDLR GENE

The LDLR and PCSK9 genes were chosen for analysis because they have a direct effect on LDL-c levels (Brown *et al.*, 1981; Abifadel *et al.*, 2003). More than a 1000 variants have been identified within the LDLR gene and some of these variants situated in the coding regions are responsible for FH (Fokkema *et al.*, 2005) whereas others have a LDL-c lowering or raising effect and are situated within regulatory regions such as the promoter (Scholtz *et al.*, 1999) or the 3'UTR (Muallem *et al.*, 2007). More than 100 variants have also been identified in the PCSK9 gene and have been identified as GOF or LOF variants, where the GOF variants may cause FH (Abifadel *et al.*, 2003) and the LOF variants have a moderate to very efficient LDL-c lowering effect (Cohen *et al.*, 2005).

Little information is available on variants that are present in the LDLR and PCSK9 genes in the black South African population and, as seen in the literature review, the few known variants, most of which are FH-causing variants, have been identified in single individuals or small sample sizes (Leitersdorf *et al.*, 1988; Peeters *et al.*, 1998; Scholtz *et al.*, 1999; Thiart *et al.*, 2000). The sequencing of the promoter, exon and flanking intron regions of these genes was thus first undertaken to determine whether there were known and novel SNPs present within these areas of the genes. Sequencing of the promoter areas, exons and flanking intron regions in both these genes revealed that a large number of the variants are situated within the intron and regulating regions such as the promoter and 3'UTR. The LDLR gene has a large 3'UTR (2.5kb) (Yamamoto *et al.*, 1984), which could explain the large number of variants found within this region. The 3'UTR of the LDLR has been identified as a region that regulates gene expression by stabilising the mRNA (Nakahara *et al.*, 2002). Variants identified within the promoter region of LDLR have been associated with both elevated (Peeters *et al.*, 1998) and lowered levels of LDL-c (Scholtz *et al.*, 1999). A novel variant at position c.-314C>T was identified in the promoter of the LDLR gene, which, together with a known rare variant (rs17249141) is associated with low LDL-c levels (Scholtz *et al.*, 1999). Ten variants were identified within the exons of the LDLR gene, eight of these being

synonymous mutations and two of them missense mutations, where the base change led to a change in amino acid. Even though the synonymous mutations do not cause a change in amino acids these changes could be identified as exon-splicing enhancers that alter the efficiency of LDLR splicing and are associated with elevated LDL-c (Rafiq *et al.*, 2011). Two variants in exon 12 and 13 are exon-splicing enhancers, namely rs688 and rs5925. Six novel variants were identified by means of sequencing the LDLR gene, five of them located in intron regions and, as previously mentioned, one identified in the promoter region. In total, 70 variants have been identified from the regions that were sequenced in the LDLR gene, of which six were novel. It is also important to determine which of the variants that were identified in the literature were detected with sequencing of the gene in the study population. The only variant identified in the literature in the black South African population that was detected in the study population was rs17249141. Most of the variants previously found in the black South African population were also variants that caused FH and the study samples were very small, ranging from one individual to members of one family (Leitersdorf *et al.*, 1988; Peeters *et al.*, 1998; Scholtz *et al.*, 1999; Thiar *et al.*, 2000). These variants are also rare and it could be that the sample size used could not detect these very rare variants although, if we take into account that the LDL-c levels were not high enough to diagnose for homozygote FH (Blom, 2011), these variants may not be present in the study population. Variants that have been previously identified and were available on the 1000 Genomes and dbSNP databases have been identified within other populations. This indicates that several variants are shared amongst populations, which raises the question of whether the frequencies of the minor alleles of each variant are comparable across populations. It is interesting to note that, while exons 4 and 9 have been identified as FH hot spots in other populations (Varret *et al.*, 1997), in the current study population no variants were detected in exons 3, 4, 5, 9 and 17. Classified as FH cold spots (Varret *et al.*, 1997), each of exons 13 and 15 revealed one variant, which is comparable with the literature. These results already give an indication of the difference between populations.

GOF and LOF mutations in the PCSK9 gene have been identified as having direct effects on LDL-c levels and it is important to investigate these (Abifadel *et al.*, 2003; Cohen *et al.*, 2005). These variants are rare, however, and in order to have a complete picture of the gene's role in determining the differences in LDL-c levels between individuals it is of great importance to include common variants in the investigation. The variants rs28362286 and rs67608943 have been detected in the African American population (Cohen *et al.*, 2005). Only rs28362286 was detected in the current study in two of the 30 individuals included in the subsample. The other variants that were identified with the sequencing of this gene have been identified in other populations and these include the

following variants that are associated with lower LDL-c levels: rs111563724, rs28362263 and rs28362286 (which have been identified as LOF variants). Known GOF variants detected in the study population were rs28362261, rs141502002 and rs28362270. Variant S127R (c.625T>A) was not detected in the subsample and, as previously mentioned in the literature review, this is the only GOF mutation that has been identified in an individual of mixed ancestry in a South African population (Homer *et al.*, 2008). As the PCSK9 gene was identified only in 2003, many of the variants that are currently known are not fully characterised and their associations with LDL-c are not known. Some variants have been identified as part of haplotypes that associate with LDL-c, one such variant being rs505151 (Chen *et al.*, 2005), which was detected in the study population. Sequencing of the regions in the PCSK9 gene, as previously indicated, revealed 71 variants, of which seven were novel. A large number of variants were found in the intron and regulating regions; five synonymous variants, 12 missense variants, two insertions and two deletions were detected in the PCSK9 gene.

6.3 GENOTYPING THE STUDY POPULATION FOR SNPS IDENTIFIED AND DETERMINING HARDY-WEINBERG EQUILIBRIUM

By using the Illumina's VeraCode™ Golden Gate Genotyping (GGGT) Assay on a BeadXpress™ Reader system using the VeraCode™ technology the analysis was limited to the genotyping of SNPs only. The insertions and deletions identified by the Sanger sequencing was thus not genotyped. Various limitations of the GGGT Assay design also lessened the number of SNPs that could be analysed. During the GGGT Assay additional SNPs failed the analysis decreasing the SNP number once again. These are limitations of the method that was chosen to genotype the population. The methods chosen for genotyping a population should be carefully considered as this could have a large impact on the significance of results.

All of the SNPs except for two were not in HWE which gives an indication that no population stratification was present in the study population such as non-random mating, natural selection, genetic drift and the presence of admixture (Cardon & Palmer, 2003). Adherence to HWE also gives an indication that the mutant alleles which were identified are representative of the population and that the apparently healthy population did not have an impact on excluding FH patients.

6.4 THE ASSOCIATION BETWEEN LDL-C LEVELS AND THE DIFFERENT GENOTYPES OF EACH SNP IN THE PCSK9 AND LDLR GENES AND MAF COMPARISONS

Five variants genotyped in the PCSK9 genes were significantly associated with LDL-c levels. The two LOF variants were associated with lower LDL-c levels, as has previously been reported by various studies (Allard *et al.*, 2005; Cohen *et al.*, 2005). These two variants are rare variants and were present in very low frequencies the study population, which was comparable to the frequencies in the African population used in the 1000 Genomes project (referred to as the African reference group) and are not present in the European population (referred to as the European reference group) used in the 1000 Genomes project.

Variant rs28362269 is also a rare variant and was significantly associated with elevated LDL-c levels. The MAF of this variant was similar to that of the African reference group but it was not present in the European reference group. The variant was detected in a Malaysian study on FH patients and was also significantly associated with higher LDL-c levels in these patients at a MAF of 0.11 (Lye *et al.*, 2013).

Variant rs499718 is a common variant that was significantly associated with LDL-c levels but the heterozygotes were the genotype with the elevated levels (association between TC and CC). Unfortunately, no information is available on the functional significance of this variant as it is also situated in intron 3. The MAF of the variant is 0.17 in the European reference population, which is less than the 0.51 and 0.46 of the study population and African reference population.

Rs7552350, a common variant, was significantly associated with elevated LDL-c levels but was not in HWE and, as previously mentioned, this association should be interpreted with caution. Functional information on this variant was not available in the literature and the MAF is not known in either of the reference groups used.

Various common SNPs were also identified in the PCSK9 gene; however, these were not significantly associated with LDL-c levels. The three novel variants identified had a small LDL-c lowering effect ranging from 3.7–7.4% when an individual had one allele and a lowering effect of 10.2–26.1% when both alleles were present. The association with LDL-c levels did not reach statistical significance, however, as the size of the homozygote variant groups was too small for statistical relevance. These results nevertheless warrant further research into the mechanisms by which these variants affect LDL-c levels and whether they could have a protective effect on those

harbouring the variants. Two variants previously reported (Kotowski *et al.*, 2006) and associated with elevated LDL-c were not significantly associated. This highlights the importance of population-specific research in genetics.

The variants genotyped in the LDLR gene which were significantly associated with LDL-c levels were found mostly in the regulatory areas of the gene such as the promoter, 3'UTR and intron regions responsible for effective splicing, as discussed in Chapter 4. Variant rs17249141, which was associated with lower levels of LDL-c, does not seem to be present in the European reference group but the MAF is similar to that of the African reference group. The MAF of rs2738447 and rs14158 is lower than the MAF in the European reference group but, again, it is quite similar to that of the African population from the 1000 Genomes project. The only variant with a MAF that was comparable across all three populations, namely rs2738465, is a common variant which was significantly associated with LDL-c levels. Even though these variants are shared amongst the various populations it can be seen that, because of the different frequencies, the contribution of each variant towards the variance in LDL-c levels would differ amongst the populations. The novel variants that were identified are rare, with MAF of less than 0.02, and there was no association with LDL-c levels.

Most of the variants detected in the exons had an elevating effect on LDL-c levels although the associations did not reach statistical significance. Nevertheless, this is an indication that variants found within the exon region do have an effect on receptor function. Even synonymous variants such as rs688 and rs5925, as previously mentioned in section 6.2, are exon splice enhancers that have an influence on the effectiveness of the splicing of mRNA. The MAF of these two variants, as previously mentioned in Chapter 4, are also very different from those of the European reference group. These variants are a good example of how genetics may differ between populations and of the importance of establishing how a SNP or haplotype would add to understanding the variance of different biological/biochemical markers for each different population. Ten of the 52 variants genotyped in this study population were significantly associated with LDL-c levels. Other studies have used larger numbers of SNPs and also found only a few variants that associated with the chosen biochemical marker or outcome (Lu *et al.*, 2010; Lye *et al.*, 2013; Talmud *et al.*, 2009). This simply indicates that although there are many variants present within each gene, not all of the variants associate with LDL-c levels.

6.5 THE PRESENCE OF HAPLOTYPES AND THE ASSOCIATION BETWEEN LDL-C LEVELS AND THE DIFFERENT HAPLOTYPES

The presence of haplotypes in the LDLR and PCSK9 genes and their associations with LDL-c have not been extensively investigated in the black South African population and therefore little information on this topic is available. Various haplotypes were determined by utilising the software program Haploview and making use of Gabriel's confidence intervals (Gabriel *et al.*, 2002). Case-control analysis was conducted to determine whether the frequency of the haplotypes differed between cases and controls. Cases were coded as individuals with LDL-c of greater than 3mmol/L and the controls were those with levels equal to or less than 3mmol/L. From the case-control analysis of the LDLR gene in Haploview it was observed that the haplotypes with the highest frequency were associated with lower LDL-c levels. From these results it could be hypothesised that the presence of these haplotypes is the reason for the low mean LDL-c levels in this population.

Haplotypes identified in the PCSK9 gene did not have such a strong association with LDL-c levels and those haplotypes that had a strong significant association were associated with elevated levels of LDL-c. One of the novel variants formed a haplotype with one other variant, where the AC haplotype associated with elevated LDL-c levels and the CC haplotype associated with lower levels of LDL-c. Variant rs505151 has previously been identified in a haplotype that associated with LDL-c levels (Chen *et al.*, 2005) and in this study population it was found that the variant also associated with elevated levels of LDL-c. Linkage disequilibrium (LD) between the variants in both genes is not very strong when r^2 is used as a measure to determine the recombination of haplotypes. In PCSK9 more historical recombination has taken place as determined by the multiallelic D' and there were only two variants in LD. The LDLR gene had more variants in LD as less recombination has taken place between the haploblocks.

6.6 FACTORS PREDICTING LDL-C LEVELS

Saturated fat intake was the only dietary component to associate with LDL-c levels in this population, although after adjustment for age, gender and BMI, this association lost its significance. No interaction was found between saturated fat intake and the variants genotyped in the two genes in the study population. This could also be explained, as LDLR and PCSK9 expression is regulated by SREBP-1, which is regulated by sterols in the circulation. Transcription is up-regulated when

plasma is depleted of sterol through the interaction of the sterol response element (SRE-1) with other regulatory elements, which promotes the synthesis of the LDL receptor mRNA. Transcription is inhibited when the sterol regulatory element-binding protein interacts with SRE-1, rendering it inactive in the presence of sterols (Soutar, 1999).

The factors that affect LDL-c levels, as identified in the literature, are diet, BMI, physical activity, gender (Grundey 2005) and genetics (Lusis 1988). In the PURE study population, factors that correlated most strongly with LDL-c levels were age, gender and BMI. The effect of genetics was also tested by using genetic risk scores (GRSs), SNPs and haplotypes from the LDLR and PCSK9 genes. Saturated fat intake had a weak correlation with mean LDL-c levels and this is most probably due to the fact that the study population has a diet that can be described as low in fat because their mean fat intake is below 25% of TE and the mean saturated fat intake is below 6% of TE. Investigating the correlation between dietary quality scores and LDL-c levels indicated that there was no relation between these two variables. This could be because of the low fat intake of the study population.

Gender also had a correlation with LDL-c, women having significantly higher mean LDL-c levels than men in the study population. The mean LDL-c levels of the study population also increase with age but this is more noticeable in women, where the percentage of individuals with LDL-c levels above 3mmol/L increases over the age categories. For men, these percentages stay constant over the age categories. The group with the highest mean LDL-c levels was older women within the highest BMI category. This association has previously been found in the Framingham Offspring study (McNamara *et al.*, 1987) and recently the CRIBSA study also found a similar association (Peer *et al.*, 2013). Catabolism of LDL-c decreases as an individual age and, because of a lower fractional catabolic rate of the lipoprotein, these levels increase (Ericsson *et al.*, 1991). This could possibly be due to the reduction in LDL receptor activity in the liver, which then causes impairment in the catabolism of LDL (Ericsson *et al.*, 1991). There is a much stronger association between age and elevated LDL-c in the women than in the men in this study population, which was also found in the Framingham Offspring study, as women with elevated LDL-c tend to be older (Schaefer *et al.*, 1994). Menopause in women also affects LDL-c associations but information on menopause for this population was not available even though the older women had higher LDL-c levels. BMI associated most strongly with LDL-c and added the best predictive value to explaining LDL-c levels. As stated in the Framingham Offspring study, this is a factor that can be modified by lifestyle changes and this information should be communicated to the specific target population. Even though the study population has a fat intake of less than 25%, it has been suggested that the

increase in fat intake of 16% to approximately 26% over the last 50 years is associated with the increased prevalence of obesity and overweight in the black South African population (Bourbon *et al.*, 2008). This study also attempted to determine whether genetic variables would add value to predicting the variance in LDL-c levels. The allele count GRS associated with elevated LDL-c comprised variants that were more common and did not have a strong association with LDL-c levels. The predictive value was small in comparison with the GRS associated with lower levels of LDL-c as this GRS consisted of variants that are rare and have a greater effect on LDL-c levels. Because the predictive value of the single SNPs was very much the same as that of the GRS, either single SNPs or a GRS can be used as the effects are the same. A GRS is beneficial if only one variable is needed. Even when using them all together, the haplotypes had very little predictive value and this might be explained by the presence of common SNPs in these haplotypes. The variant rs28362286 had the greatest predictive value in this population when it was used as a single SNP and, even when used in the GRS, associated with lower LDL-c. Rare variants with a large impact on protein function have a greater predictive effect on the variance in LDL-c levels in individuals.

An apparently healthy population

Volunteers recruited for the study should not previously have been diagnosed for any chronic diseases of lifestyle and use medication for these conditions. Definite exclusions were HIV and TB. As previously mentioned elsewhere in this thesis the HWE was used to ensure that the alleles were in equilibrium and this was used to conclude that the study population sample was representative of the community it was taken from. When the range of LDL-c is analysed in more detail it can be seen that 42.5% of the population has mean LDL-c level above 3mmol/l. There is also 5% of the population who have a LDL-c level above 5mmol/l which is the bottom of the range for heterozygote FH individuals. Analysis of the baseline data from the PURE study indicated that only 1.4% of the participants on the African continent were using statins as secondary preventative drug (Yusuf *et al.*, 2011) which could indicate that hypercholesterolemia is either underdiagnosed in this population or it is not a common chronic disease in this population. From this it can be concluded that the fact that the study population was apparently healthy would not have caused an exclusion of individuals with FH.

6.7 STRENGTHS AND LIMITATIONS

This study is one of the few studies in South Africa that investigated the frequency and association of various genotypes in the LDLR and PCSK9 genes with LDL-c in a black South African population. Various mutations previously found in European and African populations were confirmed and variant, rs148698650, was identified in an African population for the first time. Some rare variants were also identified that are not found in the European population. Novel variants were also identified in this study population which were not previously identified. One of the first studies known to use a GRS to predict LDL-c levels. Haplotypes in the LDLR gene were identified which associated with LDL-c levels. This is one of few studies that report on a more detailed genotype-phenotype association.

No functional test were done on the novel variants found in both genes to verify that the association with the variant is result of a defective protein and not with another loci which could be LD with the novel variant. A limited number of SNPs were included in the GRS derived from only the two genes that were investigated. A larger sample size could give more significant results and more homozygote individuals could be identified. All the variants could not be genotyped as the GGGT assay had various limitations. Only LDL-c were investigated and not particle sizes or other lipid fractions. This study did not investigate the cause of the elevated BMI in this population which could be contributed to inactivity, high energy consumption or genetic factors.

6.8 RECOMMENDATIONS

The following recommendations for future research stem from the discussions in this thesis:

1. Whole genes should be sequenced with new technology such as next-generation sequencing to determine all possible variants present within the regulatory regions, introns and exons as the results show that variants in all parts of the gene are important. Additional genes that are involved with LDL-c metabolism should be included in future investigations as this will give more insight into the role of genetics in determining LDL-c levels. Examples of such genes are APOB, APOE, ARH, MTP and NPC1P1.
2. Measuring PCSK9 levels for the study populations could also give more insight into the mechanisms behind the PCSK9 variants. Functional studies on the variants which had significant associations but for which the function of the variant allele is not known would also give more insight into their role in LDL-c metabolism.

3. Investigating the impact of genetics on LDL-c levels and the predictors of LDL-c levels is the start of research into lipoproteins and their role in developing CAD in this population. Measuring apoB and other non-HDL-C levels is also of great importance as these lipoproteins have been shown to have a stronger association with CAD than LDL-c alone (Pischon *et al.*, 2005). Investigating of the LDL particle size is also of great importance as the different sizes contributes differently towards CAD risk and elucidating the role thereof in the obese individuals of the population would be of great importance to improve interventions and to lower their risk of developing CAD.
4. In order to construct a better model to predict the variance in LDL-c levels other variables should also be investigated and included in the regression models and as mentioned above more genetic variants should be included. Examples of these variables may include haemostasis factors, other lipid fragments and blood pressure.

6.9 GENERAL CONCLUSION

From the genetic aspect of this study it is clear that there is a difference in the frequency of variants in the *LDLR* and *PCSK9* genes between this study population and European populations or even some African populations. This again puts an emphasis on the importance of population specific genetic research. The results from this study are also only applicable to this specific population in epidemiological transition living in the North West province of South Africa. There was no significant relationship between saturated fat and LDL-c levels which is most probably because of the low fat intake in both urban and rural groups. DQS also did not associate with LDL-c levels and this could be the result of the low intake of all foods by both the rural and urban groups. BMI turned out to be the best predictor for LDL-c where older women from urban areas with a BMI >25kg/m² had higher LDL-c levels. To ensure a larger predictive value from genetic data more variants from various other genes should be included in the analysis as previously mentioned. Increased LDL-c levels still remain one of the leading causes of CAD and in this study population more than 50% of the volunteers had LDL-c levels higher than the recommended 3mmol/L. Hypercholesterolaemia should be monitored in this population because even though their diet (low fat intake) has not yet had an effect on LDL-c levels other factors from the transition which is yet to be identified has had an impact on the BMI of the population which in turn correlates with LDL-c levels.

REFERENCES

- Abifadel, M., Rabes, J.P., Devillers, M., Munnich, A., Erlich, D., Junien, C., Varret, M. & Boileau, C. 2009. Mutations and polymorphisms in the proprotein convertase subtilisin kexin 9 (PCSK9) gene in cholesterol metabolism and disease. *Human mutation*, 30(4):520-529.
- Abifadel, M., Varret, M., Rabes, J., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., Derre, A., Villegier, L., Farnier, M., Beucler, I., Bruckert, E., Chambaz, J., Chanu, B., Lecerf, J. & Luc, G. 2003. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nature genetics*, 34(2):154.
- 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.
- Alborn, W.E., Cao, G., Careskey, H.E., Qian, Y., Subramaniam, D.R., Davies, J., Conner, E.M. & Konrad, R.J. 2007. Serum proprotein convertase subtilisin kexin type 9 is correlated directly with serum LDL cholesterol. *Clinical chemistry*, 53(10):1814-1819.
- Amento, E.P., Ehsani, N., Palmer, H. & Libby, P. 1991. Cytokines and growth factors positively and negatively regulate interstitial collagen gene expression in human vascular smooth muscle cells. *Arteriosclerosis, thrombosis, and vascular biology*, 11(5):1223-1230.
- Amsellem, S., Briffaut, D., Carrié, A., Rabès, J., Girardet, J., Fredenrich, A., Moulin, P., Krempf, M., Reznik, Y. & Vialettes, B. 2002. Intronic mutations outside of alu-repeat-rich domains of the LDL receptor gene are a cause of familial hypercholesterolemia. *Human genetics*, 111(6):501-510.
- Amuna, P. & Zotor, F.B. 2008. Epidemiological and nutrition transition in developing countries: Impact on human health and development. *The proceedings of the nutrition society*, 67(1):82-90.
- Anderson, R.G., Goldstein, J.L. & Brown, M.S. 1977. A mutation that impairs the ability of lipoprotein receptors to localise in coated pits on the cell surface of human fibroblasts. *Nature*, 270(5639):695-699.
- Arca, M. & Jokinen, E. 1998. Low density lipoprotein receptor mutations in a selected population of individuals with moderate hypercholesterolemia. *Atherosclerosis*, 136(1):187-194.

- Baass, A., Dubuc, G., Tremblay, M., Delvin, E.E., O'Loughlin, J., Levy, E., Davignon, J. & Lambert, M. 2009. Plasma PCSK9 is associated with age, sex, and multiple metabolic markers in a population-based sample of children and adolescents. *Clinical chemistry*, 55(9):1637-1645.
- Babiak, J. & Rudel, L.L. 1987. Lipoproteins and atherosclerosis. *Baillière's clinical endocrinology and metabolism*, 1(3):515-550.
- Badimón, L., Vilahur, G. & Padró, T. 2009. Lipoproteins, platelets, and atherothrombosis. *Revista Española de cardiología (English edition)*, 62(10):1161-1178.
- Barnes, M.J. & Farndale, R.W. 1999. Collagens and atherosclerosis. *Experimental gerontology*, 34(4):513-525.
- Benjannet, S., Rhainds, D., Hamelin, J., Nassoury, N. & Seidah, N.G. 2006. The proprotein convertase (PC) PCSK9 is inactivated by furin and/or PC5/6A. *Journal of biological chemistry*, 281(41):30561-30572.
- Berriot-Varoqueaux, N., Aggerbeck, L.P., Samson-Bouma, M. & Wetterau, J.R. 2000. The role of the microsomal triglyceride transfer protein in abetalipoproteinemia. *Annual review of nutrition*, 20(1):663.
- Bjornheden, T., Babyi, A., Bondjers, G. & Wiklund, O. 1996. Accumulation of lipoprotein fractions and subfractions in the arterial wall, determined in an *in vitro* perfusion system. *Atherosclerosis*, 123(1-2):43-56.
- Blom, D. 2011. Familial hypercholesterolaemia: Review article. *South African family practice*, 53(1):11-18.
- Boren, J., Wettsten, M., Rustaeus, S., Andersson, M. & Olofsson, S. 1993. The assembly and secretion of apoB-100-containing lipoproteins. *Biochemical society transactions*, 21(2):487-493.
- Bourne, L.T., Langenhoven, M.L., Steyn, K., Jooste, P.L., Laubscher, J.A. & Van der Vyver, E. 1993. Nutrient intake in the urban African population of the Cape Peninsula, South Africa. the brisk study. *The Central African journal of medicine*, 39(12):238-247.

- Brown, M.S., Anderson, R.G. & Goldstein, J.L. 1983. Recycling receptors: The round-trip itinerary of migrant membrane proteins. *Cell*, 32(3):663-667.
- Brown, M.S. & Goldstein, J.L. 2009. Cholesterol feedback: From Schoenheimer's bottle to Scap's MELADL. *Journal of lipid research*, 50(Supplement):S15-S27.
- Brown, M.S., Kovanen, P.T. & Goldstein, J.L. 1981. Regulation of plasma cholesterol by lipoprotein receptors. *Science*, 212(4495):628-635.
- Brown, M.S. & Goldstein, J.L. 1986. A receptor-mediated pathway for cholesterol homeostasis. *Science*, 232(4746):34-47.
- Bunn, C.F., Lintott, C.J., Scott, R.S. & George, P.M. 2002. Comparison of SSCP and DHPLC for the detection of LDLR mutations in a New Zealand cohort. *Human mutation*, 19(3):311-311.
- Burnett, J.R. & Hooper, A.J. 2008. Common and rare gene variants affecting plasma LDL cholesterol. *The clinical biochemist.reviews / Australian association of clinical biochemists*, 29(1):11-26.
- Callis, Jansen, Thiart, de Villiers, Raal & Kotze. 1998. Mutation analysis in familial hypercholesterolemia patients of different ancestries: Identification of three novel LDLR gene mutations. *Molecular and cellular probes*, 12(3):149-152.
- Camejo, G., Hurt-Camejo, E., Wiklund, O. & Bondjers, G. 1998. Association of apo B lipoproteins with arterial proteoglycans: Pathological significance and molecular basis. *Atherosclerosis*, 139(2):205-222.
- Cameron, J., Holla, Ø.L., Laerdahl, J.K., Kulseth, M.A., Ranheim, T., Rognes, T., Berge, K.E. & Leren, T.P. 2008. Characterization of novel mutations in the catalytic domain of the PCSK9 gene. *Journal of internal medicine*, 263(4):420-431.
- Camino-López, S., Llorente-Cortés, V., Sendra, J. & Badimon, L. 2007. Tissue factor induction by aggregated LDL depends on LDL receptor-related protein expression (LRP1) and rho A translocation in human vascular smooth muscle cells. *Cardiovascular research*, 73(1):208-216.

- Cardon, L.R. & Palmer, L.J. 2003. Population stratification and spurious allelic association. *The lancet*, 361(9357):598-604.
- Careskey, H.E., Davis, R.A., Alborn, W.E., Troutt, J.S., Cao, G. & Konrad, R.J. 2008. Atorvastatin increases human serum levels of proprotein convertase subtilisin/kexin type 9. *Journal of lipid research*, 49(2):394-398.
- Carreau, A., Kieda, C. & Grillon, C. 2011. Nitric oxide modulates the expression of endothelial cell adhesion molecules involved in angiogenesis and leukocyte recruitment. *Experimental cell research*, 317(1):29-41.
- Chadjichristos, C.E. & Kwak, B.R. 2007. Connexins: New genes in atherosclerosis. *Ann med*, 39(6):402-411.
- Chen, W., Goldstein, J.L. & Brown, M.S. 1990. NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. *Journal of biological chemistry*, 265(6):3116-3123.
- Chin, D.J., Luskey, K.L., Faust, J.R., MacDonald, R.J., Brown, M.S. & Goldstein, J.L. 1982. Molecular cloning of 3-hydroxy-3-methylglutaryl coenzyme A reductase and evidence for regulation of its mRNA. *Proceedings of the National Academy of Sciences*, 79(24):7704-7708.
- Cohen, J., Pertsemlidis, A., Kotowski, I., Graham, R., Garcia, C. & Hobbs, H.H. 2005. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nature genetics*, 37(2):161-165.
- Cohen, J.C., Pertsemlidis, A., Fahmi, S., Esmail, S., Vega, G.L., Grundy, S.M. & Hobbs, H.H. 2006. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proceedings of the national academy of sciences of the United States of America*, 103(6):1810-1815.
- Costet, Cariou, Lambert, Lalanne, Lardeux, Jarnoux, Grefhorst, Staels & Krempf. 2006. Hepatic PCSK9 expression is regulated by nutritional status via insulin and sterol regulatory element-binding protein 1c. *The journal of biological chemistry*, 281(10):6211-6218.

- Cui, Q., Ju, X., Yang, T., Zhang, M., Tang, W., Chen, Q., Hu, Y., Haas, J.V., Troutt, J.S., Pickard, R.T., Darling, R., Konrad, R.J., Zhou, H. & Cao, G. 2010. Serum PCSK9 is associated with multiple metabolic factors in a large Han Chinese population. *Atherosclerosis*, 213(2):632-636.
- Cunningham, D., Danley, D.E., Geoghegan, K.F., Griffor, M.C., Hawkins, J.L., Subashi, T.A., Varghese, A.H., Ammirati, M.J., Culp, J.S. & Hoth, L.R. 2007. Structural and biophysical studies of PCSK9 and its mutants linked to familial hypercholesterolemia. *Nature structural & molecular biology*, 14(5):413-419.
- Daub, K., Langer, H., Seizer, P., Stellos, K., May, A.E., Goyal, P., Bigalke, B., Schönberger, T., Geisler, T., Siegel-Axel, D., Oostendorp, R.A.J., Lindemann, S. & Gawaz, M. 2006. Platelets induce differentiation of human CD34+ progenitor cells into foam cells and endothelial cells. *The FASEB journal*, 20(14):2559-2561.
- Davis, C.G., Elhammer, A., Russell, D., Schneider, W., Kornfeld, S., Brown, M.S. & Goldstein, J.L. 1986a. Deletion of clustered O-linked carbohydrates does not impair function of low density lipoprotein receptor in transfected fibroblasts. *Journal of biological chemistry*, 261(6):2828-2838.
- Davis, C.G., Goldstein, J.L., Südhof, T.C., Anderson, R., Russell, D.W. & Brown, M.S. 1986b. Acid-dependent ligand dissociation and recycling of LDL receptor mediated by growth factor homology region. *Nature*, 326(6115):760-765.
- de Graaf, J., Hak-Lemmers, H.L., Hectors, M.P., Demacker, P.N., Hendriks, J.C. & Stalenhoef, A.F. 1991. Enhanced susceptibility to in vitro oxidation of the dense low density lipoprotein subfraction in healthy subjects. *Arteriosclerosis and thrombosis : A journal of vascular biology / american heart association*, 11(2):298-306.
- Deiana, L., Garuti, R., Pes, G.M., Carru, C., Errigo, A., Roller, M., Pisciotto, L., Masturzo, P., Cantafora, A., Calandra, S. & Bertolini, S. 2000. Influence of β^0 -thalassemia on the phenotypic expression of heterozygous familial hypercholesterolemia : A study of patients with familial hypercholesterolemia from Sardinia. *Arteriosclerosis, thrombosis, and vascular biology*, 20(1):236-243.
- Dewpura, T., Raymond, A., Hamelin, J., Seidah, N.G., Mbikay, M., Chrétien, M. & Mayne, J. 2008. PCSK9 is phosphorylated by a golgi casein kinase-like kinase ex vivo and circulates as a phosphoprotein in humans. *FEBS journal*, 275(13):3480-3493.

- Dolman, R.C., Jerling, J.C., Pieters, M., Ntyintyane, L. & Raal, F. 2011. Risk factor profile of coronary artery disease in black South Africans. *SA heart journal*, 8(1):4-11.
- Engelmann, B., Kogl, C., Kulschar, R. & Schaipp, B. 1996. Transfer of phosphatidylcholine, phosphatidylethanolamine and sphingomyelin from low-and high-density lipoprotein to human platelets. *Biochem.J*, 315:781-789.
- Esser, V. & Russell, D.W. 1988. Transport-deficient mutations in the low density lipoprotein receptor. alterations in the cysteine-rich and cysteine-poor regions of the protein block intracellular transport. *Journal of biological chemistry*, 263(26):13276-13281.
- Fan, D., Yancey, P.G., Qiu, S., Ding, L., Weeber, E.J., Linton, M.F. & Fazio, S. 2008. Self-association of human PCSK9 correlates with its LDLR-degrading activity. *Biochemistry*, 47(6):1631-1639.
- Feeley, A., Musenge, E., Pettifor, J.M. & Norris, S.A. 2012. Changes in dietary habits and eating practices in adolescents living in urban South Africa: The Birth to Twenty cohort. *Nutrition*, 28(7–8):e1-e6.
- Ference, B.A., Yoo, W., Alesh, I., Mahajan, N., Mirowska, K.K., Mewada, A., Kahn, J., Afonso, L., Williams, K.A. & Flack, J.M. 2012. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: A mendelian randomization analysis. *Journal of the American college of cardiology*, 60(25):2631-2639.
- Fisher, T.S., Surdo, P.L., Pandit, S., Mattu, M., Santoro, J.C., Wisniewski, D., Cummings, R.T., Calzetta, A., Cubbon, R.M., Fischer, P.A., Tarachandani, A., De Francesco, R., Wright, S.D., Sparrow, C.P., Carfi, A. & Sitlani, A. 2007. Effects of pH and low density lipoprotein (LDL) on PCSK9-dependent LDL receptor regulation. *Journal of biological chemistry*, 282(28):20502-20512.
- Fouchier, S.W., Defesche, J.C., Umans-Eckenhausen, M.A. & Kastelein, J.J. 2001. The molecular basis of familial hypercholesterolemia in the Netherlands. *Human genetics*, 109(6):602-615.
- Frayn, K. & Stanner, S. 2005. The aetiology and epidemiology of cardiovascular disease. (In Stanner, S., ed. first ed. Blackwell Publishing Ltd. p. 1-22).

- Gleissner, C.A., Leitinger, N. & Ley, K. 2007. Effects of native and modified low-density lipoproteins on monocyte recruitment in atherosclerosis. *Hypertension*, 50(2):276-283.
- Goldstein, J.L., Anderson, R.G. & Brown, M.S. 1979. Coated pits, coated vesicles, and receptor-mediated endocytosis. *Nature*, 279(5715):679-685.
- Goldstein, J.L. & Brown, M.S. 1990. Regulation of the mevalonate pathway. *Nature*, 343(6257):425.
- Goldstein, J.L., Brown, M.S., Anderson, R.G., Russell, D.W. & Schneider, W.J. 1985. Receptor-mediated endocytosis: Concepts emerging from the LDL receptor system. *Annual review of cell biology*, 1(1):1-39.
- Goldstein, J.L., Dana, S.E. & Brown, M.S. 1974. Esterification of low density lipoprotein cholesterol in human fibroblasts and its absence in homozygous familial hypercholesterolemia. *Proceedings of the national academy of sciences*, 71(11):4288-4292.
- Gould, A.L., Rossouw, J.E., Santanella, N.C., Heyse, J.F. & Furberg, C.D. 1998. Cholesterol reduction yields clinical benefit: Impact of statin trials. *Circulation*, 97(10):946-952.
- Gould, A.L., Davies, G.M., Alemao, E., Yin, D.D. & Cook, J.R. 2007. Cholesterol reduction yields clinical benefits: Meta-analysis including recent trials. *Clinical therapeutics*, 29(5):778-794.
- Grefhorst, A., McNutt, M.C., Lagace, T.A. & Horton, J.D. 2008. Plasma PCSK9 preferentially reduces liver LDL receptors in mice. *Journal of lipid research*, 49(6):1303-1311.
- Grundy, S.M. 2005. Prevention of atherosclerotic cardiovascular disease: Why are the benefits of lifestyle therapies neglected? *Dialogues in cardiovascular medicine*, 10(2):73.
- Hakim, J.G., Odwee, M.G., Siziya, S., Ternouth, I. & Matenga, J. 1995. Acute myocardial infarction in zimbabwe: The changing scene of coronary artery disease. *The Central African journal of medicine*, 41(10):303-308.
- Hampton, E.N., Knuth, M.W., Li, J., Harris, J.L., Lesley, S.A. & Spraggon, G. 2007. The self-inhibited structure of full-length PCSK9 at 1.9 Å reveals structural homology with resistin within the C-terminal domain. *Proceedings of the National Academy of Sciences*, 104(37):14604-14609.
- Haslam, D.W. & James, W.P.T. 2005. Obesity. *The lancet*, 366(9492):1197-1209.

- 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.
- Heiss, G., Tamir, I., Davis, C.E., Tyroler, H.A., Rifkind, B.M., Schonfeld, G., Jacobs, D. & Frantz Jr, I.D. 1980. Lipoprotein-cholesterol distributions in selected north american populations: The lipid research clinics program prevalence study. *Circulation*, 61(2):302-315.
- Hawkes, C. 2006. Uneven dietary development: Linking the policies and processes of globalization with the nutrition transition, obesity and diet-related chronic diseases. *Globalization and health*, 24.
- Hedstrom, L. 2002. Serine protease mechanism and specificity. *Chemical reviews*, 102(12):4501-4524.
- Hobbs, H., Brown, M.S., Goldstein, J.L. & Russell, D. 1986. Deletion of exon encoding cysteine-rich repeat of low density lipoprotein receptor alters its binding specificity in a subject with familial hypercholesterolemia. *Journal of biological chemistry*, 261(28):13114-13120.
- Hobbs, H.H., Russell, D.W., Brown, M.S. & Goldstein, J.L. 1990. The LDL receptor locus in familial hypercholesterolemia: Mutational analysis of a membrane protein. *Annual review of genetics*, 24(1):133-170.
- Hobbs, H.H., Brown, M.S. & Goldstein, J.L. 1992. Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. *Human mutation*, 1(6):445-466.
- Homer, V., Marais, A., Charlton, , Laurie, A., Hurndell, , Scott, , Mangili, , Sullivan, D., Barter, P., Rye, , George, P. & Lambert, . 2008. Identification and characterization of two non-secreted PCSK9 mutants associated with familial hypercholesterolemia in cohorts from New Zealand and South Africa. *Atherosclerosis (00219150)*, 196(2):659-666.
- Hoogendijk, , Scholtz, , Pimstone, , Ehrenborg, E., Kastelein, J.J., Defesche, , Thiart, R., Plessis, L., de Villiers, , Zaahl, , Delport, R., Rubinsztein, , Raffel, , Grim, , Benckekor, S., Amouyel, P., Brousseau, T., Steyn, K., Lombard, & Hayden, . 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 & cellular probes*, 17(4):175.

- Horton, J.D., Cohen, J.C. & Hobbs, H.H. 2009. PCSK9: A convertase that coordinates LDL catabolism. *Journal of lipid research*, 50(Supplement):S172-S177.
- Horton, J.D., Shah, N.A., Warrington, J.A., Anderson, N.N., Park, S.W., Brown, M.S. & Goldstein, J.L. 2003. Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. *Proceedings of the National Academy of Sciences*, 100(21):12027-12032.
- Horton, J.D., Cohen, J.C. & Hobbs, H.H. 2009. PCSK9: A convertase that coordinates LDL catabolism. *Journal of lipid research*, 50(Supplement):S172-S177.
- Horton, J., Cohen, J. & Hobbs, H. 2007. Molecular biology of PCSK9: Its role in LDL metabolism. *Trends in biochemical sciences - regular edition*, 32(2):71-77.
- Hui, D.Y., Innerarity, T., Milne, R., Marcel, Y. & Mahley, R. 1984. Binding of chylomicron remnants and beta-very low density lipoproteins to hepatic and extrahepatic lipoprotein receptors. A process independent of apolipoprotein B48. *Journal of biological chemistry*, 259(24):15060-15068.
- Humphries, S.E., Whittall, R.A., Hubbart, C.S., Maplebeck, S., Cooper, J.A., Soutar, A., Naoumova, R., Thompson, G.R., Seed, M. & Durrington, P.N. 2006. Genetic causes of familial hypercholesterolaemia in patients in the UK: Relation to plasma lipid levels and coronary heart disease risk. *Journal of medical genetics*, 43(12):943-949.
- Hussain, M.M., Iqbal, J., Anwar, K., Rava, P. & Dai, K. 2003. Microsomal triglyceride transfer protein: A multifunctional protein. *Frontiers in bioscience: A journal and virtual library*, 8s500-6.
- Hwang, S., Ballantyne, C.M., Sharrett, A.R., Smith, L.C., Davis, C.E., Gotto, A.M. & Boerwinkle, E. 1997. Circulating adhesion molecules VCAM-1, ICAM-1, and E-selectin in carotid atherosclerosis and incident coronary heart disease cases: The atherosclerosis risk in communities (ARIC) study. *Circulation*, 96(12):4219-4225.
- Jensen, H., Jensen, L., Hansen, P., Faergeman, O. & Gregersen, N. 1996. High sensitivity of the single-strand conformation polymorphism method for detecting sequence variations in the low-density lipoprotein receptor gene validated by DNA sequencing. *Clinical chemistry*, 42(8):1140-1146.

- Jeong, H.J., Lee, H., Kim, K., Kim, Y., Yoon, D. & Park, S.W. 2008. Sterol-dependent regulation of proprotein convertase subtilisin/kexin type 9 expression by sterol-regulatory element binding protein-2. *Journal of lipid research*, 49(2):399-409.
- Kajinami, K., Mabuchi, H., Itoh, H., Michishita, I., Takeda, M., Wakasugi, T., Koizumi, J. & Takeda, R. 1988. New variant of low density lipoprotein receptor gene. FH-tonami. *Arteriosclerosis, thrombosis, and vascular biology*, 8(2):187-192.
- Kathiresan, S., Manning, A.K., Demissie, S., D'Agostino, R.B., Surti, A., Guiducci, C., Gianniny, L., Burt, N., Melander, O., Orho-Melander, M., Arnett, D.K., Peloso, G.M., Ordovas, J.M. & Cupples, L.A. 2007. A genome-wide association study for blood lipid phenotypes in the Framingham heart study. *BMC medical genetics*, 8S17-10.
- 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. *The New England journal of medicine*, 329(2):73-78.
- Khazaei, M., Moien-afshari, F. & Laher, I. 2008. Vascular endothelial function in health and diseases. *Pathophysiology*, 15(1):49-67.
- Khoo, K., van Acker, P., Defesche, J., Tan, H., van, d.K., Heijnen-van Eijk, S., Kastelein, J. & Deslypere, J. 2000. Low-density lipoprotein receptor gene mutations in a southeast asian population with familial hypercholesterolemia. *Clinical genetics*, 58(2):98-105.
- 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.
- Kourimate, S., Le May, C., Langhi, C., Jarnoux, A.L., Ouguerram, K., Zaïr, Y., Nguyen, P., Krempf, M., Cariou, B. & Costet, P. 2008. Dual mechanisms for the fibrate-mediated repression of proprotein convertase subtilisin/kexin type 9. *Journal of biological chemistry*, 283(15):9666-9673.
- Kruger, H.S., Venter, C.S., Vorster, H.H. & Margetts, B.M. 2002. Physical inactivity is the major determinant of obesity in black women in the North West province, South Africa: The THUSA study. *Nutrition*, 18(5):422-427.

- Kruger, H.S., Venter, C.S. & Vorster, H.H. 2003. Physical inactivity as a risk factor for cardiovascular disease in communities undergoing rural to urban transition: The THUSA study. *Cardiovascular journal of South Africa*, 14:16-22.
- Kruger, H.S., Venter, C.S. & Vorster, H.H. 2001. Obesity in African women in the North West province, South Africa is associated with an increased risk of non-communicable diseases: The THUSA study. *British journal of nutrition*, 86(06):733-740.
- Kubes, P., Suzuki, M. & Granger, D. 1991. Nitric oxide: An endogenous modulator of leukocyte adhesion. *Proceedings of the National Academy of Sciences*, 88(11):4651-4655.
- Kwon, Lagace, McNutt, Horton & Deisenhofer. 2008. Molecular basis for LDL receptor recognition by PCSK9. *Proceedings of the National Academy of Sciences of the United States of America*, 105(6):1820-1825.
- Lagace, T.A., Curtis, D.E., Garuti, R., McNutt, M.C., Park, S.W., Prather, H.B., Anderson, N.N., Ho, Y., Hammer, R.E. & Horton, J.D. 2006. Secreted PCSK9 decreases the number of LDL receptors in hepatocytes and in livers of parabiotic mice. *Journal of clinical investigation*, 116(11):2995-3005.
- Lakoski, S.G., Lagace, T.A., Cohen, J.C., Horton, J.D. & Hobbs, H.H. 2009. Genetic and metabolic determinants of plasma PCSK9 levels. *Journal of clinical endocrinology metabolism*, 94(7):2537-2543.
- Lambert, Ancellin, Charlton, Comas, Pilot, Keech, Patel, Sullivan, Cohn, Rye & Barter. 2008. Plasma PCSK9 concentrations correlate with LDL and total cholesterol in diabetic patients and are decreased by fenofibrate treatment. *Clinical chemistry*, 54(6):1038-1045.
- Langhi, C., Le May, C., Kourimate, S., Caron, S., Staels, B., Krempf, M., Costet, P. & Cariou, B. 2008. Activation of the farnesoid X receptor represses PCSK9 expression in human hepatocytes. *FEBS letters*, 582(6):949-955.
- Lehrman, M.A., Schneider, W.J., Südhof, T.C., Brown, M.S., Goldstein, J.L. & Russell, D.W. 1985. Mutation in LDL receptor: Alu-alu recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science*, 227(4683):140-146.

- Leitersdorf, E., Hobbs, H.H., Fourie, A.M., Jacobs, M., van der Westhuyzen, D.R. & Coetzee, G.A. 1988. Deletion in the first cysteine-rich repeat of low density lipoprotein receptor impairs its transport but not lipoprotein binding in fibroblasts from a subject with familial hypercholesterolemia. *Proceedings of the National Academy of Sciences of the United States of America*, 85(21):7912-7916.
- Leren, T.P., Solberg, K., Rødningen, O.K., Tonstad, S. & Ose, L. 1994. Two founder mutations in the LDL receptor gene in norwegian familial hypercholesterolemia subjects. *Atherosclerosis*, 111(2):175-182.
- Lewis, G.F. & Rader, D.J. 2005. New insights into the regulation of HDL metabolism and reverse cholesterol transport. *Circulation research*, 96(12):1221-1232.
- Li, J., Tumanut, C., Gavigan, J., Huang, W., Hampton, E.N., Tumanut, R., Suen, K.F., Trauger, J.W., Spraggon, J., Lesley, S.A., Laiu, G., Yowe, D. & Harris, J.L. 2007. Secreted PCSK9 promotes LDL receptor degradation independently of proteolytic activity. *Biochemical journal*, 406(2):203-207.
- Liao, J.K., Shin, W.S., Lee, W.Y. & Clark, S.L. 1995. Oxidized low-density lipoprotein decreases the expression of endothelial nitric oxide synthase. *Journal of biological chemistry*, 270(1):319-324.
- Libby, P. 2002. Inflammation in atherosclerosis. *Nature*, Dec(420):868-874.
- Libby, P., Okamoto, Y., Rocha, V.Z. & Folco, E. 2010. Inflammation in atherosclerosis. *Circulation Journal*, 74:213-220.
- Libby, P. 2013. Mechanisms of acute coronary syndromes and their implications for therapy. *New England journal of medicine*, 368(21):2004-2013.
- Lind, S., Eriksson, M., Rystedt, E., Wiklund, O., Angelin, B. & Eggertsen, G. 1998. Low frequency of the common Norwegian and Finnish LDL-receptor mutations in Swedish patients with familial hypercholesterolaemia. *Journal of internal medicine*, 243(7):19-25.
- Lindgren, V., Luskey, K.L., Russell, D.W. & Francke, U. 1985. Human genes involved in cholesterol metabolism: Chromosomal mapping of the loci for the low density lipoprotein receptor and 3-hydroxy-3-methylglutaryl-coenzyme A reductase with cDNA probes. *Proceedings of the National Academy of Sciences*, 82(24):8567-8571.

- Llorente-Cortés, V. & Badimon, L. 2005. LDL Receptor–related protein and the vascular wall: Implications for atherothrombosis. *Arteriosclerosis, thrombosis, and vascular biology*, 25(3):497-504.
- Lombardi, P., Sijbrands, E., van de Giessen, K., Smelt, A., Kastelein, J., Frants, R. & Havekes, L. 1995. Mutations in the low density lipoprotein receptor gene of familial hypercholesterolemic patients detected by denaturing gradient gel electrophoresis and direct sequencing. *Journal of lipid research*, 36(4):860-867.
- Lopez, D. 2008. PCSK9: An enigmatic protease. *Biochimica et biophysica acta (BBA) - molecular and cell biology of lipids*, 1781(4):184-191.
- MacIntyre, U.E., Kruger, H.S., Venter, C.S. & Vorster, H.H. 2002. Dietary intakes of an african population in different stages of transition in the North West province, South Africa: The THUSA study. *Nutrition research*, 22(3):239-256.
- Mamo, Y. & Oli, K. 2001. Trends of acute myocardial infarction admissions over a decade, Tikur Anbessa hospital. *Ethiopian medical journal*, 39(3):193-202.
- Maschberger, P., Bauer, M., Baumann-Siemons, J., Zangl, K.J., Negrescu, E.V., Reininger, A.J. & Siess, W. 2000. Mildly oxidized low density lipoprotein rapidly stimulates via activation of the lysophosphatidic acid receptor src family and syk tyrosine kinases and Ca²⁺ influx in human platelets. *Journal of biological chemistry*, 275(25):19159-19166.
- Mayne, J., Raymond, A., Chaplin, A., Cousins, M., Kaefer, N., Gyamera-Acheampong, C., Seidah, N.G., Mbikay, M., Chrétien, M. & Ooi, T.C. 2007. Plasma PCSK9 levels correlate with cholesterol in men but not in women. *Biochemical and biophysical research communications*, 361(2):451-456.
- McNamara, J.R., Campos, H., Ordovas, J.M., Peterson, J., Wilson, P.W. & Schaefer, E.J. 1987. Effect of gender, age, and lipid status on low density lipoprotein subfraction distribution. results from the Framingham offspring study. *Arteriosclerosis, thrombosis, and vascular biology*, 7(5):483-490.
- Michiels, C. 2003. Endothelial cell functions. *Journal of cellular physiology*, 196(3):430-443.

- Moons, A.H., Levi, M. & Peters, R.J. 2002. Tissue factor and coronary artery disease. *Cardiovascular research*, 53(2):313-325.
- Mozas, P., Castillo, S., Tejedor, D., Reyes, G., Alonso, R., Franco, M., Saenz, P., Fuentes, F., Almagro, F., Mata, P. & Pocoví, M. 2004. Molecular characterization of familial hypercholesterolemia in Spain: Identification of 39 novel and 77 recurrent mutations in LDLR. *Human mutation*, 24(2):187-187.
- Napoli, C., de Nigris, F., Williams-Ignarro, S., Pignalosa, O., Sica, V. & Ignarro, L.J. 2006. Nitric oxide and atherosclerosis: An update. *Nitric oxide*, 15(4):265-279.
- Narcisi, T.M., Shoulders, C.C., Chester, S.A., Read, J., Brett, D.J., Harrison, G.B., Grantham, T.T., Fox, M.F., Povey, S. & de Bruin, T.W. 1995. Mutations of the microsomal triglyceride-transfer-protein gene in abetalipoproteinemia. *American journal of human genetics*, 57(6):1298-1310.
- Nassoury, N., Blasiolo, D.A., Tebon Oler, A., Benjannet, S., Hamelin, J., Poupon, V., McPherson, P.S., Attie, A.D., Prat, A. & Seidah, N.G. 2007. The cellular trafficking of the secretory proprotein convertase PCSK9 and its dependence on the LDLR. *Traffic*, 8(6):718-732.
- Nauck, M., Köster, W., Dörfer, K., Eckes, J., Scharnagl, H., Gierens, H., Nissen, H., Nauck, M., Wieland, H. & März, W. 2001. Identification of recurrent and novel mutations in the LDL receptor gene in German patients with familial hypercholesterolemia. *Human mutation*, 18(2):165-166.
- Niesen, M., Bedi, M. & Lopez, D. 2008. Diabetes alters LDL receptor and PCSK9 expression in rat liver. *Archives of biochemistry and biophysics*, 470(2):111-115.
- Nordlie, M.A., Wold, L.E. & Kloner, R.A. 2005. Genetic contributors toward increased risk for ischemic heart disease. *Journal of molecular and cellular cardiology*, 39(4):667-679.
- Olofsson, S., Asp, L. & Borén, J. 1999. The assembly and secretion of apolipoprotein B-containing lipoproteins. *Current opinion in lipidology*, 10(4):341-346.
- Omran, A.R. 1971. The epidemiologic transition: A theory of the epidemiology of population change. *The Milbank Memorial Fund quarterly*, 49(4, Part 1):pp. 509-538.

- Öörni, K., Pentikäinen, M.O., Ala-Korpela, M. & Kovanen, P.T. 2000. Aggregation, fusion, and vesicle formation of modified low density lipoprotein particles: Molecular mechanisms and effects on matrix interactions. *Journal of lipid research*, 41(11):1703-1714.
- Oosthuizen, W., Vorster, H.H., Kruger, A., Venter, C.S., Kruger, H.S. & de Ridder, J.H. 2002. Impact of urbanisation on serum lipid profiles--the THUSA survey. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*, 92(9):723-728.
- Orr, A.W., Stockton, R., Simmers, M.B., Sanders, J.M., Sarembock, I.J., Blackman, B.R. & Schwartz, M.A. 2007. Matrix-specific p21-activated kinase activation regulates vascular permeability in atherogenesis. *The journal of cell biology*, 176(5):719-727.
- Packard, C. & Shepherd, J. 1999. Physiology of the lipoprotein transport system: An overview of lipoprotein metabolism. (In Betteridge, D.J., Illingworth, D.R. & Shepherd, J., eds. *Lipoproteins in health and disease*. 1st ed. Arnold Publishers. p. 17-30).
- Peer, N., Steyn, K., Lombard, C., Gaziano, T. & Levitt, N. 2013. Alarming rise in prevalence of atherogenic dyslipidaemia in the black population of cape town: The cardiovascular risk in black South Africans (CRIBSA) study. *European journal of preventive cardiology*, .
- Peeters, A.V., Kotze, M.J., Scholtz, C.L., De Waal, L.F., Rubinsztejn, D.C., Coetzee, G.A., Zuliani, G., Streiff, R., Liu, J. & van der Westhuyzen, D.R. 1998. A 3-basepair deletion in repeat 1 of the LDL receptor promoter reduces transcriptional activity in a South African Pedi. *Journal of lipid research*, 39(5):1021-1024.
- Peterson, E.D., Pollack, C.V., Roe, M.T., Parsons, L.S., Littrell, K.A., Canto, J.G. & Barron, H.V. 2003. Early use of glycoprotein IIb/IIIa inhibitors in non-ST-elevation acute myocardial infarction Observations from the national registry of myocardial infarction 4. *Journal of the American college of cardiology*, 42(1):45-53.
- Pischon, T., Girman, C.J., Sacks, F.M., Rifai, N., Stampfer, M.J., Rimm, E.B. 2005. Non-high-density lipoprotein cholesterol and apolipoprotein B in the prediction of coronary heart disease in men. *Circulation*, 112(22):3375– 3383.
- Piper, D., Jackson, , Liu, , Romanow, W., Shetterly, , Thibault, S., Shan, & Walker, N. 2007. The crystal structure of PCSK9: A regulator of plasma LDL-cholesterol. *Structure*, 15(5):545-552.

- Poirier, S., Mayer, G., Benjannet, S., Bergeron, E., Marcinkiewicz, J., Nassoury, N., Mayer, H., Nimpf, J., Prat, A. & Seidah, N.G. 2008. The proprotein convertase PCSK9 induces the degradation of low density lipoprotein receptor (LDLR) and its closest family members VLDLR and ApoER2. *Journal of biological chemistry*, 283(4):2263-2372.
- Popkin, B.M. 2002. The shift in stages of the nutrition transition in the developing world differs from past experiences! *Public health nutrition*, 5(1A):205-214.
- Popkin, B.M. & Caballero, B., eds. 2002. The nutrition transition: Diet and disease in the developing world. 1st ed. London: Academic Press.
- Popkin, B.M. 1998. The nutrition transition and its health implications in lower-income countries. *Public health nutrition*, 1(01):5-21.
- Pritchard, K.A., Groszek, L., Smalley, D.M., Sessa, W.C., Wu, M., Villalon, P., Wolin, M.S. & Stemerman, M.B. 1995. Native low-density lipoprotein increases endothelial cell nitric oxide synthase generation of superoxide anion. *Circulation research*, 77(3):510-518.
- Reiner, Ž., Catapano, A.L., De Backer, G., Graham, I., Taskinen, M., Wiklund, O., Agewall, S., Alegria, E., Chapman, M.J. & Durrington, P. 2011. ESC/EAS guidelines for the management of dyslipidaemias the task force for the management of dyslipidaemias of the European society of cardiology (ESC) and the European atherosclerosis society (EAS). *European heart journal*, 32(14):1769-1818.
- Rudenko, G., Henry, L., Henderson, K., Ichtchenko, K., Brown, M.S., Goldstein, J.L. & Deisenhofer, J. 2002. Structure of the LDL receptor extracellular domain at endosomal pH. *Science*, 298(5602):2353-2358.
- Russell, D.W., Yamamoto, T., Schneider, W.J., Slaughter, C.J., Brown, M.S. & Goldstein, J.L. 1983. cDNA cloning of the bovine low density lipoprotein receptor: Feedback regulation of a receptor mRNA. *Proceedings of the national academy of sciences*, 80(24):7501-7505.
- Scholtz, C., Peeters, A., Hoogendijk, C., Thiart, R., Villiers, J., Hillermann, R., Liu, J., Marais, A. & Kotze, M. 1999. Mutation -59c-->t in repeat 2 of the LDL receptor promoter: Reduction in transcriptional activity and possible allelic interaction in a South African family with familial hypercholesterolaemia. *Human molecular genetics*, 8(11):2025-2030.

- Schuster, H., Keller, C., Wolfram, G. & Zollner, N. 1995. Ten LDL receptor mutants explain one third of familial hypercholesterolemia in a German sample. *Arteriosclerosis, thrombosis, and vascular biology*, 15(12):2176-2180.
- Seidah, N.G., Benjannet, S., Wickham, L., Marcinkiewicz, J., Jasmin, S.B., Stifani, S., Basak, A., Prat, A. & Chrétien, M. 2003. The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): Liver regeneration and neuronal differentiation. *Proceedings of the National Academy of Sciences*, 100(3):928-933.
- Shah, P.K., Falk, E., Badimon, J.J., Fernandez-Ortiz, A., Mailhac, A., Villareal-Levy, G., Fallon, J.T., Regnstrom, J. & Fuster, V. 1995. Human monocyte-derived macrophages induce collagen breakdown in fibrous caps of atherosclerotic plaques. potential role of matrix-degrading metalloproteinases and implications for plaque rupture. *Circulation*, 92(6):1565-1569.
- Sims, P.J. & Wiedmer, T. 2001. Unraveling the mysteries of phospholipid scrambling. *Thrombosis and haemostasis-Stuttgart-*, 86(1):266-275.
- Sliwa, K., Wilkinson, D., Hansen, C., Ntyintyane, L., Tibazarwa, 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.
- Smith, E., Thompson, W., Crosbie, L. & Stirk, C. 1992. Fibrinogen/fibrin in atherogenesis. *European journal of epidemiology*, 8(1):83-87.
- Soutar, A.K. 1999. Low-density lipoprotein receptors. (In Betteridge, D.J., Illingworth, D.R. & Shepherd, J., eds. *Lipoproteins in health and disease*. first ed. Arnold Publishers. p. 303-322).
- Soutar, A.K., Naoumova, R.P. & Traub, L.M. 2003. Genetics, clinical phenotype, and molecular cell biology of autosomal recessive hypercholesterolemia. *Arteriosclerosis, thrombosis, and vascular biology*, 23(11):1963-1970.
- Soutar, A.K. & Naoumova, R.P. 2007. Mechanisms of disease: Genetic causes of familial hypercholesterolemia. *Nature clinical practice cardiovascular medicine*, 4(4):214-225.
- Steinberg, D. 2002. Atherogenesis in perspective: Hypercholesterolemia and inflammation as partners in crime. *Nature medicine*, 8(11):1211-1217.

- Stewart, S., Carrington, M., Pretorius, S., Methusi, P. & Sliwa, K. 2011. Standing at the crossroads between new and historically prevalent heart disease: Effects of migration and socio-economic factors in the Heart of Soweto cohort study. *European heart journal*, 32(4):492-499.
- Steyn, K., Jooste, P.L., Bourne, L., Fourie, J., Badenhorst, C.J., Bourne, D.E., Langenhoven, M.L., Lombard, C.J., Truter, H. & Katzenellenbogen, J. 1991. Risk factors for coronary heart disease in the black population of the Cape Peninsula. the BRISK study. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*, 79(8):480-485.
- Steyn, K., Sliwa, K., Hawken, S., Commerford, P., Onen, C., Damasceno, A., Ounpuu, S., Yusuf, S. & for the INTERHEART Investigators in Africa,. 2005. Risk factors associated with myocardial infarction in Africa: The INTERHEART Africa study. *Circulation*, 112(23):3554-3561.
- Stubbs, C.O. & Lee, A.J. 2004. The obesity epidemic: both energy intake and physical activity contribute. *Medical journal of Australia*, 181(9):489-491.
- Südhof, T.C., Goldstein, J.L., Brown, M.S. & Russell, D.W. 1985. The LDL receptor gene: A mosaic of exons shared with different proteins. *Science*, 228(4701):815-822.
- Südhof, T.C., Russell, D.W., Brown, M.S. & Goldstein, J.L. 1987. 42 bp element from LDL receptor gene confers end-product repression by sterols when inserted into viral TK promoter. *Cell*, 48(6):1061-1069.
- Südhof, T.C., Russell, D.W., Goldstein, J.L., Brown, M.S., Sanchez-Pescador, R. & Bell, G.I. 1985. Cassette of eight exons shared by genes for LDL receptor and EGF precursor. *Science*, 228(4701):893-895.
- Temple, N.J. & Steyn, N.P. 2011. The cost of a healthy diet: A South African perspective. *Nutrition*, 27(5):505-508.
- Thiart, Scholtz, Vergotine, Hoogendijk, de Villiers, Nissen, Brusgaard, Gaffney, Hoffs, Vermaak & Kotze. 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.

- 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.
- Tolleshaug, H., Goldstein, J.L., Schneider, W.J. & Brown, M.S. 1982. Posttranslational processing of the LDL receptor and its genetic disruption in familial hypercholesterolemia. *Cell*, 30(3):715-724.
- Tosi, I., Toledo-Leiva, P., Neuwirth, C., Naoumova, R.P. & Soutar, A.K. 2007. Genetic defects causing familial hypercholesterolaemia: Identification of deletions and duplications in the LDL-receptor gene and summary of all mutations found in patients attending the Hammersmith Hospital lipid clinic. *Atherosclerosis*, 194(1):102-111.
- Van Gaal, L.F., Peeters, A.V., De Block, C.E.M., de Leeuw, I.H., Thiart, R. & Kotze, M.J. 2001. Low-density lipoprotein receptor gene mutation analysis and clinical correlation in Belgian hypercholesterolaemics. *Molecular and cellular probes*, 15(6):329-336.
- Vidal, F., Colomé, C., Martínez-González, J. & Badimon, L. 1998. Atherogenic concentrations of native low-density lipoproteins down-regulate nitric-oxide-synthase mRNA and protein levels in endothelial cells. *European journal of biochemistry*, 252(3):378-384.
- Virmani, R., Kolodgie, F.D., Burke, A.P., Finn, A.V., Gold, H.K., Tulenko, T.N., Wrenn, S.P. & Narula, J. 2005. Atherosclerotic plaque progression and vulnerability to rupture angiogenesis as a source of intraplaque hemorrhage. *Arteriosclerosis, thrombosis, and vascular biology*, 25(10):2054-2061.
- Vorster, H.H., Kruger, A. & Margetts, B.M. 2011. The nutrition transition in Africa: Can it be steered into a more positive direction? *Nutrients*, 3(4):429-441.
- Vorster, H.H., Oosthuizen, W., Jerling, J.C., Veldman, F.J., Burger, H.M. & Mclachlan, M. 1997. The nutritional status of South Africans: A review of the literature from 1975-1996.
- Vorster, H.H., Venter, C.S., Kruger, H.S., Kruger, A., Malan, N.T., Wissing, M.P., De Ridder, J.H., Veldman, F.J., Steyn, H.S. & Margetts, B.M. 2000. The impact of urbanization on physical, physiological and mental health of Africans in the North West province of South Africa: The THUSA study. *South African journal of science*, 96(9/10):505-514.

- 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*, 8(5):480-490.
- Vorster, H.H. 2002. The emergence of cardiovascular disease during urbanisation of Africans. *Public health nutrition*, 5(1A):239-243.
- Vu, T.H., Hung, D.T., Wheaton, V.I. & Coughlin, S.R. 1991. Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell*, 64(6):1057-1068.
- Walker, A.R.P. & Sareli, P. 1997. Coronary heart disease: Outlook for Africa. *Journal of the Royal Society of Medicine*, 90(Jan):23-27.
- Webb, J., Sun, X., McCarthy, S., Neuwirth, C., Thompson, G., Knight, B. & Soutar, A. 1996. Characterization of mutations in the low density lipoprotein (LDL)- receptor gene in patients with homozygous familial hypercholesterolemia, and frequency of these mutations in FH patients in the United Kingdom. *Journal of lipid research*, 37(2):368-381.
- Weisgraber, K.H., Innerarity, T.L. & Mahley, R.W. 1982. Abnormal lipoprotein receptor-binding activity of the human E apoprotein due to cysteine-arginine interchange at a single site. *The journal of biological chemistry*, 257(5):2518-2521.
- Whittall, R., Matheus, S., Cranston, T., Miller, G. & Humphries, S. 2002. The intron 14 2140 5G> A variant in the low density lipoprotein receptor gene has no effect on plasma cholesterol levels. *Journal of medical genetics*, 39(9):e57-e57.
- Whitfield, A.J., Barrett, P.H., van Bockxmeer, F.M. & Burnett, J.R. 2004. Lipid disorders and mutations in the APOB gene. *Clinical chemistry*, 50(10):1725-1732.
- WHO. 2013. Cardiovascular diseases.
<http://www.who.int/mediacentre/factsheets/fs317/en/index.html> Date of access: 08/13. 2013.
- Woolf, N. 1999. Pathology of atherosclerosis. (In Betteridge, D.J., Illingworth, D.R. & Shepherd, J., eds. Lipoproteins in health and disease. 1st ed. Arnold Publishers. p. 533-539).

- Yamamoto, T., Davis, C.G., Brown, M.S., Schneider, W.J., Casey, M.L., Goldstein, J.L. & Russell, D.W. 1984. The human LDL receptor: A cysteine-rich protein with multiple alu sequences in its mRNA. *Cell*, 39(1):27-38.
- Yusuf, S., Reddy, S., Ounpuu, S. & Anand, S. 2001. Global burden of cardiovascular diseases: Part I: General considerations, the epidemiologic transition, risk factors, and impact of urbanization. *Circulation*, 104(22):2746-2753.
- Zaid, A., Roubtsova, A., Essalmani, R., Marcinkiewicz, J., Chamberland, A., Hamelin, J., Tremblay, M., Jacques, H., Jin, W., Davignon, J., Seidah, N.G. & Prat, A. 2008. Proprotein convertase subtilisin/kexin type 9 (PCSK9): Hepatocyte-specific low-density lipoprotein receptor degradation and critical role in mouse liver regeneration. *Hepatology*, 48(2):646-654.
- Zhang, D., Garuti, R., Tang, W., Cohen, J.C. & Hobbs, H.H. 2008. Structural requirements for PCSK9-mediated degradation of the low-density lipoprotein receptor. *Proceedings of the national academy of sciences*, 105(35):13045-13050.
- Zhang, D., Lagace, T.A., Garuti, R., Zhao, Z., McDonald, M., Horton, J.D., Cohen, J.C. & Hobbs, H.H. 2007. Binding of proprotein convertase Subtilisin/Kexin type 9 to epidermal growth factor-like repeat A of low density lipoprotein receptor decreases receptor recycling and increases degradation. *Journal of biological chemistry*, 282(25):18602-18612.
- Zhao, Z., Tuakli-Wosornu, Y., Lagace, T.A., Kinch, L., Grishin, N.V., Horton, J.D., Cohen, J.C. & Hobbs, H.H. 2006a. Molecular characterization of loss-of-function mutations in PCSK9 and identification of a compound heterozygote. *American journal of human genetics*, 79(3):514.

ANNEXURE A: ETHICAL APPROVAL



YUNIBESITHI YA BOKONE-BOPHIRIMA
NORTH WEST UNIVERSITY
NOORDWES UNIVERSITEIT

Etiekomitee
Tel (018) 299 2558
Faks (018) 297 5308
E-Pos drvealn@puk.ac.za

Dr A Kruger
Bussie 594
Noordwes-Universiteit
(Potchefstroomkampus)

2 September 2004

Geagte dr Kruger

GOEDKEURING VIR EKSPERIMENTERING MET MENSE

Hiermee wens ek u in kennis te stel dat u projek getiteld "PURE study (Prospective Urban and Rural Epidemiology study)" deur die Etiekomitee goedgekeur is met nommer 04M10.

Gebruik asseblief die nommer genoem in paragraaf 1 in alle korrespondensie rakende bogenoemde projek en let daarop dat daar van projekteleiers verwag word om jaarliks in Junie aan die Etiekomitee verslag te doen insake etiese aspekte van hulle projekte asook van publikasies wat daaruit voortgespruit het. U sal in Mei 2005 die dokumentasie hieroor ontvang.

Goedkeuring van die Etiekomitee is vir 'n termyn van hoogstens 5 jaar geldig (volgens Senaatsbesluit van 4 November 1992, art 9.13.2). Vir die voortsetting van projekte na verstryking van hierdie tydperk moet opnuut goedkeuring verkry word.

Die Etiekomitee wens u alle voorspoed met u werk toe.

Vriendelike groete


PROF. NT MALAN
VOORSITTER: ETIEKKOMITEE



POTCHEFSTROOMKAMPUS
Privaatsak X6001, Potchefstroom, Suid-Afrika, 2520
Tel: (018) 299-1111 • Faks: (018) 299-2799
Internet: <http://www.nwu.ac.za>



ANNEXURE B: INFORMED CONSENT FORM

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

ANNEXURE C: INFORMED CONSENT FORM

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	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px;">Yes</td> <td style="padding: 2px 10px;">No</td> </tr> </table>	Yes	No
Yes	No		
I want to know my HIV-status	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px;">Yes</td> <td style="padding: 2px 10px;">No</td> </tr> </table>	Yes	No
Yes	No		
I agree to give a blood sample	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px;">Yes</td> <td style="padding: 2px 10px;">No</td> </tr> </table>	Yes	No
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.

ANNEXURE C (continued)

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)

ANNEXURE D: VARIANTS IDENTIFIED THROUGH SEQUENCING IN THE LDLR GENE

	rs NUMBER	DOMAIN	POSITION	DNA POSITION	FREQ ^a
1	rs369850745 - Novel	Promotor	11199911	c.-314C>T	3
2	rs17249141	Promotor	11200008	c.-217C>T	4
3	rs17242759	INTRON1	11200309	C.67+18C>A	6
4	rs12981050	INTRON1	11200412	C.67+121C>T	4
5	Novel 2	INTRON1	11200485	C.67+194C>A	25
6	rs72658855	SIGNAL SEQUENCE (Exon 1)	11210921	c.90C>T (Asn30Asn)	2
7	rs2228671	REPEAT 1 (Exon 2)	11210912	c.81C>T (Cys27Cys)	2
8	rs3745677	INTRON 2	11211077	c.190+56G>A	7
9	rs3745678	INTRON 2	11211079	c.190+58C>T	3
10	rs190990789	INTRON 2	11213237	c.191-103C>T	3
11	rs10423288	INTRON 3	11215846	c.314-50T>C	9
12	rs148698650	REPEAT 7	11218079	c.829G>A (Glu277Lys)	2
13	rs116405216	INTRON 6	11221324	c.941-4G>A	2
14	Novel 3	INTRON 6	11221306	c.941-22T>C	4
15	rs2738442	INTRON 7	11221454	c.1060+7T>C	58
16	rs12710260	INTRON 7	11221457	c.1060+10G>C	2
17	rs41301947	INTRON 7	11222108	c.1061-8T>C	1
18	rs72658861	INTRON 7	11222182	c.1061-82G>C	1
19	rs139066906	REPEAT B	11222296	c.1167G>A	2
20	rs11669576	REPEAT B	11222300	c.1171G>A (Ala391Thr)	10
21	rs72658862	INTRON 8	11223929	c.1187-25C>T	4
22	rs6413505	INTRON 9	11224157	c.1358+32C>T	16
23	rs1003723	INTRON 9	11224181	c.1359-30C>T	2
24	rs5930	EFG spacer	11224265	c.1413A>G (Arg471Arg)	54
25	rs1569372	INTRON 10	11224491	c.1586+53A>G	26
26	rs17248833	INTRON 10	11224592	c.1586+154A>G	6
27	rs142460391	INTRON 10	11224888	c.1586+450G>C	3
28	rs2738444	INTRON 10	11224931	c.1586+493C>T	30
29	rs3826807	INTRON 10	11224932	c.1586+494G>A	12
30	rs114891301	INTRON 10	11224454	c.1586+16G>A	1
31	rs5929	EFG spacer	11226800	c.1617C>T (Pro539Pro)	6
32	rs4508523	INTRON 11	11226944	c.1705+56C>T	22
33	rs41307025	INTRON 11	11227454	c.1706-81C>T	1
34	rs7259278	INTRON 11	11227466	c.1706-69G>T	20
35	rs2738447	INTRON 11	11227480	c.1706-55A>C	51
36	rs1799898	EFG spacer	11227554	c.1725C>T (Leu554Leu)	3
37	rs688	EFG spacer	11227602	c.1773C>T (Asn591Asn)	3
38	rs5925	EFG spacer	11230881	c.1959T>C (Val653Val)	14
39	rs5927	OLS	11233941	c.2232A>G (Arg744Arg)	43
40	rs369402076 - Novel 4	INTRON 15	11234029	c.2311+9T>G	1
41	rs17249358	INTRON 15	11238613	c.2312-71G>A	1

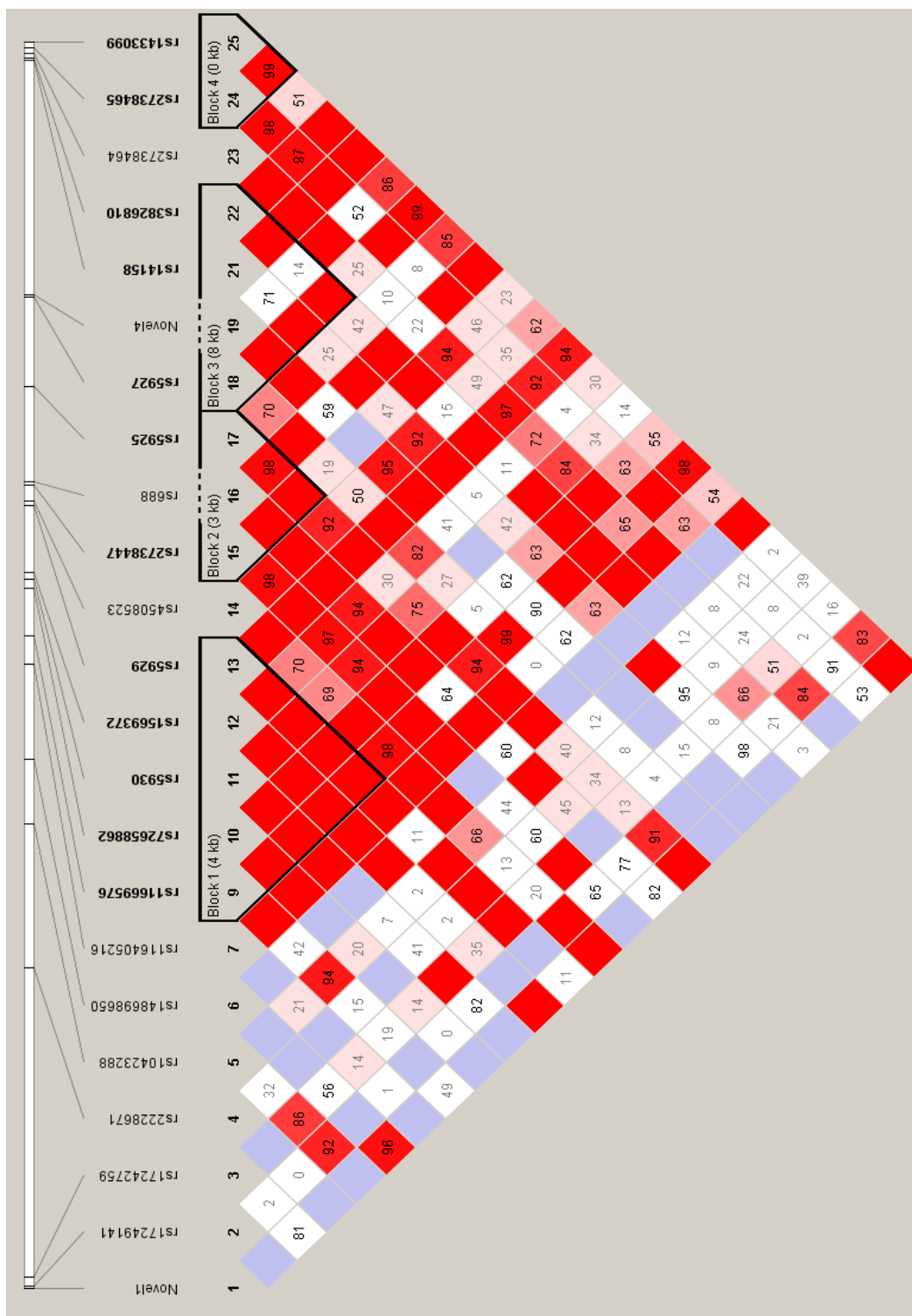
rs NUMBER	DOMAIN	POSITION	DNA POSITION	FREQ ^a
42 Novel 5	INTRON 16	11238780	c.2389+19C>T	1
43 rs72658868	INTRON 16	11238802	c.2389+41C>A	1
44 rs2738460	INTRON 16	11238807	c.2389+46C>T	7
45 rs13306501	INTRON 16	11238808	c.2389+47G>A	8
46 rs2304182	INTRON 16	11238877	c.2389+116G>A	15
47 rs6413503	INTRON 17	11241904	c.2548-42G>A	6
48 rs6413504	INTRON 17	11241915	c.2548-56A>G	16
49 rs14158	3'UTR	11242044	c.*52G>A	2
50 rs3826810	3'UTR	11242133	c.*141G>A	10
51 rs2738464	3'UTR	11242307	c.*315G>C	36
52 Novel 6	3'UTR	11242369	c.*377G>A	1
53 rs2738465	3'UTR	11242469	c.*504G>A	12
54 rs1433099	3'UTR	11242658	c.*666T>C	31
55 rs2738466	3'UTR	11242765	c.*773A>G	3
56 rs17242677	3'UTR	11242795	c.*803C>T	1
57 rs10409044	3'UTR	11242974	c.*982G>C	5
58 rs3180023	3'UTR	11243209	c.*1217C>A	2
59 NONE	3'UTR	11243211	c.*1218_1219 G insertion	1
60 rs2915966	3'UTR	11243254	c.*1262T>C	9
61 rs2978615	3'UTR	11243260	c.*1268C>T	2
62 rs10415069	3'UTR	11243324	c.*1332G>A	1
63 rs7258146	3'UTR	11243346	c.*1354T>C	6
64 rs7254521	3'UTR	11243422	c.*1430C>T	5
65 rs5742911	3'UTR	11243445	c.*1453A>G	2
66 rs17249057	3'UTR	11243502	c.*1510T>C	2
67 rs17249064	3'UTR	11243592	c.*1600G>T	2
68 rs2738467	3'UTR	11243735	c.*1743C>T	1
69 rs28398082	3'UTR	11244046	c.*2054G>A	4
70 rs17249078	3'UTR	11244202-11244203	c.*2209_2210 TA deletion	25

^aFrequency of variant in sub-sample of 60 chromosomes

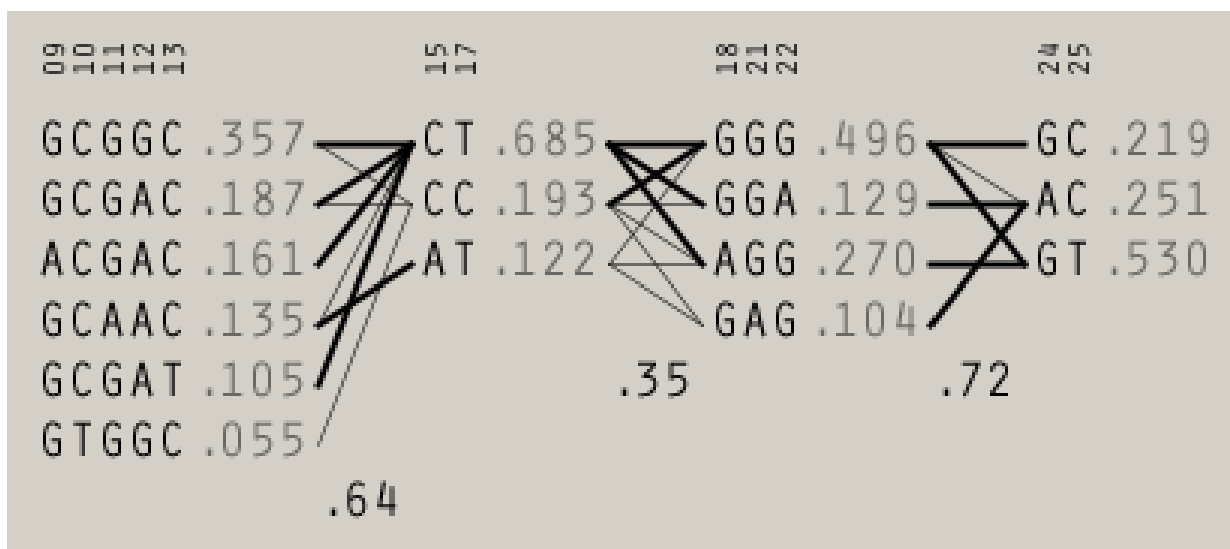
ANNEXURE E: PRIMERS DESIGNED FOR LDLR PCR AND SEQUENCING

Exon		Primer	Bases	Product length	
1	F-	TTTCAGAAGATGCGTTTC	18	632	
	R-	TTCCCTTAAATCCCTCAG	18		
2	F-	TGGAATCAACAACACTAGC	19	762	
	R-	GAACAAGATAGAGGTGGTG	19		
3	F-	AGACAGGATTGGCAAGGCCAGTG	23	812	
	R-	AGGATCACGGGGATCACGACCA	22		
4	F-	CACCTATTAGCGCACCA	18	704	
	R-	AATCCACTTCGGCACCTA	18		
5	F-	CCGGCAAAAGGCCCTGCTTCTT	22	313	
	R-	AGACGCACAGCTGGTGCAGA	20		
6	F-	GGCTCAGACACACCTGACCTTCC	23	195	
	R-	CAAGCCGCCTGCACCGAGAC	20		
7	F-	GCCAAGGTTGGCGGCGAAGG	20	246	
	R-	TGCCATGTCAGGAAGCGCAG	20		
8	F-	TCTTGTTGGGTTCCCGTGGTG	22	492	
	R-	AAGTCCTAACAGGGCGGCAGAC	22		
9	F-	TGTGCCTGTTCCCGTTGGGA	20	627	
	R-	ACGATGGCCCTTGGCTTGGGA	20		
10	F-	TGCGGATACCAAGGGCGTGAA	21	688	
	R-	TGCAAGGACAGTGTGCGTCC	20		
11	F-	GGGTTCCCAGCAGGACTATTTCCCA	25	345	
	R-	TTCAGGGAGCAGCTTGGGCTTGTCC	25		
12	F-	TGCAACTCCCCTACCTGCCCATT	22	426	
	R-	TTCTGCGTTCATCTTGGCTTGAGT	23		
13 and 14	F-	AGAGGGTGGCCTGTGTCTCATC	22	668	
	R-	TGGGTAAATGTCAAGCCCGGTGC	23		
15	F-	GCACGTGGCACTCAGAAGACGTTT	24	306	
	R-	ACGACACCTGGACTCCATCTCGT	23		
16	F-	TAAGCCCGTGTGGCCTCTCACA	22	400	
	R-	TGTCCTGTGAGCCCAAGCCA	22		
17	F-	GCCCGTGTTCCTACTCCAGCCA	22	409	
	R-	ACCAAGGCCATTGTCCGCTGA	21		
18a	F-	TGGGGGAACATGCTTGGGGATCA	23	770	
*first part of exon	R-	GCCACCACGGATTGAGCCAGAT	22		
18	F-	TGGGGGAACATGCTTGGGGATCA	23	2588	
*whole exon	R-	ACACCCATCTCCCAGAAGCCACTC	24		
18b	F-	ATCTGGCTGAATCCGTGGTGGC	22	502	
*second part of exon	R-	CGACAAGATTGGGGGAAGTGAATGGCTT	27		*nested PCR
18c	F-	CGCCCACCTAGTGCTTCCACTTCT	24	971	
*third part of exon	R-	ACCTGAAGTCCCGTCAAACGATCCA	25		*nested PCR
18d	F-	TCGTTTGACGGGACTTCAGGTTCT	24	501	
*last part of exon	R-	ACACCCATCTCCCAGAAGCCACTC	24		*nested PCR

ANNEXURE F: LD PLOT FROM HAPLOVIEW FOR THE LDLR GENE



ANNEXURE G: HAPLOTYPES IDENTIFIED IN THE LDLR GENE BY MEANS OF HAPLOVIEW



ANNEXURE H: VARIANTS IDENTIFIED THROUGH SEQUENCING IN THE PCSK9 GENE

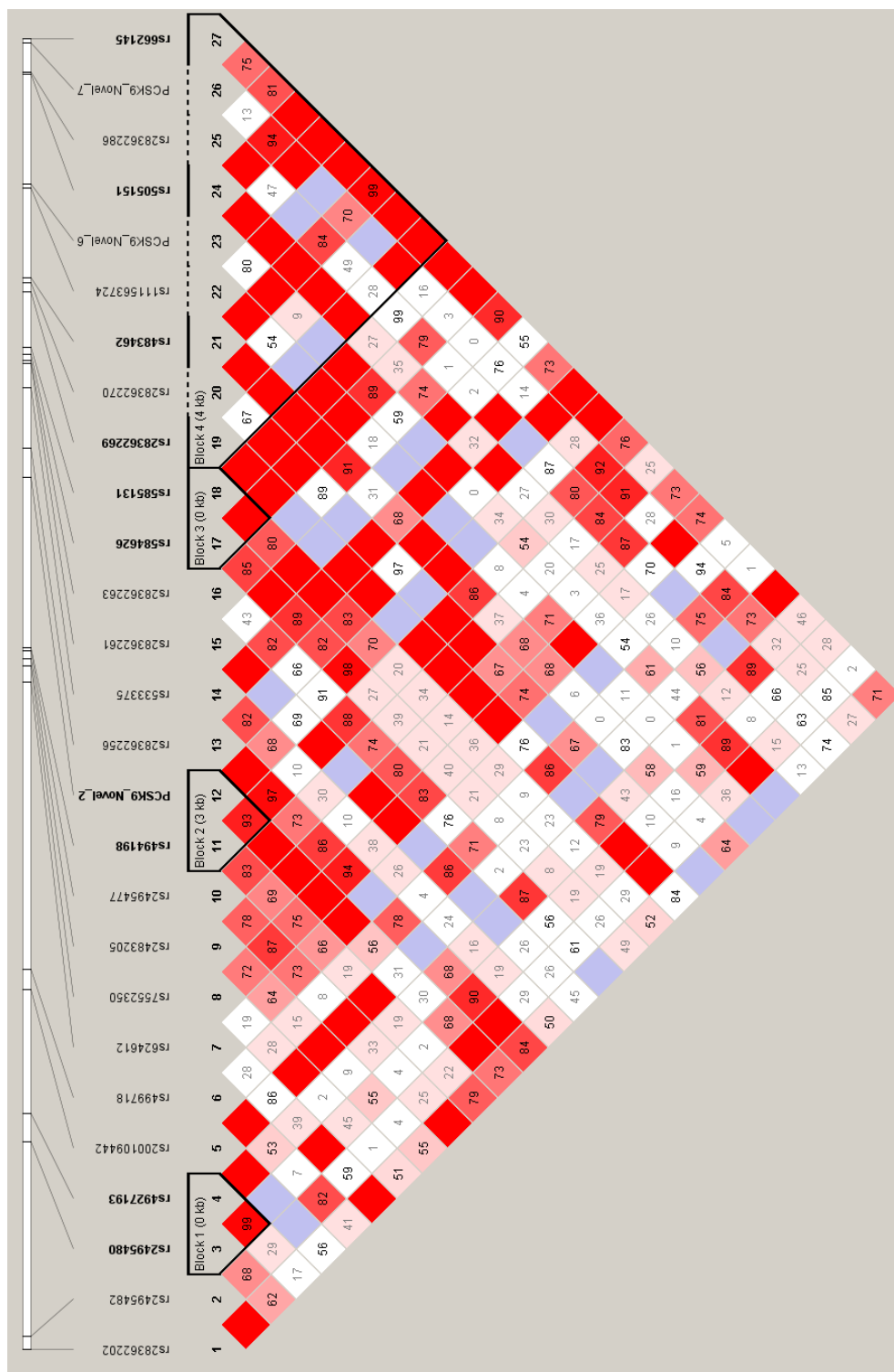
	rs NUMBER	DOMAIN	POSITION	DNA POSITION	FREQa
1	rs28362201	5'UTR	55505266	c.-245 G>T	5
2	rs28362202	5'UTR	55505485	c.-26 G>A	4
3	Novel 1	Signal peptide	55505537	c.27C>T (S27S)	4
4	rs45454392	Signal peptide	55505555- 55505554	c.43_44insGCT (Leu21_Leu22insLeu)	1
5	rs2495482	INTRON 1	55505732	c.207+15A>G	53
6	rs10888896	INTRON 1	55509219	c.208-303G>C	42
7	rs113138552	INTRON 1	55509223	c.208-293G>A	2
8	rs2495480	INTRON 1	55509355	c.208-161C>T	52
9	rs151193009	PRO-DOMAIN	55509585	c.277C>T (Arg93Cys)	2
10	rs4927193	INTRON 2	55509872	c.399+165T>C	22
11	rs2479411	INTRON 2	55509900	c.399+193C>T	14
12	rs4275490	INTRON 2	55509939	c.399+232T>C	4
13	rs200109442	INTRON 2	55512172	c.400-24C>T	3
14	rs72646505	INTRON 3	55512365	c.523+46C>A	5
15	rs499718	INTRON 3	55512549	c.523+230T>C	31
16	rs613855	INTRON 3	55517861	c.524-90C>G	14
17	rs624612	INTRON 3	55517883	c.524-68C>G	55
18	rs11800231	INTRON 3	55517940	c.524-11G>A	8
19	rs72646506	INTRON 4	55518091	c.657+7A>C	1
20	rs11800243	INTRON 4	55518093	c.657+9G>A	2
21	rs11806638	INTRON 4	55518160	c.657+76C>A	13
22	rs625619	INTRON 4	55518166	c.657+82A>G	19
23	rs45597632	INTRON 4	55518191	c.657+107 -c del	0
24	rs7552350	INTRON 4	55518198	c.657+114C>A	10
25	rs11800265	INTRON 4	55518287	c.658-36G>A	1
26	rs2483205	INTRON 4	55518316	c.658-7C>T	23
27	rs7552471	CATALYTIC	55518370	c.705C>T (Ser235Ser)	3
28	rs2495477	INTRON 5	55518467	c.799+3A>G	42
29	rs494198	INTRON 5	55518528	c.799+64C>A	12
30	Novel 2	CATALYTIC	55521695	c.829G>C (V277L)	6
32	Novel 3	INTRON 6	55522033	c.996+171C>A	1
33	rs57780528	INTRON 6	55521910	c.996+48 49 del-cggaggg	50
34	rs45598138	INTRON 6	55522083	c.996+221A>C	2
35	rs479910	INTRON 6	55522141	c.996+279A>G	24
36	rs139457113	INTRON 6	55522229	c.996+367A>G	2
37	rs28362256	INTRON 6	55522241	c.996+379C>T	3
38	Novel 4	INTRON 6	55522982	c.997-22A>C	2
39	rs509504	CATALYTIC	55523033	c.1026A>G (Gln342Gln)	57
40	rs533375	INTRON 7	55523361	c.1180+174A>G	40
41	rs28362261	CATALYTIC	55523802	c.1274A>G (Asn425Ser)	1
42	rs28362262	C-TERMINAL	55523854	c.1326C>T (Ala442Ala)	1
43	rs28362263	C-TERMINAL	55523855	c.1327G>A (Ala443Thr)	2
44	rs72646516	INTRON 8	55523891	c.1354+9G>T	1
45	rs584626	INTRON 8	55523984	c.1354+102C>T	39
46	rs28362264	INTRON 8	55524084	c.1355-88C>A	5
	rs NUMBER	DOMAIN	POSITION	DNA POSITION	FREQa

47	rs28362265	INTRON 8	55524085	c.1355-87C>T	5
48	rs585131	INTRON 8	55524116	c.1355-56C>T	44
49	rs540796	C-TERMINAL	55524197	c.1380G>A (Val460Val)	16
50	rs141502002	C-TERMINAL	55524222	c.1405C>T (Arg469Trp)	3
51	rs562556	C-TERMINAL	55524237	c.1420G>A (Val469Ile)	44
52	rs11206516	INTRON 9	55524971	c.1504-188T>C	1
53	rs28362269	INTRON 9	55525143	c.1504-16C>T	1
54	rs111400659	C-TERMINAL	55525301	c.1646G>A (Arg549His)	3
55	rs28362270	C-TERMINAL	55525313	c.1658A>G (His553Arg)	5
56	rs45439391	INTRON 10	55525399	c.1681+63C>T	4
57	rs483462	INTRON 10	55525400	c.1681+64G>A	18
58	rs73646523	INTRON 10	55526981	c.1682-67T>C	1
59	Novel 5	INTRON 10	55527001	c.1682-47G>A	1
60	rs111563724	C-TERMINAL	55527062	c.1696T>C (Trp566Arg)	1
61	Novel 6	C-TERMINAL	55527147	c.1781C>A (Ala594Asp)	7
62	rs505151	C-TERMINAL	55529187	c.2009G>A (Gly670Glu)	14
63	rs28362286	C-TERMINAL	55529215	c.2037C>A(Cys679X)	2
64	rs13376071	untranslated-3	55529671	c.*414C.T	3
65	rs28362288	untranslated-3	55529701	c.*444G>C	8
66	Novel 7	untranslated-3	55529757	c.*500C>T	3
67	rs662145	untranslated-3	55529828	c.*571C>T	21
68	rs17111557	untranslated-3	55529871	c.*614C>T	4
69	rs28362292	untranslated-3	55530106	c.*849T>C	1
70	rs28362293	untranslated-3	55530227	c.*970C>T	3
71	Novel 8	untranslated-3	55529705+G	c.*448_*449insG	1

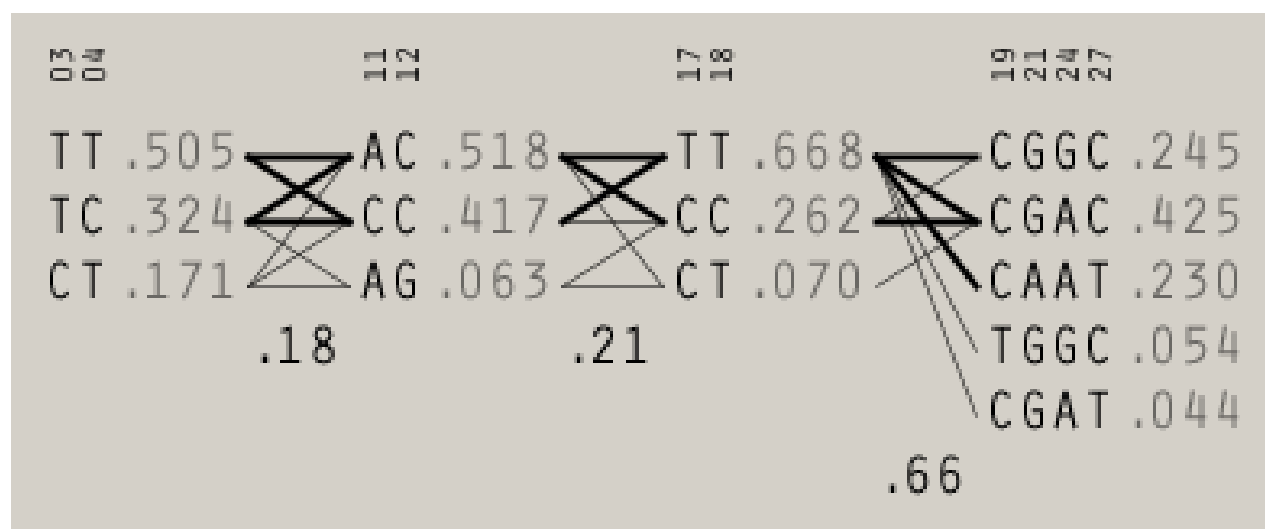
ANNEXURE I: PRIMERS DESIGNED FOR PCSK9 PCR AND SEQUENCING

Exon		Primer	Bases	Product length
1	F -	TTCCGTTAATGTTTAATCAG	20	702
	R -	CTCCACTTCCTCTCTTAC	18	702
2	F -	TTTCCCTCAATTTATTCTGAC	21	845
	R -	GTTTGATCCAGGCTCTAC	18	845
3	F -	AAACTCAGGCTCCTAGTC	21	600
	R -	TCAGATGAACCCTCCTTC	18	600
4 and 5	F -	ATTATAGAATAGGATGTAGTG	21	779
	R -	TTGGTTAGGAGACATTAG	18	779
6	F -	GAGAGTAACCTGCCTTACTCCG	25	868
	R -	TGAGCCCTCAACCGCTCCCA	20	868
7	F -	TAGAACCACTTGATACAC	18	770
	R -	AACTCCTATTCATCCATC	18	770
8 and 9	F -	ATCACCATCTTTCACCATTC	20	656
	R -	GGTACAGTCACCTCCATG	18	656
10	F -	TTAATGCTCGTGTGAGAG	18	675
	R -	GAGATGAGGAGAAGTAAGG	19	675
11	F -	TGGTTCAGACAGACCCCGGCTT	22	655
	R -	ATGGCGGTATGGTGGTGGCA	20	655
12a	F -	GGAGGGAGAAATGAAGTG	18	841
*first part of exon	R -	TAGATGCCATCCAGAAAG	18	841
12b	F -	ATTCAATCCTCAGGTCTC	18	998
*second part of exon	R -	TTGGCTTACTCAACTATTC	19	998

ANNEXURE J: LD PLOT FROM HAPLOVIEW FOR THE PCSK9 GENE



ANNEXURE K: HAPLOTYPES IDENTIFIED IN THE PCSK9 GENE BY MEANS OF HAPLOVIEW



ANNEXURE L: SUPPLEMENTARY DOCUMENTS FOR CHAPTER 5

Supplementary table1 Spearman rho's two-tailed correlation with LDL-C

Variable	Correlation Coefficient	Significance (2-tailed)
BMI (kg/m ²)	0.282	0.000
Gender	0.173	0.000
Tobacco history	0.160	0.000
Age (years)	0.111	0.000
CHO as % of TE	-0.081	0.001
Fat as % of TE	0.077	0.001
SATFAT as % of TE	0.072	0.002
CRP	0.067	0.004
MUFA as % of TE	0.062	0.008
FA C16_1 (g)	0.061	0.010
FA C18_3 (g)	0.061	0.009
Physical activity	-0.059	0.108
FA C20_4 (g)	0.056	0.018
FA C22_5 (g)	0.056	0.017
PUFA as % of TE	0.052	0.026
Cholesterol intake (g)	0.051	0.029
FA C14_0 (g)	0.051	0.031
FA C14_1 (g)	0.051	0.030
FA C10_0 (g)	0.046	0.050
FA C8_0 (g)	0.045	0.054
FA C16_0 (g)	0.044	0.061
FA C20_5 (g)	0.044	0.061

Variable	Correlation Coefficient	Significance (2-tailed)
FA C18_1 (g)	0.037	0.115
FA C18_0 (g)	0.035	0.137
FA C20_0 (g)	0.035	0.139
Fibre intake (g)	-0.031	0.189
FA C22_6 (g)	0.030	0.200
Insoluble fibre intake (g)	0.029	0.224
FA C4_0 (g)	0.029	0.221
FA C20_1 (g)	0.029	0.214
FA C6_0 (g)	0.027	0.257
FA C22_0 (g)	0.027	0.246
FA C22_1 (g)	0.026	0.275
Trans fat intake (g)	0.025	0.281
FA C18_2 (g)	0.024	0.304
FA C24_0 (g)	0.022	0.354
Soluble fibre intake (g)	0.018	0.433
FA C18_4 (g)	0.017	0.476
HDL-C (mmol/L)	-0.015	0.516
Protein intake as % of TE	0.002	0.949
Rural/Urban	0.001	0.982

BMI, body mass index; TE, total energy, SATFAT, saturated fat; CRP, C - reactive protein; MUFA, monounsaturated fat, PUFA, polyunsaturated fat; FA, fatty acid; HDL-C, high density lipoprotein cholesterol; CHO, carbohydrates

Supplementary table 2 Linear regression model for BMI and age separately according to gender to indicate the stronger relationship between older age and LDL-c in women

MEN

Step 1	B	SE B	β	p-value
Constant	0.919	0.282		0.001
BMI	0.087	0.014	0.293	0.000
$r^2 = 0.086$ (Step 1)				
Step 2	B	SE B	β	p-value
Constant	0.949	0.363		0.000
BMI	0.087	0.014	0.294	0.000
Age	-0.001	0.005	-0.006	0.896
$r^2 = 0.086$ (Step 2)				

BMI, Body mass index

WOMEN

Step 1	B	SE B	β	p-value
Constant	2.315	0.164		0.000
BMI	0.030	0.006	0.184	0.000
$r^2 = 0.034$ (Step 1)				
Step 2	B	SE B	β	p-value
Constant	1.620	0.250		0.000
BMI	0.029	0.006	0.180	0.000
Age	0.014	0.004	0.132	0.000
$r^2 = 0.051$ (Step 2)				

BMI, Body mass index

ALL with Gender x Age predictor

Step 1	B	SE B	β	p-value
Constant	2.681	0.052		0.000
Age_Gender	0.009	0.001	0.199	0.000
$r^2 = 0.040$ (Step 1)				

Supplemental table 3 Smoking status of the PURE population

Variable	Study Population ^a	Men ^a	Women ^a
Smoking status			
Never smoked			
Total	501	140	361
Rural	269	78	191
Urban	132	62	170
Formerly smoked			
Total	47	32	15
Rural	25	16	9
Urban	22	16	6
Currently smoking			
Total	617	266	3
Rural	313	116	197
Urban	304	150	154

Supplemental table 4: Comparison between the total PURE population and the sample available for analysis in this study

Variable	N=1169 ^a	N=1860 ^a	p-value
			T-test
Age (years)	49.5 (10.3)	49.3 (10.4)	0.616
BMI (kg/m ²)	24.4 (6.76)	24.6 (6.95)	0.393
TC (mmol/l)	5.03 (1.38)	5.00 (1.37)	0.532
LDL-c (mmol/l)	2.96 (1.15)	2.92 (1.16)	0.248
HDL-C (mmol/l)	1.5 (0.62)	1.52 (0.63)	0.257
TAG (mmol/l)	1.26 (0.65)	1.3 (0.79)	0.035
Total energy (kJ)	7249 (3354)	7582 (3552)	0.001
Protein % of TE	12.2 (2.06)	12.3 (2.12)	0.68
CHO % of TE	63.9 (9.15)	63.3 (9.26)	0.033
Fat % of TE	23.7 (7.61)	24.1 (7.64)	0.114
MUFA % of TE	6.18 (2.96)	6.36 (3.01)	0.035
PUFA % of TE	7.27 (2.95)	7.35 (2.94)	0.354
SATFAT % of TE	5.71 (2.76)	5.83 (2.71)	0.144
Cholesterol intake (mg)	171.1 (137.8)	187.6 (154.6)	0.000
Total fibre intake (g)	20.1 (10.00)	20.9 (10.4)	0.005

^aMeans and standard deviations are reported. BMI, body mass index; TC, total cholesterol; LDL-c, low-density lipoprotein cholesterol; HDL-C high-density lipoprotein cholesterol; TAG, triglycerides; CHO, carbohydrates; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids; SATFAT, saturated fatty acids

Supplementary table 5: A summary of the SNPs genotypes in the PCSK9 and LDLR genes**PCSK9**

SNP number	Nucleotide substitution	Position	p (HWE)	Frequency	LDL-c (mmol/l)	ANOVA	ANCOVA	Bonferroni	MAF	Novel Yes/No	Previous studies	Similar results
rs28362202	c.-26 G→A	5'UTR	0.80	n=1530								
GG				1305	2.93			0.597 (GG vs GA)	P: 0.08	No	Reported in Yoroba population from 1000 Genomes Project	No association results available
GA				218	3.03	0.139	0.260	0.670 (GA vs AA)	A: 0.02			
AA				7	2.51			0.973 (AA vs GG)	E: 0.00			
rs2495480	c.208-161 C→T	Intron 1	0.71	n=1530								
TT				1053	2.91			0.479 (TT vs TC)	P: 0.17	No	Abifadel et al, 2003 Shioji et al, 2004	Yes Yes
TC				438	2.99	0.119	0.064	0.411 (CC vs TC)	A: 0.20			
CC				39	3.27			0.132 (TT vs CC)	E: 0.06			
rs2495482	c.207+15 A→G	Intron 1	0.77	n=1523								
GG				1154	2.92			0.851 (GG vs AG)	P: 0.13	No	Leren et al, 2004 Humphries et al, 2006 Huijgen et al, 2012	No association results available
AG				347	2.99	0.157	0.244	0.906 (AG vs AA)	A: 0.18			
AA				22	3.25			0.523 (AA vs GG)	E: 0.06			
rs200109442	c.400-24 C→T	Intron 2	0.8	n=1534								
CC				1483	2.95				P: 0.02	No	Reported in African Americans in 1000 Genomes project	No association results available
CT				51	2.70	0.239	0.116	0.116	A: Not available			
TT				0	-				E: 0.001			
rs4927193	c.399+165 T→C	Intron 2	0.24	n=1535								
TT				709	2.94			1.00 (TT vs TC)	P: 0.33	No	Shioji et al, 2004 Postmus et al, 2013 Huijgen et al, 2012	No association results available
TC				647	2.97	0.383	0.287	0.344 (TC vs CC)	A: 0.20			
CC				179	2.82			0.586 (TT vs CC)	E: 0.13			
rs624612	c.524-68 C→G	Intron 3	0.99	n=1537								
CC				1351	2.95				P: 0.06	No	Abifadel et al, 2003 Blesa et al, 2008	No association results available
GC				180	2.87	0.412	0.65	1.00	A: 0.07			
GG				6	3.12				E: 0.41			
rs499718	c.523+230 T→C	Intron 3	0.95	n=1533								
TT				369	2.88			0.129 (TT vs TC)	P: 0.51	No	Ding and Kullo et al, 2008 Postmus et al, 2013	No association results available
TC				760	3.02 ^b	0.030	0.024	0.047 (TC vs CC)	A: 0.46			
CC				404	2.88 ^b			1.00 (TT vs CC)	E: 0.17			
rs2483205	c.658-7 C→T	Intron 4	0.60	n=1534								
CC				421	2.85			0.359 (CC vs CT)	P: 0.47	No	Abifadel et al, 2003 Leren et al, 2004 Postmus et al, 2013 Miyake et al, 2008 Humphries et al, 2007	No association results available
CT				784	2.95	0.159	0.072	0.814 (CT vs TT)	A: 0.49			
TT				329	3.03			0.072 (CC vs TT)	E: 0.48			

Supplementary table 5 continues

PCSK9

SNP number	Nucleotide substitution	Position	p (HWE)	Frequency	LDL-c (mmol/l)	ANOVA	ANCOVA	Bonferroni	MAF	Novel Yes/No	Previous studies	Similar results
rs7552350	c.657+114 C→A	Intron 4	0.02	n=1535								
CC				867	2.91 ^c			1.00 (CC vs CA)	P: 0.26		Huijgen et al, 2012	
CA				545	2.9 ^d	0.018	0.001	0.001 (CA vs AA)	A: NA	No		No association results available
AA				123	3.31 ^{cd}			0.001 (CC vs AA)	E: NA			
rs2495477	c.799+3 A→G	Intron 5	0.36	n=1617								
AA				121	2.89				P: 0.27		Abifadel et al, 2003	
AG				580	2.94	0.597	0.828	1.00	A: 0.28	No	Leren et al, 2004	
GG				835	2.95				E: 0.59		Miyake et al, 2008 Humphries et al, 2006 Tosi et al, 2007 Huijgen et al, 2012	No association results available
rs494198	c.799+64 C→A	Intron 5	0.22	n=1505								
CC				275	2.87			1.00 (CC vs CA)	P: 0.58		Abifadel et al, 2003	
CA				698	2.94	0.301	0.443	1.00 (CA vs AA)	A: 0.51	No	Leren et al, 2004	No association results available
AA				532	2.98			0.610 (CC vs AA)	E: 0.37			
rs28362256	c.996+379 C→T	Intron 6	0.92	n=1537								
CC				1336	2.93				P: 0.07		Reported in Bushman, Yoroba and African American populations available on dbSNP	
CT				195	3.01	0.304	0.65	1.00	A: 0.04	No		No association results available
TT				6	3.00				E: 0.00			
rs374941781	c.829 G→C (V277L)	Exon 6	0.96	n=1534								
GG				6	2.65			1.00 (GG vs GC)	P: 0.07			
GC				191	2.86	0.426	0.463	0.859 (GC vs CC)	A: NA	Yes	-	-
CC				1337	2.95			1.00 (GG vs CC)	E: NA			
rs533375	c.1180+174 A→G	Intron 7	0.31	n=1537								
AA				803	2.93			1.00 (AA vs AG)	P: 0.29		Ding and Kullo et al, 2008	
AG				599	2.98	0.231	0.42	0.654 (AG vs GG)	A: 0.25	No		No association results available
GG				135	2.85			1.00 (GG vs AA)	E: 0.18			
rs584626	c.1354+102 C→T	Intron 8	0.84	n=1536								
CC				689	2.95				P: 0.34		Abifadel et al, 2003	
CT				672	2.93	0.988	0.695	1.00	A: 0.23	No	Shioji et al, 2004	No association results available
TT				175	2.97				E: 0.18			
rs585131	c.1355-56 C→T	Intron 8	0.59	n=1536								
CC				845	2.93				P: 0.26		Abifadel et al, 2003	
CT				578	2.97	0.333	0.652	1.00	A: 0.21	No	Huijgen et al, 2012	No association results available
TT				113	2.86				E: 0.18			

Supplementary table 5 continues

PCSK9

SNP number	Nucleotide substitution	Position	p (HWE)	Frequency	LDL-c (mmol/l)	ANOVA	ANCOVA	Bonferroni	MAF	Novel Yes/No	Previous studies	Similar results
rs28362261	c.1274 A→G (N425S)	Exon 8	0.91	n=1537								
AA				1504	2.94				P: 0.01		Pisciotta et al, 2006	Yes
AG				33	2.94	0.77	0.986		A: 0.03	No	Kotowski et al, 2006	
GG				0	-				E: 0.00		Wu et al, 2013	
rs28362263	c.1327G→A (A443T)	Exon 8	0.50	n=1632								
GG				1514	2.94 ^a			0.043 (GG vs GA)	P: 0.04		Haung et al, 2009	Yes
GA				114	2.68 ^a	0.078	0.047	1.00 (GA vs AA)	A: 0.09	No	Wu et al, 2013	
AA				4	2.77			1.00 (GG vs AA)	E: 0.00		Coram et al, 2013	
rs28362269	c.1504-16 C→T	Intron 9	0.95	n=1533								
CC				1374	2.92 ^e			0.601 (CC vs CT)	P: 0.05		Lye et al, 2013	Yes, but loose sign. after Bonferroni
CT				154	3.01	0.018	0.029	0.110 (CT vs TT)	A: 0.06	No		correction in PURE population
TT				5	4.11 ^e			0.055 (CC vs TT)	E: 0.00			
rs28362270	c.1658 A→G (H553R)	Exon 10	0.99	n=1533								
AA				1425	2.92			0.094 (AA vs AG)	P: 0.04		Kotowski et al, 2006	In DHS ass with increased LDL-C, No
AG				106	3.16	0.043	0.087	1.00 (AG vs GG)	A: A: 0.01	No		ass in Cook County
GG				2	2.55			1.00 (GG vs AA)	E: 0.00			
rs483462	c.1681+64 G→A	Intron 10	0.35	n=1533								
GG				892	2.98			1.00 (GG vs GA)	P: 0.23		Abifadel et al, 2003	No association results with LDL-C
GA				568	2.90	0.11	0.17	0.541 (GA vs AA)	A: 0.39	No	Leren et al, 2004	available
AA				73	2.77			0.375 (AA vs GG)	E: 0.69			
rs111563724	c.1696 T→C (W566R)	Exon 11	0.58	n=1537								
TT				1438	2.94				P: 0.03		Found in a Bushman from	No association results available
TC				96	2.98	0.639	0.707	1.00	A: NA	No	1000 Genomes Project	
CC				3	2.45				E: NA			
rs372506466	c.1781 C→A (A594D)	Exon 11	0.37	n=1522								
CC				1382	2.97			0.097 (CC vs CA)	P: 0.05			
CA				134	2.75	0.072	0.081	1.00 (CA vs AA)	A: NA	Yes	-	-
AA				6	2.64			1.00 (AA vs CC)	E: NA			
rs28362286	c.2037C→A(C679X)	Exon 12	0.57	1428								
CC				1350	2.99 ^f				P: 0.03		Cohen et al, 2005	Yes
CA				78	2.17 ^f	0.000	0.000	0.000	A: 0.01	No	Hallman et al, 2007	
AA				0	-				E: 0.00			

Supplementary table 5 continues

PCSK9

SNP number	Nucleotide substitution	Position	p (HWE)	Frequency	LDL-c (mmol/l)	ANOVA	ANCOVA	Bonferroni	MAF	Novel Yes/No	Previous studies	Similar results
rs505151	c.2009 G→A	Exon 12	1	n=1505								
AA				737	2.93				P: 0.30		Naureckiene et al, 2003	Yes
AG				633	2.96	0.717	0.756	1.00	A: 0.29	No	Chen et al, 2005	
GG				135	3.00				E: 0.03		Humphries et al, 2006 Kotowski et al, 2006 Abifadel et al, 2003	
rs662145	c.*571C→T	3'UTR	0.47	n=1534								
CC				795	2.99			0.182 (CC vs CT)	P: 0.27		Postmus et al, 2013	No association results available
CT				632	2.88	0.184	0.168	1.00 (CT vs TT)	A: 0.42	No		
TT				107	2.97			1.00 (TT vs CC)	E: 0.75			
rs369066144	c.*500 C→T	3'UTR	0.30	n=1536								
CC				1431	2.95			1.00 (CC vs CT)	P: 0.04			
CT				101	2.85	0.25	0.278	0.718 (CT vs TT)	A: NA	Yes	-	-
TT				4	2.18			0.511 (TT vs CC)	E: NA			

LDLR

rs369850745	c.-314 C>T	Promoter	0.91									
CC				1501	2.94				P: 0.01			
CT				34	2.9	0.929	0.852	0.852	A: NA	Yes	-	-
TT				0	-				E: NA			
rs17249141	c.-217 C>T	Promoter	0.98									
CC				1521	2.95				P: 0.01			
CT				16	2.4	0.068	0.050	0.050	A: 0.02	No	Scholtz et al., 1999	Promoter activity increase to 160% in cells without sterol and 50% in cells with a sterol containing medium.
TT				0	-				E: 0.00			
rs17242759	c.67+18 C>A	Intron 1	0.42									
CC				1270	2.94			1.00 (CC vs CA)	P: 0.09			
CA				258	2.91	0.104	0.120	0.121 (CA vs AA)	A: 0.06	No	Van Zyl et al., 2013	No significant association found with LDL-C
AA				8	3.73			0.142 (CC vs AA)	E: 0.01			
rs2228671	c.81 C>T (C27C)	Exon 2	0.19								El Messal et al., 2003	
CC				1457	2.95			0.826 (CC vs CT)	P: 0.03		Chen et al., 2007	No significant association reported
CT				77	2.81	0.779	0.474	1.00 (CT vs TT)	A: 0.02	No	Al-Khateeb et al., 2011	
TT				3	2.59			1.00 (CC vs TT)	E: 0.09		Linsel-Nitschke et al., 2008	T associated with lower LDL-C levels

Supplementary table 5 continues

LDLR

SNP number	Nucleotide substitution	Position	p (HWE)	Frequency	LDL-c (mmol/l)	ANOVA	ANCOVA	Bonferroni	MAF	Novel Yes/No	Previous studies	Similar results
rs10423288	c.314-50 T>C (IVS3-50)	Intron 3	0.99									
TT				1132	2.93			1.00 (TT vs TC)	P: 0.14			
TC				373	2.97	0.426	0.444	0.646 (TC en CC)	A: 0.16	No	Van Zyl et al., 2013	No significant association reported
CC				30	2.71			0.837 (TT vs CC)	E: 0.02			
rs148698650	c.829 G>A (E277K)	Exon 6	0.93								Descamps et al., 2001	Found as a double mutant allele in FH patients
GG				1507	2.94				P: 0.01			
GA				29	3.05	0.395	0.59	0.59	A: 0.00	No	Lui et al, 2013	No significant association with LDL-C reported
AA				0	-				E: 0.00		Van Zyl et al., 2013	
rs116405216	c.941-4 G>A	Intron 6	0.9									
GG				1364	2.94				P: 0.06			
GA				164	2.92	0.997	0.945	1.00	A: 0.02	No	Van Zyl et al., 2013	No significant association reported
AA				6	3.04				E: 0.00			
rs11669576	c.1171 G>A (A391T)	Exon 8	0.77								Kotze et al., 1989 Gudnason et al., 1995	
GG				1082	2.91			0.438 (GG vs GA)	P: 0.16			
GA				412	3.01	0.233	0.32	1.00 (GA vs AA)	A: 0.24	No	Weiss et al, 1998	
AA				34	3.02			1.00 (GG vs AA)	E: 0.05		Zakharova et al., 2001 Wang te al., 2001 El Messel et al., 2003 Brusgaard et al., 2006	No effect on LDLR activity
rs72658862	c.1187-25 C>T	Intron 8	0.69									
CC				1346	2.94				P: 0.06			
CT				163	2.94	0.803	0.653	1.00	A: 0.02	No	Van Zyl et al., 2013	No significant association reported
TT				7	3.33				E: 0.00			
rs5930	c.1413 A>G (R471R)	Exon 10	0.98	n=1532								
AA				29	3.29			0.785 (AA vs AG)	P: 0.14		Tatishcheva et al., 2001	No significant association with LDL-C reported
AG				357	3.04 ^b	0.027	0.023	0.092 (AG vs GG)	A: 0.21	No	Pećin et al., 2013	
GG				1146	2.90 ^b			0.193 (AA vs GG)	E: 0.41			
rs1569372	c.1586+53 A>G	Intron 10	0.99									
AA				518	2.99			0.502 (AA vs AG)	P: 0.42			
AG				747	2.9	0.281	0.376	1.00 (AG vs GG)	A: 0.29	No	Van Zyl et al., 2013	No significant association with LDL-C reported
GG				266	2.95			1.00 (AA vs GG)	E: 0.41			

Supplementary table 5 continues

LDLR

SNP number	Nucleotide substitution	Position	p (HWE)	Frequency	LDL-c (mmol/l)	ANOVA	ANCOVA	Bonferroni	MAF	Novel Yes/No	Previous studies	Similar results
rs5929	c.1617 C>T	Exon 11	0.97									
CC	(P539P)			1239	2.97			0.102 (CC vs CT)	P: 0.10	No	Miljkovic et al., 2010 Chiou & Charng, 2012	T allele associated with lower LDL-C Used in a haplotype but no association reported for T allele
CT				281	2.82	0.072	0.077	1.00 (CT vs TT)	A: 0.12			
TT				17	2.72			1.00 (CC vs TT)	E: 0.04			
rs4508523	c.1705+56 C>T	Intron 11	0.81									
CC				493	2.99			0.824 (CC vs CT)	P: 0.44	No	Chiou&Charng, 2012	Part of haplotype; no association reported for T allele
CT				739	2.92	0.571	0.446	1.00 (CT vs TT)	A: 0.46			
TT				296	2.9			0.829 (CC vs TT)	E: 0.12			
rs2738447	c.1706-55A>C	Intron 11	0.64	n=1537								
CC				1188	2.9 ^c			0.159 (CC vs CA)	P: 0.12	No	Gopalraj, 2009	No association results with LDL-C available
CA				322	3.03 ^d	0.002	0.001	0.037 (CA vs AA)	A: 0.16			
AA				27	3.59 ^{cd}			0.004 (CC vs AA)	E: 0.43			
rs688	c.1773 C>T	Exon 12	0.82									
CC	(N591N)			1381	2.95			0.709 (CC vs CT)	P: 0.04	No	Zakharova et al., 2001 Zhu et al., 2007	Associate with increased LDL-C Affects splicing efficiency of exon 12 and associates with increased LDL-C
CT				132	2.83 ^e	0.076	0.051	0.071 (CT vs TT)	A: 0.06			
TT				2	4.62 ^e			0.101 (CC vs TT)	E: 0.47			
rs5925	c.1959 T>C	Exon 13	0.92									
TT	(V653V)			1001	2.97			0.095 (TT vs TC)	P: 0.19	No	Gopalraj, 2009 Lui et al., 2010	No association results available Part of haplotype; no association reported for C allele
TC				475	2.84 ^f	0.087	0.020	0.064 (TC vs CC)	A: 0.17			
CC				60	3.19 ^f			0.421 (TT vs CC)	E: 0.45			
rs5927	c.2232 A>G	Exon 15	0.18									
AA	(R744R)			96	2.90			0.234 (AA vs AG)	P: 0.27	No	Gopalraj, 2009 Miljkovic et al., 2010 Chiou & Charng, 2012	No association results with LDL-C G allele associated with lower LDL-C Part of haplotype; no association reported for T allele
AG				629	3.00	0.163	0.212	1.00 (AG vs GG)	A: 0.35			
GG				806	2.94			1.00 (AA vs GG)	E: 0.27			
rs369402076	c.2311+9 T>G	Intron 15	0.67									
TT				1466	2.94				P: 0.02	Yes	-	-
TG				68	2.96	0.955	0.854	0.854	A: NA			
GG				0	-				E: NA			

Supplementary table 5 continues

LDLR

SNP number	Nucleotide substitution	Position	p (HWE)	Frequency	LDL-c (mmol/l)	ANOVA	ANCOVA	Bonferroni	MAF	Novel Yes/No	Previous studies	Similar results
rs6413503	c.2548-42 G>A	Intron 17	0.00									
GG				1267	2.91			0.107 (GG vs GA)	P: 0.11	No	Andreotti et al., 2009	Significant association with total cholesterol levels
GA				166	3.1	0.107	0.089	1.00 (GA vs AA)	A: 0.16			
AA				88	3.01			1.00 (GG vs AA)	E: 0.01			
rs14158	c.*52 G>A	3'UTR	0.83	n=1521								
GG				1220	2.9 ^h			0.141 (GG vs GA)	P: 0.11	No	Muallem et al., 2007	Part of a haplotype significantly associated with LDL-c
GA				282	3.05 ^g	0.003	0.001	0.035 (GA vs AA)	A: 0.12			
AA				19	3.71 ^{gh}			0.005 (GG vs AA)	E: 0.22			
rs3826810	c.*141 G>A	3'UTR	0.13									
GG				1120	2.94				P: 0.13	No	Muallem et al., 2007	Forms part of a haplotype significantly associated with LDL-c
GA				360	2.94	0.898	0.899	1.00	A: 0.13			
AA				17	2.82				E: 0.03			
rs2738464	c.*315 G>C	3'UTR	0.59									
GG				165	2.87				P: 0.32	No	Muallem et al., 2007	Forms part of a haplotype significantly associated with LDL-c
GC				650	2.93	0.477	0.620	1.00	A: 0.38		Yi et al., 2010	
CC				722	2.96				E: 0.13			
rs2738465	c.*504 G>A	3'UTR	0.71	n=1535								
GG				855	2.90 ⁱ			1.00 (GG vs GA)	P: 0.26	No	Muallem et al., 2007	Forms part of a haplotype significantly associated with LDL-c
GA				573	2.95	0.037	0.051	0.169 (GA vs AA)	A: 0.25			
AA				107	3.18 ⁱ			0.047 (GG vs AA)	E: 0.26			
rs1433099	c.*666 T>C	3'UTR	0.20									
TT				403	2.92				P: 0.48		Muallem et al., 2007	Forms part of a haplotype significantly associated with LDL-c
TC				800	2.95	0.926	0.917	1.00	A: 0.40	No		
CC				330	2.95				E: 0.71		Anand et al., 2009	Associated with lower Apo B/A1 ratio
rs3180023	c.*1217 C>A	3'UTR	1	n=1537								
CC				1535	2.94 ^j				P: 0.001	No	Van Zyl et al., 2013	Significant association with LDL-c levels
CA				2	5.09 ^j	0.001	0.006	0.006	A: 0.01			
AA				0	-				E: 0.08			

ANNEXURE M: GUIDE FOR AUTHORS FOR THE JOURNAL OF HUMAN GENETICS

All manuscripts should be submitted to the Journal of Human Genetics via the journal's online manuscript submission and tracking system at <http://mts-jhg.nature.com>, where more detailed instructions on manuscript preparation can also be found.

AIMS AND SCOPE

The Journal of Human Genetics (JHG) is an international journal publishing articles on human genetics, including medical genetics and human genome analysis. JHG publishes articles on all aspects of human genetics, including molecular genetics, clinical genetics, behavioral genetics, immunogenetics, pharmacogenomics, population genetics, functional genomics, epigenetics, genetic counseling and gene therapy.

Articles on the following areas are especially welcome: genetic factors of monogenic and complex disorders, genome-wide association studies, genetic epidemiology, cancer genetics, personal genomics, genotype-phenotype relationships and genome diversity.

- Medical genetics
- Human genome analysis
- Gene cloning and mapping
- Linkage and association analyses
- Mutational analysis
- Susceptibility genes to multifactorial disorders
- Human Evolution
- Cancer genetics
- Gene therapy
- Genetic and functional analysis of animal models of disease or behaviour
- Genetic polymorphism of biologically important genes
- Novel mutations found in patients with hereditary diseases
- Novel mutations found in cancer cells
- Population genetics

PRIOR TO SUBMISSION**Editorial policy**

Editor-in-Chief: Katsushi Tokunaga

A manuscript will be considered for publication on the understanding that all named authors have agreed to its submission and that if accepted it will not be later published in the same or similar form in any language without the consent of the publishers. The editors also encourage submission of Reviews, Short Communications and items of Correspondence.

The editors reserve the right to reject manuscripts without review. Such rejections must be approved by the editor-in-chief, and are intended to alleviate unnecessary workload for the editorial board, as well as provide authors the opportunity to seek other publishing options as soon as possible. Articles that are selected for peer review will be reviewed by two or more referees.

To avoid unnecessary delays in the review process, please consider the following policies carefully before you submit your manuscript. Manuscripts that are not concise or do not conform to the conventions and standards of the *Journal of Human Genetics* will be returned to the authors for revision.

Conflicts of Interest

In the interests of transparency and to help reviewers assess any potential bias, Journal of Human Genetics requires authors of all submitted original research papers to declare any conflict of interest (COI) in relation to the submitted work, following the guideline and detailed regulations set by the Japan Society of Human Genetics (JSHG) in 2012.

Authors submitting their manuscripts using the journal's online manuscript tracking system are required to make their declaration as part of this process and to specify the competing interests in cases where they exist.

Criteria for COI disclosure

1. Received honoraria of 1,000,000yen or more (in one year) for lectures, article contributions or such activities to support promotional activities
2. Received 1,000,000yen or more (in one year) for Employment, supervising or advisory position
3. Given stock or stock options worth 5% or more of the total share, or generating profit of 1,000,000yen or more (in one year).
4. Received patent royalties or licensing fees of 1,000,000yen or more (in one year)
5. Received research funding of 2,000,000yen or more (in one year)

6. Outside research activities, given travels, gifts or any other benefit worth 50,000yen or more in one year.

If the self-reported COI includes the any of the above listed situation, authors should add COI statement to the end of the manuscript main text, and before the acknowledgement or the list of references. Please note that the disclosure is required only for the relationship that the author had within one year before the date of submission.

Referees are also requested to indicate any potential conflict they might have reviewing a particular paper.

Electronic manipulation of images

Digital image enhancement is acceptable practice, although it can result in the presentation of unrepresentative data as well as in the loss of meaningful signals. During manipulation of images a positive relationship between the original data and the resulting electronic image must be maintained. If a figure has been subjected to significant electronic manipulation, the specific nature of the enhancements must be noted in the figure legend or in the 'Materials and methods' section. The editors reserve the right to request original versions of figures from the authors of a paper under consideration.

Plagiarism and fabrication

Plagiarism is when an author attempts to pass off someone else's work as his or her own. Duplicate publication, sometimes called self-plagiarism, occurs when an author reuses substantial parts of his or her own published work without providing the appropriate references. Such manuscripts would not be considered for publication. But minor plagiarism without dishonest intent is relatively frequent, for example, when an author reuses parts of an introduction from an earlier paper. The editors judge any case of which they become aware (either by their own knowledge of and reading about the literature, or when alerted by referees) on its own merits.

Nature Publishing Group is part of CrossCheck, an initiative to help editors verify the originality of submitted manuscripts. As part of this process, selected submitted manuscripts are scanned and compared with the CrossCheck database.

If a case of plagiarism comes to light after a paper is published in the Journal of Human Genetics, the journal will conduct a preliminary investigation. If plagiarism is found, the journal will contact the author's institute and funding agencies. A determination of misconduct will lead Journal of Human Genetics to run a statement, bidirectionally linked online to and from the original paper, to note the plagiarism and to provide a reference to the plagiarised material. The paper containing the plagiarism will also be obviously marked on each page of the PDF. Depending on the extent of the plagiarism, the paper may also be formally retracted.

Sharing data sets

A condition of publication is that authors are required to make materials, data and associated protocols promptly available to others without preconditions. Data sets must be made freely

available to readers from the date of publication, and must be provided to editors and peer-reviewers at submission, for the purposes of evaluating the manuscript.

For the following types of data set, submission to a community-endorsed, public repository is mandatory. Accession numbers must be provided in the paper. Examples of appropriate public repositories are listed below.

[Nucleotide and protein sequences]

DNA and RNA sequences: DDBJ, Genbank, or European Nucleotide Archive(ENA).

DNA sequencing data (traces for capillary electrophoresis and short reads for next-generation sequencing): DDBJ Sequence Read Archive, NCBI Sequence Read Archive(SRA), or EBI Sequence Read Archive(ERA). Deep sequencing data: deposit in GEO or ArrayExpress upon submission to the journal. Accession numbers must be provided in the published manuscript. This policy includes even short stretches of novel sequence information such as epitopes, functional domains, genetic markers, or haplotypes. Short novel sequences must include surrounding sequence information to provide context. The sequences of all RNAi, antisense and morpholino probes must be included in the paper or deposited in a public database, with the accession number quoted. When an unpublished library is included in the paper, at minimum the sequences of the probes central to the conclusions of the paper must be presented.

Protein sequences: deposit in Protein DataBank, UniProt.

[Microarray gene expression data]

MIAME-compliant microarray data: deposit in GEO or MGED web site specifying microarray standards.

[Genome-wide association data]

We encourage authors to deposit data from genome-wide association studies in Human Genome Variation Database, dbGap, or EGA (European Genome-phenome Archive) upon submission to the journal.

Clinical trials

As defined by the International Committee of Medical Journal Editors (ICMJE), a clinical trial is any research project that prospectively assigns human subjects to intervention and comparison groups to study the cause-and-effect relationship between a medical intervention and a health outcome. A medical intervention is any intervention used to modify a health outcome and includes but is not limited to drugs, surgical procedures, devices, behavioral treatments, and process-of-care changes. A trial must have at least one prospectively assigned concurrent control or comparison group in order to trigger the requirement for registration. Nonrandomized trials are not exempt from the registration requirement if they meet the above criteria.

When reporting experiments on human subjects, indicate whether the procedures were in accordance with the ethical standards of the responsible committee on human experimentation (institutional or regional) or with the Helsinki Declaration of 1975 (as revised in 1983). Include Institutional Review Board or Animal Care and Use Committee approvals.

All clinical trials must be registered in a public registry prior to submission. Journal of Human Genetics follows the trials registration policy of the ICMJE and considers only trials that have been appropriately registered before submission, regardless of when the trial closed to enrolment. Acceptable registries must meet the following ICMJE requirements:

- be publicly available, searchable, and open to all prospective registrants
- have a validation mechanism for registration data
- be managed by a not-for-profit organization

Examples of registries that meet these criteria include (1) the registry sponsored by the United States National Library of Medicine; (2) the International Standard Randomised Controlled Trial Number Registry; (3) International Clinical Trials Registry Platform (ICTRP) (4) the National (United Kingdom) Research Register; and (5) the European Clinical Trials Database. Nature Publishing Group endorses the toolkits and guidelines produced by the following bodies:

- Committee on Publication Ethics
- Good Publication Practice
- Medical Publishing Insights and Practices Initiative

Supplementary information for the editors and the reviewers

Any manuscripts under review or accepted for publication elsewhere should accompany the submission if they are relevant to its scientific assessment. Authors should also provide upon submission any kind of supplementary material that will aid the review process.

FORMAT OF PAPERS

Preparation of manuscripts

All papers should be written in concise English but should contain sufficient detail to illustrate how the results were obtained. Manuscripts should be double-spaced with wide margins.

Manuscripts are considered with the understanding that they have not been published previously in print or electronic format and are not under consideration by another publication or electronic medium. Copies of possibly duplicative materials that have been previously published or are being considered elsewhere must be provided at the time of manuscript submission.

Manuscripts should contain a statement to the effect that all human studies have been reviewed by the appropriate ethics committee or it should be stated clearly in the text that all persons gave their informed consent prior to their inclusion in the study. Details that might disclose the identity of the subjects under study should be omitted.

Cover Letter

The uploaded covering letter must state that the material has not been submitted for publication elsewhere while under consideration for the Journal of Human Genetics. Identify the name, full postal address, and fax number, of corresponding author. The authors are free to offer suggestions of suitable expert reviewers.

CONTENT TYPES

The content types accepted by the *Journal of Human Genetics* are:

Original article

Short communication

Correspondence

Review

Commentary

Original Article

Studies that are of high scientific quality and that are of interest to the diverse readership of the journal. Manuscripts should include an abstract and appropriate experimental details to support the conclusions. Original Articles should be no more than 5000 words excluding references and figure legends and should not normally include more than six display items (tables and/or figures). They should include title, abstract, introduction, materials and methods, results and discussion sections.

Short Communication

Studies that fall short of the criteria for full research papers (eg short experimental reports limited by sample size or duration, novel hypotheses or commentaries) may be submitted as Short Communications. Experimental details are not required in Short Communications, but a general outline of experimental methods should be included. They should contain no more than 1000 words of text, a maximum of three display items (tables and/or figures) and a maximum of 20 references. Short Communications should include title, abstract, and description sections, but there is otherwise no need to divide the text into sections. In all other respects, the directions for Original Article should be followed.

Correspondence

Preliminary reports of unusual urgency, significance and interest, whose subjects may be republished in expanded form, may be submitted as items of Correspondence. They should contain no more than 900 words of text, one display item (figure or table) and a maximum of 10 references. Correspondence items **do not** contain an abstract, and apart from keywords there is no obligation to divide the text into sections. In all other respects, the directions for full papers should be followed.

Review

Reviews normally have a word limit of 5000 words including abstract but excluding references, tables and figures. A number of Reviews will be solicited by the editors; however, we also welcome timely, unsolicited Reviews. Authors with proposals for Reviews should present information concerning the proposed content and authors of their Review to the editors prior to submission. Unless otherwise informed, all changes for color images will be the authors' responsibility.

Commentary

Articles that describe issues and questions from the articles appeared in the Journal of Human Genetics, or articles describing topics of particular interest to the human genetics research community. Commentaries on published articles are normally solicited by invitation. The suggested length for a commentary article is 1000 words with a maximum of 10 references. Commentaries do not contain an abstract section, and apart from keywords there is no obligation to divide the text into sections. The use of a figure to illustrate key discussion points is encouraged.

Article section

Please make spelling consistent with current editions of either Webster's Dictionary or Oxford English Dictionary.

In general, manuscripts should be divided in to the following sections:

Title page

Abstract

Introduction

Materials and methods

Results

Discussion

Acknowledgements

References

Figure legends

Figures

Tables

Supplementary Information

[Title page]

Title page, giving a concise but informative title, and the first and last names, and other initials, and affiliations (but not degrees) of all contributors (formerly called authors). The order in which the contributors are listed should be agreed amongst the investigators, and should indicate that the first listed made the greatest contribution to the paper. Please provide a running title of no more than 50 characters including spaces.

Up to eight keywords, which may or may not appear in the title, should be given in alphabetical order, below the abstract, each separated by a slash (/).

[Abstract]

An abstract of not more than 150–200 words. The abstract should be comprehensible to readers before they have read the paper, and abbreviations and reference citations within the abstract should be avoided.

[Introduction]

This should give a short, clear account of the background and reasons for undertaking the study. It should not be a review of the literature.

[Materials and methods]

This section should contain sufficient detail so that all experimental procedures can be repeated by others, in conjunction with cited references. This section may be divided into subheadings to assist the reader. Names of products and manufacturers should be included only if alternative sources are deemed unsatisfactory.

Instruments used, as well as standard techniques and procedures applied throughout the work, should appear in a paragraph at the beginning of the Materials or Experimental section. Novel experimental procedures should be described in detail, but published procedures should be referred to by literature citation of the original article and published modifications.

Authors should use approved nomenclature for gene symbols, and use symbols rather than italicized full names (TTN, not titin). Please consult the appropriate nomenclature databases for correct gene names and symbols. A useful resource is LocusLink. Approved human gene symbols are provided by HUGO Gene Nomenclature Committee (HGNC), e-mail: nome@galton.ucl.ac.uk; see also <http://www.gene.ucl.ac.uk/nomenclature>. Approved mouse symbols are provided by The Jackson Laboratory, e-mail: nomen@informatics.jax.org; see also <http://www.informatics.jax.org/mgihome/nomen>.

Avoid listing multiple names of genes (or proteins) separated by a slash, as in 'OCT4/POU5F1', as this is ambiguous (it could mean a ratio, a complex, alternative names or different subunits). Use one name throughout and include the other at first mention: 'OCT4 (also known as POU5F1)'.

[Results]

The description of results should not simply reiterate data that appear in tables and figures and, likewise, the same data should not be displayed in both tables and figures. The results section should be concise and follow a logical sequence. If the paper describes a complex series of experiments, it is permissible to explain the protocol/experimental design before presenting the results. Do not discuss the results or draw any conclusions in this section. This section may be divided into subheadings to assist the reader. Large datasets or other cumbersome data pertinent to the manuscript may be submitted as supplementary information.

[Discussion]

Do not recapitulate the results, but discuss their significance against the background of existing knowledge, and identify clearly those aspects that are novel. The final paragraph should highlight the main conclusion(s), and provide some indication of the direction future research should take.

This section may be divided into subheadings to assist the reader. Results and Discussion may be combined.

[Acknowledgements] These should be brief, and should include sources of financial support, material (e.g. novel compounds, strains, etc.) not available commercially, personal assistance, advice from colleagues and gifts.

[References]

Authors are responsible for the accuracy of the references. Published articles as well as those in press (please state the name of the journal and enclose a copy of the manuscript) may be included. In the text of the manuscript, references to the literature should be numbered consecutively and indicated by a superscript. Each reference should be numbered individually and listed at the end of the manuscript; examples are given below. All authors should be quoted for papers with up to six authors; for papers with seven or more authors, the first six authors only should be quoted, followed by *et al.*,

Example of journal citations:

1. Nakashima, H., Hasegawa, T., Sakai, M., Inaba, R. & Imamura, T. Identification of iso(18p) marker chromosome by fluorescence in situ hybridization with single-copy DNA probe. *Jpn J. Hum. Genet.* **40**, 185–188 (1995).
2. Oishi, T., Iida, A., Otsubo, W., Kamatani, Y., Usami, M., Takei, T., Uchida, K., Tsuchiya, K. *et al.*, A functional SNP in the NKX2.5-binding site of ITPR3 promoter is associated with susceptibility to systemic lupus erythematosus in Japanese population. *J. Hum. Genet.* **53**, 151-162 (2008).

Example of book citations:

1. Vogel, F. & Motulsky, A. G. *Human Genetics — Problems and Approaches* 3rd edn (Springer, New York, 1997).

Example of an article in a book:

1. Kimura, M. in *Evolution of Life* (eds Osawa, S. & Honjo, T.) 67–78 (Springer, New York, 1991).

Journal names are italicized and abbreviated (with periods) according to common usage; refer to Index Medicus for details.

Citations of ‘unpublished data’ and ‘S. A. Waksman, personal communication’ should be written in the text parenthetically. Written proof for ‘personal communication’ and preprint for ‘in press’ may be requested for review.

Figures These should be labelled sequentially as Figure 1, Figure 2, etc. **Each figure should be saved in a separate file**, numbered and titled and cited in the text. Figure legends should be

printed, double spaced, on a separate sheet titled 'Titles and legends to figures'. Figures should be referred to specifically in the text of the paper but should not be embedded within the text. The use of three-dimensional histograms is strongly discouraged when the addition of the third dimension gives no extra information. If a table or figure has been published before, the authors must obtain written permission to reproduce the material in both print and electronic formats from the copyright owner and submit it with the manuscript. This follows for quotes, illustrations and other materials taken from previously published works not in the public domain. The original source should be cited in the figure caption or table footnote. Colour figures can be reproduced if necessary, but the authors will be expected to contribute towards the cost of publication. Please note that colour figures can be reproduced on the web for FREE (see below for further details).

At submission, ALL figures should be of a high enough quality to be assessed in the peer review process. A minimum resolution of 300 dpi is required at the size the image is to appear in print. The minimum resolution for images containing text should be 400dpi and 1000dpi for images containing line art. Please refer to the Artwork Guidelines for details of artwork (Figures and Images) preparation.

Colour figures can be reproduced if necessary, but the authors will be expected to contribute towards the cost of publication. A quote will be supplied upon acceptance of your paper.

Tables These should be labelled sequentially as Table 1, Table 2, etc. **Each table should be saved in a separate file**, numbered and titled, and cited in the text. Reference to table footnotes should be made by means of Arabic numerals. Tables should not duplicate the content of the text. They should consist of at least two columns; columns should always have headings. Authors should ensure that the data in the tables are consistent with those cited in the relevant places in the text, totals add up correctly, and percentages have been calculated correctly. Unlike figures or images, tables may be embedded into the word processing software if necessary, or supplied as separate electronic files.

Supplementary information Supplementary information is peer-reviewed material directly relevant to the conclusion of an article that cannot be included in the printed version owing to space or format constraints. It is posted on the journal's web site and linked to the article when the article is published and may consist of data files, graphics, movies or extensive tables.

The printed article must be complete and self-explanatory without the supplementary information. Supplementary information enhances a reader's understanding of the paper but is not essential to that understanding.

Supplementary information must be supplied to the editorial office in its final form for peer review. On acceptance the final version of the peer reviewed supplementary information should be submitted with the accepted paper.

To ensure that the contents of the supplementary information files can be viewed by the editor(s), referees and readers, please also submit a 'read-me' file containing brief instructions on how to use the file.

The supplementary information may not be altered, nor new supplementary information added, after the paper has been accepted for publication.

Supplying supplementary information files

Authors should ensure that supplementary information is supplied in its FINAL format because it is not subedited and will appear online exactly as originally submitted. It cannot be altered, nor new supplementary information added, after the paper has been accepted for publication.

Please supply the supplementary information via eJP, the electronic manuscript submission and tracking system, in an acceptable file format (see below). Authors should:

- Include a text summary (no more than 50 words) to describe the contents of each file.
- Identify the types of files (file formats) submitted.
- Include the text 'Supplementary information is available at (the journal's name)'s website' at the end of the article and before the references.
-

File sizes must be as small as possible, so that they can be downloaded quickly. Images should not exceed 640 x 480 pixels (9 x 6.8 inches at 72 pixels per inch) but we would recommend 480 x 360 pixels as the maximum frame size for movies. We would also recommend a frame rate of 15 frames per second. If applicable to the presentation of the supplementary information, use a 256-colour palette. Please consider the use of lower specification for all of these points if the supplementary information can still be represented clearly. Our recommended maximum data rate is 150 KB/s.

The number of files should be limited to eight, and the total file size should not exceed 8 MB. Individual files should not exceed 1 MB. Please seek advice from the editorial office before sending files larger than our maximum size to avoid delays in publication.

Further questions about the submission or preparation of supplementary information should be directed to the editorial office.

HOUSE STYLE

As the electronic submission will provide the basic material for typesetting, it is important that papers are prepared in the general editorial style of the journal.

1. See the artwork guidelines for information on labelling of figures
2. Do not make lines thinner than 1pt (0.36mm)
3. Use a coarse hatching pattern rather than shading for tints in graphs
4. Colour should be distinct when being used as an identifying tool

5. Use SI units throughout
6. Spaces, not commas, should be used to separate thousands
7. Abbreviations should be preceded by the words for which they stand in the first instance of use
8. Text should be double spaced with a wide margin

Nucleotide data

New nucleotide data must be deposited in the DDBJ/EMBL/GenBank databases and an accession number obtained before a paper can be accepted for publication. Submission to any one of the three collaborating databanks is sufficient to ensure data entry in all. The accession number should be included in the manuscript, e.g. as a footnote on the title page: The nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under the accession number(s) ----. If requested, the database will withhold release of data until publication. The most convenient method for submitting sequence data is by using the following URLs:

- DDBJ via SAKURA: <http://sakura.ddbj.nig.ac.jp/>
- EMBL via WEBIN: <http://www.ebi.ac.uk/embl/Submission/webin.html>
- GenBank via BankIt: <http://www.ncbi.nlm.nih.gov/BankIt/> or by stand-alone submission tool Sequin: <http://www.ncbi.nlm.nih.gov/Sequin/>

For special types of submissions (e.g. genomes and bulk submissions), additional submission systems are available at the following sites:

- DDBJ: Center for Information Biology and DNA Data Bank of Japan National Institute of Genetics, Yata, Mishima, Shizuoka 411-8540, JAPAN; telephone: +81-559-81-6853; fax: +81-559-81-6849; e-mail: ddbj@ddbj.nig.ac.jp URL: <http://www.ddbj.nig.ac.jp>
- EMBL: EMBL Nucleotide Sequence Submissions, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, U.K.; telephone: +44-1223-494400; fax: +44-1223-494472; e-mail: datasubs@ebi.ac.uk URL: <http://www.ebi.ac.uk>
- GenBank: National Center for Biotechnology Information, National Library of Medicine, Bldg. 38A, Rm 8N-803, Bethesda, Maryland 20894, USA; telephone: +1-301-496-2475; fax: +1-301-480-9241; e-mail: info@ncbi.nlm.nih.gov URL: <http://www.ncbi.nlm.nih.gov>

NON-NATIVE SPEAKERS OF ENGLISH

Researchers who are not native speakers of English who submit manuscripts to international journals often receive negative comments from referees or editors about the English-language usage in their manuscripts, and these problems can contribute to a decision to reject a paper. To

help reduce the possibility of such problems, we strongly encourage such authors to take at least one of the following steps:

- Have your manuscript reviewed for clarity by a colleague whose native language is English.
- Use one of the many English language editing services that are available, such as that offered by Nature Publishing Group Language Editing . An editor will improve the English to ensure that your meaning is clear and identify problems that require your review.

Please note that the use of Nature Publishing Group Language Editing is at the author's own expense and in no way implies that the article will be selected for peer review or accepted by an NPG journal (or any other journal). The decisions that the editors of any NPG journal make based on the quality and suitability of a manuscript for that journal are entirely independent of whether that manuscript has been language-edited by Nature Publishing Group Language Editing.

SUBMISSION OF PAPERS

The first thing you need to do, if you have not done so already, is register for an account. After this, please follow the instructions below to enable you to submit your article through our secure server.

NAVIGATING THE SYSTEM

When you first access our tracking system, you will be taken to your Home page, where different categories of tasks are listed. If you are required to perform a pending action item or task, there will be a red arrow → next to a 'Manuscript' link. Throughout the system, red arrows → reflect pending action items which you should address. If there are no red arrows visible on your Home page, then you are finished and have no outstanding tasks to complete.

At any time please press HOME to go to the submission home page.

PROCESS FOR MANUSCRIPT SUBMISSION

Please make sure you have gathered all the required manuscript information listed above **BEFORE** starting the submission process. The manuscript submission process starts when you press the "Submit Manuscript" link on your "Home" page. The manuscript submission process is broken down into a series of 4 primary tasks that gather detailed information about your manuscript and allow you to upload the pertinent text and figure/image files. The sequence of screens is as follows:

1. The 'Files' primary task allows you to select the actual file locations (via an open file dialogue). You will be able to 'Browse' for the relevant files on your computer. **Please include the figure number in the title line for each figure.** On the completion screen, you will be asked to specify the order in which you want the individual files to appear in the merged document. Editors and/or reviewers will also be able to look at the individual PDF files if necessary.
2. The 'Manuscript Information' primary task which asks for author details, the manuscript title, abstract, other associated manuscript information and types/number of files to be submitted. Please note, if you are the corresponding author please submit your details in the

corresponding author fields; DO NOT re-enter the same details in the contributing author fields.

3. The 'Validate' primary task gives you the opportunity to check and verify the manuscript files and manuscript information uploaded. If you are submitting manuscript files separately, we create a merged PDF containing your manuscript text, figures and tables to simplify the handling of your paper. You will need to approve the merged PDF file, and a PDF or any other file not included in the merge, to submit your manuscript. You may also update and/or change manuscript files and manuscript information by clicking on the 'Change' or 'Fix' links respectively.
4. The 'Submit' primary task is the last step in the manuscript submission process. At this stage the Manuscript Tracking System will perform a final check to ensure that all mandatory fields have been completed. Any incomplete fields will be flagged by a red arrow and highlighted by a red box. Click on the 'Fix' link to return to relevant section for completion. Once your manuscript has been finalised, click on the 'Approve Submission' button to submit your manuscript for consideration. A 'Manuscript Approved' message will display on your author desktop to confirm the submission.

You will need to have the following details for all authors before commencing online submission. Items in parenthesis are not essential for co-authors:

- Email Addresses
- First and Last Names
- Institution
- (Full Postal Address)
- (Work Telephone Numbers)
- Fax Numbers

In addition you will need:

- Covering letter
- Title and Running head (you can copy and paste this from your manuscript)
- Abstract (you can copy and paste this from your manuscript)
- Manuscript files in Word, WordPerfect, text or any RTF format
- Figures/Images in external files in TIFF or JPEG, in either grayscale or CMYK colour, not in RGB
- Tables in Excel (preferred) as separate files or embedded at the end of the manuscript file

Do not embed images and figures within the text from word processing software as embedded images are not acceptable for production. (Tables are an exception to this rule as you may be generating them using the same software and as resolution quality tends to be less important for tables.)

ANNEXURE N: GUIDE FOR AUTHORS FOR THE EUROPEAN JOURNAL OF HUMAN GENETICS

Welcome to the electronic manuscript submission website for *European Journal of Human Genetics*. The instructions below are structured so you can quickly and easily answer the following questions:

1. Is my manuscript suitable for *European Journal of Human Genetics*?
2. How do I format my manuscript for *European Journal of Human Genetics*?
3. How do I submit my manuscript to *European Journal of Human Genetics*?

Scope

The *European Journal of Human Genetics* is the official Journal of the European Society of Human Genetics, publishing high-quality, research articles, short reports and reviews in the rapidly expanding field of human genetics and genomics. It covers molecular, clinical and cytogenetics, interfacing between advanced biomedical research and the clinician, and bridging the great diversity of facilities, resources and viewpoints in the genetics community. Key areas include:

- Monogenic and multifactorial disorders
- Development and malformation
- Hereditary cancer
- Medical Genomics
- Gene mapping and functional studies
- Genotype-phenotype correlations
- Genetic variation and genome diversity
- Statistical and computational genetics
- Bioinformatics
- Advances in diagnostics
- Therapy and prevention
- Animal models
- Genetic services
- Community genetics

The journal also publishes invited editorials and commentaries, announcements of Societal and other European activities and special issues of general interest for the Human Genetics community.

Editorial Policy

Editor-in-Chief:

Professor G.-J.B. van Ommen, Department of Human Genetics, Leiden University Medical Centre

The *European Journal of Human Genetics* publishes Research Articles, review articles, short reports, scientific correspondence, news and commentary articles and announcements.

All manuscripts are subject to editorial review. Authors of accepted papers will be required to sign a Licence to Publish form.

Patient consent

For all articles that include information or clinical photographs relating to individual patients, where those patients could be identified by readers in any way, written and signed consent to publish **must** be obtained from each patient and a **patient consent form** should be mailed or faxed to the editorial office. Do not use patient's names, initials or hospital numbers, especially on any illustrative material.

Non-Native Speakers of English

Authors who are not native speakers of English who submit manuscripts to international journals often receive negative comments from referees or editors about the English-language usage in their manuscripts, and these problems can contribute to a decision to reject a paper. To help reduce the possibility of such problems, we strongly encourage such authors to take at least one of the following steps:

- Have your manuscript reviewed for clarity, and corrected by a colleague whose native language is English
- Use one of the many English language editing services that are available, such as that offered by Nature Publishing Group Language Editing, or others listed below. An editor will improve the English to ensure that your meaning is clear and identify problems that require your review
- Please note that the use of Nature Publishing Group Language Editing is at the author's own expense and in no way implies that the article will be selected for peer review or accepted by an NPG journal (or any other journal). The decisions that the editors of any NPG journal make based on the quality and suitability of a manuscript for that journal are entirely independent of whether that manuscript has been language-edited by Nature Publishing Group Language Editing

Format of Papers

Preparation of manuscripts

Page allowance: see sub-section "Types of Papers" below; totals quoted are for A4 size paper.

Margins: at least 3 cm all round.

Font/size: Times New Roman 12pt (or equivalent) throughout, including references.

Line spacing: Double-line spacing throughout, including references.

References: Numerical - see sub-section "References".

Figure Legend(s): These should be listed on a separate page following the reference section.

Figures and Tables: for format information see sub-sections "Figures" - "Tables".
Articles: maximum of 4 of each; additional and large figures/tables can be submitted as Supplemental Material.

Short Reports: a total of 3 figures/tables.

For all papers, authors are urged to write as concisely as possible. The online submission form will ask you to put forward the names and email addresses of four experts in the appropriate area of research for the manuscript. The online form will also ask you to confirm that all contributing authors have consented to publication of the material you are submitting.

Approved human gene symbols should be obtained prior to submission from:

HUGO Gene Nomenclature Committee (HGNC),
 European Bioinformatics Institute (EMBL-EBI),
 Wellcome Trust Genome Campus,
 Hinxton, Cambridgeshire
 CB10 1SA, UK
 Fax: +44 (0)1223 494 468
[E-mail: hgnc@genenames.org](mailto:hgnc@genenames.org)
<http://www.genenames.org>

Approved mouse nomenclature should be obtained prior to submission from:

Lois Maltais,
 The Jackson Laboratory,
 Bar Harbor,
 Maine 04609-0800, USA
 Tel: +1 207 288 6429
 Fax: +1 207 288 6132
[E-mail: nomen@informatics.jax.org](mailto:nomen@informatics.jax.org)
<http://www.informatics.jax.org/mgihome/nomen/>

Mutations described should adhere to nomenclature developed by the Human Genome Variation Society:

Describing sequence variants

For acceptance, manuscripts should be compliant with the HGVS rules to describe sequence variants (mutations, polymorphisms), see <http://www.hgvs.org/mutnomen/> and Taschner and Den Dunnen, 2011. Hum Mutat 32:507-511. If alternative nomenclature schemes are common in a specific field, they may be used in addition to the approved nomenclature, but they must be defined unambiguously. Variants may be described using dbSNP identifiers (e.g., rs123456:A>G). All reference sequences used should be given (Material and Methods section and as a footnote in tables listing variants) using GenBank Accession and version number (e.g.: RefSeq NG_001234.2, NM_123456.3 or U654321.1). Acceptance and/or publication may be delayed if the guidelines are not followed properly. Authors are advised to check sequence variant descriptions using software tools like the Mutalyzer program (<http://www.lovd.nl/mutalyzer/>), which includes a batch mode.

Submission of Data to Genetic Database

The European Journal of Human Genetics supports the recommendations of the Human Variome Project (Cotton RG *et al.*, 2007. Nat Genet. 39:433 <http://www.nature.com/npg/journal/v39/n4/full/ng2024.html>). Authors are required to submit all variants described to a public database, e.g. the relevant gene variant database (or LSDB), prior to acceptance. Authors must confirm the status of database submission in their cover letter. In addition, authors should note in the manuscript (e.g., Material and Methods section) the database(s) to which they have submitted the variants and provide the URL. Further information and links to gene variant databases (or LSDB) can be obtained from the HGVS (<http://www.hgvs.org/dblist/dblist.html>).

Reporting Guidelines for Systematic Reviews and Meta-Analyses

Reports of systematic reviews and meta-analyses should use the PRISMA statement as a guide, and include a completed PRISMA checklist and flow diagram to accompany the main text. Blank templates of the checklist and flow diagram can be downloaded from the PRISMA website.

Guidelines about what is to be expected of a manuscript that is analysing biomedical data with bioinformatics

1. Appropriate validation of the method using an independent dataset. New methods should be compared with the most important previous methods, and in general, an advantage of the new method in at least one substantial aspect should be demonstrated.
2. Detailed explanation and motivation of the computational and statistical methods.
3. Desirable to have website with demonstration of method or downloadable code with self-contained example.
4. Results of paper should ideally be such that they can help to plan further experiments, generate testable hypotheses, or should be of major biological or medical interest. Manuscripts that only present anecdotal usages of descriptive analyses (e.g., Gene Ontology over-representation) are discouraged.

Types of Papers

Articles

Full Length Manuscripts should not exceed 20 pages of A4 including references. Manuscript length may influence decision to publish or not.

Short

Reports

These manuscripts should not exceed 4 printed pages (ie 6 double-spaced pages or approximately 1500 words), including an abstract, essential references and not more than 3 tables or figures. Such communications should represent complete, original studies and should be arranged in the same way as full length manuscripts.

Correspondence

Letters to the Editor are encouraged. They may deal with material in published papers or they may raise new issues. In the former case the Editor may send the letter first to the authors of the original paper so that their comments may be published at the same time as the letter.

Arrangement of Articles

Cover Letter: Please include the name, institution and e-mail address of four suggested referees in your covering letter. You must also include a conflict of interest statement for each contributing author in the manuscript Cover Letter (for guidelines on writing a Conflict of Interest statement please see the end of the “Format of Papers” section).

Title page: This should bear the title of the paper (length of no more than 145 characters including spaces), the full names of the authors and the composition of consortia if applicable and their affiliations together with the name, full postal address, telephone and fax numbers, and e-mail address of the author to whom correspondence and reprint requests are to be sent. Please do *not* capitalise the title. There should be a running title of not more than 50 characters, including spaces. Some of this information will need to be additionally copied into the online submission form.

Abstract: This is an unstructured paragraph which states the purpose, basic procedures, main findings and principal conclusions of the study. The abstract should not exceed 250 words or contain abbreviations or references.

Keywords: For indexing purposes, a list of 3-6 keywords is essential. You will be prompted for these on the online submission form, but should include them with the manuscript too.

Introduction: This should assume that the reader is knowledgeable in the field and should therefore be as brief as possible.

Materials and Methods: Methods that have been published in detail elsewhere should not be described in detail. Any equipment mentioned should specify the manufacturer and their location.

Results: These should be presented succinctly in the same order as the experiments are described in materials and methods. Tables and especially graphics are encouraged for quantitative information.

Discussion: Do not recapitulate the results, but discuss their significance against the background of existing knowledge, and identify clearly those aspects that are novel. The final paragraph should highlight the main conclusion(s), and provide some indication of the direction future research should take.

Acknowledgements: These should be brief.

Conflict of Interest Statement: A conflict of interest statement must be included for each contributing author. Please see the Conflict of Interest guidelines at the end of the “Format of Papers” section for more information and for guidelines on what constitutes a conflict of interest.

References: Only papers closely related to the author's work should be quoted. Exhaustive lists should be avoided. References should appear as superscript numbers starting at 1. At the end of the paper they should be listed (double-spaced) in numerical order corresponding to the order of citation in the text. If there are 6 or fewer authors list them all; if there are 7 or more list the first 3 followed by *et al.*, Abbreviations for titles of medical periodicals should conform to those used in the latest edition of *Index Medicus*. The first and last page numbers for each reference should be provided. Abstracts and letters must be identified as such. Papers in press may be included in the list of references. Papers submitted for publication and papers

presented at meetings should not be included as references; nor should abstracts of papers presented at meetings not in the public domain. These should be cited as a personal communication in the text.

Examples of References

Journal article:

1 Brown CJ, Ballabio A, Rupert JL *et al*: A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 1991; **349**: 38-44.

Monographs:

2 Heim S, Mitelman F: Cancer Cytogenetics. New York, Liss, 1987.

Edited books:

3 Goddard AD, Solomon S: Genetic aspects of cancer; in Harris H, Hirschhorn K (eds): Advances in Human Genetics. New York, 1993, vol 21, pp 321-376.

Where a reference is to appear next to a number in the text, for example following an equation, chemical formula or biological acronym, citations should be written as (ref. X).

Example. “detectable levels of endogenous Bcl-2 (ref. 3), as confirmed by western blot”

Figures

Figures and images should be labelled sequentially, numbered and cited in the text. Figure legends should be printed, double spaced, on a separate sheet titled ‘Titles and legends to figures’. Figures should be referred to specifically in the text of the paper but should not be embedded within the text. The use of three-dimensional histograms is strongly discouraged when the addition of the third dimension gives no extra information. If a table or figure has been published before, the authors must obtain written permission to reproduce the material in both print and electronic formats from the copyright owner and submit it with the manuscript. This follows for quotes, illustrations and other materials taken from previously published works not in the public domain. The original source should be cited in the figure caption or table footnote. Colour figures can be reproduced if necessary, but the authors will be expected to contribute towards the cost of publication. A quote will be supplied upon acceptance of your paper.

Artwork Guidelines

Detailed guidelines for submitting artwork can be found by downloading the [guidelines PDF](#). Using the guidelines, please submit production quality artwork with your initial online submission. If you have followed the guidelines, we will not require the artwork to be resubmitted following the peer-review process, if your paper is accepted for publication.

Colour on the web

Authors who wish their articles to have FREE colour figures on the web (only available in the HTML (full text) version of manuscripts) must supply separate files in the following format. These files should be submitted as supplementary information and authors are asked to mention they would like colour figures on the web in their submission letter.

Tables

These should be labelled sequentially as Table 1, Table 2, etc. Each table should be typed on a separate page, numbered and titled, and cited in the text. Reference to table footnotes should be made by means of Arabic numerals. Tables should not duplicate the content of the text. They should consist of at least two columns; columns should always have headings. Authors should ensure that the data in the tables are consistent with those cited in the relevant places in the text, totals add up correctly, and percentages have been calculated correctly. Unlike figures or images, tables may be embedded into the word processing software if necessary, or supplied as separate electronic files.

House Style

As the electronic submission will provide the basic material for typesetting, it is important that papers are prepared in the general editorial style of the journal.

1. See the [artwork guidelines](#) for information on labeling of figures
2. Do not make rules thinner than 1pt (0.36mm)
3. Use a coarse hatching pattern rather than shading for tints in graphs
4. Color should be distinct when being used as an identifying tool
5. Use Si units throughout
6. Spaces, not commas should be used to separate thousands
7. Abbreviations should be preceded by the words for which they stand in the first instance of use
8. Text should be double spacing with a wide margin
9. At first mention of a manufacturer, the town (state if USA) and country should be provided

File Formats

File formats for manuscript files, figures and tables that are acceptable for our electronic manuscript submission process are given on the online forms. Further advice on file types is also available from the [Tips](#) webpage. Please follow our [artwork guidelines](#) for submitting figures, and use a common word-processing package (such as Microsoft Word) for the text. Either embed tables converted into images at the end of your Word document, or as a separate files in which ever program you used to generate them. If you submit raw data, this can be done in Excel, or tab/comma delimited format.

Supplementary information

Supplementary information is material directly relevant to the conclusion of an article that cannot be included in the printed version owing to space or format constraints. It is posted on the journal's web site and linked to the article when the article is published and may consist of data files, graphics, movies or extensive tables.

The printed article must be complete and self-explanatory without the supplementary information. Supplementary information enhances a reader's understanding of the paper but is not essential to that understanding.

Supplementary information must be supplied to the editorial office in its final form for peer review. On acceptance the final version of the peer reviewed supplementary information should be submitted with the accepted paper.

To ensure that the contents of the supplementary information files can be viewed by the editor(s), referees and readers, please also submit a 'read-me' file containing brief instructions on how to use the file.

Please note: currently, data submitted under Supplemental Material is not visible to reviewers and if it is essential that reviewers have access to this information, it should initially be included (as Supplemental Material) as either part of the Article file or under Tables or Figures with the headings Supplemental Table ... or Supplemental Figure

Supplying supplementary information files

Authors should ensure that supplementary information is supplied in its FINAL format because it is not subedited and will appear online exactly as originally submitted. It cannot be altered, nor new supplementary information added, after the paper has been accepted for publication.

Please supply the supplementary information via eJP, the electronic manuscript submission and tracking system, in an acceptable file format (see below).
Authors should:

- ☐ Include a text summary (no more than 50 words) to describe the contents of each file.
- ☐ Identify the types of files (file formats) submitted.
- ☐ Include the text 'Supplementary information is available at (the journal's name)'s website' at the end of the article and before the references.

File sizes must be as small as possible, so that they can be downloaded quickly. Images should not exceed 640 x 480 pixels (9 x 6.8 inches at 72 pixels per inch) but we would recommend 480 x 360 pixels as the maximum frame size for movies. We would also recommend a frame rate of 15 frames per second. If applicable to the presentation of the supplementary information, use a 256 colour palette. Please consider the use of lower specification for all of these points if the supplementary information can still be represented clearly. Our recommended maximum data rate is 150 KB/s.

The number of files should be limited to eight, and the total file size should not exceed 8 MB. Individual files should not exceed 1 MB. Please seek advice from the editorial office before sending files larger than our maximum size to avoid delays in publication.

Further questions about the submission or preparation of supplementary information should be directed to the editorial office.

Conflict of interest

In the interests of transparency and to help readers form their own judgments of potential bias authors must declare whether or not there is any competing financial interests in relation to the work described. This information must be included in their cover letter and in the conflict of interest section of their manuscript. In cases where the authors declare a competing financial interest, a statement to that effect is published as part of the article. If no such conflict exists, the statement will simply read that the authors have nothing to disclose.

For the purposes of this statement, competing interests are defined as those of a financial nature that, through their potential influence on behaviour or content, or from perception of such potential influences, could undermine the objectivity, integrity or perceived value of a publication. They can include any of the following:

- Funding: Research support (including salaries, equipment, supplies, reimbursement for attending symposia, and other expenses) by organizations that may gain or lose financially through this publication. The role of the funding body in the design of the study, collection and analysis of data and decision to publish should be stated.
- Employment: Recent (while engaged in the research project), present or anticipated employment by any organization that may gain or lose financially through this publication.
- Personal financial interests: Stocks or shares in companies that may gain or lose financially through publication; consultation fees or other forms of remuneration from organizations that may gain or lose financially; patents or patent applications whose value may be affected by publication.

It is difficult to specify a threshold at which a financial interest becomes significant, but note that many US universities require faculty members to disclose interests exceeding \$10,000 or 5% equity in a company. Any such figure is arbitrary, so we offer as one possible practical alternative guideline: "Declare all interests that could embarrass you were they to become publicly known after your work was published." We do not consider diversified mutual funds or investment trusts to constitute a competing financial interest.

The statement must contain an explicit and unambiguous statement describing any potential conflict of interest, or lack thereof, for any of the authors as it relates to the subject of the report. Examples include "Dr. Smith receives compensation as a consultant for XYZ Company," "Dr. Jones and Dr. Smith have financial holdings in ABC Company," or "Dr. Jones owns a patent on the diagnostic device described in this report." These statements acknowledging or denying conflicts of interest must be included in the manuscript under the heading Conflict of Interest. The Conflict of Interest disclosure appears in the cover letter, in the manuscript submission process and before the References section in the manuscript.

Following the Conflict of Interest heading, there must be a listing for each author, detailing the professional services relevant to the submission. Neither the precise amount received from each entity nor the aggregate income from these sources needs to be provided. Professional services include any activities for which the individual is, has been, or will be compensated with cash, royalties, fees, stock or stock options in exchange for work performed, advice or counsel provided, or for other services related to the author's professional knowledge and skills. This would include, but not necessarily be limited to, the identification of organizations from which the author received contracts or in which he or she holds an equity stake if professional services were provided in conjunction with the transaction.

Examples of declarations are:

Conflict of interest.

The authors declare no conflict of interest.

Conflict of interest.

Dr Caron's work has been funded by the NIH. He has received compensation as a member of the scientific advisory board of Acadia Pharmaceutical and owns stock in the company. He also has consulted for Lundbeck and received compensation. Dr Rothman and Dr Jensen declare no potential conflict of interest.

Submission of papers

The first thing you need to do, if you have not done so already, is [register for an account](#). After this, please consult the instructions below to enable you to submit your article through our secure server.

Process for Manuscript Submission

Please make sure you have gathered all the required manuscript information listed above **BEFORE** starting the submission process. The manuscript submission process starts by pressing the "Submit Manuscript" link on your "Home" page.

The manuscript submission process is broken down into a series of 4 primary tasks that gather detailed information about your manuscript and allow you to upload the pertinent text and figure/image files. The sequence of screens is as follows:

1. The 'Files' primary task allows you to select the actual file locations (via an open file dialogue). You will be able to 'Browse' for the relevant files on your computer. **Please include the figure number in the title line for each figure.** On the completion screen, you will be asked to specify the order in which you want the individual files to appear in the merged document. Editors and/or reviewers will also be able to look at the individual PDF files if necessary.
2. The 'Manuscript Information' primary task which asks for author details, the manuscript title, abstract, other associated manuscript information and types/number of files to be submitted. Please note, if you are the corresponding author please submit your details in the corresponding author fields; DO NOT re-enter the same details in the contributing author fields.
3. The 'Validate' primary task gives you the opportunity to check and verify the manuscript files and manuscript information uploaded. If you are submitting manuscript files separately, we create a merged PDF containing your manuscript text, figures and tables to simplify the handling of your paper. You will need to approve the merged PDF file, and a PDF or any other file not included in the merge, to submit your manuscript. You may also update and/or change manuscript files and manuscript information by clicking on the 'Change' or 'Fix' links respectively.
4. The 'Submit' primary task is the last step in the manuscript submission process. At this stage the Manuscript Tracking System will perform a final check to ensure that all mandatory fields have been completed. Any incomplete fields will be flagged by a red arrow and highlighted by a red box. Click on the 'Fix' link to return to relevant section for completion. Once your manuscript has been finalised, click on the 'Approve Submission' button to submit your manuscript for consideration. A 'Manuscript Approved' message will display on your author desktop to confirm the submission.

You will need to have the following details for all authors before commencing online submission. Items in parenthesis may not be compulsory for co-authors:

- Email Addresses
- First and Last Names
- Institution
- (Full Postal Address)
- (Work Telephone Numbers)
- Fax Numbers

In addition you will need:

- Covering letter (including Conflict of Interest statement)
- Title and Running Title (you can copy and paste this from your manuscript)
- Abstract (you can copy and paste this from your manuscript)
- Manuscript files in Word, WordPerfect, text or any RTF format
- Figures/Images in external files in TIFF or JPG, in either grayscale or CMYK colour, not in RGB
- Tables in Excel (preferred) as separate files or embedded at the end of the manuscript file

Do not embed images and figures within the text from word processing software as embedded images are not acceptable for production. (Tables are an exception to this rule as you may be generating them using the same software and as resolution quality tends to be less important for tables.)

Saving files with Microsoft Office 2007

Microsoft Office 2007 saves files in an XML format by default (file extensions .docx, .pptx and .xlsx). Files saved in this format cannot be accepted for publication.

Save Word documents using the file extension .doc

- Select the Office Button in the upper left corner of the Word 2007 Window and choose "Save As"
- Select "Word 97-2003 Document"
- Enter a file name and select "Save"

These instructions also apply for the new versions of Excel and PowerPoint.

Equations in Word must be created using Equation Editor 3.0

Equations created using the new equation editor in Word 2007 and saved as a "Word 97-2003 Document" (.doc) are converted to graphics and can no longer be edited. To insert or change an equation with the previous equation editor:

- Select "Object" on the "Text" section of the "Insert" tab
- In the drop-down menu - select "Equation Editor 3.0"

Do not use the "Equation" button in the "Symbols" section of the "Insert" tab.

Conflict of interest

It is essential that you note whether or not there is any conflict of interest in the submission form. This does not act as a substitute for the written statement that must be provided in the manuscript and the cover letter.

Manuscript Status

After you approve your manuscript it is submitted and you will receive an acknowledgement email. You can check the status of your manuscript at any time in the review process by:

1. Accessing the system with your password or link sent to you in the acknowledgement email
2. Clicking on the link represented by your manuscript tracking number and abbreviated title.
3. Clicking on the "Check Status" link at the bottom of the displayed page.

This procedure will display tracking information about where your manuscript is in the submission/peer review process.

Licence to Publish

The corresponding author must complete and sign the Licence to Publish form upon acceptance of the manuscript and return it to the editorial office. Failure to do so will result in delays to the publication of your paper. A copy of the Licence to Publish form can be found at <http://www.nature.com/licenceforms/ejhg/ejhg-ltp.pdf>

European Journal of Human Genetics Open

Upon submission of an original research paper, authors can indicate within the manuscript tracking system whether they wish to pay an article processing charge for their article to be published open access. Publishing open access will mean the paper is freely accessible online immediately upon publication. By paying this charge authors are also permitted to post the final, published PDF of their article on a website, institutional repository or other free public server, immediately on publication.

Open Access articles can be published under one of three Creative Commons licences, at the free choice of the authors.

Under the **Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported Licence** (CC BY-NC-SA), users are free to share (copy, distribute and transmit) and remix (adapt) the contribution under the following conditions ([read full legal code](#)):

- Attribution. Users must attribute the contribution in the manner specified by the author or licensor (but not in any way that suggests that they or their use of the contribution is endorsed by the author or licensor).
- Noncommercial. Users may not use this work for commercial purposes.
- Share Alike. If users alter, transform, or build upon this work, they may distribute the resulting work only under the same or similar licence to this one.

Under the **Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported Licence** (CC BY-NC-ND), users are free to share (copy, distribute and transmit) the contribution under the following conditions ([read full legal code](#)):

- Attribution. Users must attribute the contribution in the manner specified by the author or licensor (but not in any way that suggests that they or their use of the contribution is endorsed by the author or licensor).
- Non-commercial. Users may not use this contribution for commercial purposes.
- No derivative works. Users may not alter, transform, or build upon this work.

Under the **Creative Commons Attribution 3.0 Unported Licence**, users are free to share (copy, distribute and transmit) and remix (adapt) the contribution, even commercially, under the following conditions ([read full legal code](#)):

- Attribution. Users must attribute the contribution in the manner specified by the author or licensor (but not in any way that suggests that they or their use of the contribution is endorsed by the author or licensor)

Authors should note that some funders require papers to be published under a specific licence and so should check the funder mandate to ensure compliance.

For further information on Open Access please see the [FAQs](#). Please note that editors and reviewers will be blind to your choice to ensure an impartial decision on your paper.

ANNEXURE O: GUIDE FOR AUTHORS FOR GENES AND NUTRITION

TYPES OF PAPERS

1. Research Papers
2. Reviews
3. Commentaries
4. Editorials
5. Literature Highlights

MANUSCRIPT SUBMISSION

Submission of a manuscript implies: that the work described has not been published before; that it is not under consideration for publication anywhere else; that its publication has been approved by all co-authors, if any, as well as by the responsible authorities – tacitly or explicitly – at the institute where the work has been carried out. The publisher will not be held legally responsible should there be any claims for compensation.

- Each paper submitted must be accompanied by the “Conflict of Interest Disclosure Form.” This form can be obtained at <http://www.springer.com/12663>.
- Each paper submitted must also include the following:

Compliance with Ethics Guidelines

In order to comply with ethical requirements, every paper must include the following passages in a separate section immediately preceding the reference list. The section should be called “Compliance with Ethics Guidelines.”

Conflict of Interest:

The Conflict of Interest statements should list each author separately by name:

John Smith declares that he has no conflict of interest.

Paula Taylor has received research grants from Drug Company A.

Mike Schultz has received a speaker honorarium from Drug Company B and owns stock in Drug Company C.

If multiple authors declare no conflict, this can be done in one sentence:

John Smith, Paula Taylor, and Mike Schultz declare that they have no conflict of interest.

For studies with human subjects include the following:

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000 (5). Informed consent was obtained from all patients for being included in the study.

If doubt exists whether the research was conducted in accordance with the Helsinki Declaration, the authors must explain the rationale for their approach, and demonstrate that the institutional review body explicitly approved the doubtful aspects of the study.

If any identifying information about patients is included in the article, the following sentence should also be included:

Additional informed consent was obtained from all patients for which identifying information is included in this article.

For studies with animals include the following sentence:

All institutional and national guidelines for the care and use of laboratory animals were followed.

For articles that do not contain studies with human or animal subjects performed by any of the authors:

This article does not contain any studies with human or animal subjects performed by the any of the authors.

- For more general guidance on "Ethical standards" and "Conflict of interest," please see the Section on "Integrity of Research and Reporting" below.

TITLE PAGE

Title Page

The title page should include:

- The name(s) of the author(s)
- A concise and informative title
- The affiliation(s) and address(es) of the author(s)
- The e-mail address, telephone and fax numbers of the corresponding author

Abstract

Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references.

Keywords

Please provide 4 to 6 keywords which can be used for indexing purposes.

TEXT

Text Formatting

Manuscripts should be submitted in Word.

- Use a normal, plain font (e.g., 10-point Times Roman) for text.
- Use italics for emphasis.
- Use the automatic page numbering function to number the pages.
- Do not use field functions.
- Use tab stops or other commands for indents, not the space bar.
- Use the table function, not spreadsheets, to make tables.

- Use the equation editor or MathType for equations.
- Save your file in docx format (Word 2007 or higher) or doc format (older Word versions). Manuscripts with mathematical content can also be submitted in LaTeX.
- LaTeX macro package (zip, 182 kB)

Headings

Please use no more than three levels of displayed headings.

Abbreviations

Abbreviations should be defined at first mention and used consistently thereafter.

Footnotes

Footnotes can be used to give additional information, which may include the citation of a reference included in the reference list. They should not consist solely of a reference citation, and they should never include the bibliographic details of a reference. They should also not contain any figures or tables.

Footnotes to the text are numbered consecutively; those to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data). Footnotes to the title or the authors of the article are not given reference symbols.

Always use footnotes instead of endnotes.

Acknowledgments

Acknowledgments of people, grants, funds, etc. should be placed in a separate section before the reference list. The names of funding organizations should be written in full.

SCIENTIFIC STYLE

- Please always use internationally accepted signs and symbols for units (SI units).
- Nomenclature: Insofar as possible, authors should use systematic names similar to those used by Chemical Abstract Service or IUPAC.
- Genus and species names should be in italics.
- Generic names of drugs and pesticides are preferred; if trade names are used, the generic name should be given at first mention.

REFERENCES

Citation

Cite references in the text by name and year in parentheses. Some examples:

- Negotiation research spans many disciplines (Thompson 1990).
- This result was later contradicted by Becker and Seligman (1996).
- This effect has been widely studied (Abbott 1991; Barakat *et al.*, 1995; Kelso and Smith 1998; Medvec *et al.*, 1999).

Reference list

The list of references should only include works that are cited in the text and that have been published or accepted for publication. Personal communications and unpublished works should only be mentioned in the text. Do not use footnotes or endnotes as a substitute for a reference list.

Reference list entries should be alphabetized by the last names of the first author of each work.

- Journal article
Gamelin FX, Baquet G, Berthoin S, Thevenet D, Nourry C, Nottin S, Bosquet L (2009) Effect of high intensity intermittent training on heart rate variability in prepubescent children. *Eur J Appl Physiol* 105:731-738. doi: 10.1007/s00421-008-0955-8
Ideally, the names of all authors should be provided, but the usage of “et al” in long author lists will also be accepted:
Smith J, Jones M Jr, Houghton L et al (1999) Future of health insurance. *N Engl J Med* 341:325–329
- Article by DOI
Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. *J Mol Med*. doi:10.1007/s001090000086
- Book
South J, Blass B (2001) *The future of modern genomics*. Blackwell, London
- Book chapter
Brown B, Aaron M (2001) The politics of nature. In: Smith J (ed) *The rise of modern genomics*, 3rd edn. Wiley, New York, pp 230-257
- Online document
Cartwright J (2007) Big stars have weather too. IOP Publishing PhysicsWeb. <http://physicsweb.org/articles/news/11/6/16/1>. Accessed 26 June 2007
- Dissertation
Trent JW (1975) *Experimental acute renal failure*. Dissertation, University of California
Always use the standard abbreviation of a journal’s name according to the ISSN List of Title Word Abbreviations, see
www.issn.org/2-22661-LTWA-online.php
For authors using EndNote, Springer provides an output style that supports the formatting of in-text citations and reference list.
- EndNote style (zip, 3 kB)

TABLES

- All tables are to be numbered using Arabic numerals.
- Tables should always be cited in text in consecutive numerical order.
- For each table, please supply a table caption (title) explaining the components of the table.
- Identify any previously published material by giving the original source in the form of a reference at the end of the table caption.
- Footnotes to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data) and included beneath the table body.

ARTWORK AND ILLUSTRATIONS GUIDELINES

For the best quality final product, it is highly recommended that you submit all of your artwork – photographs, line drawings, etc. – in an electronic format. Your art will then be produced to the highest standards with the greatest accuracy to detail. The published work will directly reflect the quality of the artwork provided.

Electronic Figure Submission

- Supply all figures electronically.
- Indicate what graphics program was used to create the artwork.
- For vector graphics, the preferred format is EPS; for halftones, please use TIFF format. MS Office files are also acceptable.
- Vector graphics containing fonts must have the fonts embedded in the files.
- Name your figure files with "Fig" and the figure number, e.g., Fig1.eps.

Line Art

- Definition: Black and white graphic with no shading.
- Do not use faint lines and/or lettering and check that all lines and lettering within the figures are legible at final size.
- All lines should be at least 0.1 mm (0.3 pt) wide.
- Scanned line drawings and line drawings in bitmap format should have a minimum resolution of 1200 dpi.
- Vector graphics containing fonts must have the fonts embedded in the files.

Halftone Art

- Definition: Photographs, drawings, or paintings with fine shading, etc.
- If any magnification is used in the photographs, indicate this by using scale bars within the figures themselves.
- Halftones should have a minimum resolution of 300 dpi.

Combination Art

- Definition: a combination of halftone and line art, e.g., halftones containing line drawing, extensive lettering, color diagrams, etc.
- Combination artwork should have a minimum resolution of 600 dpi.

Color Art

- Color art is free of charge for online publication.
- If black and white will be shown in the print version, make sure that the main information will still be visible. Many colors are not distinguishable from one another when converted to black and white. A simple way to check this is to make a xerographic copy to see if the necessary distinctions between the different colors are still apparent.
- If the figures will be printed in black and white, do not refer to color in the captions.
- Color illustrations should be submitted as RGB (8 bits per channel).

Figure Lettering

- To add lettering, it is best to use Helvetica or Arial (sans serif fonts).
- Keep lettering consistently sized throughout your final-sized artwork, usually about 2–3 mm (8–12 pt).
- Variance of type size within an illustration should be minimal, e.g., do not use 8-pt type on an axis and 20-pt type for the axis label.
- Avoid effects such as shading, outline letters, etc.
- Do not include titles or captions within your illustrations.

Figure Numbering

- All figures are to be numbered using Arabic numerals.
- Figures should always be cited in text in consecutive numerical order.
- Figure parts should be denoted by lowercase letters (a, b, c, etc.).
- If an appendix appears in your article and it contains one or more figures, continue the consecutive numbering of the main text. Do not number the appendix figures, "A1, A2, A3, etc." Figures in online appendices (Electronic Supplementary Material) should, however, be numbered separately.

Figure Captions

- Each figure should have a concise caption describing accurately what the figure depicts. Include the captions in the text file of the manuscript, not in the figure file.
- Figure captions begin with the term Fig. in bold type, followed by the figure number, also in bold type.
- No punctuation is to be included after the number, nor is any punctuation to be placed at the end of the caption.
- Identify all elements found in the figure in the figure caption; and use boxes, circles, etc., as coordinate points in graphs.
- Identify previously published material by giving the original source in the form of a reference citation at the end of the figure caption.

Figure Placement and Size

- When preparing your figures, size figures to fit in the column width.
- For most journals the figures should be 39 mm, 84 mm, 129 mm, or 174 mm wide and not higher than 234 mm.
- For books and book-sized journals, the figures should be 80 mm or 122 mm wide and not higher than 198 mm.

Permissions

If you include figures that have already been published elsewhere, you must obtain permission from the copyright owner(s) for both the print and online format. Please be aware that some publishers do not grant electronic rights for free and that Springer will not be able to refund any costs that may have occurred to receive these permissions. In such cases, material from other sources should be used.

Accessibility

In order to give people of all abilities and disabilities access to the content of your figures, please make sure that

- All figures have descriptive captions (blind users could then use a text-to-speech software or a text-to-Braille hardware)
- Patterns are used instead of or in addition to colors for conveying information (color-blind users would then be able to distinguish the visual elements)
- Any figure lettering has a contrast ratio of at least 4.5:1

ELECTRONIC SUPPLEMENTARY MATERIAL

Springer accepts electronic multimedia files (animations, movies, audio, etc.) and other supplementary files to be published online along with an article or a book chapter. This feature can add dimension to the author's article, as certain information cannot be printed or is more convenient in electronic form.

Submission

- Supply all supplementary material in standard file formats.
- Please include in each file the following information: article title, journal name, author names; affiliation and e-mail address of the corresponding author.
- To accommodate user downloads, please keep in mind that larger-sized files may require very long download times and that some users may experience other problems during downloading.

Audio, Video, and Animations

- Always use MPEG-1 (.mpg) format.

Text and Presentations

- Submit your material in PDF format; .doc or .ppt files are not suitable for long-term viability.
- A collection of figures may also be combined in a PDF file.

Spreadsheets

- Spreadsheets should be converted to PDF if no interaction with the data is intended.
- If the readers should be encouraged to make their own calculations, spreadsheets should be submitted as .xls files (MS Excel).

Specialized Formats

- Specialized format such as .pdb (chemical), .vrl (VRML), .nb (Mathematica notebook), and .tex can also be supplied.

Collecting Multiple Files

- It is possible to collect multiple files in a .zip or .gz file.

Numbering

- If supplying any supplementary material, the text must make specific mention of the material as a citation, similar to that of figures and tables.
- Refer to the supplementary files as “Online Resource”, e.g., “... as shown in the animation (Online Resource 3)”, “... additional data are given in Online Resource 4”.
- Name the files consecutively, e.g. “ESM_3.mpg”, “ESM_4.pdf”.

Captions

- For each supplementary material, please supply a concise caption describing the content of the file.

Processing of supplementary files

- Electronic supplementary material will be published as received from the author without any conversion, editing, or reformatting.

Accessibility

In order to give people of all abilities and disabilities access to the content of your supplementary files, please make sure that

- The manuscript contains a descriptive caption for each supplementary material
- Video files do not contain anything that flashes more than three times per second (so that users prone to seizures caused by such effects are not put at risk)

AFTER ACCEPTANCE

Upon acceptance of your article you will receive a link to the special Author Query Application at Springer's web page where you can sign the Copyright Transfer Statement online and indicate whether you wish to order OpenChoice, offprints, or printing of figures in color.

Once the Author Query Application has been completed, your article will be processed and you will receive the proofs.

Open Choice

In addition to the normal publication process (whereby an article is submitted to the journal and access to that article is granted to customers who have purchased a subscription), Springer provides an alternative publishing option: Springer Open Choice. A Springer Open Choice article receives all the benefits of a regular subscription-based article, but in addition is made available publicly through Springer's online platform SpringerLink.

- Springer Open Choice

Copyright transfer

Authors will be asked to transfer copyright of the article to the Publisher (or grant the Publisher exclusive publication and dissemination rights). This will ensure the widest possible protection and dissemination of information under copyright laws.

Open Choice articles do not require transfer of copyright as the copyright remains with the author. In opting for open access, the author(s) agree to publish the article under the Creative Commons Attribution License.

Offprints

Offprints can be ordered by the corresponding author.

Color illustrations

Online publication of color illustrations is free of charge. For color in the print version, authors will be expected to make a contribution towards the extra costs.

Proof reading

The purpose of the proof is to check for typesetting or conversion errors and the completeness and accuracy of the text, tables and figures. Substantial changes in content, e.g., new results, corrected values, title and authorship, are not allowed without the approval of the Editor.

After online publication, further changes can only be made in the form of an Erratum, which will be hyperlinked to the article.

Online First

The article will be published online after receipt of the corrected proofs. This is the official first publication citable with the DOI. After release of the printed version, the paper can also be cited by issue and page numbers.