The relevance of specific c-reactive protein genetic variants towards cardiovascular disease risk in a black South African population undergoing an epidemiological transition

Bianca Swanepoel
20546025

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

Supervisor: Dr GW Towers
Co-supervisors: Dr KR Conradie and Dr C Nienaber-Rousseau

September (2013)

Die relevansie van spesifieke genetiese variante van c-reaktiewe proteïen teenoor kardiovaskulêre risiko in ’n swart Suid-Afrikaanse bevolking in ’n epidemiologiese oorgangsfase

Bianca Swanepoel
20546025

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

Studieleier: Dr GW Towers
Medestudieleiers: Dr KR Conradie and Dr C Nienaber-Rousseau

September (2013)

“Today is the day to start living your best life, to accept only the best, to only spend energy on the things that make you the best, and to create the best possible world around you. Life is short. Create the absolute best!”

ABSTRACT

Introduction: In Africa, it is estimated that cardiovascular disease (CVD) will affect approximately 1.3 million people per annum over the following 20 years. C-reactive protein (CRP) is a predictor of CVD risk and certain *CRP* gene polymorphisms can result in altered CRP concentrations. The distribution of *CRP* gene polymorphisms is ethnic-specific and extrapolating information from other populations to the black South African population, reported to harbour considerable genetic variation, should be avoided. This highlights the fact that genetic research among black South Africans is necessary.

Objectives: The main aim of this dissertation was to determine the association between various polymorphisms (reported and novel [single nucleotide polymorphisms (SNPs)] within the *CRP* gene with CRP concentrations [measured as high sensitivity (hs)-CRP concentrations] in a black South African population undergoing an epidemiological transition. Interactions between specific *CRP* polymorphisms and certain environmental factors on hs-CRP concentrations were also investigated.

Methods: This cross-sectional study (n=1,588) was nested within the Prospective Urban and Rural Epidemiological (PURE) study. Genotyping was performed using Illumina VeraCode technology on the BeadXpress® platform. Hs-CRP concentrations were measured by the use of a sequential multiple analyser computer (SMAC) through a particle-enhanced immunoturbidometric assay.

Results: All the SNPs adhered to the assumptions of Hardy-Weinberg equilibrium, although the distribution of several SNPs differed from that reported in other population groups. Three SNPs (rs3093058, rs3093062 and rs3093068) were associated with a significant ($p \leq 0.05$) increase in CRP concentrations. Five SNPs (rs1205, rs1341665, rs2794520, rs7553007 and rs2027471) were associated with a significant ($p \leq 0.05$) decrease in CRP concentrations. This difference in effect was most probably due to changes in gene function brought about by the localisation of these SNPs in the *CRP* gene. Men and urban individuals were more likely to present with significant associations between the SNPs investigated and CRP concentrations. The difference in the prevalence of the alleles associated with higher CRP concentrations in this population compared to non-African populations could possibly explain the increased CRP concentrations that are observed in the black South African population. Gene-gender (rs1205, rs1341665 and rs2027474) as well as gene-environmental (rs3093068) interactions were also observed.

Conclusions: CRP concentrations are in themselves a complex trait and there are many factors at play that influence their expression. Numerous factors (both genetic and environmental) are involved and no single factor acting alone is likely to have enough of an

influence to be used as a clinical diagnostic test of CRP concentrations. These results provide valuable information on the regulation of CRP in a black South African population as well as contribute to the literature of CRP on a global level.

Key words: cardiovascular disease; C-reactive protein; *CRP* polymorphisms; BeadXpress®; South African black population

OPSOMMING

Agtergrond: In Afrika word daar beraam dat kardiovaskulêre siekte (KVS) in die volgende 20 jaar ongeveer 1.3 miljoen mense per jaar gaan affekteer. C-reaktiewe proteïen (CRP) is 'n voorspeller van KVS-risiko en sekere enkelnukleotiedpolimorfismes (SNPs) in die *CRP*-geen kan veranderde CRP-konsentrasies tot gevolg hê. Die verspreiding van *CRP*-mutasies in verskillende bevolkingsgroepe verskil en daarom is dit belangrik om spesifiek in die swart Suid-Afrikaanse bevolking ondersoek in te stel.

Doelwit: Die hoofdoel van hierdie verhandeling was om die effek van verskeie *CRP*-mutasies op CRP-konsentrasies [gemeet as hoogs sensitiewe (hs)-CRP] te bepaal in 'n swart Suid-Afrikaanse bevolking in 'n proses van epidemiologiese verandering. Assosiasies tussen *CRP*-polimorfismes en sekere omgewingsfaktore op hs-CRP konsentrasies sal ook bepaal word.

Studieontwerp en metodes: Hierdie is 'n dwarssnitstudie (n=1,588) wat deel vorm van die internasionale Prospektiewe Stedelike en Landelike Epidemiologiese (PURE) studie. Die genotipering is met behulp van die Illumina® VeraCode-tegnologie gedoen op die BeadXpress®-platform. Hs-CRP konsentrasies is gemeet met behulp van 'n sekwensiële meervoudige analiseringsrekenaar (SMAC) deur middel van 'n partikel-versterkende immunoturbidometriese toets.

Resultate: Al die SNPs het voldoen aan die aannames van die Hardy-Weinberg-ekwilibrium, maar die verspreiding van sekere SNPs was anders as wat in ander bevolkinggroepe gerapporteer is. Drie van die SNPs (rs3093058, rs3093062 en rs3093068) is geassosieer met betekenisvol ($p \leq 0.05$) hoër CRP-konsentrasies, terwyl vyf SNPs (rs1205, rs1341665, rs2794520, rs7553007 en rs2027471) weer betekenisvol ($p \leq 0.05$) geassosieer is met laer CRP-konsentrasies. Betekenisvolle assosiasies tussen die SNPs wat ondersoek is en CRP-konsentrasies het meer dikwels voorgekom by mans en individue wat in 'n stedelike area gewoon het. Die waarskynlikste rede hiervoor is dat die funksie van die geen verander, afhangend van die area van die SNPs in die geen. Hierdie hoë voorkoms van die algemene alleel kan dan as verduideliking dien vir die hoë CRP-konsentrasies wat in hierdie swart Suid-Afrikaanse bevolking gesien is. Sekere SNPs het interaksies met geslag (rs1205, rs1341665 en rs2027474) getoon, asook met die omgewing (rs3093068).

Gevolgtrekking: CRP is 'n komplekse molekule en daar is baie faktore wat 'n invloed het op die uitdrukking van CRP. Talle faktore (geneties en omgewings) is betrokke en nie een enkele faktor kan alleenlik 'n groot genoeg invloed hê om gebruik te kan word as 'n

kliniese diagnostiese toets vir CRP konsentrasies. As navorsers uiteindelik die patogenetiese meganisme wil verstaan wat geassosieer word met CRP, moet alle faktore oorweeg word en in diepte bestudeer word. Hierdie resultate verskaf waardevolle inligting aangaande CRP in die swart Suid Afrikaanse populasie en dra by tot die literatuur op 'n globale vlak.

Sleutelwoorde: kardiovaskulêre siekte; C-reaktiewe proteïen; *CRP*-geen; polimorfisme; BeadXpress®; swart Suid-Afrikaanse bevolking

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A	adenine
ADT	assay design tool
ANCOVA	analyses of covariance
Ang II	angiotensin II
ASO	allele-specific oligonucleotides
AT1	angiotensin type-1
AT1R	angiotensin type-1 receptor
BMI	body mass index
bp	base pair
C	cytosine
cDNA	complementary DNA
CHD	coronary heart disease
CHO	carbohydrates
CRP	C-reactive protein
CVD	cardiovascular disease
DBP	diastolic blood pressure
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
ddNTPs	2’,3’-dideoxyribonucleotides triphosphate
EC	endothelial cells
EDTA	ethylenediamine tetra-acetic acid
eNOS	endothelial nitric oxide synthase
F	forward
ET-1	endothelin-1
G	guanine
GCKR	glucokinase regulatory protein
gDNA	genomic deoxyribonucleic acid
GI	glyceamic index
GL	glyceamic load
GWAS	genome wide association studies
HART	Hypertension in Africa Research Team
HbA1C	glycated haemoglobin
HDL-C	high-density lipoprotein-cholesterol
HIV	human immunodeficiency virus
HNF1A	hepatocyte nuclear factor 1
hs-CRP	high sensitivity C-reactive protein
HWE	Hardy-Weinberg equilibrium
ICAM	intercellular adhesion molecule
IDT	Integrated DNA Technologies
IL-1	interleukin 1
IL-6	Interleukin-6
IL-8	interleukin 8
iNOS	inducible nitric oxide synthase
Kbp	kilobasepairs
LDL-C	low density lipoprotein-cholesterol
LSO	locus-specific oligonucleotides
MAP	mitogen-activated protein
MCH	maternal and child health
MI	myocardial infarction
MMP-1	matrix metalloproteinase
mRNA	messenger ribonucleic acid
MUFA	mono-unsaturated fatty acids

NCD	non-communicable disease
NF-κB	Nuclear Factor-KappaB
NHLS	National Health Laboratory Service
NO	nitric oxide
NR-NCD	nutrition related non-communicable disease
ox-LDL	oxidised low density lipoprotein
PAI-1	plasminogen activator inhibitor type 1
PCR	polymerase chain reaction
PURE	Prospective Urban and Rural Epidemiological
QFFQ	qualitative food frequency questionnaires
R	reverse
ROS	reactive oxygen species
RNA	ribonucleic acid
rs	reference sequence
SBP	systolic blood pressure
SFA	saturated fatty acids
SMAC	Sequential Multiple Analyser Computer
SNP	single nucleotide polymorphism
T	thymine
Ta	annealing temperature
TC	total cholesterol
TE	total energy
TNF	tumor necrosis factor
UCSC	University of California Santa Cruz
USF1	upstream stimulatory factor 1
VSMC	vascular smooth muscle cell
VCAM	vascular cell adhesion molecule
VLDL	very low density lipoprotein
WHO	World Health Organization

LIST OF SYMOBOLS AND UNITS

β	beta
$X^{2`}$	Chi square
Δ	delta
°C	degree centigrade
=	equal
g	gravitational force
g	gram
>	greater than
≥	greater than or equal
$IU.L^{-1}$	international units per litre
L	litre
kg	kilogram
kJ	kilojoules
kg/m^2	kilogram per meter squared, unit of body mass index
<	less than
≤	less than or equal to
μ	micro
μL	microlitre
m	milli
mL	milliliter
mmHg	millimeters of mercury
$mmol.L^{-1}$	millimoll per litre
mol	mole
M	molecular weight
x	multiply
-	negative minus
n	number of subjects
p	p-value, indicates statistical significance
%	percentage
±	plus minus
U	unit

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

ACKNOWLEDGEMENTS

I am forever grateful to the following people for making this dissertation possible and for keeping my feet firmly on the ground throughout this study. I would like to take this opportunity to thank each and every one of you.

To my supervisor, **Dr Wayne Towers**, for his guidance and advice during my studies and for equipping me with all the qualities to reach my goals. I would also like to thank him for his support, patience and faith in me. I have learned so much from him and am forever indebted. My co-supervisor, **Dr Karin Conradie**, for her excellent laboratory skills and troubleshooting abilities. You have taught me everything I know regarding a Nutrigenetics laboratory and I am so grateful. Thank you for your kindness every day. My second co-supervisor and dear friend, **Dr Cornelie Nienaber-Rousseau**, for all your support and editing skills. Thank you for always being willing to help me, day or night, and always with a kind heart.

Dr Suria Ellis, for her help with the statistical analyses. **Tertia van Zyl**, for her excellent skills and assistance with the BeadXpress® analysis and always being willing to listen. **Barbara Bradley** for her excellent language and editing skills.

To my two best friends, **Lize Slabbert**, who has been there from the beginning and always motivated me to carry on no matter what. Thank you for the endless late-night support. I could never have done this without your help and support. **Karien Bothma**, for all the coffee dates and your warm heart and spirit. Thank you for always being excited with me and for all your love and support.

To **Riaan Reay,** thank you for always being ready to help me no matter what. You were my soundboard, and this dissertation would not have been possible without your support.

To the best parents in the world, **Charl** and **Hanlie Swanepoel**, for not only making my studies possible, but for giving me the opportunity to follow my dreams and supporting me. I am forever indebted to you, Mom and Dad. I love you more than I can hold in my heart and would like to dedicate this dissertation to you. To my two brothers, **Frikkie** and **Charlie,** for your love and support and always putting a smile on my face when things got serious.

All the **PURE study participants**, and all the staff involved in making this study possible. Without you this research would not have been possible.

Centre of Excellence for Nutrition, for making my studies possible and providing me with the opportunity to become a young scientist. The **National Research Fund**, for funding.

Finally I would like to thank my **Saviour, Jesus Christ,** for granting me this opportunity and for giving me the strength to finish what I started. Also for placing people in my life to help me complete this dissertation and learning life lessons in the process. Thank You for the grace to carry on every day.

CHAPTER ONE

Introduction

Currently, cardiovascular disease (CVD) constitutes, on average, 60% of all deaths due to chronic non-communicable disease (NCD) in the world (Yach *et al*., 2004). In South Africa, an average of 195 people die daily as a result of CVD events (Steyn, 2007). The significance of certain risk factors for CVD development, including among others, dyslipidaemia, hypertension and tobacco use, are well recognised and yet cardiovascular events occur in many individuals who do not present with these traditional risk factors (Greenland *et al*., 2003; Ridker *et al*., 2004). To improve risk prediction, it is important to explore other possible risk factors and to explain the determinants of these risk factors in order to prevent CVD by identifying and treating individuals presenting with non-traditional risk factors more effectively. Gelehrter *et al*. (1998) suggest that many biomarkers (recognised risk factors as well as non-conventional risk factors) that are related to CVD have their own layout of environmental and genetic elements that contribute to CVD aetiology, thus both these elements should be investigated simultaneously.

Markers of inflammation have emerged as possible risk factors of CVD, with several studies reporting that C-reactive protein (CRP) is a strong independent predictor of future CVD events in both men and women (Ridker *et al*., 1997; Ridker *et al*., 2003; Rost *et al*., 2001). Crawford *et al*. (2006) and Lange *et al*. (2006) established that a relationship exists not only between *CRP* genetic variants and CRP concentrations, but also with CVD risk. CVD is a multifactorial disease influenced by both environmental and genetic determinants, as well as the interplay of these variables with each other. Therefore, individuals who are genetically susceptible to increased CRP concentrations may or may not develop CVD, depending on environmental exposure and the possible interplay between these factors (Gelehrter *et al*., 1998).

Median concentrations of CRP vary between 1.5 and 1.7 $mg.L^{-1}$ in healthy American and European populations (Rifai & Ridker, 2003). Results from other ethnic groups suggest that there are differences in CRP concentrations, especially between African *versus* caucasian individuals (Albert *et al*., 2004; Danner *et al*., 2003; Khera *et al*., 2005). In a systematic review conducted by Nazmi and Victora (2007), it was concluded that

individuals of African descent have higher CRP concentrations than individuals of European descent. CRP concentrations were noted to be significantly higher (51.2%) in black women than in white women participating in the Women's Health Study (Albert *et al*., 2004). Similar results were reported in the multi-ethnic Dallas Heart Study, where black participants had higher CRP concentrations than white participants (Khera *et al*., 2005).

Genetics may, to some extent, define the ethnic variations in CRP concentrations. Numerous studies have reported that CRP concentrations are influenced by single nucleotide polymorphisms (SNPs) within the *CRP* gene that predispose an individual to either increased or decreased CRP concentrations (Crawford *et al*., 2006; Lange *et al*., 2006; Wang *et al*., 2006). The most frequent type of genetic difference is the SNP and it has been suggested that it shapes the genetic foundation for numerous complex human diseases. It should, however, also be noted that the interplay of various other components, together with this genetic foundation, ultimately gives way to the formation of the complex disease (Prokunina & Alarcón-Riquelme, 2004). By discovering the genetic variants that may be responsible for the differences in CRP concentrations between different ethnicities, it will be possible to develop a merged multifactorial model to predict the increased CVD risk associated with CRP.

Data regarding elevations in inflammatory markers in relation to CVD risk principally came from caucasian populations. Therefore, little data is available in other ethnic groups, especially for black South Africans. One can hypothesise that there would also be ethnic differences between black South Africans and other ethnicities within Africa, since Africa is one of the most ethnically and genetically diverse regions of the world (Schuster *et al*., 2010). Currently most urban and rural areas of Sub-Saharan Africa have a low frequency of CVD, but with urbanisation an increase in CVD is anticipated (Sliwa *et al*., 2008; Tibazarwa *et al*., 2009). It is important to investigate the possible protective mechanisms that are at play, which may be protecting this population. This highlights the need to study traditional and non-traditional CVD risk factors (such as CRP) as well as their genetic determinants in South African populations. This information can also be used to curb the possible rise in CVD risk in Sub-Saharan Africa.

The fact that Africa consists of genetically diverse populations is often ignored when strategies are being developed for the understanding and prediction of the NCD risk of a population, as well as when developing treatment modalities. Extrapolating knowledge of disease phenotypes associated with single gene variations, which are derived from studies

performed on non-African populations, should therefore be avoided in African populations (Sing *et al*., 2003) because implementing regimes developed and tested on other population groups in African populations will have unknown and unfavourable outcomes.

It is hypothesised that modern-day human beings originated in Africa about 200,000 years ago and then spread across the world (Campbell & Tishkoff, 2008). The theory is that modern humans have been residing in Africa permanently for a longer period of time than in any other anthropological area where they have held a large population size. This has resulted in high levels of within-population genetic variation (Campbell & Tishkoff, 2008; Reed & Tishkoff, 2006). Another explanation for the increased diversity within African populations could be that new deoxyribonucleic acid (DNA) variations arose and owing to the effects of genetic drift, selection and migration, these have altered the distribution of certain genetic mutations. This resulted in populations harbouring different combinations of DNA variants, which implies that there is a difference in the spectrum of alleles and genotypes displayed, for any specific susceptibility locus (Sing *et al*., 2003).

As mentioned, *CRP* polymorphisms are associated with altered CRP concentrations as well as with CVD risk, and their distributions are different in different ethnic groups (Hage & Szalai, 2007). The distribution of these *CRP* gene polymorphisms indicate prominent ethnic-specific effects (Ranjit *et al*., 2007, Albert *et al*., 2004). Investigating CRP in the black South African population is therefore relevant, valuable and necessary.

1.1 AIMS AND OBJECTIVES OF THIS STUDY

The primary aim of this dissertation was to determine the association between various polymorphisms (reported and novel SNPs) in the *CRP* gene and CRP concentrations [measured as high-sensitivity (hs)-CRP concentrations] in a black South African population undergoing an epidemiological transition. The secondary aim was to investigate the interaction effects between the different genotypes for these SNPs and demographic (e.g. gender) or environmental factors [e.g. area of residence (rural/urban)] on the CRP concentrations within this population.

The objectives are as follows:

a) To establish the genotype distribution of specific polymorphisms under investigation in the *CRP* gene in a black South African population;

b) To determine whether the various identified polymorphisms in the *CRP* gene are associated with CRP concentrations;

c) To investigate whether demographic (e.g. gender) or environmental factors [e.g location (urban/rural)] have a superimposed effect on the influence of the various *CRP* polymorphisms on CRP concentrations.

Very little, if any, research has been conducted to characterise the SNP/phenotype outcomes of the various *CRP* genetic variants or the frequencies thereof in a black South African population residing either in rural or urban areas. Data regarding CRP concentrations in different ethnic groups is lacking at present and this study will provide novel and valuable information pertaining to the determinants of CRP concentrations in a black South African population. In addition, this study will also evaluate whether urbanisation has an effect on these genotype-phenotype associations, which will be a further original contribution to the existing literature.

1.2 STRUCTURE OF THIS DISSERTATION

Directives in terms of language usage, formatting and quotation of sources of the North-West University were strictly followed in the writing of this chapter style dissertation. Chapter 1 provides a general introduction to the research problem addressed in this dissertation, presents an overview of the format and content of the dissertation and lists the outputs that have resulted from this work. Chapter 2 consists of a detailed review of the literature on CRP, to convey an integrated view of all the possible determinants of CRP concentrations (including pathogenic, biochemical and genetic factors) in order to facilitate the understanding and interpretation of the results that will be presented in the ensuing chapters of this dissertation. Chapter 3 encapsulates the methodologies used, *i.e.* the manner in which blood samples were collected, informed consent, assessment of nutrient intake and the statistical analyses performed to obtain the results necessary to answer the research question.

In Chapter 4 the study results are presented and discussed. Each of the investigated SNPs is discussed separately, followed by a summary of the results. Chapter 5 provides a recapitulation of the results, followed by the conclusions and recommendations on the findings of the research that was conducted. This chapter will complete the dissertation.

1.3 LIST OF RESEARCH OUTPUTS EMANATING FROM THIS STUDY TO DATE

"The effect of the A790T polymorphism in the C-reactive protein gene on cardiovascular disease risk in a black South African population". Joint Congress of the Southern African Society of Human Genetics and the African Society for Human Genetics (2011) in Cape Town, South Africa. (Poster presentation)

"*Population-specific association of certain CRP genetic variants with hs-CRP concentration in black South Africans*". Nutritional Congress Africa (2012) in Bloemfontein, South Africa. (Oral presentation)

CHAPTER TWO

Literature overview

At the start of the third millennium, NCDs appear to be sweeping the entire globe, with an increasing incidence occurring in developing countries (World Health Organisation [WHO], 2002). NCDs include, among others, CVD, type 2 diabetes mellitus and metabolic syndrome, and are commonly referred to as chronic diseases (Reddy & Yusuf, 1998). In the past, CVD was a disease that mainly occurred in the so-called first world countries. However, the global burden of CVD is now considered to be a problem not only in affluent countries, but also a major problem of developing countries (Gaziano, 2005; Boutayeb & Boutayeb, 2005).

The development of CVD is complicated and occurs in response to various aetiological pathways involving numerous risk factors. In order for CVDs to be effectively treated and prevented, these risk factors or predictors of disease risk need to be identified. CRP is considered to be one of the major predictors of CVD risk (Ridker *et al*., 1997; Ridker *et al*., 2002; Ridker *et al*., 2003). In the study by Ridker and co-workers (2002) it was reported that CRP was a stronger predictor of future CVD events than low-density lipoprotein cholesterol (LDL-C) concentrations. The predictive property of CRP is related to CRP being a marker of systemic inflammation and, therefore, most probably plays a role in the atherosclerotic process (Libby, 2006), which could in turn lead to a CVD event. Thus, the regulation of CRP concentrations in the body is important in understanding the pathological role of this protein. It has been determined that certain *CRP* gene polymorphisms can result in an individual having either high or low CRP concentrations, which in turn may have an impact on the individual's CVD risk (Hage & Szalai, 2007). This fact highlights the importance of genetic variability in CVD susceptibility.

However, one has to be cautious in inferring that *CRP* genetic variants are causally related with CVD. This concept has been questioned by Mendelian randomisation studies and this are discussed in Section 2.4.2.2. This overview of the literature will give a broad summary of the aspects that must be considered in order to answer the research question of the investigation, which is to determine the association between various *CRP*

polymorphisms with CRP concentrations in a black South African population undergoing an epidemiological transition.

2.1 CARDIOVASCULAR DISEASE AS A GLOBAL BURDEN AND AS A BURDEN IN AFRICA

According to the WHO, CVD can be defined as a group of disorders of the heart and blood vessels. Some of the most familiar disorders under the banner of this term include coronary heart disease (CHD), also referred to as coronary artery disease, as well as cerebrovascular disease. In most countries, CVD has become a widespread cause of morbidity and one of the leading causes of mortality (Murray & Lopez, 1997; Yusuf *et al*., 2001; Gersh *et al*., 2010). In Africa alone it is estimated that CVD will affect 1.3 million people per annum in the next 20 years (Murray *et al*., 1996). It is now predicted that by the year 2020, 40% of all deaths worldwide will be caused by CVD as a direct result of an unhealthy lifestyle pattern, caused by industrialisation as well as urbanisation in specific populations that are experiencing demographic and socio-economic transition (Lenfant, 2001; Willerson & Ridker, 2004). Therefore, developing countries carry a great share of the global burden of CVD (Reddy & Yusuf, 1998; Gersh *et al*., 2010) and CVD can no longer be classified as a problem of only affluent countries (Gaziano, 2005; Boutayeb & Boutayeb, 2005; Deaton *et al*., 2011). Urbanisation, together with the high prevalence of certain risk factors such as obesity, diabetes, dyslipidaemia and hypertension, can be the cause of the increasing occurrence of atherosclerotic diseases that is observed in developing countries (Yusuf *et al*., 2001). Focussing on South Africa, it is in the middle of a health transition that is characterised by the coinciding prevalence of infectious diseases together with the rise in NCDs. The burden of disease which is caused by NCDs in South Africa is predicted to increase over the next decade if measures are not taken to understand and combat this trend (Mayosi *et al*., 2009).

2.2 ORIGINS OF CARDIOVASCULAR DISEASE

It is evident that all cases of CVD have a complex multifactorial aetiology and neither genetic nor environmental agents acting independently are responsible for CVD (Sing *et al*., 2003). Some of the major non-genetic factors responsible for CVD development are obesity, diabetes, dyslipidaemia, hypertension (Yusuf *et al*., 2001) and smoking (Greenland *et al*., 2003). However, more than half of all CVD events occur in individuals without obvious hyperlipidaemia or any of the above-mentioned risk factors (Ridker *et al*., 2004), which suggests that other variables with their own set of environmental and genetic

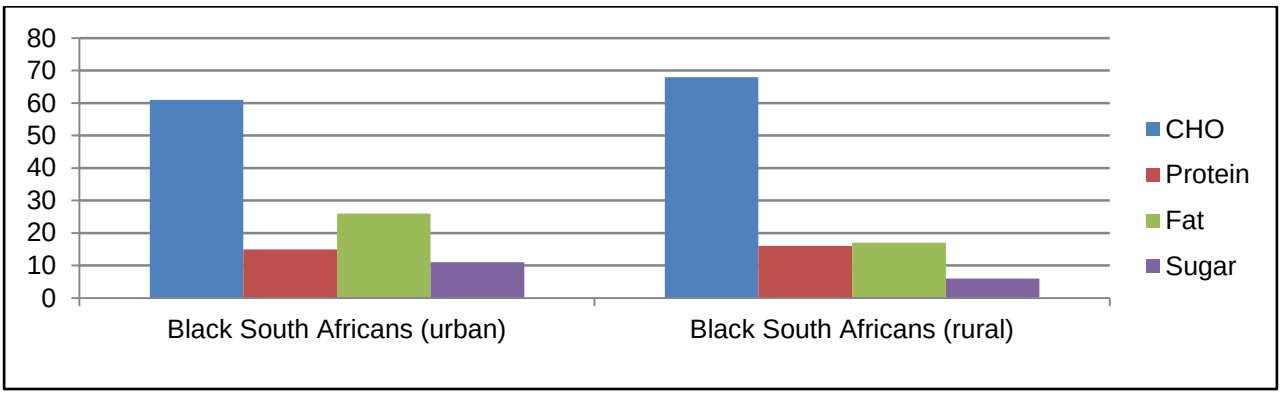
determinants could be involved. The next sections discuss the possible environmental, foetal and genetic origins of CVD.

2.2.1 The dietary transition and its role in CVD development specifically in the black South African population

Dietary patterns around the world are changing rapidly; high-fibre foods are being substituted for processed foods and the diet as a whole is becoming more energy dense (Popkin, 2006). In South Africa, the black population outnumbers the other population groups in the country, as it represents 79.5% of the population. However, it is also the most impoverished group (Statistics South Africa, 2011). The majority of the black population in South Africa reside in urban areas, *i.e*. 60.7% (Statistics South Africa, 2011). When comparing the number of urban and rural individuals in South Africa, approximately 60.7% of South Africans reside in an urban setting compared to the 39.3% who reside in a rural setting, indicating increased levels of urbanisation in South Africa.

The urban and rural populations have different eating patterns (Figure 2.1). The rural population still follows a conventional diet, of which the macronutrient distribution is high in carbohydrates (>65% of total energy [TE]) and low in fat (<25% of TE). Overall sugar intake is also lower (<10% of TE), while fibre intake is moderately higher (Steyn *et al*., 2001) than in urban populations. The diet of the urban population, on the other hand, reveals the adoption of the Westernised dietary pattern, which includes lower carbohydrate (<65% of TE) and fibre intakes, with the fat intake being higher than 25% of the TE (Bourne *et al*., 1993).

Figure 2.1 Macronutrient distribution as a percentage of total energy consumed per day by black South African adult men

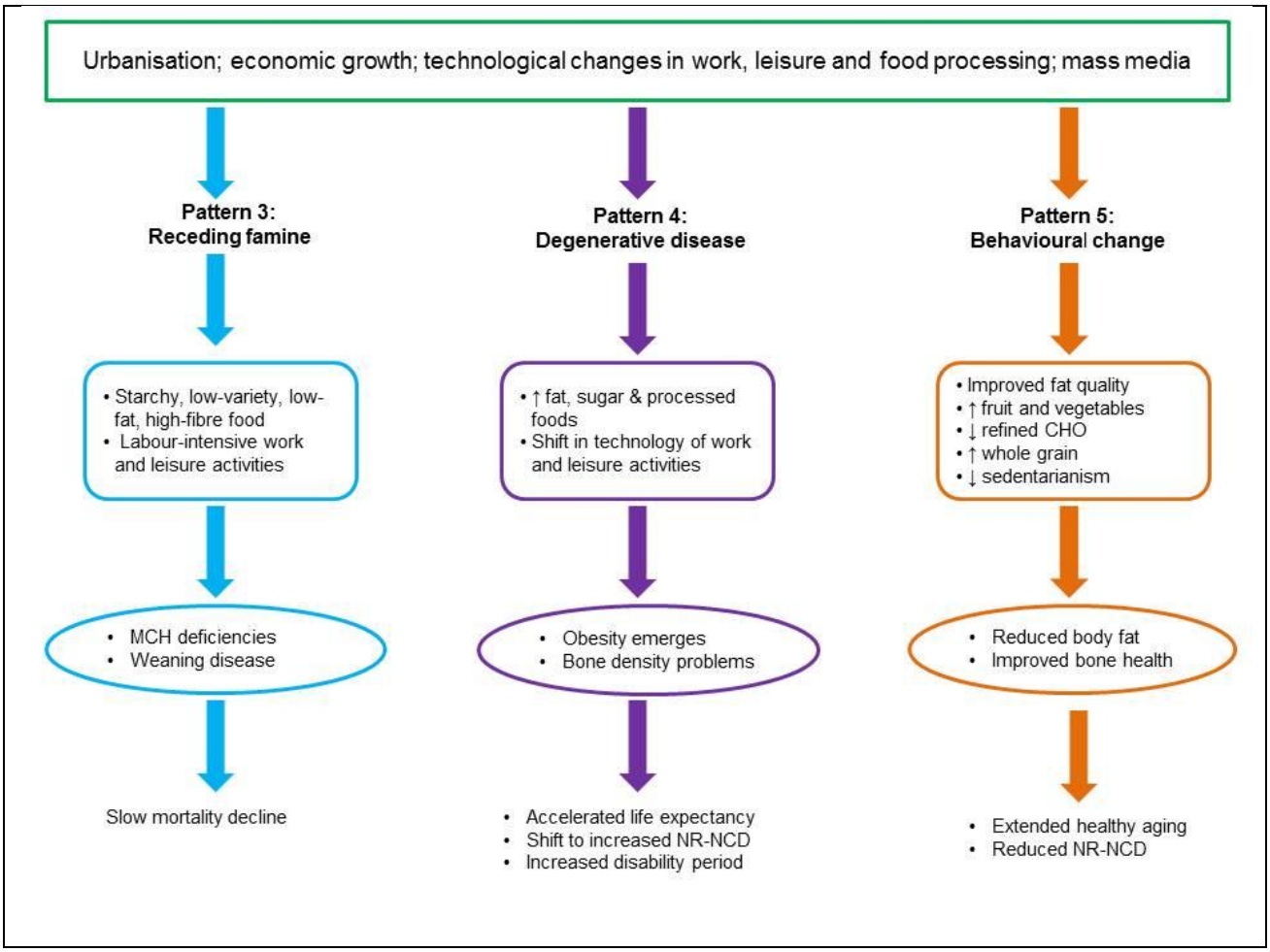


CHO = carbohydrates. Adapted from Steyn *et al*. (2006a)

The typical human diet, as well as its nutritional status, has been altered over the past few years. This includes changes in food use and ultimately leads to increased risk of nutrition-related diseases (Popkin, 2006). This stereotypical pattern is referred to as the nutritional transition. These changes occur as populations are undergoing demographic and socioeconomic changes, which will then bring about changes in body composition, dietary and activity patterns (Popkin, 2006). One of the assumptions of the nutritional transition is supported by rural and urban comparisons of African populations, where it can be seen that the traditional diet is replaced during urbanisation with a more Westernised diet (fewer carbohydrates and lower fibre intake and an increase in fat intake). The traditional diet is related to a low prevalence of NCDs, whereas the Westernised diet is associated with an increased prevalence of NCDs (Vorster *et al*., 1999).

The shift towards increased obesity and NCDs seen in the African populations is only the latest pattern of this transition. The five patterns of the nutrition transition are stipulated by Popkin (1993) as follows: The first pattern was associated with hunter-gatherer societies and was a pattern in which the diet was very healthy, but infectious diseases and other natural causes led to a very short life span. When modern agriculture and a period of famine emerged, the second pattern became evident. In the second pattern the nutritional status worsened in comparison with the first pattern because of the emergent famine, which would suggest that nutrition was not widely available. Considering the global population of today, attention is focused more intently on the last three patterns, namely receding famine, degenerative disease and behavioural change during the nutritional transition (represented in Figure 2.2). The financial income of the population increases in pattern three and as a result of this, famine begins to recede. Pattern four gives way to changes in diet as well as activity patterns, which then leads to the appearance of new disorders such as nutrition-related NCDs. In pattern five behavioural changes take place, which consequently reverse the negative occurrences taking place in the previous patterns. This pattern allows a population to follow a process of successful aging and is ultimately where the population wishes to be (Manton & Soldo, 1985; Crimmins *et al*., 1989). A series of factors, *i.e*. urbanisation, economic growth, food processing *etc*., drives all the changes observed in these patterns. Although this process is relatively consistent, the occurrence of specific NCDs differs in various areas of the world. Reddy and Yusuf (1998) point out that part of this variance might be associated with the different stages of the epidemiological transition, but that genetic-environmental interactions probably contribute most to this variability.

Figure 2.2 Patterns of the nutritional transition



CHO = carbohydrates; MCH = maternal and child health; NR-NCD = nutrition related non-communicable disease. Adapted from Popkin (2006)

In summary, two historical processes of transformation take place at the same time during the nutritional transition. The first process is the demographic transition and entails a modification from a pattern with a high death rate and short life expectancy, to a pattern of lower death rates and a longer life expectancy, which can be ascribed to a decrease in communicable disease. The second process is the epidemiological transition, which is the shift from a pattern of high prevalence of infectious disease (disease caused by malnutrition, famine and poor environmental hygiene) to one with a high prevalence of chronic and degenerative disease, associated with urbanised and industrial lifestyles (Popkin, 2003).

2.2.2 Foetal origins of cardiovascular disease

The risk of having increased susceptibility to developing a complex disease, such as CVD or type 2 diabetes, or the potential to achieve the most favourable health is decided at conception by the grouping of genetic variants obtained from each parent. Various

environmental exposures throughout life can influence the manner in which these genetic variants are expressed (Mathers & McKay, 2009). The theory that foetal and infant nutrition have an effect on adult response to environmental change (such as diet), which will then ultimately influence CVD risk, arose because one cannot completely clarify the frequency, prevalence, geographic differences, trends and individual variations of CVD by simply evaluating lifestyle and genetic factors alone (Barker, 1996; Barker, 2001). The explanation for the supposed foetal origins of adult disease was proposed by Hales and Barker (1992) and was called the "thrifty phenotype" hypothesis. This theory states that an adverse intra-uterine environment, due to poor foetal nutrition, imposes mechanisms that programme foetal metabolism to cope with future nutritional thrift. If this state of nutritional hardship should continue, the physiological adaptation would be suitable. The opposite, however, is also true. Should the individual subsequently be exposed to over-abundant nutrition, there would be a state of physiological maladaptation (mismatch), which would eventually give way to the onset of chronic adult disease, such as CVD (Gluckman *et al*., 2005).

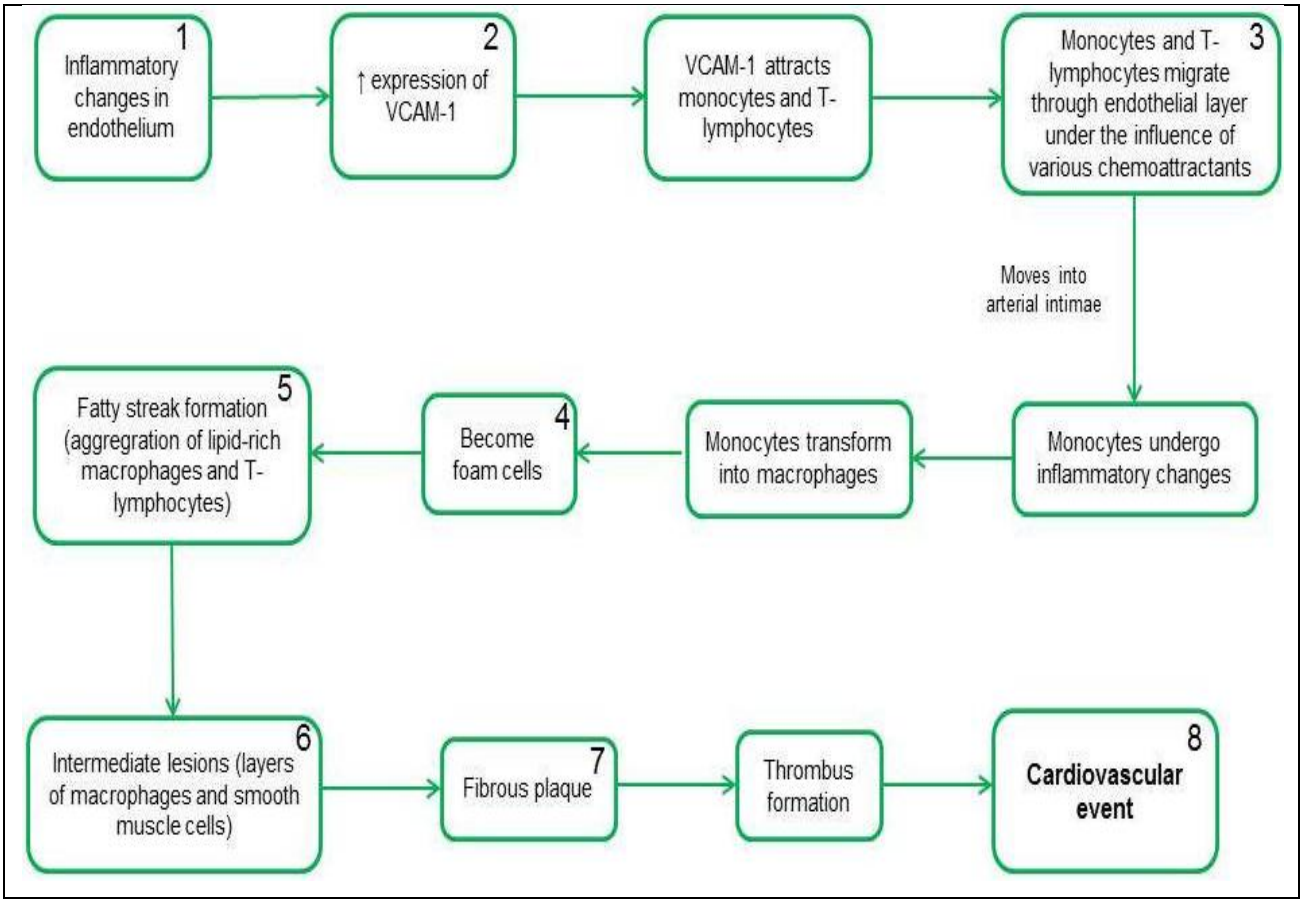
Singhal and Lucas (2004) also reported that the main factors of metabolic syndrome, as well as the risk of developing CVD in the future, can be permanently influenced by foetal nutrition. They also indicated that in the past 40 years CVD risk has been affected by early nutrition as well as growth (Singhal & Lucas, 2004). As the importance of the foetal development of CVD risk is realised, it is also being discovered that various environmental factors play an equally important role. If intrauterine malnutrition is considered an additional risk factor for developing CVD in adulthood, the countries with modifications in dietary and activity patterns (developing populations) are at greater risk than developed countries (Vorster, 1999).

2.2.3 Molecular origins of cardiovascular disease

Atherosclerosis is one of the major underlying causes of CVD. Thus, the commencement and development of the mechanism responsible for atherosclerosis needs to be understood and is critical for prevention as well as treatment of CVD. It is currently well known that inflammation plays a key role (Libby, 2006) in this process. Figure 2.3 illustrates an overview of the molecular origins of CVD, which are discussed in the subsequent section.

The origins of atherosclerosis involve the initiation of endothelial dysfunction by atherogenic triggers, of which the best recognised are modified or oxidised LDL-C. Therefore, one can say that initially inflammatory changes occur in the endothelium (step one). This functional modification causes an increased expression of atherogenic signal molecules, which include vascular cell adhesion molecule 1 (VCAM-1), chemoattractants (monocyte chemoattractant protein 1), as well as a host of growth factors and cytokines, such as CD40 ligand (CD154), macrophage colony stimulating factor, tumour necrosis factor–α (TNF-α), interleukin 1 (IL-1), and interleukin 6 (IL-6), which are indicated in step two (Libby *et al*., 2002; De Caterina *et al*., 2004). All the aforementioned signalling molecules allow monocytes and T lymphocytes to bond to the arterial endothelium. These then migrate through the endothelial layer under the influence of various chemoattractants (step three).

Figure 2.3 Breakdown of the molecular origins of cardiovascular disease



VCAM-1 = vascular cell adhesion molecule 1; adapted from Libby, (2006)

Within the intimae, monocytes mature into macrophages and ultimately form lipid-rich foam cells (step four), which accumulate cholesterol esters in the cytoplasm (Libby, 2006). These foam cells are a trait of the first morphologically recognisable precursor of atherosclerotic plaque, also called the fatty streak (step five). T-lymphocytes, which are

now in the intimae, as mentioned in step three, release proinflammatory cytokines that amplify the inflammatory activity through which the intermediate lesions from the fatty streak are formed (step six). The same aforementioned signalling molecules are responsible for the growth and ultimately the destabilisation of the plaque (step seven). This process will ultimately result in a cardiovascular event (step eight). The role of CRP in this process is discussed in Section 2.3.3.

2.3 C-REACTIVE PROTEIN

CRP was discovered in 1930 by William Tillet and Thomas Francis while they were studying patients with acute *Streptococcus pneumonia* infections (Tillett & Francis, 1930) and it was in turn studied thereafter by Abernethy and Avery (1941). CRP was at first defined as a substance seen in the plasma of patients with acute infections that reacted with the C polysaccharide of the *Pneumococcus* bacteria. The ligand-binding activity of CRP is calcium-dependent and it binds with highest affinity to phosphocholine, which is an integral part of cell membrane phospholipids, such as phosphatidylcholine. Under normal conditions phosphocholine is not exposed; however, once a cell has been damaged it becomes accessible to CRP (Du-Clos, 2000; Volanakis, 2001). This exposure then results in the binding of CRP to phosphocholine to activate the complement system. The binding of CRP to phosphocholine also enhances phagocytosis by macrophages. Kilpatrick and Volanakis (1991) conducted *in vitro* as well as *in vivo* experiments and suggested that the ability of CRP to identify unknown pathogens as well as damaged cells and then to initiate the elimination of these pathogens is the main function of CRP. CRP can eliminate the pathogens by interacting with the humoral and cellular effector systems in the blood. The assumption can, therefore, be made that CRP has both recognition and effector functions. CRP also binds with a high affinity to chromatin. One of the major physiological functions of CRP is to act as a scavenger for chromatin released by dead cells during acute inflammatory processes (Robey *et al*., 1984).

CRP falls into two major protein classes. CRP concentrations increase in the plasma during inflammation, therefore the first group that CRP belongs to is the acute phase proteins. Secondly, CRP can be classified as a pentraxin protein because of its structure and its calcium-dependent binding specificities (Klipatrick & Volanakis, 1991). CRP consists of five non-covalently associated protomers arranged symmetrically around a central core, and has a molecular weight of 118 kDa (Thompson *et al*., 1999). With regard to the secretion of CRP in the liver, it is commonly agreed that its production is primarily

under the control of IL-6 and to a lesser extent of IL-1 and TNF, which are all inflammatory cytokines (Blake & Ridker, 2002). The half life of CRP is estimated at 19 hours and appears to be stable in healthy and disease states (Jialal *et al*., 2004).

2.3.1 C-reactive protein and cardiovascular disease risk

CRP is a powerful predictor of CVD risk at all levels of the Framingham risk score, as well as metabolic syndrome (Bisoendial *et al*., 2010; Ridker *et al*., 1997; Ridker *et al*., 2003). CRP is also believed to amplify the anti-inflammatory response through complement activation, tissue damage and activation of endothelial cells (Libby *et al*., 2002). Kushner *et al*. (1963) demonstrated deposition of rabbit CRP in experimental myocardial infarction (MI) or stroke lesions in one of the first applications of immunofluorescence.

A long history of association between CRP and CVD events exists. In 1963, Kushner *et al.* reported the kinetics of the acute phase CRP response in relation to human acute MI and thereafter the behaviour of CRP in clinical CHD and MI was investigated (de Beer *et al*., 1982). The present phase of interest regarding CRP and CVD events began in the 1990's when observations were made that CRP concentrations were elevated in patients with acute MI tested shortly after the start of pain, before the acute phase reaction to the infarction could have started (Eklund, 2007). In this time, a large prospective European study revealed that baseline CRP concentrations significantly predicted future coronary events in patients with stable and unstable angina (Thompson *et al*., 1995). Ridker and Cook (2004) also reported that CRP concentrations predict future CVD events across the full spectrum of disease seen in clinical practice. A meta-analysis, which involved over 7,000 patients with coronary events, reported that patients with CRP concentrations in the upper tertile have a 50% increased risk of developing acute CVD events opposed to the patients with lower CRP concentrations (Danesh *et al*., 2004). A meta-analysis, which included 54 prospective cohort studies, was also published and confirmed CRP as an independent risk marker for CVD (Emerging risk factors collaboration, 2010). In conclusion, it has now been recognised that elevated CRP concentrations can be used to predict CVD events such as MI in patients with stable or unstable CHD (Thompson *et al*., 1995).

2.3.2 Possible confounding and risk factors of CVD and their associations with C-reactive protein

Major contributions to the prevention of CVD have been made in the past few years by the identification of modifiable risk factors. In the past, various studies have reported that healthy changes in lifestyle factors result in beneficial effects, lowering the CVD risk factor burden on CVD outcomes and longevity (Lloyd-Jones *et al*., 2010). According to Grundy and co-workers (1999), major independent modifiable risk factors include cigarette smoking of any amount, hypertension, elevated total cholesterol (TC) as well as LDL-C, low serum HDL-C and type 2 diabetes mellitus. The Framingham Heart Study has explained the analytical association between these risk factors and CVD risk (D'Agostino *et al*., 2008). As this research field expands, novel CVD risk factors are being reported. This section evaluates the possible roles that CVD may play in the regulation of these different factors.

Although the most common CVD risk factors, such as elevated cholesterol concentrations, contain atherogenic properties, inflammatory processes also contribute to CVD risk development independently. CRP concentrations have recently been reported to contribute to the development of CVD owing to their inflammatory properties, resulting in an increased risk of acute thrombotic events *i.e*. MI and stroke (Ridker *et al*., 1998). CRP is currently suggested to be the most robust inflammatory marker of CVD risk (Pearson *et al.*, 2003). Furthermore, it has become apparent that CRP is centrally involved in the pathogenic mechanism of numerous traditional risk factors such as hypertension, age and gender, as well as non-traditional risk factors such as IL-6 concentrations and human immunodeficiency virus (HIV) infection.

A widespread condition and major risk factor for stroke and heart attacks in South Africa is hypertension, or high blood pressure (Steyn *et al*., 2006a). Hypertension was identified as the most frequent of all the CVD risk factors in a cross-sectional CVD risk factor survey, which was conducted on a black South African population (Connor *et al*., 2008). Several investigators have observed higher CRP concentrations in individuals with hypertension (Blake *et al*., 2003; Schillaci *et al*., 2003). In a study conducted by Lakoski *et al*. (2005) systolic blood pressure was associated with CRP concentrations. Lakoski and co-workers (2005) also reported that ethnic differences in the relationship between CRP and hypertension were evident in their study.

IL-6 is a plasma cytokine that participates in mediating inflammation and is a central stimulus for the acute-phase response (Papanicolaou *et al*., 1998). Circulating concentrations of IL-6 and CRP are physiologically linked, but it remains unclear whether these markers of systemic inflammation indicate a relationship with one another with regard to various CVD risk factors in healthy individuals (Bermudez *et al*., 2002).

Individuals infected with HIV demonstrate increased rates of CHD compared to uninfected individuals (Triant *et al*., 2007). This can be attributable to the fact that traditional CVD risk factors are often similar among individuals with or without HIV (Triant *et al*., 2007); however, the role of inflammation in the infection process and the value of related biomarkers have not been studied, but seem to be of importance (Triant *et al*., 2009). Noursadeghi and Miller (2005) reported that CRP concentrations were elevated in individuals with HIV compared to the general population. In a study conducted by Traint and co-workers (2009), their results suggested that increased both CRP and HIV infection are independently associated with acute MI and that individuals with HIV and an additional elevated CRP concentration are at a fourfold increased risk of developing an acute inflammatory state.

Two risk factors in addition to the genetic make-up, which cannot be modified, but should be taken into account when investigating CRP concentrations, are age and gender. Age is one of the most influential risk factors for CVD, as the risk of stroke doubles each decade after the age of 55 (Mackay *et al*., 2004). Lower CRP concentrations were reported for younger individuals in a study conducted by Holmes *et al*. (2009). Deary *et al*. (2009) stated that a high variability in acute inflammatory biomarker levels (such as CRP concentrations) were present in older individuals. Therefore, it has been determined that CRP concentrations increases with age. Gender also plays an important role in CVD development. The risk of stroke is similar between both genders (Kelly-Hayes *et al*., 2003), but according to Mackay and co-workers (2004), premenopausal women experience lower CHD rates than men of the same age. It is important to note that the risk increases significantly for women after the protective effect of oestrogen is lost (Bothig, 1989). Various studies have determined CRP concentrations to be higher in women than in men (Khera *et al*., 2005).

2.3.3 Mechanistic involvement of CRP in the vascular disease process

If the role of CRP in the vascular disease process could be explained, a broader understanding would emerge with regard to its cause-and-effect association with CVD, even though sufficient evidence from well-designed *in vitro* as well as *in vivo* studies (Szalai *et al*., 1995; Szalai & McCrory, 2002; Paul *et al*., 2004) is available. CRP is responsible for the activation of monocytes, endothelial cells and vascular smooth muscle cells (VSMC), which then causes a decrease in nitric oxide (NO) production and an increase in angiotensin II signalling, plasminogen activator inhibitor-1 (PAI-1) concentrations as well as endothelin-1 (ET-1) activity. These molecules are all indicators of the pro-atherogenic, pro-thrombotic, pro-inflammatory and pro-endothelial functions of CRP.

The proof of CRP's role in the pathogenesis of vascular disease was illustrated in CRP transgenic mice, which express human CRP in a way that closely imitates the pattern in humans. In these mice, CRP promotes arterial thrombosis (which has been verified by different independent laboratories [Danenberg *et al*., 2003]), accelerates atherosclerosis (Paul *et al*., 2004), induces endothelial dysfunction (Teoh *et al*., 2008) and exacerbates the vascular injury response (Xing *et al*., 2008). All the mechanisms of CRP mentioned are summarised in Figure 2.4.

2.3.3.1 Endothelial cells and CRP

CRP concentrations have been determined to correlate inversely with endothelial vasoreactivity (Cleland *et al*., 2000; Fichtlscherer *et al*., 2000). Venugopal and co-workers (2002) established that a significant reduction in the messenger ribonucleic acid (mRNA) and protein production of endothelial nitric oxide synthase (eNOS) in human aortic endothelial cells was brought about by CRP. In addition, they determined that eNOS activity, as well as eNOS bioactivity, was decreased in these aortic endothelial cells. The effect of CRP seems to be at the level of decreasing the stability of eNOS mRNA (Verma *et al*., 2002). CRP blocks NO-dependent processes through the reduction of eNOS expression as well as NO release, which then facilitates endothelial cell apoptosis, revealing a pro-atherogenic and pro-inflammatory phenotype (Jialal *et al*., 2004).

Another noteworthy product of endothelial cells is prostacyclin, which is a powerful vasodilator, an inhibitor of smooth muscle cell proliferation as well as platelet aggregation.

Venugopal *et al*. (2003) reported that the addition of CRP, even at low doses, brought about decreases in the release of the stable metabolite of prostacyclin in both human aortic endothelial cells and human coronary artery endothelial cells. They also determined that CRP stimulated superoxide anion release and because it also inhibits eNOS, this suggests that CRP could decrease prostacyclin synthase activity.

PAI-1 is regarded as part of the serine protease inhibitors and appears to be synthesised in the liver, adipose tissue, endothelial cells, VSMC and macrophages (Jialal *et al*., 2004). PAI-1 is considered a marker of impaired fibrinolysis as well as atherothrombosis and is elevated in CVD patients. Devaraj and co-workers (2003) have indicated that CRP induces PAI-1 mRNA production, antigen levels and enzymatic activity in human aortic endothelial cells. Therefore, CRP may imitate the role of a procoagulant, because of the effects reported *in vitro*, *i.e.* increased PAI-1 and tissue factor, as well as reduction in eNOS and prostacyclin (Jialal *et al*., 2004).

Venugopal *et al*. (2002) observed that in aortic endothelial cells, CRP increases intercellular adhesion molecule (ICAM) as well as VCAM expression and in turn the adhesion of monocytes to the endothelium. Furthermore, it has been reported that CRP triggers the increased release of endothelial-derived contracting factor, ET-1 (Verma *et al*., 2002). ET-1 is a powerful vasoconstrictor, and has also emerged as an intermediary of CRP-induced upregulation of adhesion molecules and monocyte chemoattractant protein-1 in venous endothelial cells (Jialal *et al*., 2004). The chemokine interleukin-8 (IL-8) is considered to be one of the triggers of the adhesion of monocytes to endothelium (Jialal *et al*., 2004). Devaraj and co-workers (2003) reported that CRP brings about an increase in the expression of IL-8 in human aortic endothelial cells, as well as human coronary artery endothelial cells. In summary, it is clear that CRP is involved in the endothelial cells and plays a role in the biochemical factors (eNOS, PAI-1, ICAM, VCAM, ET-1 and IL-8) associated with these endothelial cells.

2.3.3.2 CRP and monocyte-macrophages

Foam cells, which are commonly seen in atherosclerotic lesions, are mainly macrophages loaded with cholesterol ester-rich cytoplasmic lipid inclusions (Kruth, 2001; Faruqi & DiCorleto, 1993). In response to inadequate signalling of certain factors (Figure 2.4), circulating monocytes will attach to the endothelial surface of muscle arteries and in response to chemotactic factors in the arterial intimae and media layers, enter the

sub-endothelial area and differentiate into macrophages (Faruqi & DiCorleto, 1993). The cholesterol-rich lipid region of the foam cells is primarily from LDL-C, which will move into the intimae as well as the media layers of the arteries, where it is kept in the extracellular matrix and finally taken up by monocyte-derived macrophages (Williams & Tabas, 1998).

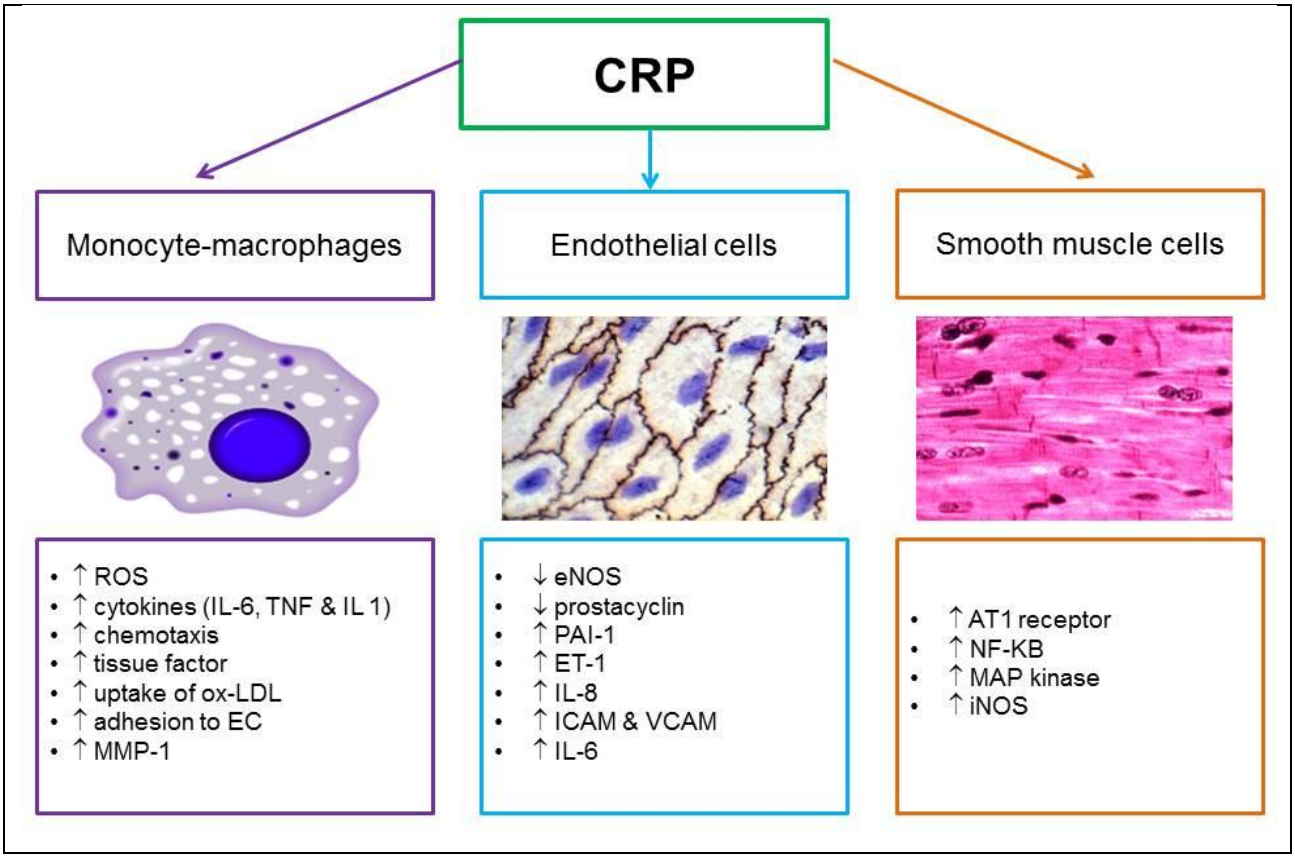
CRP has been determined to be present in the intimae and the media layers of arteries (Torzewski *et al*., 1998), as well as on the surface of foam cells (Li *et al*., 2004). In a study conducted by Yasojima and co-workers (2001), they reported that CRP can be synthesised by macrophages that occur in human atherosclerotic plaques. The aggregation of CRP molecules arises *via* a calcium-dependent process, where each of the five subunits of the CRP molecule can bind to two molecules of calcium. The aforementioned process is reported to bring about conformational changes in the CRP molecule, which is believed to be responsible for the aggregation of CRP (Pepys *et al*., 1985). It is, therefore, plausible that in the presence of elevated calcium concentrations in the arterial wall, CRP derived from filtration of plasma CRP or synthesised *in situ* by macrophages may form aggregates. These aggregated CRP molecules will possibly bind and cluster LDL-C particles, allowing them to be phagocytosed by sub-endothelial macrophages, leading to the formation of foam cells. The specific mechanism of macrophage uptake of LDL-C bound to immobilised CRP aggregates remains to be clarified (Fu & Borensztajn, 2002). Jialal *et al*. (2004) reported CRP as being proartherogenic in monocyte-macrophages because of its ability to increase tissue factor expression, promote monocyte chemotaxis and, thus, adhesion to endothelial cells, as well as its ability to increase reactive oxygen species (ROS) release and to support oxidised LDL-C uptake, thus also leading to foam cell formation.

2.3.3.3 CRP and smooth muscle cells

One of the main atherosclerotic changes that assist angiotensin II (Ang II)-induced ROS production, VSMC migration, proliferation, and remodelling is the angiotensin type-1 receptor (AT1R). The essential significance of AT1R in the development of atherosclerosis is apparent (Nickenig & Harrison, 2002), and led to the demonstration by Verma *et al*. (2002) that the higher number of AT1R-binding sites in VSMC, which influences the risk of atherogenesis, could be explained by the activity of CRP, which upregulates AT1R mRNA and protein activity in VSMC (Wang *et al*., 2003). The functional association between CRP and Ang II in mediating vascular smooth muscle cellular pathology may be due to the fact that CRP also increases Ang II-induced VSMC migration

and proliferation. An *in vivo* model of carotid balloon angioplasty demonstrated that AT1R expression was induced by CRP exposure. This expression resulted in increases in neointimal formation, VSMC migration and proliferation, and promoted collagen and elastin production. These last-mentioned proteins (collagen and elastin) form an essential part of the vessel wall. Therefore, CRP is directly involved in the proatherosclerotic effects at the level of VSMC. Inducible nitric oxide synthase (iNOS) and certain cell signal transduction pathways, including mitogen-activated protein (MAP) kinase pathway and nuclear factor-kappa β (NF-κβ), are upregulated by CRP within the VSMC (Hattori *et al*., 2003).

Figure 2.4 Potential atherothrombotic effects of CRP on vascular cells



AT1 = angiotensin type-1; CRP = C-reactive protein; EC = endothelial cells; eNOS = endothelial nitric oxide synthase; ET-1 = endothelin-1; ICAM = intercellular adhesion molecule; IL-1 = interleukin 1; IL-6 = interleukin 6; IL-8 = interleukin 8; iNOS = inducible nitric oxide synthase; MAP = Mitogen-activated protein; MMP-1 = matrix metalloproteinase-1; NF-Kb = Nuclear Factor-Kappa β; ox-LDL = oxidised low density lipoprotein; PAI-1 = plasminogen activator inhibitor type 1; ROS = reactive oxygen species; TNF = tumour necrosis factor; VCAM = vascular cell adhesion molecule. Adapted from Jialal et al. (2004)

2.3.4 C-reactive protein and the diet

The links between diet, inflammatory processes and NCDs are currently topics of intense research. With regard to the effect that the macronutrient composition of the diet has on CRP concentrations, limited data is available (Clifton, 2003). Dietary glycaemic index (GI) is the tendency of carbohydrates in the diet to increase blood glucose levels, whereas dietary glycaemic load (GL) incorporates dietary GI and the amount of carbohydrates

consumed (Salmeron *et al*., 1997). Dietary GI as well as dietary GL have both presented with inconsistent relationships to CRP concentrations in observational studies (Lui *et al*., 2001 & Van Dam *et al*., 2000). Dietary GI and GL may possibly increase the risk for a number of NCDs through adverse effects on the blood lipid profile and systemic inflammation (Liu *et al*., 2002).

It might be expected that diets that minimise postprandial glucose and insulin would reduce CRP concentrations because of the dietary glycaemic load, which influences insulin resistance. Swift postprandial glucose and insulin response can be caused by increased intakes of carbohydrates where these carbohydrates are quickly digested and absorbed (Jenkins *et al*., 1987; Miller, 1994). Liu *et al*. (2002) investigated whether the glycaemic load of carbohydrates correlated with concentrations of CRP in middle-aged women. They concluded that the glycaemic load correlated directly with the plasma concentration of CRP when adjusting for body mass index (BMI), TE intake as well as other reported risk factors for ischaemic heart disease (Liu *et al*., 2002). It was observed that both the quality and the quantity of carbohydrate intake were directly associated with CRP concentrations. Their findings imply that the increase in CRP concentrations is initiated by a pro-inflammatory process caused by an increased glycaemic load, which affects insulin resistance. Long-term exposure to high dietary glycaemic load can result in repeated postprandial hyperglycaemia, hyperinsulinaemia or insulin resistance, which in turn may give way to high CRP concentrations. It is also suggested that the liver can be stimulated to increase the production of end acute phase reactants, such as CRP, by means of hyperglycaemia, which leads to the production of advanced glycation end products. Studies have related hyperglycaemia to inflammation by demonstrating simultaneous inflammation, endothelial dysfunction and insulin resistance at a physiological level (Yudkin *et al*., 1999). Oxidative stress on the endothelium is one of the numerous proposed mechanisms that promote inflammation and is enhanced by hyperglycaemia (Mohanty *et al*., 2002). Levitan and co-workers (2008) reported that the quality as well as the quantity of carbohydrates consumed may influence inflammation (measured as CRP concentrations) in non-diabetic women. CRP presented with a small but progressive increase across quintiles of dietary GI, with a highly significant difference between first and fifth quintiles. They further concluded that diets which are characterised by a lower GI and GL were associated with a lower CRP concentration. In summary, a diet that reduces postprandial glucose, through intakes of the correct glycaemic load, could prevent an increase in CRP concentrations. However, results regarding carbohydrate intake and CRP concentrations are still inconclusive and further research is

needed. The current study will also give insight into the effects of urbanisation among the study population, as the more refined carbohydrates will be expected to be present in the urban group and, therefore, higher CRP concentrations are expected (see Section 4.1).

Biomarkers of inflammation, including CRP, have indicated a positive correlation with diets containing high saturated and trans-fatty acids (Fung *et al*., 2001; King *et al*., 2003; Lopez-Garcia *et al*., 2004). Fung *et al*. (2001) identified two dietary patterns, *i.e*. the prudent dietary pattern, which would typically contain high intakes of whole grains, fruit, vegetables, fish and legumes, and the Westernised dietary pattern, which is characterised by increased intakes of red meat, processed meat, high-fat dairy products and sugar-containing beverages. It is important to note that Fung and co-workers (2001) did not correlate CRP with the individual nutrients in the diet; the overall discussion rather focused on the positive correlation between high fat intake, especially saturated and trans-fatty acids, and inflammation. King *et al*. (2003) made use of the 1999 to 2000 National Health and Nutrition Examination Survey data and reported a modest correlation between saturated fat consumption and high CRP concentrations.

In a study conducted by Esposito *et al*. (2004), it was reported that subjects who randomly received the Mediterranean diet presented with an associated decrease in CRP concentrations as well as other inflammatory markers over a two year period. The Mediterranean diet consisted of high intakes of fruit, vegetables, nuts and whole grains and had a high content of mono-unsaturated fatty acids (MUFA). One limitation of this study was the inability to determine whether individual components of the diet were responsible for the changes observed or if it was the sum of all the dietary modifications. In another similar study conducted by Estruch *et al*. (2006) it was reported that high-risk patients, who after nutritional education and advice changed to a baseline Mediterranean diet supplemented with virgin olive oil presented with significant differences ($p < 0.003$) in CRP concentrations compared to only a low-fat diet. In summary, the clinical usefulness of a whole-diet approach in the prevention of CVD, through a reduction in inflammatory markers, is being emphasised rather than individual food components (Esposito *et al*., 2004).

In an experimental study conducted by Blum and co-workers (2006) they reported a postprandial decrease in CRP concentrations of 6%, two hours after the subjects ingested a 1000 kcal meal containing 45% MUFA. In an 8-week dietary intervention study, overweight subjects received a high-SFA or MUFA diet and then had biopsies of

abdominal fat pads taken. Consumption of the SFA diet resulted in increased expression of adipose tissue genes promoting inflammation whereas the MUFA diet led to an anti-inflammatory gene expression profile (van Dijk *et al*., 2009). Evidence indicates that consumption of *i.e*. PUFAs, MUFAs, flavonoids, and carotenoids from food is associated with a decrease in inflammatory markers, whereas SFA, trans-fatty acids, high-GI carbohydrates, and a high n6/n3 PUFA ratio are associated with increased levels of inflammation. The Mediterranean dietary pattern may best fulfil requirements for an anti-inflammatory diet.

Albert and co-workers (2002) reported that CRP concentrations were lower with moderate alcohol intake compared with no or occasional alcohol consumption. This mimics the U or J-shaped correlation between alcohol consumption and CVD mortality. They concluded that alcohol may attenuate CVD mortality in part through an anti-inflammatory mechanism, therefore supporting the findings of Imhof *et al*. (2001), in which they suggested that the anti-inflammatory action of alcohol, with regard to CRP, could contribute to the link between moderate alcohol consumption and lower CVD mortality.

The effect of dietary changes on an expressed phenotype varies significantly between individuals (Ordovas *et al*., 2007). The differences in dietary habits are also seen between rural and urban groups (as discussed in detail in Section 2.2.1). Certain individuals can be classified as hypo-responders to dietary intervention and others as hyper-responders (Ordovas *et al*., 2007). In the field of nutrigenetics, gene-diet interactions are being studied in an effort to improve the understanding of various factors mediating individual responses to different dietary interventions. There is a growing amount of literature on inter-individual differences in response to certain dietary factors, for example high or low intakes of saturated fatty acids (Ordovas, 2006; Prentice *et al*., 2006). Investigation of the interactions between genetic variation and the diet are providing substantial evidence to support the concept that more individualised nutrition, or in this case population-specific nutrition, is required to address the interaction of different dietary factors and genetic variation on CVD risk (Ordovas, 2006).

Vigilance is needed before one can implement these results into a clinical environment for three major reasons: (i) the interpretation of the significant results is dependent on a specifically chosen significance level, which is frequently dependent on the study design, meaning that different analyses may produce different significant results, (ii) preliminary analyses, which have reported associations between a phenotypic trait and specific

gene-nutrient interactions, are often not repeatable in succeeding studies, and (iii) gene-gene interactions, as well as differences in ethnicities, are often ignored when interpreting gene variations that cause different phenotypic outcomes. If all of the aforementioned is considered, it is evident that there is a place for nutrigenetics in CVD prevention and that it could play a vital role in its future treatment if implemented in a population-specific manner.

2.4 GENETICS OF C-REACTIVE PROTEIN

In 1983 Whitehead and co-workers reported that they had isolated a complementary DNA (cDNA) clone for CRP from an adult human liver cDNA library. The single copy of the human *CRP* gene was localised to the proximal arm of chromosome 1 by means of studying somatic cell hybrids (Whitehead *et al*., 1983) and through *in situ* hybridisation, the *CRP* gene was specifically localised to 1q21-1q23 (Floyd-Smith *et al*., 1986). CRP spans approximately 2.5 kilobases (kb) of DNA and consists of two exons separated by a single 278 base pair (bp) intron (Lei *et al*., 1985).

The regulation of the *CRP* gene is suggested to occur principally at transcriptional level. The regulation of *CRP* transcription is relatively complicated because in the proximal part of the *CRP* promoter there are two locations for IL-6-inducible CCAAT/enhancer binding proteins (Ramji & Foka, 2002), two hepatocyte nuclear factor-1 binding sites (Toniatti *et al*., 1990), a signal transducer and activator of transcription type 3 response element (Zhang *et al*., 1996), and an NF-κβ site (Voleti & Agrawal., 2005). The control of baseline CRP expression is also multifaceted and entails other regulatory elements, such as the transcription factor termed upstream stimulatory factor 1 (USF1), which has been determined to bind to E-box elements in the proximal *CRP* promoter (Szalai *et al*., 2005). The E-box elements are DNA sequences, which usually lie upstream of a gene in the promoter region. Differences in the sequences of these E-box elements are suggested to be associated with differences in CRP concentrations as well as CVD risk and in addition the variations observed interact with environmental exposure (Hage & Szalai, 2007).

Carlson *et al*. (2005) resequenced the entire *CRP* gene plus an additional 1,700 kb upstream and 2,800 kb downstream of the gene in 47 individuals and discovered only 31 SNPs. Of the 31 SNPs that were determined, 13 were polymorphic in caucasians and 30 were polymorphic in black African American individuals (Carlson *et al*., 2005). Hage and Szalai (2007) examined evidence that the coding and promoter regions of the *CRP* gene

are polymorphic and determined ample evidence that certain *CRP* gene polymorphisms affect the amount of CRP produced (Crawford *et al*., 2006; Hage & Szalai, 2007; Hage *&* Szalai, 2009; Lawlor *et al*., 2008a; Lange *et al*., 2006; Rhodes *et al*., 2008; Shen *et al*., 2008; Suk *et al*., 2005; Wang *et al*., 2006) and that this in turn will have an impact on CVD risk.

2.4.1 *CRP* gene polymorphisms

It is hypothesised that genetic variation in the *CRP* gene may affect the basal CRP concentrations of an individual and thus increase the risk of developing CVD in the future (Pai *et al*., 2008; Eklund 2007). Pankow and co-workers (2001) reported evidence of family studies that indicated heritability estimates of basal CRP concentrations varying from 35 - 40%.

The amount of research on *CRP* polymorphisms has increased in the past decade. A large number of studies conducted independently by different research groups have established the association between different *CRP* polymorphisms and baseline CRP concentrations (Crawford *et al*., 2006; Hage & Szalai., 2009; Hage & Szalai, 2007; Lawlor *et al*., 2008a; Lange *et al*., 2006; Rhodes *et al*., 2008; Shen *et al*., 2008; Suk *et al*., 2005; Wang *et al*., 2006). These findings are supported by a number of genome wide association studies (GWAS), which investigated a large number of SNPs across the genome and determined that *CRP* SNPs are indeed important determinants of CRP concentrations (Benjamin *et al*., 2007; Melzer *et al*., 2008; Ridker *et al*., 2008a). The populations investigated included Europeans, African Americans and caucasian Americans.

2.4.1.1 *CRP* gene polymorphisms associated with serum CRP concentrations

The minor alleles of a number of different SNPs (reference sequence [rs] 7553007, rs1341665, rs2027471, rs1205 and rs2794520) have been reported to have an association with lower basal CRP concentrations. Seven *CRP* SNPs were investigated in more than 1,000 MI survivors, across cities in Europe, in an investigation by Kolz *et al*. (2008) and it was determined that a significant association was present between the minor alleles of two SNPs (rs1800947 and rs1205) and lower CRP concentrations (Kolz *et al*., 2008 & Schumacher *et al*. 2009).

Another two SNPs, namely rs1205 and rs2808630, were determined by Crawford *et al*. (2006) to have an association with decreased CRP concentrations. These were determined in a non-Hispanic American population. Suk *et al*. (2005) determined that CRP concentrations were significantly higher with the additional presence of the mutant allele for rs1417938, where it was observed that the CRP concentrations were highest for the mutant homozygous genotype, followed by the heterozygous genotype and lowest for the wildtype homozygous genotype. These associations persisted strongly even after controlling for traditional cardiovascular risk factors.

Recent data obtained from the Cardiovascular Risk in Young Finns Study (Kivimäki *et al*., 2007) confirmed the fact that baseline CRP concentrations are fairly constant over the lifetime of an individual. The researchers of this study went about genotyping 1,560 individuals for five *CRP* SNPs that had previously indicated an association with serum CRP. The SNPs included rs2794521 and rs3091244 in the promoter region, rs1800947 in exon 2, as well as rs1130864 (Lawlor *et al*. 2008a) and rs1205 in the 3'-untranslated region. Haplotyping was conducted using these five SNPs and certain haplotypes were determined to correlate with circulating CRP concentrations in childhood as well as in adulthood. If one looked across the lifetime of the individuals, the haplotypes indicated an even stronger effect on the concentrations of CRP, which would explain up to 5% of the 35 - 40% variation in serum CRP concentrations (Kivimäki *et al*., 2007). Rhodes *et al*. (2008) conducted a large study, which included participants who had systemic lupus erythematosus and they reported that the heritability of basal CRP concentrations was around 28%.

Lange *et al*. (2006) reported associations between *CRP* SNPs and future CVD events, specifically with rs3093058, where the mutant allele was associated with an increased risk of MI in African Americans. Furthermore, rs3093062 was reported to have a CRP-lowering effect, according to Szalai and co-workers (2005). All the SNPs mentioned in this section were included in this research project.

Table 2.1 summarises the *CRP* SNPs discussed in the above mentioned paragraphs, and highlights the different population groups investigated in these different studies. The fact that the population groups were mostly caucasian or African American highlights the uniqueness of the current study population

Table 2.1 Summary of the investigated *CRP* SNPs with the target population

CRP SNP	ΔCRP	Target population	Reference
rs3093058	Increase	African American	Lange, *et al*. (2006)
rs3093062	Decrease	African American	Szalai, *et al*. (2005)
rs1800947	Increase	European	Schumacher, *et al*. (2009)
rs1130864	Decrease	Caucasian	Lawlor, *et al*. (2008a)
rs1205	Decrease	non-Hispanic American population	Crawford, *et al*. (2006)
rs1417938	Decrease	Mainly Caucasian, but included African Americans, Hispanic and Asians	Suk, *et al*. (2005)
rs2808630	Decrease	non-Hispanic American population	Crawford, *et al*. (2006)
rs1341665	Decrease	Caucasian	Wang, *et al*. (2006)
rs3093068	Increase	Caucasian	Hage and Szalai (2007)
rs2794520	Decrease	Filipinos	Rhodes, *et al*. (2008)
rs7553007	Decrease	Filipinos	Rhodes, *et al*. (2008)
rs2027471	Decrease	Filipinos	Rhodes, *et al*. (2008)

CRP = C-reactive protein; rs = reference sequence; SNP = single nucleotide polymorphism.

2.4.1.2 *CRP* gene polymorphisms associated with cardiovascular disease - Mendelian randomisation studies

Genetic association studies form the basis on which Mendelian randomisation studies are performed and genetic factors such as linkage disequilibrium, pleiotropy, genetic heterogeneity or population stratification can cause false conclusions. The logical approach suggested by Mendelian randomisation is appealing because of the argument that all the possible confounding variables should also then be spread randomly across the genotypes (Hage & Szalia, 2009). If CRP is in fact a true risk factor for CVD, the presence of an SNP that predisposes an individual to higher CRP concentrations should then be associated with higher disease risk (Hage & Szalai, 2009). Lawlor *et al*. (2008a) combined data from five large trials (a total of 18,637 individuals) and reported that the SNP rs1130864 was associated with CRP concentrations, and CRP in turn with CHD. However, there was no direct relationship between rs1130864 and the disease. Similar results were determined when the data of the Nurses' Health Study and the Health Professionals Follow-up Study was used (Pai *et al*., 2008). In African American individuals, lower CRP concentrations were reported not only for a specific *CRP* haplotype, but also a significant association with a decrease in early CVD deaths with similar (non-significant) results in white American individuals (Hindorff *et al*., 2008). In a Mendelian randomisation study conducted by Timpson *et al*. (2005) they studied associations between CRP concentrations and metabolic syndrome phenotypes and reported that plasma CRP concentrations were associated with all major metabolic syndrome constituents, but results that used the genetic determinants of plasma CRP as a tool to establish a role in the

metabolic syndrome phenotypes were not significant, suggesting non-casual associations. Most of these studies were unsuccessful in reporting associations between *CRP* SNPs and CVD risk, which raises doubt regarding the causal role of CRP in CVD risk. It is also important to note that it is estimated that to establish some sort of association between *CRP* SNPs and CVD, a total of 15,000 cases and 15,000 controls is needed, suggesting that all of the studies to date have been underpowered (Danesh *et al*., 2008). While the concept that CRP concentrations are causally associated with CVD has been challenged by Mendelian randomisation studies, these studies do not take the multifactorial nature of CVD into consideration and there is ample evidence that CRP could still be considered to be a reliable marker for low-grade inflammation that identifies individuals at higher risk of future CVD events (Casas *et al*., 2006; Timpson *et al*., 2005; Lawlor *et al*., 2008b; Smith *et al*., 2005).

2.5 OTHER GENES OF IMPORTANCE

GWASs are considered an unbiased method of association, because no prior knowledge regarding which gene or genetic marker is involved in the process being investigated is needed to establish the genes that are responsible for the disease or trait under investigation (Ben-Assayag *et al*., 2007). Although high concentrations of CRP predict increased risk of the development of metabolic syndrome, diabetes, MI and stroke (Laaksonen *et al*., 2004), data on the influence of genetic variation on CRP is not generally available. GWASs have, therefore, been used to study genetic determinants of CRP. Ridker and co-workers (2008b) stated that notable genetic effects on CRP are located outside the *CRP* locus and that the detection of these loci might be useful in further explaining mechanisms that emphasise relationships between CRP, the metabolic syndrome and other vascular events. In order to clarify the last-mentioned problem, Ridker and co-workers (2008a) conducted a GWAS which included 6,345 apparently healthy women. 336,108 SNPs were investigated as potential determinants of plasma CRP concentrations in these women.

The results of this study indicated seven loci which were associated with plasma CRP, which were statistically significant at the genome-wide level. Two of the loci which were identified, *i.e*. the *glucokinase regulatory protein (GCKR)* and the *hepatocyte nuclear factor 1 (HNF1A)* gene, have previously been reported, because of their association with early onset diabetes. One locus is a gene-poor region, which indicates large regions separated by protein coding genes, and is located on chromosome 12q23.2. The last four loci are

located in or close to the *leptin* receptor gene, the *apolipoprotein E* gene, the *IL-6* receptor gene and the *CRP* gene itself. Of the seven loci identified, the protein products of six are reported to be involved in the metabolic syndrome and atherogenesis (Ridker *et al*., 2008b). Ridker and co-workers (2008b) concluded that genes that are involved in metabolic and inflammatory regulation have significant effects on CRP concentration. Most of the polymorphisms are related to insulin resistance and the metabolic syndrome, which could suggest a pathophysiological link between CRP and CVD. Benjamin *et al*. (2007) genotyped more than 1,000 participants from the Framingham offspring cohort for 112,990 SNPs. In that study, two SNPs (rs2794520; rs2808629) were highly associated with serum CRP concentrations. It is thus clear that expected as well as unexpected genes have an influence on CRP concentrations and can assist in understanding the genetic mechanisms regarding atherosclerosis and prevention thereof.

2.6 SUMMARY OF THE LITERATURE

It is now widely recognised that the CVD burden is not only a problem of affluent countries, but is also rapidly increasing in the developing world. This increasing prevalence of CVD can be attributable to a number of factors, as discussed in the above sections, but it seems that a large contribution in the black South African population may possibly be ascribed to the interaction between the increasing urbanisation of black South Africans and their heterogeneous genetic background, on which very little research is available. This is due to the huge challenges regarding financial availability, infrastructure and similar restrictions, which hamper research. Developing strategies to curb the increasing prevalence of NCDs in black South Africans is, therefore, essential. For these strategies to be effective, they need to be transdisciplinary and should be founded on evidence-based theories (“molecules to society”). This research project has aimed to implement this strategy in elucidating the role of genetic variants in CRP on CVD risk in a black South African population undergoing a nutritional transition.

CHAPTER THREE

Materials and Methods

The following chapter is a discussion of the various techniques and methodologies used in this study. Section 3.1 gives information on the ethical aspects of this study, while in Section 3.2, detailed information on the study design and subject selection is given. Various biochemical analyses were performed, as discussed in Section 3.3. Biochemical analytical protocols were followed according to the specifications of the manufacturers' of the kits that were used. Any deviations from these protocols are reported in the text. Anthropometric measurements were undertaken to determine weight and height and ultimately to calculate BMI (Section 3.4). Sections 3.5 and 3.6 describe the manner in which blood pressure was measured, as well as the manner in which the nutrient intake data was collected by means of the quantitative food frequency questionnaire (QFFQ). CRP concentrations were determined to address the aims of this project while the other biochemical analyses, anthropometric measurements, blood pressure measurements and dietary intake analyses were performed to assess association with CRP and to address possible confounding interactions between these variables and CRP. Since the main research question of this dissertation relates to the genetic variants of CRP, Section 3.7 is devoted to the determination thereof. Section 3.9 includes a discussion of the statistical analyses undertaken, which were determined by using the *Statistica* ®[1] version 10 program.

Because of the mulitdisciplinary nature of the present study the following paragraph describes the role of the M.Sc candidate within this research. The candidate was involved in most of the phases of this project. This includes fieldwork, labarotory work, data capturing, statistical analysis and interpretation of results. In particular the student played a paramount role during the data collection phase which included the preparation of biological samples, the genetic analysis of the obtained samples as well as the quality control of the data obtained from the BeadXpress® platform, statistical analysis and interpretation thereof and writing up of the data.

[1] *Statistica* ® is a registered trademark of Statsoft Inc., Tulsa, Oklahoma, USA

3.1 ETHICS COMMITTEE APPROVAL

Permission to conduct the South African arm of the Prospective Urban and Rural Epidemiological (PURE) study was obtained from the North West Provincial Department of Health, tribal chiefs, community leaders, employers and mayors of the towns where subjects were recruited. The study was conducted in agreement with the ethical principles of the Declaration of Helsinki and was also approved by the Ethics Committee of the North-West University (Ethics number: NWU-00016-10-A1). Furthermore, the research project presented in this dissertation was evaluated and approved as a sub-study by the Ethics Committee of the North-West University. All eligible individuals gave written informed consent after all the study procedures had been explained to them in their home language. All the volunteers who took part in the PURE study had the option of withdrawing at any stage during the study. The collected data was treated confidentially and all statistical analyses were performed with anonymised datasets.

3.2 STUDY DESIGN AND POPULATION

This cross-sectional epidemiological study was embedded in the international multi-centre 12-year PURE study and was performed on the samples and data collected in 2005. The aims of the PURE study include the longitudinal follow-up of individuals with different lifestyles, to determine the differential occurrence of chronic diseases and risk factors using standardised data collection in the urban and rural areas of 17 low, middle and high income countries from around the globe. Data is collected on a national, community, household and individual level (Teo *et al*., 2009).

For the South African arm of the study it was decided to include 6,000 African households (3,000 each from urban and rural settlements) that were selected from randomly chosen addresses in the allocated areas. The choice of the sampling frame was based on representativeness of both urban and rural groups and the feasibility of long-term follow-up. The urban group was chosen on the basis of being a stable population and cluster sampling was implemented. The rural population group was also chosen on the basis of being a stable population. Thus, individuals residing in the rural area were unlikely to undergo large migrations to urban areas. Participants in the urban group were defined as belonging to an established household in a developed city area, whereas the participants of the rural group were defined as being from an area of low development, where the community is still subject to tribal laws. From the 6,000 randomly selected households, a

total of 4,000 subjects met the inclusion criteria, *i.e*. apparently healthy black South Africans between the ages of 35 and 60 years with no reported usage of medication for chronic diseases of lifestyle, tuberculosis or known infection with HIV. Of these, 2,792 (rural = 1,444, urban = 1,348) agreed to take part in the study and indicated their availability during the dates scheduled for blood collection. However, during the 12-week blood-collection period, blood was collected from only 1,006 rural and 1,004 urban apparently healthy black South Africans who gave written informed consent for the procedure.

3.3 BIOCHEMICAL ANALYSES

Fasting venous blood samples were collected by registered nurses from the ante-cubital vein of the right arm of the subjects. For the collection of serum, blood was collected in a tube without anti-coagulants and allowed to coagulate at room temperature for 30 minutes, after which it was centrifuged at 2,000 x gravitational force (g) for 15 minutes at 10°C (degrees centigrade). The resultant serum was aliquotted and stored at –70°C until the serum was analysed. For plasma samples, blood was collected in ethylenediamine tetra-acetic acid (EDTA) tubes and centrifuged at 2,000 x g for 15 minutes at 4°C. The resultant supernatant was aliquotted and stored at -70°C until analysis. For the collection of the buffy coat used for DNA isolation, citrate-treated whole blood was centrifuged at 2,000 x g for 15 minutes at room temperature and the leukocyte layer was transferred to a storage tube and stored at -70°C until DNA was extracted. All of the above-mentioned samples were centrifuged within 30 minutes after collection.

3.3.1 Measurement of CRP concentrations

CRP concentrations were measured as high sensitivity (hs)-CRP with a sequential multiple analyser computer (SMAC) by using a particle-enhanced immunoturbidimetric assay (Konelab™[2] auto analyser). The principle of the procedure is based on the measurement of turbidity due to immunoprecipitation of the CRP molecule at 540 nm (Price *et al*., 1987). Micro-particles coated with anti-human CRP were added to buffered serum samples, which attached to the free CRP molecules causing these molecules to precipitate out of solution. The increase in absorbance caused by the immunoprecipitation is recorded when the reaction has reached its end point. The change in absorbance is directly

[2] Konelab™ is a trademark of Thermo Fisher Scientific Oy, Vantaa, Finland and Cobas Integra 400 Plus, Roche, Switzerland

proportional to the amount of antigen (CRP) in solution[3]. The coefficient of variance for all assays was <10%.

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

Quantitative determination of HDL-C in the serum of the participants was analysed using the Konelab20i^{TM} auto analyser, which is a clinical chemistry analyser for colorimetric, immunoturbidometric and ion-selective electrode analyses[4]. The measurement is conducted according to the homogeneous enzymatic colorimetric assay principle (Gordon *et al*., 1977). When magnesium ions and dextran sulphate are present, water-soluble complexes containing LDL-C, very low density lipoprotein (VLDL) as well as chylomicrons are formed. These complexes are resistant to polyethylene glycol (PEG)-modified enzymes. The HDL-C concentration is then measured enzymatically by cholesterol esterase and cholesterol oxidase coupled with PEG to the amino groups. Cholesterol esters are broken down quantitatively into free cholesterol and fatty acids (by cholesterol esterase). When oxygen is present, the cholesterol is oxidised by cholesterol oxidase to form Δ4-cholestenone and hydrogen peroxide. This hydrogen peroxide, together with aminoantipyrine, sodium N-(2-hydro-3-sulfopropyl)-3,5-dimethoxyaniline, hydrogen and water, reacts with peroxidase to form a blue pigment. The colour intensity of the blue quinoneimine dye formed is directly proportional to the HDL-C concentration. The absorbance is then measured at 583 nm. TC was analysed by oxidase-peroxidase and phenol aminoantipyrine reagents on the Konelab20i ^{TM} auto analyser (Thermo Fisher Scientific, Vantaa, Finland). Cholesterol esters are enzymatically hydrolysed by cholesterol esterase to cholesterol and free fatty acids. Free cholesterol, including that originally present, is then oxidised by cholesterol oxidase to cholest-4-en-3-one and hydrogen peroxide. The hydrogen peroxide combines with hydroxybenzoic acid and 4-aminoantipyrine to form a chromophore (quinoneimine dye) which may be quantitated at 500-550 nm (Allain *et al*., 1974).

The Konelab20i^{TM} auto-analyser does not measure LDL-C particles directly, but instead estimates them using the Friedewald equation by subtracting the amount of cholesterol associated with other particles, such as HDL-C and VLDL, assuming a prolonged fasting state: $H = C – L – kT$, where H is HDL-C, L is LDL-C, C is TC, T is triglycerides, and k is a

[3] CRP measurement was conducted by Dr P.T. Pisa, Centre of Excellence for Nutrition, North-West University, Potchefstroom campus
[4] HDL-C levels were measured in the laboratory of Dr Christa Grobler of the North-West University, Vaal Triangle Campus

constant which is 0.20 if the quantities are measured in $mg.dL^{-1}$ and 0.45 if in $mmol.L^{-1}$ (Johnson *et al*., 1997).

3.3.3 Determination of triglyceride concentrations

Triglycerides were also measured with the SMAC, using the Konelab™ auto analyser. The principle of this procedure is based on the activity of lipase, which hydrolyses triglycerides to glycerol and fatty acids (Young, 2000). Glycerol is phosphorylated to glycerol-3-phosphate, which is in turn oxidised to dihydroxyacetone phosphate and hydrogen peroxide. When the hydrogen peroxide reacts with the 4-aminoantopyrine and 4-chlorophenol, a quinoneimine dye is formed. The absorbance is then measured at 510 nm. Finally, the triglyceride concentration is calculated by using a calibration curve (Thermo Electron Corporation, Vantaa, Finland)[5].

3.3.4 Determination of fibrinogen concentrations

The fibrinogen concentrations of the participants were measured by the use of the modified Clauss method using the Multifibrin U-test kit on the Dade Behring BCS coagulation analyser (Multifibrin U-test, Dade Behring, Deerfield, USA). This method is based on coagulation of citrated plasma by adding a large amount of thrombin and where the coagulation time per blood sample is determined by the amount of fibrinogen in the blood sample (Mackie *et al*., 2003). A calibration curve was calculated by a fibrinogen calibrator kit and the concentration of fibrinogen determined from a reference pooled blood sample.

3.3.5 Determination of human immunodeficiency virus status

Subjects each signed informed consent to be tested for HIV after receiving pre-test counselling in groups of ten by a trained counsellor. EDTA-treated whole blood was used for the First Response rapid HIV card test 1-2.0 (PMC Medical, Nani Doman, India) which is based on the principle of immunochromatography in which a nitrocellulose membrane is precoated with recombinant HIV-1 capture antigens as well as HIV-2 antigens. When the test sample, along with assay diluent, flows through the nitrocellulose membrane, the recombinant HIV-1 and -2 antigens conjugated with colloidal gold particles bind to the HIV antibodies present in the test sample. The antigen-antibody complex travels through the

[5] Triglycerides were measured in the laboratory of Dr. Christa Grobler of the North-West University, Vaal Triangle Campus

nitrocellulose membrane and binds to the corresponding immobilised HIV-1 antigens and HIV-2 antigens, which will lead to the formation of the colour, indicating reactive results[6]. When the result is positive, it is confirmed with a Pareeshak test (BHAT Bio tech India) card test. The card test distinguishes between HIV-1 and HIV-2, but not between subtypes. Both these tests were conducted by research nurses who were trained in Voluntary Counselling and Testing and complied to the UNAIDS/WHO Policy statement on HIV-testing (UNAIDS/WHO, 2004), as well as to the National Department of Health protocol. Nurses provided individual post-test counselling to subjects who wanted to be informed of their HIV status.

3.4 ANTHROPOMETRIC MEASUREMENTS

In this investigation only weight and height were used. Weight was measured in minimal clothing using a portable electronic scale (Precision Health Scale, A&D Company, Tokyo, Japan) and the participants had to be barefoot with arms hanging freely at their sides. The scales were calibrated prior to each session. Weight was measured and recorded in duplicate. The mean of the two measurements was used. The height was measured with a stadiometer (IP 1465, Invicta, London, UK) while the head was held in the Frankfort plane (represented by a line between the margin of the orbit of the eye and the tragion) while the height was recorded to the nearest 0.1 cm after the subject had inhaled fully and maintained the erect position. Height was also measured twice and the mean was used. BMI was calculated by dividing body mass (in kilogram [kg]) by the height squared (in metres [m]) and used as an estimate of body composition (measured in $kg.m^{-2}$).

3.5 BLOOD PRESSURE

Seated blood pressure was measured on the right arm in duplicate, using the OMRON HEM-757 automatic digital blood pressure monitor (Omron Healthcare, Kyoto, Japan). The subject should not have smoked, exercised or eaten in the last 30 minutes, as well as be rested and calm for 5 minutes. The subject should also not have climbed stairs in the last 15 - 30 minutes prior to the measurement being taken. The subject was seated upright and relaxed with his/her right arm supported at heart level while the measurement was taken using the brachial artery. The exact readings for systolic and diastolic pressures in millimetres of mercury (mmHg) were reported[7].

[6] HIV was measured by Dr. M.J. Watson, School of nursing, North-West University, Potchefstroom campus
[7] Blood pressure measurements were conducted by members of Hypertension in Africa Research Team (HART), North-West University, Potchefstroom campus.

3.6 QUESTIONNAIRES

All the subjects participating in the PURE study completed a set of questionnaires during individual interviews conducted by the researchers and specially trained field workers in the language of the subjects' choice. The data recorded for each individual, within the adult questionnaire, included age, gender, medical history (stroke and diabetes incidence), inhabitance area (urban or rural) and habits of tobacco use. All questionnaire data was captured electronically using the Microsoft EXCEL® spread sheet program, which was exported to the *Statistica* ® version 10 program, for statistical analyses.

Fieldworkers who had been trained to ensure accurate completion of the QFFQ were responsible for the collection of dietary intake data. The following data was used from the QFFQ: percentage of carbohydrates, total protein, total fat, MUFA, PUFA and saturated fatty acids. The grams of total fibre intake and total sugar intake were also used from this questionnaire. The total energy consumption (total kJ intake) for each individual was also collected from the QFFQ. Food portion books designed, standardised and validated for use in the North West Province of South Africa were used to ensure more accurate estimations of portion sizes and household measures. The QFFQ used was previously validated for use in South Africa, and collected dietary data regarding the frequency, amounts and preparation methods of foods and beverages consumed (MacIntyre *et al*., 2001(a); MacIntyre *et al*., 2001(b)). This QFFQ was validated against 7-day weighed food records, 24-hour urinary nitrogen excretion and estimated basal metabolic rate. Even though this QFFQ is deemed to be reliable it is important to remember that it relies on memory as well as the accurate estimation of portion sizes. The *Foodfinder3*®[8] program was used to computerise the dietary data, which was sent to the Medical Research Council of South Africa for nutrient analysis[9].

3.7 GENETIC ANALYSES

For the genetic analyses to be conducted, the DNA of the PURE individuals was isolated using two methods, which are discussed in Section 3.7.1. The *CRP* gene was in turn sequenced in 30 randomly selected black South African individuals from the PURE study to determine possible novel polymorphisms within the *CRP* gene (Section 3.7.2) in the population under investigation. Alignment of the 30 individuals with the reference gene

[8] *Foodfinder3*® is a registered trademark of the Medical Research Council, Tygerberg, 2007
[9] Dietary data was computerised and analysed by Ria Laubscher, at the Medical Research Council, South Africa

sequence (Genbank accession AF449713)[10] was achieved by means of the BioEdit program Version 7.1.3.0 (Hall, 1999) mentioned in Section 3.7.3. Table 3.6 is a summary of the polymorphisms identified in the *CRP* gene, both novel as well as SNPs reported in the literature. After the rs numbers for the different SNPs had been identified *via* an *in silico* search strategy, the PURE DNA samples were sent off for genotyping of the SNP set using the BeadXpress®[11] platform (Section 3.7.3).

3.7.1 Deoxyribonucleic acid isolation

Two protocols were followed for the isolation of DNA[12]. The first protocol allowed for the isolation of genomic deoxyribonucleic acid (gDNA) from citrate-treated buffy coat using the modified protocol of the FlexiGene™[13] DNA extraction kit (QIAGEN). In a 15 mL centrifuge tube, 2.5 mL of lysis buffer (FG1 Cell Lysis Solution) was mixed with one mL of buffy coat. The tube was gently inverted until thoroughly mixed, followed by centrifugation at 2,000 x g for 5 minutes, at 4°C. As much supernatant as possible was removed, without disrupting the visible white pellet. A solution of 0.5 mL denaturation buffer (FG2) and 5 µL QIAGEN protease was added to the tube that contained the pellet. The tube was vortexed until the white blood cells were resuspended (10 - 15 seconds). The tube was incubated at 65°C for 10 minutes. Following incubation, 0.5 mL isopropanol was added to the tube and mixed by inversion, until visualisation of the DNA precipitate as a clump. The sample was centrifuged at 2,000 x g for three minutes, after which the DNA was visible as a small white pellet. The supernatant was decanted and 0.5 mL of a room temperature 70% ethanol solution was added to the DNA. The tube was inverted several times to wash the DNA pellet and the sides of the centrifuge tube. Centrifugation was repeated, followed by the ethanol being aspirated. After the addition of 200 µL of DNA rehydration solution, the tube was vortexed at a low speed. The DNA was incubated at 65°C for one hour, after which the DNA yield and purity were determined. DNA purity (A_{260}/A_{280}) and yield were determined with the Nano-Drop spectrophotometer (ND-1000). An A_{260}/A_{280} ratio between 1.7 and 1.9 was considered acceptable, as the DNA was then deemed to be without any contamination of proteins, phenol or ribonucleic acid (RNA). Equation 3.1 was used to calculate the absorbance of a sample.

[10] The sequences were aligned by Dr K.R Conradie and the student preparing this research project, Centre of Excellence for Nutrition, North-West University, Potchefstroom campus
[11] BeadXpress® is a registered trademark of Illumina® Inc., San Diego, CA, USA
[12] DNA was isolated by the students in the Nutrigenetics laboratory, Centre of Excellence for Nutrition, North-West University, Potchefstroom campus
[13] FlexiGene™ is a trademark of the QIAGEN Pty. Ltd., Australia

Equation 3.1 Calculation to determine absorbance of DNA sample

$$\text{Absorbance} = -\log\left(\text{Intensity}_{sample} / \text{Intensity}_{blank}\right)$$

The concentration of the DNA was calculated by applying Equation 3.2. A final DNA working dilution of 10 ng.μL^{-1} was prepared and was stored at -20°C.

Equation 3.2 Relationship of double stranded DNA concentration to ultraviolet sample absorbance

$$[\text{double stranded DNA}] = A_{260} \times (50\ \text{ng}.\mu l^{-1} \times \text{dilution factor})$$

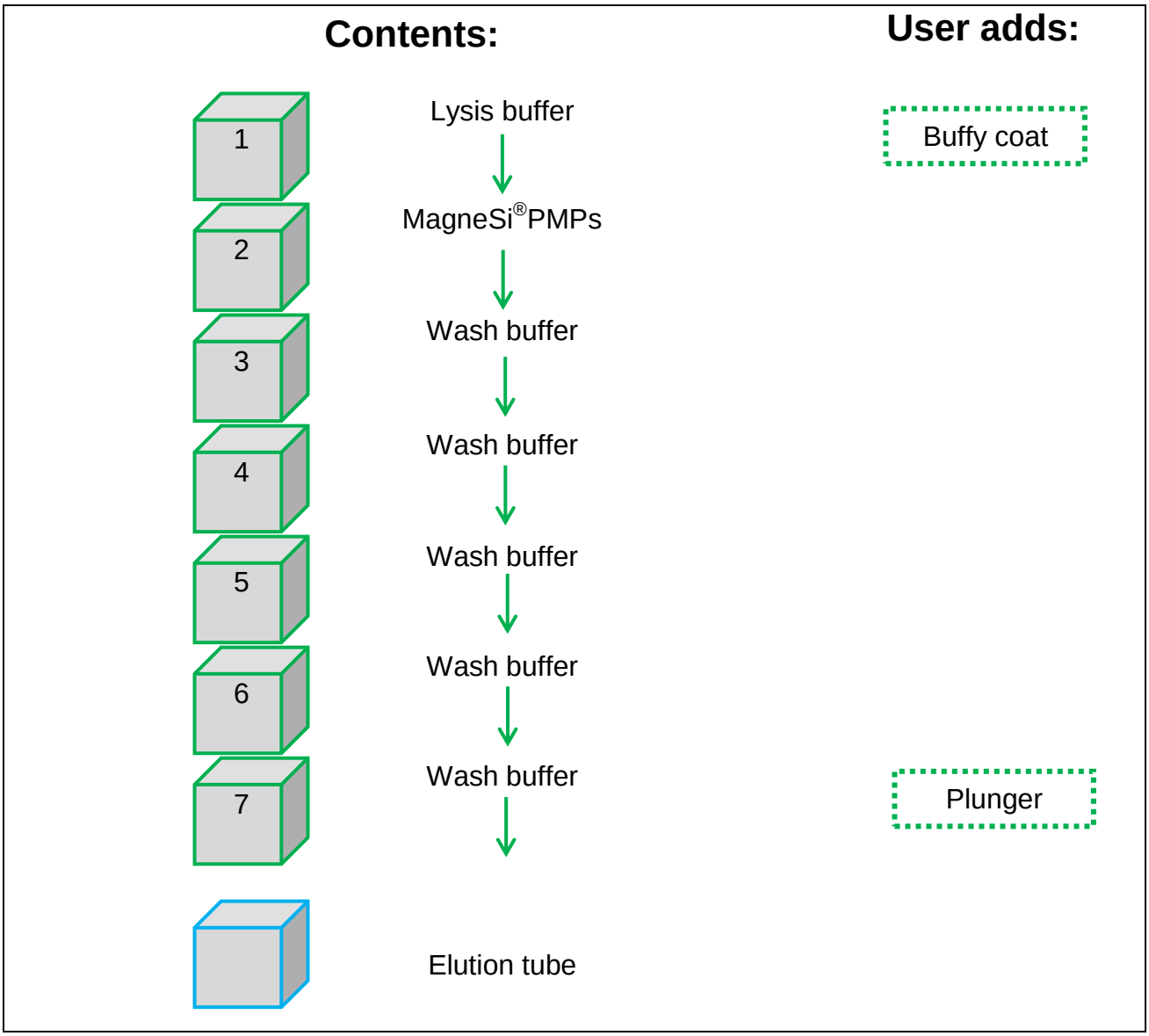
A_{260} = absorbance of sample at 260 nanometers; DNA = deoxyribonucleic acid

The second round of DNA isolation was performed using the Maxwell®[14] 16 blood DNA purification kit (Figure 3.1). This was only performed on 350 samples, of which the DNA yield was not sufficient from the initial DNA isolation method using the Flexigene kit (QIAGEN). This instrument is supplied with preprogrammed purification procedures and is designed for use with predispensed reagent cartridges, maximising simplicity and convenience. The Maxwell® system was chosen as it is a more effective isolation method than the Flexigene kit in terms of DNA yield. This was viewed as an advantage, as the 350 samples to be isolated via this method had given low yields using the Flexigene kit.

The Maxwell® 16 blood DNA purification kit is a magnetic bead-based (MagneSi®PMPs) cartridge DNA extraction system that uses paramagnetic-particle technology for faster separation of the DNA from all the other cellular debris and enables the extraction of up to 16 samples in a single run. The unique design offers mixing, capturing of paramagnetic particles, binding, then purification through a series of capture and release washes. The extraction was performed on approximately 50 mg of buffy coat sample. The general principle is that the magnetic core of the bead is coated with a material that will bind nucleic acids under specific conditions.

[14] Maxwell® is a registered trademark of Promega Corporation, Madison, WI, USA

Figure 3.1 Maxwell® 16 Blood DNA Purification System Cartridge



For DNA isolation using the Maxwell® 16 blood DNA purification kit system, the standard elution volume plate was used and the DNA buffy coat protocol was selected. Buffy coat (500 μL) was pipetted into a cartridge while plungers were added to the cartridge before it was placed in the Maxwell® 16 blood DNA purification kit system (Figure 3.1). Elution buffer (500 μL) was added to each blue elution tube and was placed in the elution tube slots at the front of the Maxwell® platform. The buffy coat sample undergoes a sequence of lysis, extraction, purification and elution steps in the chambers of the cartridge. The lysis buffer step (step one) is formulated to give the correct concentration of salts, pH and cofactors so that the target molecule (DNA) will preferentially bind to the beads. It therefore breaks open the cells by destroying the lipid membranes that enclose the cells, as well as the nucleic membranes within the cells. The MagneSi®PMPs[15] are used for

[15] MagneSi®PMPs is a registered trademark of Promega Corporation, Madison, WI, USA

mixing, capturing, binding and ultimately purifying the sample. The movement of the plungers keeps the beads in suspension, which improves the rate of binding and washing. At the end of each step the magnetic rods move to the bottom of the plungers so that the magnetic particles are captured out of the solution. The beads are moved from well to well with very little carryover of solution. The wash solutions (step three to step seven) are formulated so that the target molecule (DNA) stays on the beads and other molecules will wash off into the solution. For DNA extraction there is a drying step between the last wash (step seven) and the elution. The drying step allows the residual alcohol on the beads to evaporate. Once the run has been completed, the blue elution tubes (which contain the low salt buffer) are transferred to the magnetic elution rack and the isolated DNA is then pipetted into a storage tube.

The DNA had to be normalised to 50 ng.μL^{-1} because of the requirements of the BeadXpress® reader. The normalisation process was as follows: Each DNA sample's concentration was measured twice, as described previously. The mean of the two measurements was used in Equation 4.1 to determine the amount of double distilled water and DNA from the stock solution that was needed for setting up a working dilution of 50 ng.μL^{-1}.

Equation 3.3 Normalisation of DNA to 50 ng.μL^{-1}

(C X V)/A = B	**V – B = D**

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

After the concentrated DNA aliquot had been diluted with the calculated amount of double distilled water, another absorbance measurement was performed to ensure that the final working dilution of the samples was 50 ng.μL^{-1}. The A_{260}/A_{280} ratio was also recorded to ensure that the DNA quality was sufficient and without any contamination of, for example, ethanol and other proteins. DNA samples were plated in a 96-well plate after the normalisation process had been completed.

3.7.2 Determination of novel polymorphisms in the *CRP* gene of the black South African population

To determine polymorphisms of interest in the *CRP* gene, several processes had to be completed in a specific order. The subsequent section will discuss in detail the process that was followed in order identify the polymorphisms to be investigated in the black South African population.

The four regions of interest were located in the promotor, intron, exon 2 and downstream region and covered the area in the *CRP* gene that was of importance in order to identify the relevant *CRP* polymorphisms. The length of each region is outlined in Table 3.1. Results of the optimisation of each region are discussed in Section 4.5.1 – 4.5.4.

Primers sets specifically for the four regions of interest in the *CRP* gene were designed[16] and used for the amplification of these regions, using the polymerase chain reaction (PCR). Primers were specifically designed *via* the Primer-BLAST program (Table 3.5), which is a software program provided by Integrated DNA Technologies. The principles to which the primers were designed are discussed in Section 4.4. The optimal annealing temperature for each region, as well as the region sizes, are summarised in Table 3.5. The BioRad CFX96 Touch™[17] Real-Time PCR Detection System was used for amplification of the aforementioned regions in the CRP gene. The optimal annealing temperature of each primer was determined (results presented in Section 4.4.2 – 4.4.4) *via* optimisation and the findings are summarised in Table 3.5. For each reaction, Bioline My Taq™[18] reaction buffer (5 μL) and HS DNA Polymerase (1.25 U) were used, comprising the following components: 25 nmol dNTPs, 75 nmol $MgCl_2$ and 0.4 pmol of the forward and reverse primer. A final working volume of 15 μL was used.

Amplification was achieved *via* the following thermal cycling conditions:

1. Initial denaturation at 95°C for 60 seconds
2. Cycling conditions (30 cycles):
    a. denaturation at 95°C for 15 seconds
    b. annealing at the specific annealing temperature (T_a) for 15 seconds
    c. elongation at 72°C for 10 seconds.

A final holding step was also included at 20°C.

[16] Primers were designed by Dr K.R. Conradie, Centre of Excellence for Nutrition, North-West University, Potchefstroom campus
[17] BioRad CFX96 Touch™ is a trademark of Bio-Rad Laboratories, Inc., California, USA
[18] My Taq™ is a trademark of Bioline, London, UK

Table 3.1 *CRP* primers used for the amplification of the *CRP* gene

Primer	Primer sequence	Annealing temperature	Region	Length of regions (bp)
CRP 1F	5'-tga aga gtg agt taa gta ggg aac tg-3'	58.0°C	Promotor	980
CRP 1R	5'-taa ggg agt ttg cgc cac ta-3'		Promotor	
CRP 2F	5'-tgt tct gaa ata att ttg ctt cc-3'	55.0°C	Intron	997
CRP 2R	5'-ctc ctg ccc caa gat gat-3'		Intron	
CRP 3F	5'-gga tcg tgg agt tct ggg ta-3'	54.5°C	Exon 2	943
CRP 3R	5'-ctc tgc tgg ggc aat tct aa-3'		Exon 2	
CRP 4F	5'-ttt aat tcc cca ccc atg ac-3'	59.0°C	Exon 2	979
CRP 4R	5'-agg gga ctc ttg gac agg tt-3'		Downstream	

bp = base pairs; CRP = C-reactive protein; t = thymine; a = adenine; c = cytosine; g = guanine; F = forward; R = reverse; °C = degree centigrade

After completion of the PCR process, the success of the reaction was tested *via* 2% agarose gel electrophoresis of the PCR products. The gels consisted of 2% molecular grade agarose (BioRad), 1 X Tris/Borate/EDTA (TBE) buffer and 0.5 $\mu g.ml^{-1}$Actual quantity! ethidium bromide (EtBr). Samples were loaded with 2 μl of Bioline loading buffer and were electrophoresed at 100 volts for 60 minutes. Fragments were visualised using an ultraviolet transilluminator and electronic copies of the image were captured.

The *CRP* gene was sequenced in 30 randomly selected samples from the 2,010 PURE individuals *via* fluorescently labelled cycle sequencing, which is based on the basic principles of Sanger sequencing (Sanger *et al*., 1977) at the Central Analytical Facility, Stellenbosch University. NucleoFast®[19] 96 PCR Clean-Up Kits were used to clean the PCR products prior to the sequencing process being performed. NucleoFast® technology is based on ultrafiltration where the nucleic acids from the PCR reaction are collected on a filter membrane, while the contaminants are filtered directly to the waste. Contaminants, such as primers, deoxyribonucleotide triphosphate (dNTPs) and salts, pass through the membrane under vacuum or centrifugal pressure. The amplified products are retained on the membrane. The products can be recovered directly from the membrane after incubating the membrane for 1 to 3 minutes with either nuclease-free water or a low salt buffer.

[19] NucleoFast® is a registered trademark of Machery-Nagel, Germany

After the PCR cleanup had been done, the Central Analytical Facility at the Stellenbosch University conducted the sequencing. The standard ABI Prism®[20], BIGDye®[21] Terminator version 3 Ready Reaction Cycle Sequencing Kit supplied by Applied Biosystems was used as a platform for the sequencing reactions. Sanger sequencing works on the basis of chain termination sequencing (Sanger *et al*., 1977). The double-stranded DNA denatures to form a single strand from which the sequencing reaction can take place. A primer specific to the DNA fragment to be sequenced, an enzyme (DNA polymerase), dNTPs and fluorescently labelled 2',3'-dideoxyribonucleotide triphosphates (ddNTPs) are added to the reaction mixture. The DNA polymerase binds to the primer and starts to generate a new strand of DNA by incorporating complementary dNTPs. The enzyme continues to extend the new strand until it incorporates fluorescently labelled ddNTPs. These ddNTPs are chemically altered to terminate the elongation of the DNA strand and the enzyme falls away. This action is repeated several times and generates a large number of fragments of different lengths ending in fluorescently labelled nucleotide bases.

The sequenced product is purified using a modification of the Princeton Separations post-sequencing kit, which is used to remove unincorporated fluorescently labelled ddNTPs. The principle of the kit is based on perforated beads that retain the ddNTPs and dNTPs so that the larger fragments can pass through. The purified product is in turn electrophoresed on a 50 cm capillary on either an ABI3730xl DNA analyser or an ABI3120xl genetic analyser according to standard protocols.

After the sequences of the 30 samples had been received back (both forward and reverse), they were aligned using BioEdit program Version 7.1.3.0 (Hall, 1999) according to the reference *CRP* gene (Genbank accession AF449713) and these alignments were used to identify alterations *via* manual inspection. Alignment was conducted electronically by using the alignment functions of the BioEdit program. SNPs that were identified, *via* the 30 individuals sequenced, were grouped into reported SNPs and novel SNPs. Rs numbers were then established for all the SNPs included by means of the University of California Santa Cruz (UCSC) Human Genome Assembly software. The "blat" function in the program was used to determine the positions of the SNPs, after which the "Genome Browser" function was used to determine the rs numbers. In addition to the SNPs determined in the 30 individuals, an *in silico* review of the literature on *CRP* SNPs that have an effect on CRP concentrations was also undertaken, using the following search

[20] ABI Prism® is a registered trademark of Applied Biosystems, California, USA
[21] BIGDye® is a registered trademark of Applied Biosystems, California, USA

terms: “c-reactive protein” and “genetics”, “c-reactive protein” and “CVD”, “*CRP* polymorphisms” as well as “c-reactive protein” and “single nucleotide polymorphisms”. The following datasets were searched: Medline, EBSCO host, Scopus and Web of science. These SNPs were confirmed to be either reported or novel by use of the dbSNP website (http://www.ncbi.nlm.nih.gov/projects/SNP/) to identify if the SNP had an rs number.

3.7.3 Process of SNP identification using the BeadXpress® platform

After the identification of the rs numbers of all the SNPs which were to be investigated, they were compiled into a list which was submitted to Illumina® for validation to ensure accurate assay design and to identify the SNPs that were most likely to give successful results. Table 3.2 is a summary of the final SNPs that were included in the analysis of this research project after taking into account all the variables and criteria mentioned above.

Table 3.2 Summary of the identified SNPs in the *CRP* gene and rs numbers

Position on chromosome 1	Rs number	Change	Type of mutation	Reference
159685315	rs3093058	A/T	Transversion	Lange *et al*. (2006)
159685096	rs2794521	A/G	Transition	Kathiresan *et al*. (2006)
159684684	rs3093062	G/A	Transition	Szalai *et al*. (2005)
159683811	Novel	A/C	Transversion	This investigation
159683438	rs1800947	G/C	Transversion	Schumacher *et al*. (2009)
159683091	rs1130864	C/T	Transition	Lawlor *et al*. (2008a)
159682234	rs1205	G/A	Transition	Hage and Szalai (2009)
159684186	rs1417938	A/T	Transversion	Suk *et al*. (2005)
159680868	rs2808630	C/T	Transition	Crawford *et al*. (2006)
159691559	rs1341665	A/G	Transition	Wang *et al*. (2006)
159681364	rs3093068	C/G	Transversion	Hage and Szalai (2007)
159678816	rs2794520	C/T	Transition	Rhodes *et al*. (2008)
159698549	rs7553007	A/G	Transition	Rhodes *et al*. (2008)
159689388	rs2027471	A/T	Transversion	Rhodes *et al*. (2008)

a = adenine; c = cytosine; g = guanine; rs = reference sequence; t = thymine

The Illumina® GoldenGate Assay is a widely used and robust assay for multiplexed genotyping and can be used on the BeadXpress® platform using VeraCode®[22] technology for up to 384-plex reactions. The BeadXpress® platform will be fully discussed in Section 4.5. The first step in designing the assay was to select and submit a requested list of loci to Illumina®. This list was sent using the identity (RSList) format for the reported SNPs,

[22] VeraCode® is a registered trademark of Illumina® Inc., San Diego, CA, USA

whereas the sequence (SequenceList) format was used for the novel alterations. In order to ensure fast and successful assay development, these lists were evaluated with the Assay Design Tool (ADT). An SNP_Score file was generated, which was used to design an initial assay panel according to the SNP assays that were predicted to have a high likelihood of success. This SNP_Score file also provided predicted success information, validation status, minor allele frequencies from published studies and other useful information metrics (Table 4.5). The metrics were used for the selection of the SNPs most likely to result in successful genetic analyses. The actual analyses were performed by the National Health Laboratory Service (NHLS) at the University of Witwatersrand, Johannesburg. DNA was normalised as described in Section 4.3, Equation 3.3.

The typical workflow that was followed to generate the genotypic data was divided over two days. On the first day, the first step was DNA activation, which allowed for the binding of the genomic DNA samples to paramagnetic particles. This activation step was highly robust and required only a minimum input of DNA. The second step involved the combination of the assay oligonucleotides, the hybridisation buffer and the paramagnetic particles with the activated DNA in a hybridisation step. Three oligonucleotides are designed for each SNP locus, where two are specific to each allele of the SNP site and are referred to as allele-specific oligonucleotides (ASOs) and the third oligonucleotide, which hybridises to several bases downstream from the SNP site, called the locus-specific oligonucleotide (LSO). Regions of genomic complementarity and universal PCR primer sites are included in all three oligonucleotide sequences. A particular bead type is targeted by the LSO, which contains a unique address sequence. During the primer hybridisation process, the assay oligonucleotides hybridise to the genomic DNA sample bound to the paramagnetic particles. Subsequent to the hybridisation step, step three involves several wash steps, which are designed to remove excess and mis-hybridised oligonucleotides, to ensure that only the hybridised oligonucleotides are included. After the clean-up, extension of the appropriate ASO and ligation of the extended product to the LSO join the information about the genotype present at the SNP site to the address sequence on the LSO. In step four, the ligation products serve as the PCR templates for the universal PCR primers, namely P1, P2 and P3.

Day two of the determination of the genotypic data consists of a further five steps. Step one entails the labelling of primers P1 and P2 with Cy3 and Cy5 dyes, respectively. Step two involves the hybridisation of the single-stranded, dye-labelled PCR products to their complementary bead types through their unique address sequence. The BeadXpress®

Reader draws up to eight samples at a time from a standard 96-well plate and introduces them into an eight-chambered transparent groove plate that forms the bottom of the system's fluidic cell. Through a unique combination of fluid flow, gravity and capillary force, the VeraCode microbeads efficiently populate the groove plate and align closely within the grooves. Once the alignment is done, the entire fluidic cell is activated across the optical system and scanned for fluorescent intensity and code classification. The BeadXpress® Reader makes use of a dual-colour detection system to identify the unique holographic code, which is inscribed into each VeraCode microbead, and to detect the signal intensity associated with each bead. Hybridisation of the GoldenGate®[23] assay products onto the VeraCode beads separates the assay products for individual SNP genotype readout. Step three for the second day is to wash the VeraCode bead plate, which is followed by the use of the BeadXpress® reader for microbead code identification and fluorescent signal detection (step four). Therefore, for each SNP, the amplification product for homozygous genotypes displays a signal in either the Cy3 or Cy5 channels, whereas the heterozygous genotype at this locus should display a signal in both channels. The final step is to analyse the generated genotypic data through Illumina's Genome Studio®[24] Genotyping Module v1.0 program, which performs automated genotype clustering and calling. Controls are incorporated into each run to ensure accuracy and are fully discussed in Section 4.6.2.

3.8 STATISTICAL ANALYSIS

For the statistical analysis of the data, the computer software package *Statistica*® 10 was used. Identified variables were divided into continuous and categorical variables. Since the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will normally be distributed according to the central limit theorem, the large sample used in this study permitted the use of parametric statistical analysis (Rice, 2007). Descriptive statistics were conducted to obtain a broad overview of the population under investigation (Table 4.1 and Table 4.2). All the continuous variables were reported as the mean and standard deviation (±SD). The descriptive statistics were divided for urban and rural comparisons, as well as for comparing men and women. For the continuous data, the differences in the means of the identified variable were calculated for men *versus* women as well as rural *versus* urban. To achieve this, an independent T-test was performed. The differences in the mean CRP concentrations between the various

[23] GoldenGate® is a registered trademark of Illumina® Inc., San Diego, CA, USA
[24] GenomeStudio® is a registered trademark of Illumina® Inc., San Diego, CA, USA

categories of the categorical data were calculated by means of the analysis of variance test. Confounding variables for CRP were identified by means of the Spearman Rank test. A *p*-value of less or equal to 0.05 was regarded as statistically significant for all the statistical tests.

The genotype frequencies of the different SNPs under investigation were determined for the whole group and were also calculated for the urban/rural and men/women categories. The expected genotype frequencies when a population adheres to the assumptions of the Hardy-Weinberg equilibrium (HWE) were calculated and compared to the observed frequencies by using a Chi-square (χ^2) test to determine adherence to or significant differences from the assumptions of HWE for the genotype distribution of the various CRP polymorphisms. A low Chi-square value and high p-value (> 0.05) were regarded as adhering to the assumptions of HWE. The HWE also states that both allele and genotype frequencies in a population remain constant from generation to generation unless a specific disturbing influence is introduced. These influences can be non-random mating, mutations, natural selection, genetic drift and migration (Hardy, 1908).

To establish if there is a significant difference between the three genotype groups of a specific SNP and the CRP concentrations, the analysis of covariance (ANCOVA) was used. The CRP concentrations, which are reported for all the SNPs, are therefore adjusted means. An advanced general linear model was used to detect interaction effects between area of residence or gender, CRP concentrations and the specific *CRP* SNP. Post-hoc tests were not conducted because of the use of adjusted means. However, when these results were interpreted, the tendency of the results were discussed.

When investigating genetic factors influencing CRP concentrations one expects high CRP concentrations in individuals harbouring certain genotypes. Therefore, excluding individuals with high CRP concentrations might result in losing individuals carrying the allele responsible for the increase in CRP concentrations. Furthermore, one of the inclusion criteria of the participants was to be apparently healthy, however, to be cautious statistical analysis was conducted excluding those with a CRP concentration above 10 $mg.L^{-1}$. Therefore, in addition to the above-mentioned statistical analysis performed, the researcher also wanted to eliminate possible incorrect findings that could have resulted in reaction to the inclusion of individuals with acute levels of inflammation, as defined by CRP concentrations above 10 $mg.L^{-1}$ (Kushner *et al*., 2006). Therefore, the researcher

re-analysed the data but excluded participants with acute inflammation. These results are discussed in detail in Section 4.1 and Section 4.8.

CHAPTER FOUR

Results and discussion

This research project was, undertaken firstly to determine if specific reported and novel genetic variants within the *CRP* gene were associated with altered CRP concentrations and secondly, to determine the genotypic distribution of these SNPs in a black South African population to determine the possible origins of the higher CRP concentrations the researcher encountered in this population. This study is unique in that it is the first study to investigate the effects of *CRP* genetic polymorphisms on CRP concentrations in a black South African population.

The results obtained from the methods followed in Chapter 3 will be presented in the subsequent section. Section 4.1 will focus on the demographic and biochemical characteristics of the population under investigation. Sections 4.3 and 4.4 include a discussion of the DNA isolation procedure, as well as the PCR and sequencing results of the *CRP* gene undertaken in this investigation. The polymorphisms identified in the four regions within the *CRP* gene investigated in this study, as well as the polymorphisms identified in the literature, are discussed in Section 4.5. The results obtained from the BeadXpress® are discussed in Section 4.6. Finally, the association analyses of the SNPs are discussed (Section 4.7), followed by the results excluding individuals with possible acute inflammation. Section 4.9 will summarise the most important results determined from this investigation. One should bear in mind that using mean values are only a way to characterize the population, but that the data of population was representative of all the possible ranges e.g. although mean BMI suggests that the average individual in the population is overweight, there are still several individuals which were underweight, normal weight or obese.

4.1 DEMOGRAPHIC AND BIOCHEMICAL CHARACTERISTICS OF THE PURE STUDY POPULATION

Descriptive statistics for all the identified variables (as indicated in Section 3.8) were determined for the entire group, as well as for specific population subgroups *i.e*. rural *versus* urban and men *versus* women. The variables were separated into categorical and continuous variables and are presented in Table 4.1 and Table 4.2, respectively. Subjects

with missing genotypic data, as well as missing CRP concentration data, were removed from the dataset to avoid any bias occurring owing to this missing data. A total of 1,588 individuals were ultimately included for all statistical analyses. As mentioned in Section 3.8, additional statistical analyses were conducted, excluding individuals with a CRP concentration above 10 $mg.L^{-1}$, and all the means of the different variables were approximately the same (Addendum A). Those variables that differed significantly between the different subpopulations in the analyses among all the participants remained statistically significant despite the exclusion of subjects with acute inflammation (Table 4.21 in Section 4.8).

Table 4.1 Continuous baseline characteristics in the whole group, as well as between the rural/urban and men/women groups in the PURE study population

Variables	Whole group mean ± SD	Rural mean ± SD	Urban mean ± SD	p-value	Men mean ± SD	Women mean ± SD	p-value
Number of individuals	1587	775	812	--	593	994	--
CRP (mg.L^{-1})	8.44 ± 12.42	8.21 ± 12.42	8.67 ± 12.43	0.46	8.15 ± 13.27	8.62 ± 11.90	0.47
LDL-c (mmol.L^{-1})	2.93 ± 1.17	2.92 ± 1.16	2.94 ± 1.18	0.72	2.72 ± 1.15	3.06 ± 1.16	<0.01
HDL-c (mmol.L^{-1})	1.52 ± 0.64	1.52 ± 0.62	1.52 ± 0.65	0.95	1.57 ± 0.64	1.49 ± 0.63	<0.01
TC (mmol.L^{-1})	5.02 ± 1.38	4.97 ± 1.36	5.06 ± 1.40	0.19	4.82 ± 1.34	5.14 ± 1.39	<0.01
TG (mmol.L^{-1})	1.30 ± 0.80	1.21 ± 0.65	1.39 ± 0.92	<0.01	1.23 ± 0.87	1.34 ± 0.76	<0.01
Age (years)	49.24 ± 10.34	48.19 ± 9.78	50.23 ± 10.77	<0.01	49.83 ± 10.40	48.88 ± 10.30	0.08
Height (cm)	160.7 ± 1.91	160.1 ± 0.08	161.2 ± 0.09	0.12	167.6 ± 0.06	157.6 ± 0.06	<0.01
Weight (kg)	63.02 ± 16.61	61.70 ± 15.93	64.42 ± 17.19	<0.01	58.43 ± 12.05	65.75 ± 18.28	<0.01
BMI (kg.m^{-2})	24.55 ± 6.92	24.09 ± 6.53	25.03 ± 7.28	0.01	20.78 ± 4.06	26.79 ± 7.28	<0.01
Dietary variables:							
Total energy intake (kJ)	7,919.24 ± 3,763.77	6,480.69 ± 2,745.39	9,262.00 ± 4,078.14	<0.01	8,684.58 ± 4,105.57	7,464.00 ± 3,467.80	<0.01
Total protein intake (% of TE)	11.77 ± 2.03	10.96 ± 1.80	12.54 ± 1.93	<0.01	11.84 ± 2.16	11.74 ± 1.97	0.36
Total CHO intake (% of TE)	60.72 ± 9.23	65.64 ± 8.76	56.14 ± 7.03	<0.01	59.97 ± 9.07	61.17 ± 9.30	0.01
Total fibre intake (g)	20.97 ± 10.35	17.77 ± 7.47	23.96 ± 11.70	<0.01	23.07 ± 11.72	19.72 ± 9.23	<0.01
Total sugar intake (g)	44.77 ± 31.64	33.31 ± 25.59	55.46 ± 32.98	<0.01	45.50 ± 34.76	44.33 ± 29.63	0.48
Total fat intake (% of TE)	22.92 ± 7.54	19.24 ± 6.80	26.36 ± 6.51	<0.01	21.89 ± 7.33	23.53 ± 7.60	<0.01
MUFA (%)	6.05 ± 2.90	4.63 ± 2.33	7.52 ± 2.59	<0.01	5.79 ± 2.82	6.20 ± 2.94	0.01
PUFA (%)	6.99 ± 2.85	6.33 ± 3.05	7.61 ± 2.51	<0.01	6.56 ± 2.66	7.25 ± 2.93	<0.01
SFA (%)	5.54 ± 2.63	4.27 ± 2.40	6.72 ± 2.25	<0.01	5.25 ± 2.48	5.71 ± 2.70	<0.01
SBP (mmHg)	132.98 ± 24.02	128.64 ± 22.18	137.09 ± 24.97	<0.01	135.05 ± 23.68	131.73 ± 24.14	0.01
DBP (mmHg)	87.23 ± 14.33	85.36 ± 14.12	89.01 ± 14.32	<0.01	86.26 ± 14.70	87.81 ± 14.08	0.04
Fibrinogen (g.L^{-1})	3.70 ± 2.22	3.85 ± 2.16	3.54 ± 2.16	0.01	3.33 ± 2.15	3.91 ± 2.22	<0.01

CHO = carbohydrates; DBP = diastolic blood pressure; g = grams; g.L^{-1} = grams per litre; CRP = C-reactive protein; HDL-C = high density lipoprotein cholesterol; kJ = kilojoules; LDL-C = low density lipoprotein cholesterol; mmol.L^{-1} = millimole per litre; mmHG = millimetre of mercury; MUFA = monounsaturated fatty acids; PUFA = polyunsaturated fatty acids; SBP = systolic blood pressure; SFA = saturated fatty acids; TC = total cholesterol; TE = total energy; TG = triglycerides; IU.L^{-1} = international units per litre; U.L^{-1} = units per litre; a p-value of <0.05 is considered significant and is indicated in red text in the table.

Table 4.1 summarises the baseline characteristics of the continuous variables for the population under investigation. Comparisons were made between the rural and urban, as well as between the men and women subgroups, as discussed in Section 3.8. Significant differences were observed and indicated in Table 4.1 ($p \leq 0.05$).

As mentioned, CRP concentrations can indicate risk of a future CVD event, with concentrations of greater than 3 $mg.L^{-1}$ indicating high risk, 1-3 $mg.L^{-1}$ indicating intermediate risk and less than 1 $mg.L^{-1}$ indicating low risk (Ledue & Rifai, 2003). The CRP concentrations did not differ significantly between the rural and urban groups or between the genders, but it is important to note that the CRP concentrations (8.15-8.67 $mg.L^{-1}$) are higher than the recommended cut-off limits in all these groups. CRP concentrations tend to be higher in women than in men (Khera *et al*., 2005), and similar results were observed in this population, although to a lesser extent. As expected, the exclusion of subjects with CRP concentrations of above 10 $mg.L^{-1}$ influenced the mean of the population's CRP concentrations, as well as the different groups (men, women, rural and urban). After exclusion of those with acute inflammation, the mean for the whole group (*i.e*. 2.88 $mg.L^{-1}$) remained higher than recommended and can be related to an intermediate risk of the development of CVD. There were still no significant differences between the CRP concentrations of either the rural *versus* urban groups ($p = 0.22$) or the men *versus* women ($p = 0.27$) after removal of individuals with an CRP>10 $mg.L^{-1}$ (see Addendum A).

Weight differed significantly (<0.01) between the rural and urban group with the rural group weighing less on average than the urban individuals. This could be expected as the kilojoule (kJ) of the urban group is more than the rural group. Other factors could be at play here such as exercise differences in the different location groups. Because of the difference in weight one would expect a significant difference (0.01) in BMI values between the two groups. Height, weight and BMI differed significantly between the men and women, with the women being shorter of stature and weighing more than the men. This will then result in the significant difference between the BMI's of the two groups. The women in this population had a higher BMI (26.79 $kg.m^{-2}$) than the men (20.78 $kg.m^{-2}$).

The blood lipid profile determinants of the rural and urban groups did not differ significantly, except for triglyceride concentrations ($p < 0.01$). Even though the values differed significantly, both groups' triglyceride values were within the recommended range, *i.e*. < 2.82 $mmol.L^{-1}$ (Porter & Kaplan, 2011). The remaining components of the blood lipid

profile also presented with the healthy expected values in the rural and urban groups, as defined by Porter and Kaplan (2011), *i.e*. LDL-C ≤ 3.36 mmol.L^{-1}, HDL-C ≥ 1.04 mmol.L^{-1} and TC between 3.88-5.15 mmol.L^{-1}. For the men as well as for the women, all the components of the blood lipid profile were in accordance with the recommendations. The TC of the women in this population was, however, at borderline risk levels with the upper level of the range being higher than 5.15 mmol.L^{-1} (Porter & Kaplan, 2011). Significant differences were observed between the genders regarding the blood lipid profile, with women having the more less favourable profile compared to men in this population (although still within the recommended normal values). Overall, the lipid profile of this population fell within the healthy ranges.

The recommended daily kJ intakes for men and women were 8,400 kJ and 6,720 kJ, respectively (Vorster & Nel, 2001; Wolmarans & Oosthuizen, 2001). The kJ intake of the men and women in this study differed significantly, as well as between the rural and urban groups, as indicated in Table 4.1. The women had a higher kJ consumption (7,464 kJ) than the recommended 6,720 kJ. On average, the women in this population, as well as the urban individuals, consumed more kJ than recommended. This might explain the high level of obesity observed among the women in this population. More refined carbohydrates (more energy dense) were also consumed by the women, which could contribute to the large number of overweight individuals among the women as well, but this did not result in significantly different CRP concentrations. The urban group consumed a higher kJ diet than recommended (9,262 kJ), and this could most probably be ascribed to the fact that as urbanisation increases in South Africa, more black South Africans are exposed to a Westernised diet that includes more energy dense foods. This is due to the associated “nutrition transition” that occurs in response to urbanisation (Bourne *et al*., 1993), hence the increase in dietary intake.

Protein intake is recommended to be between 15% and 25% of the TE per day (Vorster & Nel, 2001; Wolmarans & Oosthuizen, 2001). The rural and urban groups, as well as the men and women, consumed less than the required percentage of protein per day. Food sources rich in protein are not freely available in some of the communities included in this study and are considered expensive. The participants mainly rely on plant sources to meet their protein requirements, which can be insufficient. Total carbohydrate intake (45-65% of TE) and total fat intake (20-35% of TE) in this cohort were within the normal ranges recommended for men and women (Vorster & Nel, 2001; Wolmarans & Oosthuizen, 2001), as well as for the rural and urban subgroups. As carbohydrates are

more freely available, one would expect the intakes to be within the recommended ranges. Total fibre intake also differed significantly between the groups and was below the suggested 28 g per day in the whole groups as well as all the subgroups investigated (Vorster & Nel, 2001). This could possibly be ascribed to the consumption of more refined carbohydrates than high-fibre products, again because of the availability and affordability of the former. Total sugar intake is recommended to be less than 10% of TE (Vorster & Nel, 2001; Wolmarans & Oosthuizen, 2001). All subgroups were within this range, but the urban group consumed on average almost double the amount of sugar than the rural group did. This could be attributed to the exposure to a more Westernised diet in the urban population, which is also a possible explanation for the difference in the percentage of fat consumed. Urban subjects consumed on average more total dietary fat than rural subjects. The suggested ranges are 20 - 30% fat as part of TE (Wolmarans & Oosthuizen, 2001) and all four groups were within this range. The different fatty acids were all within the recommended ranges (FAO, 2010) in this population *i.e.* MUFA (15-20%), PUFA (6-11%) and SFA (<10%). The overall dietary intake of this group is within the expected ranges, except for protein intake, most probably for the reasons discussed previously.

In conclusion with regard to the dietary data from this population, a detailed investigation of the PURE population's diet undertaken by Dolman (2012), reported that urban dwelling individuals consume a diet that was more likely to comply to the micronutrient requirements, since the variety of foods consumed increased. The rural individuals are at risk of developing several micronutrient deficiencies. Urban living individuals' diets also provided more fibre than that of the rural living individuals. In terms of dietary quality, the urban group was better off than the rural group. We are of the opinion that the benefits of the urban diet will change as the urban dwellers adopt a more westernised diet. It is also important to note that although the diets of the urban dwellers cannot yet be fully characterised as "westernised" it is clear that it is rapidly changing, when analysing the 2005 data. One can speculate that the diet will change even more or evolve into a fully characterised westernised diet in the next 10 years. Therefore, the individuals from the urban group had a more westernised diet than that of the rural group, but cannot yet be classified as a fully westernised diet pattern.

A systolic blood pressure value of between 120 and 139 mmHg and a diastolic blood pressure of between 80 – 89 mmHg are considered pre-hypertensive (Porter & Kaplan, 2011). All four groups investigated in this population were, therefore, pre-hypertensive. Urban men and women had a slightly higher systolic blood pressure and diastolic blood

pressure value. This could be due to the effect of urbanisation in this population and the adoption of an unhealthy lifestyle, increasing hypertensive risk.

Table 4.2 is a summary of the differences in CRP concentrations between the four groups outlined previously for the different categorical variables investigated. Significant differences were observed between the different categories in BMI and smoking status. There seems to be a U-shaped relationship between CRP concentrations and the BMI categories for the whole group, urban, rural and women groups. This emulates the relationship between BMI and mortality, which shows that both underweight and overweight present a higher risk of mortality (Zheng *et al*., 2011). The highest CRP concentrations are observed in the underweight and obese group, with the normal and overweight group being the lowest. This could possibly indicate that both too little and too much body fat results in increased levels of inflammation. Another school of thought could be that any illness associated with infection could lead to weight loss and result in a low BMI, while simultaneously causing increased inflammation due to the infection and therefore increasing CRP concentrations.

Table 4.2 CRP concentrations for categorical baseline characteristics for the total group and by gender and location groups

Categorical variables	Whole group CRP mean ($mg.L^{-1}$) ± SD (n)	P-value	Rural CRP mean ($mg.L^{-1}$) ± SD (n)	Urban CRP mean ($mg.L^{-1}$) ± SD (n)	p-value	Men CRP mean ($mg.L^{-1}$) ± SD (n)	Women CRP mean ($mg.L^{-1}$) ± SD (n)	p-value
BMI:								
< 18.5 (Underweight)	10.48 ± 15.86 (276)	<0.01	10.22 ± 16.28 (148)	10.78 ± 15.41 (128)	0.28	10.86 ± 15.86 (161)	9.96 ± 15.91 (115)	0.16
18.5-24.9 (Normal weight)	6.58 ± 11.82 (653)		6.89 ± 12.33 (352)	6.21 ± 11.20 (301)		7.04 ± 12.34 (326)	6.12 ± 11.28 (327)	
25-29.9 (Overweight)	6.81 ± 9.25 (256)		7.06 ± 9.84 (124)	6.56 ± 8.70 (132)		6.67 ± 8.38 (55)	6.84 ± 9.50 (201)	
≥ 30 (Obese)	11.72 ± 11.38 (317)		10.41 ± 9.37 (148)	12.86 ± 12.81 (169)		6.20 ± 4.41 (19)	12.07 ± 11.60 (298)	
HIV status:								
Positive	10.17 ± 15.52 (263)	0.22	6.98 ± 11.18 (66)	9.12 ± 13.96 (84)	0.58	7.15 ± 13.11 (52)	8.72 ± 12.69 (98)	0.74
Negative	8.07 ± 11.63 (1,317)		8.27 ± 12.39 (706)	8.60 ± 12.24 (726)		8.22 ± 13.28 (539)	8.57 ± 11.69 (893)	
Smoking:								
Previously used tobacco products	12.35 ± 16.70 (58)	0.02	17.32 ± 20.44 (26)	8.54 ± 11.85 (32)	0.04	13.78 ± 19.18 (37)	10.20 ± 11.16 (21)	0.44
Currently use tobacco products	8.55 ± 12.94 (828)		8.04 ± 12.55 (397)	9.05 ± 13.28 (431)		8.36 ± 13.20 (354)	8.71 ± 12.75 (474)	
Never used tobacco products	8.07 ± 11.35 (693)		7.75 ± 11.27 (350)	8.32 ± 11.43 (343)		6.82 ± 11.87 (198)	8.52 ± 11.11 (495)	

BMI = body mass index; CRP = C-reactive protein; HIV = human immunodeficiency virus; n = number of individuals; SD = standard deviation; BMI categories were adapted from the WHO (2002) recommendations; a p-value of < 0.05 is considered as significant and is indicated in red.

No significant differences were observed between HIV positive and HIV negative individuals. However, in a subset of the PURE population, HIV positive and negative subjects matched according to age, gender, BMI and locality (urban and rural) differed significantly in terms of CRP concentrations between the HIV positive and negative group (Fourie *et al*., 2010). This is in line with findings determined in other studies (Hsue *et al*., 2004; Dolan *et al*., 2005).

The CRP concentrations in the different tobacco use categories differed significantly for the whole group ($p = 0.02$) as well as for the rural *versus* urban groups ($p = 0.04$). In the whole group, the rural group as well as in the men and women, it was observed that CRP concentrations were highest in subjects that reported having "previously used tobacco products". It seems that cessation of tobacco use has no beneficial effect on CRP concentrations, as CRP was significantly higher in this group. From the current study the reason for this is unclear and further research that focuses on tobacco use and CRP is necessary. Contrastingly to this finding, Stępień *et al*. (2011) reported no significant difference in CRP concentrations between current, former or non-smokers.

4.2 VARIABLES CORRELATING WITH CRP CONCENTRATIONS

As mentioned in Section 3.8, confounding variables for CRP were identified by means of the Spearman Rank test. Continuous variables with large correlation values, *i.e*. $r > 0.2$ (Steyn, 2002), with CRP concentrations were considered to be possible confounders *i.e*. BMI and fibrinogen. These two variables both had the highest correlation value (Table 4.3). Systolic blood pressure, diastolic blood pressure, age, LDL-C, HDL-C, triglycerides, total sugar intake and MUFA intake also correlated significantly with CRP. However, the correlations were smaller than 0.15 and thus not of practical significance. The determination of significant results despite very small r-values could be due to the large sample size of the present study (Hair & Anderson., 2010).

Table 4.3 Correlations between identified variables and CRP concentrations in the PURE population

Variable	R	p-value	Variable	R	p-value
SBP	0.056	0.025	Total energy intake	0.026	0.300
DBP	0.083	0.001	Total protein intake	0.030	0.236
Age	0.119	0.001	Total CHO intake	0.027	0.284
Cholesterol	0.044	0.081	Total fibre intake	0.011	0.672
LDL-C	0.085	0.001	Total sugar intake	0.069	0.007
HDL-C	-0.145	0.001	Total fat intake	0.038	0.137
TG	0.146	0.001	MUFA	0.052	0.042
Fibrinogen	0.417	0.001	PUFA	0.041	0.109
BMI	0.222	0.001	SFA	0.043	0.095

CHO = carbohydrates; BMI = body mass index; DBP = diastolic blood pressure; HDL-C = high density lipoprotein cholesterol; LDL-C = low density lipoprotein cholesterol; MUFA = monounsaturated fatty acids; PUFA = polyunsaturated fatty acids; SBP = systolic blood pressure; SFA = saturated fatty acids; TC = total cholesterol; TG = triglyceride; the total number of individuals were 1587.

The variables identified as possible confounders were adjusted for during the statistical analyses. In the following sections, however, the determination of the genetic variations within the *CRP* gene is discussed.

4.3 DNA ISOLATION

As described in Section 3.7.1, two DNA isolation methods were used in order to isolate gDNA from either EDTA or citrate-treated buffy coat. The reason for using the two different DNA isolation methods was that the DNA yield of 350 of the samples was not sufficient after isolation using the Flexigene kit (most likely because of low lymphocyte concentrations). Thus, the Maxwell® 16 System was used, as it is a more efficient automated purification system. The Maxwell® 16 System uses paramagnetic particle technology (Section 3.7.1) for faster separation of DNA, which makes it more effective at isolating the DNA. Purity and yield of the isolated gDNA was determined as discussed in Section 3.7.1. DNA isolated using the FlexiGene™ DNA extraction kit (QIAGEN) had yields that ranged from 152-1,424 ng.μL^{-1} and A_{260}/A_{280} ratios between 1.74 and 1.88. DNA obtained *via* the Maxwell® 16 System had yields that ranged from 82-232 ng.μL^{-1} and A_{260}/A_{280} ratios ranging from 1.70 - 1.87. The reason for the seemingly lower yields obtained through the Maxwell® 16 System can be ascribed to the fact that the second round of DNA isolation was performed on samples that had previously given low yields, therefore these samples most probably had reduced lymphocyte concentrations. Therefore, the yields of the two methods should not be used to compare the efficiency of the different isolation methods. A final working DNA dilution of 50 ng.μL^{-1} was prepared,

which was used for the genotyping analysis *via* the BeadXpress® Reader, as discussed in Section 4.6.

4.4 POLYMERASE CHAIN REACTION AMPLIFICATION AND AUTOMATED SEQUENCING RESULTS

Thirty individuals were randomly selected from the PURE study population (Section 3.7.2), in which the four regions of the *CRP* gene were sequenced and polymorphisms present within these regions were identified. The main purpose for sequencing these 30 individuals was to determine novel polymorphisms in the *CRP* gene in this population, as well as to determine whether the polymorphisms reported in the literature (in other ethnic populations) are also present in the black South African population under investigation. Primers were designed *via* the Primer-BLAST program, which is a general-purpose target-specific PCR primer design tool that offers a high level of sensitivity and usability (Ye *et al*., 2012).

A few conditions were satisfied in order to design the specific primers successfully for the amplification of the four regions. These conditions included that the primer length was between 18-22 bp, which was long enough for adequate specificity, but short enough for the primers to bind easily. Secondly, the primer melting temperature (T_m) was in the range of 52 - 65°C. This T_m produces the best results, whereas a primer with a T_m higher than 65°C has a tendency to undergo secondary annealing. Both the forward and reverse primers had to have similar T_m and a balanced G/C content (40-60%) to avoid mispriming. More than three G's or C's in the last five bases at the 3' end of the primer were avoided, again to avoid mispriming. Hairpin formation was avoided by ensuring that the primer adhered to a delta (Δ) G of greater than -2 kcal.mol^{-1} at the 3' end. Optimally a 3' end self-dimer with a ΔG of > -5 kcal.mol^{-1} or an internal self-dimer with a ΔG of > -6 kcal.mol^{-1} was accepted. To prevent cross-dimer formation a ΔG of > -5 kcal.mol^{-1} at the 3' end and an internal cross-dimer with a ΔG of > -6 kcal.mol^{-1} were allowed (Ye *et al*., 2012).

Specific protocols (Section 3.7.2) for each primer set were followed, as indicated in Table 3.1, in order to amplify the specific region of the *CRP* gene under investigation. A total of four regions were amplified (promotor, intron, exon 2 and the downstream region). The different nucleotide numbers to which the primers bind in relation to the reference sequence (Genbank accession AF449713) are indicated in Table 4.4.

Table 4.4 Binding sites of the *CRP* region-specific primer sets used in this study

Primer	Region	Nucleotide number*	Primer	Region	Nucleotide number*
CRP 1F	Promotor	707	*CRP* 3F	Exon 2	2483
CRP 1R	Promotor	1667	*CRP* 3R	Exon 2	3406
CRP 2F	Intron	1587	*CRP* 4F	Exon 2	3189
CRP 2R	Intron	2566	*CRP* 4R	Downstream	4148

F = forward; R = reverse; *nucleotide numbering according to the reference sequence (Genbank accession # AF449713).

A positive control sample (containing control DNA) and a negative control sample (containing double distilled water) were included in each run to ensure that the reaction procedure was successful and that there was no contamination. Once the optimal T_a had been determined, by means of a temperature gradient for each primer, the samples were amplified using a standardised PCR reaction, as indicated in Section 3.7.2, in order to amplify the fragment of interest. The success of the PCR amplification was determined by electrophoresing the amplified product on a 2% agarose gel. If successful, the PCR product underwent purification using the NucleoFast® 96 PCR Clean-Up Kits from Machery-Nagel (Section 3.7.2), after which automated cycle sequencing of each region was undertaken, as described in Section 3.7.2, at the Central Analytical Facility, Stellenbosch University.

4.5 POLYMORPHISMS IDENTIFIED IN THE *CRP* GENE

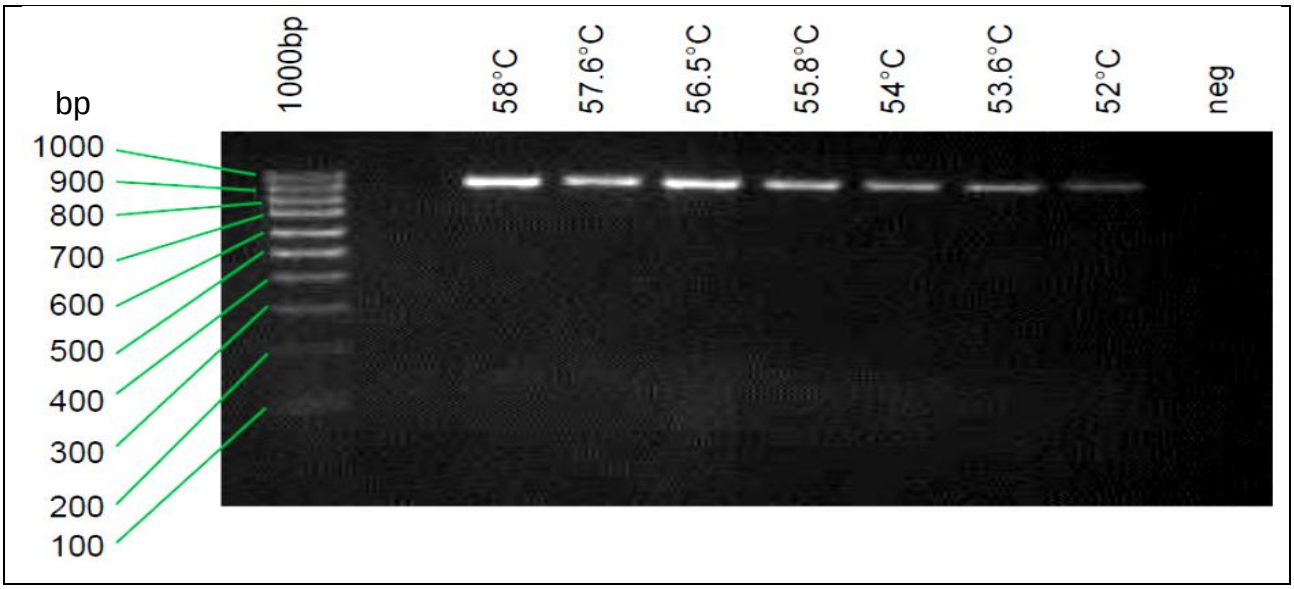
After all the processes had been followed as described in Section 3.7.2 and Section 4.3, the sequencing electropherogram results were obtained from the Central Analytical Facility. The electropherograms were inspected and aligned *via* the BioEdit program version 7.1.3.0 (Hall, 1999), according to the *CRP* gene reference sequence (Genbank accession # AF449713) by making use of the “alignment” function of the BioEdit program. The alignment was also checked afterwards to ensure accurate and correct alignment with the reference sequence. As mentioned, four regions were amplified and are discussed subsequently.

4.5.1 Amplification of region one in the *CRP* gene

For amplification of region one within the *CRP* gene, the protocol mentioned in Section 3.7.2 was followed and the primers indicated in Table 3.5 were used. Region one is 980 bp long and located in the promoter region of the *CRP* gene. SNP rs3093058 and rs3093062 were located in this region. A temperature gradient was conducted for region one of the *CRP* gene and a T_a of 58°C was determined to be optimal for amplification.

Figure 4.1 is a photographic representation of the temperature gradient analysis. During electrophoresis a 1,000 bp marker (Bioline Hyperladder ™ IV) was used to estimate the fragment size. No amplification product was seen in the negative control, thus ensuring the absence of contamination.

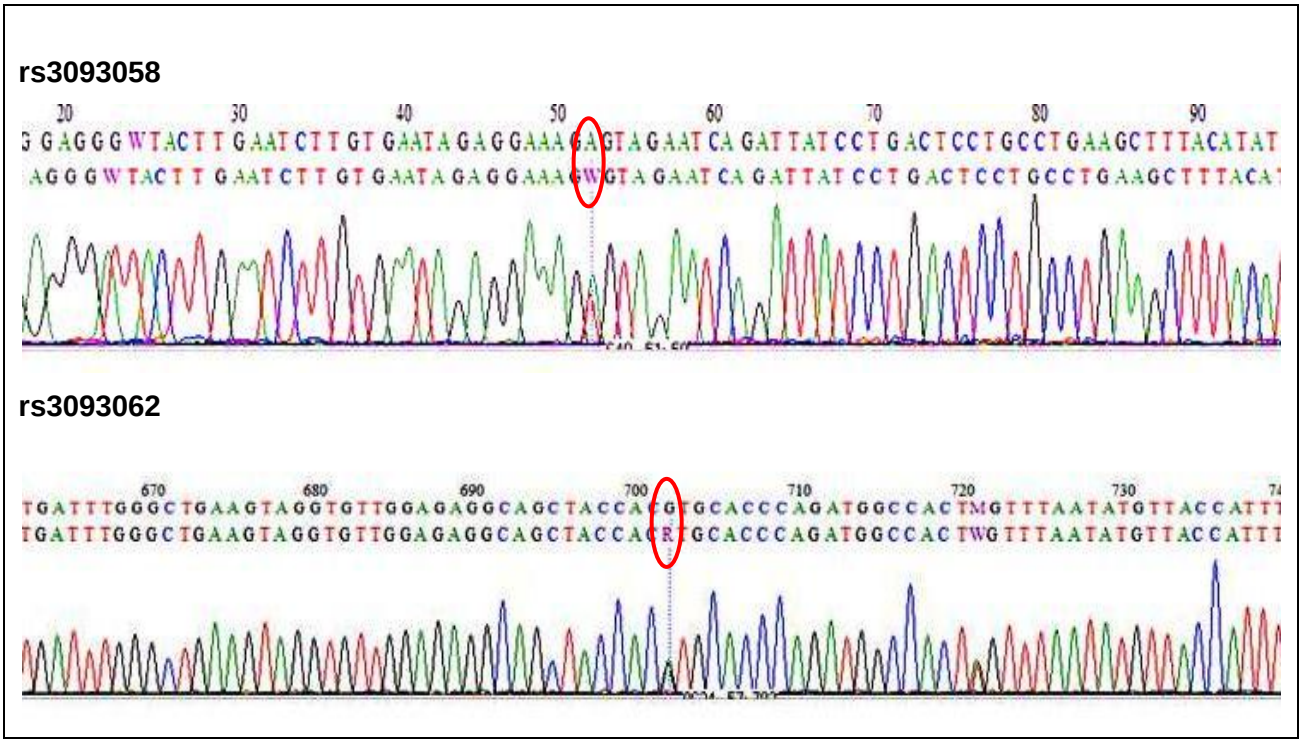
Figure 4.1 Photographic representation of the temperature gradient of region one of the *CRP* gene



2% agarose gel electrophoresed at 100 volts for 60 minutes in 1% Tris/Borate/EDTA (TBE) buffer; 1000 bp = 1000 base pair molecular weight marker (Bioline Hyperladder ™ IV); neg = negative control; °C = degrees centigrade

Following successful amplification, the 30 participants analysed were sequenced as described in Section 3.7.1. Following alignment of the generated electropherograms, it was determined that in this region, two polymorphisms were present (rs3093058 and rs3093062). The rs numbers of the identified polymorphism were identified using the UCSC website, as mentioned in Section 3.7.1, and are indicated in Figure 4.2.

Figure 4.2 Representative electropherogram of the two SNPs identified in region one of the *CRP* gene

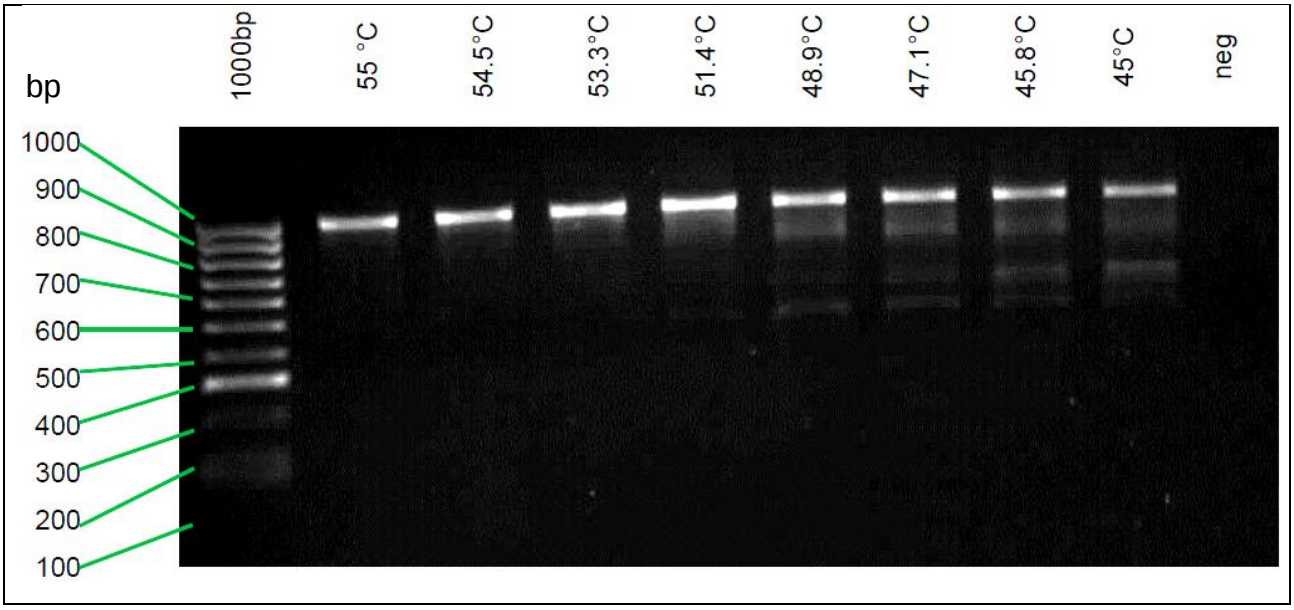


A = adenine; C = cytosine; G = guanine; rs = reference sequence; T = thymine; nucleotides circled in red = polymorphisms

4.5.2 Amplification of region two in the *CRP* gene

After a temperature gradient had been conducted on the primer set of region two of the *CRP* gene, it was determined that the optimal T_a for this region was 55°C (Figure 4.3). Protocols were followed, as mentioned in Section 3.7.1, using the primers indicated in Table 3.5. Region two was determined to be 997 bp long and was located in the intron area of the *CRP* gene. The negative control gave no amplification, therefore no contamination was present. Figure 4.3 is a photographic representation of the temperature gradient analysis.

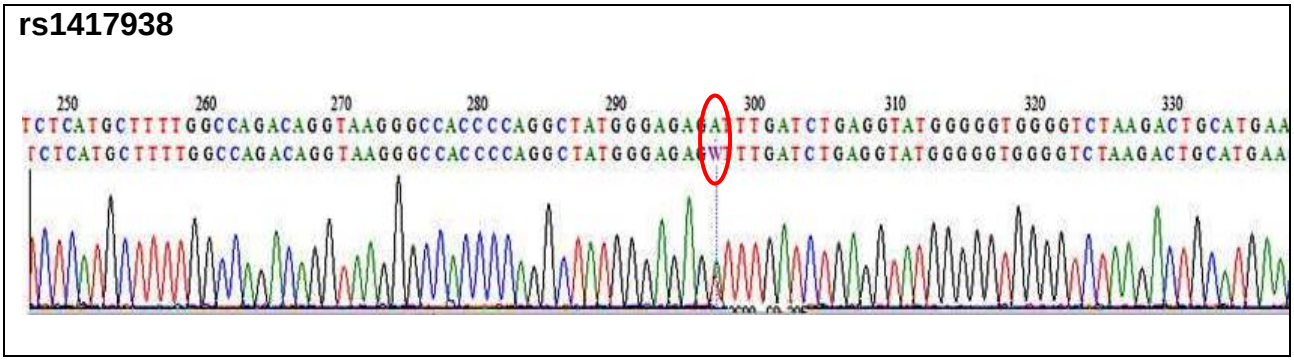
Figure 4.3 Photographic representation of the temperature gradient of region two of the *CRP* gene



2% agarose gel electrophoresed at 100 volts for 60 minutes in 1% Tris/Borate/EDTA (TBE) buffer; 1000 bp = 1000 base pair molecular weight marker (Bioline Hyperladder ™ IV); neg = negative control; °C = degrees centigrade.

Following alignment of the 30 individuals, it was determined that only one existing polymorphism (rs1417938) was present in the region of primer 2 (Figure 4.4) in the 30 individuals sequenced. One of the 30 sequenced subjects harbouring the heterozygote genotype is indicated in Figure 4.4.

Figure 4.4 Representative electropherogram of the one SNP identified in region two of the *CRP* gene



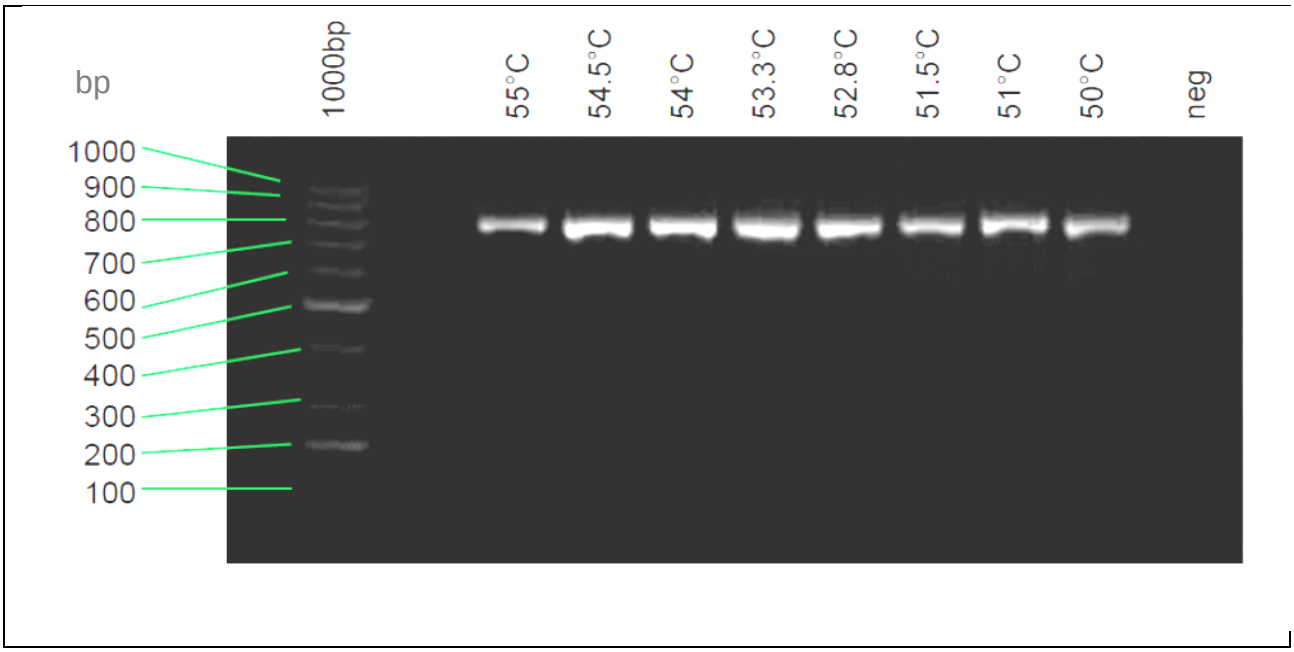
A = adenine; C = cytosine; G = guanine; rs = reference sequence; T = thymine; nucleotides circled in red = polymorphism.

4.5.3 Amplification of region three in the *CRP* gene

Region three of the *CRP* gene was located in the second exonic region and was determined to be 943 bp long. After a temperature gradient had been conducted, following the protocol in Section 3.7.2, a T_a of 54.5°C was established to be the optimal temperature. Figure 4.5 is a photographic representation of the results of the temperature gradient. As in all the regions, a 1,000 bp marker (Bioline Hyperladder ™ IV) was used to

estimate the fragment size. As seen in Figure 4.5, the negative control indicated no amplification, suggesting the absence of contamination.

Figure 4.5 Photographic representation of the temperature gradient of region three of the *CRP* gene



2% agarose gel electrophoresed at 100 volts for 60 minutes in 1% Tris/Borate/EDTA (TBE) buffer; 1000 bp = 1000 base pair molecular weight marker (Bioline Hyperladder ™ IV); neg = negative control; °C = degrees centigrade.

After successful amplification of the 30 sequenced subjects, alignment was conducted and used to determine two previously reported polymorphisms (rs1800947 and rs1130864), which were identified using the BioEdit program (Figure 4.6). The polymorphism is indicated in red in Figure 4.6.

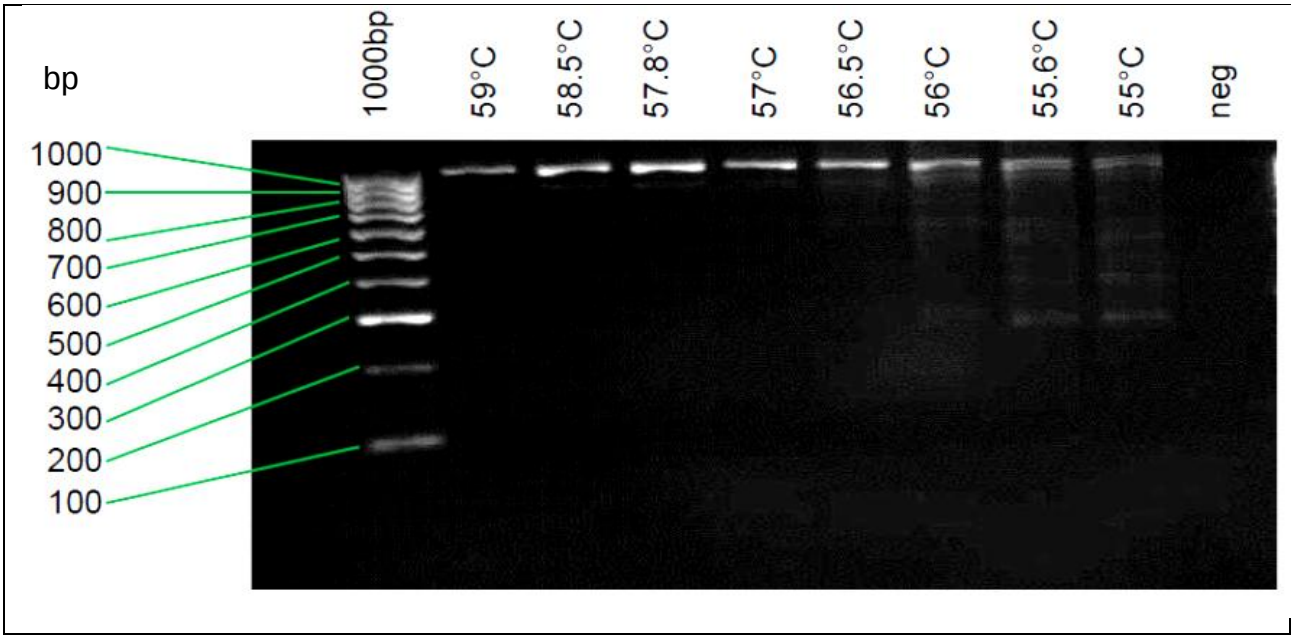
Figure 4.6 Representative electropherogram of the two SNPs identified in region three of the *CRP* gene

A = adenine; C = cytosine; G = guanine; rs = reference sequence; T = thymine; nucleotides circled in red = polymorphisms.

4.5.4 Amplification of region four in the *CRP* gene

After following the protocol mentioned in Section 3.7.2 and using the primers as tabulated in Table 3.5, a temperature gradient was conducted and it was determined that the optimal T_a was 59°C for the amplification of region 4 (Figure 4.7). Region 4 was 979 bp long and was partly located in exon 2 and the downstream region of the *CRP* gene. Figure 4.7 is a photographic representation of the temperature gradient analysis. The negative control gave no amplification, indicating that no contamination was present.

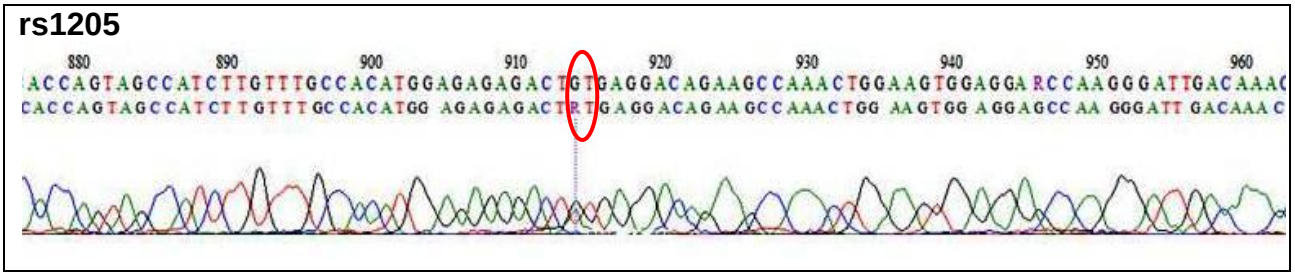
Figure 4.7 Photographic representation of the temperature gradient of region four of the *CRP* gene



2% agarose gel electrophoresed at 100 volts for 60 minutes in 1% Tris/Borate/EDTA (TBE) buffer; 1000 bp = 1000 base pair molecular weight marker (Bioline Hyperladder ™ IV); neg = negative control; °C = degree centigrade

Following alignment of the 30 individuals by using the BioEdit program, only one existing polymorphism (rs1205) was determined in region 4. Figure 4.8 is a representation of rs1205 identified in the current population. Although the electropherogram at this position is difficult to discern, this SNP was included, since it is a well-reported SNP in the CRP literature and the BeadXpress® platform enabled accurate determination thereof.

Figure 4.8 Representative electropherogram of the one SNP identified in region four of the *CRP* gene



A = adenine; C = cytosine; G = guanine; rs = reference sequence; T = thymine; nucleotides circled in red = polymorphism.

4.5.5 Polymorphism in the *CRP* gene reported in the literature

Additional *CRP* polymorphisms, which were not identified by sequencing, were determined by gathering information from electronic sources regarding *CRP* polymorphisms (Table 4.5). The following medical subject heading terms were used during the literature search: "c-reactive protein" and "cardiovascular disease"; "*CRP* polymorphisms" and "CVD"; "*CRP* SNPs"; "*CRP* polymorphisms" and "Africans". The following databases were used for this search: Medline, EBSCO host, PubMed, Web of science and Scopus. The SNPs that had a significant association (p value < 0.05) with CRP concentrations were then included. The identified SNPs had to comply with the BeadXpress® criteria as well, as discussed in Section 4.6.1. The final SNPs that were included from the literature are indicated in Table 4.5 with the reference article.

Table 4.5 *CRP* polymorphisms identified from the literature

rs number	Reference	rs number	Reference
rs2808630	Crawford *et al*. (2006)	rs2794520	Rhodes *et al*. (2008)
rs1341665	Wang *et al*. (2006)	rs7553007	Rhodes *et al*. (2008)
rs3093068	Hage and Szalai (2007)	rs2027471	Rhodes *et al*. (2008)

rs = reference sequence

4.6 BEADXPRESS® ANALYSIS OF THE SNPS IN THE *CRP* GENE

The participants included in this study were screened for the SNPs outlined in Table 3.2 *via* the BeadXpress® platform. As mentioned in Section 3.7.2, the BeadXpress® platform uses the Illumina® Golden Gate assay technology to genotype the different *CRP* polymorphisms under investigation (see Table 4.7). The following sections will describe the process of designing the assays (Section 4.6.1), as well as how the genotypic data was generated and analysed (Section 4.6.2).

4.6.1 Designing custom GoldenGate® genotyping assays

Section 3.7.2 describes the process that was followed in order to establish which polymorphisms could be included in this research project, as well as the manner in which the rs numbers were determined. The SNPs that were included in this investigation were submitted to the Illumina® company in the Identity (RSList) file format, which is used for the validation of reported alterations that are available in the current version of dbSNP in order to determine whether an assay for a specific SNP has been or can be designed using the GoldenGate® assay technology. After submission of the Identity file to Illumina®, a SNPScore file was generated using the ADT, which is described in Section 3.7.3. This SNPScore file provided an important set of informative metrics with regard to the designability of the different SNP assays, as indicated in Table 4.6, which in turn was used to create the final SNP order file to be used for the genotyping analysis. These metrics were used to select the SNP assays that were most likely to be successful in the design of the final BeadXpress® product.

Table 4.6 Description of the different information metrics generated by the ADT analysis

Informative metrics	Description
SNP_Score value	Indicates the expected success of the designed assay. The SNP_Score is given on a scale of 0.000 – 1.000, and in addition an SNP_Score of 1.1 indicates that the particular SNP is a GoldenGate® validated assay and that these assay oligonucleotides are available for use. The higher the SNP_Score the greater the probability of a successful assay design.
Failure_Codes: Critical failures (undesignable):	
101	Flanking sequence is too short
102	SNP or sequencing formatting error • Space in submitted sequence • More than one set of brackets in sequence • Missing bracket around SNP • SNP alleles not separated by "/"
103	TOP/BOTTOM strand cannot be determined because of low sequence complexity
104	SNP is not appropriate for Illumina® platform. Possible causes: • Tri- or quad-allelic SNP • Insertion or deletion polymorphism • SNP contains characters other than A, G, C and T
105	SNP is located in the mitochondrial genome
106	Degenerate nucleotides are in assay design region
Warnings (Designable):	
301	SNP is in duplicated/repetitive region
302	Polymorphism is outside assay limits
340	Another SNP in the list is closer than 61 nucleotides away
399	Multiple contributing issues
Validation_status:	
GoldenGate validated	SNP has previously been designed and has successfully generated polymorphic results on the Illumina® platform.
Two-hit or HapMap validated	Both alleles of the SNP have been seen in two independent methods and populations, or have been validated by the HapMap project.
Non-validated	SNP has been seen in only one method or population. Even if it has a high design score, there is an increased chance that it is monomorphic
Unknown	SNP is not reported within Illumina's® database based on SNP name

T = thymine; A = adenine; C = cytosine; G = guanine; SNP = single nucleotide polymorphisms

Table 4.7 is a summary of the 15 SNPs with a reported rs-number (excluding the novel SNP) that were initially submitted for BeadXpress® assay design *via* the ADT software. After careful consideration of all the aforementioned information metrics, two SNPs, *i.e*. rs3093061 and rs3093066, were excluded from the analysis because of their designability ranks and failure codes (Table 4.6).

Table 4.7 Initial 15 SNPs with the results from the SNPScore file

Locus Name	Source	SNP Score	Designability rank	Failure Codes	Validation Bin
rs3093058	dbSNP	0.879	1	--	OneKGenome Validated
rs3093059	dbSNP	0.831	1	340	OneKGenome Validated
rs3093061	dbSNP	0.15	0	399	HapMap Validated
rs3093062	dbSNP	0.736	1	--	OneKGenomeValidated
rs1800947	dbSNP	1.1	1	--	GoldenGate Validated
rs3093066	dbSNP	0.756	1	340	OneKGenome Validated
rs1130864	dbSNP	1.1	1	340	GoldenGate Validated
rs1205	dbSNP	1.1	1	--	GoldenGate Validated
rs1417938	dbSNP	0.801	1	--	HapMap Validated
rs2808630	dbSNP	1.1	1	--	GoldenGate Validated
rs1341665	dbSNP	1.1	1	--	GoldenGate Validated
rs2794521	dbSNP	0.431	0.5	340	HapMap Validated
rs3093068	dbSNP	0.476	0.5	--	OneKGenome Validated
rs2794520	dbSNP	0.908	1	--	HapMap Validated
rs7553007	dbSNP	1.1	1	--	GoldenGate Validated

rs = reference sequence; SNP = single nucleotide polymorphism; dbSNP version 135 were used during these analysis

A final SNPScore file was compiled and the SNPs mentioned in Table 3.2 were then sent off for final assay design and analysis. Therefore, the SNPs that were analysed included the remaining 13 SNPs, as well as a novel SNP.

4.6.2 Analysing GoldenGate® genotyping data

After the assay design had been performed and the final assay design had been generated by Illumina® (Section 3.7.2), the BeadXpress® analyses were performed at the NHLS at the University of the Witwatersrand, Johannesburg. Once the BeadXpress® assay had been set up, a step-by-step process was followed in order to generate the genotyping results of the SNPs investigated, as outlined in Section 3.7.2. Random samples of selected clusters in certain SNPs were sequenced (similar process as described in Section 3.7.2) to determine if the genotypic data of the BeadXpress® correlated with that of the sequencing results. All the samples which were sequenced, matched with the data generated *via* the BeadXpress® (data not presented) except for the novel SNP and rs2794521 which was then excluded from the remaining analyses, therefore, only 12 SNPs remained that were included statistically analysed.

The GoldenGate® Genotyping Assay included 48 assay controls, which offer a high level of assurance and the ability to troubleshoot errors such as PCR and primer hybridisation failures. Quality control of the results began with an overall evaluation of the assay

performance and determination of which samples, if any, required reprocessing or removal. In order to assess the overall performance of the samples, the reagents, the equipment and the BeadChips, various internal controls were included in each GoldenGate® assay. These controls included sample-dependent and sample-independent controls, as well as controls that could indicate contamination when present. Table 4.8 lists the IllumiCode sequence IDs of the different controls included in the assay, which are beads that are included in the analysis, along with a description and the expected outcome for each. The U3 and U5 match, which is frequently referred to, represents the Cy3 and Cy5 fluorescent channels, respectively. Therefore, a U3 match indicates that in order for the control to be successful, it needs to give a signal in the Cy3 channel.

Table 4.8 IllumiCode Sequence IDs used as controls and expected outcomes

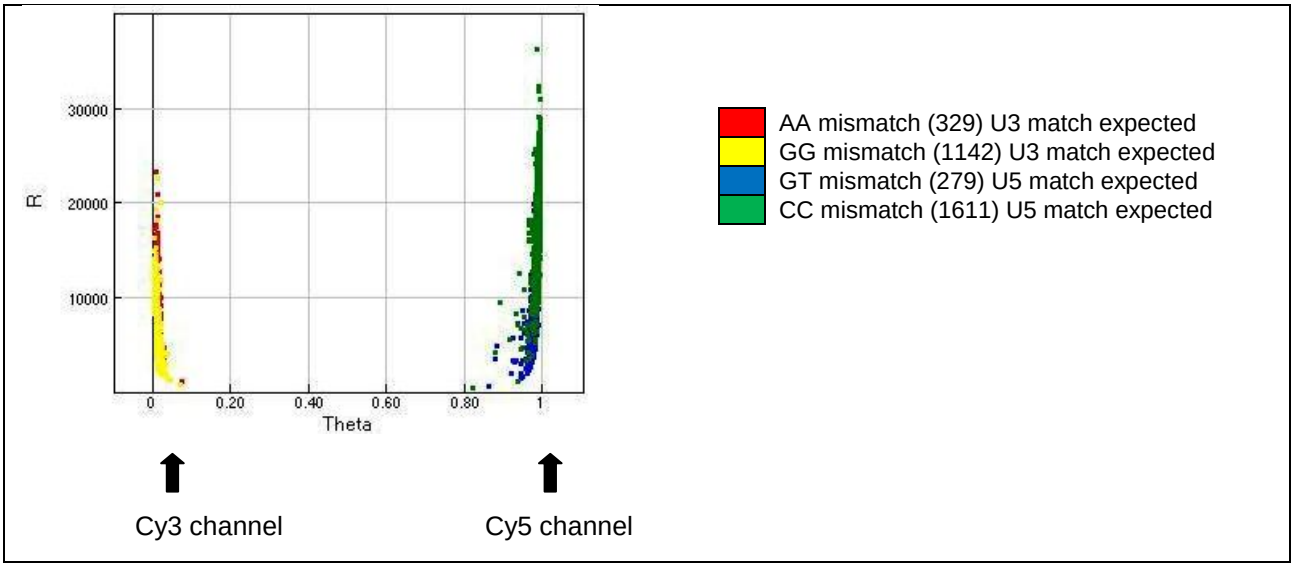
IllumiCode sequence ID	Description	Expected outcome
329	AA mismatch	U3 match
1611	CC mismatch	U5 match
1142	GG mismatch	U3 match
279	GT mismatch	U5 match
1742	High AT (31% GC)	U3 match
4824	High GC (62% GC)	U5 match
658	15-bp gap	U3 and U5 match
962	First hybridisation controls, 42/57 T_m	U5 match
1209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

a = adenine; bp = base pairs; c = cytosine; g = guanine; t = thymine; U5 = refers to the Cy5 channel; U3 = refers to the Cy3 channel; T_m = melting temperature; T_a = annealing temperature

Firstly, the allele-specific extension controls measured the extension efficiency of the properly matched ASO *versus* mismatched ASO. These controls test for A-A, C-C, G-G and G-T mismatches corresponding with IllumiCode sequence IDs 329, 1611, 1142 and 279, respectively (Table 4.8). It is expected that sequence ID 329 (indicated in red) and 1142 (indicated in yellow) should give a signal that is predominately in the Cy3 channel (U3 match), while the sequence IDs 279 (indicated in blue) and 1611 (indicated in green) should give a signal predominately in the Cy5 channel (U5 match). This is what is seen in the allele-specific extension control determined during this investigation (Figure 4.9). The “mismatches” have to do with the complementarity of the template DNA with the

oligonucleotide. These mismatched oligonucleotides would, therefore, not bind to the template DNA and a signal in the opposite channel would be generated.

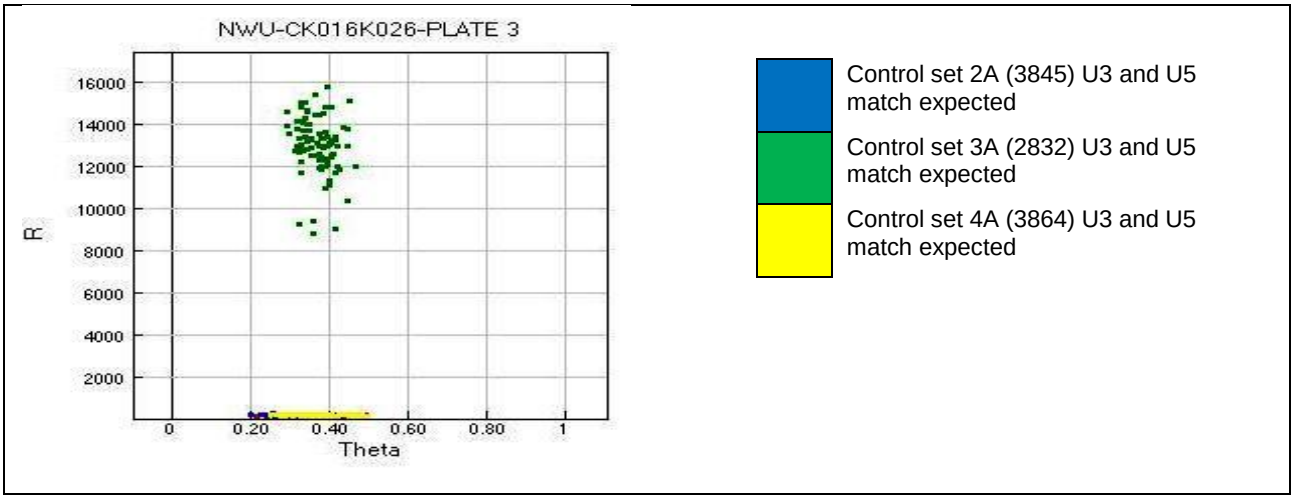
Figure 4.9 Allele-specific extension control of the BeadXpress® analysis



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

Contamination controls are also built into the assay to ensure and assess that no contamination is present in each plate. Figure 4.10 is a representation of the contamination control dashboard of a single plate. The lack of contamination is indicated by the amplification of only one colour (green). This figure is representative of the graphs for all the plates and, therefore, each plate had no carry-over contamination.

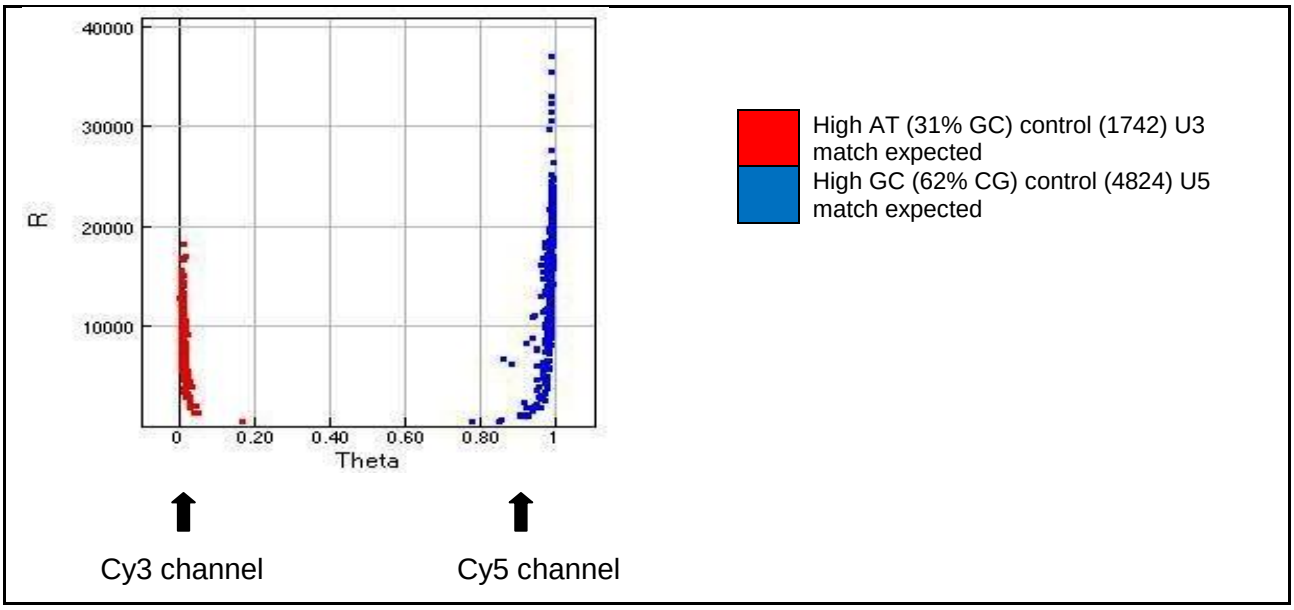
Figure 4.10 Contamination control of plate 3



The PCR uniformity control tests the PCR amplification efficiency for high AT and high GC-rich regions of the DNA. IllumiCode sequence ID 1742 evaluates the amplification efficiency for high AT-rich regions (31% GC) and should result in a high Cy3 channel

signal. IllumiCode sequence ID 4824 amplifies over a high GC-rich (62% GC) region and should result in a high Cy5 channel signal. Therefore, the controls are expected to display intensity in the Cy3 channels (red) and in the Cy5 channel (blue) respectively. This is indeed the pattern that is seen the controls of the current analysis (Figure 4.11).

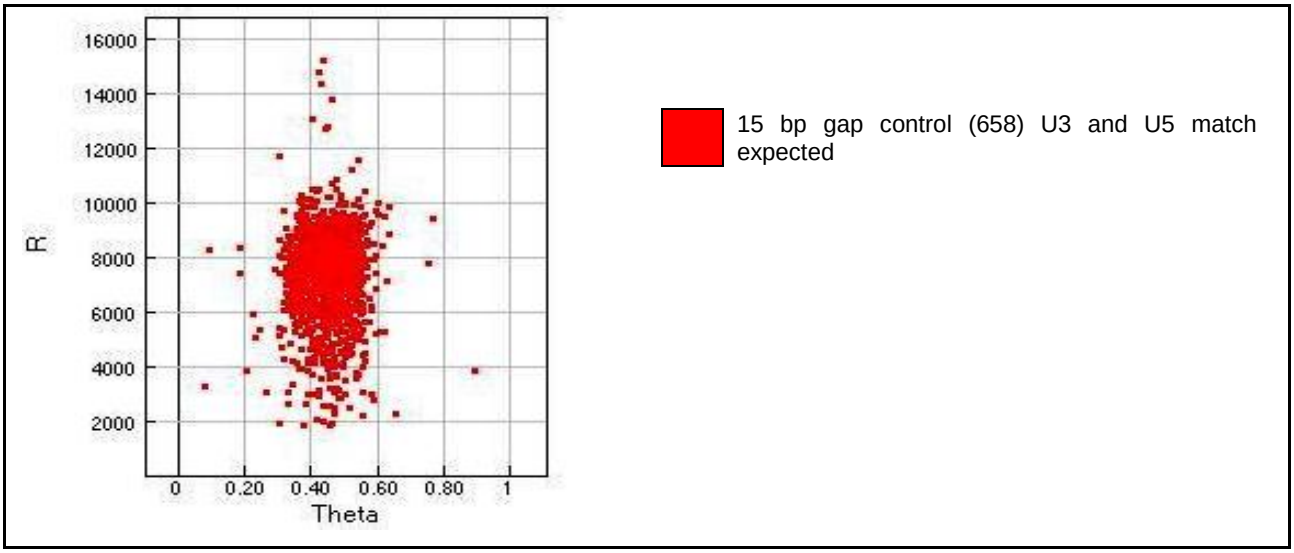
Figure 4.11 Polymerase chain reaction uniformity controls



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

The extension gap control (IllumiCode sequence ID 658) tests for the efficiency of extending the 15 bases from the 3' end of the ASO to the 5' end of the LSO. Signals should be seen from both the Cy3 and Cy5 fluorophores and should be somewhere in the heterozygote theta range. This trend is seen in the controls throughout the investigation (Figure 4.12).

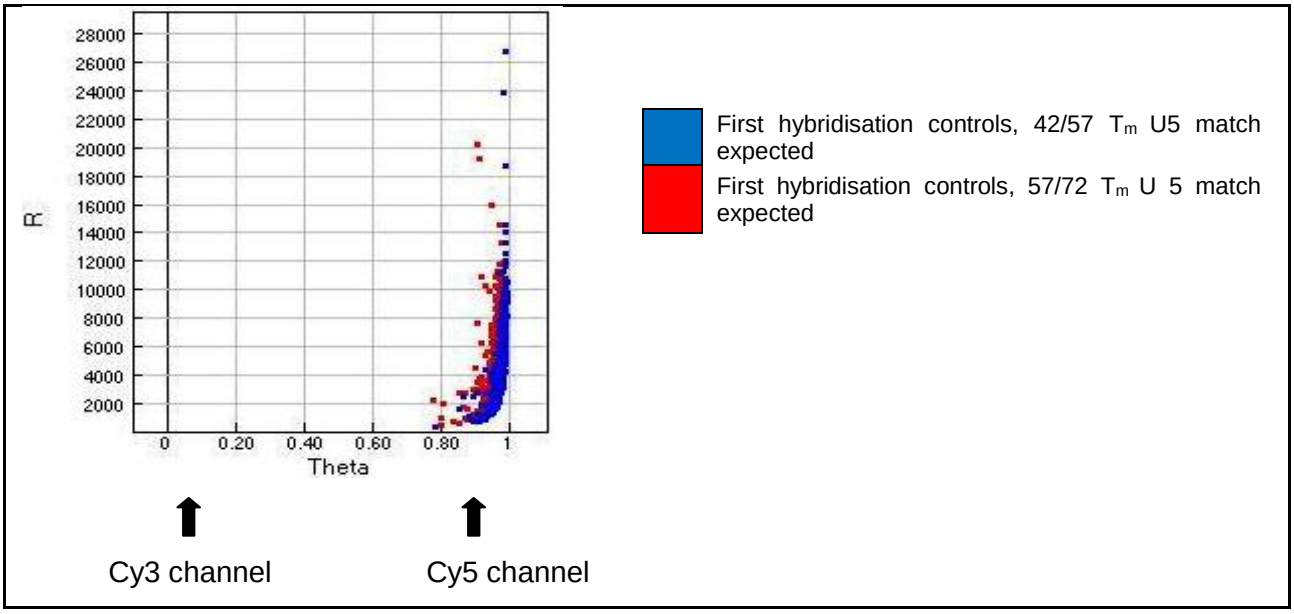
Figure 4.12 Extension gap control



bp = base pair; U5 = refers to the Cy5 channel; U3 = refers to the Cy3 channel

The first hybridisation control measures how well the two ASOs anneal to a specific target at different T_a. Both IllumiCode sequence ID 962 and 1209 should result in a Cy5 match. As seen in Figure 4.13, both the mentioned IllumiCode sequence IDs (red and blue) signals are in the Cy5 channel, illustrating that this control was successful.

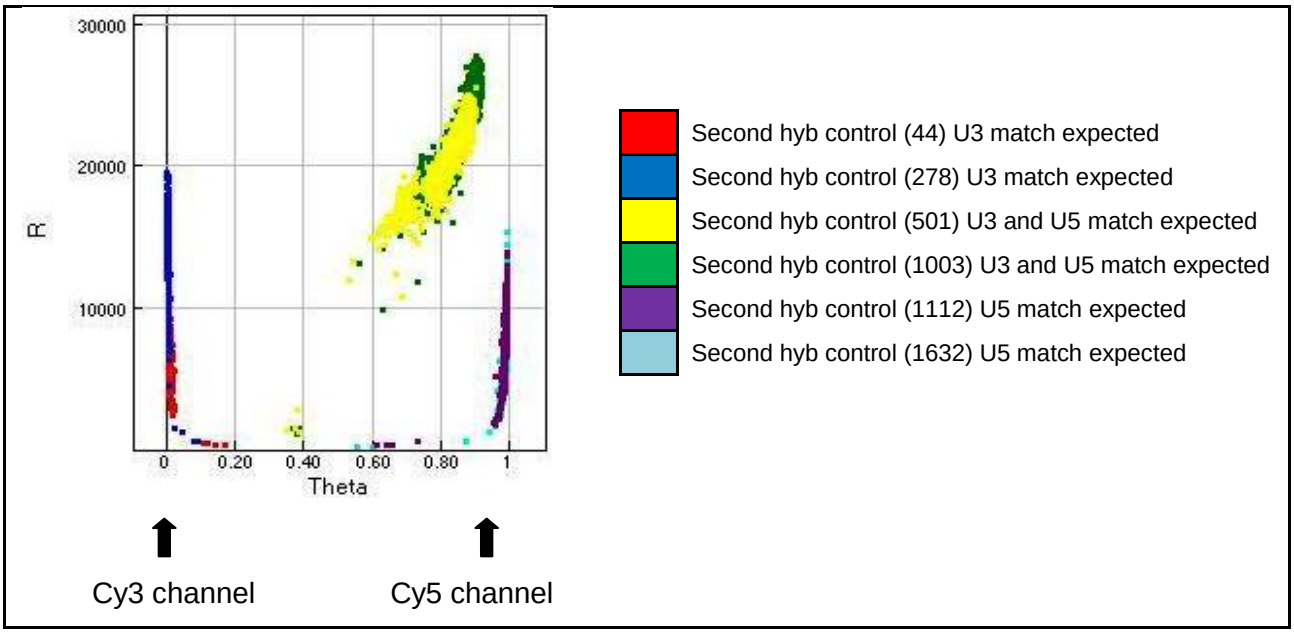
Figure 4.13 First hybridisation controls



T_m = melting temperature; U5 = refers to the Cy5 channel; U3 = refers to the Cy3 channel

The second hybridisation controls test the hybridisation of the single-stranded assay products to the IllumiCode sequences on the array beads. IllumiCode sequence IDs 44 and 278 should result in a Cy3 signal only (red and blue), sequence IDs 1112 and 1632 should result in only a Cy5 signal (cyan and purple), and sequence IDs 501 and 1003 should have signals contributed by both Cy3 and Cy5 (yellow and green). As illustrated in Figure 4.14, all the mentioned IllumiCode sequence IDs signals are in accordance with what they should signal, therefore this control is in line with what is expected.

Figure 4.14 Second hybridisation controls



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

The quality of the SNP clusters, which are used to determine the actual genotype of a participant, is based on the so-called GenCall score, which is a quality metric that indicates the reliability of each genotype call. This GenCall Score is a value between 0 and 1 that is assigned to each genotype called by the GenomeStudio® program. Genotypes with lower GenCall scores are situated further from the centre of a cluster and have lower reliability. Each GenCall score is calculated using information from the clustering of samples and is based on four characteristics of a cluster, namely angle, dispersion, overlap and intensity. The size of the shade call regions is defined by the GenCall score cut-off. A GenCall score below 0.5 was considered to be unreliable and, therefore all samples included in the analysis had to have a GenCall score of above 0.5. Random samples of selected clusters in certain SNPs were sequenced (similar process as described in Section 3.7.2) to determine if the genotypic data of the BeadXpress® correlated with that of the sequencing results. All the samples that were sequenced, matched the data generated *via* the BeadXpress® (data not presented). This confirmed that the clustering calls from the BeadXpress® were correct.

Another metric that provided information to be used to ensure that the data was of the highest quality, was the call rate. The call rate value represents the proportion of all the samples with call scores above the no call threshold at each locus. Values varied from 0-1. After reviewing all of the BeadXpress® data, it was concluded that all SNPs that had a call rate value of below 0.9 should be excluded from the analysis.

The mean normalised intensity (AB R Mean) of the heterozygote cluster is another metric that assists in identifying SNPs with low intensity data and has values increasing from 0. SNPs with an AB R mean below 0.2 were flagged and evaluated to determine whether or not the SNP should be included. The mean of the normalised theta (AB T Mean) values of the heterozygote cluster is another measurement used for determining correct clustering calls. This value ranges from 0–1 and is used to identify SNPs where the heterozygote cluster has shifted toward either of the homozygote clusters. SNPs that presented with an AB T mean of between 0-0.2 and 1-0.8 were flagged and evaluated further through sequencing to be certain of the clustering calls.

4.7 GENETIC ASSOCIATION ANALYSES BETWEEN SPECIFIC *CRP* SNPs AND CRP CONCENTRATIONS

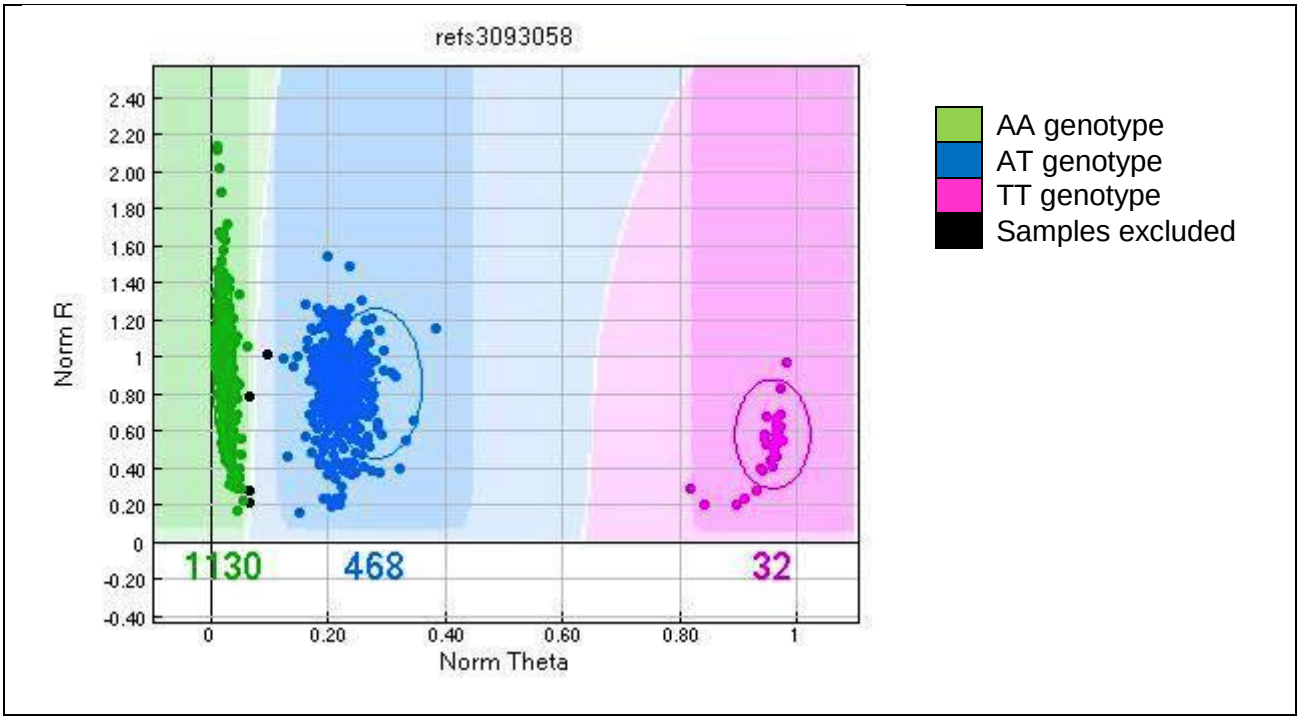
Genetic association analyses were undertaken for each of the SNPs investigated in this study. Adherence of the specific SNPs to the assumptions of HWE was determined for each SNP by means of the Chi-square test for goodness-of-fit. Differences in CRP concentrations were also determined between each genotype group for the whole population, as well as for the rural *versus* urban groups and the men *versus* women groups using the ANCOVA test and adjusting for BMI and fibrinogen. BMI and fibrinogen presented with medium correlations ($r = 0.222$ and $r = 0.479$, respectively; $p < 0.02$) with CRP concentrations and differed significantly between rural/urban groups as well as between the genders and was, therefore, adjusted for in the genetic association analyses. Differences between the genotype groups were also illustrated in graphic format and significant differences ($p \leq 0.05$) were indicated where applicable. Table 4.9 to Table 4.20 are a summary of the findings for each SNP individually.

4.7.1 SNP rs3093058

Figure 4.15 represents the call region shading of the rs3093058 SNP as determined through the GenomeStudio® program. The call rate for this SNP was 0.998, indicating a reliable SNP (Section 4.5.2). The samples indicated in black in Figure 4.15 were excluded because their GenCall score was < 0.5. This means that the samples were too far from the middle of a specific genotype cluster, indicating an unreliable clustering call. The AB R mean of this particular SNP is 0.858, indicating the data was reliable. The AB T mean value of this SNP was 0.282 (Section 4.6.2). To ensure that the alleles are labelled correctly from the BeadXpress® data, caution was given to the strand on which the assay

was design and was then correlated with that of the SNP's actual alleles according to the NCBI website. This process was followed for each of the investigated SNPs.

Figure 4.15 GenomeStudio® shade call regions for SNP rs3093058



A = adenine; refs = reference sequence; T = thymine; Norm R = Norm intensity

The wild-type allele for rs3093058 was determined to be the A allele. The minor allele frequency (MAF) in this population was determined to be 16.5%. In the Sub-Saharan African populations previously investigated it was reported to be 50%, while in African Americans it was 23%. It has, however, been determined to be absent in European populations (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=3093058). Research suggests that the African population consists of the most genetically diverse ethnic groups in the world and it has also been suggested that all other populations arose from this population (Campbell & Tishkoff, 2008). The high MAF in Sub-Saharan Africans could therefore be ascribed to the fact that allele frequencies can vary widely throughout the African population owing to the effects of genetic drift and natural selection (Cavalli-Sforza *et al*., 1994). The sample size of the current black South African population was much larger than that in the previously reported analyses, and it could therefore give a better estimation of the MAF for this population. This must be kept in mind when interpreting these MAF results and when comparing the frequencies of the genotypes between the different population groups.

The Chi-square test indicated in Table 4.9 was conducted to ensure the black South African cohort investigated adhered to the assumptions of HWE (Section 3.8) at the rs3093058 locus. The p-value of the Chi-square test was determined to be non-significant ($p = 0.21$), suggesting that the population was in HWE at the rs3093058 locus, thus indicating that this population is undergoing random mating, no consanguineous breeding, no migration and no selective survival and that the population is sufficiently large (Hardy, 1908).

Table 4.9 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects of the rs3093058 SNP with CRP concentrations in the PURE study population

rs3093058 (A\T)							
		AA	**AT**	**TT**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed number	1130	468	32	1630	0.21	
	Frequency (%)	69	29	2	100		
	Expected number	1141.4	445.2	43.4	1630		
	$X^2 = (O - E)^2 / E$	3.00	1.17	0.11	4.28		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	7.08	9.43	10.83	--	≤0.01	--
	Men	7.17	11.31	11.83	--	≤0.01	0.10*
	Women	6.62	7.94	9.76	--	0.13	
	Rural	6.98	8.36	7.45	--	0.30	0.08*
	Urban	7.10	10.59	14.69	--	≤0.01	

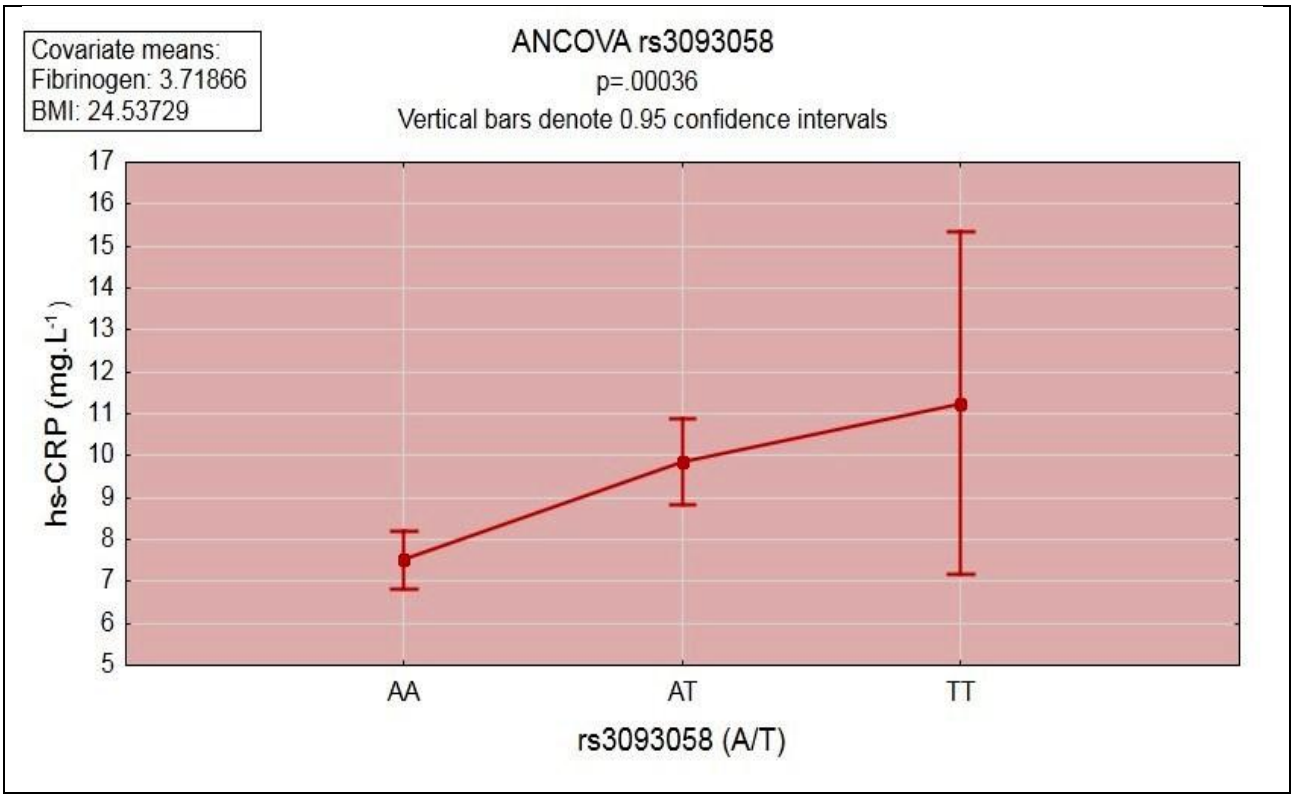
A = adenine; E = expected numbers; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; O = observed numbers; T = thymine; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between rs3093058 and gender and locality on CRP concentrations. **Men (n)** AA = 410; AT = 164; TT = 14; **Women (n)** AA = 683; AT = 289; TT = 16; **Rural (n)** AA = 526; AT = 226; TT = 14; **Urban (n)** AA = 567; AT = 227; TT = 16

Figure 4.16 is a graphic representation of the CRP concentrations in the different genotype groups of the rs3093058 SNP for the whole group while adjusting for fibrinogen and BMI. A significant increase in CRP concentrations was observed in individuals harbouring the TT genotype compared to individuals harbouring the AA genotype (Table 4.9). The larger 95% confidence interval for CRP concentrations for subjects harbouring the homozygote mutant genotype, as indicated in Figure 4.16, could be ascribed to the lower number of subjects harbouring this genotype. However, it is important to note that there are still a significant number of individuals harbouring the mutant allele (T), *i.e*. 16.5%, in the black South African population. Therefore, screening for this SNP may have a significant public health impact.

The CRP concentrations adjusted for fibrinogen and BMI of the rural and female subjects did not differ significantly between the different genotype classes, which was in contrast to the urban and male subjects, who had significantly different ($p \leq 0.01$) CRP

concentrations. Therefore, one could argue that when residing in an urban community and adopting the concomitant lifestyle, harbouring the T allele has an increased influence on CRP concentration, but when residing in a rural community the genotype harboured at the rs3093058 locus makes no difference to the CRP phenotype. However, no interaction effect ($p = 0.08$) was observed for this genotype between subjects residing in the different communities in relation to CRP concentrations, although this might be due to lack of power. Significant differences in CRP concentrations were observed in men over the different genotype groups; this was, however, not true for women and are discussed in detail in Section 4.9. The same applies to the interaction effect between gender and the genotype groups at the rs3093058 locus, where no interaction effect was observed ($p = 0.10$). This is discussed in Section 4.9.

Figure 4.16 CRP concentrations in the three genotype groups of rs3093058 for the whole PURE cohort investigated



A = adenine; BMI = body mass index; hs-CRP = high sensitivity C-reactive protein; rs = reference sequence; T = thymine

Similar results were reported for rs3093058 and rs3093062. After testing for linkage disequilibrium (LD), these SNPs were determined to be in LD within the study population (data not presented). It is, therefore, difficult to determine which of these SNPs is the functional SNP affecting CRP concentrations. This may indicate that there is a specific haplotype associated with increased CRP; however, further study is required. This

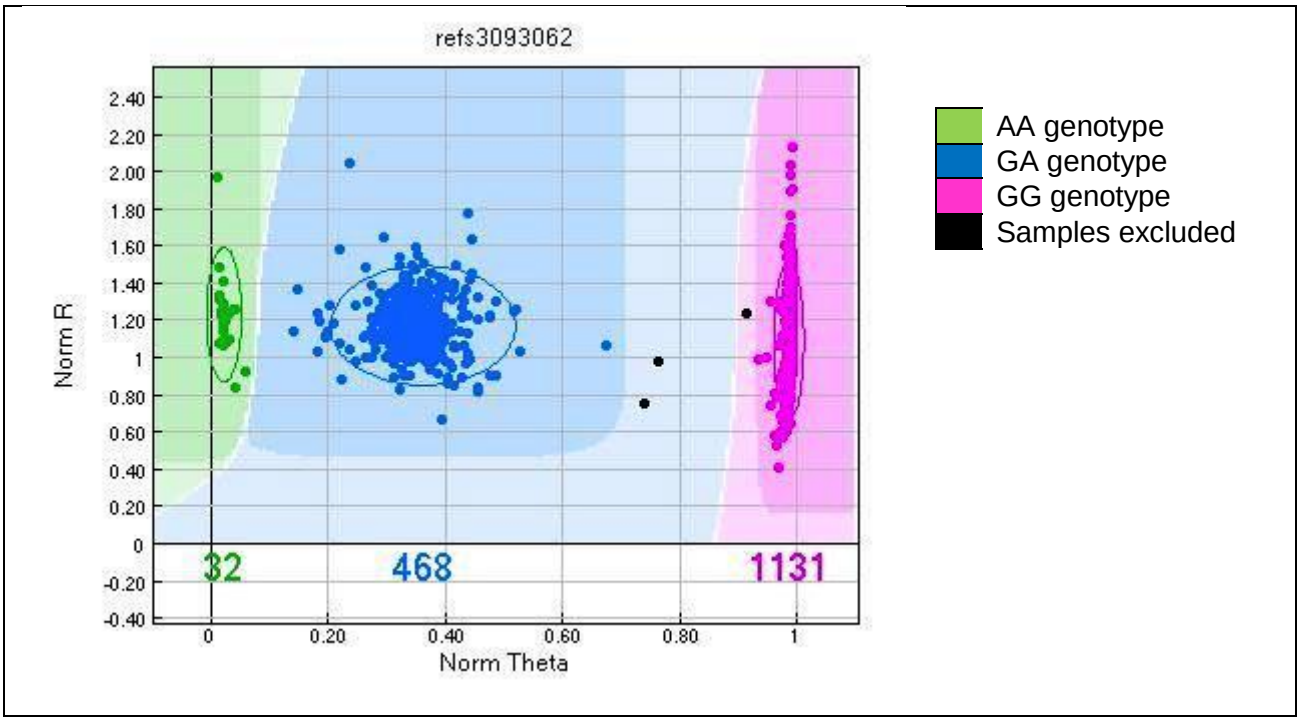
indicates that in future when screening for these SNPs in the black South African population, only one needs be included.

In numerous published studies, rs3093058 was reported to have an association with increased CRP concentrations (Crawford *et al*., 2006). The T allele was furthermore associated with an increased risk of MI in black African Americans (Lange *et al*., 2006). The observation of the current investigation is in agreement with what is reported in literature, *i.e*. increasing CRP concentrations are associated with this SNP.

4.7.2 SNP rs3093062

The shade call regions for the rs3093062 locus are illustrated in Figure 4.17, with the different genotype groups represented by the different colours indicated in the legend. As mentioned in the previous section, the samples in black were excluded from all analyses because of their call rate < 0.5. The call rate of this SNP was 0.998, indicating that the largest proportion of all the samples at this locus had a call score above the no call threshold. The AB R mean and AB T mean were 1.171 and 0.362 respectively, indicating a reliable cluster call (Section 4.6.2).

Figure 4.17 GenomeStudio® shade call region for SNP rs3093062



A = adenine; G = guanine; refs = reference sequence; Norm R = Norm intensity

The ancestral allele for rs3093062 is the G allele. The MAF for the current population was calculated to be 16.5%. The MAF for this SNP in previously investigated Sub-Saharan African, African American and European populations was reported to be 50%, 24% and 0%, respectively (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=3093062). This would imply that the factors affecting genotype distribution, such as migration and natural selection, are different in the different population groups, resulting in the large differences in the MAF. It is also important to bear in mind that the number of subjects used to determine the MAF of the reported populations differed, *i.e*. the MAF of the Sub-Saharan Africans was only based on two individuals, African Americans on 46 individuals and the Europeans on 120 individuals. The sample set of the current black South African population was much larger (1,587) and therefore the same reasoning described in Section 4.7.1 applies.

A Chi-square test was conducted to determine whether the distribution of this genetic variation within the current study population was in HWE as indicated in Table 4.10. The p-value of the test was not significant (p = 0.21), suggesting that the population is in HWE for rs3093062 and complies with the assumptions of Hardy Weinberg equilibrium, as described in Section 3.8.

Table 4.10 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects with CRP concentrations of the rs3093062 locus in the PURE study population

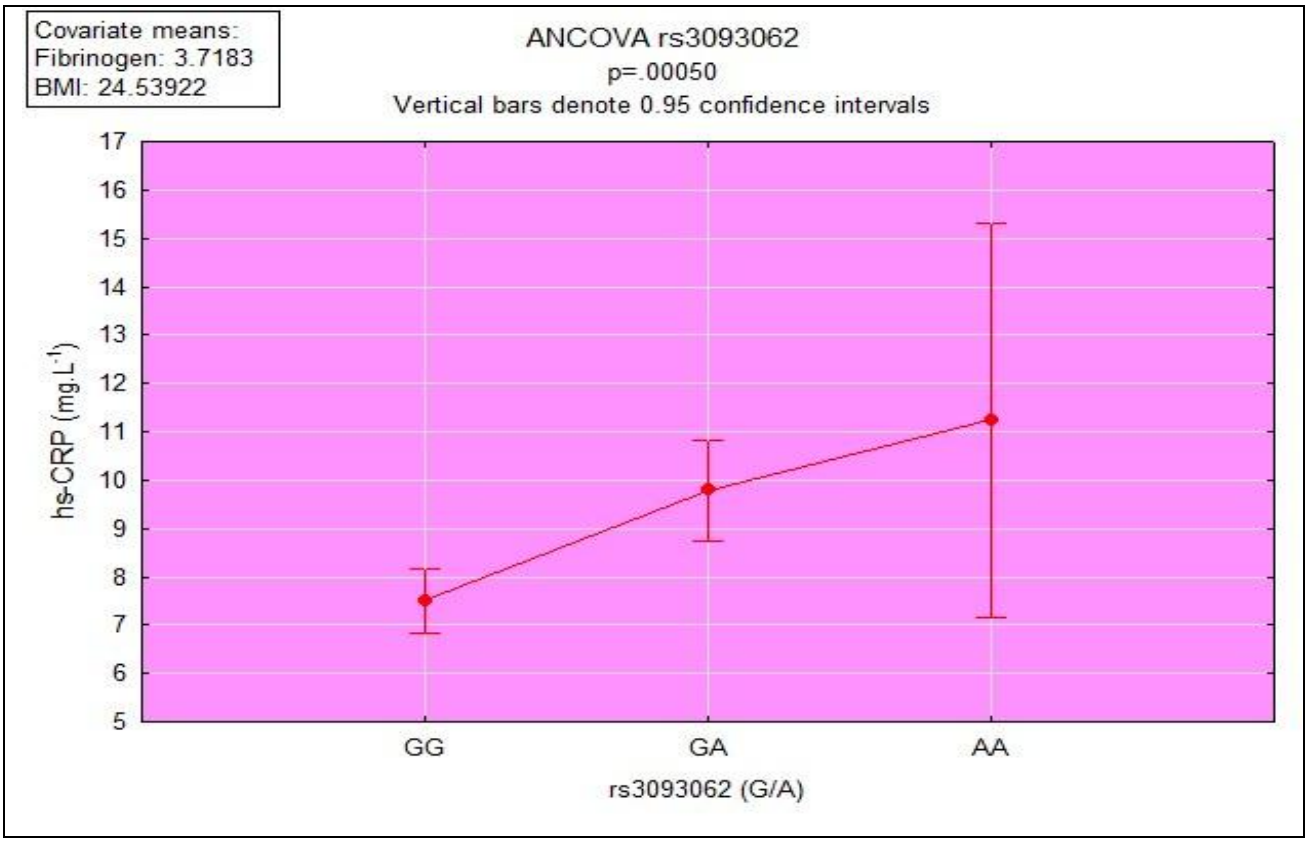
rs3093062 (G/A)							
		GG	**GA**	**AA**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	1131	468	32	1631	0.21	
	Frequency (%)	69	29	2	100		
	Expected numbers	1141.4	445.2	43.4	1630		
	$X^2 = (O – E)^2 / E$	0.11	1.16	2.99	4.26		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	7.07	9.36	10.81	--	<0.01	
	Men	7.17	11.30	11.80	--	<0.01	0.08*
	Women	6.62	7.84	9.76	--	0.17	
	Rural	6.97	8.22	7.44	--	0.36	0.06*
	Urban	7.11	10.60	14.66	--	<0.01	

A = adenine; G = guanine; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; rs = reference sequence; X^2 = Chi-square value; O = observed; E = expected; * = indicates the p-values of the interaction effect between rs3093062, gender and locality on CRP concentrations; **Men (n)** GG = 410; GA = 164; AA = 14; **Women (n)** GG = 684; GA = 289; AA =16; **Rural (n)** GG = 527; GA = 225; AA = 14; **Urban (n)** GG = 567; GA = 228; AA =16

The rs3093062 SNP indicated a significant (p < 0.01) association with CRP concentrations (Figure 4.18), when adjusting for BMI and fibrinogen in the whole group. A significant increase in CRP concentrations was observed between the different genotype groups, with

the largest difference being observed between the heterozygote GA genotype and the homozygote GG genotype. No interaction effect was determined between the rural/urban groups or for the gender groups and CRP concentrations. Significant differences ($p < 0.01$) were, however, observed for the CRP concentrations between the different genotype groups of the individuals residing in the urban group (Table 4.10). No significant differences were observed in the rural group ($p = 0.36$). Once again it seems that when residing in an urban community and adopting the associated lifestyle, the genotype harboured, at this particular locus, has an influence on CRP concentrations, but when residing in a rural community it has no significant effect on the CRP phenotype. Furthermore, CRP concentrations for men were reported to be significantly different between the different genotype groups ($p < 0.01$), whereas those of women were observed not to have a significant association ($p = 0.17$). These significant observations in the urban and male groups are discussed in more detail in Section 4.9.

Figure 4.18 CRP concentrations in the three genotype groups of rs3093062 for the whole PURE cohort



A = adenine; ANCOVA = analysis of covariance; BMI = body mass index; G = guanine; hs-CRP = high sensitivity C-reactive protein; rs = reference sequence

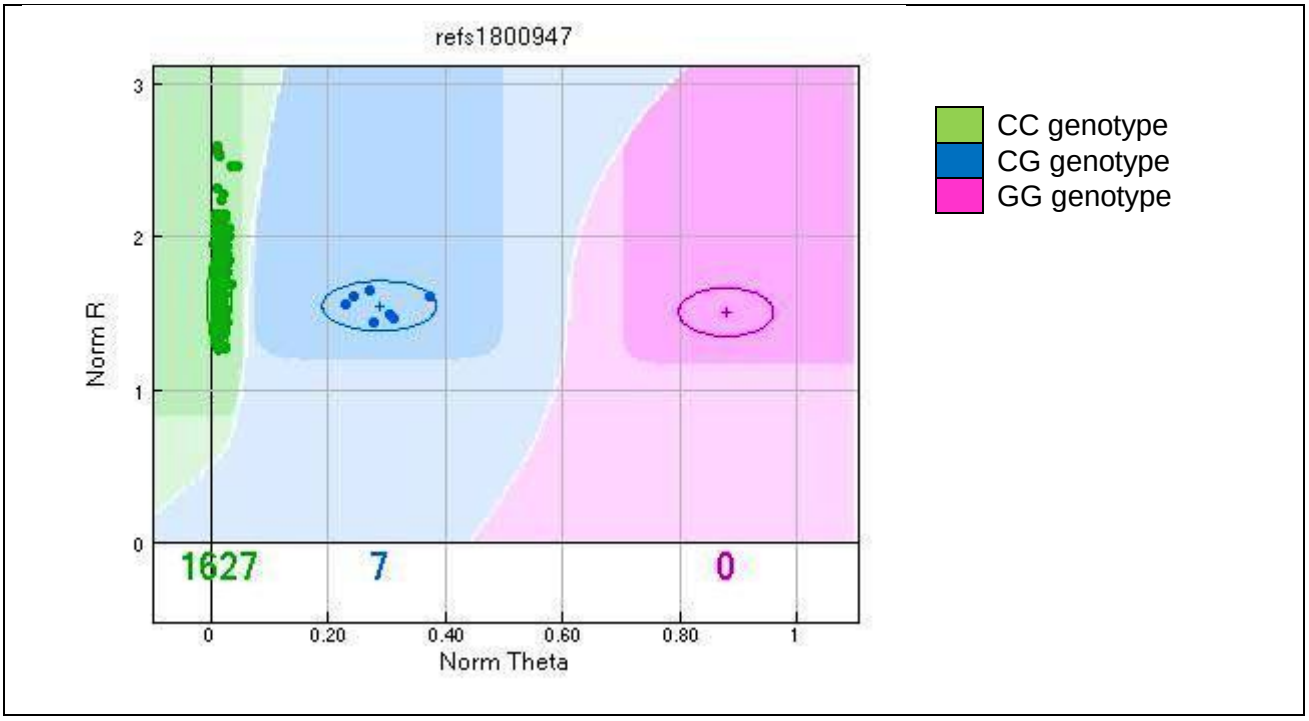
Szalai *et al*. (2005) reported that in African-Americans, the average serum CRP was lowest in individuals harbouring the AA homozygote at the rs3093062 locus, intermediate in those harbouring the GA heterozygote, and highest in individuals harbouring the GG

homozygote. In a non-Hispanic black sample studied by Crawford *et al*. (2006), similar results were reported in that there was an increase in CRP concentrations with the addition of the mutant G allele. Contrastingly, in this black South African population the opposite was observed. This observation could possibly be ascribed to the genetic distribution observed within this large sample size which could be a more robust distribution profile for the Tswana population.

4.7.3 SNP rs1800947

As can be seen in Figure 4.19, the majority of this population harboured the wild-type genotype (CC) for SNP rs1800947. The shade call regions, represented by the different colours, are clearly illustrated in Figure 4.19. The call rate for the rs1800947 SNP was determined to be 1 and both the AB R mean (1.547) and the AB T mean (0.287) were also within desirable ranges (Section 4.6.2).

Figure 4.19 GenomeStudio® shade call region for SNP rs1800947



C = cytosine; G = guanine; refs = reference sequence; Norm R = Norm intensity

For rs1800947, the reference allele is the C allele. The MAF for rs1800947 in African Americans and European populations is estimated to be 2.3% and 6.2%, respectively. The major C allele was absent in the previously reported Sub-Saharan African population (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1800947). This is not consistent with the observations in the current population, where the C allele had a 99.7%

prevalence. This particular SNP was determined to adhere to the assumptions of HWE (p = 0.93) as discussed in Section 3.8.

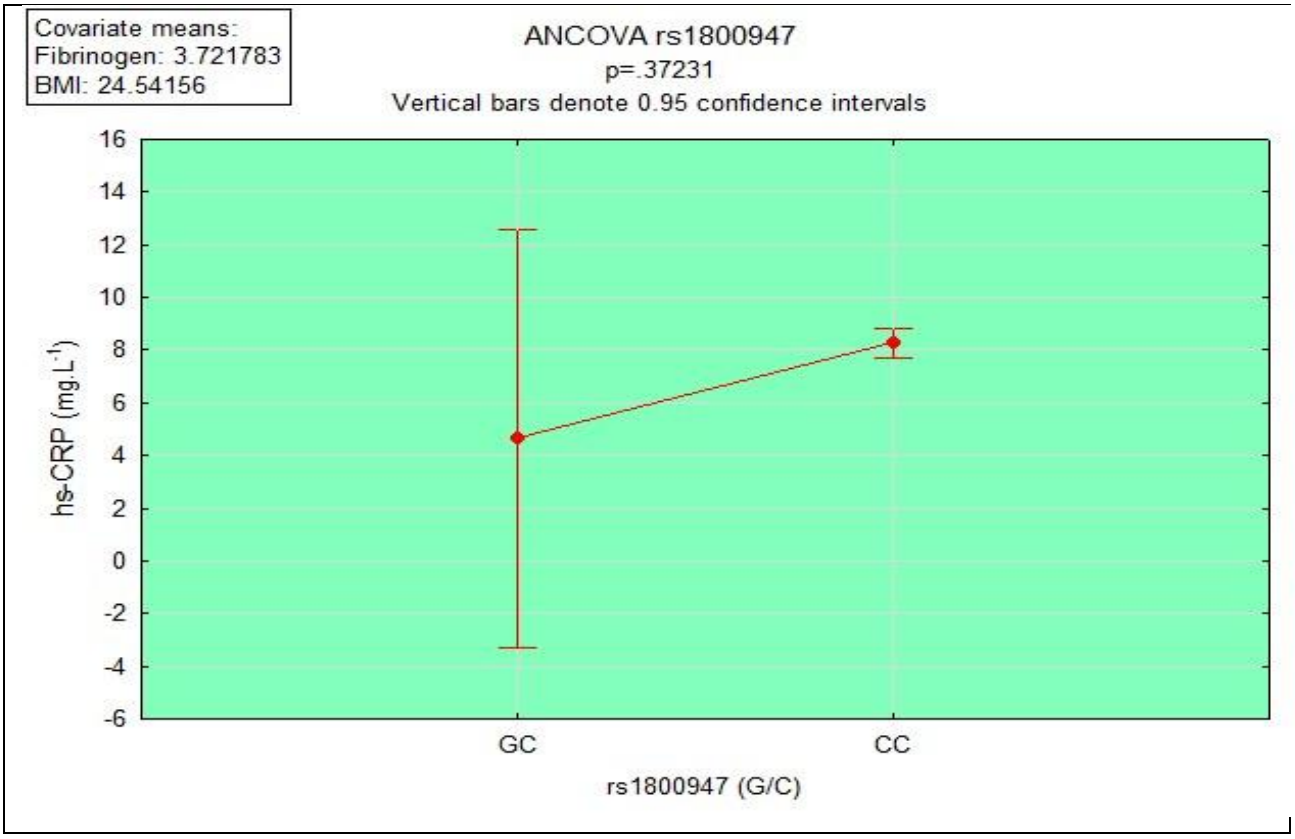
Table 4.11 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects of the rs1800947 SNP with CRP concentrations in the PURE study population

rs1800947(G/C)							
		GG	**GC**	**CC**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	0	7	1627	1634	0.93	
	Frequency (%)	0	0	100	100		
	Expected numbers	0.01	6.99	1627.0	1634		
	$X^2 = (O - E)^2 / E$	0.01	0.00	0.00	0.01		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	--	4.39	8.01	--	0.37	
	Men	--	3.67	8.64	--	0.51	0.81*
	Women	--	4.22	7.22	--	0.53	
	Rural	--	4.94	7.46	--	0.71	0.77*
	Urban	--	3.38	8.34	--	0.43	

G = guanine; C = cytosine; E = expected; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; O = observed; rs = reference sequence; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between rs1800947, gender and locality on CRP concentrations; **Men (n)** GC = 3; CC = 585; **Women (n)** GC = 4; CC = 988; **Rural (n)** GC = 4 CC =764; **Urban (n)** GC = 3 CC = 809

As seen in Figure 4.20, no-one harboured the homozygous wild-type genotype within the black South African cohort, and no significant differences were observed between the heterozygote and the homozygous mutant genotype (p = 0.37) for CRP concentrations. The same applies to the interaction effect between the rural/urban groups and the gender groups (Table 4.11). Since there were no subjects harbouring the homozygous wild-type genotype, it has limited the observations the researcher could make regarding the different genotype associations to CRP concentrations. Furthermore, the lack of statistical significance when comparing the heterozygote individuals with those harbouring the homozygous mutant genotype could be due to a lack of power in view of the low number of individuals harbouring the heterozygote genotype. Too little statistical power owing to small sample sizes might result in erroneous conclusions and future studies aiming to investigate rs1800947 should enrol many more subjects than the current study. However, because of the low frequency of the G allele, it is unlikely that this SNP will be of importance for the health of the population.

Figure 4.20 CRP concentrations in the two genotype groups of rs1800947 for the whole PURE cohort



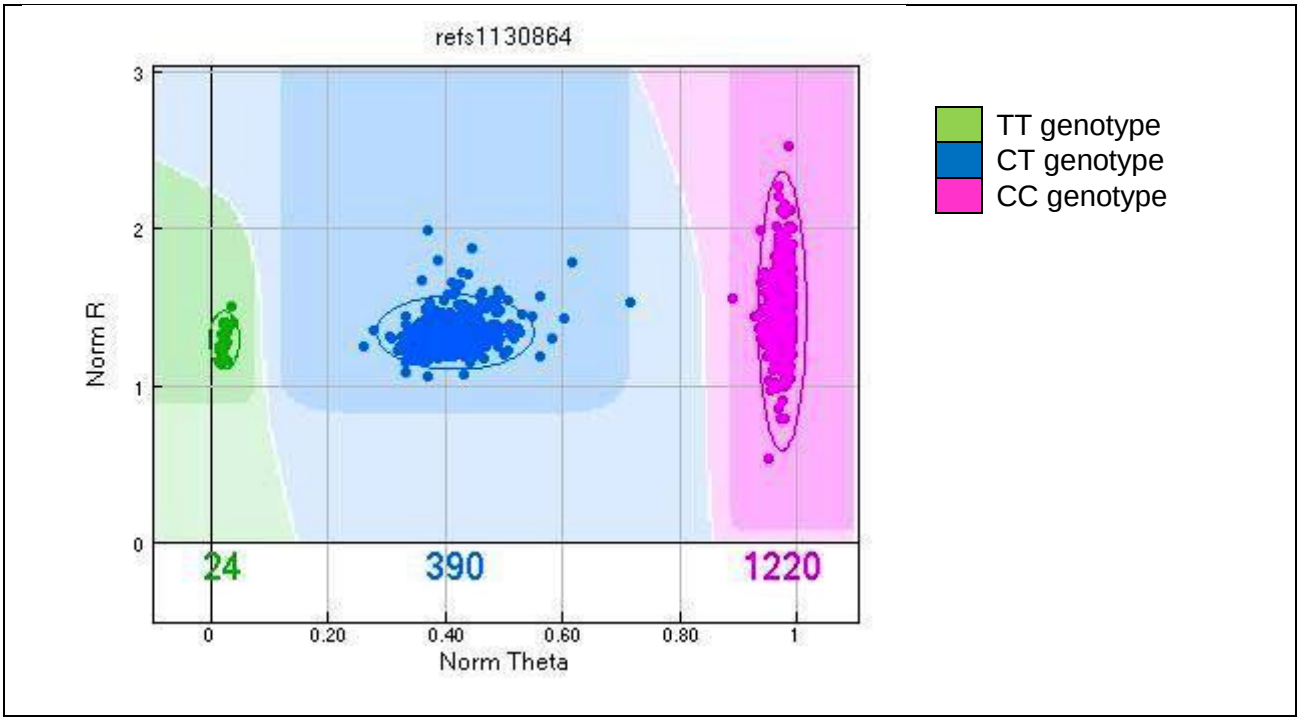
ANCOVA = analysis of covariance; BMI = body mass index; C = cytosine; G = guanine; hs-CRP = high sensitivity C-reactive protein; rs = reference sequence

Schumacher and co-workers (2009), reported a significant (p = 0.04) increase in CRP concentrations in individuals harbouring the heterozygous GC genotype compared to those harbouring the homozygous CC genotype. They also reported a low frequency of the wild-type allele. Rhodes *et al*. (2008) and Wensley *et al*. (2011) reported similar results, *i.e*. an increase in CRP concentrations. The finding from the current research project regarding rs1800947, although not significant, is similar to what is found in the literature.

4.7.4 SNP rs1130864

As mentioned, the GenomeStudio® program was used for determining the shade call regions for all the SNPs investigated in this study and Figure 4.21 is a visual representation of the genotype groups at the rs1130864 locus. The call rate for SNP rs1130864 was determined to be 1. The other two measurements that were used for determining the correct clustering calls were the AB T mean and AB R mean, which were 0.417 and 1.343, respectively. These three measurements are all within the desirable ranges (Section 4.6.2).

Figure 4.21 GenomeStudio® shade call region for SNP rs1130864



C = cytosine; refs = reference sequence; T = thymine; Norm R = Norm intensity

The C allele was determined to be the reference allele for rs1130864. The MAF (13%) of rs1130864 in this black South African cohort is in agreement with other population groups such as African Americans (14.3%) and Sub-Saharan Africans (10.6%), but differs from that of the European population, which was 30.4% (http://www.ncbi.nlm.nih.gov/projects/SNP/ snp_ref.cgi?rs=1130864). The fact that the mutant allele has a low prevalence in all the African populations and a slightly higher prevalence in the European populations highlights the fact that populations are exposed to different factors, such as natural selection, inbreeding and migration, which affect their genotype distribution. This is most likely due to the bottleneck effect. SNP rs1130864 was determined to adhere to the assumptions of the Hardy-Weinberg equilibrium (Section 3.8), as a non-significant p-value ($p = 0.25$) was determined after Chi-square testing, as indicated in Table 4.12.

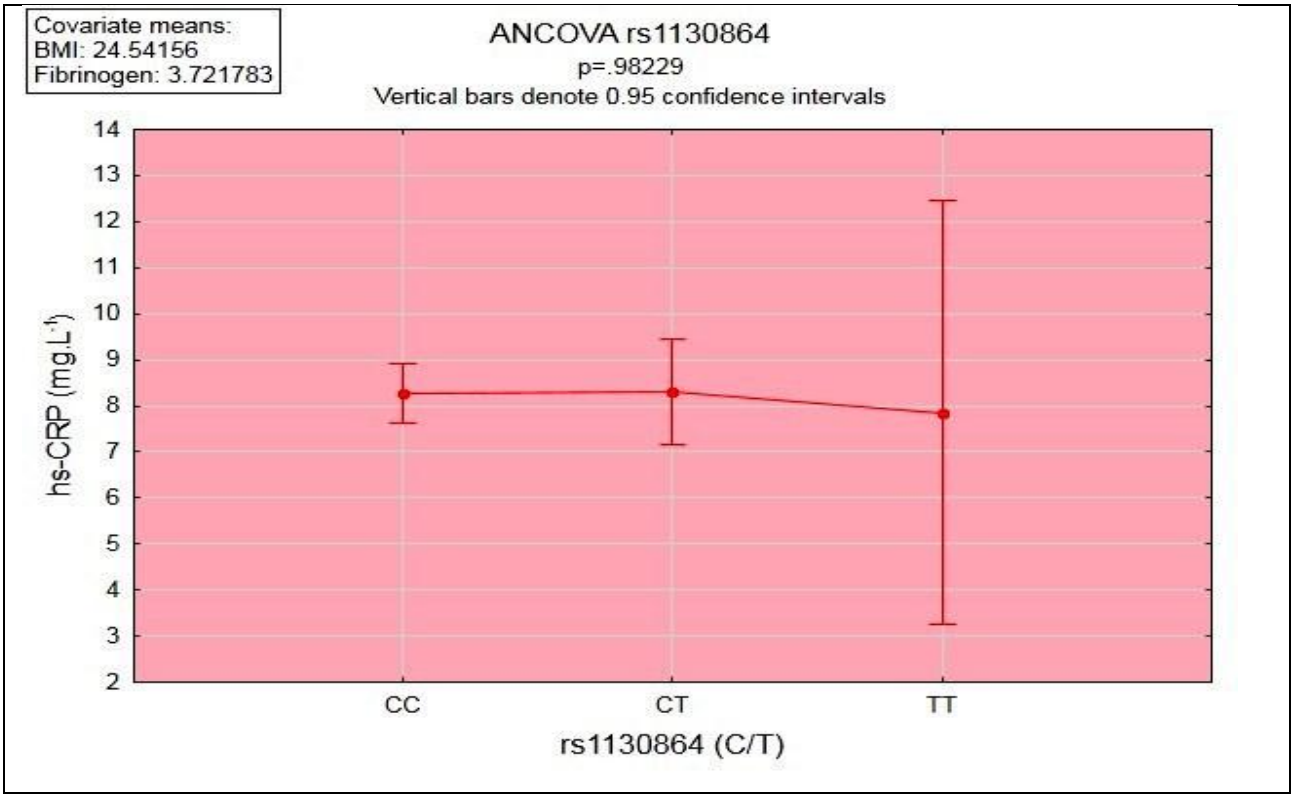
Table 4.12 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects with CRP concentrations of the rs1130864 locus in the PURE study population

rs1130864 (C/T)							
		CC	**CT**	**TT**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	1220	390	24	1634	0.25	
	Frequency (%)	75	24	1	100		
	Expected numbers	1225.35	379.30	29.35	1634		
	$X^2 = (O - E)^2 / E$	0.02	0.30	0.98	1.30		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	8.96	9.01	8.56	--	0.98	
	Men	9.29	10.29	9.63	--	0.74	0.48*
	Women	8.33	7.68	7.57	--	0.79	
	Rural	8.47	8.10	8.18	--	0.93	0.75*
	Urban	9.10	9.76	8.71	--	0.73	

C = cytosine; E = expected; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; O = observed; rs = reference sequence; T = thymine; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between rs1130864, gender and locality on CRP concentrations; **Men (n)** CC = 424; CT = 155; TT = 9; **Women (n)** CC = 754; CT = 224; TT = 14; **Rural (n)** CC = 560; CT = 193; TT = 15; **Urban (n)** CC = 618; CT = 186; TT = 8

In the men and urban subgroups, as well as in the entire group, the highest CRP concentrations were observed in the heterozygote genotype group (CT). This was, however, not a significant effect. No significant interaction effect was observed either, as can be seen in Table 4.12 and Figure 4.22. Therefore, in addition to the SNP having no significant effect on CRP concentrations in the whole group, neither gender nor the level of urbanisation had an interaction effect with rs1130864 on CRP concentrations. This would imply that this SNP is unlikely to have an effect on CRP concentrations.

Figure 4.22 CRP concentrations in the three genotype groups of rs1130864 for the whole PURE cohort



ANCOVA = analysis of covariance; BMI = body mass index; C = cytosine; hs-CRP = high sensitivity C-reactive protein; rs = reference sequence; T = thymine

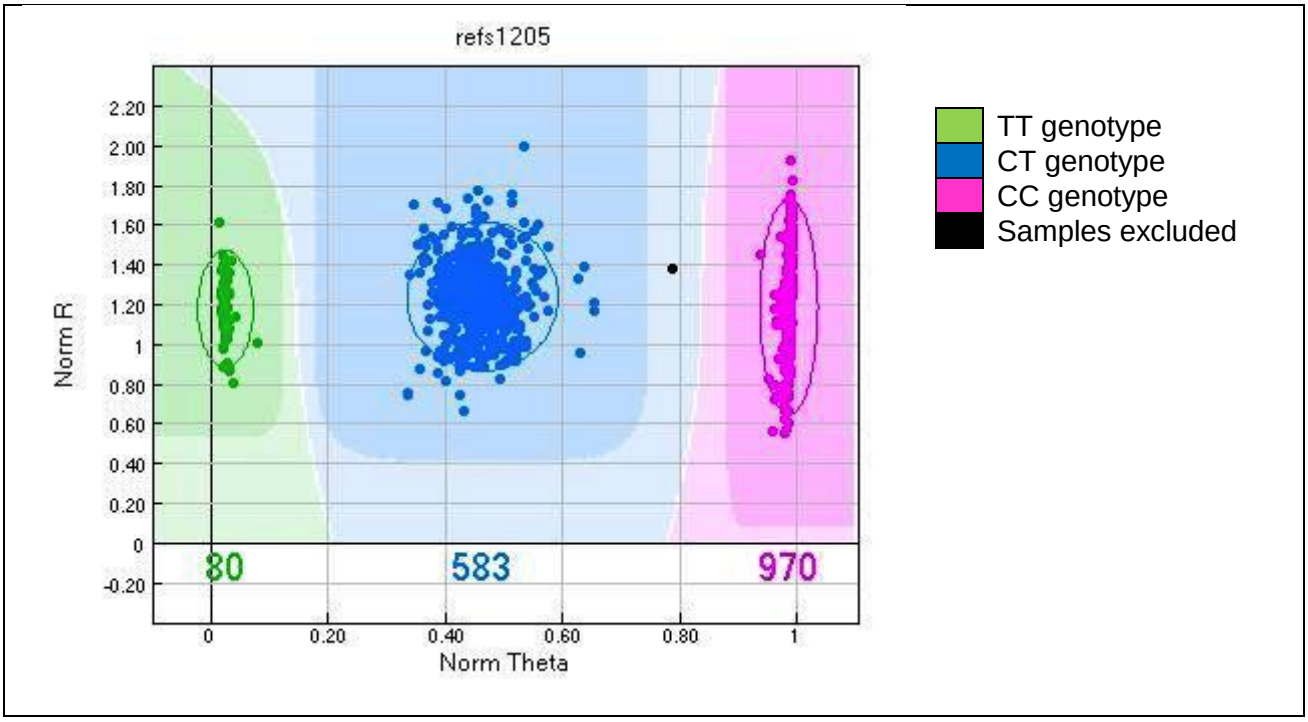
Circulating CRP concentrations were significantly associated with the TT genotype at the rs1130864 locus compared to the individuals harbouring the CC and CT genotypes in five studies for which the analyses were pooled together (Lawlor *et al*., 2008a). Lee *et al*. (2009) reported that individuals harbouring the CC genotype had higher CRP concentrations than the other two genotype groups in European individuals. Although a slight decrease in CRP concentrations were observed in the current population, for individuals harbouring the minor allele of this SNP, it was not a significant effect. This SNP, therefore, has no public health significance and is unlikely to be included as a possible public health screening tool. However, it does highlight the differential associations that are present in different populations.

4.7.5 SNP rs1205

The shading call regions for the rs1205 SNP are illustrated in Figure 4.23, where the samples in black were excluded because of a GenCall score of less than 0.5. The call rate of this SNP was reported to be 0.999, indicating that the largest proportion of the samples at each locus had a call score above the no call threshold. The AB R mean and AB T

mean were 1.242 and 0.463, respectively indicating a good reliable clustering call (Section 4.6.2).

Figure 4.23 GenomeStudio® shade call region for SNP rs1205



A = adenine; G = guanine; refs = reference sequence; Norm R = Norm intensity

For the rs1205 locus, the C allele was determined to be the reference or ancestral allele. The MAF at the rs1205 locus in this population (23%) is slightly higher than that reported in African Americans (14%) as well as Sub-Saharan Africans (15%). It seems to be in agreement with the European population of 25% (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1205). A possible reason for the MAF being lower in comparison with that of the other African populations is similar to the reasons given in Section 4.7.1. The cohorts of the European population and the Sub-Saharan Africans consisted of 226 individuals and the African Americans consisted of 44 individuals. Therefore, the reported MAF established for this SNP should be interpreted with caution.

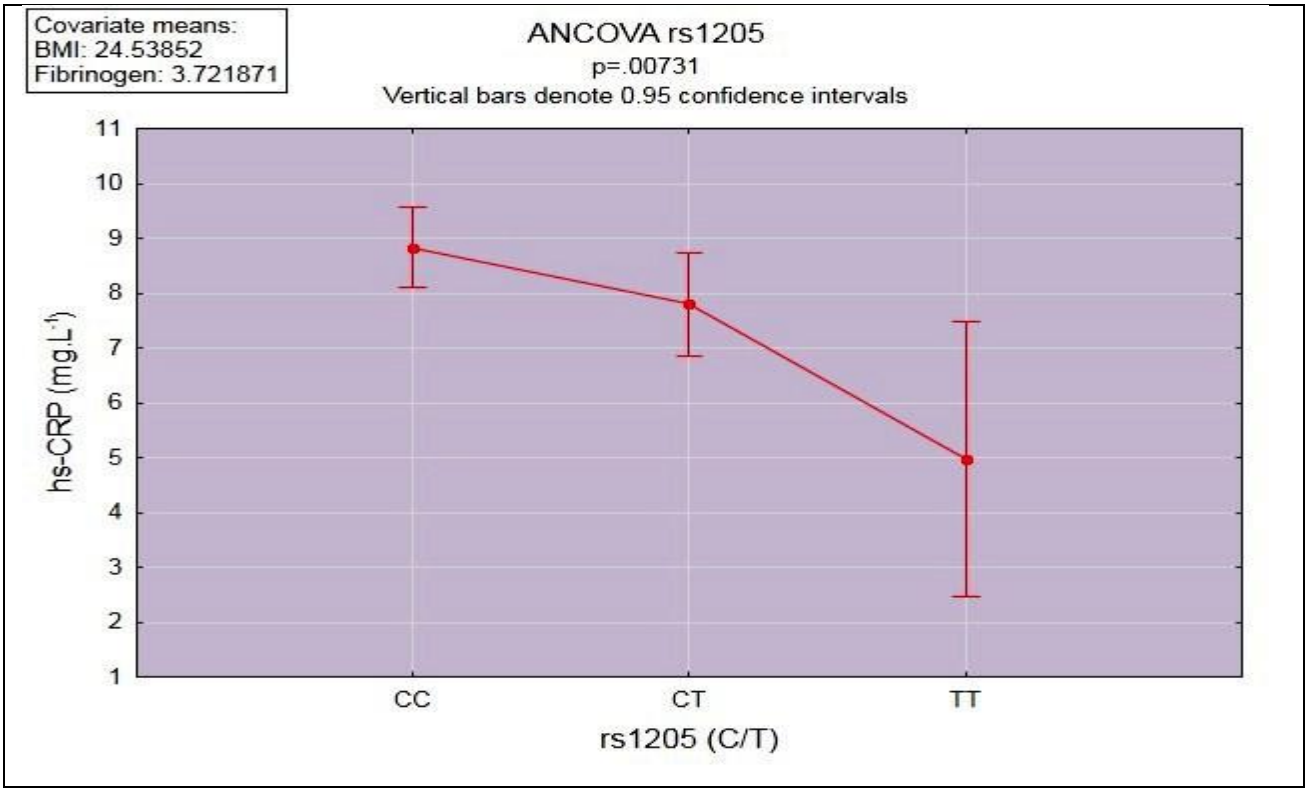
Table 4.13 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects of the rs1205 SNP with CRP concentrations in the PURE study population

rs1205 (C/T)							
		CC	**CT**	**TT**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	970	583	80	1633	0.52	
	Frequency (%)	59	36	5	100		
	Expected numbers	974.51	573.97	84.51	1633		
	$X^2 = (O - E)^2 / E$	0.02	0.14	0.24	0.4		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	8.74	7.71	4.88	--	<0.01	
	Men	10.42	7.25	4.25	--	<0.01	0.02*
	Women	7.63	7.90	4.97	--	0.16	
	Rural	8.06	7.65	5.08	--	0.22	0.46*
	Urban	9.48	7.72	4.52	--	0.01	

A = adenine; E = expected; G = guanine; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; O = observed; rs = reference sequence; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between rs1205, gender and locality on CRP concentrations; **Men (n)** CC = 340; CT = 228; TT = 20; **Women (n)** CC = 597; CT = 337; TT = 57; **Rural (n)** CC = 462; CT = 260; TT = 45; **Urban (n)** CC = 475; CT = 305; TT = 32

This SNP adheres to the assumptions of HWE (Section 3.8) with a p value of 0.52 determined via Chi-square testing (Table 4.13). Significant differences in CRP concentrations were observed between the different ($p < 0.01$) genotype groups at the rs1205 locus when investigating the whole group (Figure 4.24). The greatest difference was observed between those harbouring the heterozygote CT genotype and those harbouring the homozygote TT genotype. In this case, it seems the SNP imparts a protective effect, as it is associated with a decrease in CRP concentrations. For the men ($p < 0.01$), as well as in the urban group ($p = 0.01$), significant differences were also observed for CRP concentrations between the genotype groups.

Figure 4.24 CRP concentrations in the three genotype groups of rs1205 for the whole group

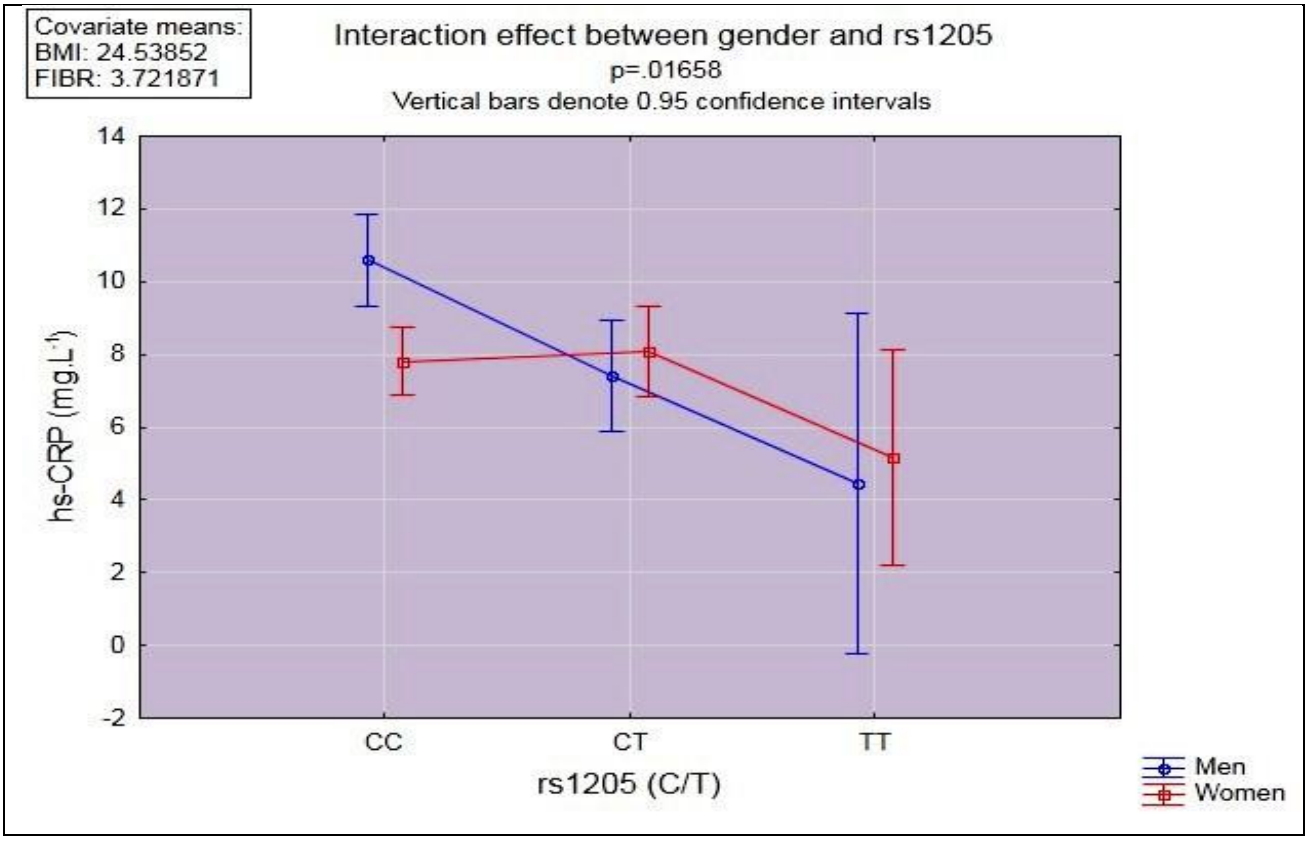


ANCOVA = analysis of covariance; BMI = body mass index; C = cytosine; hs-CRP = high sensitivity C-reactive protein; rs = reference sequence; T = thymine

For SNP rs1205 only men and individuals residing in the urban setting reported significant differences between genotype classes and a decrease in CRP concentrations and are discussed in Section 4.9. A significant ($p = 0.02$) interaction effect between gender and CRP concentrations was observed for this SNP (Figure 4.25). The mutant allele in this population is present at a percentage of 39%, which indicates possible public health significance and could potentially be used in the future as a genetic public health screening tool in the black population of South Africa.

As can be observed from Figure 4.25, among individuals harbouring the CC genotype, men had a higher CRP concentration than women, but for individuals harbouring either the CT or TT genotype, women had higher CRP concentrations. Therefore, with the addition of the mutant allele, women had higher CRP concentrations than men. There seems to be an underlying factor that causes women to have higher CRP concentrations than those of men. This is a possible gene-gender interaction and reasons for it should be investigated in future studies (discussed in detail Section 4.9).

Figure 4.25 The effect of the interaction between genotype and gender on CRP concentrations for rs1205



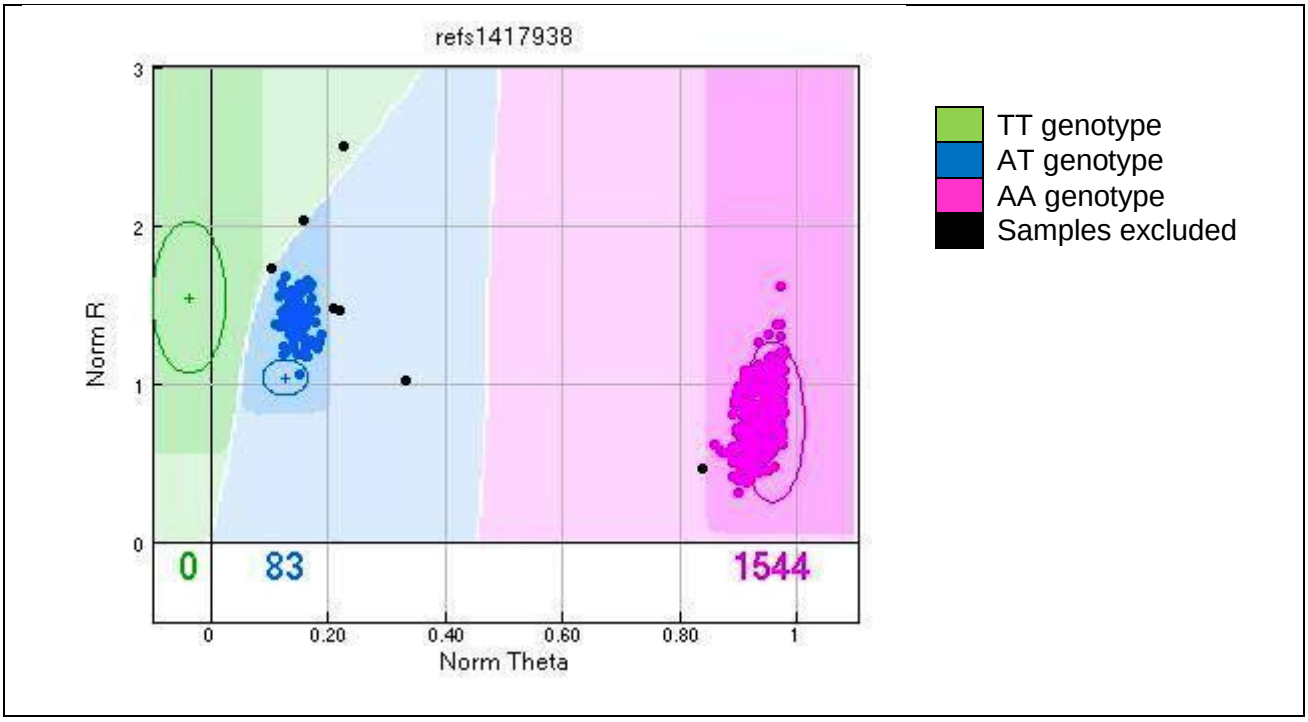
BMI = body mass index; C = cytosine; hs-CRP = high sensitivity C-reactive protein; rs = reference sequence; T = thymine

Similar lowering effects were reported by numerous other studies, with the addition of the mutant T allele at the rs1205 locus. Flores-Alfaro and co-workers (2012) reported a significant 0.5 $mg.L^{-1}$ reduction in CRP for individuals harbouring the TT genotype at the rs1205 locus. The same results were reported in other studies by Schumacher *et al*. (2008) and Eiriksdottir *et al*. (2009), in that there was a reduction in CRP concentrations in those harbouring the T allele. These studies are, therefore, in line with what is observed in the current study.

4.7.6 SNP rs1417938

The call rate for this particular SNP was calculated to be 0.996 (Section 4.6.2). The shading call regions are illustrated in Figure 4.26, where it was determined that the majority of this population harboured the wild-type allele for rs1417938. The AB R mean was determined to be 1, whereas the AB T mean was determined to be 0.127. When evaluating this AB T mean, the clustering call could be interpreted to be weak, therefore the clustering calls determined *via* the BeadXpress® was additionally assessed by means of sequencing of certain random samples to ensure correct clustering calls. Samples highlighted in black were excluded because of a GenCall score of less than 0.5.

Figure 4.26 GenomeStudio® shade call region for SNP rs1417938



A = adenine; refs = reference sequence; T = thymine; Norm R = Norm intensity

The reference allele for rs1417938 was reported to be the T allele. The MAF of this particular SNP in the investigated population is high (97.5%) compared to African Americans (17%) as well as Europeans (33%). Contrastingly, in Sub-Saharan Africans the MAF was previously reported to be 4% (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1417938). The Chi-square test for rs1417938 gave a p-value of 0.29, which indicates that this SNP adheres to the assumptions of HWE (Section 3.8). Table 4.14 is a summary of the results for SNP rs1417938.

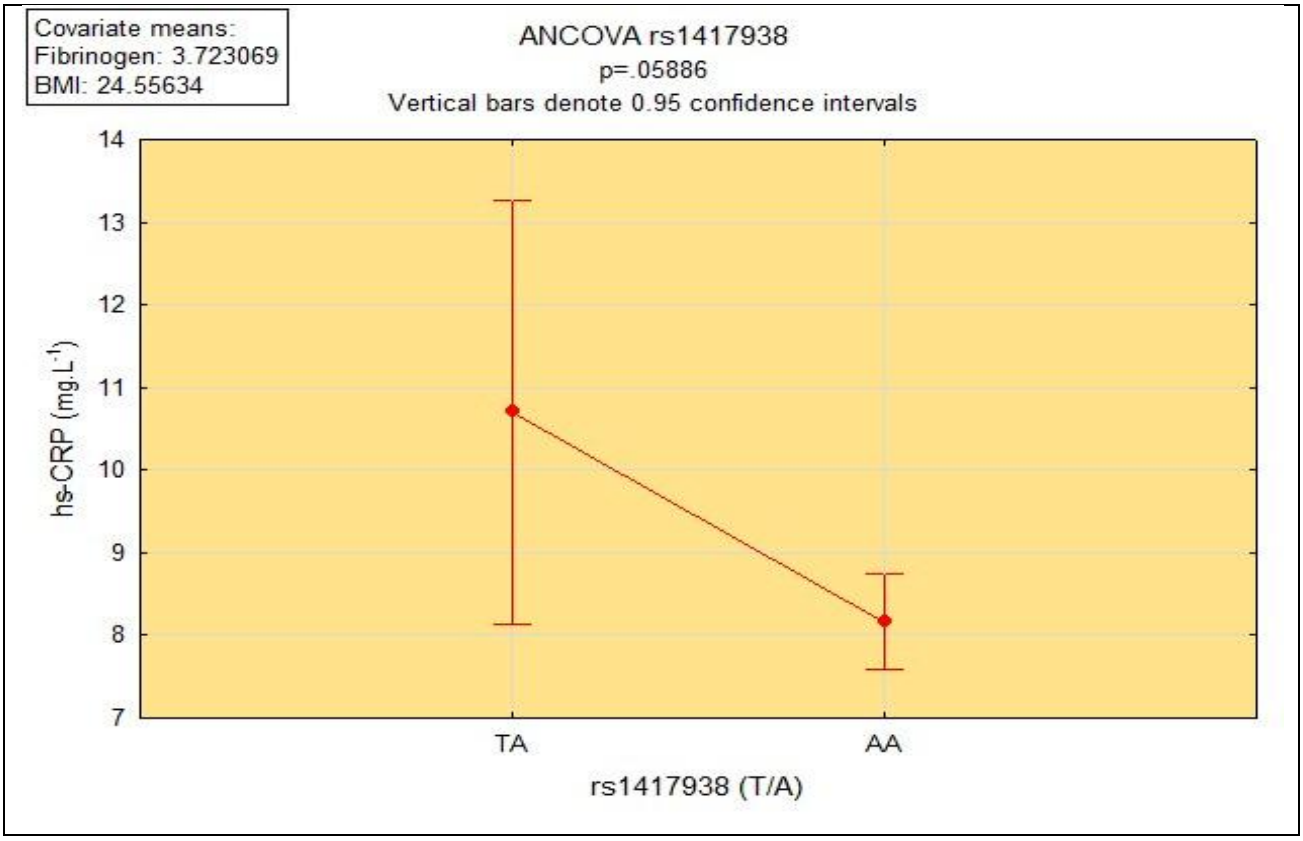
Table 4.14 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects of the rs1417938 SNP with CRP concentrations in the PURE study population

rs1417938 (T/A)							
		TT	**TA**	**AA**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	0	83	1544	1627	0.29	
	Frequency (%)	0	5	95	100		
	Expected numbers	1.06	80.88	1545.06	1627		
	$X^2 = (O - E)^2 / E$	1.06	0.06	0.00	1.12		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	--	10.57	8.03	--	0.06	
	Men	--	12.01	8.63	--	0.13	0.63*
	Women	--	9.33	7.26	--	0.17	
	Rural	--	10.12	7.64	--	0.23	0.99*
	Urban	--	11.01	8.49	--	0.16	

A = adenine; E = expected; CRP = C-reactive protein; O = observed; $mg.L^{-1}$ = milligrams per litre; rs = reference sequence; T = thymine; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between rs1417938 , gender and locality on CRP concentrations; **Men (n)** TA = 32; AA = 553 ; **Women (n)** TA = 45; AA = 944; **Rural (n)** TA = 35; AA =731; **Urban (n)** TA = 42; AA = 766

Increases in CRP concentrations can be seen when comparing individuals harbouring the AA genotype to those harbouring the TA genotype, but this difference was not determined to be significant in the whole group (Figure 4.27), or in the groups split for gender or level of urbanisation (Table 4.14). No interaction effect was observed between either the gender groups or rural/urban groups in relation to the SNP. From Figure 4.27 one can observe a difference between the two genotype groups; however, it is not significant. A possible reason for this observation could be the small number of individuals harbouring the TA genotype. Because of this fact it is unlikely that this SNP has any public health implication in this population.

Figure 4.27 CRP concentrations in the two genotype groups of rs1417938 for the whole group



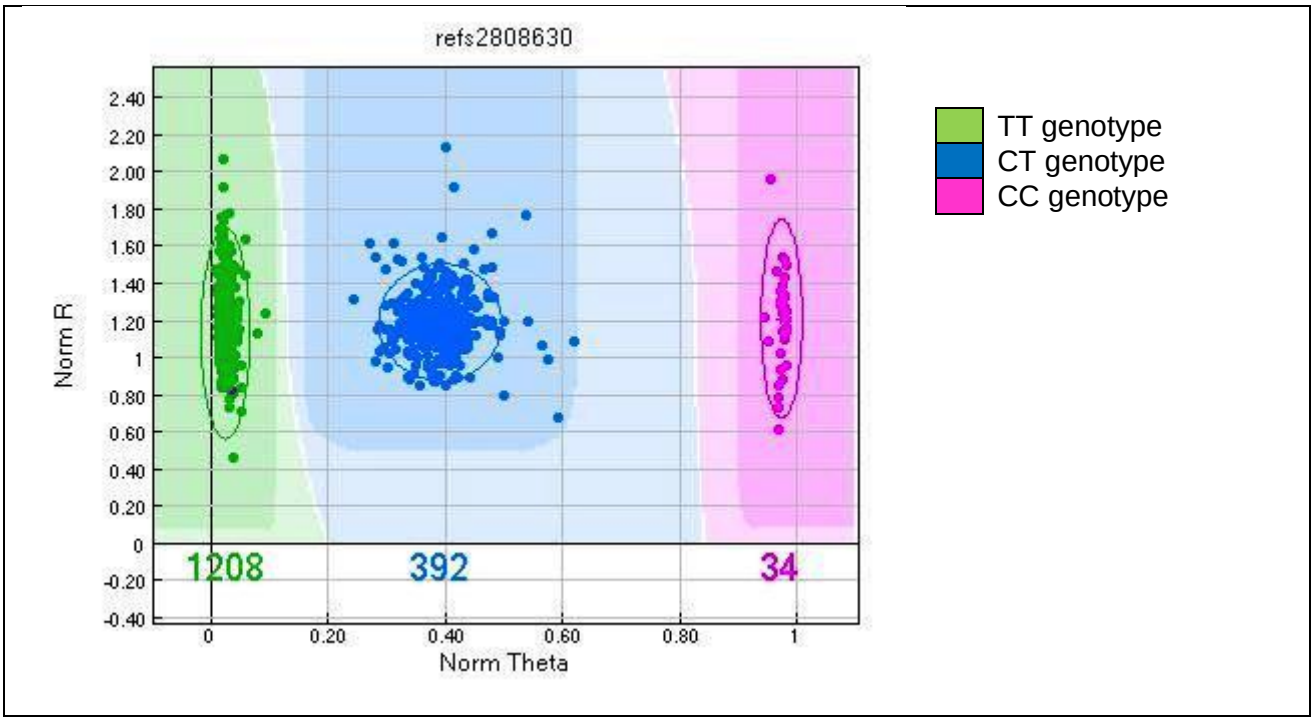
ANCOVA = analysis of covariance; BMI = body mass index; A = adenine; hs-CRP = high sensitivity C-reactive protein; T = thymine; rs = reference sequence

In a study conducted by Suk *et al*. (2005), they reported an incremental decrease in CRP concentrations in individuals harbouring the TT genotype compared to those harbouring the AA genotype. In the current study a decrease in CRP concentrations was observed between the TA genotype and the AA genotype; however, it was not significant. The TT genotype was absent from this population.

4.7.7 SNP rs2808630

The shading call regions for the rs2808630 locus, as determined by the GenomeStudio® program, are represented in Figure 4.28, where the different genotype groups are represented by the different colours. The call rate, the AB R mean value and the AB T mean value were 1, 1.191 and 0.391, respectively. These values indicate that the cluster separation and cluster calls are both reliable and that the samples were correctly called (Section 4.6.2).

Figure 4.28 GenomeStudio® shade call region for SNP rs2808630



C = cytosine; refs = reference sequence; T = thymine; Norm R = Norm intensity

The T allele was determined to be the reference allele for rs2808630. The MAF of the rs2808630 SNP in this population (14%) is similar to the MAF reported in other populations, such as African Americans (9%) and Sub-Saharan Africans (17%), but differs from the MAF reported for a European population (30%) (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=2808630). The variations observed between the different groups with regard to the MAF may be due to ethnic diversity, but could also be due to the different sample size sets used to calculate the MAF for the different reported populations. The MAF from the previously reported populations was calculated from 44 European individuals, 62 Sub-Saharan African individuals and 60 African American individuals. The MAF comparisons should therefore be interpreted with caution (Section 4.7.1). This SNP was also determined to adhere to the assumptions of HWE (Section 3.8) with a p-value of 0.74 for the Chi-square test (Table 4.15).

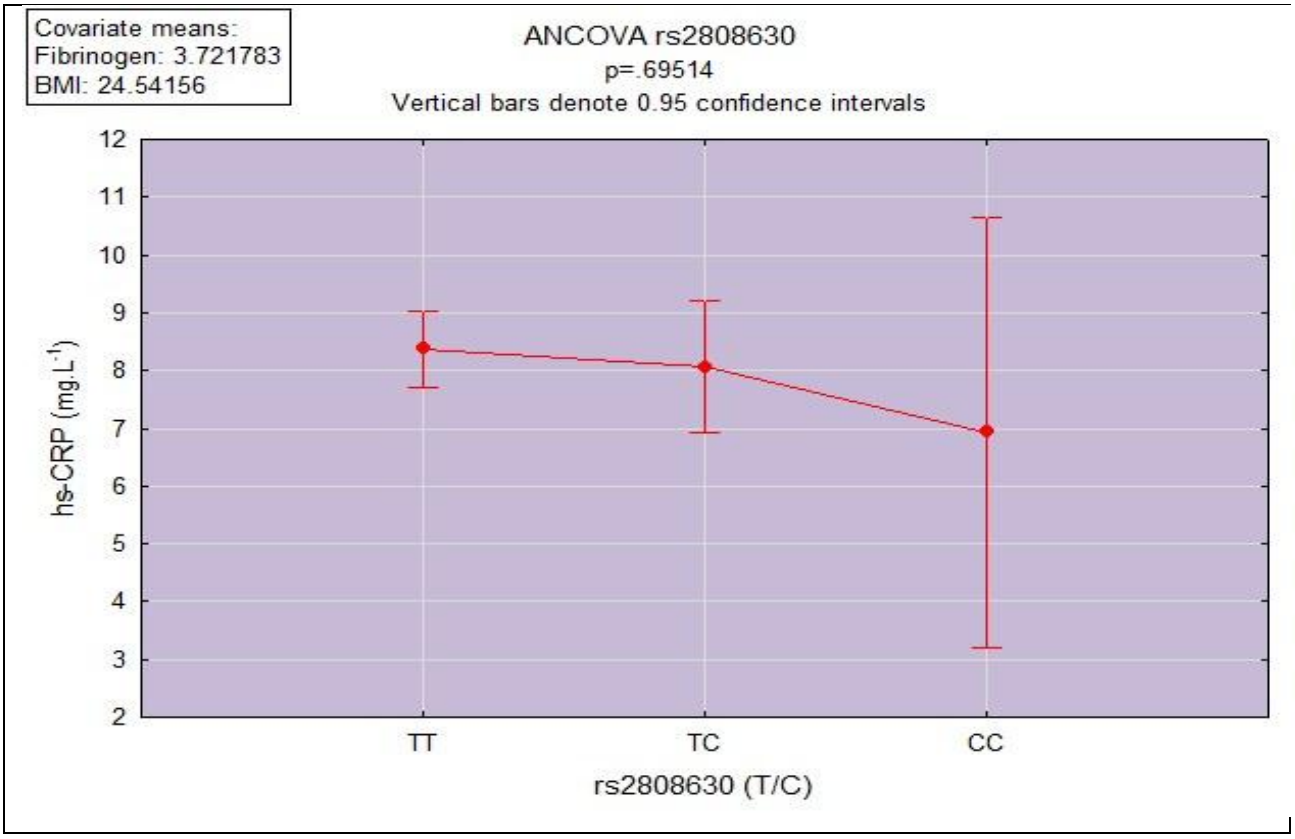
Table 4.15 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects of the rs2808630 SNP with CRP concentrations in the PURE study population

rs2808630 (T/C)							
		TT	**TC**	**CC**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	1208	392	34	1634	0.74	
	Frequency (%)	74	24	2	100		
	Expected numbers	1206.37	395.25	32.37	1634		
	$X^2 = (O - E)^2 / E$	0.00	0.03	0.08	0.11		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	9.14	8.83	7.70	--	0.70	
	Men	10.18	9.21	4.55	--	0.09	0.21*
	Women	8.39	8.51	9.28	--	0.93	
	Rural	8.76	8.76	7.83	--	0.95	0.82*
	Urban	9.87	9.11	7.77	--	0.54	

C = cytosine; E = expected; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; O = observed; rs = reference sequence; T = thymine; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between rs2808630, gender and locality on CRP concentrations; **Men (n)** TT = 445; TC = 130; CC = 13; **Women (n)** TT = 717; TC = 254; CC = 21; **Rural (n)** TT = 588; TC = 162; CC = 18; **Urban (n)** TT = 574; TC = 222; CC = 16

CRP concentrations decreased in individuals harbouring the CC genotype compared to those harbouring the TT genotype when investigating the whole group (Figure 4.29); however, this finding was non-significant. Possible reasons for this could be the small number of individuals harbouring the mutant C allele. This SNP, therefore, is expected to be of no public health relevance in this population. No significant difference was observed between the genders or the rural/urban groups (Table 4.15). No interaction effect was determined between the different genotype classes between the genders and the rural/urban group on the CRP concentrations.

Figure 4.29 CRP concentrations in the three genotype groups of rs2808630 for the whole PURE cohort



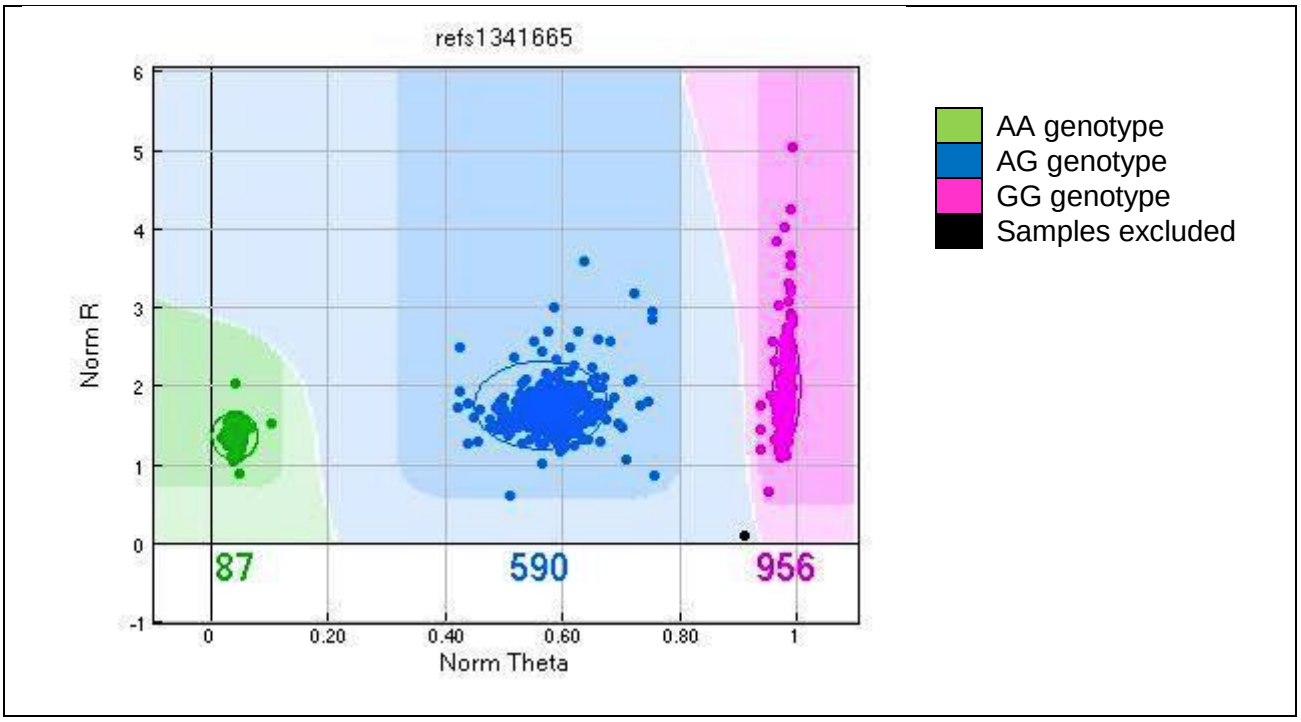
ANCOVA = analysis of covariance; BMI = body mass index; C = cytosine; hs-CRP = high sensitivity C-reactive protein; T = thymine; rs = reference sequence

In a non-hispanic black population it was reported that rs2808630 was associated with a decrease in CRP concentrations (Crawford *et al*., 2006). In this study a decrease in the CRP concentrations was also observed from the TT genotype to the CC genotype; however, this trend was not significant.

4.7.8 SNP rs1341665

The genotype calls were clustered separately for SNP rs1341665 and are represented by the different colours indicated in Figure 4.30. The call rate for this SNP was 0.999, the AB T mean was 0.563 and the AB R mean was 1.761. These indicated good, reliable clustering calls. Samples highlighted in black were excluded because of a GenCall score of less than 0.5 (Section 4.6.2).

Figure 4.30 GenomeStudio® shade call region for SNP rs1341665



A = adenine; G = guanine; refs = reference sequence; Norm R = Norm intensity

For rs1341665, the reference allele was determined to be the G allele. The MAF of rs1341665 did not differ much from that reported in other populations (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1341665), where the MAF in African Americans and Sub-Saharan Africans was 18%, Europeans were at 27% and the MAF in the current population was determined to be 23%.

A Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions was conducted (Table 4.16) to determine whether or not the SNP distribution adhered to the assumptions of HWE. A p-value of 0.74 was determined and indicates that the SNP adheres to all the assumptions of the HWE (Section 3.8).

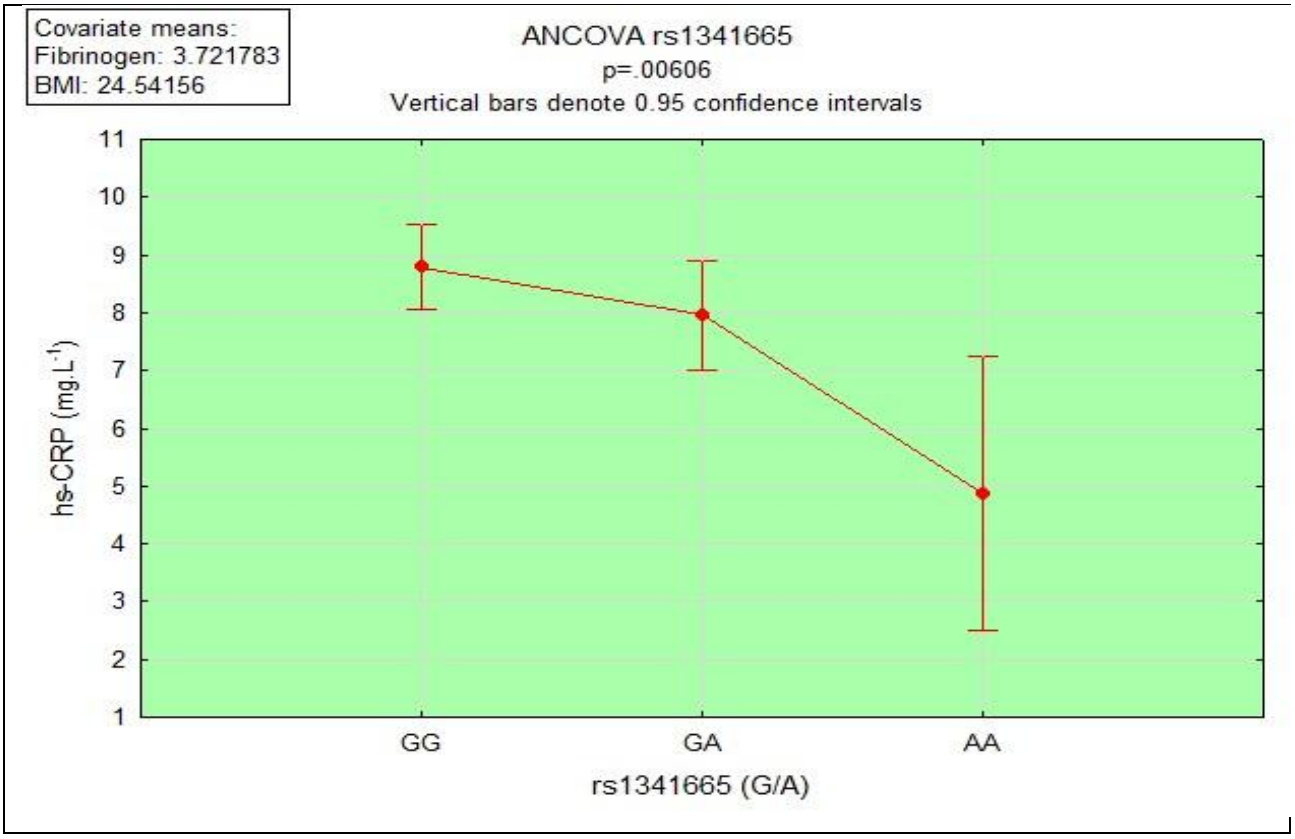
Table 4.16 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects of the rs1341665 SNP with CRP concentrations in the PURE study population

rs1341665 (G/A)							
		GG	**GA**	**AA**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	956	590	87	1633	0.74	
	Frequency (%)	59	36	5	100		
	Expected numbers	958.36	585.28	89.36	1633		
	$X^2 = (O - E)^2 / E$	0.01	0.04	0.06	0.11		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	8.64	7.81	4.72	--	<0.01	
	Men	10.12	7.46	4.55	--	<0.01	0.05*
	Women	7.55	7.82	4.56	--	0.08	
	Rural	8.06	7.57	4.77	--	0.14	0.75*
	Urban	9.24	7.99	4.57	--	0.03	

A = adenine; E = expected; G = guanine; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; O = observed; rs = reference sequence; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between rs1341665, gender and locality on CRP concentrations; **Men (n)** GG = 339; GA = 224; AA = 24; **Women (n)** GG = 583; GA = 348; AA = 61; **Rural (n)** GG = 450; GA = 269 AA = 49; **Urban (n)** GG = 472; GA = 303; AA = 36

The CRP concentrations were significantly decreased across the different genotype groups, with the CRP concentrations in individuals harbouring the homozygote wild-type (GG) being double those of the CRP concentrations in individuals harbouring the homozygote mutant genotype ($p < 0.01$), as illustrated in Figure 4.31. The men ($p < 0.01$) and the urban group ($p = 0.03$) presented with a significant decrease in CRP concentrations in the AA genotype group (Table 4.16). An interaction effect was also observed between the gender groups and the genotype classes with regard to CRP concentrations. Men and urban individuals harboured this CRP lowering SNP together with other *CRP* SNPs which have shown increases in CRP concentrations and are discussed in detail in Section 4.9.

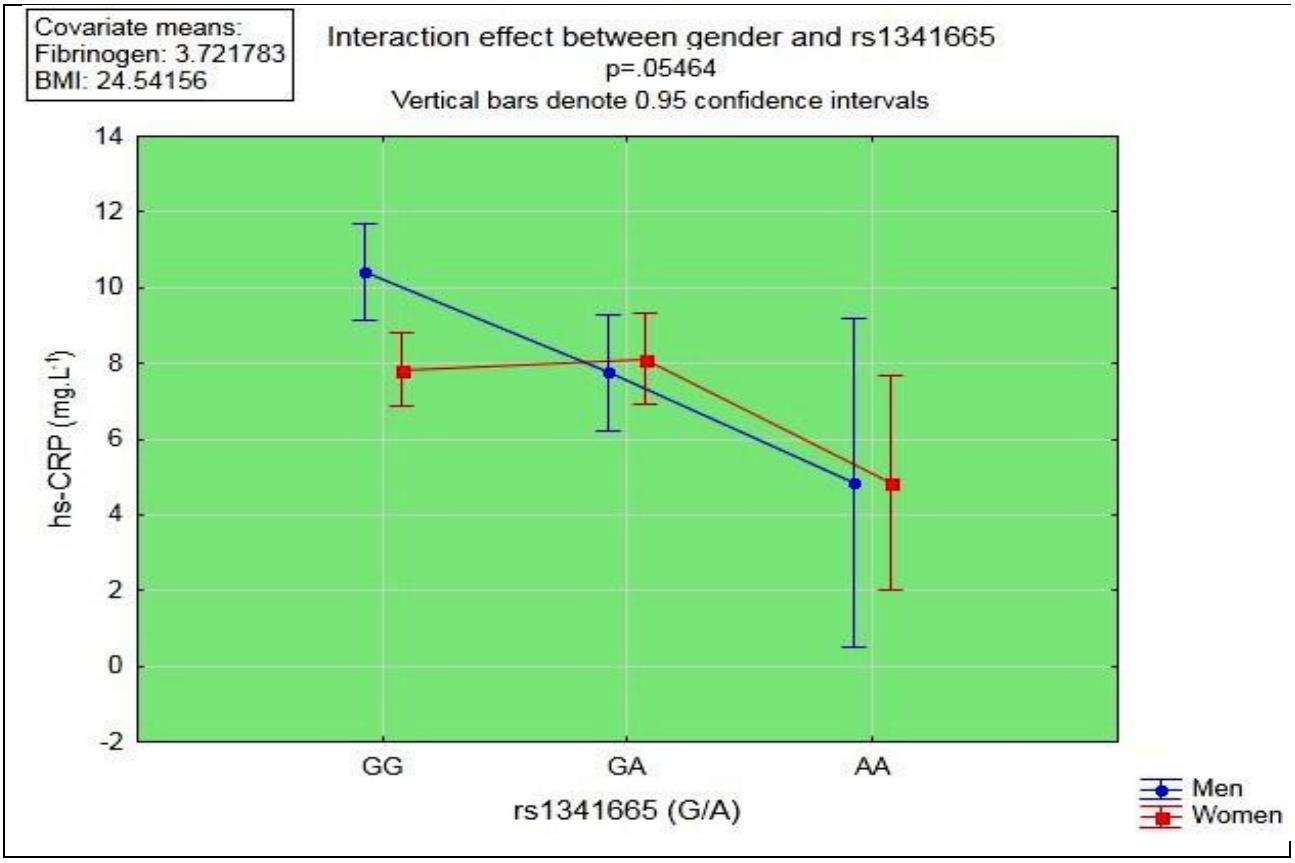
Figure 4.31 CRP concentrations in the three genotype groups of rs1341665 for the whole PURE cohort



A = adenine; ANCOVA = analysis of covariance; BMI = body mass index; G = guanine; hs-CRP = high sensitivity C-reactive protein; rs = reference sequence

Among individuals harbouring the homozygous wild-type genotype, women had a lower CRP concentration than men harbouring the same genotype (Figure 4.32). With the addition of the mutant allele women had higher CRP concentrations than men. Furthermore, it is important to bear in mind that a single SNP does not determine the CRP phenotype.

Figure 4.32 The effect of the interaction between genotype and gender on CRP concentrations for rs1341665



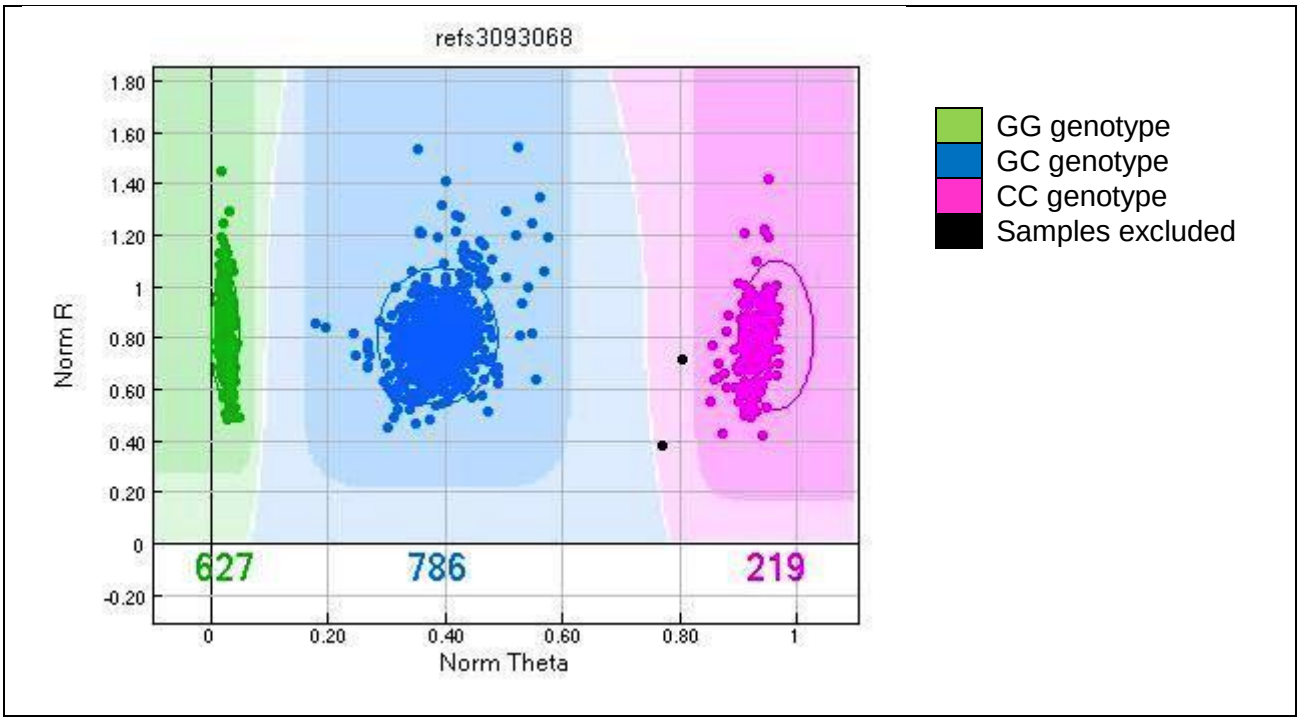
A = adenine; BMI = body mass index; G = guanine; hs-CRP = high sensitivity C-reactive protein; rs = reference sequence

SNP rs1341665 was significantly associated with a decrease in CRP concentrations in a study conducted by Wang *et al*. (2006) as well as a more recent study by Doumatey *et al*. (2012). These observations are in line with those of the current study, where significant associations were also observed between the genotypes of rs1341665 and CRP concentrations. In the current population, the frequency of the mutant allele was 23% and therefore the observed association between this allele and CRP concentrations could possibly be of significance at a public health level.

4.7.9 SNP rs3093068

Figure 4.33 represents the call region shading of the rs3093068 SNP as determined through the GenomeStudio® program. The call rate for this SNP was 0.999. The samples indicated in black in Figure 4.33 were excluded because their GenCall score was < 0.5. The AB R mean of this particular SNP was 0.806, indicating the intensity of the data to be reliable. The AB T mean value of this SNP was 0.386, indicating a reliable clustering call (Section 4.6.2).

Figure 4.33 GenomeStudio® shade call region for SNP rs3093068



C = cytosine; G = guanine; refs = reference sequence; Norm R = Norm intensity

The ancestral allele for SNP rs3093068 was reported to be the C allele. This population reported a higher MAF of 62% in comparison with Europeans and African Americans, with a MAF of 8.4% and 31%, respectively (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=3093068), but it is comparable to the MAF of 50% previously reported for the Sub-Saharan population. As previously discussed, the sample size used to determine the MAF for the different ethnic groups was small. Therefore, there is a chance that the small sample is not fully representative of the larger population from which the sample was taken and the MAF that was determined cannot be extrapolated to the whole population.

The Chi-square test was conducted to ensure that the population adhered to the assumptions of HWE (Table 4.17). The p-value of the Chi-square test was determined to be non-significant ($p = 0.27$), suggesting that the population adheres to the assumptions of HWE at the rs3093068 locus and it is, among others, most likely that the population is undergoing random mating, no inbreeding and no migration, as discussed in Section 3.8.

Table 4.17 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects of the rs3093068 SNP with CRP concentrations in the PURE study population

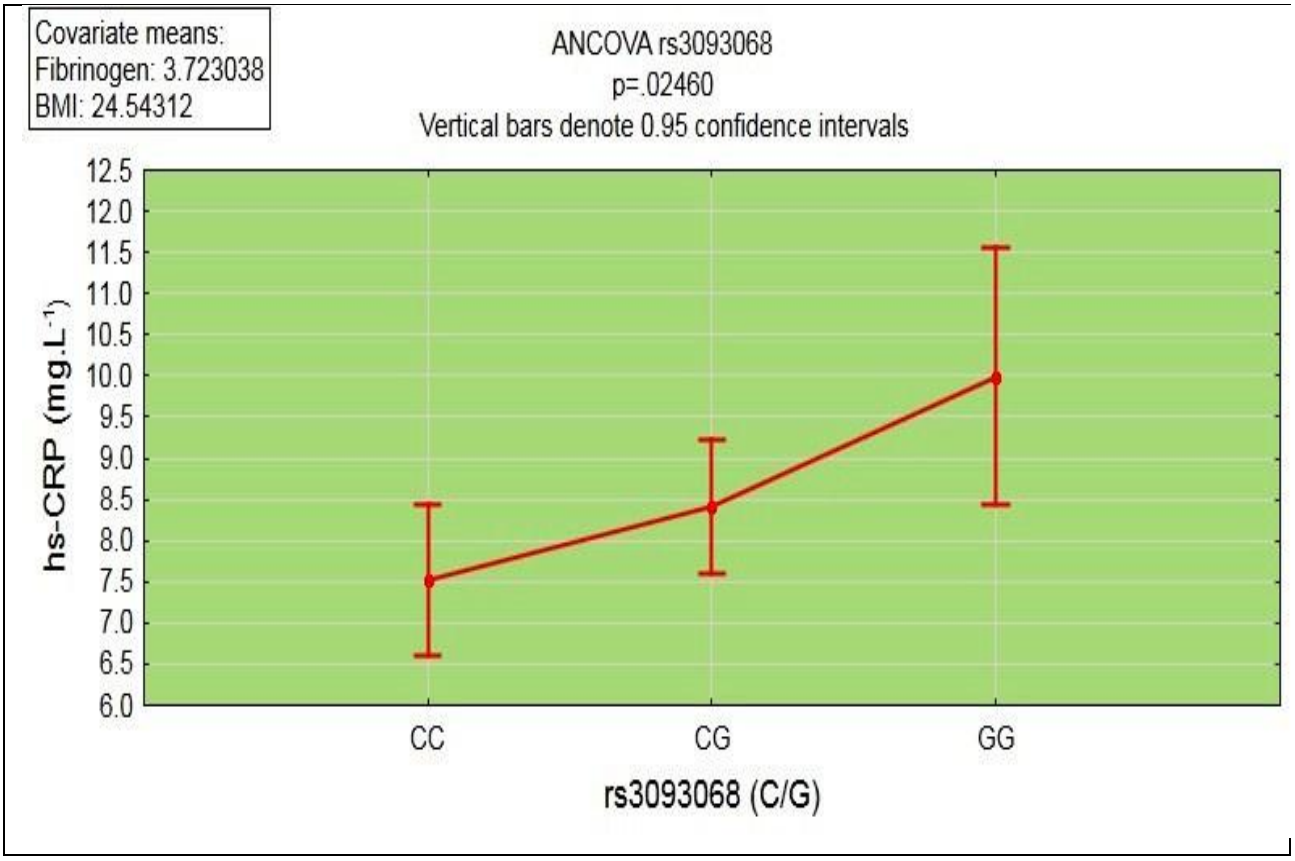
rs3093068 (C/G)							
		CC	**CG**	**GG**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	219	786	627	1632	0.27	
	Frequency (%)	13	48	38	100		
	Expected numbers	229.50	765.00	637.50	1632		
	$X^2 = (O - E)^2 / E$	0.48	0.58	0.17	1.23		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	7.62	8.50	10.09	--	0.02	
	Men	7.58	9.46	11.50	--	0.04	0.32*
	Women	7.31	7.52	8.97	--	0.26	
	Rural	8.02	7.60	8.89	--	0.58	0.04*
	Urban	7.14	9.44	11.66	--	<0.01	

C = cytosine; E = expected; G = guanine; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; O = observed; rs = reference sequence; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between, gender and locality on CRP concentrations; **Men (n)** CC = 78; CG = 296; GG = 214; **Women (n)** CC = 130; CG = 470; GG = 390; **Rural (n)** CC = 292; CG = 367; GG = 292; **Urban (n)** CC = 100; CG = 399 GG = 312

Figure 4.34 is a graphic representation of the CRP concentrations in the different genotype groups at the rs3093068 locus for the whole group, adjusting for fibrinogen and BMI. A significant difference was observed between the genotype groups as indicated in Table 4.17, where the greatest increase in CRP concentrations was observed between the CG genotype and the GG genotype ($p = 0.02$). The CRP concentrations of the rural and female subjects did not differ significantly between the different genotype subgroups, which was in contrast to the urban group ($p < 0.01$) and the male ($p = 0.04$) subjects harbouring these genotypes, which had significantly higher CRP concentrations (discussed in detail in Section 4.9).

Once again, one could argue that when residing in an urban community and adopting the concomitant lifestyle, the genotype harboured has a greater influence on CRP concentrations, but when residing in a rural community it makes no difference to the CRP phenotype. Individuals residing in the different communities are exposed to different factors, which can influence the expression of the different genotypes and therefore have an influence on the CRP concentrations (Section 4.9).

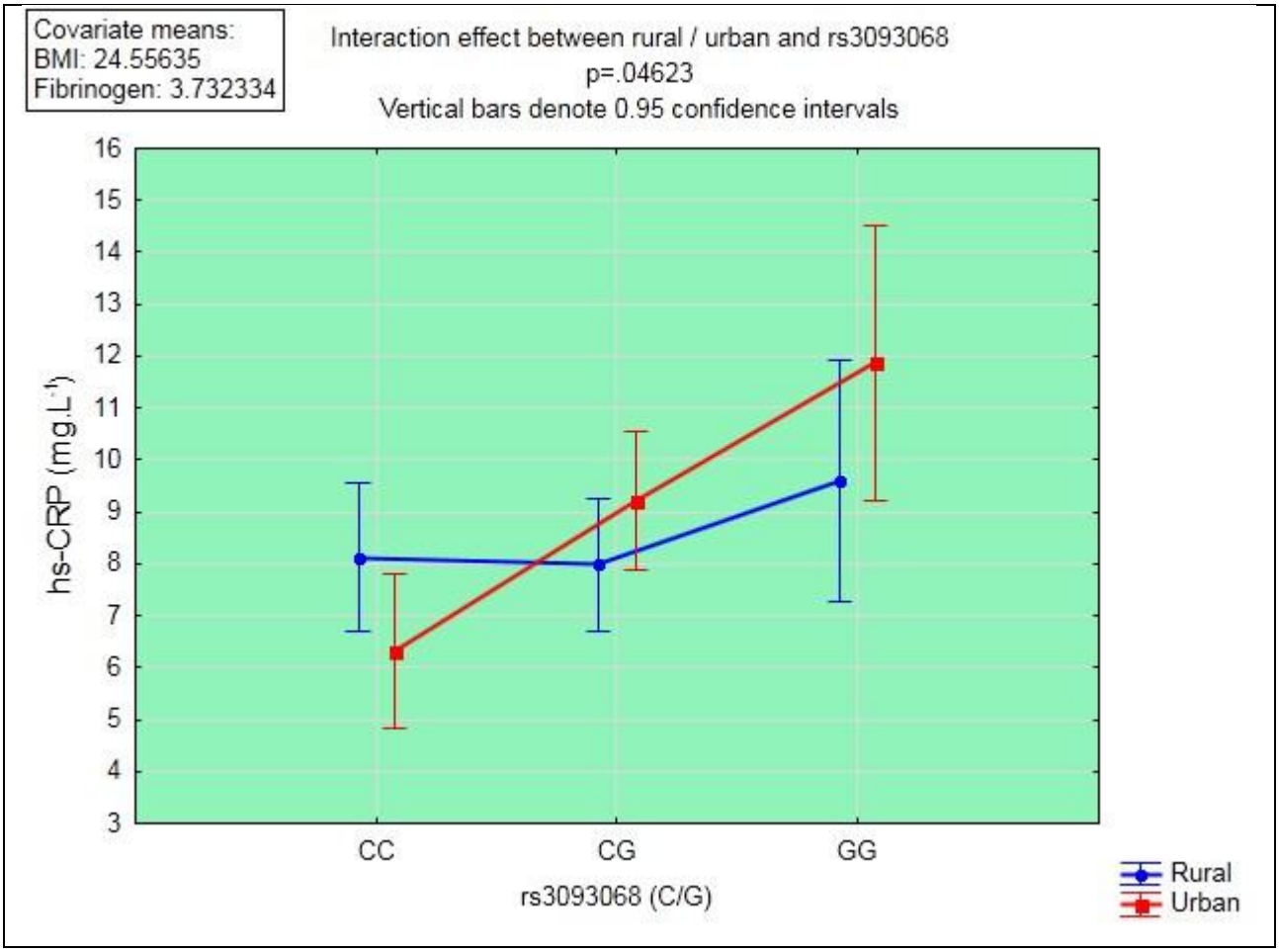
Figure 4.34 CRP concentrations in the three genotype groups of rs3093068 for the whole PURE cohort



ANCOVA = analysis of covariance; BMI = body mass index; C = cytosine; G = guanine; CRP = high sensitivity C=reactive protein; rs = reference sequence

Figure 4.35 represents a significant interaction effect between the rural and urban groups in the three genotype classes (p = 0.04). Among individuals harbouring the CC genotype, those who resided in the urban community had a lower CRP concentration than those residing in the rural community, but for individuals harbouring either the CG or GG genotype, the urban group showed higher CRP concentrations than the individuals in the rural group who carried the same genotype. This could possibly be a gene-environment interaction where people in the different residential areas have different lifestyles. This is discussed in Section 4.9. No interaction effect was observed between subjects in the different gender classes (p = 0.32).

Figure 4.35 The effect of the interaction between genotype and location on CRP concentrations for rs3093068



BMI = body mass index; c = cytosine; g = guanine; hs-CRP = high sensitivity C-reactive protein

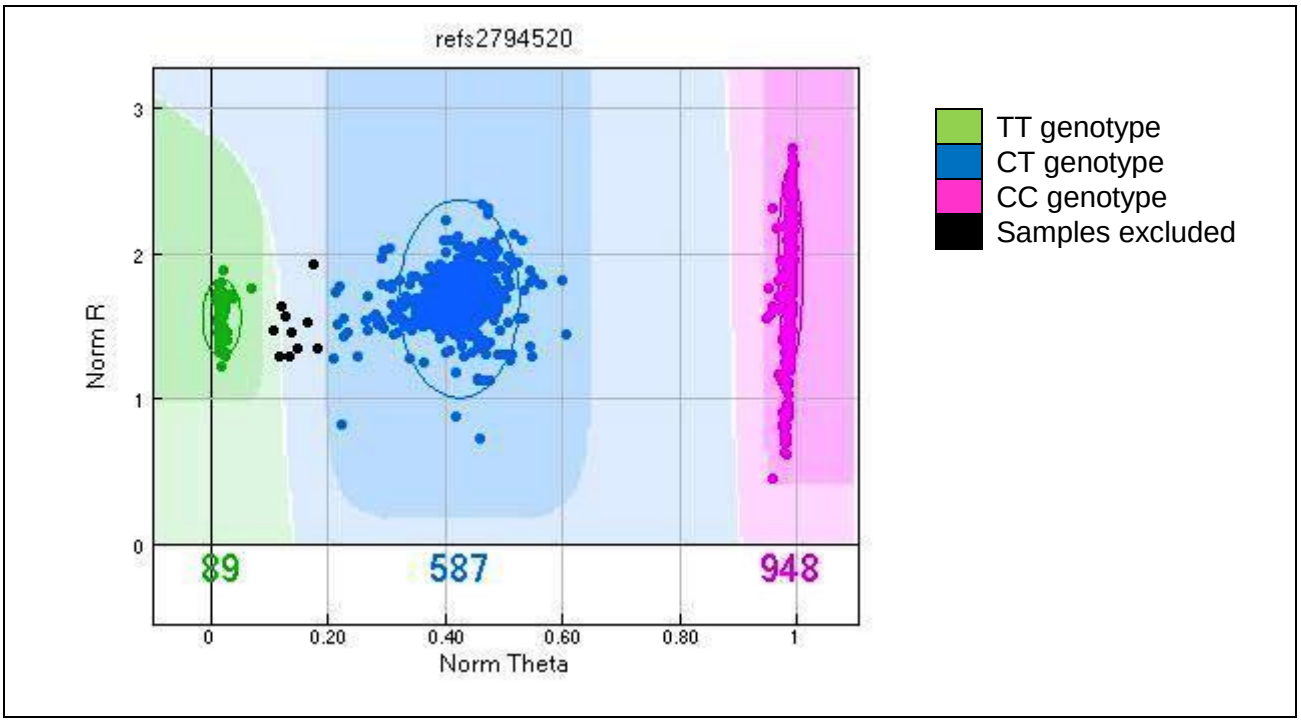
Lee *et al*. (2009) reported an association ($p < 0.05$) with increased CRP concentrations in the different genotype groups (combination of caucasian, black, Hispanic and Asian population) and Kardys *et al*. (2007) reported an allele dose effect for SNP rs3093068, where the CRP concentrations increased with the addition of the mutant G allele in Europeans. However, significant differences were only observed between the individuals harbouring the homozygote CC genotype and those harbouring the heterozygote CG genotype. This is similar to that which is observed in the current population regarding SNP rs3093068. The frequency of the mutant allele in this population was 62%, which implies that it could be of public health significance and could in future possibly be used as a genetic screening tool within this population with regard to CRP concentrations.

4.7.10 SNP rs2794520

Figure 4.36 represents the genotype groups for rs2794520 as illustrated in the GenomeStudio® program. Some samples (indicated in black) were excluded because of a

GenCall score of below 0.5. The call rate for the rs2794520 SNP was 0.994, the AB R mean value was 1.691 and the AB T mean value was 0.423, indicating a good and reliable SNP (Section 4.6.2).

Figure 4.36 GenomeStudio® shade call region for SNP rs2794520



C = cytosine; refs = reference sequence; T = thymine; Norm R = Norm intensity

The reference allele for rs2794520 is the C allele. The MAF for the rs2794520 SNP in this cohort (24%) was similar to that reported in Sub-Saharan Africans (17%) as well as the African American population (17%), but was determined to be higher in the European population which had a MAF of 34% (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=2794520). This again highlights the effect that a more representative sample set could have on extrapolated values, as well as the diversity among different populations. This SNP was determined to adhere to the assumptions of the HWE with a p-value of 0.88 for the Chi-square test (Section 3.8).

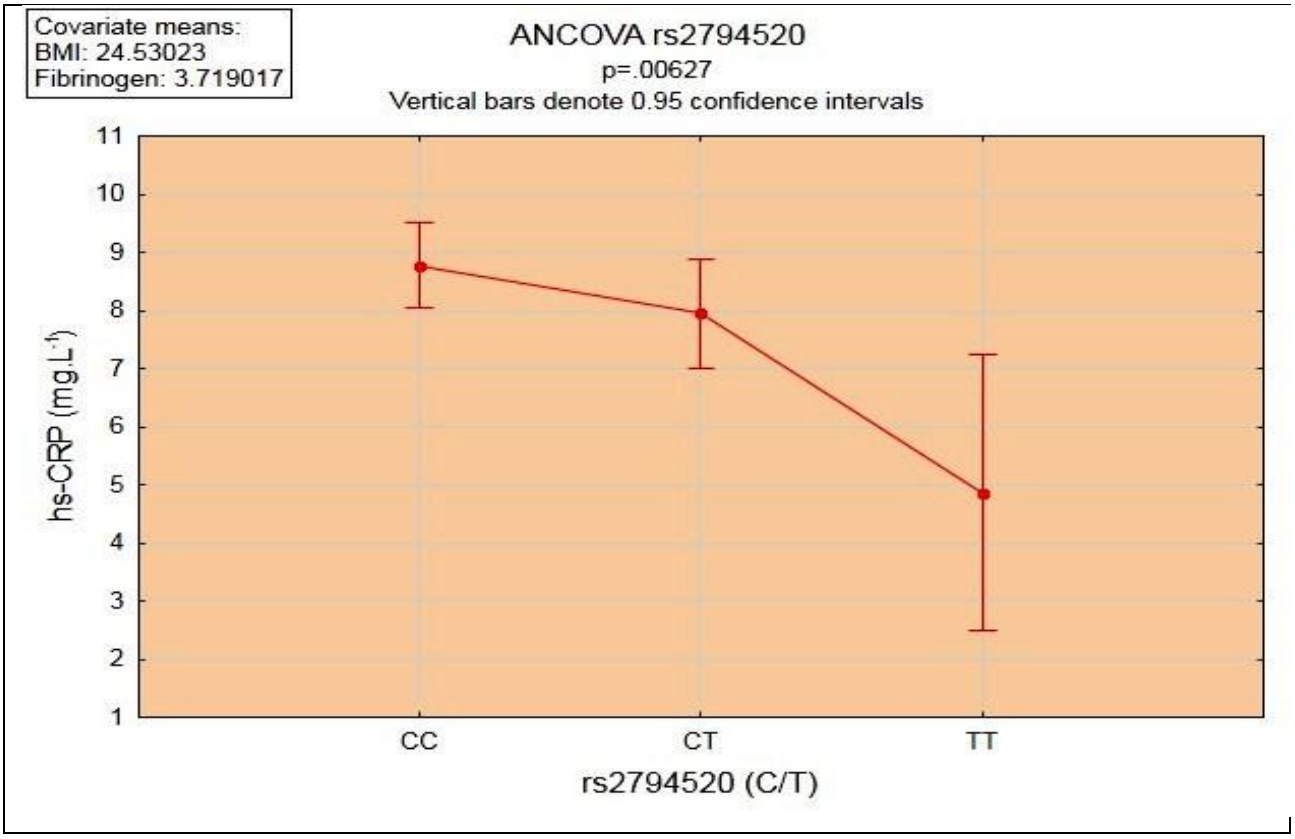
Table 4.18 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects of the rs2794520 SNPs with CRP concentrations in the PURE study population

rs2794520 (C/T)							
		CC	**CT**	**TT**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	948	587	89	1624	0.88	
	Frequency (%)	58	36	5	100		
	Expected numbers	949.09	584.82	90.09	1624		
	$X^2 = (O - E)^2 / E$	0.00	0.01	0.01	0.02		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	8.63	7.80	4.71	--	<0.01	
	Men	10.12	7.47	4.54	--	0.01	0.06*
	Women	7.74	7.80	4.55	--	0.09	
	Rural	8.06	7.53	4.76	--	0.13	0.78*
	Urban	9.21	8.01	4.56	--	0.03	

C = cytosine; E = expected; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; O = observed; rs = reference sequence; T = thymine; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between rs2794520, gender and locality on CRP concentrations; **Men (n)** CC = 334; CT = 224; TT = 24; **Women (n)** CC = 581; CT = 345; TT = 62; **Rural (n)** CC = 449; CT = 268; TT = 49; **Urban (n)** CC = 466; CT = 301; TT = 37

CRP concentrations decreased significantly ($p < 0.01$) across the genotype groups (Figure 4.37). This SNP seems to present with a protective effect in that it decreases CRP concentrations. The largest decrease in CRP concentrations was observed when comparing subjects harbouring the homozygous wild-type (CC) genotype and those harbouring the homozygous mutant genotype (TT). However, the TT genotype's distribution within this population was 5% and that of the CC genotype was 58%. Therefore, the public health significance with regard to the lowered disease risk as a result of the lowered CRP concentrations in individuals harbouring the TT is expected to be very small (Table 4.18). Again it was observed that men ($p = 0.01$) and individuals in the urban group ($p = 0.03$) showed significant decreases in CRP concentrations with the addition of the mutant allele, whereas the women and rural group indicated no significant differences. Possible reasons for men and urban individuals harbouring SNPs which increases and decreases CRP concentrations are discussed in Section 4.9. No interaction effects were observed between the genders and the rural/urban groups.

Figure 4.37 CRP concentrations in the three genotype groups of rs2794520 for the whole PURE cohort



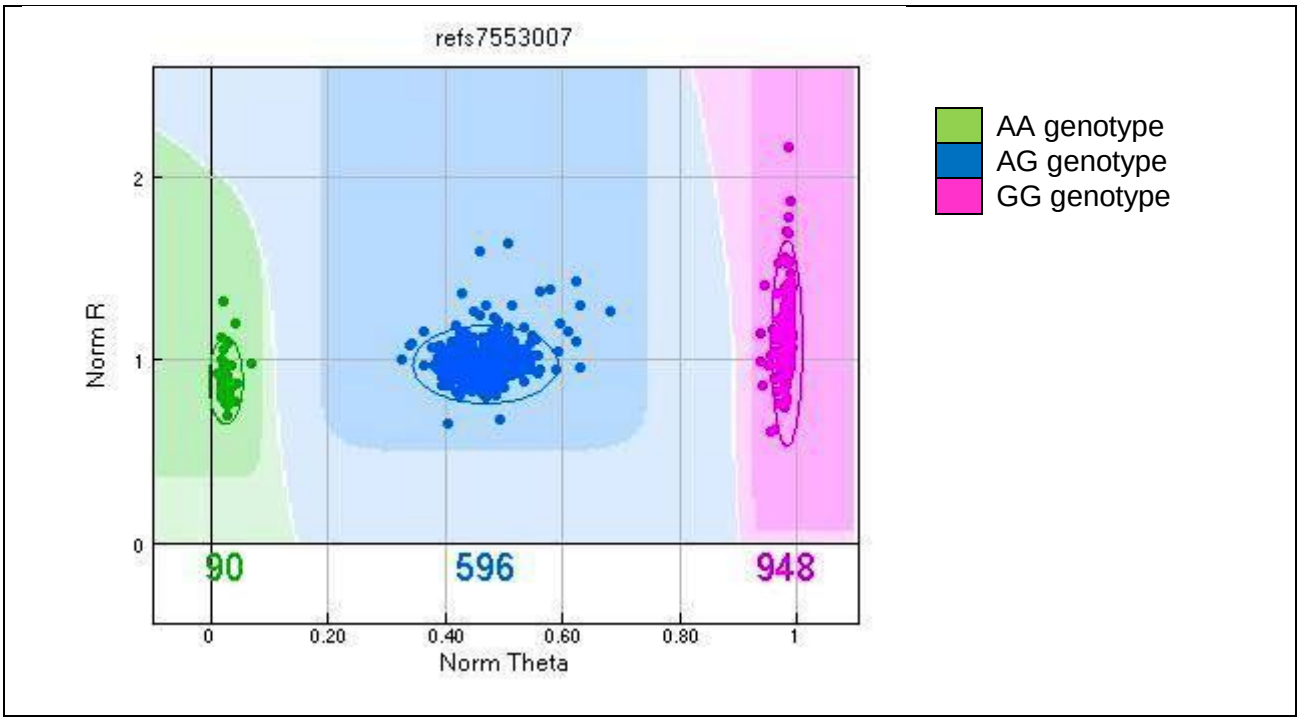
ANCOVA = analysis of covariance; BMI – body mass index; C = cytosine; hs-CRP = high sensitivity C-reactive protein; rs = reference sequence

Rhodes and co-workers (2008) reported a significant reduction in CRP concentrations in a Caucasian population with the addition of the rare allele (T). The results the researcher obtained in this black South African population were similar to those reported by Rhodes *et al*. (2008) for individuals of caucasian descent.

4.7.11 SNP rs7553007

Figure 4.38 illustrates the genotype distribution of rs7553007 as determined with the GenomeStudio® program. The call rate was 1, the AB R mean was 0.977 and the AB T mean was 0.468, indicating good reliable clustering calls (Section 4.6.2).

Figure 4.38 GenomeStudio® shade call region for SNP rs7553007



A = adenine; G = guanine; refs = reference sequence; Norm R = Norm intensity

The reference allele for SNP rs7553007 was determined to be the G allele. The MAF of the black South African population (24%) is in line with that of other previously reported populations, as the European population had an MAF of 37% and the Sub-Saharan African population had an MAF of 20% (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=7553007). To date, no MAF has been reported in the African American population.

The Chi-square test was conducted for the rs7553007 SNP to ensure that this SNP adheres to all the assumptions of the HWE. A p-value of 0.77 was determined and it was therefore determined that the SNP adheres to the assumptions of HWE.

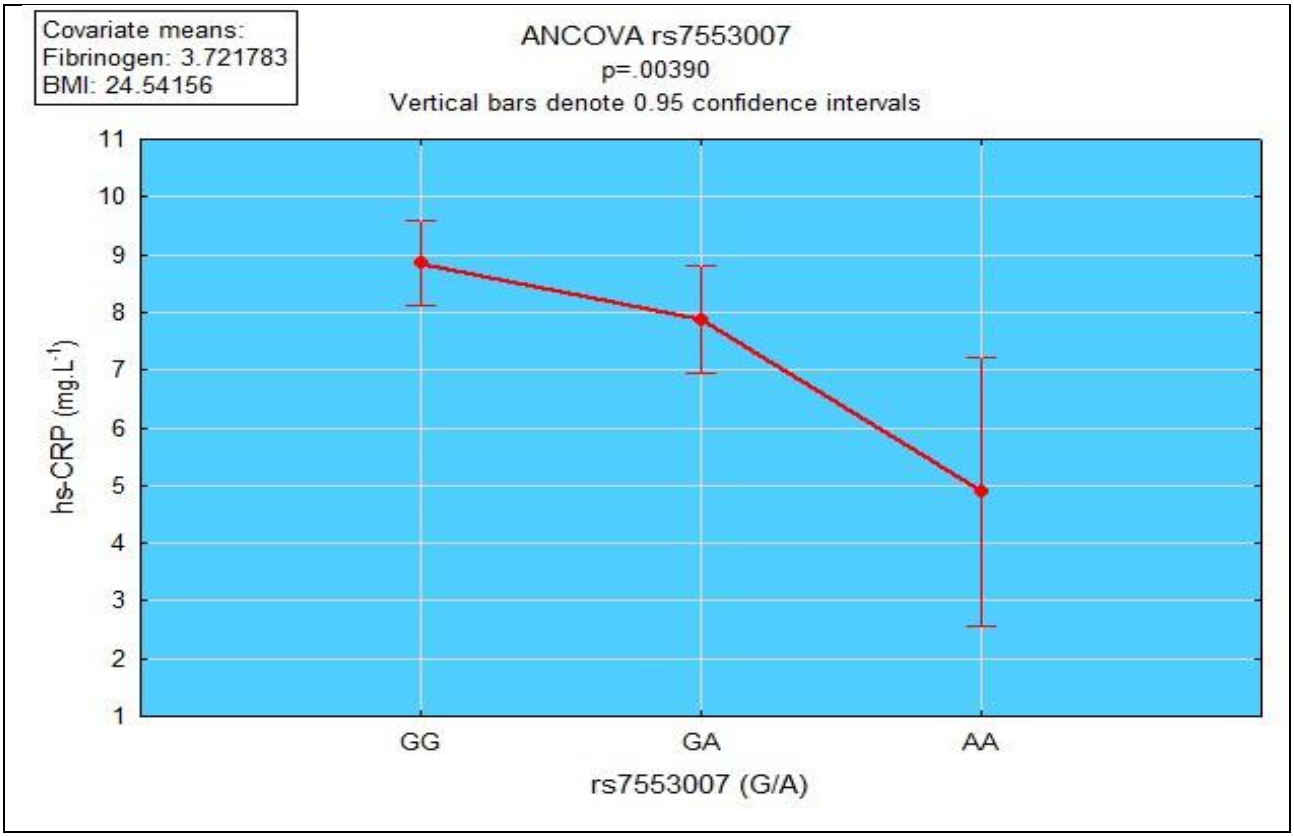
Table 4.19 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects with CRP concentrations of the rs7553007 locus in the PURE study population

rs7553007 (G/A)							
		GG	**GA**	**AA**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	948	596	90	1634	0.77	
	Frequency (%)	58	36	6	100		
	Expected numbers	950.13	591.74	92.13	1634		
	$X^2 = (O - E)^2 / E$	0.00	0.03	0.05	0.08		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	8.67	7.70	4.71	--	<0.01	
	Men	10.08	7.44	4.49	--	<0.01	0.08*
	Women	7.60	7.65	4.57	--	0.08	
	Rural	8.09	7.44	4.69	--	0.11	0.78*
	Urban	9.26	7.89	4.66	--	0.03	

A = adenine; E = expected ;G = guanine; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; O = observed; rs = reference sequence; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between rs1341665, gender and locality on CRP concentrations; **Men (n)** GG = 340; GA = 223; AA = 25; **Women (n)** GG = 575; GA = 354; AA = 63; **Rural (n)** GG = 445; GA = 273; AA = 50; **Urban (n)** GG = 470; GA = 304; AA = 38

As can be seen in Table 4.19 and Figure 4.39, the whole group ($p < 0.01$), men ($p < 0.01$) as well as the urban group ($p = 0.03$) presented with a significant decrease in CRP concentrations in individuals harbouring the wild-type (GG) homozygote genotype compared to those harbouring the variant homozygote AA genotype. Once again it seems that when residing in an urban community and adopting the related lifestyle, the genotype harboured, for this particular SNP, has an influence on CRP concentration, but when residing in a rural community it makes no difference to the CRP phenotype. Males presented with significant differences in CRP concentrations between the different genotype groups, but females presented with no significant difference. Possible reasons for this are discussed in Section 4.9.

Figure 4.39 CRP concentrations in the three genotype groups of rs7553007 for the whole PURE cohort



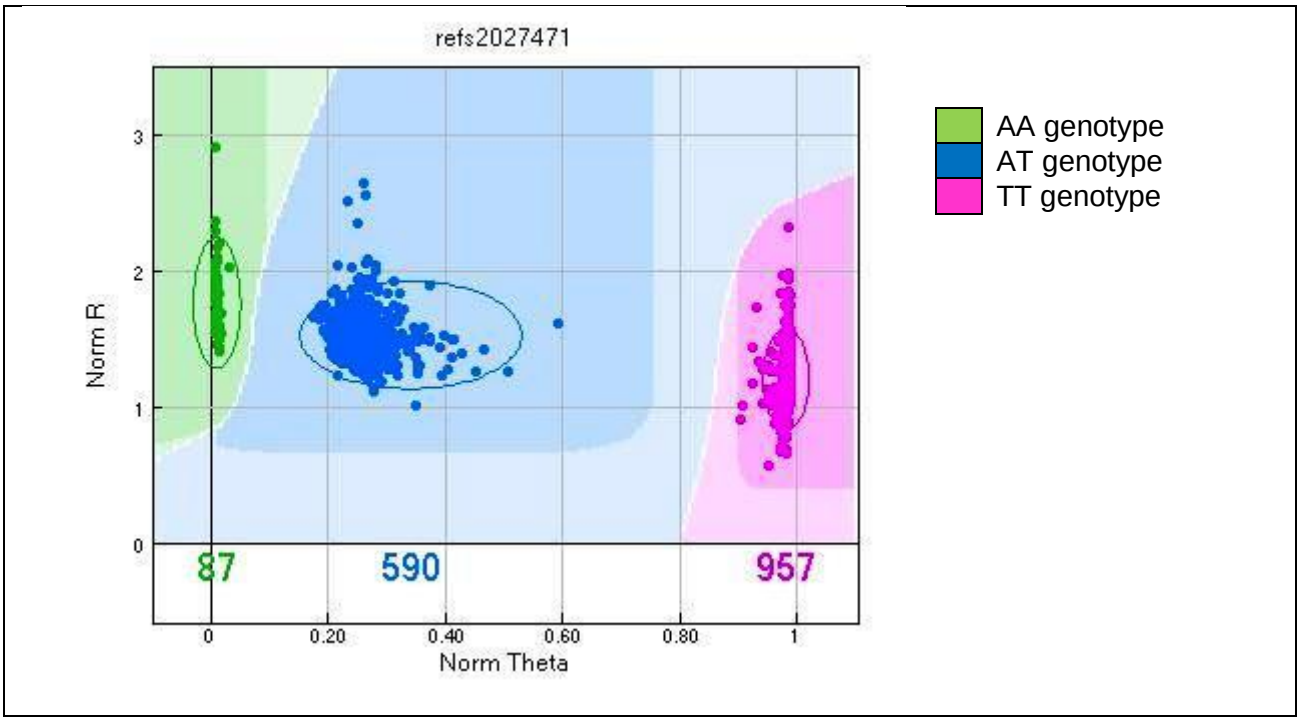
A = adenine; ANCOVA = analysis of covariance; BMI = body mass index; G = guanine; hs-CRP = high sensitivity C-reactive protein; rs = reference sequence

Three studies reported similar results to those of the current investigation regarding the rs7553007 SNP. In a study conducted by Elliot *et al*. (2009), CRP concentrations were lowered by 21% per minor allele harboured. The same was reported by Rhodes *et al*. (2008) and a more recent report by Reiner *et al*. (2012). The study populations that were investigated included African Americans and Hispanic Americans. This SNP is in accordance with that reported in literature.

4.7.12 SNP rs2027471

As mentioned in all the previous sections, the GenomeStudio® program was used to illustrate the genotypic data visually, as indicated in Figure 4.40, with the different genotype groups being represented by the different colours. The call rate for the SNP was 1.00, the AB R mean value was 1.531 and the AB T mean value was 0.341. These values all indicate a reliable and good clustering call for this SNP (Section 4.6.2).

Figure 4.40 GenomeStudio® shade call region for SNP rs2027471



A = adenine; refs = reference sequence; T = thymine; Norm R = Norm intensity

In the population in the current study, the present SNP was determined to meet the assumptions of HWE (p-value = 0.75), as discussed in Section 3.8. The ancestral allele for rs2027471 was determined to be the T allele. The MAF of rs2027471 was reported to be 77% in the population under investigation, which is not in accordance with the other populations, as Europeans presented with a MAF of 34% and the Sub-Saharan African population with an MAF of 18%. The African American population reported no MAF for this SNP (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=2027471). The variation that is observed between the MAF in the current population when compared to the other ethnic groups, could possibly be due to factors that are reported to influence the genetic make-up of the different ethnic groups. Table 4.20 is a summary of the results for SNP rs2027471.

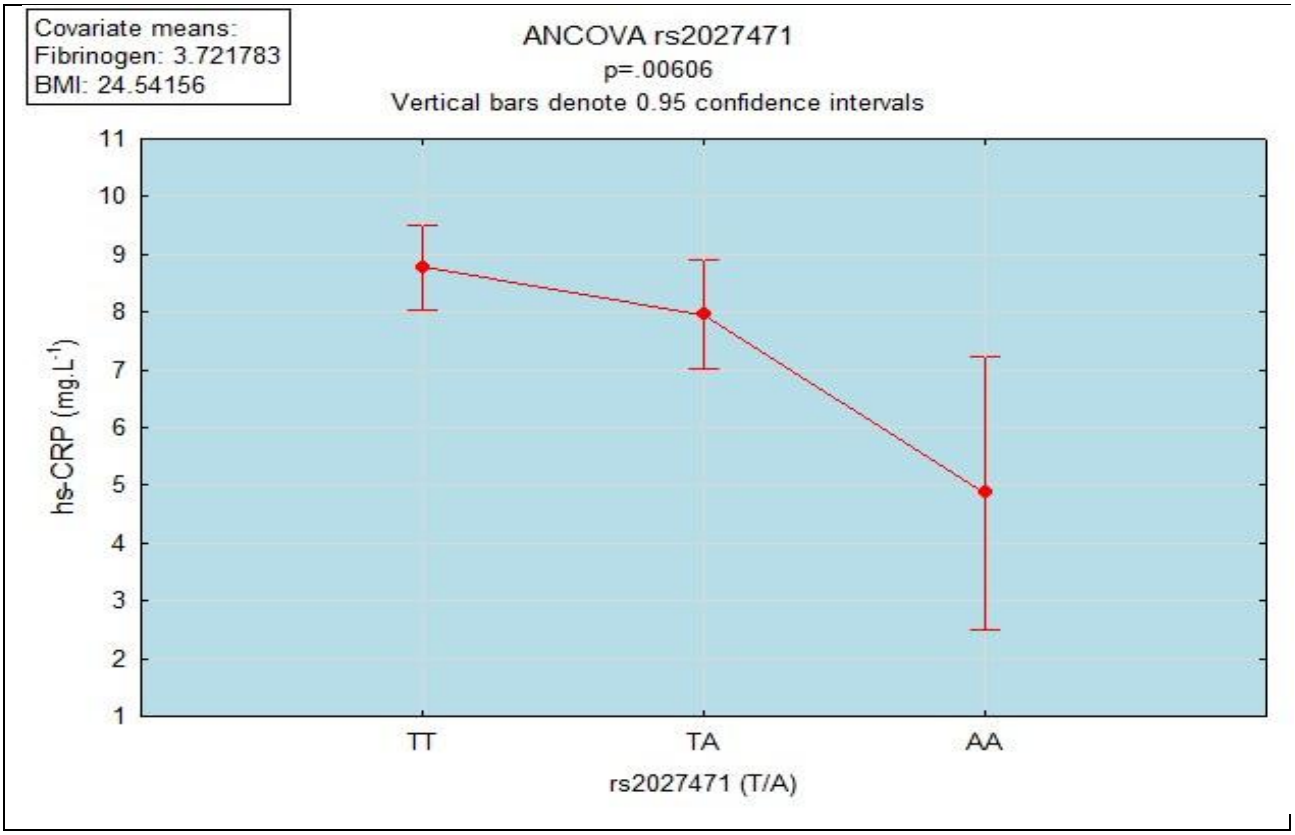
Table 4.20 Determination of adherence to Hardy-Weinberg equilibrium and genetic associations as well as interaction effects of the rs2027471 SNP with CRP concentrations in the PURE study population

rs2027471 (T/A)							
		TT	**TA**	**AA**	**Total**	**p-value**	
Chi-square test for goodness-of-fit to the Hardy-Weinberg proportions	Observed numbers	957	590	87	1634	0.75	
	Frequency (%)	59	36	5	100		
	Expected numbers	959.30	585.39	89.30	1634		
	$X^2 = (O - E)^2 / E$	0.01	0.04	0.06	0.11		
Adjusted means for CRP concentrations ($mg.L^{-1}$)	Whole group	9.36	7.50	4.50	--	<0.01	
	Men	10.12	7.46	4.55	--	<0.01	0.05*
	Women	7.55	7.82	4.56	--	0.09	
	Rural	8.06	7.57	4.77	--	0.14	0.75*
	Urban	9.24	7.99	4.57	--	0.03	

A = adenine; E = expected; CRP = C-reactive protein; $mg.L^{-1}$ = milligrams per litre; O = observed; rs = reference sequence; T = thymine; X^2 = Chi-square value; * = indicates the p-values of the interaction effect between rs1341665, gender and locality on CRP concentrations; **Men (n)** TT = 340; TA = 224; AA = 24; **Women (n)** TT = 583; TA = 348; AA = 61; **Rural (n)** TT = 450; TA = 269; AA = 49; **Urban (n)** TT = 473; TA = 303; AA = 36

Figure 4.41 represents the CRP concentrations in the different genotype groups for the rs2024741 SNP. A significant ($p < 0.01$) decrease in the CRP concentrations was observed between the different genotype groups among the whole group. In the men's group ($p < 0.01$) as well as in the urban group ($p = 0.03$) a significant difference was also observed regarding the decrease in CRP concentrations across the different genotype classes (Table 4.20). A significant interaction effect was also observed between the SNP and gender (Figure 4.42). As with most of the SNPs discussed, it seems that when residing in an urban setting and adopting the concomitant lifestyle, the genotype harboured has an effect on the CRP concentrations as well as the fact that CRP decreasing and increasing SNPs are harboured by urban individuals. This is discussed in detail in Section 4.9.

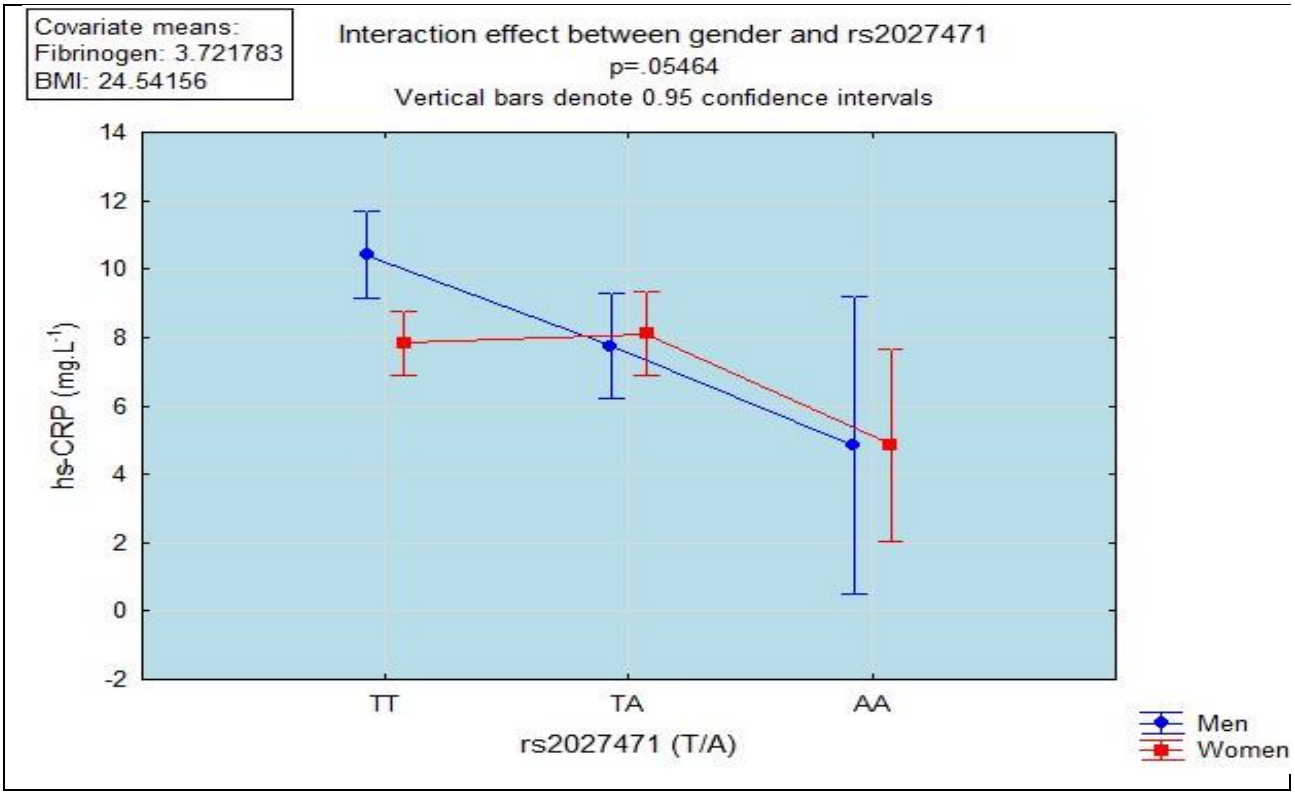
Figure 4.41 CRP concentrations in the three genotype groups of rs2027471 for the whole PURE group



A = adenine; ANCOVA = analysis of covariance; BMI = body mass index; rs = reference sequence; T = thymine; rs = reference sequence

Among individuals harbouring the TT genotype, men had higher CRP concentrations than women. When harbouring the AT genotype, women had a higher CRP concentration than men (Figure 4.42). Other factors could also cause women to have higher CRP concentrations than those of men for this specific SNP. One could speculate that the different gender hormones might have an influence (discussed in Section 4.9).

Figure 4.42 The effect of the interaction between genotype and gender on CRP concentrations for rs2027471



A = adenine; BMI = body mass index; rs = reference sequence; T = thymine

In accordance with the current study's findings, Rhodes *et al*. (2008) reported a decrease in CRP concentrations associated with the rare A allele. In this study, the largest decrease in CRP concentrations was observed between the TT genotype and the AA genotype; however, the AA genotype had a presence of 5% in this population. This SNP is, therefore, not of significance to public health in this population.

4.8 ASSOCIATIONS AND INTERACTION EFFECTS OF CRP CONCENTRATIONS WITH SNPS, EXCLUDING INDIVIDUALS WITH POSSIBLE ACUTE INFLAMMATION

As mentioned in Section 3.8, the researcher in addition reanalysed the significant results and excluded those individuals with a CRP concentration of above 10 $mg.L^{-1}$, as these individuals could have had increased inflammation due to infection (Clyne & Olshaker, 1999). Table 4.21 represents the results of all the SNPs and their association with CRP concentrations in the cohort investigated, excluding the individuals with a CRP of higher than 10 $mg.L^{-1}$.

Table 4.21 Genetic associations of *CRP* SNPs with CRP concentrations, excluding individuals with CRP concentrations >10 mg.L^{-1} in the PURE study population

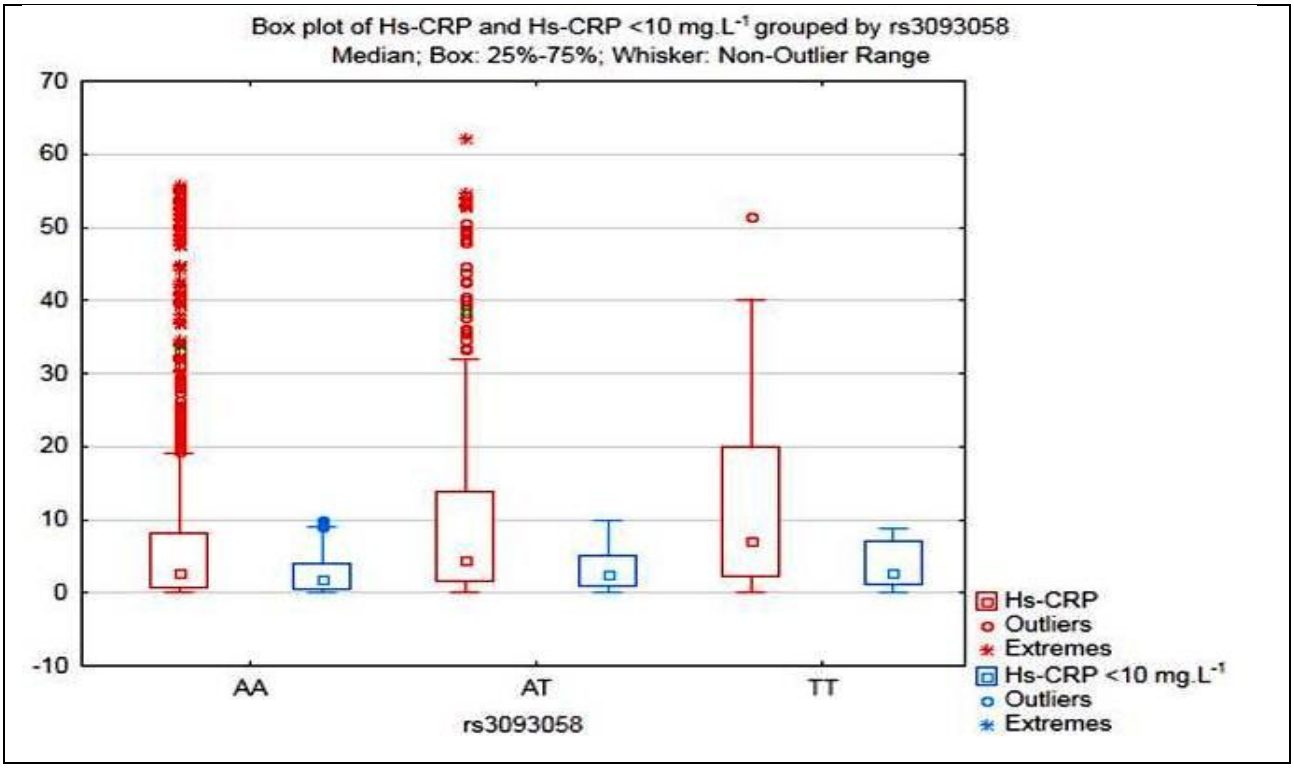
SNP designation		CRP concentrations (mg.L^{-1})						
		Homozygote (wildtype)	N	Heterozygote	N	Homozygote (mutant)	N	p-value
rs3093058:	Whole group	2.72	770	3.29	286	4.15	19	0.01
	Men	2.36	298	2.96	106	3.96	9	0.02
	Women	2.94	472	3.52	180	4.27	10	0.01
	Rural	2.77	404	3.29	159	4.26	12	0.02
	Urban	2.66	366	3.30	127	3.96	7	0.02
rs1205:	Whole group	3.06	608	2.76	407	2.15	61	0.01
	Men	2.85	223	2.29	173	1.26	17	0.01
	Women	3.21	385	3.06	234	2.56	44	0.24
	Rural	3.07	337	2.86	201	2.19	37	0.12
	Urban	3.05	271	2.67	206	2.03	24	0.05
rs7553007:	Whole group	3.10	589	2.75	416	2.15	72	0.01
	Men	2.85	222	2.29	170	1.42	21	0.01
	Women	3.25	367	3.04	246	2.53	51	0.12
	Rural	3.10	323	2.85	211	2.22	42	0.08
	Urban	3.07	266	2.66	205	2.04	30	0.03
rs2794520:	Whole group	3.07	593	2.75	409	2.17	69	0.01
	Men	2.85	220	2.27	170	1.47	20	0.01
	Women	3.22	373	3.04	239	2.53	49	0.17
	Rural	3.09	326	2.82	207	2.24	41	0.09
	Urban	3.04	267	2.68	202	2.04	28	0.05
rs3093068:	Whole group	2.51	432	3.11	517	3.40	126	0.01
	Men	2.13	156	2.68	208	3.32	49	0.01
	Women	2.75	276	3.37	309	3.47	77	0.01
	Rural	2.43	219	3.20	282	3.55	74	0.01
	Urban	2.60	213	2.99	235	3.21	51	0.12
rs3093062:	Whole group	2.72	770	3.29	286	4.15	19	0.01
	Men	2.36	298	2.96	106	3.97	9	0.02
	Women	2.94	472	3.52	180	4.27	10	0.01
	Rural	2.77	404	3.29	159	4.26	12	0.02
	Urban	2.66	366	3.30	127	3.96	7	0.02
rs2027471:	Whole group	3.08	596	2.76	412	2.17	69	0.01
	Men	2.85	222	2.28	171	1.47	20	0.01
	Women	3.23	374	3.07	241	2.54	49	0.17
	Rural	3.09	327	2.87	208	2.24	41	0.11
	Urban	3.06	269	2.67	204	2.05	28	0.04
rs1341665:	Whole group	3.08	596	2.76	412	2.17	69	0.01
	Men	2.85	222	2.28	171	1.47	20	0.01
	Women	3.23	374	3.07	241	2.54	49	0.17
	Rural	3.09	327	2.87	208	2.24	41	0.11
	Urban	3.06	269	2.66	204	2.04	28	0.04

CRP = C-reactive protein; N = number of individuals; rs = reference sequence;.

All the polymorphisms that presented with a significant association to CRP concentrations remained significant (indicated in red in Table 4.21) even after removal of the individuals with CRP concentrations of above 10 mg.L^{-1}, indicating that these associations are not

dependent on the effects of high-grade inflammation. After the exclusion of these subjects, the main observations determined in this dissertation remained unchanged, with the addition of three SNPs (rs3093058, rs3093068 and rs3093062), indicating a significant association with CRP concentrations, which previously did not present with significant differences in CRP concentrations between the rural and female subjects. A possible reason for this observation could be the difference in the distribution of CRP concentrations between the whole group and the group following exclusion of the individuals with CRP concentrations > 10 $mg.L^{-1}$. Figure 4.43 is an illustration of the difference in the distribution of CRP concentrations for the different genotypes at the rs3093058 locus when investigating the entire group compared to investigating only those individuals with CRP concentration < 10 $mg.L^{-1}$. The interaction effect of each SNP with gender and locality were also reanalysed, for these individuals (Table 4.22).

Figure 4.43 Distribution of CRP concentrations for whole group and individuals with CRP concentrations <10 $mg.L^{-1}$ in the PURE population



A = adenine; Hs-CRP = high sensitivity C-reactive protein; rs = reference sequence; T = thymine

As can be seen from Figure 4.43, the individuals with acute inflammation were randomly distributed across the different genotypes. With the full sample set, the range of the data is larger, therefore the probability of overlap in values between the groups is greater, thus possibly masking significant differences, but in the smaller sample set, the variable range becomes more defined between the groups and therefore significant differences are more likely to be observed.

Table 4.22 Interaction effects of genotype and gender and locality with CRP concentrations, excluding individuals with CRP concentrations > 10 mg.L^{-1} in the PURE study population

CRP concentrations <u>including</u> >10 mg.L^{-1}			CRP concentrations <u>excluding</u> < 10 mg.L^{-1}		
CRP SNP	**Gender interaction** (p-value)	**Location interaction** (p-value)	**CRP SNP**	**Gender interaction** (p-value)	**Location interaction** (p-value)
rs3093058	0.10	0.08	**rs3093058**	0.98	0.68
rs1205	0.02	0.46	**rs1205**	0.43	0.55
rs7553007	0.08	0.78	**rs7553007**	0.54	0.83
rs2794520	0.06	0.78	**rs2794520**	0.33	0.95
rs3093068	0.32	0.04	**rs3093068**	0.84	0.35
rs3093062	0.08	0.06	**rs3093062**	0.98	0.90
rs2027471	0.05	0.75	**rs2027471**	0.49	0.68
rs1341665	0.05	0.75	**rs1341665**	0.32	0.78

CRP = C-reactive protein; rs = reference sequence; SNP = single nucleotide polymorphism

The differences observed in the significance of the interaction effects indicated in Table 4.22 could possibly be ascribed to the fact there are fewer individuals in each genotype group because of the exclusion criteria thus masking the interaction effects seen before exclusion of those with high CRP. Those excluded were representative of all three genotypes of the SNPs that had significant interaction effects before their exclusion.

4.9 SUMMARY OF THE RESULTS

To summarise the results determined in this research project, the objectives (mentioned in Chapter 1) will be briefly discussed to evaluate the manner in which they have been achieved. With regard to the first objective (to establish the genotype distribution of specific polymorphisms in the *CRP* gene in a black South African population), of the 16 SNPs the researcher initially attempted to determine, four SNPs were unsuccessful (two were lost owing to issues with the design of the BeadXpress® assay and the other two SNPs were lost owing to failed genotyping). This is typical of most medium throughput genotyping experiments and even though there were fallouts, this research project is still the largest investigation of *CRP* polymorphisms and their association with CRP concentrations in a black South African population.

The distributions of the genotype classes of the 12 SNPs investigated were determined to be in agreement with those expected in a population adhering to the assumptions of HWE, indicating, therefore, that the differences in genotypic distribution in the population that was studied were not the result of inbreeding, migration, selective survival or small

population size (Hardy, 1908). Furthermore, there is considerable ethnic variation in the frequency of the SNPs reported in this dissertation compared to those in other ethnic groups, as indicated in Section 4.7. This gives an indication of the unique genetic profiles present in the black South African population, as well as a possible explanation for the increased CRP concentrations within this population. A summary of the results of the second objective, which was to determine the associations between the 12 SNPs and CRP concentrations in the black South African population, is indicated in Table 4.22.

Table 4.23 Summary of the MAF of the SNPs investigated in this study, as well as their association with CRP levels in the black South African population

CRP SNP	MAF %	ΔCRP	Significant	*CRP* SNP	MAF %	ΔCRP	Significant
rs3093058	T = 16.5%	Increase	Yes	rs2808630	C = 14%	Decrease	No
rs3093062	A = 16.5%	Increase	Yes	rs1341665	A = 23%	Decrease	Yes
rs1800947	C = 100%	Increase	No	rs3093068	G = 62%	Increase	Yes
rs1130864	T = 13.0%	Decrease	No	rs2794520	T = 23%	Decrease	Yes
rs1205	T = 23.0%	Decrease	Yes	rs7553007	A = 24%	Decrease	Yes
rs1417938	A = 97.5%	Decrease	No	rs2027471	A = 24%	Decrease	Yes

a = adenine; c = cytosine; CRP = c-reactive protein; g = guanine; MAF = minor allele frequency; rs = reference sequence; t = thymine

As can be seen in Table 4.22, three of the investigated SNPs (rs3093058, rs3093062 and rs3093068) were all associated with a significant increase in CRP concentrations. Five SNPs (rs1205, rs1341665, rs2794520, rs7553007 and rs2027471) were associated with a significant decrease in CRP concentrations. All the SNPs that presented with significant associations with CRP concentrations, except for rs3093062, are in agreement with that which was previously reported for other ethnic groups in the literature. The increase in CRP concentrations associated with rs3093062 is in contrast to the associations with CRP determined in an African American as well as in a European cohort. As previously mentioned, this observation could possibly be ascribed to the genetic distribution observed within this large sample size which could be a more robust distribution profile for the Tswana population.

It has been determined that individuals of African, Latin American or South Asian ancestry have higher CRP concentrations than individuals of European descent (Nazmi & Victora, 2007). Furthermore, CRP concentrations in black South African women have been determined to be much higher on average than those of caucasian South African women (Schutte *et al*., 2006). These findings are most likely due to population-specific genetic associations. The MAFs of the different SNPs are also summarised in Table 4.22. All the *CRP* SNPs which were significantly associated with a decrease in CRP

concentrations had a MAF of approximately 23%. This would then suggest that the prevalence of the major alleles at these loci in the black South African population that was investigated was approximately 67%. As the major alleles were associated with higher CRP concentrations, this could be a possible explanation for the increased CRP concentrations that have been determined in the black South African population. As mentioned, the majority of the investigated SNPs significantly decreased CRP concentrations in the study population, but considering the fact that the major allele frequencies of these SNPs and the outcomes thereof (higher CRP concentrations) were at higher levels than determined in the European population (as indicated in Table 4.23), this could partly explain the high CRP concentrations observed in this black South African population. Furthermore, the SNPs that were reported to increase CRP concentrations (rs3093058, rs3093062 and rs3093068) had an MAF of 16.5% for both rs3093058 and rs3093062 and 62% for rs3093068. This also contributes to the increased CRP concentrations observed in this population. Table 4.23 is a summary of the MAFs in the European population. With regard to the three SNPs previously discussed *i.e*. rs3093058, rs3093062 and rs3093068, which increased CRP concentrations in the study population, the MAFs of these SNPs were 8.4% for rs3093068 and 0% for rs3093058 and rs3093062. In the European population, rs3093062 was associated with a decrease in CRP concentrations, which is the opposite of what the researcher observed in the current study population. This would further explain the increased CRP concentrations seen in black population groups when compared to European populations. Furthermore, certain of the SNPs (rs2808630, rs2794520, rs7553007 and rs2027471) that were associated with lower CRP concentrations had a higher prevalence in the European population compared to the study population. This suggests that when considering the MAFs of the different SNPs in the different population groups, the black South African population is expected to have higher CRP concentrations.

Table 4.24 Summary of the MAF of the SNPs investigated in the European population

***CRP* SNP**	**MAF %**	**ΔCRP**	***CRP* SNP**	**MAF %**	**ΔCRP**
rs3093058	T = 0.0%	Increase	rs2808630	C = 30.0%	Decrease
rs3093062	A = 0.0%	Decrease	rs1341665	A = 27.0%	Decrease
rs1800947	C = 6.2%	Increase	rs3093068	G = 8.4%	Increase
rs1130864	T = 30.4%	Decrease	rs2794520	T = 34.0%	Decrease
rs1205	T = 25.0%	Decrease	rs7553007	A = 37.0%	Decrease
rs1417938	A = 33.0%	Decrease	rs2027471	A = 34.0%	Decrease

a = adenine; c = cytosine; g = guanine; MAF = minor allele frequency; rs = reference sequence; t = thymine

It is important to keep in mind that phenotypes, such as high CRP concentrations, are often affected by more than one gene (epistasis) and that phenotypic expression is not always the direct reflection of a single genotype. In addition, phenotypic expression can also be modified by factors other than genetic background, such as environmental factors, including dietary habits (Klug *et al*., 2010). Furthermore, it has been determined that alterations in the intron, exon 2, and the 3' UTR are in strong linkage disequilibrium with each other and with the proximal CRP promoter (Szalai *et al*., 2005). Therefore, it cannot be ruled out that the observed associations between the CRP concentrations and the SNPs investigated may be due to linkage disequilibrium between these SNPs and possible functional SNPs that are as yet unidentified.

The researcher observed several associations between certain SNPs and CRP concentrations that were dependent on the level of urbanisation of the participants (namely rs3093058, rs3093062, rs1205, rs1341665, rs3093068, rs2794520, rs7553007 and rs2027471), thus implying that when adopting the related lifestyle associated with living in an urban environment, the genotype harboured has an influence on CRP concentrations, but when residing in a rural community it makes no difference to the CRP phenotype. Three SNPs (rs3093058, rs3093062 and rs303068) were associated with an increase in CRP concentrations and five SNPs (rs1205, rs1341665, rs2794520, rs7553007 and rs2027471) were associated with a significant decrease in CRP concentrations when present in the urban group. Possible reasons for this observation, *i.e*. that urban individuals harboured SNPs that increased CRP concentrations as well as SNPs that decreased CRP concentrations, could be ascribed to the possible functional effect of the SNP dependent on the region in the *CRP* gene where the SNP was localised. Two of the SNPs (rs3093058 and rs3093062) associated with increased CRP concentrations are located in the promotor region of the *CRP* gene. The promotor region is the site where transcription is initiated and it harbours binding sites for specific transcription regulatory proteins, known as transcription factors (Klug *et al*., 2010). If a polymorphism is present in the promotor region, it may destroy or create one of these transcription factor binding sites, thus resulting in an altered pattern of CRP mRNA expression. Future functional analyses will have to be undertaken to determine if these SNPs do indeed alter the binding of certain transcription factors and in so doing up-regulate the expression of this gene.

Four of the SNPs that were determined to decrease CRP concentrations are located within the intronic region of the *CRP* gene. Introns are DNA sequences that are not represented in the final mature RNA product of a gene (Klug *et al*., 2010). During the transcription

process, the introns are removed by means of splicing in order to generate a mature mRNA molecule, which can be translated into a functional protein (Klug *et al*., 2010). Therefore, if a polymorphism is present in the intronic region of a gene, splicing may not be initiated properly, which could cause the mRNA molecule not to mature correctly, therefore preventing expression of the protein (in this case CRP), resulting in the decreasing effect that these SNPs have on CRP concentrations.

Other possible causes for the differences in the association patterns seen between *CRP* SNPs and CRP concentrations in the rural and urban group could also be attributed to environmental factors such as smoking. The CRP concentrations in the different tobacco use categories differed significantly for the rural *versus* urban groups ($p = 0.04$). Smoking has a well-reported effect on CRP concentrations (Danesh *et al*., 1999).

Furthermore, in investigating the possible causes of the differences in association patterns between the rural and the urban groups, the dietary data of the two groups was compared. All the dietary variables differed significantly ($p \leq 0.01$) between the rural and urban groups. The urban group had a more inflammatory diet (high SFA intake, protein intake and high refined carbohydrates, as discussed in Section 2.3.4) than the rural group. The more inflammatory diet followed by the urban group could possibly have played a role in the differences seen in CRP concentrations. Although the dietary variables were significantly different between the groups, none of these variables correlated very well with CRP concentrations, suggesting that the diet could only have had a small influence on the differences in the CRP concentrations seen between these two sub-groups. Age was also determined to differ significantly between rural and urban groups, but again this variable only had a weak correlation with CRP concentrations. Many environmental factors in the different levels of urbanisation could be at play and influence CRP concentration differently in individuals residing in a rural community compared to individuals in an urban community. These factors should be studied in more detail in future research to establish possible mechanisms by which area of residence influences CRP concentrations.

Associations between the SNPs at the rs3093058, rs3093062, rs1205, rs1341665, rs3093068, rs2794520, rs7553007 and rs2027471 and CRP concentrations revealed that males presented different association patterns across the different genotypes compared to the female group, which generally did not present with significant associations between *CRP* SNPs and CRP concentrations. Again, three SNPs (rs3093058, rs3093062 and rs303068) were associated with an increase in CRP concentrations and five SNPs

(rs1205, rs1341665, rs2794520, rs7553007 and rs2027471) were associated with a significant decrease in CRP concentrations. As discussed previously, the localisation of the SNPs within the *CRP* gene may have an influence on whether the expression of the protein (CRP) will be up-regulated or down-regulated. The differences identified between the CRP concentrations and the different genotype classes observed in the men were initially hypothesised to be due to the BMI differences between men and women. Women in this cohort were significantly more obese than the men, as indicated in Section 4.1. The high level of inflammation among the women (due to the high level of obesity) could have been a possible explanation as to why significant differences in CRP concentrations were not seen between the different genotype groups, as the increased inflammation may have masked the differences in CRP expression. However, this was unlikely, as BMI presented with a medium correlation with CRP concentrations and was adjusted for in all these analyses. Furthermore, it was determined that although the men had a higher kJ intake than the women, the women in the cohort had a more inflammatory diet (high SFA intake, protein intake and high refined carbohydrates) than the men. This could suggest that the CRP concentrations of the women would be higher than those of the men, but in Section 4.1 the researcher reported that the CRP concentrations of the genders did not differ. Therefore it is difficult to determine the possible reasons for the difference in association patterns determined between the two gender groups. Other factors that the researcher did not measure, such as hormonal factors, could affect CRP concentrations and should be investigated in future studies to determine the possible mechanisms of action at play. Limited data is available on the comparison of the distribution of CRP concentrations between the different genders and this should also be investigated in future studies.

Interaction effect analyses were undertaken to determine if any interaction effects were at play in CRP regulation, *i.e*. to establish whether level of urbanisation or gender influenced the associations between the different genotypes and CRP concentrations. Only one SNP (rs3093068) presented with an interaction effect with the level of urbanisation. This observed gene-environment interaction (Figure 4.34) could indicate the environmental modification of a phenotypic expression and could be due to underlying infection in the rural group causing increased levels of CRP in the individuals harbouring the wild-type homozygous mutant. However, it is more likely that this could have been a spurious association, since the p-value for this association was borderline significant, *i.e*. 0.046, and there was a great deal of overlap in the ranges of the CRP concentrations between the different genotype groups.

Furthermore, gene-gender interaction effects were observed within this population for three SNPs, namely rs1205, rs1341665 and rs2027471. Women harbouring the wild-type homozygous genotypes at the three SNPs had lower CRP concentrations than men harbouring the wild-type homozygous genotype; however, in respect of the heterozygote and the homozygote mutant genotypes the CRP concentrations did not differ between the genders. One could speculate that these interactions could be ascribed to other factors that also influence CRP concentrations that differ between the genders, such as hormones, alcohol intake or psychosocial stress or any other factor that differs between the genders that the researcher did not measure, but that the effect of this variable is lost when harbouring either the heterozygous or homozygous mutant genotype. The effect that alcohol could have is plausible, since it has been reported that the alcohol intake of men was higher than that of women in the PURE cohort (Nienaber-Rousseau *et al*., 2013). Differences in alcohol intake could influence the response of the homozygous wild-type genotype in men and thus increase their CRP concentrations. When excluding individuals with a CRP concentration that can be considered to be an indication of acute inflammation, CRP concentrations also differed significantly across the genotypes, similar to those determined for the whole group, indicating that increased inflammation due to infection is unlikely to be a factor in this process. Further functional analyses may allow researchers to determine the mechanistic nature of these differences.

After this research had been conducted, many questions on the role of the genetic determinants of CRP concentrations in health and disease remain unanswered and should be investigated. These questions include determining the manner in which environment and gender play a role, together with genetics, in the determination of CRP concentrations. CRP in itself is a complex molecule and there are many factors at play that influence its expression.

CHAPTER FIVE

Conclusion

It is well known that CVD has a complex multifactorial aetiology and neither genetic nor environmental agents acting independently cause CVD. Based on this, it is therefore important to realise that an individual's risk of developing CVD is dependent on the interaction between numerous genetic variants and on several environmental factors. From Chapter 2, it is evident that CRP concentrations are important biomarkers of the development of CVDs. CRP concentrations are one of those factors that are influenced by both genetic and environmental issues. For this reason the researcher decided to follow a multi-factorial approach in this project to understand the determinants of CRP concentrations in a black South African population better.

Several aspects of this research were novel in that it is the first study of this magnitude to investigate the genetic determinants of CRP concentrations in a black South African population (n = 1,588). The large sample size of this investigation allowed accurate estimation of the frequency of the genetic determinants in the *CRP* gene. Since the frequencies of the SNPs investigated differed from those of the non-African ethnic groups, previously reported in the literature, this study contributed to the evidence that approaching public health genetics in a population-specific manner is critical to developing effective screening strategies. It is important not to extrapolate knowledge of phenotypes associated with genetic variation derived from studies performed on non-African populations, to African populations. There is, therefore, a need for establishing population-specific risk factors for NCD aetiology. The value of studying diverse ethnic groups is not only in elucidating the diversity in the genetic profiles, but also evaluating the diversity in dietary habits and other lifestyle factors.

In this study CRP concentrations only differed significantly across the genotype classes of certain SNPs (rs3093059, rs3093062, rs1205, rs1341665, rs3093068, rs2794520, rs7553007 and rs2027471) when the participants adopted an urban lifestyle, therefore one can conclude that the level of urbanisation has a significant impact on the phenotypic expression of the genetic variation in a population. This highlights that genetics should no longer be looked at in isolation, but researchers should be aware of the environment in

which the study population is living and take into account various environmental factors that may play a role. Further research should aim to determine exactly which factors associated with the level of urbanisation affect the phenotypic expression of CRP. A possible factor to take into account might be increased psychosocial stress, which may be influenced by socio-economic status and level of education. Once these lifestyle factors are identified, awareness should be created and behavioural change campaigns initiated in populations with high frequencies of the genotypes of the SNPs identified to be influenced by these lifestyle factors.

A similar effect was determined for men, whereby the phenotypic expression of the genotypes, at the rs3093059, rs3093062, rs1205, rs1341665, rs3093068, rs2794520, rs7553007 and rs2027471 loci, differed significantly, but not for women in relation to CRP concentrations. Although the male gender is a demographic determinant that cannot be modified, awareness campaigns should focus on this gender group. It might be more important for men to follow a diet associated with a decrease in inflammatory markers (including restricted alcohol intake) and follow a lifestyle that does not increase CRP concentrations than for women, to prevent the effect of higher circulating CRP concentrations brought about by genetic variants.

One of the major findings of this research project is that for a complex phenotype, one SNP is unlikely to demonstrate a major contribution to a phenotype. As the researcher measured 12 SNPs within the *CRP* gene in this study, she could more accurately predict the CRP phenotype than by only investigating a single SNP. Three of the 12 SNPs (rs3093058, rs3093062 and rs303068) were associated with an increase in CRP concentrations and five SNPs (rs1205, rs1341665, rs2794520, rs7553007 and rs2027471) were associated with a significant decrease in CRP concentrations and were observed in the men as well as urban individuals. As discussed in Section 4.9, this could most probably be ascribed to changes in gene function brought about by the specific localisation of these SNPs in the gene itself. To determine if this is actually the mechanism of action, a systems biological approach is needed to conduct functional studies of these SNPs in order to identify possible mechanisms by which the level of urbanisation may influence CRP concentrations. Systems biology incorporates data from genomics, transcriptomics, proteomics and other areas of biology to get a broader overview of the effects of genetic variation in an individual. Other future studies should also include the use of GWAS, which is an approach that involves scanning for genetic markers across the whole genome of an individual or a group of people to detect genetic variations associated with a

particular disease. This information can be used to identify other genes involved in the regulation of CRP concentrations to develop better strategies for detection, treatment and prevention of high CRP concentrations. Next generation sequencing (to identify more allele variants present in the whole *CRP* gene), epigenomics (where one can investigate influences of environmental factors on the epigenome and in turn the effect on the expression of certain genes) and also knockout mice studies (mice that have had one or more genes experimentally removed from their genome, which can be used to determine what the effect will be of removing parts of, or the whole *CRP* gene on the expressed phenotype) should also be investigated in future and will make a huge contribution to the understanding of CRP and its determinants. This study presented significant insights into the genetic determinants of CRP concentrations and future studies, as mentioned above, would generate more information and insight into the determinants of CRP concentrations. Although determining the role of CRP in CVD risk was not one of the aims of this project, it is definitely an important avenue of investigation that stems from the project. Mendelian randomisation studies, as discussed in Section 2.4.2.2, should be undertaken with regard to CRP in this cohort to prove whether or not these CRP SNPs are causally associated with CVD. In order for Mendelian studies to be conducted, prospective data is needed for this study population regarding their CRP concentrations. Studies should also be undertaken to investigate whether these *CRP* SNPs and the differences in CRP concentrations that result from harbouring different combinations of these SNPs ultimately contribute to differences in CVD outcomes.

It is, however, important to realise that the science of nutrigenetics is still in its infancy and, therefore, making dietary recommendations based on an individual's genetic make-up should not be done until more research has been conducted. Numerous factors (both genetic and environmental) are involved and no single factor acting alone is likely to have enough of an influence to be used as a clinical diagnostic test of CRP concentrations. Therefore, the research conducted for this dissertation can only partly explain the increased CRP concentrations often reported in black South Africans, as there are other factors at play in CRP regulation, such as the determination that the CRP concentrations indicate a polygenic inheritance pattern, meaning that a number of genes are involved in the expression of altered CRP concentrations. Many more genetic loci remain to be discovered that contribute to the genetic basis of plasma CRP concentrations, but these results provide valuable information on the regulation of CRP in a black South African population as well as contribute to the literature regarding CRP concentrations on a global level.

REFERENCE LIST

Abernethy, T.J. & Avery, O.T. 1941. The occurrence during acute infections of a protein not normally present in the blood. *Journal of experimental medicine,* 73(2):173-182.

Albert, C.M., Ma, J., Rifai, N., Stampfer, M.J. & Ridker, P.M. 2002. Prospective study of C-reactive protein, homocysteine, and plasma lipid levels as predictors of sudden cardiac death. *Circulation,* 105(22):2595-2599.

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

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

Barker, D.J.P. 1996. Growth in utero and coronary heart disease. *Nutrition reviews,* 54(2):1-7.

Barker, D.J.P. 2001. Fetal and infant origins of adult disease. *Monatsschrift Kinderheilkunde,* 149(1):2-6.

Ben-Assayag, E., Shenhar-Tsarfaty, S., Bova, I., Berliner, S., Shopin, L., Peretz, H., Usher, S., Shapira, I. & Bornstein, N. 2007. Triggered C-reactive protein (CRP) concentrations and the CRP gene− 717A> G polymorphism in acute stroke or transient ischemic attack. *European journal of neurology,* 14(3):315-320.

Benjamin, E., Dupuis, J., Larson, M., Lunetta, K., Booth, S., Govindaraju, D., Kathiresan, S., Keaney, J., Keyes, M. & Lin, J.P. 2007. Genome-wide association with select biomarker traits in the Framingham heart study. *BMC medical genetics,* 8(1):11-23.

Bermudez, E.A., Rifai, N., Buring, J., Manson, J.A.E. & Ridker, P.M. 2002. Interrelationships among circulating interleukin-6, C-reactive protein, and traditional cardiovascular risk factors in women. *Arteriosclerosis, thrombosis, and vascular biology,* 22(10):1668-1673.

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

Blake, G.J., Rifai, N., Buring, J.E. & Ridker, P.M. 2003. Blood pressure, C-reactive protein, and risk of future cardiovascular events. *Circulation,* 108(24):2993-2999.

Blake, G. & Ridker, P. 2002. Inflammatory bio-markers and cardiovascular risk prediction. *Journal of internal medicine,* 252(4):283-294.

Bothig, S. 1989. Who MONICA project: objectives and design. *International journal of epidemiology*, 18(3):29–37.

Bourne, L., Langenhoven, M., Steyn, K., Jooste, P., Laubscher, J. & Van der Vyver, E. 1993. Nutrient intake in the urban African population of the Cape Peninsula, South Africa. The BRISK study. *Central African journal of medicine,* 39(12):238-247.

Boutayeb, A. & Boutayeb, S. 2005. The burden of non communicable diseases in developing countries. *International journal for equity in health,* 4(1):2-10.

Campbell, M.C. & Tishkoff, S.A. 2008. African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annual review of genomics and human genetics,* 9(1):403-433.

Carlson, C.S., Aldred, S.F., Lee, P.K., Tracy, R.P., Schwartz, S.M., Rieder, M., Liu, K., Williams, O.D., Iribarren, C. & Lewis, E.C. 2005. Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *American journal of human genetics,* 77(1):64-77.

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

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

Cleland, S.J., Sattar, N., Petrie, J.R., Forouhi, N.G., Elliott, H.L. & Connell, J.M.C. 2000. Endothelial dysfunction as a possible link between C-reactive protein levels and cardiovascular disease. *Clinical science,* 98(5):531-535.

Clifton, P.M. 2003. Diet and C-reactive protein. *Current atherosclerosis reports,* 5(6):431-436.

Clyne, B. & Olshaker, J.S. 1999. The C-reactive protein. *Journal of emergency medicine*, 17 (6):1019-1025

Connor, M., Rheeder, P., Bryer, A., Meredith, M., Dubb, A. & Fritz, V. 2008. The South African stroke risk in general practice study. *South African medical journal,* 95(5):334-340.

Crawford, D.C., Sanders, C.L., Qin, X., Smith, J.D., Shephard, C., Wong, M., Witrak, L., Rieder, M.J. & Nickerson, D.A. 2006. Genetic variation is associated with C-reactive protein levels in the third national health and nutrition examination survey. *Circulation,* 114(23): 2458-2465.

Crimmins, E.M., Saito, Y. & Ingegneri, D. 1989. Changes in life expectancy and disability-free life expectancy in the United States. *Population and development review,* 15(2):235-267.

D'Agostino Sr, R.B., Vasan, R.S., Pencina, M.J., Wolf, P.A., Cobain, M., Massaro, J.M. & Kannel, W.B. 2008. General cardiovascular risk profile for use in primary care. *Circulation,* 117(6):743-753.

Danenberg, H.D., Szalai, A.J., Swaminathan, R.V., Peng, L., Chen, Z., Seifert, P., Fay, W.P., Simon, D.I. & Edelman, E.R. 2003. Increased thrombosis after arterial injury in human C-reactive protein–transgenic mice. *Circulation,* 108(5):512-515.

Danesh, J., Muir, J., Wong, Y.K., Ward, M., Gallimore, J.R. & Pepys, M.B. 1999. Risk factors for coronary heart disease and acute-phase proteins. A population-based study. *European heart journal*, 20(13):954–959.

Danesh, J., Wheeler, J.G., Hirschfield, G.M., Eda, S., Eiriksdottir, G., Rumley, A., Lowe, G., Pepys, M.B. & Gudnason, V. 2004. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *New England journal of medicine,* 350(14):1387-1397.

Danesh, J., Hingorani, A. & Hallmans, G. 2008. Collaborative pooled analysis of data on C-reactive protein gene variants and coronary disease: Judging causality by Mendelian randomisation. *European journal of epidemiology,* 23(8):531-540.

Danner, M., Kasl, S.V., Abramson, J.L. & Vaccarino, V. 2003. Association between depression and elevated C-reactive protein. *Psychosomatric Medicine*, 65(3)**:**347-356.

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R.E., Rafnsson, L. P, & Starr, J. M. 2009. Age-associated cognitive decline. *British Medical Bulletin* 92: 135-152.

De Beer, F., Soutar, A.K., Baltz, M.L., Trayner, I., Feinstein, A. & Pepys, M. 1982. Low density lipoprotein and very low density lipoprotein are selectively bound by aggregated C-reactive protein. *The journal of experimental medicine,* 156(1):230-242.

De Caterina, R., Madonna, R. & Massaro, M. 2004. Effects of omega-3 fatty acids on cytokines and adhesion molecules. *Current atherosclerosis reports,* 6(6):485-491.

Devaraj, S., Xu, D.Y. & Jialal, I. 2003. C-reactive protein increases plasminogen activator inhibitor-1 expression and activity in human aortic endothelial cells: Implications for the metabolic syndrome and atherothrombosis. *Circulation,* 107(3):398-404.

Dolan, S.E., Hadigan, C., Killilea, K.M., Sullivan, M.P., Hemphill, L., Lees, R.S., Schoenfeld, D. & Grinspoon, S. 2005. Increased cardiovascular disease risk indices in HIV-infected women. *Journal of acquired immune deficiency syndromes*, 39 (1):44-54

Dolman, R. 2013. The role of diet in Cardiovascular disease in black South Africans: both sides of the story. North West University – Potchefstroom campus. (Thesis – Ph.D.)

Doumatey, A.P., Chen, G., Tekola Ayele, F., Zhou, J., Erdos, M., Shriner, D., Huang, H., Adeleye, J., Balogun, W. & Fasanmade, O. 2012. C-reactive protein (CRP) promoter polymorphisms influence circulating CRP levels in a genome-wide association study of African Americans. *Human molecular genetics,* 21(13):3063-3072.

Du Clos, T.W. 2000. Function of C-reactive protein. *Annals of medicine,* 32(4):274-278.

Eiriksdottir, G., Smith, A., Aspelund, T., Hafsteinsdottir, S., Olafsdottir, E., Launer, L., Harris, T. & Gudnason, V. 2009. The interaction of adiposity with the CRP gene affects CRP levels: Age, gene/environment susceptibility-Reykjavik study. *International journal of obesity,* 33(2):267-272.

Eklund, C. 2007. Gene polymorphisms affecting the production of C-reactive protein. University of Tampere. (Dissertation – M.Sc.).

Elliott, P., Chambers, J.C., Zhang, W., Clarke, R., Hopewell, J.C., Peden, J.F., Erdmann, J., Braund, P., Engert, J.C. & Bennett, D. 2009. Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *Journal of the American medical association,* 302(1):37-48.

The emerging risk factors collaboration. 2010. C-reactive protein concentration and risk of coronary heart disease, stroke and mortality: an individual participant meta-analysis. *The Lancet*, 375 (9709):132-140.

Esposito, K., Marfella, R., Ciotola, M., Di Palo, C., Giugliano, F., Giugliano, G., D'Armiento, M., D'Andrea, F. & Giugliano, D. 2004. Effect of a Mediterranean-style diet on endothelial dysfunction and markers of vascular inflammation in the metabolic syndrome. *Journal of the American medical association,* 292(12):1440-1446.

Estruch, R., Martínez-González, M.A., Corella, D., Salas-Salvadó, J., Ruiz-Gutiérrez, V., Covas, M.I., Fiol, M., Gómez-Gracia, E., López-Sabater, M.C. & Vinyoles, E. 2006. Effects of a Mediterranean-style diet on cardiovascular risk factors. *Annals of internal medicine,* 145(1):1-11.

FAO. 2010. Fats and fatty acids in human nutrition – report of an expert consultation. 10-14 Nov. 2009. Geneva. Food and agriculture organization of the United Nation (FOA) Food and nutrition Paper 91. Rome.

Faruqi, R.M. & DiCorleto, P.E. 1993. Mechanisms of monocyte recruitment and accumulation. *British heart journal,* 69(1):19-29.

Fichtlscherer, S., Rosenberger, G., Walter, D.H., Breuer, S., Dimmeler, S. & Zeiher, A.M. 2000. Elevated C-reactive protein levels and impaired endothelial vasoreactivity in patients with coronary artery disease. *Circulation,* 102(9):1000-1006.

Flores-Alfaro, E., Fernández-Tilapa, G., Salazar-Martínez, E., Cruz, M., Illades-Aguiar, B. & Parra-Rojas, I. 2012. Common variants in the CRP gene are associated with serum C-reactive protein levels and body mass index in healthy individuals in Mexico. *Genetics and molecular research,* 11(3):1658-2267.

Floyd-Smith, G., Whitehead, A.S., Colten, H.R. & Francke, U. 1986. The human C-reactive protein gene (CRP) and serum amyloid P component gene (APCS) are located on the proximal long arm of chromosome 1. *Immunogenetics,* 24(3):171-176.

Fourie, C.M., Van Rooyen, J.M., Kruger, A. & Schutte, A.E. 2010. Lipid abnormalities in a never-treated HIV-1 subtype C-infected African population. *Lipids,* 45(1):73-80.

Fu, T. & Borensztajn, J. 2002. Macrophage uptake of low-density lipoprotein bound to aggregated C-reactive protein: Possible mechanism of foam-cell formation in atherosclerotic lesions. *Biochemical journal,* 366(1):195-201.

Fung, T.T., Rimm, E.B., Spiegelman, D., Rifai, N., Tofler, G.H., Willett, W.C. & Hu, F.B. 2001. Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *American journal of clinical nutrition,* 73(1):61-67.

Gaziano, T.A. 2005. Cardiovascular disease in the developing world and its cost-effective management. *Circulation,* 112(23):3547-3553.

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

Gordon, T., Castelli, W. P., Marthana, C. H., William, B. K. & Thomas, R. D. 1977. High density lipoprotein as a protective factor against coronary heart disease: The Framingham study. *The American journal of medicine,* 65(5):707-714.

Gluckman, P.D., Hanson, M.A. and Pinal, C. 2005. The developmental origins of adult disease. *Maternal & Child Nutrition*, 1(3):130–141.

Greenland, P., Knoll, M.D., Stamler, J., Neaton, J.D., Dyer, A.R., Garside, D.B. & Wilson, P.W. 2003. Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. *Journal of the American medical association,* 290(7):891-897.

Grundy, S.M., Pasternak, R., Greenland, P., Smith Jr, S. & Fuster, V. 1999. Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: A statement for healthcare professionals from the American heart association and the American college of cardiology. *Circulation,* 100(13):1481-1492.

Hage, F.G. & Szalai, A.J. 2007. C-reactive protein gene polymorphisms, C-reactive protein blood levels, and cardiovascular disease risk. *Journal of the American college of cardiology,* 50(12):1115-1122.

Hage, F.G. & Szalai, A.J. 2009. The role of C-reactive protein polymorphisms in inflammation and cardiovascular risk. *Current atherosclerosis reports,* 11(2):124-130.

Hair, J.F. & Anderson, R.E. 2010. Multivariate Data Analysis. 7th ed. NJ: Prentice Hall.

Hales, C.N. & Barker, D.J.P. 1992. Type 2 (non-insulin-dependent) diabetes mellitus: The thrifty phenotype hypothesis. *Diabetologia,* 35(7):595-601.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. *Nucleotide Acids Symposium*, (41):95-98.

Hardy, G.H. 1908. Mendelian proportions in a mixed population. *Science*, 28(706), 49-50.

Hattori, Y., Matsumura, M. & Kasai, K. 2003. Vascular smooth muscle cell activation by C-reactive protein. *Cardiovascular research,* 58(1):186-195.

Hindorff, L.A., Rice, K.M., Lange, L.A., Diehr, P., Halder, I., Walston, J., Kwok, P., Ziv, E., Nievergelt, C. & Cummings, S.R. 2008. Common variants in the CRP gene in relation to longevity and cause-specific mortality in older adults: The cardiovascular health study. *Atherosclerosis,* 197(2):922-930.

Holmes, C., Cunningham C., Zotova, E., Woolford, J., Dean, C., Kerr, S., Culliford, D. & Perry, V. H. 2009. Systemic inflammation and disease progression in Alzheimer disease. *Neurology,* 73 (10):768-774.

Hsue, P.Y., Lo, J.C., Franklin, A., Bolger, A.F., Martin, J.N., Deeks, S.G. & Waters, D.D. 2004. Progression of atherosclerosis as assessed by carotid intima-media thickness in patients with HIV infection. *Circulation*, 109 (13):1603-1608.

Imhof, A., Froehlich, M., Brenner, H., Boeing, H., Pepys, M.B. & Koenig, W. 2001. Effect of alcohol consumption on systemic markers of inflammation. *The Lancet,* 357(9258):763-767.

Jenkins, D., Wolever, T., Collier, G.R., Ocana, A., Rao, A.V., Buckley, G., Lam, Y., Mayer, A. & Thompson, L.U. 1987. Metabolic effects of a low-glycemic-index diet. *The American journal of clinical nutrition,* 46(6):968-975.

Jialal, I., Devaraj, S. & Venugopal, S.K. 2004. C-reactive protein: Risk marker or mediator in atherothrombosis? *Hypertension,* 44(1):6-11.

Johnson, R., McNutt, P., MacMahon, S. & Robson, R. 1997. Use of the Friedewald formula to estimate LDL-cholesterol in patients with chronic renal failure on dialysis. *Clinical chemistry,* 43(11):2183-2184.

Kardys, I., de Maat, M.P., Klaver, C.C., Despriet, D.D., Uitterlinden, A.G., Hofman, A., de Jong, P.T. & Witteman, J.C. 2007. Usefulness of combining complement factor H and C-reactive protein genetic profiles for predicting myocardial infarction (from the Rotterdam Study). *American Journal of Cardiology*, 100(4):646-648.

Kathiresan, S., Larson, M.G., Vasan, R.S., Guo, C., Gona, P., Keaney Jr, J.F., Wilson, P.W., Newton-Cheh, C., Musone, S.L. & Camargo, A.L. 2006. Contribution of clinical correlates and 13 C-reactive protein gene polymorphisms to inter individual variability in serum C-reactive protein level. *Circulation,* 113(11):1415-1423.

Kelly-Hayes, M., Beiser, A., Kase, C.S., Scaramucci, A., D'Agostino, R.B. & Wolf, P.A. 2003. The influence of gender and age on disability following ischemic stroke: The Framingham study. *Journal of stroke and cerebrovascular diseases,* 12(3):119-126.

Khera, A., McGuire, D.K., Murphy, S.A., Stanek, H.G., Das, S.R., Vongpatanasin, W., Wians, F.H., Grundy, S.M. & de Lemos, J.A. 2005. Race and gender differences in C-reactive protein levels. *Journal of the American college of cardiology,* 46(3):464-469.

Kilpatrick, J.M. & Volanakis, J.E. 1991. Molecular genetics, structure, and function of C-reactive protein. *Immunologic research,* 10(1):43-53.

King, D.E., Egan, B.M. & Geesey, M.E. 2003. Relation of dietary fat and fiber to elevation of C-reactive protein. *American journal of cardiology,* 92(11):1335-1339.

Kivimäki, M., Lawlor, D.A., Smith, G.D., Eklund, C., Hurme, M., Lehtimäki, T., Viikari, J.S.A. & Raitakari, O.T. 2007. Variants in the CRP gene as a measure of lifelong differences in average C-reactive protein levels the cardiovascular risk in young Finns study, 1980–2001. *American journal of epidemiology,* 166(7):760-764.

Klug, W.S., Cummings, M.R., Sencer, C.A. & Palladino, M.A. 2009. Concepts of genetics. 9th ed. San Francisco CA: Pearson, Benjamin Cummings.

Kolz, M., Koenig, W., Müller, M., Andreani, M., Greven, S., Illig, T., Khuseyinova, N., Panagiotakos, D., Pershagen, G. & Salomaa, V. 2008. DNA variants, plasma levels and variability of C-reactive protein in myocardial infarction survivors: Results from the AIRGENE study. *European heart journal,* 29(10):1250-1258.

Kruth, H.S. 2001. Macrophage foam cells and atherosclerosis. *Front bioscience,* 6(3):429-455.

Kushner, I., Rakita, L. & Kaplan, M.H. 1963. Studies of acute-phase protein. II. localization of Cx-reactive protein in heart in induced myocardial infarction in rabbits. *Journal of clinical investigation,* 42(2):286-292.

Kushner, I., Rzewnicki, D. & Samols, D. 2006. What does minor elevation of C-reactive protein signify? *American journal of medicine*, 199(166):e17-e28.

Laaksonen, D., Niskanen, L., Nyyssönen, K., Punnonen, K., Tuomainen, T.P., Valkonen, V.P., Salonen, R. & Salonen, J. 2004. C-reactive protein and the development of the metabolic syndrome and diabetes in middle-aged men. *Diabetologia,* 47(8):1403-1410.

Lakoski, S.G., Cushman, M., Palmas, W., Blumenthal, R., D'Agostino Jr, R.B. & Herrington, D.M. 2005. The relationship between blood pressure and C-reactive protein in the multi-ethnic study of atherosclerosis (MESA). *Journal of the American college of cardiology,* 46(10):1869-1874.

Lange, L.A., Carlson, C.S., Hindorff, L.A., Lange, E.M., Walston, J., Durda, J.P., Cushman, M., Bis, J.C., Zeng, D., Lin, D., Kuller, L.H., Nickerson, D.A., Psaty, B.M., Tracy, R.P. & Reiner, A.P. 2006. Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. J*ournal of the American medical association,* 296(22):2703-2711.

Lawlor, D.A., Harbord, R.M., Timpson, N.J., Lowe, G.D.O., Rumley, A., Gaunt, T.R., Baker, I., Yarnell, J.W.G., Kivimäki, M. & Kumari, M. 2008(a). The association of C-reactive protein and CRP genotype with coronary heart disease: Findings from five studies with 4,610 cases amongst 18,637 participants. *PLoS one,* 3(8):e3011-e3025.

Lawlor, D.A., Harbord, R.M., Sterne, J.A., Timpson, N. & Davey Smith, G. 2008(b). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in medicine,* 27(8):1133-1163.

Lee, C.C., You, N.Y., Song, Y., Hsu, Y.H., Manson, J.A., Nathan, L., Tinker, L. & Liu, S. 2009. Relation of genetic variation in the gene coding for C-reactive protein with its plasma protein concentrations: Findings from the Women's health initiative observational cohort. *Clinical chemistry,* 55(2):351-360.

Ledue T.B. & Rifai, N. Preanalytic. 2003. Analytic Sources of Variations in C-reactive Protein Measurement: Implications for cardiovascular disease risk assessment. *Clinical Chemistry,* 49(8):1258–1271.

Lei, K.J., Liu, T., Zon, G., Soravia, E., Liu, T.Y. & Goldman, N. 1985. Genomic DNA sequence for human C-reactive protein. *Journal of biological chemistry,* 260(24):13377-13383.

Lenfant, C. 2001. Can we prevent cardiovascular diseases in low-and middle-income countries? *Bulletin of the World Health Organization,* 79(10):980-982.

Li, Z., Hong, K., Saltsman, P., DeShields, S., Bellman, M., Thames, G., Liu, Y., Wang, H., Elashoff, R. & Heber, D. 2004. Long-term efficacy of soy-based meal replacements *vs*. an individualised diet plan in obese type II DM patients: Relative effects on weight loss, metabolic parameters, and C-reactive protein. *European journal of clinical nutrition,* 59(3):411-418.

Libby, P., Ridker, P.M. & Maseri, A. 2002. Inflammation and atherosclerosis. *Circulation,* 105(9):1135-1143.

Libby, P. 2006. Inflammation and cardiovascular disease mechanisms. *The American journal of clinical nutrition,* 83(2):456-460.

Liu, S., Manson, J.A.E., Buring, J.E., Stampfer, M.J., Willett, W.C. & Ridker, P.M. 2002. Relation between a diet with a high glycemic load and plasma concentrations of high-sensitivity C-reactive protein in middle-aged women. *American journal of clinical nutrition,* 75(3):492-498.

Lloyd-Jones, D., Adams, R.J., Brown, T.M., Carnethon, M., Dai, S., De Simone, G., Ferguson, T.B., Ford, E., Furie, K. & Gillespie, C. 2010. Heart disease and stroke statistics—2010 update. *Circulation,* 121(7):948-954.

Lopez-Garcia, E., Schulze, M.B., Fung, T.T., Meigs, J.B., Rifai, N., Manson, J.A.E. & Hu, F.B. 2004. Major dietary patterns are related to plasma concentrations of markers of inflammation and endothelial dysfunction. *American journal of clinical nutrition,* 80(4):1029-1035.

MacIntyre U.E., Venter C.S., Vorster H.H., Steyn H.S. 2001(a). A combination of statistical methods for the analysis of the relative validation data of the quantitative food frequency questionnaire used in the THUSA study. Transition, health and urbanisation in South Africa. *Public health nutrition*, 4(1):45-51.

MacIntyre U.E., Venter C.S., Vorster H.H. 2001(b). A culture-sensitive quantitative food frequency questionnaire used in an African population. Relative validation by 7-day weighted records and biomarkers. *Public health nutrition*, 4(1):63-71.

Mackie, I.J., Kitchen, S., Machin S.J. &. Lowe G.D.O. 2003. Guidelines on fibrinogen assays. *British journal of haematology*, 121(3):396–404.

Mackay, J., Mensah, G.A. & Greenlund, K. 2004. The atlas of heart disease and stroke. World Health Organization.

Manton, K.G. & Soldo, B.J. 1985. Dynamics of health changes in the oldest old: New perspectives and evidence. *The Milbank memorial fund quarterly health and society,* 63(2):206-285.

Mathers, J.C. & McKay, J.A. 2009. Epigenetics–potential contribution to fetal programming. *Advances in experimental medicine and biology, 646:*119-123.

Melzer, D., Perry, J.R.B., Hernandez, D., Corsi, A.M., Stevens, K., Rafferty, I., Lauretani, F., Murray, A., Gibbs, J.R., Paolisso, G., Rafiq, S., Simon-Sanchez, J.. Lango, H., Scholz, S., Weedon, M.N., Arepalli, S., Rice, N., Washecka, N., Hurst, A., Britton, A., Henley, W., van de Leemput, J., Li, R., Newman, A.B., Tranah, G., Harris, T., Panicker, V., Dayan, C., Bennett, A., McCarthy, M.I., Ruokonen, A., Jarvelin, M.R., Guralnik, J., Bandinelli, S., Frayling, T.M., Singleton, A. & Ferrucci, L. 2008. A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS genetics,* 4(5):e1000072-e1000082.

Miller, J.C. 1994. Importance of glycemic index in diabetes. *American journal of clinical nutrition,* 59(3):747-752.

Mohanty, P., Ghanim, H., Hamouda, W., Aljada, A., Garg, R. & Dandona, P. 2002. Both lipid and protein intakes stimulate increased generation of reactive oxygen species by polymorphonuclear leukocytes and mononuclear cells. *American journal of clinical nutrition,* 75(4):767-772.

Murray, C.J.L. & Lopez, A.D. 1997. Alternative projections of mortality and disability by cause 1990-2020: Global burden of disease study. *The Lancet,* 349(9064):1498-1504.

Murray, C.J.L., Lopez, A.D. & Mundial, B. 1996. The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Cambridge, MA: Harvard University Press.

Nazmi, A. & Victora, C.G. 2007. Socioeconomic and racial/ethnic differentials of C-reactive protein levels: A systematic review of population-based studies. *BMC public health,* 7(1):212-214.

Nickenig, G. & Harrison, D.G. 2002. The AT1-type angiotensin receptor in oxidative stress and atherogenesis. *Circulation,* 105(4):530-536.

Nienaber-Rousseau, C., Venter, C.S., Ellis, S.M., Kruger, A., Hoekstra, T., Moss, S.J., Melse-Boonstra, A. & Towers, G.W. 2013. Gene-gene interactions of specific MTHFR, MTR and CBS gene variants in relation to homocysteine in a black South African population. *Journal of nutrigenetics and nutrigenomics*, 6(2):61-72.

Noursadeghi, M. & Miller, R.F. 2005. Clinical value of C-reactive protein measurements in HIV-positive patients. *International journal of STD & AIDS,* 16(6):438-441.

Ordovas, J.M. 2006. Nutrigenetics, plasma lipids, and cardiovascular risk. *Journal of the American dietetic association,* 106(7):1074-1081.

Ordovas, J.M., Kaput, J. & Corella, D. 2007. Nutrition in the genomics era: Cardiovascular disease risk and the Mediterranean diet. *Molecular nutrition & food research,* 51(10):1293-1299.

Pai, J.K., Mukamal, K.J., Rexrode, K.M. & Rimm, E.B. 2008. C-reactive protein (CRP) gene polymorphisms, CRP levels, and risk of incident coronary heart disease in two nested case-control studies. *PLoS one,* 3(1):e1395.

Pankow, J.S., Folsom, A.R., Cushman, M., Borecki, I.B., Hopkins, P.N., Eckfeldt, J.H. & Tracy, R.P. 2001. Familial and genetic determinants of systemic markers of inflammation: The NHLBI family heart study. *Atherosclerosis,* 154(3):681-689.

Papanicolaou, D.A., Wilder, R.L., Manolagas, S.C. & Chrousos, G.P. 1998. The pathophysiologic roles of interleukin-6 in human disease. *Annals of internal medicine,* 128(2):127-137.

Paul, A., Ko, K.W.S., Li, L., Yechoor, V., McCrory, M.A., Szalai, A.J. & Chan, L. 2004. C-reactive protein accelerates the progression of atherosclerosis in apolipoprotein E–deficient mice. *Circulation,* 109(5):647-655.

Pearson, T.A., Mensah, G., Alexander, W.R., Anderson, J.L.; Cannon, R.O., Criqui, M., Fadl, Y.Y., Fortmann, S.P., Hong, Y., Myers, G.L., Rifai, N., Smith, S.C. Jr, Taubert, K., Tracy, R.P. & Vinicor, F. 2003. Markers of inflammation and cardiovascular disease: application to clinical and public health practice. A statement for healthcare professionals from the Centers for

Disease Control and Prevention and the American Heart Association. *Circulation*, 107(3):499–511.

Pepys, M., Rowe, I. & Baltz, M.L. 1985. C-reactive protein: Binding to lipids and lipoproteins. *International review of experimental pathology,* 27:83-111.

Popkin, B.M. 1993. Nutritional patterns and transitions. *Population and development review,* 19(1):138-157.

Popkin, B.M. 2003. The nutrition transition in the developing world. *Development policy review,* 21(5-6):581-597.

Popkin, B.M. 2006. Global nutrition dynamics: The world is shifting rapidly toward a diet linked with non-communicable diseases. *American journal of clinical nutrition,* 84(2):289-298.

Porter, R.S and Kaplan, J.L. 2011. *The Merck Manual of Diagnosis and Therapy*. 19th ed. Whitehouse Station, NJ. Merck Sharp & Dohme Crop.

Prentice, R.L., Caan, B., Chlebowski, R.T., Patterson, R., Kuller, L.H., Ockene, J.K., Margolis, K.L., Limacher, M.C., Manson, J.A.E., Parker, L.M., Paskett, E., Phillips, L. Robbins, J., Rossouw, J.E., Sarto, G.E., Shikany, J.M., Stefanick, M.L., Thomson, C.A., Van Horn, L., Vitolins, M.Z., Wactawski-Wende, J., Wallace, R.B., Wassertheil-Smoller, S., Whitlock, E., Yano, K., Adams-Campbell, L., Anderson, G.L., Assaf, A.R., Beresford, S.A., Black, H.R., Brunner, R.L., Brzyski, R.G., Ford, L., Gass, M., Hays, J., Heber, D., Heiss, G., Hendrix, S.L., Hsia, J., Hubbell, F.A., Jackson, R.D., Johnson, K.C., Kotchen, J.M., LaCroix, A.Z., Lane, D.S., Langer, R.D., Lasser, N.L. & Henderson, M.M. 2006. Low-fat dietary pattern and risk of invasive breast cancer: The Women's Health Initiative Randomized Controlled Dietary Modification Trial. *Journal of the American medical association,* 295(6):629-642.

Price, C. P., Trull, A. K., Berry, D. &. Gorman, E. G. 1987. Development and validation of a particle-enhanced turbidimetric immunoassay for C-reactive protein. *Journal of immunological methods,* 99(2):205-211.

Prokunina, L. &. Alarcón-Riquelme, M.E. 2004. Regulatory SNPs in complex diseases: their identification and functional validation. *Expert Reviews in Molecular Medicine*, 6(10):1-15.

Ramji, D.P. & Foka, P. 2002. CCAAT/enhancer-binding proteins: Structure, function and regulation. *Biochemical journal,* 365(Pt 3):561-575.

Ranjit, N., Diez-Roux, A.V., Shea, S., Cushman, M., Ni, H. & Seeman, T. 2007. Socioeconomic position, race/ethnicity, and inflammation in the multi-ethnic study of atherosclerosis. *Circulation,* 116(21):2383-2390.

Reddy, K.S. & Yusuf, S. 1998. Emerging epidemic of cardiovascular disease in developing countries. *Circulation,* 97(6):596-601.

Reed, F.A. & Tishkoff, S.A. 2006. African human diversity, origins and migrations. *Current opinion in genetics & development,* 16(6):597-605.

Reiner, A.P., Beleza, S., Franceschini, N., Auer, P.L., Robinson, J.G., Kooperberg, C., Peters, U. & Tang, H. 2012. Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. *American journal of human genetics,* 19(3):502-512.

Rhodes, B., Wong, A., Navarra, S.V., Villamin, C. & Vyse, T.J. 2008. Genetic determinants of basal C-reactive protein expression in Filipino systemic lupus erythematosus families. *Genes and immunity,* 9(2):153-160.

Rice, J.A. 2007. Mathematical statistics and data analysis. 3rd ed. Belmont,CA: Thomson, Duxbury press.

Ridker, P.M., Cushman, M., Stampfer, M.J., Tracy, R.P. & Hennekens, C.H. 1997. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *New England journal of medicine,* 336(14):973-979.

Ridker, P.M., Glynn, R.J. & Hennekens, C.H. 1998. C-reactive protein adds to the predictive value of total and HDL cholesterol in determining risk of first myocardial infarction. *Circulation,* 97(20):2007-2011.

Ridker, P.M., Rifai, N., Rose, L., Buring, J.E. & Cook, N.R. 2002. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. *New England journal of medicine,* 347(20):1557-1565.

Ridker, P.M., Buring, J.E., Cook, N.R. & Rifai, N. 2003. C-reactive protein, the metabolic syndrome, and risk of incident cardiovascular events: An 8-year follow-up of 14 719 initially healthy American women. *Circulation,* 107(3):391-397.

Ridker, P.M., Brown, N.J., Vaughan, D.E., Harrison, D.G. & Mehta, J.L. 2004. Established and emerging plasma biomarkers in the prediction of first atherothrombotic events. *Circulation,* 109(25 suppl. 1):6-19.

Ridker, P.M. & Cook, N. 2004. Clinical usefulness of very high and very low levels of C-reactive protein across the full range of Framingham risk scores. *Circulation,* 109(16):1955-1959.

Ridker, P.M., Paynter, N.P., Rifai, N., Gaziano, J.M. & Cook, N.R. 2008(a). C-reactive protein and parental history improve global cardiovascular risk prediction the Reynolds risk score for men. *Circulation,* 118(22):2243-2251.

Ridker, P.M., Pare, G., Parker, A., Zee, R.Y.L., Danik, J.S., Buring, J.E., Kwiatkowski, D., Cook, N.R., Miletich, J.P. & Chasman, D.I. 2008(b). Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: The women's genome health study. *American journal of human genetics,* 82(5):1185-1192.

Rifai, N. & Ridker, P.M. 2003. Population distributions of C-reactive protein in apparently healthy men and women in the United States: implication for clinical interpretation. *Clinical chemistry*, 49 (4), 666-669.

Robey, F.A., Jones, K.D., Tanaka, T. & Liu, T.Y. 1984. Binding of C-reactive protein to chromatin and nucleosome core particles. A possible physiological role of C-reactive protein. *Journal of biological chemistry,* 259(11):7311-7316.

Rost, N.S., Wolf, P.A., Kase, C.S., Kelly-Hayes, M., Silbershatz, H., Massaro, J.M., D'Agostino, R.B., Franzblau, C. & Wilson, P.W. 2001. Plasma concentration of C-reactive protein and risk of ischemic stroke and transient ischemic attack: the Framingham study. *Journal of cerebral circulation*, 32(11):2575-2579.

Sanger, F., Nicklen, S. & Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National academy of sciences,* 74(12):5463-5467.

Schillaci, G., Pirro, M., Gemelli, F., Pasqualini, L., Vaudo, G., Marchesi, S., Siepi, D., Bagaglia, F. & Mannarino, E. 2003. Increased C-reactive protein concentrations in never-treated hypertension: The role of systolic and pulse pressures. *Journal of hypertension,* 21(10):1841-1846.

Schumacher, W., Cockcroft, J., Timpson, N.J., McEniery, C.M., Gallacher, J., Rumley, A., Lowe, G., Smith, G.D., Wilkinson, I.B. & Ben-Shlomo, Y. 2009. Association between C-reactive protein genotype, circulating levels, and aortic pulse wave velocity. *Hypertension,* 53(2):150-157.

Schuster, S.C., Miller, W., Ratan, A., Tomsho, L.P., Giardine, B., Kasson, L.R., Harris, R.S., Petersen, D.C., Zhao, F., QI, J., Alkan, C., Kidd, J.M., Sun, Y., Drautz, D.I., Bouffard, P., Muzny, D.M., Reid, J.G., Nazareth, L.V., Wang, Q., Burhans, R., Riemer, C., Wittekindt, N.E., Moorjani, P., Tindall, E.A., Danko, C.G., Teo, W.S., Buboltz, A.M., Zhang, Z., MA, Q., Oosthuysen, A., Steenkamp, A.W., Oostuisen, H., Venter, P., Gajewski, J., Zhang, Y., Pugh, B.F., Makova, K.D., Nekrutenko, A., Mardis, E.R., Patterson, N., Pringle, T.H., Chiaromonte, F., Mullikin, J.C., Eichler, E.E., Hardison, R.C., Gibbs, R.A., Harkins, T.T. & Hayes, V.M. 2010. Complete Khoisan and Bantu genomes from Southern Africa. *Nature*, 463(7283):943-947.

Schutte, A., Van Vuuren, D., Van Rooyen, J., Huisman, H., Schutte, R., Malan, L. & Malan, N. 2006. Inflammation, obesity and cardiovascular function in african and caucasian women from South Africa: The POWIRS study. *Journal of human hypertension,* 20(11):850-859.

Shen, J., Arnett, D.K., Parnell, L.D., Peacock, J.M., Lai, C.Q., Hixson, J.E., Tsai, M.Y., Province, M.A., Straka, R.J. & Ordovas, J.M. 2008. Association of common C-reactive protein (CRP) gene polymorphisms with baseline plasma CRP levels and fenofibrate response. *Diabetes care,* 31(5):910-915.

Sing, C.F., Stengard, J.H. & Kardia, S.L.R. 2003. Genes, environment, and cardiovascular disease. *Arteriosclerosis, thrombosis, and vascular biology,* 23(7):1190-1196.

Singhal, A. & Lucas, A. 2004. Early origins of cardiovascular disease: Is there a unifying hypothesis? *The Lancet,* 363(9421):1642-1645.

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, G.D., Lawlor, D.A., Harbord, R., Timpson, N., Rumley, A., Lowe, G.D.O., Day, I.N.M. & Ebrahim, S. 2005. Association of C-reactive protein with blood pressure and hypertension life

course confounding and Mendelian randomization tests of causality. *Arteriosclerosis, thrombosis, and vascular biology,* 25(5):1051-1056.

Stępień, E., Miszalski-Jamka, T., Kapusta, P., Tylko, G. & Pasowicz, M. 2011. Beneficial effect of cigarette smoking cessation on fibrin clot properties. *Journal of thrombosis and thrombolysis*, 32 (2), 177-82.

Stewart, S., Wilkinson, D., Becker, A., Askew, D., Ntyintyane, L., McMurray, J.J.V. & Sliwa, K. 2006. Mapping the emergence of heart disease in a black, urban population in Africa: The heart of Soweto study. *International journal of cardiology,* 108(1):101-108.

Steyn, H.S. (jr.). 2002. Practically significant relationships between two variables, *South African Journal of Industrial Psychology*, 28(3):10-15.

Steyn, K. 2007. Heart disease in South Africa. Editor, Heart and Stroke Foundation South Africa.

Steyn, K., Fourie, J. & Temple, N. 2006(a). Chronic diseases of lifestyle in South Africa: 1995-2005. *Cape Town: South African medical research council.*

Steyn, N., Bradshaw, D., Norman, R., Joubert, J., Schneider, M. & Steyn, K. 2006(b). Dietary changes and the health transition in South Africa: Implications for health policy. *The double burden of malnutrition. Case studies from six developing countries. FAO - food and nutrition paper,* 259-299.

Steyn, N., Burger, S., Monyeki, K., Alberts, M. & Nthangeni, G. 2001. Seasonal variation in dietary intake of the adult population of Dikgale. *South African journal of clinical nutrition,* 14(4):140-145.

Suk, H.J., Ridker, P.M., Cook, N.R. & Zee, R.Y.L. 2005. Relation of polymorphism within the C-reactive protein gene and plasma CRP levels. *Atherosclerosis,* 178(1):139-145.

Szalai, A.J., Briles, D.E. & Volanakis, J.E. 1995. Human C-reactive protein is protective against fatal streptococcus pneumoniae infection in transgenic mice. *Journal of immunology,* 155(5):2557-2563.

Szalai, A.J. & McCrory, M.A. 2002. Varied biologic functions of c-reactive protein. *Immunologic research,* 26(1):279-287.

Szalai, A., Wu, J., Lange, E., McCrory, M., Langefeld, C., Williams, A., Zakharkin, S., George, V., Allison, D. & Cooper, G. 2005. Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. *Journal of molecular medicine,* 83(6):440-447.

Teo, K., Chow, C.K., Vaz, M., Rangarajan, S. & Yusuf, S. 2009. The prospective urban rural epidemiology (PURE) study: Examining the impact of societal influences on chronic non-communicable diseases in low-, middle-, and high-income countries. *American heart journal,* 158(1):1-7.

Teoh, H., Quan, A., Lovren, F., Wang, G., Tirgari, S., Szmitko, P.E., Szalai, A.J., Ward, M.E. & Verma, S. 2008. Impaired endothelial function in C-reactive protein overexpressing mice. *Atherosclerosis,* 201(2):318-325.

Thompson, S.G., Kienast, J., Pyke, S.D.M. & Haverkate, F. 1995. Hemostatic factors and the risk of myocardial infarction or sudden death in patients with angina pectoris European concerted action on thrombosis and disabilities angina pectoris study group. *New England journal of medicine,* 332(10):635-641.

Thompson, D., Pepys, M.B. & Wood, S.P. 1999. The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure,* 7(2):169-177.

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.

Tillett, W.S. & Francis, T. 1930. Serological reactions in pneumonia with a non-protein somatic fraction of pneumococcus. J*ournal of experimental medicine,* 52(4):561-571.

Timpson, N.J., Lawlor, D.A., Harbord, R.M., Gaunt, T.R., Day, I.N.M., Palmer, L.J., Hattersley, A.T., Ebrahim, S., Lowe, G. & Rumley, A. 2005. C-reactive protein and its role in metabolic syndrome: Mendelian randomisation study. *The Lancet,* 366(9501):1954-1959.

Toniatti, C., Demartis, A., Monaci, P., Nicosia, A. & Ciliberto, G. 1990. Synergistic trans-activation of the human C-reactive protein promoter by transcription factor HNF-1 binding at two distinct sites. *EMBO journal,* 9(13):4467-4475.

Torzewski, J., Torzewski, M., Bowyer, D.E., Fröhlich, M., Koenig, W., Waltenberger, J., Fitzsimmons, C. & Hombach, V. 1998. C-reactive protein frequently colocalizes with the

terminal complement complex in the intima of early atherosclerotic lesions of human coronary arteries. *Arteriosclerosis, thrombosis, and vascular biology,* 18(9):1386-1392.

Triant, V.A., Lee, H., Hadigan, C. & Grinspoon, S.K. 2007. Increased acute myocardial infarction rates and cardiovascular risk factors among patients with human immunodeficiency virus disease. *Journal of clinical endocrinology & metabolism,* 92(7):2506-2512.

Triant, V.A., Meigs, J.B. & Grinspoon, S.K. 2009. Association of C-reactive protein and HIV infection with acute myocardial infarction. *Journal of acquired immune deficiency syndromes,* 51(3):268-273.

UNAIDS, W.H.O.U. *2004. WHO policy statement on HIV testing*.

Van Dam, R.M., Visscher, A.W., Feskens, E.J., Verhoef, P. & Kromhout, D. 2000. Dietary glycemic index in relation to metabolic risk factors and incidence of coronary heart disease: the Zutphen Elderly Study. *European journal clinical nutrition*, 54(9):726 –31.

Venugopal, S.K., Devaraj, S., Yuhanna, I., Shaul, P. & Jialal, I. 2002. Demonstration that C-reactive protein decreases eNOS expression and bioactivity in human aortic endothelial cells. *Circulation,* 106(12):1439-1441.

Venugopal, S.K., Devaraj, S. & Jialal, I. 2003. C-reactive protein decreases prostacyclin release from human aortic endothelial cells. *Circulation,* 108(14):1676-1678.

Verma, S., Li, S.H., Badiwala, M.V., Weisel, R.D., Fedak, P.W.M., Li, R.K., Dhillon, B. & Mickle, D.A.G. 2002. Endothelin antagonism and interleukin-6 inhibition attenuate the proatherogenic effects of C-reactive protein. *Circulation,* 105(16):1890-1896.

Volanakis, J.E. 2001. Human C-reactive protein: Expression, structure, and function. *Molecular immunology,* 38(2-3):189-197.

Voleti, B. & Agrawal, A. 2005. Regulation of basal and induced expression of C-reactive protein through an overlapping element for OCT-1 and NF-κB on the proximal promoter. *Journal of immunology,* 175(5):3386-3390.

Vorster, H.H. & Nell, T.A. 2001. Make starchy foods the basis of most meals. *South African journal of clinical nutrition*, 14(3):17-23.

Vorster, H.H., Bourne, L.T., Venter, C.S. & Oosthuizen, W. 1999. Contribution of nutrition to the health transition in developing countries: A framework for research and intervention. *Nutrition reviews,* 57(11):341-349.

Wang, C.H., Li, S.H., Weisel, R.D., Fedak, P.W.M., Dumont, A.S., Szmitko, P., Li, R.K., Mickle, D.A.G. & Verma, S. 2003. C-reactive protein upregulates angiotensin type 1 receptors in vascular smooth muscle. *Circulation,* 107(13):1783-1790.

Wang, Q., Hunt, S.C., Xu, Q., Chen, Y.E., Province, M.A., Eckfeldt, J.H., Pankow, J.S. & Song, Q. 2006. Association study of CRP gene polymorphisms with serum CRP level and cardiovascular risk in the NHLBI family heart study. *American journal of physiology-heart and circulatory physiology,* 291(6):2752-2757.

Wensley, F., Gao, P., Burgess, S., Kaptoge, S., Di Angelantonio, E., Shah, T., Engert, J.C., Clarke, R., Davey-Smith, G. & Nordestgaard, B.G. 2011. Association between C-reactive protein and coronary heart disease: Mendelian randomisation analysis based on individual participant data. *British medical journal,* 342:d548.

Whitehead, A.S., Bruns, G., Markham, A.F., Colten, H.R. & Woods, D.E. 1983. Isolation of human C-reactive protein complementary DNA and localization of the gene to chromosome 1. *Science,* 221(4605):69-71.

Willerson, J.T. & Ridker, P.M. 2004. Inflammation as a cardiovascular risk factor. *Circulation,* 109(21 suppl. 1):2-10.

Williams, K.J. & Tabas, I. 1998. The response-to-retention hypothesis of atherogenesis reinforced. *Current opinion in lipidology,* 9(5):471-474.

Wolmarans, P. & Oosthuizen, W. 2001. Eat fats sparingly - implications for health and disease. *South African journal of clinical nutrition,* 14(3):48-54.

World Health Organization. 2002. The world health report 2002 - Reducing Risks, Promoting Healthy Life.

Xing, D., Hage, F.G., Chen, Y.F., McCrory, M.A., Feng, W., Skibinski, G.A., Majid-Hassan, E., Oparil, S. & Szalai, A.J. 2008. Exaggerated neointima formation in human C-reactive protein transgenic mice is IgG Fc receptor type I (Fc gamma RI)-dependent. *American journal of pathology,* 172(1):22-30.

Yach, D., Hawkes, C., Gould, C.L. & Hofman, K.J. 2004. The global burden of chronic diseases. *Journal of the American medical association,* 291(21):2616-2622.

Yasojima, K., Schwab, C., McGeer, E.G. & McGeer, P.L. 2001. Generation of C-reactive protein and complement components in atherosclerotic plaques. *American journal of pathology,* 158(3):1039-1051.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T.L. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics,* 13(1):134.

Young, D. S. 2000. Effects of Drugs on Clinical Laboratory Tests. 5th ed. Washington, D.C: AACC Press.

Yudkin, J.S., Stehouwer, C.D.A., Emeis, J.J. & Coppack, S.W. 1999. C-reactive protein in healthy subjects: Associations with obesity, insulin resistance, and endothelial dysfunction: A potential role for cytokines originating from adipose tissue. *Arteriosclerosis, thrombosis, and vascular biology,* 19(4):972-978.

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.

Zhang, D., Sun, M., Samols, D. & Kushner, I. 1996. STAT3 participates in transcriptional activation of the C-reactive protein gene by interleukin-6. *Journal of biological chemistry,* 271(16):9503-9509.

Zheng, W., McLerran, D.F., Rolland, B., Zhang, X., Inoue, M., Matsuo, K., He, J., Gupta, P.C., Ramadas,K., Tsugane, S., Irie, F., Tamakoshi. A., Gao, Y.T., Wang, R., Shu, X.O., Tsuji. I,. Kuriyama, S., Tanaka, H., Satoh, H., Chen, C.J., Yuan, J.M., Yoo, K.Y., Ahsan, H., Pan, W.H., Gu, D., Pednekar, M.S., Sauvaget, C., Sasazuki, S., Sairenchi, T., Yang, G., Xiang, Y.B., Nagai, M., Suzuki, T., Nishino, Y., You, S.L., Koh, W.P., Park, S.K., Chen, Y., Shen, C.Y., Thornquist, M., Feng, Z., Kang, D., Boffetta, P. & Potter, J.D. 2011. Association between body-mass index and risk of death in more than 1 million Asians. *New England journal of medicine*, 364 (8), 719-729.

Internet references:

Statistical release (P0302). Mid-year population estimates 2011**:**
(http://www.statssa.gov.za/publications/P0302/P03022011.pdf
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=3093058
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=3093062
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1800947
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1130864
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1205
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1417938
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=2808630
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1341665
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=3093068
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=2794520
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=7553007
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=2027471
http://www.ncbi.nlm.nih.gov/projects/SNP/

ADDENDUM

Addenda A refers to the additional statistical analyses which were conducted (as mentioned in Section 3.8), excluding individuals with a CRP concentration above 10 $mg.L^{-1}$. The significant differences for the variables investigated remained significant even after those individuals with CRP concentration of above 10 $mg.L^{-1}$ were excluded.

Addenda A

Table A1: Numerical baseline characteristics in the whole group, as well as between the rural/urban and men/women groups in the PURE study population, excluding those individuals with a CRP concentration of above 10 mg.L^{-1}

Variables	Whole group mean ± SD	Rural mean ± SD	Urban mean ± SD	p-value	Men mean ± SD	Women mean ± SD	p-value
Number of individuals	1,220	608	612	--	476	744	--
CRP (mg.L^{-1})	2.88 ± 2.65	2.94 ± 2.58	2.83 ± 2.58	0.48	2.58 ± 2.57	3.07 ± 2.39	0.01
LDL-c (mmol.L^{-1})	2.92 ± 1.23	2.94 ± 1.13	2.90 ± 1.12	0.55	2.74 ± 1.12	3.03 ± 1.12	<0.01
HDL-c (mmol.L^{-1})	1.64 ± 0.63	1.57 ± 0.63	1.56 ± 0.64	0.87	1.54 ± 0.64	1.54 ± 0.62	0.06
TC (mmol.L^{-1})	5.04 ± 1.34	5.03 ± 1.32	5.05 ± 1.36	0.74	4.88 ± 1.31	5.14 ± 1.35	<0.01
TG (mmol.L^{-1})	1.27 ± 0.79	1.19 ± 0.64	1.36 ± 0.64	<0.01	1.26 ± 0.93	1.28 ± 0.69	0.56
Age (years)	49.01 ± 10.32	48.06 ± 9.58	49.95 ± 10.93	<0.01	49.52 ± 10.20	48.67 ± 10.39	0.16
Height (cm)	161.0 ± 0.08	160.6 ± 0.09	161.5 ± 0.09	0.07	167.8 ± 0.07	157.7 ± 0.06	<0.01
Weight (kg)	61.80 ± 15.12	60.90 ± 15.15	62.79 ± 15.04	0.03	58.14 ± 11.79	63.50 ± 16.70	<0.01
BMI (kg.m^{-2})	23.94 ± 6.14	23.69 ± 6.08	24.21 ± 6.19	0.15	20.99 ± 4.00	25.82 ± 6.52	<0.01
Dietary variables:							
Total energy intake (kJ)	7,910.00 ± 3,749.70	6,518.46 ± 2,756.29	9,266.04 ± 4,079.78	<0.01	8,622.89 ± 4,038.52	7,453.14 ± 3,478.39	<0.01
Total protein intake (% of TE)	11.73 ± 2.03	10.96 ± 1.87	12.48 ± 1.89	<0.01	11.79 ± 2.14	11.69 ± 2.00	0.43
Total CHO intake (% of TE)	60.88 ± 9.30	65.73 ± 8.88	56.15 ± 6.99	<0.01	60.00 ± 9.24	61.44 ± 9.31	0.01
Total fibre intake (g)	21.08 ± 10.32	18.04 ± 7.49	24.03 ± 11.75	<0.01	22.93 ± 11.49	19.89 ± 9.31	<0.01
Total sugar intake (g)	44.33 ± 32.08	33.39 ± 26.16	54.98 ± 33.70	<0.01	45.92 ± 36.05	43.31 ± 29.22	0.17
Total fat intake (% of TE)	22.76 ± 7.52	19.26 ± 6.90	26.18 ± 6.45	<0.01	21.76 ± 7.32	23.41 ± 7.58	<0.01
MUFA (%)	5.99 ± 2.88	4.49 ± 2.39	7.45 ± 2.56	<0.01	5.77 ± 2.82	6.13 ± 2.92	0.01
PUFA (%)	6.91 ± 2.86	6.25 ± 3.06	7.56 ± 2.50	<0.01	6.50 ± 2.63	7.17 ± 2.98	<0.01
SFA (%)	5.49 ± 2.63	4.27 ± 2.41	6.67 ± 2.26	<0.01	5.22 ± 2.50	5.66 ± 2.70	0.01
SBP (mmHg)	133.52 ± 24.09	129.59 ± 22.23	137.40 ± 25.22	<0.01	136.65 ± 23.76	131.52 ± 24.11	0.01
DBP (mmHg)	87.33 ± 14.41	85.64 ± 14.12	88.99 ± 14.12	<0.01	87.13 ± 14.81	87.45 ± 14.16	0.70
Fibrinogen (g.L^{-1})	3.19 ± 1.77	3.35 ± 1.85	3.03 ± 1.17	<0.01	2.82 ± 1.51	3.43 ± 1.87	<0.01

CHO = carbohydrates; DBP = diastolic blood pressure; g = grams; g.L^{-1} = grams per litre; CRP = C-reactive protein; HDL-C = high density lipoprotein cholesterol; kJ = kilojoules; LDL-C = low density lipoprotein cholesterol; mmol.L^{-1} = millimole per litre; mmHG = millimetre of mercury; MUFA = monounsaturated fatty acids; PUFA = polyunsaturated fatty acids; SBP = systolic blood pressure; SFA = saturated fatty acids; TC = total cholesterol; TE = total energy; TG = triglycerides; IU.L^{-1} = international units per litre; U.L^{-1} = units per litre; a p-value of <0.05 is considered significant and is indicated in red text in the table

Table A2: Mean CRP concentrations for categorical baseline characteristics for the total group and by gender and location groups, excluding those individuals with CRP concentration of above 10 mg.L^{-1}

Categorical variables	**Whole group** CRP mean (mg.L^{-1}) ± SD (n)	**p-value**	**Rural** CRP mean (mg.L^{-1}) ± SD (n)	**Urban** CRP mean (mg.L^{-1}) ± SD (n)	**p-value**	**Men** CRP mean (mg.L^{-1}) ± SD (n)	**Women** CRP mean (mg.L^{-1}) ± SD (n)	**p-value**
BMI:								
18.5 (Underweight)	2.30 ± 2.46 (201)	<0.01	2.89 ± 2.65 (112)	2.20 ± 2.21 (89)	0.95	2.36 ± 2.38 (115)	2.23 ± 2.57 (86)	0.72
18.5-24.9 (Normal weight)	2.39 ± 2.47 (557)		2.44 ± 2.49 (299)	2.34 ± 2.45 (258)		2.41 ± 2.56 (274)	2.37 ± 2.39 (283)	
25-29.9 (Overweight)	3.15 ± 2.55 (206)		3.23 ± 2.60 (99)	3.07 ± 2.51 (107)		3.59 ± 2.64 (45)	3.03 ± 2.51 (161)	
≥ 30 (Obese)	4.69 ± 2.69 (193)		4.91 ± 2.69 (95)	4.52 ± 2.66 (97)		6.20 ± 4.41 (19)	4.72 ± 2.68 (176)	
HIV status:								
Positive	2.48 ± 2.29 (194)	0.06	2.66 ± 2.48 (55)	2.11 ± 2.25 (61)	0.47	1.95 ± 1.89 (43)	2.61 ± 2.59 (73)	0.52
Negative	2.96 ± 2.70 (1,023)		2.97 ± 2.74 (551)	2.91 ± 2.61 (550)		2.65 ± 2.62 (432)	3.13 ± 2.69 (669)	
Smoking:								
Previously used tobacco products	3.06 ± 2.71 (41)	0.35	2.87± 2.66 (16)	3.18 ± 2.78 (25)	0.70	2.53 ± 2.29 (26)	3.98 ± 3.19 (15)	0.39
Currently use tobacco products	2.75 ± 2.59 (641)		2.74 ± 2.55 (311)	2.77 ± 2.63 (329)		2.59 ± 2.57 (280)	2.88 ± 2.60 (360)	
Never used tobacco products	3.03 ± 2.72 (531)		3.16 ± 2.88 (279)	2.89 ± 2.52 (252)		2.58 ± 2.61 (166)	3.24 ± 2.74 (365)	

BMI = body mass index; CRP = c-reactive protein; HIV = human immunodeficiency virus; n = number of individuals; SD = standard deviation; BMI categories were adapted from the WHO (2002) recommendations; a p-value of < 0.05 is considered as significant and is indicated in red.